

# **Inferencia estadística**

Curso intermedio



# Inferencia estadística

Curso intermedio

Luz Mery González García y  
José Alfredo Jiménez Moscoso



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

Bogotá, D. C., Colombia, 2024



Catalogación en la publicación Universidad Nacional de Colombia

González García, Luz Mery, 1975-

Inferencia estadística : curso intermedio / Luz Mery González García y José Alfredo Jiménez Moscoso. – Primera edición. – Bogotá : Universidad Nacional de Colombia. Facultad de Ciencias : Coordinación de publicaciones Facultad de Ciencias, 2024.

1 recurso en línea (xxx páginas) : ilustraciones (algunas a color), diagramas. – (Colección Notas de clase)

Incluye referencias bibliográficas e índice analítico  
ISBN 978-958-505-768-5 (digital)

1. Estadística matemática – Problemas, ejercicios, etc. 2. Estimación estadística – Metodología – Ejercicios y problemas 3. Probabilidades 4. Distribución (Teoría de probabilidades) 5. Teoría bayesiana de decisiones estadísticas – Problemas, ejercicios, etc. 6. Teoremas límite (Teoría de probabilidades) 7. Convergencia 8. Intervalos de confianza 9. Contrastación de hipótesis (Estadística) 10. Prueba de hipótesis estadística 11. Prueba de bondad de ajuste 12. Muestreo (Estadística) 13. Procesos puntuales 14. Mínimos cuadrados 15. Procesos estocásticos 16. Teoría de la estimación 17. Estimación de parámetros 18. Correlación (Estadística) 19. Estadística matemática – Procesamiento de datos I. Jiménez Moscoso, José Alfredo, 1973-, autor II. Título III. Serie

CDD-23 519.54 / 2024

© Universidad Nacional de Colombia

Facultad de Ciencias

© Luz Mery González García

© José Alfredo Jiménez Moscoso

Primera edición, 2024

ISBN (digital): 978-958-505-768-5

**Edición**

Daniela Guerrero Acosta

Coordinación de publicaciones - Facultad de Ciencias

coorpub\_fcbog@unal.edu.co

**Corrección de estilo:**

**Diseño de la colección**

**Maqueta **

Camilo Cubides

Prohibida la reproducción total o parcial por cualquier medio sin la autorización escrita del titular de los derechos patrimoniales

Impreso y hecho en Bogotá, D. C., Colombia



A mis hijos, esposo y padres, que hacen que cada día sea especial.

Luz Mery

Con amor y cariño, este libro es dedicado a mi esposa, mi hija, mi madre y a  
la memoria de mi padre.

José Alfredo



# Contenido

<b>Lista de figuras</b>	<b>IX</b>
<b>Lista de tablas</b>	<b>XIII</b>
<b>Lista de siglas</b>	<b>XVII</b>
<b>Prefacio</b>	<b>XIX</b>

## Capítulo uno

<b>Variables aleatorias y distribuciones univariadas</b>	<b>1</b>
1.1. Conceptos básicos de espacio muestral	3
1.2. Concepto de probabilidad	8
1.3. Probabilidad condicional	17
1.4. Regla de multiplicación	18
1.5. Independencia	20
1.6. Regla de Bayes	25
1.7. Variable aleatoria	28
1.8. Distribuciones de probabilidad	35
1.8.1. Algunas distribuciones discretas	35
1.8.2. Algunas distribuciones continuas	45
1.9. Aproximaciones de distribuciones de probabilidad	66
1.10. Desigualdad de Chebyshev	69
1.11. Funciones generadoras	72
1.11.1. Función generadora de probabilidad	72
1.11.2. Función generadora de momentos	75
1.11.3. Función característica	84
1.12. Ejercicios	91

## Capítulo dos

<b>Vectores aleatorios y distribuciones multivariadas</b>	<b>99</b>
2.1. Distribución de probabilidad conjunta	102
2.2. Función de distribución conjunta	105
2.3. Valores esperados y momentos	106
2.4. Valor esperado condicional	107

2.5. Varianza condicional . . . . .	110
2.6. Correlación entre variables . . . . .	111
2.7. Distribución de suma de variables aleatorias independientes . .	115
2.8. Transformaciones de variables aleatorias . . . . .	118
2.9. Normal bivariada . . . . .	120
2.9.1. Distribución normal bivariada estándar . . . . .	120
2.9.2. Distribución normal bivariada . . . . .	120
2.10. Familia exponencial lineal . . . . .	125
2.10.1. Familia exponencial lineal (uniparamétrica) . . . . .	125
2.10.2. Familia exponencial de un parámetro en forma natural . .	126
2.11. Probabilidad y estadística multivariada . . . . .	127
2.11.1. Función de distribución acumulada conjunta . . . . .	128
2.12. Valor esperado y matriz de covarianzas . . . . .	129
2.13. Funciones generadoras multidimensionales . . . . .	131
2.13.1. Función generadora de momentos multidimensional . . . .	131
2.13.2. Función característica multidimensional . . . . .	134
2.14. Distribución de suma de variables aleatorias independientes . .	136
2.15. Transformaciones de dos o más variables aleatorias . . . . .	138
2.16. Normal multivariada . . . . .	145
2.17. Ejercicios . . . . .	147

## Capítulo tres

<b>Propiedades de una muestra aleatoria</b>	<b>149</b>
3.1. Conceptos básicos de muestra aleatoria . . . . .	151
3.2. Estadísticos de orden . . . . .	153
3.3. Órdenes de magnitud . . . . .	163
3.3.1. Órdenes de magnitud de sucesiones de números reales y vectores . . . . .	163
3.3.2. Órdenes de magnitud de funciones . . . . .	169
3.3.3. Órdenes de magnitud en probabilidad de sucesiones estocásticas . . . . .	171
3.4. Convergencia estocástica . . . . .	176
3.4.1. Tipos de convergencia estocástica . . . . .	176
3.4.2. Convergencia en distribución o convergencia en ley . . . .	182
3.4.3. Teoremas límites . . . . .	191
3.5. Ejercicios . . . . .	201

Capítulo cuatro

<b>Métodos de estimación puntual</b>	<b>205</b>
4.1. Principio de sustitución	207
4.1.1. Estimación por sustitución de frecuencias	207
4.1.2. Estimación por método de momentos	209
4.2. Estimación por máxima verosimilitud	213
4.3. Estimación bayesiana	227
4.3.1. Distribuciones <i>a priori</i> y <i>a posteriori</i>	227
4.3.2. Estimación puntual bayesiana	234
4.4. Estimación por el método de mínimos cuadrados	236
4.4.1. Mínimos cuadrados ordinarios	236
4.4.2. Mínimos cuadrados ponderados	246
4.5. Conceptos básicos para estimación minimax	250
4.6. Ejercicios	257

Capítulo cinco

<b>Propiedades de los estimadores</b>	<b>261</b>
5.1. Suficiencia	263
5.2. Estadística suficiente minimal	269
5.3. Completez	272
5.4. Error cuadrático medio e insesgamiento	274
5.5. Estimador insesgado de mínima varianza (UMVUE)	277
5.6. Cota inferior de Cramér-Rao	282
5.7. Consistencia	293
5.8. Eficiencia	297
5.9. Invarianza para estimadores de máxima verosimilitud	298
5.10. Familia exponencial	300
5.10.1. Estimación de máxima verosimilitud en la familia exponencial	302
5.10.2. Familia exponencial de un parámetro en forma natural	304
5.10.3. Familia exponencial de $k$ parámetros	308
5.10.4. Estimación de máxima verosimilitud en la familia exponencial de $k$ parámetros	309
5.11. Ejercicios	311

Capítulo seis

<b>Intervalos de confianza</b>	<b>315</b>
6.1. Conceptos básicos de intervalos de confianza	317
6.2. Intervalos de confianza basados en una única estadística	322

6.3. Intervalos de confianza basados en la FVR .....	333
6.4. Método pivotal .....	338
6.5. Algunos casos particulares .....	347
6.5.1. Intervalo de confianza para una proporción .....	347
6.5.2. Intervalo de confianza para diferencia de medias de dos poblaciones independientes .....	349
6.5.3. Intervalo de confianza para diferencia de medias de dos poblaciones no independientes .....	360
6.5.4. Intervalo de confianza para diferencia de proporciones de dos poblaciones independientes .....	362
6.6. Estimación bayesiana por intervalo .....	363
6.7. Tamaño de muestra simple bajo normalidad .....	371
6.7.1. Tamaño de muestra para estimar la media con varianza conocida .....	371
6.7.2. Tamaño de muestra para la estimación de una proporción	373
6.8. Ejercicios .....	374

## Capítulo siete

<b>Prueba de hipótesis</b> .....	<b>377</b>
7.1. Elementos básicos de pruebas de hipótesis .....	379
7.2. Evaluación de tests .....	380
7.3. Test uniformemente más potente .....	383
7.3.1. Test de razón de verosimilitud .....	383
7.3.2. Lema de Neyman-Pearson .....	391
7.4. Algunos casos particulares .....	407
7.4.1. Prueba de hipótesis sobre una varianza cuando $\mu$ es una constante desconocida .....	411
7.4.2. Prueba de hipótesis sobre diferencia de medias de dos poblaciones independientes .....	414
7.4.3. Pruebas de hipótesis para la igualdad de varianzas .....	421
7.4.4. Prueba de hipótesis para diferencia de medias de dos poblaciones no independientes .....	422
7.4.5. Pruebas de hipótesis para diferencia entre dos propor- ciones poblacionales, muestras independientes .....	425
7.5. Bondad de ajuste .....	427
7.5.1. Test de Pearson para bondad de ajuste .....	427
7.5.2. Test de Kolmogorov para bondad de ajuste .....	435
7.6. Ejercicios .....	441

Capítulo ocho

<b>Comportamiento asintótico de estadísticas de prueba</b>	<b>443</b>
8.1. Función de estimación. Caso uniparamétrico	445
8.1.1. Propiedades de una función de estimación	446
8.2. Función de estimación. Caso multiparamétrico y vectorial	451
8.2.1. Propiedades de una función de estimación multivariada	453
8.3. Estadísticas de prueba	459
8.4. Ejercicios	467

Apéndice A

<b>Series de Taylor</b>	<b>469</b>
A.1. Expansiones de Taylor	471
A.1.1. Fórmula de Taylor con resto	474
A.2. Ejercicios	479

Apéndice B

<b>Tópicos especiales de matemáticas</b>	<b>481</b>
B.1. Tópicos de álgebra matricial	483
B.1.1. Transformación ortogonal	484
B.1.2. Forma cuadrática	487
B.1.3. Inversa generalizada	487
B.2. Tópicos de cálculo	490
B.2.1. Máximo o mínimo local	492
B.2.2. Desigualdad de Jensen	494
B.3. Distribuciones de algunas formas cuadráticas	496
B.4. Otros tópicos	500
B.4.1. Condiciones de regularidad	500
B.4.2. Ejemplo numérico de máxima verosimilitud	501
B.5. Ejercicios	504

Apéndice C

<b>Tablas</b>	<b>505</b>
<b>Referencias</b>	<b>513</b>
<b>Índice analítico</b>	<b>521</b>



## Lista de figuras

1.1. fmp uniforme discreta. (a) $n = 5$ y (b) $n = 10$ .....	36
1.2. fda uniforme discreta. (a) $n = 5$ y (b) $n = 10$ .....	36
1.3. fmp binomial, $m = 10$ . (a) $p = 0,2$ y (b) $p = 0,5$ .....	38
1.4. fda binomial, $m = 10$ . (a) $p = 0,2$ y (b) $p = 0,5$ .....	38
1.5. fmp y fda Poisson. (a) $\lambda = 2,93$ y (b) $\lambda = 19,90$ .....	41
1.6. fmp y fda geométrica. (a) $p = 0,19$ y (b) $p = 0,41$ .....	43
1.7. fdp uniforme sobre $[a, b]$ . (a) $[1, 8]$ y (b) $[4, 7; 7, 8]$ .....	46
1.8. fda uniforme sobre $[a, b]$ . (a) $[1, 8]$ y (b) $[4, 7; 7, 8]$ .....	46
1.9. fdp y fda normal con $\sigma^2 = 4$ . (a) $\mu = 4$ y (b) $\mu = -3$ .....	48
1.10. fdp y fda normal con $\mu = 2$ . (a) $\sigma^2 = 2^2$ y (b) $\sigma^2 = 0.7^2$ ...	49
1.11. fdp y fda normal estándar .....	49
1.12. fdp y la fda exponencial. (a) $\lambda = 0,2$ y (b) $\lambda = 0,1$ .....	52
1.13. fdp y la fda chi-cuadrada. (a) $r = 2,5$ ( $k = 5$ ) y (b) $r = 7,5$ ( $k = 15$ ) .....	53
1.14. fdp y fda Gamma. (a) $r = 11,11$ y $\lambda = 2,22$ y (b) $r = 4$ y $\lambda = 0,8$ .....	53
1.15. fdp y la fda gamma. (a) $r = 11,11$ y $\lambda = 5,55$ y (b) $r = 277,78$ y $\lambda = 27,78$ .....	54
1.16. fmp binomial con $m = 20$ y $p = 0,5$ .....	68
2.1. Coeficiente de correlación .....	112
2.2. fdp de la normal bivariada en 3D .....	124
2.3. fdp $t$ -Student para diferentes grados de libertad .....	143
3.1. Ejemplos de sucesiones de números $a_n$ .....	164
3.2. Ejemplos de sucesiones de normas de vectores en $\mathbb{R}^3$ .....	167
4.1. Funciones de verosimilitud para la fdp uniforme en $\mathcal{F}_\theta$ , considerando diferentes valores de $x_{(n)}$ .....	223
4.2. Distribución <i>a priori</i> , verosimilitud y fdp <i>a posteriori</i> para $n = 10$ , $\sum_{i=1}^{10} x_i = 3$ , $B(1,5)$ (1. <sup>a</sup> fila), $B(6,6)$ (2. <sup>a</sup> fila) y $B(6,0,2)$ (3. <sup>a</sup> fila) .....	230

4.3.	Distribución <i>a priori</i> , verosimilitud y fdp <i>a posteriori</i> para $n = 10$ , $\sum_{i=1}^{10} x_i = 9$ , $B(1,5)$ (1. <sup>a</sup> fila), $B(6,6)$ (2. <sup>a</sup> fila) y $B(6,0,2)$ (3. <sup>a</sup> fila) .....	230
4.4.	Gráfico de dispersión para los datos del ejemplo 4.23 y línea ajustada .....	240
4.5.	Matriz de <i>scatter plot</i> para los datos del ejemplo 4.25 .....	244
4.6.	Gráfico de dispersión para los datos del ejemplo 4.25 en 3D .....	245
4.7.	Gráfico de dispersión para los datos del ejemplo 4.25 en 3D y plano ajustado .....	245
6.1.	Probabilidad de cobertura versus $\theta$ , intervalo de la forma $[Y + a, Y + b]$ , $a = 2$ , $b = 10$ , (a) $n = 20$ y (b) $n = 30$ , fdp uniforme $(0, \theta)$ .....	321
6.2.	Familia estocásticamente creciente. fda $X \sim N(\mu, 1)$ considerando diferentes valores de $\mu$ .....	323
6.3.	Familia estocásticamente decreciente. fda $X \sim Exp(\lambda)$ considerando diferentes valores de $\lambda$ .....	324
6.4.	fda versus $\mu$ , $X \sim N(\mu, 1)$ , para $x = -2$ en (a) y $x = 0$ en (b). Familia estocásticamente creciente .....	324
6.5.	fda versus $\lambda$ , $X \sim Exp(\lambda)$ , para $x = 0,5$ en (a) y $x = 1$ en (b). Familia estocásticamente decreciente .....	325
6.6.	$\lambda_L(t)$ y $\lambda_U(t)$ de una fda $T \sim Exp(\lambda)$ para $\alpha_1 = \alpha_2 = 0,10$ , con $t = 0,5$ en (a) y $t = 1$ en (b) Familia estocásticamente decreciente .....	325
6.7.	$\mu_L(t)$ y $\mu_U(t)$ de una fda $T \sim N(\mu, 1)$ para $\alpha_1 = \alpha_2 = 0,10$ , con $t = -2$ en (a) y $t = 0$ en (b). Familia estocásticamente creciente .....	326
6.8.	Región de no-rechazo para una fda $X \sim N(\mu, 1)$ para $\alpha_1 = \alpha_2 = 0,10$ .....	326
6.9.	$\theta_L(t)$ y $\theta_U(t)$ para una fda $X \sim N(\mu, 1)$ para $\alpha_1 = \alpha_2 = 0,10$ .....	327
6.10.	fda versus $p$ , $X \sim B(10, p)$ , para $x = 2$ en (a) y $x = 4,5$ en (b). Familia estocásticamente creciente .....	329
6.11.	fda versus $p$ , $X \sim G(p)$ , para $x = 2$ en (a) y $x = 6,5$ en (b). Familia estocásticamente decreciente .....	329
6.12.	fda versus $p$ , $X \sim B(10, p)$ , para $x = 2$ en (a) y $x = 4,5$ en (b). Familia estocásticamente creciente .....	330
6.13.	Región de no-rechazo para una fda binomial de parámetros $n = 10$ y $\theta_0 = p_0 = 0,75$ para $\alpha_1 = \alpha_2 = 0,10$ .....	331
7.1.	Función de potencia $\beta(\theta)$ , ejemplo 7.1 .....	381

7.2.	Función de potencia $\beta(\theta)$ .....	382
7.3.	Función de potencia $\beta(\theta)$ para $\theta_0 = 2$ , $n = 10$ y considerando dos valores para $c$ , $c = 0,10$ y $c = 0,70$ , ejemplo 2 .....	385
7.4.	$\theta$ vs. $L(\theta x)$ cuando $x_{(1)} = 1$ , $\bar{x} = 1,02$ y $n = 10$ , ejemplo 7.4	387
7.5.	$\theta$ vs. $L(\theta x)$ cuando $x_{(1)} = 1$ , $\bar{x} = 1,02$ y $n = 10$ . $\theta_0 = 0,9$ , ejemplo 7.4 .....	388
7.6.	$\theta$ vs. $L(\theta x)$ cuando $x_{(1)} = 1$ , $\bar{x} = 1,02$ y $n = 10$ . $\theta_0 = 1,2$ . Ejemplo 7.4 .....	388
7.7.	Zona de rechazo y $P$ -valor, ejemplo 7.14, ítem 1 .....	408
7.8.	Zona de rechazo y $P$ -valor, ejemplo 7.14, ítem 2 .....	410
7.9.	Zona de rechazo y $P$ -valor, ejemplo 7.16 .....	414
7.10.	Distribución empírica, dieta A, ejemplo 7.23 .....	435
7.11.	Distribución empírica (línea continua) y teórica (línea intermitente), dieta A, ejemplo 7.23 .....	437
7.12.	Gráfico Q-Q para los datos de la dieta A considerando una fdp normal .....	438
7.13.	Envelope para los datos de la dieta A considerando una fdp normal con media 274,7 y varianza 124,9 .....	440
B.1.	Rotación de ejes en $\mathbb{R}^2$ .....	485
B.2.	Ilustración de función cóncava y convexa .....	494



# Lista de tablas

1.1.	Probabilidades de caras en 3 lanzadas de una moneda.....	9
1.2.	Clasificación de los pacientes según su nivel de HIA .....	11
1.3.	N.º de compradores de VCR por marca y N.º de cabezas ...	12
1.4.	Mutaciones en una especie animal .....	16
1.5.	Tabla de contingencia problemas de obesidad vs. hipertensión	17
1.6.	Número de personas según tratamiento y cura .....	18
1.7.	Tabla de contingencia problemas auditivos vs. visuales .....	23
1.8.	Distribución de pelotas de acuerdo a color y caja .....	27
1.9.	Probabilidad para el ejemplo sobre el lanzamiento de un dado	37
1.10.	Método de integración tabular .....	62
1.11.	Corrección por continuidad .....	68
1.12.	Resumen $\mathbb{E}(X)$ , $\mathbb{V}(X)$ y $m_X(t)$ para algunas distribuciones ..	83
1.13.	Resumen $m_X(t)$ y $\psi_X(t)$ para algunas distribuciones .....	89
1.14.	Admitidos en pregrado por Facultad, 2007.....	92
2.1.	Funciones asociadas a algunas distribuciones miembros de la familia exponencial (con $T(x) = x$ ) .....	126
3.1.	Datos ilustrativos sobre altura (cm) de árboles tung luego de siete meses de plantados .....	151
3.2.	Datos ilustrativos sobre altura (cm) de árboles tung luego de siete meses de plantados, datos ordenados .....	152
3.3.	Función de frecuencia, $P(10)$ .....	155
3.4.	Muestra aleatoria de tamaño 20, ejemplo 3.6 .....	155
3.5.	Muestra aleatoria de tamaño 20, ejemplo 3.6, datos ordenados	156
3.6.	Función de frecuencia, $P(10)$ .....	157
3.7.	Muestra aleatoria de tamaño 15, fmp Poisson de pará- metro $\lambda = 10$ .....	158
3.8.	Muestra aleatoria de tamaño 15, fmp Poisson de pará- metro $\lambda = 10$ , datos ordenados .....	158
3.9.	Ejemplo de sucesiones de números $a_n$ .....	163
3.10.	Ejemplo de sucesiones de vectores en $\mathbb{R}^3$ .....	166

3.11. Sucesión $a_n$ aproximada por (a) y por (b) . . . . .	169
3.12. Órdenes de magnitud de números $a_{nj}$ . . . . .	178
4.1. EMM de los parámetros de algunas distribuciones . . . . .	212
4.2. EMV de los parámetros de algunas distribuciones . . . . .	225
4.3. Algunas distribuciones conjugadas . . . . .	234
4.4. Ejemplo ilustrativo sobre un modelo de regresión. N.º de reclamaciones y valor pagado para diferentes ramos . . . . .	239
4.5. Ejemplo ilustrativo sobre un modelo de regresión múltiple. N.º de reclamaciones, valor pagado para diferentes ramos y N.º de expuestos año póliza . . . . .	243
4.6. Puntajes de las 4 canciones en el festival de música . . . . .	251
4.7. Puntajes de las 3 canciones en el festival de música . . . . .	251
4.8. Ingresos (\$) del evento de acuerdo al clima . . . . .	252
4.9. Ingresos (\$) del evento según el clima si se compra una póliza contra la lluvia . . . . .	252
4.10. Lo que dejarían de ganar . . . . .	252
4.11. Ingresos y pérdidas (entre paréntesis) para las 4 opciones . . . . .	253
4.12. Ingreso y monto del seguro de vida del jefe de familia . . . . .	259
4.13. Datos sobre el valor unitario, tasa de producción y costos . . . . .	260
6.1. Pesos en gramos de ratas bajo dos dietas . . . . .	343
6.2. Pesos promedios del fruto de pitahaya en gramos, dos cosechas del año 1989 . . . . .	353
6.3. Ingestión de sodio en la presión arterial sistólica (mmHg) . . . . .	361
6.4. Contenido de cobre en tomates . . . . .	376
7.1. Casos posibles para probar una hipótesis estadística . . . . .	380
7.2. Decisiones en la comprobación de hipótesis . . . . .	382
7.3. Datos codificados, ejemplo 7.19 . . . . .	423
7.4. Tablas de frecuencias esperadas y observadas . . . . .	428
7.5. Distribución del número de ventas de vestidos por porcentaje de disminución en su precio de venta . . . . .	431
7.6. Distribución empírica, dieta A, ejemplo 7.23 . . . . .	435
7.7. fda empírica, $S_{10}(x)$ , y fda teórica, $\Phi(x)$ , dieta A, ejemplo 7.23 . . . . .	436
7.8. Datos dieta A ordenados, ejemplo 6.17 . . . . .	438
7.9. Organización de los residuales estandarizados de las $m$ simulaciones . . . . .	439
7.10. N.º de nacimientos por mes en un año concreto en el hospital A . . . . .	442

7.11. Porcentaje de base seca, ejercicio 6 .....	442
A.1. Coeficientes de la serie de Taylor .....	473
A.2. Fórmulas del resto .....	479
B.1. Número de accidentes de los operarios de maquinaria .....	501
B.2. Iteraciones de Newton Raphson para encontrar $r$ .....	502
B.3. Comparación de distribuciones ajustadas .....	503
C.1. Valores de la fda normal estándar .....	507
C.2. Valores de cuantiles de la fda $t$ -Student .....	509
C.3. Valores de cuantiles de la fda chi-cuadrada .....	511



# Lista de siglas

- BN** binomial negativa
- ECM** error cuadrático medio
- EMM** estimador por el método de los momentos
- EMV** estimador vía máxima verosimilitud
- fd<sub>a</sub>** función de distribución acumulada
- fd<sub>p</sub>** función de densidad de probabilidad
- fg<sub>m</sub>** función generadora de momentos
- fg<sub>p</sub>** función generadora de probabilidad
- fmp** función de masa de probabilidad
- FVR** función de verosimilitud relativa
- iid** independientes e idénticamente distribuidas
- LRT** test de razón de verosimilitud
- MCO** mínimos cuadrados ordinarios
- MCP** mínimos cuadrados ponderados
- RVM** razón de verosimilitud monótona
- UMP** uniformemente más potente
- UMVUE** estimador insesgado de mínima varianza



# Prefacio

La inferencia estadística es, en la actualidad, un elemento esencial de los conocimientos estadísticos necesarios para estadísticos y científicos de datos. Además, la comprensión de las temáticas abordadas en inferencia (métodos de estimación, intervalos de confianza y pruebas de hipótesis) son apropiadas para estadísticos, economistas, físicos y profesionales que necesitan realizar pronósticos.

Pese a las diversas utilidades de inferencia estadística, la mayoría de los textos de inferencia no introducen algunos temas; por eso, en muchos casos, no se encuentra un libro que se ajuste a los requerimientos y necesidades de ciertas materias. Este material está basado en los cursos de inferencia estadística y estadística matemática de los programas académicos en estadística de la Universidad Nacional, sede Bogotá; y se redactó usando diferentes textos, se resaltan, principalmente, los libros referenciados en la bibliografía: [12], [20], [48], [65], [67], [77] y [82], entre otros.

Este texto sirve de ayuda para aquellos estudiantes que toman diversas asignaturas en las cuales deben tener o les serían útiles las metodologías de inferencia. Aunque es recomendable tomar un curso de probabilidad antes de comenzar un curso de inferencia, este material le permitirá al lector adquirir la práctica necesaria en el manejo de métodos de estimación y pruebas de hipótesis.

El objetivo principal de este material es el de capacitar al lector para que adquiera la habilidad de usar herramientas de estadística inferencial en diferentes ámbitos, proporcionándole conceptos como la estimación puntual y por intervalo, para que pueda realizar pruebas de hipótesis de una manera óptima; durante su desarrollo, se plantean ejemplos y ejercicios relacionados con la teoría. Algunos ejemplos se pueden replicar siguiendo las instrucciones del paquete estadístico R (se empleó [73] versión 4.3.1). Se incluyeron dos capítulos a manera de recopilación de conceptos, procedimientos y resultados básicos que, por lo general, forman parte del curso de probabilidad, con el propósito de que el material sea autocontenido. Si el lector tiene amplios conocimientos de probabilidad, puede pasar de inmediato al capítulo 3.

Este material está escrito de forma secuencial, pues los contenidos previos son importantes para tener una mejor comprensión del desarrollo de cada sección posterior, lo cual ayudará al lector a alcanzar su principal objetivo. Asimismo, proporciona un medio individual para estudiar el tema expuesto y es muy práctico como texto autodidáctico. Además, permitirá que el lector avance a su propio ritmo. Por este motivo, pueden usarlo los estudiantes con diferentes aptitudes, conocimientos y velocidades de lectura.

Agradecemos la colaboración de la Facultad de Ciencias, que por medio de su oficina de publicaciones nos permitió la divulgación de este material. También queremos dar las gracias tanto a los colegas que evaluaron este manuscrito, como a nuestros estudiantes del curso de inferencia de la carrera de Estadística, por sus sugerencias y comentarios, los cuales fueron muy útiles en la redacción de este libro. Además, queremos agradecer al equipo editorial de la Universidad Nacional de Colombia, sede Bogotá, por la corrección de estilo.

Capítulo  
*uno*  
**Variables  
aleatorias y  
distribuciones  
univariadas**

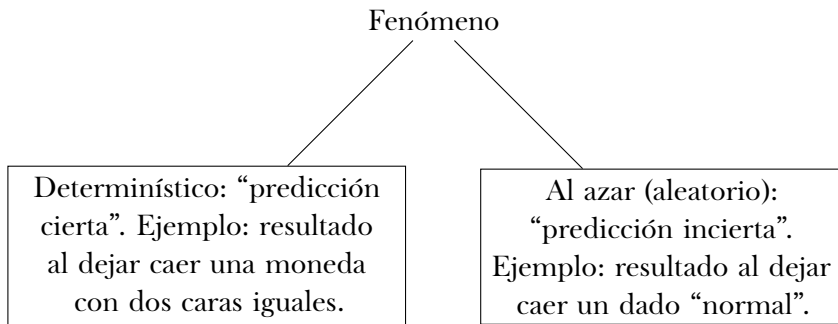


Este capítulo es una recopilación de conceptos, procedimientos y resultados básicos que, por lo general, forman parte del curso de probabilidad. Por consiguiente, una gran parte de estos resultados aparecen sin prueba; además, en algunos casos se consideran temas que el lector debe manejar y que por su importancia son retomados posteriormente.

El propósito fundamental de este capítulo es servir como prerrequisito para los siguientes capítulos y, como ya se mencionó, no se profundizará en los temas considerados aquí.

## 1.1. Conceptos básicos de espacio muestral

1. *Fenómeno*: un suceso puede ser



2. *Experimento aleatorio*: teniendo en cuenta que *alea* es una palabra del latín que significa "dado o suerte" o "la suerte está echada", experimento aleatorio es aquel proceso que tiene más de un posible resultado y cumple con las siguientes condiciones:
  - a) Todos los posibles resultados son conocidos antes de realizar el experimento.
  - b) En cualquier ejecución del experimento el resultado no se puede conocer por anticipado.
  - c) El experimento se debe poder repetir bajo idénticas condiciones.

**Ejemplo 1.1.** (Aleatorio). *Seleccionar un estudiante aleatoriamente de la Universidad Nacional de Colombia, sede Bogotá, y preguntarle su edad.*

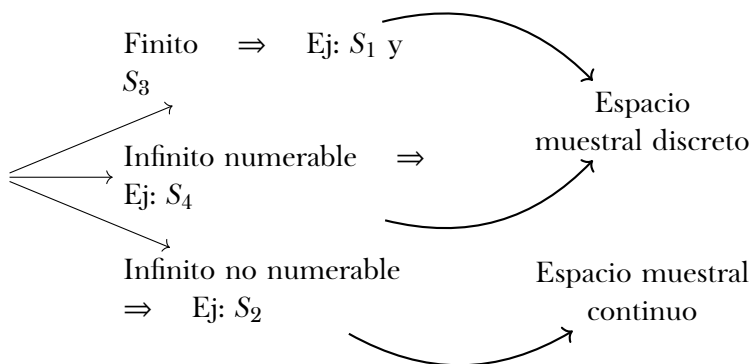
**Ejemplo 1.2.** (No aleatorio). *En un colegio femenino preguntar el sexo a los estudiantes de un salón o grupo.*

3. *Espacio muestral*: es el conjunto de todos los posibles resultados de un experimento estadístico, se denota por el símbolo  $S$  o  $\Omega$  (omega).

**Ejemplo 1.3.** Algunos ejemplos de espacio muestral son:

- Clasificar un pollo al nacer por su sexo.  
 $S_1 = \{\text{macho, hembra}\}$
- Anotar el peso de un pollo al nacer en gramos.  
 $S_2 = \{x : x > 0\}$
- Lanzar un dado “normal” una vez y observar su resultado.  
 $S_3 = \{1, 2, 3, 4, 5, 6\}$
- Número de accidentes automovilísticos un sábado dado en la vía Bogotá-Tunja.  
 $S_4 = \{0, 1, 2, 3, 4, 5, 6, \dots\}$

#### 4. Tipos de espacio muestral



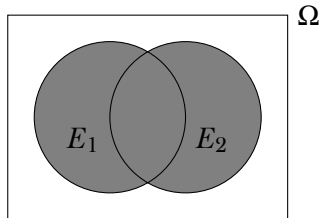
Sea  $\omega$  el resultado de un *experimento aleatorio*  $E$ , a  $\omega$  se le llama un *punto muestral*. Los puntos muestrales pueden ser números, vectores, símbolos, etc., los cuales están determinados por el experimento considerado. Luego, el número de puntos muestrales puede ser finito, enumerable o no enumerable (ver [11, p. 1]).

A la totalidad de resultados posibles del experimento aleatorio  $E$  se denota por  $\Omega$  y es llamado el *espacio muestral*. Estos conjuntos pueden ser expresados por comprensión o por extensión. Una colección de resultados del experimento aleatorio  $E$  en la cual se este interesado se llama un *evento*. Entonces, un *evento* es un subconjunto de  $\Omega$ .

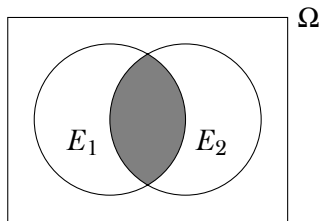
Para ver gráficamente la relación entre varios eventos, sus uniones e intersecciones, se usan *Diagramas de Venn*. Este diagrama se denomina así en honor de John Venn (1834-1923). En él se representa el conjunto de posibilidades para un experimento mediante un rectángulo, por ejemplo.

Sean dos eventos  $E_1$  y  $E_2$  en un espacio muestral  $\Omega$  [19, p.33]:

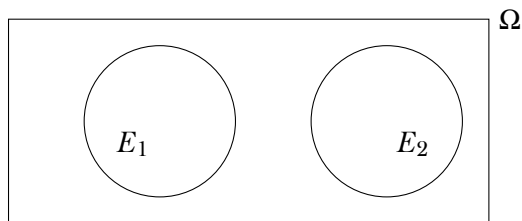
- El evento formado por todos los posibles resultados en  $E_1$  o en  $E_2$  o en ambos recibe el nombre de la unión de  $E_1$  y  $E_2$ , y se denota por  $E_1 \cup E_2$ .



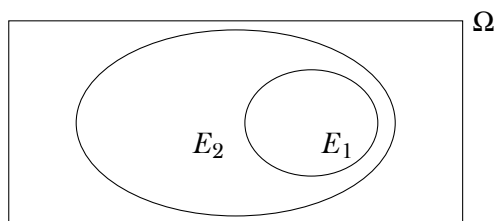
- El evento formado por los posibles resultados comunes tanto a  $E_1$  como a  $E_2$  recibe el nombre de la intersección de  $E_1$  y  $E_2$  y se denota por  $E_1 \cap E_2$ .



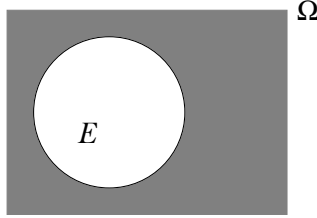
- Se dice que los eventos  $E_1$  y  $E_2$  son *mutuamente excluyentes* o *disyuntos* si no tienen resultados en común; o sea,  $E_1 \cap E_2 = \emptyset \equiv$  evento vacío.



- Si cualquier resultado de  $E_1$  también es resultado de  $E_2$ , se dice que el evento  $E_1$  está contenido en  $E_2$  y se denota  $E_1 \subset E_2$ .



- El complemento de un evento  $E$  con respecto al espacio muestral  $\Omega$ , es aquel que contiene a todos los resultados de  $\Omega$  que no se encuentran en  $E$  y se denota por  $E^c$  ó  $E'$ .



**Ejemplo 1.4.** Suponga que una moneda “legal” se lanza 3 veces. Determine el espacio muestral.

*Solución.* Los posibles resultados son:

$$\underbrace{CCC}_{\omega_1}, \underbrace{CCS}_{\omega_2}, \underbrace{CSC}_{\omega_3}, \underbrace{SCC}_{\omega_4}, \underbrace{CSS}_{\omega_5}, \underbrace{SCS}_{\omega_6}, \underbrace{SSC}_{\omega_7}, \underbrace{SSS}_{\omega_8}$$

Si  $B$  es el evento que salga exactamente una cara, entonces,

$$B = \{\omega_5, \omega_6, \omega_7\} \quad \text{y} \quad B^c = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_8\},$$

note que  $B^c$ , también, es un evento.  $\Omega$  queda conformado por todos los posibles resultados:

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\},$$

y se llama un **evento seguro**. Al evento (vacío)  $\emptyset$  se le llama **evento imposible**. ☑

**DEFINICIÓN 1.1.** ( $\sigma$ -álgebra). Sea  $\Omega \neq \emptyset$ . Una colección  $\mathcal{F}$  de subconjuntos de  $\Omega$  es una  $\sigma$ -álgebra sobre  $\Omega$ , si y solo si, (ver [16, p. 9]):

- $\Omega \in \mathcal{F}$ .
- Si  $A \in \mathcal{F}$  entonces  $A^c \in \mathcal{F}$  ( $\mathcal{F}$  es cerrado bajo complementos).
- Si  $A_1, A_2, A_3, \dots \in \mathcal{F}$  entonces  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ . Es decir, cualquier colección arbitraria de  $\mathcal{F}$  pertenece a  $\mathcal{F}$  ( $\mathcal{F}$  es cerrado bajo uniones numerables).

Los elementos de  $\mathcal{F}$  se llaman eventos.

**Ejemplo 1.5.** Considere los datos del ejemplo 1.4. Determine  $\mathcal{F}_T = \mathcal{P}(\Omega)$ .

*Solución.* La **cardinalidad** de  $\Omega$ , que corresponde al número de puntos en  $\Omega$ , es 8. El número de todos los posibles subconjuntos de  $\Omega$  en este caso es finito y es:

$$\eta(\mathcal{P}(\Omega)) = 2^8 = 256$$

donde  $\eta(\cdot)$  denota la cardinalidad. Veamos:

$\emptyset$	1						
$\{\omega_1\}$	8	$\{\omega_1, \omega_2, \omega_3\}$	56	$\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$	56		
$\{\omega_2\}$		$\{\omega_1, \omega_2, \omega_4\}$		$\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_6\}$			
$\vdots$		$\vdots$		$\vdots$			
$\{\omega_8\}$		$\{\omega_6, \omega_7, \omega_8\}$		$\{\omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}$			
$\{\omega_1, \omega_2\}$	28	$\{\omega_1, \omega_2, \omega_3, \omega_4\}$	70	$\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$	28		✓
$\{\omega_1, \omega_3\}$		$\{\omega_1, \omega_2, \omega_3, \omega_5\}$		$\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_7\}$			
$\vdots$		$\vdots$		$\vdots$			
$\{\omega_7, \omega_8\}$		$\{\omega_5, \omega_6, \omega_7, \omega_8\}$		$\{\omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}$			
		$\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7\}$				8	
		$\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_8\}$					
		$\vdots$					
		$\{\omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}$					
		$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}$					1

**Ejemplo 1.6.** En el caso de  $\Omega' = \{1, 2, 3\}$ , se tiene que  $\mathcal{F}$  es la colección de  $2^3 = 8$  conjuntos

$$\begin{matrix} \{1\} & \{1, 2\} & \{1, 2, 3\} \\ \{2\} & \{1, 3\} & \emptyset \\ \{3\} & \{2, 3\} & \end{matrix}$$

**Ejemplo 1.7.** De acuerdo a [11, p. 5], sea  $\Omega = (-\infty, \infty)$  la recta real. En este caso  $\mathcal{F}$  es seleccionado para que contenga todos los conjuntos de la forma

$$[a, b], (a, b), (a, b), [a, b),$$

para todo número real  $a$  y  $b$ .

**Ejemplo 1.8.** Suponga que  $\Omega$  es la recta real,  $\mathbb{R}$ , consistiendo de todos los puntos,  $\omega$ , en la recta real,  $\mathbb{R}$ .

$$\Omega = \{\omega : -\infty < \omega < \infty\}.$$

Sea:

$$\begin{aligned} A &= (-\infty, a) = \{\omega \in \mathbb{R} : \omega < a\} \\ B &= (c, d) = \{\omega \in \mathbb{R} : c < \omega < d\}, \end{aligned}$$

entonces:

$$A^c = [a, \infty), \quad B^c = (-\infty, c] \cup [d, \infty)$$

$$A \cap B = \begin{cases} \emptyset & \text{si } a < c < d \\ (c, a) & \text{si } c < a < d \\ (c, d) & \text{si } c < d < a \end{cases}$$

$$A \cup B = \{\omega \in \mathbb{R} : \omega < a \text{ ó } c < \omega < d\}.$$

Si  $a < c$  se tiene que

$$A^c \supset B, \quad A^c \cup B = A^c, \quad A^c \cap B = B.$$

Una  $\sigma$ -álgebra que es de interés es la  $\sigma$ -álgebra de Borel, denominada así en honor de Émile Borel (1871-1956) (ver [16, p. 10]).

**DEFINICIÓN 1.2.** La menor  $\sigma$ -álgebra sobre  $\mathbb{R}$  se denomina  $\sigma$ -álgebra de Borel y se denota por:

$$\mathfrak{B}(\mathbb{R}) = \sigma\{(-\infty, a] \subseteq \mathbb{R} : a \in \mathbb{R}\}.$$

**Ejemplo 1.9.** ( $\sigma$ -álgebra de Borel en  $\mathbb{R}^n$ ). Según [16, p. 11], dados  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  y  $\mathbf{b} = (b_1, b_2, \dots, b_n)$  elementos de  $\mathbb{R}^n$  con  $\mathbf{a} \leq \mathbf{b}$ , esto es,  $a_i \leq b_i$  para todo  $i = 1, 2, \dots, n$ . La  $\sigma$ -álgebra generada por todos los intervalos de la forma:

$$(\mathbf{a}, \mathbf{b}] := \{x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n : a_i < x_i \leq b_i, i = 1, 2, \dots, n\},$$

se llama  $\sigma$ -álgebra de Borel en  $\mathbb{R}^n$ .

**DEFINICIÓN 1.3.** (Espacio medible). Sean  $\Omega \neq \emptyset$  y  $\mathcal{F}$  una  $\sigma$ -álgebra sobre  $\Omega$ . La pareja  $(\Omega, \mathcal{F})$  se llama **espacio medible** (ver [16, p. 11]).

**Nota 1.1.** Se tiene un caso especial cuando  $\Omega = \mathbb{R}$  y  $\mathcal{F} = \mathfrak{B}$ . Luego  $(\mathbb{R}, \mathfrak{B})$  es un espacio medible. De forma similar, si  $\Omega = \mathbb{R}^n$  y  $\mathcal{F} = \mathfrak{B}^n$  se tendrá que  $(\mathbb{R}^n, \mathfrak{B}^n)$  es un espacio medible.

## 1.2. Concepto de probabilidad

El objetivo ahora es asignar a cada evento  $E$  un número real no negativo que indique el “chance” que tiene  $E$  de ocurrir. Hay tres formas de calcular o estimar la probabilidad.

**DEFINICIÓN 1.4.** *Definición clásica, ley de Laplace o a priori:* este método es de uso limitado, ya que necesita las siguientes características:

- El espacio muestral es finito.
- Los resultados del espacio muestral deben ser equiprobables.

Bajo estas condiciones, si un evento  $E$  tiene  $h$  posibilidades de ocurrir entre un total de  $k$  posibilidades igualmente posibles, entonces la probabilidad de que ocurra  $E$  se denota por:

$$P(E) = \frac{\eta(E)}{\eta(\Omega)} = \frac{h}{k}.$$

Y la probabilidad de que no ocurra  $E$  le corresponde:

$$P(E^c) = \frac{\eta(E^c)}{\eta(\Omega)} = \frac{k-h}{k} = 1 - \frac{h}{k} = 1 - P(E).$$

Por tanto,

$$P(E) + P(E^c) = \frac{h}{k} + 1 - \frac{h}{k} = 1.$$

**Ejemplo 1.10.** Retomando el ejemplo 1.4, encuentre la probabilidad de obtener solo caras en los tres lanzamientos de una moneda.

*Solución.* Sea  $X :=$  “número de caras obtenidas”, entonces:

Tabla 1.1. Probabilidades de caras en 3 lanzadas de una moneda

$X(\omega)$	$P(X(\omega))$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

Esta tabla se construyó usando el enfoque clásico de probabilidad. ✓

**Ejemplo 1.11.** Sea  $E$  el resultado de lanzar un dado “legal” una vez y que salga 2 o 4. Existen seis formas de caer el dado,  $\{1, 2, 3, 4, 5, 6\}$ . Si el dado es bueno o legal, cada cara del dado tiene la misma oportunidad de caer. Como el evento  $E$  puede ocurrir de dos formas, entonces:

$$h = \eta(E) = 2; \quad k = \eta(\Omega) = 6; \quad \Rightarrow \quad P(E) = \frac{2}{6} = \frac{1}{3}.$$

La probabilidad de que no ocurra  $E$ , es decir, que no salga ni 2 ni 4 es:

$$P(E^c) = 1 - \frac{2}{6} = \frac{2}{3}.$$

**DEFINICIÓN 1.5. Definición empírica, frecuencial o a posteriori:** en muchas situaciones, los resultados de un experimento no son igualmente posibles. En este caso, no es correcto estimar la probabilidad usando el método clásico. En lugar de esto, se emplea la interpretación de la probabilidad como una frecuencia relativa, que se basa en que un experimento aleatorio se debe poder efectuar bajo idénticas condiciones varias veces. Cada vez que el experimento aleatorio se ejecuta se observa un resultado, el cual no se conoce por anticipado por la aleatoriedad. Así, la probabilidad de un suceso se aproxima por la frecuencia relativa de los resultados. Conforme aumenta la repetición del experimento, la frecuencia relativa de los resultados favorables se aproxima al verdadero valor de la probabilidad del evento.

Si un experimento aleatorio se repite del mismo modo  $n$  veces y  $n_B$  de los resultados son favorables a un evento  $B$ , la probabilidad de que ocurra  $B$  definida desde la frecuencia relativa será el límite de  $n_B/n$  cuando  $n$  es grande:

$$P(B) = \lim_{n \rightarrow \infty} \frac{n_B}{n}.$$

**Ejemplo 1.12.** Un investigador trabaja en un nuevo fármaco para insensibilizar a los pacientes frente a picaduras de abejas. De 200 sujetos sometidos a prueba, 180 presentaron una disminución en la gravedad de los síntomas tras sufrir una picadura después de ser sometidos al tratamiento. Es natural suponer, entonces, que la probabilidad de que ocurra lo mismo en otro paciente que reciba el mismo tratamiento es aproximadamente:

$$P(B) = \frac{n_B}{n} = \frac{180}{200} = 0,90.$$

Según este estudio, se establece que el fármaco en un 90% de los casos es efectivo para disminuir la reacción de sensibilidad a las picaduras de abejas.

**DEFINICIÓN 1.6. Definición axiomática o matemática de la probabilidad (Kolmogorov).** Sea  $(\Omega, \mathcal{F})$  un espacio medible. Una función  $P : \mathcal{F} \rightarrow \mathbb{R}$  que satisface las siguientes condiciones (ver [16, p. 14]):

- a)  $P(A) \geq 0$  para todo  $A \in \mathcal{F}$ .
- b)  $P(\Omega) = 1$ .

c) Si  $A_1, A_2, A_3, \dots$  son elementos de  $\mathcal{F}$  mutuamente excluyentes, es decir,  $A_i \cap A_j = \emptyset$  para todo  $i \neq j$ , entonces:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i), \tag{1.1}$$

se llama **medida de probabilidad** sobre  $(\Omega, \mathcal{F})$ . La tripla  $(\Omega, \mathcal{F}, P)$  se llama **espacio de probabilidad**.

**Nota 1.2.**

- Los subconjuntos  $A$  del espacio muestral  $\Omega$  se llaman **eventos**. El conjunto  $\{\omega\}$  se llama **evento elemental**.
- El evento  $A$  ocurre si  $\omega \in A$  se observa.
- A cada conjunto  $A \in \mathcal{F}$  se le asigna un número real no negativo  $P(A)$ , ( $P(A) \geq 0$ ), el cual es llamado la probabilidad del evento  $A$ .

**Ejemplo 1.13.** En un estudio retrospectivo de una nueva recombinación del bacilo para influenza, al paciente se le aplica el bacilo y se mide la respuesta, la cual se clasifica en tres niveles de anticuerpos inhibidores de la hemaglutinación (HIA) que se encuentra en la sangre después de seis semanas de aplicado. Los datos se presentan en la tabla 1.2.

Tabla 1.2. Clasificación de los pacientes según su nivel de HIA

Bajo	Moderado	Alto	Total
31	26	16	73

Fuente: adaptada de [26, p. 207]

Supóngase que se selecciona un paciente al azar y se le aplica el bacilo. ¿Cuál es la probabilidad de que su nivel de respuesta sea alto?

*Solución.* Sean los eventos:

$E_1$ : El paciente fue clasificado en nivel bajo.

$E_2$ : El paciente fue clasificado en nivel moderado.

$E_3$ : El paciente fue clasificado en nivel alto.

Luego,

$$P(E_3) = \frac{16}{73} = 0,2192,$$

es decir, al seleccionar un paciente al azar y aplicarle el bacilo, se espera con probabilidad 0,22 que su nivel de HIA sea alto. ✓

**Ejemplo 1.14.** Una tienda de video vende 3 marcas diferentes de videograbadoras (o video cassette recorder-VCR), cada una con 2 o 4 cabezas. En la tabla 1.3 se indica el número de compradores que adquirieron solo una VCR recientemente.

Tabla 1.3. N.º de compradores de VCR por marca y N.º de cabezas

Marca	Número de cabezas		Total
	Dos	Cuatro	
Sony	10	20	30
Panasonic	5	15	20
Samsung	8	13	21
Total	23	48	71

Suponga que se selecciona un comprador reciente.

- ¿Cuál es la probabilidad de que este elija una VCR de la marca Sony?
- ¿Cuál es la probabilidad de que este elija una VCR de cuatro cabezas?
- ¿Cuál es la probabilidad de que este elija una VCR de la marca Panasonic o Samsung?

*Solución.* Sean los eventos:

$E_1$ : El comprador elige una VCR de la marca Sony.

$E_2$ : El comprador elige una VCR de la marca Panasonic.

$E_3$ : El comprador elige una VCR de la marca Samsung.

$E_4$ : El comprador elige una VCR de dos cabezas.

$E_5$ : El comprador elige una VCR de cuatro cabezas.

- La probabilidad de que el comprador elija la marca Sony es

$$P(E_1) = \frac{30}{71} = 0,4225,$$

es decir, él elige la marca Sony el 42% de las veces.

- La probabilidad de que se elija una VCR de 4 cabezas es

$$P(E_5) = \frac{48}{71} = 0,6761,$$

es decir, el comprador elige una VCR de cuatro cabezas el 68% de las veces.

- c) La probabilidad de que el comprador elija la marca Panasonic o Samsung, es decir,  $P(E_2 \cup E_3)$ , pero como  $E_2 \cap E_3 = \emptyset$ , entonces:

$$P(E_2 \cup E_3) = P(E_2) + P(E_3) = \frac{20}{71} + \frac{21}{71} = 0,5775,$$

es decir, él elige la marca Panasonic o Samsung el 58 % de las veces. ☑

**Teorema 1.1. Propiedades elementales del espacio de probabilidad.**  
 Si  $P(\cdot)$  es una medida de probabilidad sobre  $(\Omega, \mathcal{F})$ . Entonces:

1.  $P(\emptyset) = 0$ .
2.  $0 \leq P(A) \leq 1$ , para todo  $A \subseteq \Omega$ .
3.  $P(A^c) = 1 - P(A)$ .
4. Si  $A, B \subseteq \Omega$  entonces:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (1.2)$$

5. Si  $A, B \subseteq \Omega$  y  $A \subseteq B$  entonces  $P(A) \leq P(B)$  (esto es,  $P$  es monótona creciente).

*Demostración.*

3. Si  $A$  y  $A^c$  son dos eventos complementarios, entonces:

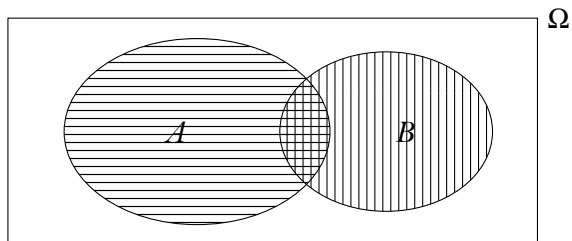
$$(A \cup A^c) = \Omega \quad \text{y} \quad (A \cap A^c) = \emptyset,$$

por lo que por el axioma 4c:

$$P(A \cup A^c) = P(A) + P(A^c) = 1. \quad (1.3)$$

Despejando  $P(A^c)$  se obtiene la prueba.

4. Sean  $A$  y  $B$  eventos del espacio muestral  $\Omega$ , entonces



Luego,  $A \cup B = (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$ . Nótese que estos eventos son excluyentes y por el axioma 4c:

$$P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(A^c \cap B).$$

Por otra parte,

$$A = (A \cap B^c) \cup (A \cap B) \quad \text{y} \quad B = (A^c \cap B) \cup (A \cap B),$$

nuevamente por el axioma 4c:

$$\begin{aligned} P(A) &= P(A \cap B^c) + P(A \cap B) & \text{y} \\ P(B) &= P(A^c \cap B) + P(A \cap B). \end{aligned}$$

Sumando se tiene que

$$\begin{aligned} P(A) + P(B) &= \underbrace{P(A \cap B^c) + P(A \cap B) + P(A^c \cap B)}_{P(A \cup B)} \\ &\quad + P(A \cap B), \end{aligned}$$

por consiguiente:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Como consecuencia, si  $A$  y  $B$  son mutuamente excluyentes, es decir, si  $A \cap B = \emptyset$ , entonces:

$$P(A \cup B) = P(A) + P(B). \quad \checkmark$$

**Ejemplo 1.15.** La probabilidad de que una industria estadounidense se ubique en Múnich ( $M$ ) es 0,7; que se localice en Bruselas ( $B$ ), de 0,4, y que se encuentre ya sea en Bruselas o en Múnich, o en ambas, de 0,8 (ver [89, p. 59]).

1. ¿Cuál es la probabilidad de que la industria se localice en ambas ciudades?
2. ¿Cuál es la probabilidad de que la industria no se localice en ninguna ciudad?

*Solución.* De la información dada se tiene que:

$$P(M) = 0,7, \quad P(B) = 0,4 \quad \text{y} \quad P(M \cup B) = 0,8.$$

1. Están preguntando  $P(M \cap B) = ?$

De la expresión (1.2) se tiene que para dos eventos  $A$  y  $B$ ,

$$P(A \cap B) = P(A) + P(B) - P(A \cup B). \quad (1.4)$$

Luego, al reemplazar se obtiene:

$$P(M \cap B) = P(M) + P(B) - P(M \cup B) = 0,7 + 0,4 - 0,8 = 0,3.$$

Por tanto, la probabilidad de que la industria esté en ambas ciudades es 0,3.

2. Están preguntando  $P((M \cup B)^c) = ?$

A partir de la expresión (1.3), despejando se obtiene:

$$\begin{aligned} P((M \cup B)^c) + P(M \cup B) &= 1 \\ P((M \cup B)^c) &= 1 - P(M \cup B) \\ P((M \cup B)^c) &= 1 - 0,8 \\ P((M \cup B)^c) &= 0,2. \end{aligned}$$

Luego, la probabilidad de que la industria no esté localizada en ninguna de las ciudades es 0,2.  $\checkmark$

**Teorema 1.2. Principio de inclusión-exclusión.** *Dados  $A_1, \dots, A_n$  eventos en un espacio probabilístico  $(\Omega, \mathcal{F}, P)$ , entonces (ver [20, p. 45]):*

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{i,j:i<j} P(A_i \cap A_j) + \\ &\sum_{i,j,k:i<j<k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} P\left(\bigcap_{i=1}^n A_i\right). \end{aligned} \quad (1.5)$$

*Demostración.* La prueba es por inducción matemática. El caso  $n = 2$  fue dado en (1.2). Si  $n = 3$  usando (1.2) se llega a:

$$\begin{aligned} P((A_1 \cup A_2) \cup A_3) &= P(A_1 \cup A_2) + P(A_3) - P((A_1 \cup A_2) \cap A_3) \\ &= P(A_1) + P(A_2) - P(A_1 \cap A_2) + P(A_3) \\ &\quad - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3). \end{aligned} \quad (1.6)$$

Supongamos que el principio (1.5) es cierto para  $n$  y probemos su validez para  $n + 1$ . Usando de nuevo (1.2) para  $n + 1$  se tiene que:

$$P\left(\bigcup_{i=1}^{n+1} A_i\right) = P\left(\bigcup_{i=1}^n A_i\right) + P(A_{n+1}) - P\left(\bigcup_{i=1}^n (A_i \cap A_{n+1})\right).$$

Empleando (1.5) en el primer y último término y agrupando términos se obtiene el resultado.  $\checkmark$

**Ejemplo 1.16. Mutaciones en una especie animal.** Se estudian en una especie animal tres mutaciones no relacionadas,  $A$ ,  $B$  y  $C$ . La probabilidad de tener la mutación  $A$  es 0,13, la  $B$  es 0,11 y la  $C$  es 0,14. Calcular las probabilidades de que:

- 1) Un animal no tenga ninguna de las mutaciones.
- 2) Un animal tenga alguna de las mutaciones.
- 3) Un animal tenga la mutación  $A$  y  $C$ , pero no la  $B$ .

*Solución.* En este caso, los eventos que representan las mutaciones  $A$ ,  $B$  y  $C$  son independientes y se tiene de la información suministrada que:

Tabla 1.4. Mutaciones en una especie animal

Mutación	$A$	$B$	$C$
Presente	0,13	0,11	0,14
Ausente	0,87	0,89	0,86

- 1) Dado que los eventos son independientes, entonces:

$$\begin{aligned} P(A^c \cap B^c \cap C^c) &= P(A^c)P(B^c)P(C^c) \\ &= 0,87 \times 0,89 \times 0,86 = 0,665898. \end{aligned}$$

- 2) Puesto que los eventos no son excluyentes, usando (1.6):

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) \\ &\quad - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \\ &= 0,13 + 0,11 + 0,14 - (0,13 \times 0,11) - (0,13 \times 0,14) \\ &\quad - (0,11 \times 0,14) + (0,13 \times 0,11 \times 0,14) = 0,334102. \end{aligned}$$

Nótese que, esta probabilidad también se pudo establecer como:

$$\begin{aligned} P(A \cup B \cup C) &= 1 - P[(A \cup B \cup C)^c] = 1 - P(A^c \cap B^c \cap C^c) \\ &= 1 - 0,665898 = 0,334102. \end{aligned}$$

- 3) Finalmente, la probabilidad de que un animal tenga la mutación  $A$  y  $C$ , pero no la  $B$  está dada por:

$$\begin{aligned} P(A \cap B^c \cap C) &= P(A)P(B^c)P(C) = P(A)(1 - P(B))P(C) \\ &= 0,13 \times 0,89 \times 0,14 = 0,016198. \quad \square \end{aligned}$$

### 1.3. Probabilidad condicional

La probabilidad de que un evento  $B$  ocurra cuando se sabe que ya ocurrió un evento  $A$  se llama probabilidad condicional y se indica por  $P(B | A)$ , que se lee como probabilidad de que “ocurra  $B$  dado que ya ocurrió  $A$ ”. Se define como:

$$P(B | A) = \frac{P(A \cap B)}{P(A)} \quad \text{si } P(A) > 0.$$

**Ejemplo 1.17.** *Se estima que entre la población total de Estados Unidos, el 55 % padece de obesidad ( $A$ ), el 20 % es hipertensa ( $B$ ) y el 60 % es obesa o hipertensa. ¿La probabilidad de que una persona obesa padezca hipertensión es igual a la probabilidad de que una persona padezca hipertensión?*

*Solución.* Sean los eventos

$A$  : “Población que tiene problemas de obesidad”.

$B$  : “Población que tiene problemas de hipertensión”.

En la tabla de contingencia 1.5 se muestra la relación entre estos dos eventos.

Tabla 1.5. Tabla de contingencia problemas de obesidad vs. hipertensión

	$B$	$B^c$	Total
$A$	0,15	0,40	0,55
$A^c$	0,05	0,40	0,45
Total	0,20	0,80	1,00

En la construcción de esta tabla se empleó la expresión (1.4), luego:

$$P(A \cap B) = 0,55 + 0,2 - 0,6 = 0,15.$$

A partir de los datos de la tabla,  $P(\text{Hipertenso}|\text{Obeso})$  queda

$$P(B | A) = \frac{P(A \cap B)}{P(A)} = \frac{0,15}{0,55} = 0,27 \neq P(B) = P(\text{Hipertenso}).$$

Luego, se concluye que las dos probabilidades no son iguales, es más, si una persona es obesa, aumenta la probabilidad de tener hipertensión. ☑

**Ejemplo 1.18.** *Suponga que personas afectadas de cierta enfermedad fueron asignadas a dos tratamientos: un grupo al que se le aplicó suero y el otro sin suero. Los datos están presentados en la tabla 1.6.*

Tabla 1.6. Número de personas según tratamiento y cura

	Curados (C)	No curados (NC)	Total
Grupo A (Usando suero)	72	23	95
Grupo B (Sin suero)	67	38	105
Total	139	61	200

1. ¿Cuál es la probabilidad de que una persona curada estuviera asignada al grupo sin suero?
2. ¿Cuál es la probabilidad de que a una persona que no se curó se le hubiera asignado al grupo sin suero?
3. Si una persona fue asignada al grupo con suero, ¿cuál es la probabilidad de que no se cure?

*Solución.* De los datos suministrados en la tabla se tiene que

$$1. P(\text{Sin suero}|\text{Curado}) = P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{\frac{67}{200}}{\frac{139}{200}} = 0,4820.$$

Así, la probabilidad de que una persona curada estuviera asignada al grupo sin suero es 0,48.

$$2. P(\text{Sin suero}|\text{No curado}) = P(B|NC) = \frac{P(B \cap NC)}{P(NC)} = \frac{\frac{38}{200}}{\frac{61}{200}} = 0,62295.$$

Entonces, la probabilidad de que una persona que no se curó estuviera asignada al grupo sin suero es 0,62.

$$3. P(\text{No curado}|\text{Con suero}) = P(NC|A) = \frac{P(NC \cap A)}{P(A)} = \frac{\frac{23}{200}}{\frac{95}{200}} = 0,2421.$$

Si una persona fue asignada al grupo con suero, la probabilidad de que no se cure es 0,24.  $\square$

## 1.4. Regla de multiplicación

Si  $A$  y  $B$  son eventos de  $\Omega$  y  $P(A) > 0$  y  $P(B) > 0$ , entonces:

$$P(A \cap B) = P(A | B)P(B) = P(B | A)P(A). \quad (1.7)$$

**Ejemplo 1.19.** [63, p. 130] La denominada prospección geobotánica se basa en el estudio de las plantas que aparecen en depósitos de minerales; por ejemplo, una pequeña planta de menta con una flor de color malva es un indicador del cobre. Suponga que en una región dada existe un 30 % de probabilidad de que el suelo tenga un alto contenido de cobre y un 23 % de probabilidad de que la menta esté presente en ese lugar. Si el contenido de cobre es alto, existe un 70 % de probabilidad de que esté presente la menta. ¿Cuál es la probabilidad de que el contenido de cobre sea alto y que la menta esté presente?

*Solución.* Sean los eventos:

- $A$ : “El suelo contiene un alto contenido de cobre”.
- $B$ : “La menta está presente en ese lugar”.

Por la información que da el ejemplo se tiene que:

$$P(A) = 0,3, \quad P(B) = 0,23 \quad \text{y} \quad P(B | A) = 0,7.$$

Están preguntando  $P(A \cap B) = ?$  Puesto que:

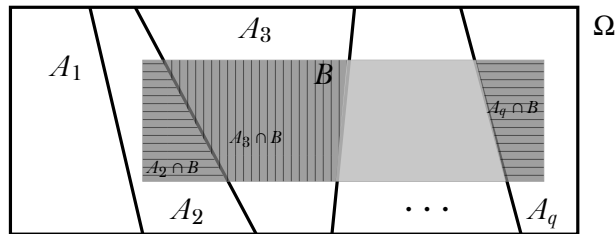
$$P(B | A) = \frac{P(A \cap B)}{P(A)} \Rightarrow P(A \cap B) = P(B | A)P(A) = 0,7 \times 0,3 = 0,21.$$

Por tanto, la probabilidad de que el contenido de cobre sea alto y la menta esté presente es 0,21. ✓

**Teorema 1.3. Teorema de probabilidad total.** Sean  $A_1, A_2, A_3, \dots, A_q$ ,  $q$  eventos mutuamente excluyentes de  $\Omega$ , tal que por lo menos uno ocurre  $\left(\sum_{k=1}^q P(A_k) = 1\right)$ . Entonces para cualquier  $B \in \mathcal{F}$  se satisface:

$$P(B) = \sum_{k=1}^q P(B | A_k)P(A_k).$$

*Demostración.* Nótese que  $A_1, A_2, A_3, \dots, A_q$  es una partición finita de  $\Omega$ .



Luego,

$$P(B) = P(B \cap \Omega) = P\left[B \cap \left(\bigcup_{k=1}^q A_k\right)\right] = P\left[\bigcup_{k=1}^q (B \cap A_k)\right].$$

Estos eventos son mutuamente excluyentes, entonces por el axioma 4c:

$$P(B) = \sum_{k=1}^q P(B \cap A_k).$$

Por la expresión (1.7) se tiene que:

$$P(B) = \sum_{k=1}^q P(A_k)P(B | A_k),$$

con lo cual queda probado el teorema.  $\square$

## 1.5. Independencia

Se dirá que dos eventos son independientes si uno puede producirse con independencia de otro, es decir, la realización o no realización de uno no tiene efecto alguno sobre la probabilidad de la realización o no del otro. Por tal motivo, sean  $A$  y  $B$  dos eventos tales que  $P(A) > 0$  y  $P(B) > 0$ . Estos eventos son independientes, si y solo si:

$$P(B | A) = P(B) \quad \text{y} \quad P(A | B) = P(A).$$

También se dirá que dos eventos  $A$  y  $B$  son independientes, si y solo si:

$$P(A \cap B) = P(A)P(B). \quad (1.8)$$

En caso contrario se dice que los eventos son dependientes.

**Ejemplo 1.20.** *En un pequeño pueblo se dispone de un carro de bomberos y una ambulancia para casos de emergencia. La probabilidad de que el primero esté disponible cuando se le necesite es 0,98 y que la ambulancia lo esté cuando se le llame, es 0,92. En el caso que resulte un herido al quemarse un edificio, encuentre la probabilidad de que tanto el carro de bomberos como la ambulancia estén disponibles.*

*Solución.* Sean los eventos:

$A$ : “El carro de bomberos esta disponible”.

$B$ : “La ambulancia esta disponible”.

Luego,

$$P(A \cap B) = P(A)P(B) = 0,98 \times 0,92 = 0,9016.$$

Así, la probabilidad de que tanto el carro de bomberos como la ambulancia estén disponibles es 0,9016.  $\square$

**Teorema 1.4.** *Si los eventos  $A$  y  $B$  son independientes, se satisface que:*

1. *La independencia es una propiedad recíproca.*
2. *Los eventos  $A$  y  $B^c$  son independientes.*

3. Los eventos  $A^c$  y  $B$  son independientes.
4. Los eventos  $A^c$  y  $B^c$  son independientes.

*Demostración.*

1. Por la definición de probabilidad condicional

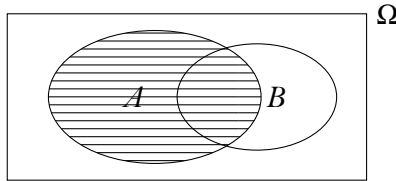
$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad \text{y} \quad P(B | A) = \frac{P(A \cap B)}{P(A)}.$$

Al despejar:

$$P(A \cap B) = P(B)P(A | B) = P(B | A)P(A), \quad (1.9)$$

como por hipótesis  $A$  es independiente de  $B$ ,  $P(A | B) = P(A)$ , sustituyendo en (1.9) se tiene:  $P(A \cap B) = P(B)P(A | B) = P(B)P(A)$ .

2. Usando el diagrama de Venn



$$A = (A \cap B^c) \cup (A \cap B).$$

Luego

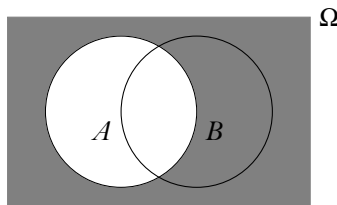
$$\begin{aligned} P(A) &= P(A \cap B^c) + \underbrace{P(A \cap B)} \\ &= P(A \cap B^c) + P(A)P(B). \end{aligned}$$

Por lo tanto, despejando:

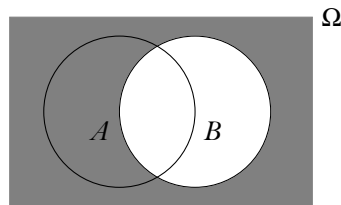
$$P(A \cap B^c) = P(A) - P(A)P(B) = P(A)[1 - P(B)] = P(A)P(B^c),$$

como se satisface (1.8), se tiene lo que se quería demostrar.

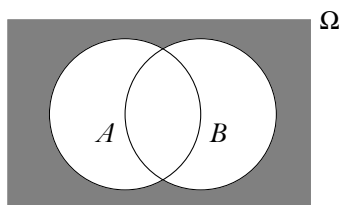
3. La prueba es análoga al ítem 2.
4. Por las leyes de De Morgan:



$A^c$



$B^c$



$(A \cup B)^c$

Luego,  $(A \cup B)^c = A^c \cap B^c$ , por lo tanto

$$P(A^c \cap B^c) = P[(A \cup B)^c] = 1 - P(A \cup B).$$

Sustituyendo la expresión (1.2) se tiene que

$$\begin{aligned} P(A^c \cap B^c) &= 1 - [P(A) + P(B) - P(A \cap B)] \\ &= (1 - P(A)) - P(B)(1 - P(A)) \\ &= P(A^c)(1 - P(B)) = P(A^c)P(B^c), \end{aligned}$$

como se satisface (1.8), se tiene lo que se quería demostrar.  $\checkmark$

**DEFINICIÓN 1.7. Independencia de eventos 2 a 2.** Una familia de eventos  $\mathcal{A} = \{A_k, 1 \leq k \leq n\} = \{A_1, \dots, A_n\}$  se dice independiente dos a dos si

$$P(A_i \cap A_j) = P(A_i)P(A_j) \quad i \neq j, \quad 1 \leq i, j \leq n.$$

La independencia dos a dos no implica la independencia mutua de la familia.

**DEFINICIÓN 1.8. Colección mutuamente independiente.** Una familia de eventos  $\mathcal{A} = \{A_k, 1 \leq k \leq n\} = \{A_1, \dots, A_n\}$  se dice independiente si

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = \prod_{j=1}^n P(A_j),$$

por lo que se denomina colección de eventos mutuamente independientes.

**Nota 1.3.** En un experimento, si los eventos  $A_1, A_2, \dots, A_k$  ocurren, entonces:

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3 \cap \dots \cap A_k) &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots \\ &P(A_k|A_1 \cap A_2 \cap A_3 \cap \dots \cap A_{k-1}) \end{aligned}$$

Si los eventos  $A_1, A_2, A_3, \dots, A_k$  son independientes, entonces:

$$P(A_1 \cap A_2 \cap A_3 \cap \dots \cap A_k) = P(A_1)P(A_2)P(A_3) \dots P(A_k).$$

**Ejemplo 1.21.** Según [61, p. 65], en una escuela el 20 % de los estudiantes tienen problemas visuales, el 8 % problemas auditivos y el 4 % tienen ambos problemas, visuales y auditivos. Determine:

1. ¿Los dos eventos, tener problemas visuales y auditivos, son independientes?
2. ¿Cuál es la probabilidad de que un estudiante tenga problemas auditivos si tiene problemas visuales?
3. ¿Cuál es la probabilidad de que un estudiante no tenga problemas auditivos si tiene problemas visuales?

*Solución.* Sean los eventos

$A$  : “Estudiantes que tienen problemas auditivos”.

$V$  : “Estudiantes que tienen problemas visuales”.

En la tabla de contingencia 1.7 se muestra la relación entre estos dos eventos.

Tabla 1.7. Tabla de contingencia problemas auditivos vs. visuales

	$V$	$V^c$	Total
$A$	0,04	0,04	0,08
$A^c$	0,16	0,76	0,92
Total	0,20	0,80	1,00

Fuente: adaptada de [61, p. 66].

Luego:

- $P(A)P(V) = (0,08).(0,20) = 0,016$  y  $P(A \cap V) = 0,04$ . Como estos resultados son distintos,  $A$  y  $V$  no son independientes.
- $P(A|V) = \frac{P(A \cap V)}{P(V)} = \frac{0,04}{0,20} = \frac{1}{5} = 0,20$ .
- $P(A^c|V) = \frac{P(A^c \cap V)}{P(V)} = \frac{0,16}{0,20} = \frac{4}{5} = 0,80$ . ☑

**Ejemplo 1.22.** Se sacan tres cartas de un juego en sucesión, sin reemplazo, de una baraja francesa ordinaria. Sea  $A_1$  el evento la primera carta es un As rojo,  $A_2$  la segunda carta es un 10 o una Jota y  $A_3$  la tercera es mayor que 3 pero menor que 7. Encuentre la probabilidad de obtener  $P(A_1 \cap A_2 \cap A_3)$ .

*Solución.* Sean los eventos:

$A_1$  : “La primera carta es un As rojo”,

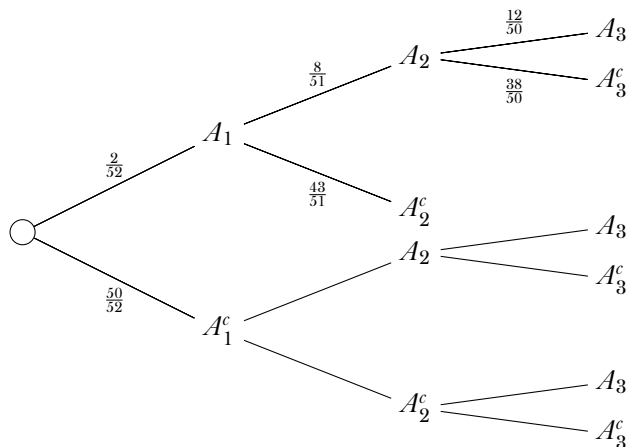
$A_2$  : “La segunda carta es 10 o Jota”,

$A_3$  : “La tercera carta es mayor que 3 y menor que 7, es decir, {4, 5, 6}”,

entonces:

$$P(A_1) = \frac{2}{52}; \quad P(A_2|A_1) = \frac{8}{51}; \quad P(A_3|A_1 \cap A_2) = \frac{12}{50}$$

$$\Rightarrow P(A_1 \cap A_2 \cap A_3) = \frac{192}{132600} = 0,0014.$$



Luego, la probabilidad de sacar tres cartas de una baraja en el orden propuesto (As rojo, 10 o Jota, mayor que 3 y menor que 7) es 0,0014 o 0,14 %.  $\square$

**Ejemplo 1.23.** Una urna contiene bolas de tres colores: 4 blancas ( $B$ ), 6 rojas ( $R$ ) y 5 negras ( $N$ ). Se sacan tres bolas al azar sin reponerlas. Hallar la probabilidad de que salgan en el orden  $R, B, R$ .

*Solución.* Sean los eventos:

$A_1$  : "La primera bola es  $R$ ",

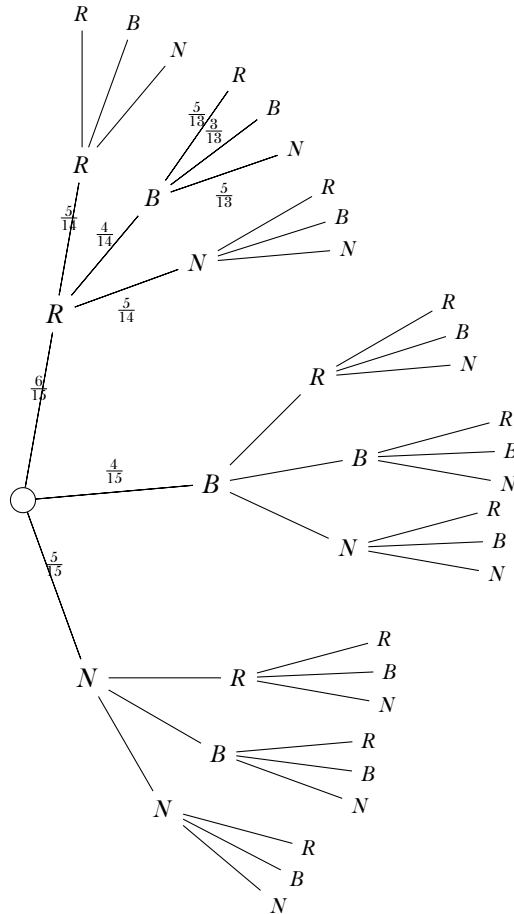
$A_3$  : "La tercera bola es  $R$ ",

$A_2$  : "La segunda bola es  $B$ ",

entonces:

$$P(A_1) = \frac{6}{15}; \quad P(A_2|A_1) = \frac{4}{14}; \quad P(A_3|A_1 \cap A_2) = \frac{5}{13}$$

$$\Rightarrow P(A_1 \cap A_2 \cap A_3) = \frac{120}{2730} = 0,0440$$

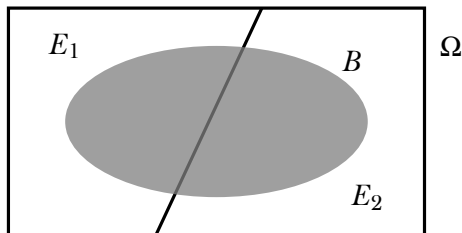


Así, la probabilidad de que al sacar tres bolas de una caja que contiene 4 bolas blancas, 6 rojas y 5 negras, sin reponerlas, salgan en el orden roja, blanca y roja es 0,044 o 4%. ✓

## 1.6. Regla de Bayes

Sean  $E_1$  y  $E_2$  una partición de  $\Omega$  (espacio muestral), es decir,  $E_1 \cup E_2 = \Omega$  y  $E_1 \cap E_2 = \emptyset$ , de los cuales se conoce sus probabilidades  $P(E_1)$  y  $P(E_2)$ , y las probabilidades condicionadas  $P(B|E_1)$  y  $P(B|E_2)$ , entonces:

$$\begin{aligned}
 P(E_1|B) &= \frac{P(E_1 \cap B)}{P(B)} = \frac{P(E_1 \cap B)}{P(B \cap E_1) + P(B \cap E_2)} \\
 &= \frac{P(B|E_1)P(E_1)}{P(B|E_1)P(E_1) + P(B|E_2)P(E_2)}.
 \end{aligned}$$



En general, usando el teorema de probabilidad total se tiene que:

$$P(E_j|B) = \frac{P(B|E_j)P(E_j)}{\sum_{k=1}^q P(B|E_k)P(E_k)} \text{ para todo } j.$$

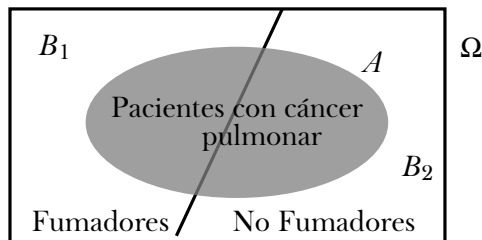
**Ejemplo 1.24.** Supóngase, como en [19, p. 44], que en un centro médico, de todos los fumadores de quienes se sospecha que tenían cáncer pulmonar, el 90 % lo tenía, mientras que únicamente el 5% de los no fumadores lo padecían. Si la proporción de fumadores es 0,45, ¿cuál es la probabilidad de que un paciente con cáncer pulmonar, seleccionado al azar, sea fumador?

*Solución.* Sean los eventos:

$B_1$ : “El paciente es fumador”.

$B_2$ : “El paciente no es fumador”.

$A$ : “El paciente tiene cáncer pulmonar”.



Del enunciado se sabe que:

$$P(B_1) = 0,45, \quad \text{entonces} \quad P(B_2) = 1 - P(B_1) = 0,55$$

$$P(A | B_1) = 0,9 \quad \text{y} \quad P(A | B_2) = 0,05,$$

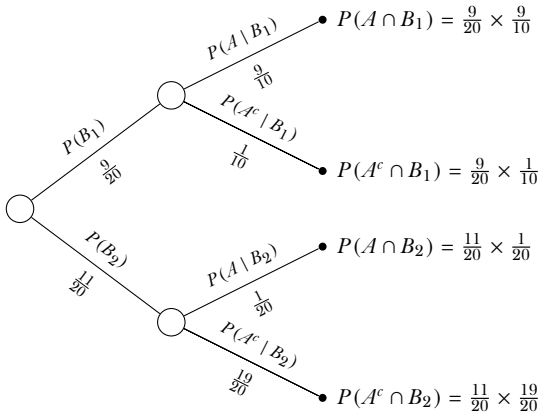
y preguntan  $P(B_1|A) = ?$

Usando el teorema de Bayes,

$$P(B_1|A) = \frac{P(A \cap B_1)}{P(A \cap B_1) + P(A \cap B_2)}$$

$$\begin{aligned}
 &= \frac{P(A | B_1)P(B_1)}{P(A | B_1)P(B_1) + P(A | B_2)P(B_2)} \\
 &= \frac{(0,90)(0,45)}{(0,9)(0,45) + (0,05)(0,55)} = 0,9364.
 \end{aligned}$$

Luego, la probabilidad de que un paciente con cáncer pulmonar, seleccionado aleatoriamente, sea fumador es 0,9364 o de un 94 %. Gráficamente:



**Ejemplo 1.25.** Suponga que se distribuyen pelotas de colores en tres cajas idénticas, de la siguiente manera:

Tabla 1.8. Distribución de pelotas de acuerdo a color y caja

	Caja	1	2	3
Color	Roja	2	4	3
Blanca	Negra	3	1	4
		5	3	3

Una caja se elige al azar y se saca una pelota de ella, también aleatoriamente, y se observa que es roja. ¿Cuál es la probabilidad de que la caja 3 sea la que se escogió?

**Solución.** Sean los eventos:

$B$ : “La pelota seleccionada es roja”.

$C_k$ : “Se selecciona la  $k$ -ésima caja”.

De la tabla se tiene que:

$$\begin{array}{lll}
 P(C_1) = 1/3 & P(C_2) = 1/3 & P(C_3) = 1/3 \\
 P(B|C_1) = 2/10 & P(B|C_2) = 4/8 & P(B|C_3) = 3/10
 \end{array}$$

Preguntan  $P(C_3|B) = ?$

$$\begin{aligned} P(C_3|B) &= \frac{P(B \cap C_3)}{P(B \cap C_1) + P(B \cap C_2) + P(B \cap C_3)} \\ &= \frac{P(B|C_3)P(C_3)}{P(B|C_1)P(C_1) + P(B|C_2)P(C_2) + P(B|C_3)P(C_3)} \\ &= \frac{(3/10)(1/3)}{(2/10)(1/3) + (4/8)(1/3) + (3/10)(1/3)} = 0,3 \end{aligned}$$

Así, la probabilidad de que la caja 3 sea la que se escogió es 0,3 o del 30%.  $\square$

## 1.7. Variable aleatoria

**DEFINICIÓN 1.9. Variable aleatoria.** Una *variable aleatoria*  $X$  en un espacio de probabilidad  $(\Omega, \mathcal{F}, P)$  es una función real definida en el espacio  $\Omega$ , tal que  $[X \leq x]$  es un evento aleatorio para todo  $x \in \mathbb{R}$ ; es decir,  $X : \Omega \rightarrow \mathbb{R}$  es *variable aleatoria* si  $[X \leq x] \in \mathcal{F} \quad \forall x \in \mathbb{R}$  (ver [43, p. 36]).

**Ejemplo 1.26.** Considérese que los hoteles ubicados en la cabecera del municipio  $A$  son cinco:

$HotelAA$	$HotelAD$
$HotelAB$	$HotelAE$
$HotelAC$	

Se desea estudiar las inversiones, costos y gastos por categoría de protección ambiental, durante el año, de estos hoteles, pero no es posible tener información de los cinco hoteles; se puede hacer una encuesta por muestreo, en la que la población finita corresponde a los hoteles ubicados en la cabecera del municipio  $A$  en el año de interés.

*Solución.* En este caso,  $\Omega = U = \{HAA, HAB, HAC, HAD, HAE\}$  y  $\mathcal{S}$  es una colección de subconjuntos de  $U$ , que corresponde a un conjunto de muestras,

$$\mathcal{S} = \{s_1, s_2, s_3, \dots, s_j, \dots, s_M\},$$

que se puedan obtener de un procedimiento de muestreo sin reemplazo de los cinco hoteles. En este caso,  $\eta(\mathcal{S}) = 2^5 = 32 = M$ , veamos:

$s_1 = \emptyset$	$s_{17} = \{HAA, HAB, HAC\}$
$s_2 = \{HAA\}$	$s_{18} = \{HAA, HAB, HAD\}$
$s_3 = \{HAB\}$	$s_{19} = \{HAA, HAB, HAE\}$
$s_4 = \{HAC\}$	$s_{20} = \{HAA, HAC, HAD\}$
$s_5 = \{HAD\}$	$s_{21} = \{HAA, HAC, HAE\}$
$s_6 = \{HAE\}$	$s_{22} = \{HAA, HAD, HAE\}$
$s_7 = \{HAA, HAB\}$	$s_{23} = \{HAB, HAC, HAD\}$
$s_8 = \{HAA, HAC\}$	$s_{24} = \{HAB, HAC, HAE\}$
$s_9 = \{HAA, HAD\}$	$s_{25} = \{HAB, HAD, HAE\}$
$s_{10} = \{HAA, HAE\}$	$s_{26} = \{HAC, HAD, HAE\}$
$s_{11} = \{HAB, HAC\}$	$s_{27} = \{HAA, HAB, HAC, HAD\}$
$s_{12} = \{HAB, HAD\}$	$s_{28} = \{HAA, HAB, HAC, HAE\}$
$s_{13} = \{HAB, HAE\}$	$s_{29} = \{HAA, HAB, HAD, HAE\}$
$s_{14} = \{HAC, HAD\}$	$s_{30} = \{HAA, HAC, HAD, HAE\}$
$s_{15} = \{HAC, HAE\}$	$s_{31} = \{HAB, HAC, HAD, HAE\}$
$s_{16} = \{HAD, HAE\}$	$s_{32} = \{HAA, HAB, HAC, HAD, HAE\}$

Note que los eventos corresponden a muestras. Usando la notación de muestreo estadístico, a cada muestra se le asigna una probabilidad conocida de selección  $P(s)$  (en muestreo se le llama diseño muestral):

$P(s_1) = 0$	$P(s_{12}) = 0$	$P(s_{23}) = 9/90$
$P(s_2) = 0$	$P(s_{13}) = 0$	$P(s_{24}) = 10/90$
$P(s_3) = 0$	$P(s_{14}) = 0$	$P(s_{25}) = 11/90$
$P(s_4) = 0$	$P(s_{15}) = 0$	$P(s_{26}) = 12/90$
$P(s_5) = 0$	$P(s_{16}) = 0$	$P(s_{27}) = 0$
$P(s_6) = 0$	$P(s_{17}) = 6/90$	$P(s_{28}) = 0$
$P(s_7) = 0$	$P(s_{18}) = 7/90$	$P(s_{29}) = 0$
$P(s_8) = 0$	$P(s_{19}) = 8/90$	$P(s_{30}) = 0$
$P(s_9) = 0$	$P(s_{20}) = 8/90$	$P(s_{31}) = 0$
$P(s_{10}) = 0$	$P(s_{21}) = 9/90$	$P(s_{32}) = 0$
$P(s_{11}) = 0$	$P(s_{22}) = 10/90$	

**Nota 1.4.**

- $X(\omega)$  es la imagen del argumento  $\omega$  bajo la función  $X$ , que también se conoce como el valor de  $X$  en  $\omega$ .
- $X$  es también una aplicación de  $\Omega$  hacia  $\Omega'$ , denotada por  $\Omega \xrightarrow{X} \Omega'$ .
- $X$  establece una relación de correspondencia entre los puntos de  $\Omega$  con los puntos de  $\Omega'$ .
- $\Omega$  es llamado el **dominio** de  $X$  y  $\Omega'$  es llamado el **rango**.
- Si el espacio del rango es la recta real,  $\mathbb{R}$ , o un subconjunto de ella, la función se dice que es numérica o función de valor real.

Considerando la información del ejemplo 1.4, para cada punto  $\omega' \in \Omega'$ , puede existir uno o más puntos de  $\Omega$  cuya imagen bajo  $X$  es  $\omega'$ . El conjunto de todos los puntos  $\omega \in \Omega$  cuya imagen bajo  $X$  es  $\omega'$  se llama la *imagen inversa* de  $\{\omega'\}$  y se denota por  $X^{-1}(\{\omega'\})$ , así:

$$X^{-1}(\{\omega'\}) := \{\omega \in \Omega : X(\omega) = \omega'\}.$$

Si se trata de un evento  $B'$ , con  $B' \subset \Omega'$ , el conjunto de todos los puntos de  $\Omega$  para los cuales  $X(\omega) \in B'$  se llama la imagen inversa de  $B'$  bajo  $X$ , denotada por  $X^{-1}(B')$ :

$$X^{-1}(B') := \{\omega : X(\omega) \in B'\}.$$

$X^{-1}$  es llamada la *función inversa* de  $X$ . La notación es:

$$\begin{aligned} X(B) &:= \{X(\omega) : \omega \in B\}, \quad B \subset \Omega, \\ X^{-1}(\mathcal{F}') &:= \{X^{-1}(B') : B' \in \mathcal{F}'\}, \end{aligned}$$

donde  $\mathcal{F}'$  es una colección de subconjuntos de  $\Omega'$  (ver [11, p. 21]).

**Ejemplo 1.27.** Considerando la información del ejemplo 1.4, supóngase que  $X(\omega)$  es el número de caras, es decir, la **función** dada por:

$$\begin{array}{ccc} X : & \Omega & \rightarrow & \Omega' \\ & \omega & \rightarrow & X(\omega) = \omega' \end{array}$$

donde  $\Omega' = \{\omega'_1, \omega'_2, \omega'_3, \omega'_4\} = \{0, 1, 2, 3\}$  y  $\mathcal{F}' = \mathcal{P}(\Omega')$ . Determine la función inversa de esta aplicación.

**Solución.** Note que para cada valor  $\omega$ ,  $X(\omega)$  es único, y está dado por:

$$\begin{array}{ll} \omega_1 = CCC \rightarrow X(\omega_1) = 3 = \omega'_4 & \omega_2 = CCS \rightarrow X(\omega_2) = 2 = \omega'_3 \\ \omega_3 = CSC \rightarrow X(\omega_3) = 2 = \omega'_3 & \omega_4 = SCC \rightarrow X(\omega_4) = 2 = \omega'_3 \\ \omega_5 = CSS \rightarrow X(\omega_5) = 1 = \omega'_2 & \omega_6 = SCS \rightarrow X(\omega_6) = 1 = \omega'_2 \\ \omega_7 = SSC \rightarrow X(\omega_7) = 1 = \omega'_2 & \omega_8 = SSS \rightarrow X(\omega_8) = 0 = \omega'_1 \end{array}$$

Luego, la colección de subconjuntos de  $\Omega'$  es dada por:

$$\begin{aligned} \mathcal{F}' = \{ & \emptyset, \{0\}, \{1\}, \{2\}, \{3\}, \{0, 1\}, \{0, 2\}, \{0, 3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \\ & \{0, 1, 2\}, \{0, 1, 3\}, \{0, 2, 3\}, \{1, 2, 3\}, \Omega' \}. \end{aligned}$$

En este caso, la imagen inversa de  $\{\omega'\}$  está dada por:

$$\begin{aligned}\omega'_1 = 0 &\quad \rightarrow \quad X^{-1}(\{\omega'_1\}) = \{\omega_8\} \\ \omega'_2 = 1 &\quad \rightarrow \quad X^{-1}(\{\omega'_2\}) = \{\omega_5, \omega_6, \omega_7\} \\ \omega'_3 = 2 &\quad \rightarrow \quad X^{-1}(\{\omega'_3\}) = \{\omega_2, \omega_3, \omega_4\} \\ \omega'_4 = 3 &\quad \rightarrow \quad X^{-1}(\{\omega'_4\}) = \{\omega_1\}.\end{aligned}$$

Si  $B' = \{0, 2\}$  entonces  $X^{-1}(B') = \{\omega_2, \omega_3, \omega_4, \omega_8\}$ . ☑

**Ejemplo 1.28.** Considere la función de valor real  $I_A$  o  $I(A)$  definida sobre  $\Omega$ , es decir,  $I_A : \Omega \rightarrow \mathbb{R}$  dada por

$$I_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \in A^c \end{cases} \quad (1.10)$$

A la función  $I_A$  o  $I(A)$  se le llama **indicadora** o **característica**. Determine la función inversa de esta aplicación.

*Solución.* El rango estricto de  $I_A$  es

$$I_A(\Omega) = \{I_A(\omega) : \omega \in \Omega\} = \{0, 1\}.$$

Si  $B \subset \mathbb{R}$ , entonces:

$$I_A^{-1}(B) = \begin{cases} \emptyset & \text{si } B \text{ no contiene ni a } 0, \text{ ni a } 1 \\ A & \text{si } B \text{ contiene a } 1, \text{ pero no a } 0 \\ A^c & \text{si } B \text{ contiene a } 0, \text{ pero no a } 1 \\ \Omega & \text{si } B \text{ contiene a } 0 \text{ y a } 1 \end{cases}$$

luego:

$$I_A^{-1}(\mathfrak{B}) = \{\emptyset, A, A^c, \Omega\}. \quad \text{☑}$$

**Nota 1.5.** En particular,

$$X^{-1}(\Omega') = \{\omega : X(\omega) \in \Omega'\} = \Omega.$$

**Nota 1.6.** (Espacio de probabilidad inducido por  $X$ ). Sea  $X$  una variable aleatoria real definida sobre  $(\Omega, \mathcal{F}, P)$ . Para cualquier  $B \in \mathfrak{B}(\mathbb{R})$ , se puede definir

$$P_X(B) := P[\{\omega : X(\omega) \in B\}] = P[X^{-1}(B)].$$

$P_X$  sobre  $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$  es una medida de probabilidad y entonces  $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P_X)$  es el espacio de probabilidad inducido por  $X$ . A  $P_X$  se le llama la **distribución de probabilidad de  $X$**  (ver [11, p. 59]).

**Ejemplo 1.29.** Supóngase, como en [19, p. 53], que se lanzan dos dados, uno rojo y otro azul. Sea la variable aleatoria  $X$  que representa la suma de las caras de los dados al ser lanzados:

Resultado ( $\omega$ )	Vr. de la Variable Aleatoria ( $X(\omega)$ )	$h$	Probabilidad
(1,1)	2	1	$P(X = 2) = 1/36$
(1,2) (2,1)	3	2	$P(X = 3) = 2/36$
(1,3) (2,2) (3,1)	4	3	$P(X = 4) = 3/36$
(1,4) (2,3) (3,2) (4,1)	5	4	$P(X = 5) = 4/36$
(1,5) (2,4) (3,3) (4,2) (5,1)	6	5	$P(X = 6) = 5/36$
(1,6) (2,5) (3,4) (4,3) (5,2) (6,1)	7	6	$P(X = 7) = 6/36$
(2,6) (3,5) (4,4) (5,3) (6,2)	8	5	$P(X = 8) = 5/36$
(3,6) (4,5) (5,4) (6,3)	9	4	$P(X = 9) = 4/36$
(4,6) (5,5) (6,4)	10	3	$P(X = 10) = 3/36$
(5,6) (6,5)	11	2	$P(X = 11) = 2/36$
(6,6)	12	1	$P(X = 12) = 1/36$
	Total	$k = 36$	

donde  $h$  denota la frecuencia absoluta, es decir, el número de ocurrencias.

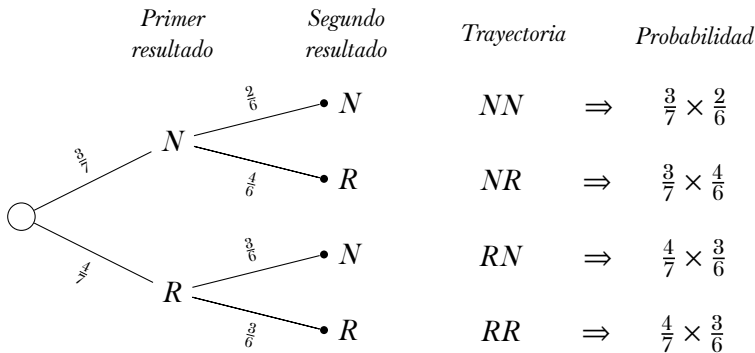
- Si se está interesado en el evento en el que la suma de las caras al ser lanzados dos dados, uno rojo y otro azul, es igual a 4, se tiene que:

$$P_X([4,4]) := P[\{\omega : X(\omega) \in [4,4]\}] = P[X^{-1}([4,4])] = 3/36.$$

- Si nos interesa el evento en el que la suma de las caras al ser lanzados dos dados, uno rojo y otro azul, sea menor a 5, se tiene que:

$$P_X((-\infty,5]) := P[\{\omega : X(\omega) \in (-\infty,5)\}] = P[X^{-1}((-\infty,5))] = 6/36.$$

**Ejemplo 1.30.** Se sacan dos pelotas en sucesión, sin reemplazos, de una urna que contiene 4 pelotas rojas (R) y 3 negras (N). Los resultados posibles,  $y$ , y los valores de la variable aleatoria  $Y$ , donde  $Y$  es el número de pelotas rojas, son:



Resultado ( $\omega$ 's)	$Y(\omega)$	Probabilidad
RR	2	$P(Y = 2) = \frac{4}{7} \times \frac{3}{6} = \frac{12}{42}$
RN - NR	1	$P(Y = 1) = \frac{4}{7} \times \frac{3}{6} + \frac{3}{7} \times \frac{4}{6} = \frac{24}{42}$
NN	0	$P(Y = 0) = \frac{3}{7} \times \frac{2}{6} = \frac{6}{42}$

**Ejemplo 1.31.** Según [64, pg. 107], un estudio de mercadotecnia estima que un nuevo instrumento para el análisis de muestras de suelo tendrá mucho, poco o ningún éxito, con probabilidades 0,3, 0,6 y 0,1, respectivamente. Las ganancias anuales asociadas con un producto muy exitoso, poco exitoso o no exitoso son 10 millones, 5 millones y 1 millón de dólares, respectivamente. La variable aleatoria es la ganancia anual del producto. Determine  $P(X = x)$ .

*Solución.* Sea  $X$  la ganancia anual del producto, entonces

$$x = 10, \quad 5 \quad \text{ó} \quad 1$$

Una forma reducida y adecuada de expresar la función de probabilidad anterior es:

$$P(X = x) = \begin{cases} 0,1 & \text{si } x = 1 \\ 0,6 & \text{si } x = 5 \\ 0,3 & \text{si } x = 10 \\ 0 & \text{en otro caso} \end{cases} \quad \checkmark$$

**Ejemplo 1.32.** Según [16, p. 67], sea  $X$  una variable aleatoria tal que:

$$f_X(x) = \begin{cases} 1 & \text{si } 0 \leq x < 1 \\ 0 & \text{en otro caso,} \end{cases}$$

encuentre  $P[X \leq 0,2]$ .

*Solución.* Sea  $B = (-\infty, 0,2)$ , entonces

$$\begin{aligned} P_X [(-\infty, 0,2)] &= P[\{\omega : X(\omega) \in (-\infty, 0,2)\}] = P [X^{-1}((-\infty, 0,2))] \\ &= \int_{-\infty}^{0,2} f_X(x) dx = \int_{-\infty}^0 f_X(x) dx + \int_0^{0,2} f_X(x) dx \\ &= \int_0^{0,2} dx = x|_0^{0,2} = 0,2 - 0 = 0,2. \end{aligned} \quad \checkmark$$

**Nota 1.7.** El soporte de una “medida” corresponde a todos los conjuntos que no tiene medida cero.

**DEFINICIÓN 1.10.** (Función de distribución acumulada). Sea  $X$  una variable aleatoria real. La función  $F_X$  definida sobre  $\mathbb{R}$  mediante:

$$F_X(x) := P_X((-\infty, x]) = P(X \leq x),$$

se denomina función de distribución acumulada (*fda*) de la variable aleatoria  $X$  (ver [16, p. 56]).

**Teorema 1.5. Propiedades de la fda.** Sea  $X$  una variable aleatoria real definida sobre  $(\Omega, \mathcal{F}, P)$ . La *fda*  $F_X$  satisface las siguientes condiciones:

1. Si  $x \leq y$  entonces  $F_X(x) \leq F_X(y)$ .
2.  $F_X(x^+) := \lim_{h \rightarrow 0^+} F_X(x+h) = F_X(x)$  para todo  $x \in \mathbb{R}$ .
3.  $\lim_{x \rightarrow \infty} F_X(x) = 1$ .
4.  $\lim_{x \rightarrow -\infty} F_X(x) = 0$ .

*Demostración.* Ver detalles de la prueba en [16, p. 59]. ☑

**Ejemplo 1.33.** Considerando la información del ejemplo 1.10, determine la  $F_X(x)$ .

*Solución.* La *fda* correspondiente es dada por

$$\underbrace{P(X \leq x)}_{F_X(x)} = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{8} & \text{si } 0 \leq x < 1 \\ \frac{4}{8} & \text{si } 1 \leq x < 2 \\ \frac{7}{8} & \text{si } 2 \leq x < 3 \\ 1 & \text{si } x \geq 3. \end{cases} \quad \text{☑}$$

**Ejemplo 1.34.** Considerando la información del ejemplo 1.32, determine la  $F_X(x)$ .

*Solución.* La *fda* correspondiente es dada por

$$\underbrace{P(X \leq x)}_{F_X(x)} = \begin{cases} 0 & \text{si } x < 0 \\ \int_0^x dt = x & \text{si } 0 \leq x < 1 \\ 0 & \\ 1 & \text{si } x \geq 1. \end{cases} \quad \text{☑}$$

## 1.8. Distribuciones de probabilidad

En la sección anterior se dio en la definición 1.9 el concepto de *variable aleatoria*, en esta sección se presentan las distribuciones asociadas a las variables aleatorias más empleadas.

**DEFINICIÓN 1.11.** (Tipos de variables aleatorias). Según [43, p. 41]:

- Una variable aleatoria  $X$  es **discreta** si toma un número finito o numerable de valores, es decir, si existe un conjunto finito o numerable  $\{x_1, x_2, \dots\} \subset \mathbb{R}$  tal que  $X(\omega) \in \{x_1, x_2, x_3, \dots\}$  para todo  $\omega \in \Omega$ . La función  $p(x_i)$  definida por  $p(x_i) = P(X = x_i)$ ,  $i = 1, 2, 3, \dots$ , como sigue,

$$f_X(x) = \begin{cases} P(X = x_i) & \text{si } x = x_1, x_2, x_3, \dots \\ 0 & \text{en otro caso} \end{cases}$$

se llama función de masa de probabilidad (**fmp**) (por sus siglas o función de frecuencia) de la variable aleatoria  $X$ . Esta función satisface que:

- $f_X(x) \geq 0$  para todos los valores de  $x$ .
  - $\sum_x f_X(x) = 1$ .
- Una variable aleatoria  $X$  es (absolutamente) **continua** si existe una función  $f_X(x) \geq 0$  tal que:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \quad \forall x \in \mathbb{R}.$$

En este caso, se dice que  $f_X(x)$  es función de densidad de probabilidad (función de densidad de probabilidad (**fdp**), por sus siglas) de la variable aleatoria  $X$  o simplemente **densidad** de  $X$ . La **fdp** cumple que

- $f_X(x) \geq 0$  para todos los valores de  $x$ .
- $\int_x f_X(x) dx = 1$ .

### 1.8.1. Algunas distribuciones discretas

**Ejemplo 1.35.** (Distribución uniforme discreta). Para [16, p. 108], una variable aleatoria  $X$  tiene distribución uniforme discreta de parámetro  $n$ , con  $n \in \mathbb{Z}^+$ , si su **fmp** es:

$$P(X = x) = \begin{cases} \frac{1}{n} & \text{si } x = 1, 2, 3, \dots, n \\ 0 & \text{en otro caso.} \end{cases} \quad (1.11)$$

Determine su *fda*  $F_X$ .

*Solución.* En este caso,  $F_X$  es dada por:

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{1}{n} & \text{si } 1 \leq x < 2 \\ \vdots & \vdots \\ \frac{n-1}{n} & \text{si } n-1 \leq x < n \\ 1 & \text{si } x \geq n. \end{cases}$$

En las figuras 1.1 y 1.2 se presentan la *fmp* y la *fda*, respectivamente, para valores de  $n$  iguales a 5 y 10.

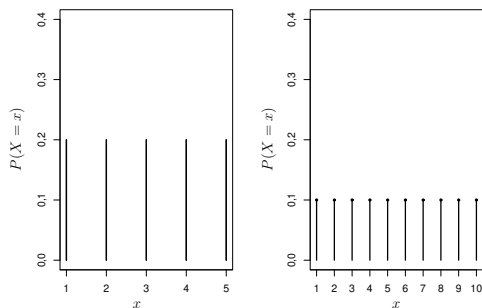


Figura 1.1. *fmp* uniforme discreta. (a)  $n = 5$  y (b)  $n = 10$

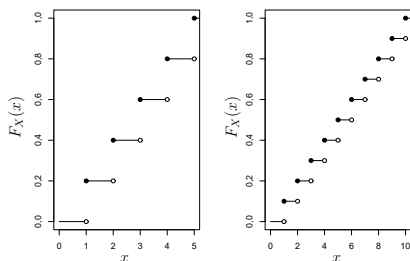


Figura 1.2. *fda* uniforme discreta. (a)  $n = 5$  y (b)  $n = 10$

**Ejemplo 1.36.** Considere el experimento de lanzar un dado “normal” y observe el valor de la cara que queda hacia arriba.

*Solución.* En este caso la variable aleatoria  $X$  tiene distribución uniforme discreta. Luego, los posibles valores de la variable aleatoria  $X$  y su probabilidad se presentan en la tabla 1.9.

Tabla 1.9. Probabilidad para el ejemplo sobre el lanzamiento de un dado

Resultado ( $\omega$ )	Vr. de la variable aleatoria ( $X(\omega)$ )	Probabilidad
1	1	$P(X = 1) = 1/6$
2	2	$P(X = 2) = 1/6$
3	3	$P(X = 3) = 1/6$
4	4	$P(X = 4) = 1/6$
5	5	$P(X = 5) = 1/6$
6	6	$P(X = 6) = 1/6$

La probabilidad de que el dado caiga máximo en 3 es dada por:

$$P(X \leq 3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = 0,5. \quad \checkmark$$

Antes de presentar el siguiente ejemplo, se presentará el concepto de un experimento (proceso) Bernoulli.

**Nota 1.8.** (Proceso Bernoulli). *Un proceso Bernoulli satisface que:*

1. El experimento consiste de  $m$  ensayos repetidos.
2. El resultado de cada uno de los ensayos puede clasificarse como un éxito o como un fracaso.
3. La probabilidad de éxito, representada por  $p$ , permanece constante para todos los ensayos.
4. Hay independencia entre los  $m$  ensayos.

Sea  $X$  la variable aleatoria que determina el número de éxitos en un experimento Bernoulli, se dice que  $X$  tiene distribución binomial. Notación:  $X \sim B(m, p)$ <sup>1</sup>.

**Ejemplo 1.37.** (Distribución binomial). Para [16, p. 109], una variable aleatoria  $X$  tiene distribución binomial de parámetros  $m$  y  $p$ , con  $m \in \mathbb{Z}^+$  y  $0 < p < 1$ , si su fmp es:

$$P(X = x) = \begin{cases} \binom{m}{x} p^x (1 - p)^{m-x} & \text{si } x = 0, 1, 2, 3, \dots, m \\ 0 & \text{en otro caso.} \end{cases} \quad (1.12)$$

Determine su fda  $F_X$ .

<sup>1</sup>Cuando en la distribución binomial  $m = 1$ , la distribución se llama distribución Bernoulli y se denota por  $X \sim Ber(p)$ .

*Solución.* En este caso  $F_X$  es dada por:

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 0 \\ \binom{m}{0} p^0 (1-p)^{m-0} & \text{si } 0 \leq x < 1 \\ (1-p)^m + \binom{m}{1} p (1-p)^{m-1} & \text{si } 1 \leq x < 2 \\ \vdots & \vdots \\ \sum_{x=0}^{m-1} \binom{m}{x} p^x (1-p)^{m-x} & \text{si } m-1 \leq x < m \\ 1 & \text{si } x \geq m. \end{cases}$$

En las figuras 1.3 y 1.4 se presentan la **fmp** y la **fda**, respectivamente, para  $m = 10$  y valores de  $p$  iguales a 0,2 y 0,5. ☑

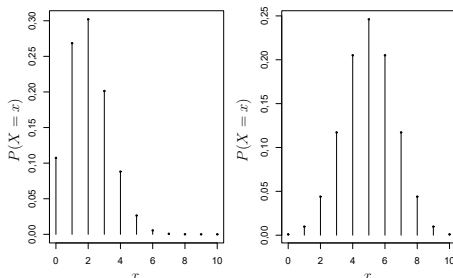


Figura 1.3. **fmp** binomial,  $m = 10$ . (a)  $p = 0,2$  y (b)  $p = 0,5$

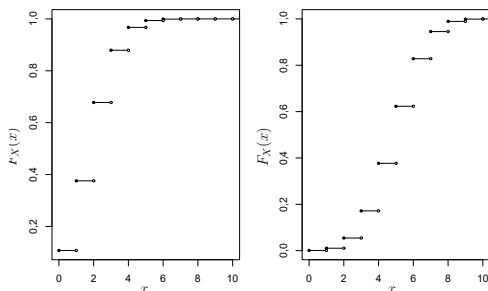


Figura 1.4. **fda** binomial,  $m = 10$ . (a)  $p = 0,2$  y (b)  $p = 0,5$

**Ejemplo 1.38.** Un agricultor que siembra fruta afirma que  $2/3$  de su cosecha de duraznos se contaminó con una especie de mosca de fruta (mediterránea). Encuentre la probabilidad de que al examinar 4 duraznos:

1. Los 4 duraznos estén contaminados por la mosca del mediterráneo.
2. Menos de 2 duraznos estén contaminados por la mosca del mediterráneo.
3. Cualquier cantidad entre 1 y 3 duraznos, inclusive, esté contaminada por la mosca del mediterráneo.

*Solución.* Sea la variable aleatoria  $X$  que indica si “un durazno está contaminado por la larva de la mosca mediterránea”. Considerando que:

- La probabilidad de que en cada inspección un durazno esté contaminado por la mosca del mediterráneo, “éxito”,  $p$ , permanece constante.
- Las  $m = 4$  selecciones son independientes entre sí.

Suponiendo que  $X$  tiene fmp binomial  $B(m, p)$  con  $m = 4$  y  $p = 2/3$ .

1. Los 4 duraznos estén contaminados por la mosca del mediterráneo, es decir,  $P(X = 4) = ?$

$$P(X = 4) = \binom{4}{4} \left(\frac{2}{3}\right)^4 \left(1 - \frac{2}{3}\right)^{4-4} = \left(\frac{2}{3}\right)^4 = 0,1975.$$

Por tanto, la probabilidad de que los cuatro duraznos estén contaminados por la mosca del mediterráneo es 0,1975 o de un 20 %.

2. Menos de 2 duraznos estén contaminados por la mosca del mediterráneo, es decir,  $P(X < 2)$ .

$$\begin{aligned} P(X < 2) &= \binom{4}{0} \left(\frac{2}{3}\right)^0 \left(1 - \frac{2}{3}\right)^{4-0} + \binom{4}{1} \left(\frac{2}{3}\right)^1 \left(1 - \frac{2}{3}\right)^{4-1} \\ &= 0,01235 + 0,09877 = 0,1111. \end{aligned}$$

Por tanto, la probabilidad de que menos de 2 duraznos estén contaminados por la mosca del mediterráneo es 0,1111 o 11 %.

3. Cualquier cantidad entre 1 y 3 duraznos, inclusive, esté contaminada por la mosca del mediterráneo, es decir,  $P(1 \leq X \leq 3)$ :

$$\begin{aligned} P(1 \leq X \leq 3) &= \binom{4}{1} \left(\frac{2}{3}\right)^1 \left(1 - \frac{2}{3}\right)^{4-1} + \binom{4}{2} \left(\frac{2}{3}\right)^2 \left(1 - \frac{2}{3}\right)^{4-2} + \\ &\quad \binom{4}{3} \left(\frac{2}{3}\right)^3 \left(1 - \frac{2}{3}\right)^{4-3} = 0,09877 + 0,2963 + 0,3951 \\ &= 0,7901. \end{aligned}$$

Luego, la probabilidad de que entre 1 y 3 duraznos, inclusive, estén contaminados por la mosca mediterránea es 0,7901 o 79%.  $\square$

Antes de presentar el siguiente ejemplo, se presentará el concepto de un experimento (proceso) Poisson.

**Nota 1.9.** (Proceso Poisson). *Un proceso Poisson en particular considera que un intervalo en los números reales puede subdividirse en subintervalos suficientemente pequeños (ver [64, p. 148]) tales que:*

1. La probabilidad de que ocurra más de un resultado en un intervalo de tiempo muy corto es despreciable.
2. La probabilidad de que un resultado sencillo ocurra en un intervalo de tiempo o en una región específica es proporcional a la longitud del intervalo de tiempo.
3. El proceso Poisson no tiene memoria, esto significa que el número de resultados que ocurren en un intervalo de tiempo o en una región específica es independiente del número que ocurre en otro intervalo disyunto de tiempo.

**Ejemplo 1.39.** (Distribución Poisson). *Para [16, p. 121], una variable aleatoria  $X$  tiene distribución Poisson de parámetro  $\lambda$ ,  $\lambda > 0$ , si su fmp es:*

$$P(X = x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & \text{si } x = 0, 1, 2, 3, \dots \\ 0 & \text{en otro caso.} \end{cases} \quad (1.13)$$

Determine su fda  $F_X$ .

*Solución.* En este caso,  $F_X$  es dada por:

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{e^{-\lambda} \lambda^0}{0!} & \text{si } 0 \leq x < 1 \\ e^{-\lambda} + \frac{e^{-\lambda} \lambda^1}{1!} & \text{si } 1 \leq x < 2 \\ e^{-\lambda} + e^{-\lambda} \lambda + \frac{e^{-\lambda} \lambda^2}{2!} & \text{si } 2 \leq x < 3 \\ \vdots & \vdots \end{cases}$$

El parámetro  $\lambda$  indica la tasa de ocurrencia por unidad de medida bajo un proceso Poisson. En la figura 1.5 se presentan la fmp y la fda para  $\lambda = 2, 93$  y  $\lambda = 19, 90$ .

La distribución de Poisson es el principal modelo de probabilidad usado para analizar problemas de línea de espera. Notación:  $X \sim \mathcal{P}(\lambda)$ .  $\square$

**Ejemplo 1.40.** *Se sospecha que muchas muestras de agua, todas del mismo tamaño y tomadas del río A a 2 kilómetros de su nacimiento, han sido contaminadas por*

operarios irresponsables de una planta de tratamiento de aguas. Se contó el número de organismos coliformes en cada muestra. El número medio de organismos por muestra fue de 15. Suponiendo que la distribución de probabilidad del número de organismos se pueda considerar que sigue una distribución Poisson, calcular la probabilidad de que el número de organismos en la siguiente muestra sea:

1. Más de 16.
2. 18 o menos.
3. Exactamente dos.

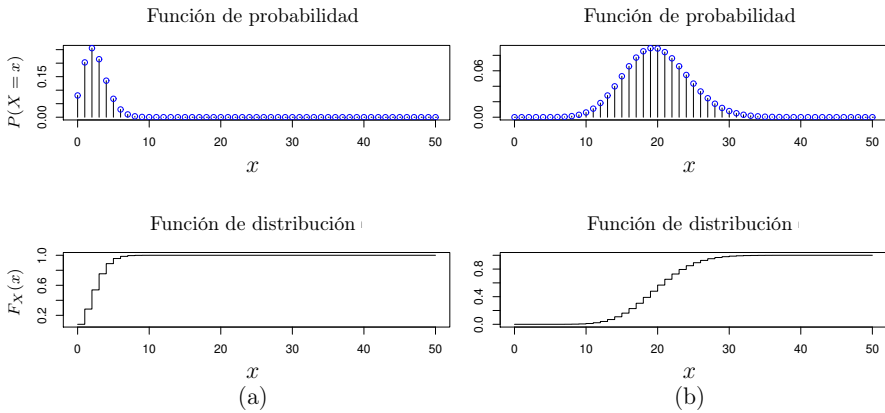


Figura 1.5. fmp y fda Poisson. (a)  $\lambda = 2,93$  y (b)  $\lambda = 19,90$

*Solución.* Sea  $X$  la variable aleatoria que cuenta el “número de organismos coliformes en cada muestra”, supóngase que cada muestra se recoge a 2 kilómetros del nacimiento del río A. Si se considera que las condiciones son propias para un proceso Poisson, se tiene que  $X \sim P(\lambda)$  con  $\lambda = 15$ .

1.  $P(X > 16) = P(X \geq 17) = 1 - P(X \leq 16)$ , es decir,

$$\begin{aligned}
 P(X > 16) &= 1 - [P(X = 0) + P(X = 1) + P(X = 2) \\
 &\quad + P(X = 3) + \dots + P(X = 15) + P(X = 16)] \\
 &= 1 - [0 + 0 + 0 + 0,0002 + \dots + 0,1024 + 0,0960] \\
 &= 1 - 0,6641 = 0,3359.
 \end{aligned}$$

Así, la probabilidad de que la siguiente muestra contenga más de 16 organismos es de 0,34.

$$2. P(X \leq 18) = \sum_{k=0}^{18} P(X = k), \text{ luego}$$

$$P(X \leq 18) = 0 + 0 + 0 + 0,0002 + \dots + 0,0847 + 0,0706 = 0,8195.$$

Entonces, la probabilidad de que la siguiente muestra contenga 18 o menos organismos es 0,8195 o 82%.

$$3. P(X = 2) = \frac{e^{-\lambda} \lambda^2}{2!} = \frac{e^{-15} 15^2}{2!} = 0,000034.$$

Por tanto, la probabilidad de que la siguiente muestra contenga exactamente dos organismos es 0,000034.  $\checkmark$

**Ejemplo 1.41.** (Distribución geométrica). Para [16, p. 126], una variable aleatoria  $X$  tiene distribución geométrica de parámetro  $p$ ,  $0 < p < 1$ , si su fmp es:

$$P(X = x) = \begin{cases} p(1-p)^x & \text{si } x = 0, 1, 2, 3, \dots \\ 0 & \text{en otro caso.} \end{cases} \quad (1.14)$$

Determine su fda  $F_X$ .

*Solución.* Su fda  $F_X$  es dada por:

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < 0 \\ p(1-p)^0 & \text{si } 0 \leq x < 1 \\ p + p(1-p)^1 & \text{si } 1 \leq x < 2 \\ p + p(1-p) + p(1-p)^2 & \text{si } 2 \leq x < 3 \\ \vdots & \vdots \end{cases}$$

Siendo  $X$  la variable aleatoria que determina el número de fracasos en un experimento Bernoulli, antes de aparecer un éxito, se dice que  $X$  tiene distribución geométrica. Notación:  $X \sim G(p)$ .  $\checkmark$

**Ejemplo 1.42.** Muestre que (1.14) satisface las propiedades de fmp (definición 1.11, ítems 1.11 y 1.11).

*Solución.* La parte a) se satisface de manera trivial. A partir de la expresión (A.6) se tiene que:

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r} \quad \text{si } |r| < 1. \quad (1.15)$$

Si  $|r| \geq 1$ , la serie diverge. Para probar b) en la expresión anterior se toma  $r = 1 - p$ , luego:

$$\sum_{k=1}^{\infty} P(X = x_k) = \sum_{k=0}^{\infty} p(1 - p)^k = p \frac{1}{1 - (1 - p)} = 1. \quad \checkmark$$

En la figura 1.6 se presentan la fmp y la fda para  $p = 0,19$  y  $p = 0,41$ .

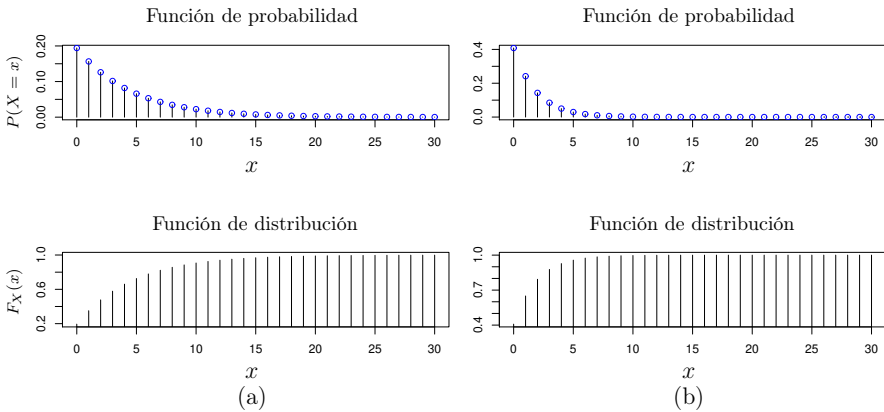


Figura 1.6. fmp y fda geométrica. (a)  $p = 0,19$  y (b)  $p = 0,41$

**Ejemplo 1.43.** Según [64, p. 132], la probabilidad de que una muestra de aire contenga una molécula rara es 0,02. Si se supone que las muestras son independientes con respecto a la molécula rara, ¿cuál es la probabilidad de que sea necesario analizar exactamente 125 muestras antes de detectar una molécula rara?

*Solución.* Sea  $X$  la variable aleatoria que cuenta el número de muestras que se analizan antes de detectar una molécula rara, si se considera que las condiciones son propias para un proceso Bernoulli se podría asumir que  $X \sim G(0,02)$ . Preguntan por  $P(X = 125)$ .

$$P(X = 125) = p(1 - p)^x = 0,02 \times (1 - 0,02)^{125} = 0,0016.$$

Entonces, la probabilidad de que sea necesario analizar exactamente 125 muestras antes de detectar una molécula rara es 0,0016.  $\checkmark$

**DEFINICIÓN 1.12. Distribución de probabilidad binomial negativa.** Sea  $X$  la variable aleatoria que determina el número de fracasos antes de aparecer  $r$ -éxitos.

Si  $X$  tiene *fmp* dada por:

$$P(X = k) = \begin{cases} \binom{r+k-1}{k} p^r q^k & \text{si } k = 0, 1, 2, 3, \dots \\ 0 & \text{en otro caso,} \end{cases} \quad (1.16)$$

donde  $r > 0$ ,  $0 < p < 1$  y  $q = 1 - p$ . Entonces se dice que  $X$  tiene distribución binomial negativa con parámetros  $r$  y  $p$ , y se denota " $X \sim BN(r, p)$ ". Esta *fmp* cuando  $r \in \mathbb{N}$  se conoce como la *fmp* de Pascal.

La fórmula del binomio de Newton con exponente negativo está dada por (A.5):

$$f(x) = (1 - x)^{-n} = \sum_{k=0}^{\infty} \binom{n+k-1}{k} x^k,$$

usando la expresión dada en [32] se tiene que:

$$(\alpha - \beta)^{-r} = \alpha^{-r} \left(1 - \frac{\beta}{\alpha}\right)^{-r} = \alpha^{-r} \sum_{k=0}^{\infty} \binom{r+k-1}{k} \left(\frac{\beta}{\alpha}\right)^k, \quad (1.17)$$

donde  $\alpha = 1 + \beta$ . Luego, la *fmp* binomial negativa se expresa como:

$$P(X = k) = \binom{r+k-1}{k} \left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta}{1+\beta}\right)^k, \quad (1.18)$$

donde  $p = (1 + \beta)^{-1}$ . Puesto que  $r > 0$ , ( $r \in \mathbb{R}$ ), para encontrar la combinatoria, se usa el símbolo de Pochhammer dado en [1] como sigue:

$$\binom{r+k-1}{k} = \frac{(r)_k}{k!} = \begin{cases} 1, & \text{si } k = 0 \\ \prod_{m=0}^{k-1} \frac{m+r}{m+1}, & \text{si } k \geq 1. \end{cases} \quad (1.19)$$

**Ejemplo 1.44.** De acuerdo con [16, p. 126], en un departamento de control de calidad se inspeccionan las unidades terminadas que provienen de una línea de ensamble. Si la proporción de unidades defectuosas es de 0,03, ¿cuál es la probabilidad de que la vigésima unidad inspeccionada sea la tercera que se encuentra defectuosa?

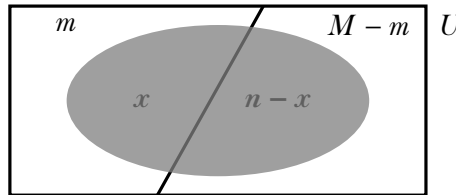
*Solución.* Sea  $Y$  = "número de unidades que es necesario inspeccionar hasta obtener exactamente tres unidades defectuosas". Denotando  $X = Y - r$ , es claro que  $X \sim BN(3, 0,03)$ . Por lo tanto,

$$P(X = 17) = \binom{19}{2} (0,03)^3 (1 - 0,03)^{17} = 2,7509 \times 10^{-3}. \quad \checkmark$$

**DEFINICIÓN 1.13. Distribución de probabilidad hipergeométrica.** Sea  $M$  el número total de objetos en una población finita, con  $m$  de estos de un tipo y  $M - m$  de otro. Al elegir una muestra aleatoria de la población constituida por  $n$  objetos, la probabilidad de que sean exactamente  $x$  de un tipo y  $n - x$  sean del otro, está dada por la *fmp* hipergeométrica:

$$P(X = x) = \begin{cases} \frac{\binom{m}{x} \binom{M-m}{n-x}}{\binom{M}{n}} & \text{si } x = 0, 1, 2, \dots, n \\ 0 & \text{en otro caso,} \end{cases} \quad (1.20)$$

donde los parámetros  $(M, m, n)$  satisfacen  $n \leq M$ ,  $m \leq M - n$ .



**Ejemplo 1.45.** De acuerdo con [16, p. 115], la división de vigilancia de una Institución de Educación Superior (IES) ha adquirido 50 equipos de comunicación para optimizar el servicio de sus predios. Se selecciona aleatoriamente 8 equipos y se ensayan para encontrar posibles defectos. Si 3 de los 50 equipos están defectuosos, ¿cuál es la probabilidad de que la muestra contenga a lo más dos equipos defectuosos?

*Solución.* Sea  $X :=$  “número de equipos defectuosos en la muestra”. Luego,  $X \sim Hg(50, 3, 8)$ . Por lo tanto,

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \frac{\binom{3}{0} \binom{47}{8}}{\binom{50}{8}} + \frac{\binom{3}{1} \binom{47}{7}}{\binom{50}{8}} + \frac{\binom{3}{2} \binom{47}{6}}{\binom{50}{8}} = 0,99714. \quad \checkmark \end{aligned}$$

### 1.8.2. Algunas distribuciones continuas

**Ejemplo 1.46.** (Distribución uniforme continua). Según [16, p. 140], una variable aleatoria  $X$  tiene distribución uniforme sobre el intervalo  $[a, b]$ , con  $a < b$  números reales, si su *fdp* es:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{en otro caso.} \end{cases} \quad (1.21)$$

*Solución.* Su fda  $F_X$  es dada por:

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x \geq b. \end{cases}$$

En las figuras 1.7 y 1.8 se presentan la fdp y la fda, respectivamente, para los intervalos  $[a, b] = [1, 8]$  y  $[a, b] = [4, 7; 7, 8]$ .

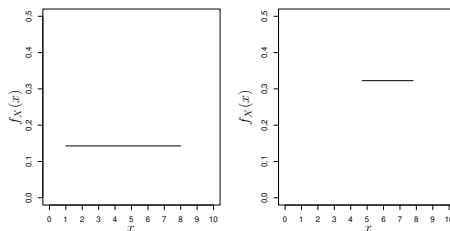


Figura 1.7. fdp uniforme sobre  $[a, b]$ . (a)  $[1, 8]$  y (b)  $[4, 7; 7, 8]$

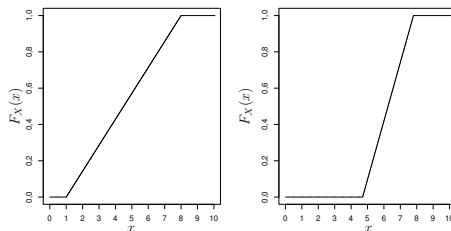


Figura 1.8. fda uniforme sobre  $[a, b]$ . (a)  $[1, 8]$  y (b)  $[4, 7; 7, 8]$



**Ejemplo 1.47.** Suponga que las ventas diarias de una gasolinera  $A$  está entre 12 y 17 millones. Si se considera que una distribución uniforme es apropiada para la variable aleatoria “volumen de ventas diarias de la gasolinera”:

1. ¿Qué porcentaje de días venderá más de 15 millones?
2. ¿Qué porcentaje de días venderá entre 13 y 14 millones?

*Solución.*

1. El porcentaje de días que venderá más de 15 millones será:

$$P(X > 15) = \int_{15}^{17} \frac{1}{17-12} dt = \frac{1}{5} t \Big|_{t=15}^{t=17} = \frac{17-15}{5} = \frac{2}{5} = 0,40.$$

Es decir, la gasolinera venderá más de 15 millones el 40% de los días.

2. El porcentaje de días que venderá entre 13 y 14 millones será:

$$P(13 < X < 14) = \int_{13}^{14} \frac{1}{17-12} dt = \frac{1}{5} t \Big|_{t=13}^{t=14} = \frac{14-13}{5} = \frac{1}{5} = 0,20.$$

Luego, se venderán entre 13 y 14 millones el 20% de los días.    ✓

**DEFINICIÓN 1.14.** (Distribución beta). Para [16, p. 161], una variable aleatoria  $X$  tiene distribución beta de parámetros  $\alpha$  y  $\beta$ , si su fdp es dada por:

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad 0 < x < 1, \quad \alpha > 0, \quad \beta > 0, \tag{1.22}$$

donde  $\Gamma(\cdot)$  es la función gamma<sup>2</sup>, es decir,

$$\Gamma(r) = \int_0^\infty t^{r-1} e^{-t} dt. \tag{1.23}$$

**DEFINICIÓN 1.15.** (Distribución normal). Para [16, p. 145], una variable aleatoria  $X$  tiene distribución normal de parámetros  $\mu$  y  $\sigma^2$ , con  $\mu$  número real y  $\sigma^2 > 0$ , si su fdp es:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} (x - \mu)^2\right\}; x \in \mathbb{R}. \tag{1.24}$$

Notación:  $X \sim N(\mu, \sigma^2)$ . Luego, su fda  $F_X$  es dada por:

$$F_X(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp\left\{-\frac{1}{2\sigma^2} (u - \mu)^2\right\} du.$$

<sup>2</sup>La función gamma satisface en particular que [20, p. 99]:

$$\Gamma(r+1) = \begin{cases} 1, & r = 0, \\ r\Gamma(r), & r > 0, \end{cases} \quad \begin{matrix} \Gamma(k) = (k-1)! & \text{para } k \in \mathbb{Z}^+ \\ \Gamma(1/2) = \sqrt{\pi}. \end{matrix}$$

La distribución normal es muy usada en la literatura estadística, ya que no solo una gran variedad de variables aleatorias se considera que siguen esta **fdp**, sino que al tener en cuenta muestras grandes (teoría asintótica), en variados casos se tiene convergencia a la distribución normal.

**Nota 1.10.** La demostración que (1.24) satisface las propiedades de **fdp** (definición 1.11, ítems 1.11 y 1.11) se puede consultar en [21, p. 12].

En la figura 1.9 se presentan la **fdp** y la **fda** para  $\sigma^2 = 2^2$ , y valores de  $\mu$  igual a  $\mu = 4$  y  $\mu = -3$ .

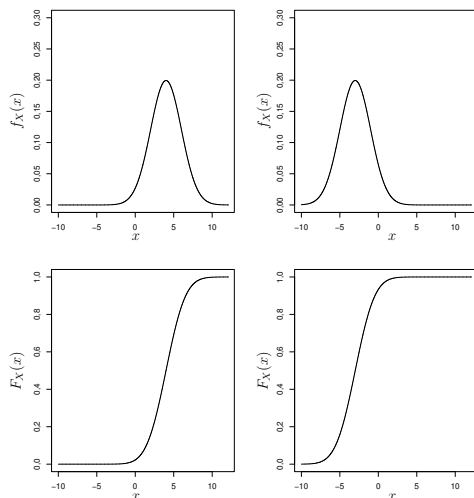


Figura 1.9. **fdp** y **fda** normal con  $\sigma^2 = 4$ . (a)  $\mu = 4$  y (b)  $\mu = -3$

En la figura 1.10 se presentan la **fdp** y la **fda** para  $\mu = 2$ , y valores de  $\sigma$  igual a  $\sigma^2 = 2^2$  y  $\sigma^2 = 0,7^2$ .

Un caso especial de la distribución normal es cuando  $\mu = 0$  y  $\sigma^2 = 1$  (**distribución normal estándar**). En este caso<sup>3</sup> es usual usar la notación  $Z \sim N(0, 1)$  y denotar  $P(Z \leq z) = \Phi(z)$ .

<sup>3</sup>Una notación matemática unificada utilizada en este material es:

$$\varphi(x, 0, 1) = \varphi(x) \quad \text{y} \quad \Phi(x, 0, 1) = \Phi(x).$$

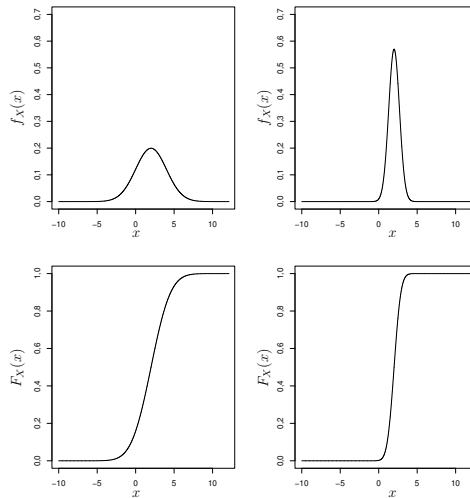


Figura 1.10. **fdp** y **fda** normal con  $\mu = 2$ . (a)  $\sigma^2 = 2^2$  y (b)  $\sigma^2 = 0.7^2$

En la figura 1.11 se presentan la **fdp** y la **fda** para la distribución normal estándar.

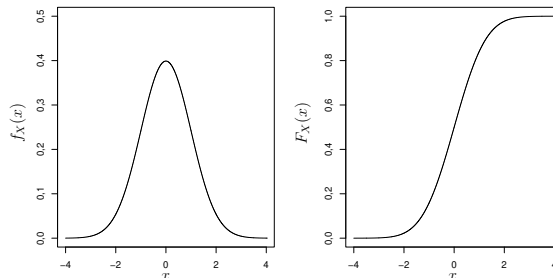


Figura 1.11. **fdp** y **fda** normal estándar

**Ejemplo 1.48.** Suponga que entre los diabéticos, el nivel de glucosa en la sangre, en ayunas, puede considerarse que tiene distribución aproximadamente normal, con media de 126 mg/100ml y desviación estándar de 8 mg/100ml.

1. Hallar la probabilidad de que el nivel de glucosa en la sangre, en ayunas, sea menor de 140 mg/100ml.

2. ¿Qué porcentaje de diabéticos tendrá niveles de glucosa entre 110 mg/100ml y 140 mg/100ml?
3. Encontrar un punto  $x_0$  que tenga la propiedad que el 25% de los diabéticos tenga un nivel de glucosa,  $X$ , por debajo de este valor.

*Solución.* En este caso,  $X \sim N(126, 64)$ , luego:

1. La probabilidad de que el nivel de glucosa en la sangre, en ayunas, sea menor de 140 mg/100ml es:

$$\begin{aligned} P(X < 140) &= P\left(\frac{X - \mu}{\sigma} < \frac{140 - 126}{8}\right) = P\left(\underbrace{\frac{X - \mu}{\sigma}}_Z < \frac{14}{8}\right) \\ &= P(Z < 1,75) = \Phi(1,75) = 0,9599. \end{aligned}$$

Entonces, la probabilidad de que un diabético tenga niveles de glucosa, en ayunas, por debajo de 140 mg/100ml es del 96 %.

2. El porcentaje de diabéticos que tendrán niveles entre 110 mg/100ml y 140 mg/100ml es:

$$\begin{aligned} P(110 < X < 140) &= P\left(\frac{110 - 126}{8} < \frac{X - \mu}{\sigma} < \frac{140 - 126}{8}\right) \\ &= P\left(\frac{-16}{8} < \underbrace{\frac{X - \mu}{\sigma}}_Z < \frac{14}{8}\right) = P(-2 < Z < 1,75) \\ &= \Phi(1,75) - \Phi(-2) = 0,9599 - 0,0228 = 0,9371. \end{aligned}$$

Entonces, el porcentaje de diabéticos que tendrá niveles de glucosa, en ayunas, entre 110 mg/100ml y 140 mg/100ml es del 94 %.

3. El punto  $x_0$  se establece como:

$$\begin{aligned} P(X < x_0) &= 0,25 \\ P\left(\underbrace{\frac{X - \mu}{\sigma}}_Z < \underbrace{\frac{x_0 - 126}{8}}_{z_0}\right) &= 0,25 \\ P(Z < z_0) &= 0,25 \quad \rightarrow \quad z_0 = -0,675 \\ &\rightarrow \quad \frac{x_0 - 126}{8} = -0,675, \\ \text{luego } x_0 &= (-0,675)(8) + 126 = 120,6. \end{aligned}$$

Por tanto, el 25 % de los diabéticos tienen un nivel de glucosa, en ayunas, por debajo de 120,6 mg/100ml.  $\square$

**Nota 1.11. Interpolación.** Suponga que  $x + \alpha = 100z$

$$\Phi\left(\frac{x + \alpha}{100}\right) = \alpha\Phi\left(\frac{x + 1}{100}\right) + (1 - \alpha)\Phi\left(\frac{x}{100}\right)$$

donde  $x$  es entero y  $\alpha$  es decimal. Ejemplo:

$$\Phi(1,3678) = 0,78\Phi(1,37) + (1 - 0,78)\Phi(1,36).$$

**DEFINICIÓN 1.16.** De acuerdo con [21, p. 16], sea  $Z_i \sim N(0, 1)$ ,  $i = 1, 2, \dots, n$ ,  $n$  variables aleatorias independientes. Si  $R = \sum_{i=1}^n Z_i^2$ , entonces,  $R$  tiene distribución chi-cuadrada con  $n$  grados de libertad.

**Ejemplo 1.49.** (Distribución gamma). Para [16, p. 156], una variable aleatoria  $X$  tiene distribución gamma de parámetros  $r$  y  $\lambda$ , con  $r > 0$  y  $\lambda > 0$ , si su *fdp* está dada por:

$$f_X(x) = \frac{\lambda}{\Gamma(r)} (\lambda x)^{r-1} e^{-\lambda x}; x > 0, \quad (1.25)$$

donde  $\Gamma(\cdot)$  es dada en (1.23). Notación:  $X \sim \text{gamma}(r, \lambda)$ . Determine su *fda*.

*Solución.* Su *fda*  $F_X$  es dada por:

$$F_X(x) = P(X \leq x) = \frac{\lambda}{\Gamma(r)} \int_0^x (\lambda t)^{r-1} e^{-\lambda t} dt.$$

Para  $r = n$  con  $n \in \mathbb{N}$ , realizando la integral anterior por partes, se obtiene la siguiente *fda* para la correspondiente variable aleatoria gamma:

$$F_X(x) = \frac{\lambda}{\Gamma(n)} \int_0^x (\lambda t)^{n-1} e^{-\lambda t} dt = 1 - \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!} e^{-\lambda x}. \quad (1.26)$$

Esta expresión es similar a la *fda* de una Poisson. ☑

**Nota 1.12.** Casos particulares de la distribución gamma:

- Cuando  $r = 1$  y  $\lambda > 0$ , la variable aleatoria  $X$  tiene **distribución exponencial** de parámetro  $\lambda$ . Notación:  $X \sim \text{Exp}(\lambda)$ .

$$f_X(x) = \frac{\lambda}{\Gamma(1)} (\lambda x)^{1-1} e^{-\lambda x} = \lambda e^{-\lambda x}; x > 0. \quad (1.27)$$

En la figura 1.12 se presentan la *fdp* y la *fda* para  $\lambda = 0,2$  y  $\lambda = 0,1$ , respectivamente.

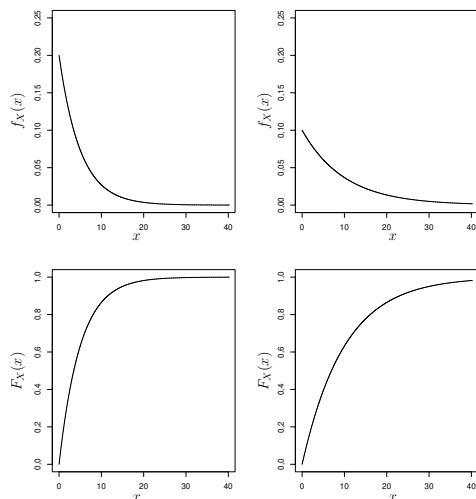


Figura 1.12. **fdp** y la **fda** exponencial. (a)  $\lambda = 0,2$  y (b)  $\lambda = 0,1$

- Cuando  $r = \frac{k}{2}$  y  $\lambda = \frac{1}{2}$ , la variable aleatoria  $X$  tiene **distribución chi-cuadrada** con  $k$  grados de libertad. Notación:  $X \sim \chi^2_{(k)}$ .

$$\begin{aligned}
 f_X(x) &= \frac{\frac{1}{2}}{\Gamma\left(\frac{k}{2}\right)} \left(\frac{1}{2}x\right)^{\frac{k}{2}-1} e^{-1/2x} \\
 &= \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} x^{k/2-1} e^{-x/2}; x > 0.
 \end{aligned} \tag{1.28}$$

En la figura 1.13 se presentan la **fdp** y la **fda** para  $r = 2,5$  ( $k = 5$ ) y  $r = 7,5$  ( $k = 15$ ), respectivamente.

En la figura 1.14 se presentan la **fdp** y la **fda** para  $r = 11,11$  y  $\lambda = 2,22$ ,  $r = 4$  y  $\lambda = 0,8$ , respectivamente.

En la figura 1.15 se presentan la **fdp** y la **fda** para  $r = 11,11$  y  $\lambda = 5,55$ ,  $r = 277,78$  y  $\lambda = 27,78$ , respectivamente.

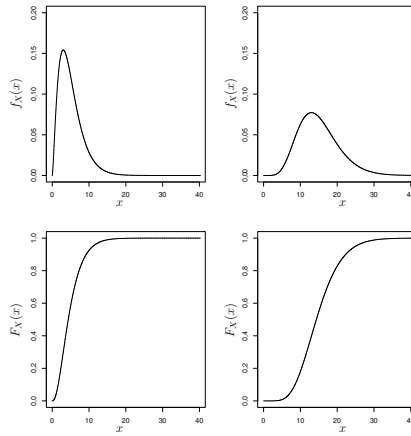


Figura 1.13. **pdf** y **fda** chi-cuadrada. (a)  $r = 2,5$  ( $k = 5$ ) y (b)  $r = 7,5$  ( $k = 15$ )

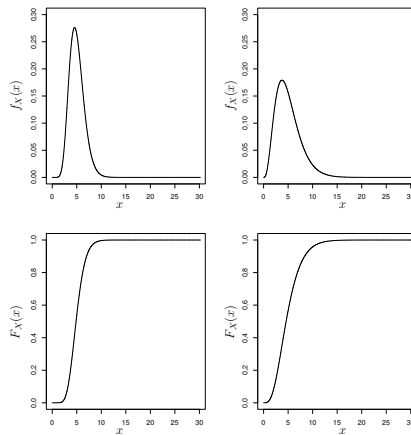


Figura 1.14. **pdf** y **fda** Gamma. (a)  $r = 11,11$  y  $\lambda = 2,22$  y (b)  $r = 4$  y  $\lambda = 0,8$

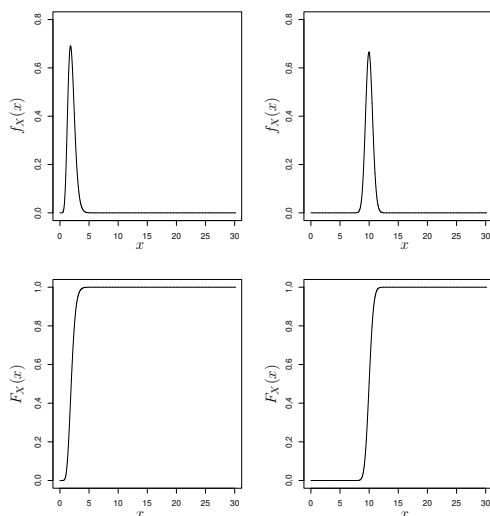


Figura 1.15. **fdp** y la **fda** gamma. (a)  $r = 11,11$  y  $\lambda = 5,55$  y (b)  $r = 277,78$  y  $\lambda = 27,78$

**DEFINICIÓN 1.17. Distribución log-normal.** Si  $X$  tiene **fdp** log-normal con parámetros  $\mu$  y  $\sigma^2$ , su **fdp** está dada por:

$$f_X(x) = \frac{1}{x\sqrt{2\pi}\sigma \ln C} \exp\left\{-\frac{1}{2}\left(\frac{\log_C x - \mu}{\sigma}\right)^2\right\}, \quad (1.29)$$

donde  $x > 0$ ,  $\sigma > 0$  y  $\log_C x$  es el logaritmo en base  $C$  ( $C > 1$ ) y se denota como  $X \sim \mathcal{LN}(\mu, \sigma^2)$ . La mayoría de textos considera  $C = e$ .

**Nota 1.13.** Para graficar diferentes distribuciones cuando su(s) parámetro(s) cambian, se puede usar la instrucción en R:

```
remove(list=ls())
library(gamlss.demo)
demoDist()
```

**Ejemplo 1.50.** En los ejemplos 1.35, 1.37, 1.39 y 1.41, la variable aleatoria considerada es discreta. En los ejemplos 1.46, 1.15 y 1.49, la variable aleatoria considerada es continua.

**DEFINICIÓN 1.18.** (Valor esperado). Según [16, p. 74], dada  $X$  una variable aleatoria real definida sobre el espacio de probabilidad  $(\Omega, \mathcal{F}, P)$ .

- Si  $X$  es una variable aleatoria discreta, con valores  $x_1, x_2, x_3, \dots$ , se dice que  $X$  posee un valor esperado si:

$$\sum_{k=1}^{\infty} |x_k| P(X = x_k) < \infty. \quad (1.30)$$

El valor esperado  $\mathbb{E}(X)$  (esperanza matemática, media) de  $X$  se define como:

$$\mathbb{E}(X) := \sum_{k=1}^{\infty} x_k P(X = x_k). \quad (1.31)$$

- Si  $X$  es una variable aleatoria continua con *fdp*  $f_X$ , se dice que  $X$  posee un valor esperado si:

$$\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty. \quad (1.32)$$

El valor esperado  $\mathbb{E}(X)$  (esperanza matemática, media) de  $X$  se define como:

$$\mathbb{E}(X) := \int_{-\infty}^{\infty} x f_X(x) dx. \quad (1.33)$$

**Teorema 1.6. Propiedades del valor esperado.** Dadas las variables aleatorias  $X, Y$ , la función  $\mathbb{E}(\cdot)$  tiene las siguientes propiedades:

- i) Si  $X \geq 0$  entonces  $\mathbb{E}(X) \geq 0$ .
- ii)  $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$ .
- iii)  $\mathbb{E}(cX) = c\mathbb{E}(X)$ , para toda  $c \in \mathbb{R}$ .
- iv)  $\mathbb{E}(X \pm Y) = \mathbb{E}(X) \pm \mathbb{E}(Y)$ .
- v) Si  $X \leq Y$  entonces  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ , (esto es, el valor esperado es monótonamente creciente).
- vi) Si  $X, Y$  son independientes y existe  $\mathbb{E}(XY)$ ,

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

*Demostración.* Ver detalles de la prueba en [43]. ☑

**Ejemplo 1.51.** Considerando la información del ejemplo 1.35, es decir, una variable aleatoria  $X$  con distribución uniforme discreta de parámetro  $n$ , determine su valor esperado.

*Solución.* Su valor esperado es dado por la expresión (1.31)

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} x_k P(X = x_k) = \sum_{k=1}^n k \frac{1}{n} = \frac{1}{n} \sum_{k=1}^n k = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}. \quad \checkmark$$

**Ejemplo 1.52.** Considerando la información del ejemplo 1.37, es decir, una variable aleatoria  $X$  con distribución binomial de parámetros  $m$  y  $p$ , determine su valor esperado.

*Solución.* Su valor esperado es dado por la expresión (1.31)

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=1}^{\infty} x_k P(X = x_k) = \sum_{k=0}^m k \binom{m}{k} p^k (1-p)^{m-k} = \sum_{k=1}^m \frac{km! p^k (1-p)^{m-k}}{(m-k)! k!} \\ &= \sum_{k=1}^m \frac{m! p^k (1-p)^{m-k}}{(m-k)! (k-1)!} = \sum_{k=1}^m \frac{m(m-1)!}{(m-k)! (k-1)!} p p^{k-1} (1-p)^{m-k} \\ &= mp \sum_{k=1}^m \frac{(m-1)!}{(m-k)! (k-1)!} p^{k-1} (1-p)^{m-k},\end{aligned}$$

considerando  $j = k - 1$ , cabe notar que los nuevos límites de la sumatoria son: cuando  $k = 1$  se tiene que  $j = 1 - 1 = 0$  y cuando  $k = m$  se tiene que  $j = m - 1 = n$ , luego:

$$\mathbb{E}(X) = mp \underbrace{\sum_{j=0}^n \frac{n!}{(n-j)! j!} p^j (1-p)^{n-j}}_1 = mp. \quad \checkmark$$

**Ejemplo 1.53.** Considerando la información del ejemplo 1.39, es decir, una variable aleatoria  $X$  con distribución Poisson de parámetro  $\lambda$ , determine su valor esperado.

*Solución.* Su valor esperado es dado por la expresión (1.31)

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=1}^{\infty} x_k P(X = x_k) = \sum_{k=0}^{\infty} k \frac{e^{-\lambda} \lambda^k}{k!} = \sum_{k=1}^{\infty} \frac{k e^{-\lambda} \lambda^k}{k!} \\ &= \sum_{k=1}^{\infty} \frac{e^{-\lambda} \lambda \lambda^{k-1}}{(k-1)!},\end{aligned}$$

considerando  $j = k - 1$  y haciendo los respectivos cambios, se tiene que:

$$\mathbb{E}(X) = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda e^{-\lambda} e^{\lambda} = \lambda,$$

empleando la expresión (A.2) dada en el apéndice A. \checkmark

**Ejemplo 1.54.** Considerando la información del ejemplo 1.41, es decir, una variable aleatoria  $X$  con distribución geométrica de parámetro  $p$ , determine su valor esperado.

*Solución.* Su valor esperado es dado por la expresión (1.31)

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=1}^{\infty} x_k P(X = x_k) = \sum_{k=0}^{\infty} k p (1-p)^k = p \sum_{k=0}^{\infty} k (1-p)^{k-1} (1-p) \\ &= p(1-p) \sum_{k=1}^{\infty} \underbrace{k (1-p)^{k-1}}_r.\end{aligned}$$

Derivando<sup>4</sup> ambos lados de la expresión (1.15) se llega a:

$$\sum_{k=1}^{\infty} k r^{k-1} = \frac{1}{(1-r)^2} \quad \text{si } |r| < 1. \quad (1.34)$$

Al sustituir esta última expresión en  $\mathbb{E}(X)$  se obtiene:

$$\mathbb{E}(X) = p(1-p) \frac{1}{(1-(1-p))^2} = p(1-p) \frac{1}{p^2} = \frac{(1-p)}{p}. \quad \checkmark$$

**Ejemplo 1.55.** Considerando la información del ejemplo 1.45, es decir, una variable aleatoria  $X$  con distribución hipergeométrica de parámetros  $M$ ,  $m$  y  $n$ , determine su valor esperado.

*Solución.* Su valor esperado es dado por la expresión (1.31)

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=1}^{\infty} x_k P(X = x_k) = \sum_{k=0}^n k \frac{\binom{m}{k} \binom{M-m}{n-k}}{\binom{M}{n}} = \sum_{k=1}^n \frac{km!}{(m-k)!k!} \frac{\binom{M-m}{n-k}}{\binom{M}{n}} \\ &= \sum_{k=1}^n \frac{m(m-1)!}{(m-k)!(k-1)!} \frac{\binom{M-m}{n-k}}{\binom{M}{n}} = m \sum_{k=1}^n \frac{\binom{m-1}{k-1} \binom{M-m}{n-k}}{\frac{M}{n} \binom{M-1}{n-1}} = \frac{nm}{M} \sum_{k=1}^n \frac{\binom{m-1}{k-1} \binom{M-m}{n-k}}{\binom{M-1}{n-1}}.\end{aligned}$$

Sea  $j = k - 1$  los nuevos límites de la sumatoria son: cuando  $k = 1$  se tiene que  $j = 1 - 1 = 0$  y cuando  $k = n$  se tiene que  $j = n - 1$ , es decir,

$$\mathbb{E}(X) = \frac{nm}{M} \sum_{j=0}^{n-1} \frac{\binom{m-1}{j} \binom{M-m}{n-(j+1)}}{\binom{M-1}{n-1}} = \frac{nm}{M} \frac{1}{\binom{M-1}{n-1}} \sum_{j=0}^{n-1} \binom{m-1}{j} \binom{M-m}{n-1-j}.$$

<sup>4</sup>En [20, p. 70], teorema 2.4.3 se puede consultar las condiciones para diferenciar bajo el operador de integración. Para condiciones para diferenciar bajo el operador de sumatoria consultar [20, p. 74], teorema 2.4.8.

Por la “identidad de Vandermonde” o “convolución de Vandermonde”, cuyo nombre le fue dado en honor del matemático francés Alexandre-Théophile Vandermonde (1772), se tiene que:

$$\binom{n}{r} = \sum_{k=0}^r \binom{m}{k} \binom{n-m}{r-k}, \quad m, n, r \in \mathbb{N}_0. \quad (1.35)$$

Luego, el valor esperado:

$$\mathbb{E}(X) = \frac{nm}{M} \underbrace{\frac{1}{\binom{M-1}{n-1}} \sum_{j=0}^{n-1} \binom{m-1}{j} \binom{M-m}{n-1-j}}_1 = \frac{nm}{M}. \quad \square$$

**Ejemplo 1.56.** Considerando la información del ejemplo 1.46, es decir, una variable aleatoria  $X$  con distribución uniforme sobre el intervalo  $[a, b]$ , con  $a < b$  números reales, determine su valor esperado.

*Solución.* Su valor esperado es dado por la expresión (1.33)

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x dx = \left( \frac{1}{b-a} \right) \frac{x^2}{2} \Big|_a^b \\ &= \frac{1}{b-a} \left( \frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{(b-a)(a+b)}{2(b-a)} = \frac{a+b}{2}. \quad \square \end{aligned}$$

**Ejemplo 1.57.** Sea una variable aleatoria  $X$  con distribución exponencial de parámetro  $\lambda$ , ver (1.27), determine su valor esperado.

*Solución.* El valor esperado de  $X$  es dado por la expresión (1.33)

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} \lambda x e^{-\lambda x} dx.$$

Esta integral se evalúa mediante cambio de variable, sea  $u = \lambda x$  entonces  $du = \lambda dx$ , ahora si  $x = 0$  entonces  $u = 0$  y si  $x \rightarrow \infty$  entonces  $u \rightarrow \infty$ ; luego:

$$\int_0^{\infty} \lambda x e^{-\lambda x} dx = \int_0^{\infty} u e^{-u} \left( \frac{1}{\lambda} \right) du = \frac{1}{\lambda} \int_0^{\infty} u e^{-u} du,$$

usando la integración por partes (ver [5, p. 266]), sea  $v = u$  y  $dw = e^{-u} du$ , entonces  $dv = du$  y  $w = -e^{-u}$

$$\int_0^{\infty} u e^{-u} du = -u e^{-u} \Big|_{u=0}^{u=\infty} + \int_0^{\infty} e^{-u} du = -\frac{u}{e^u} \Big|_{u=0}^{u=\infty} - e^{-u} \Big|_{u=0}^{u=\infty}, \quad (1.36)$$

por tanto:

$$\mathbb{E}(X) = -\frac{1}{\lambda} \lim_{k \rightarrow \infty} \left( \frac{u+1}{e^u} \Big|_{u=0}^{u=k} \right) = -\frac{1}{\lambda} \left( \lim_{k \rightarrow \infty} \frac{k+1}{e^k} \right) + \frac{1}{\lambda},$$

usando la regla de L'Hôpital (ver nota B.9):

$$\mathbb{E}(X) = -\frac{1}{\lambda} \lim_{k \rightarrow \infty} \left( \frac{k+1}{e^k} \right) + \frac{1}{\lambda} = -\frac{1}{\lambda} \lim_{k \rightarrow \infty} \frac{1}{e^k} + \frac{1}{\lambda} = \frac{1}{\lambda}. \quad \checkmark$$

**DEFINICIÓN 1.19.** De acuerdo con [20, p. 55], si  $X$  es una variable aleatoria y  $g$  una función real tal que  $g(X)$  es también una variable aleatoria. Entonces el valor esperado de  $g(X)$  está dado por:

$$\mathbb{E}[g(X)] = \begin{cases} \sum_x g(x)P[X = x], & \text{si } X \text{ es discreta,} \\ \int_{-\infty}^{\infty} g(x)f_X(x)dx, & \text{si } X \text{ es continua.} \end{cases} \quad (1.37)$$

El valor esperado de una variable aleatoria  $X$  es un número fijo y una característica de la *fdp* de  $X$ , luego, no es una función de  $X$ . Nótese que el valor esperado existe si la respectiva suma (o integral) converge a un valor finito. En algunos textos, esto se conoce como la ley del estadístico inconsciente.

**DEFINICIÓN 1.20.** ( $n$ -ésimo momento de  $X$  (o  $F_X(x)$ )). Según [20, p. 59], para cada entero  $n$ , el  $n$ -ésimo momento ordinario de  $X$  (o  $F_X(x)$ ),  $\mu'_n(X)$ , es

$$\mu'_n(X) = \mathbb{E}(X^n).$$

El  $n$ -ésimo momento central de  $X$ ,  $\mu_n(X)$ , es

$$\mu_n(X) = \mathbb{E}[(X - \mu)^n],$$

con  $\mu = \mu'_1(X) = \mathbb{E}(X)$ .

**DEFINICIÓN 1.21.** Según [88, p. 144], el momento factorial de orden  $r$  de una variable aleatoria de valores enteros,  $X$ , se define como:

$$\mu_{[r]}(X) = \mathbb{E} \left[ \frac{\Gamma(X+1)}{\Gamma(X+1-r)} \right] = \sum_{k=r}^{\infty} \frac{\Gamma(k+1)}{\Gamma(k+1-r)} P[X = k], \quad (1.38)$$

donde  $r \in \mathbb{N}$  y  $\Gamma(\cdot)$  es la función gamma dada en (1.23).

**Ejemplo 1.58.** Considerando la información del ejemplo 1.35, es decir, una variable aleatoria  $X$  con *fmp* uniforme de parámetro  $n$ , determine  $\mu'_2(X)$ .

*Solución.*

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{k=1}^{\infty} x_k^2 P(X = x_k) = \sum_{k=1}^n k^2 \frac{1}{n} = \frac{1}{n} \sum_{k=1}^n k^2 = \frac{1}{n} \frac{n(n+1)(2n+1)}{6} \\ &= \frac{(n+1)(2n+1)}{6}.\end{aligned}\quad \square$$

**Ejemplo 1.59.** Considerando la información del ejemplo 1.37, es decir, una variable aleatoria  $X$  con distribución binomial de parámetros  $m$  y  $p$  (ver [19, p. 95]), determine  $\mu_{[2]}(X)$ .

*Solución.*

$$\begin{aligned}\underbrace{\mathbb{E}(X^2 - X)} &= \sum_{k=1}^{\infty} (x_k^2 - x_k) P(X = x_k) = \sum_{k=0}^m (k^2 - k) \binom{m}{k} p^k (1-p)^{m-k} \\ \mu_{[2]}(X) &= \sum_{k=2}^m \frac{k(k-1)m!}{(m-k)!k!} p^k (1-p)^{m-k} = \sum_{k=2}^m \frac{m!}{(m-k)!(k-2)!} p^k (1-p)^{m-k} \\ &= \sum_{k=2}^m \frac{m(m-1)(m-2)!}{(m-k)!(k-2)!} p^2 p^{k-2} (1-p)^{m-k},\end{aligned}$$

y tomando  $j = k - 2$ , los nuevos límites de la sumatoria son: cuando  $k = 2$  se tiene que  $j = 2 - 2 = 0$  y cuando  $k = m$  se tiene que  $j = m - 2 = n$ , luego:

$$\mu_{[2]}(X) = m(m-1)p^2 \underbrace{\sum_{j=0}^n \frac{n!}{(n-j)!j!} p^j (1-p)^{n-j}}_1 = m(m-1)p^2, \quad (1.39)$$

aquí se usó la fórmula del binomio de Newton dada en (A.4). \(\square\)

**Ejemplo 1.60.** Considerando la información del ejemplo 1.39, es decir, una variable aleatoria  $X$  con distribución Poisson de parámetro  $\lambda$  (ver [19, p. 106]), determine  $\mu_{[2]}(X)$ .

*Solución.*

$$\begin{aligned}\mu_{[2]}(X) &= \mathbb{E}(X^2 - X) = \sum_{k=1}^{\infty} (x_k^2 - x_k) P(X = x_k) = \sum_{k=0}^{\infty} (k^2 - k) \frac{e^{-\lambda} \lambda^k}{k!} \\ &= \sum_{k=2}^{\infty} \frac{k(k-1)e^{-\lambda} \lambda^k}{k!} = \sum_{k=2}^{\infty} \frac{e^{-\lambda} \lambda^2 \lambda^{k-2}}{(k-2)!} = e^{-\lambda} \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!},\end{aligned}$$

considerando  $j = k - 2$  y haciendo los respectivos cambios, se tiene:

$$\mu_{[2]}(X) = \mathbb{E}(X^2 - X) = \lambda^2 e^{-\lambda} \underbrace{\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}}_{e^{\lambda}} = \lambda^2, \quad (1.40)$$

aquí se empleó la serie de Taylor de la exponencial, expresión (A.2).  $\square$

**Ejemplo 1.61.** Considerando la información del ejemplo 1.41, es decir, una variable aleatoria  $X$  con distribución geométrica de parámetro  $p$ , determine  $\mu_{[2]}(X)$ .

*Solución.*

$$\begin{aligned} \mu_{[2]}(X) &= \mathbb{E}(X^2 - X) = \sum_{k=1}^{\infty} (x_k^2 - x_k) P(X = x_k) = \sum_{k=0}^{\infty} k(k-1)p(1-p)^k \\ &= p \sum_{k=1}^{\infty} k(k-1)(1-p)^{k-1}(1-p) = p(1-p)^2 \sum_{k=2}^{\infty} k(k-1) \underbrace{(1-p)^{k-2}}_r. \end{aligned}$$

Derivando ambos lados de la expresión (1.34) se llega a:

$$\sum_{k=2}^{\infty} k(k-1)r^{k-2} = \frac{2}{(1-r)^3} \quad \text{si } |r| < 1. \quad (1.41)$$

Al sustituir esta última expresión en  $\mu_{[2]}(X)$  se obtiene:

$$\mu_{[2]}(X) = p(1-p)^2 \frac{2}{(1-(1-p))^3} = \frac{2(1-p)^2}{p^2}. \quad \square$$

**Ejemplo 1.62.** Considerando la información del ejemplo 1.46, es decir, una variable aleatoria  $X$  con distribución uniforme sobre el intervalo  $[a, b]$ , con  $a < b$  números reales, determine  $\mu'_2(X)$ .

*Solución.*

$$\begin{aligned} \mathbb{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x^2 dx = \left( \frac{1}{b-a} \right) \frac{x^3}{3} \Big|_a^b \\ &= \frac{1}{b-a} \left( \frac{b^3}{3} - \frac{a^3}{3} \right) = \frac{(b-a)(a^2 + ab + b^2)}{3(b-a)} = \frac{a^2 + ab + b^2}{3}. \quad \square \end{aligned}$$

**Ejemplo 1.63.** Considerando la información del ejemplo 1.57, es decir, una variable aleatoria  $X$  con distribución exponencial de parámetro  $\lambda$ , determine  $\mu'_2(X)$ .

*Solución.*

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^{\infty} \lambda x^2 e^{-\lambda x} dx.$$

Haciendo cambio de variable:  $u = \lambda x$  entonces  $du = \lambda dx$ , ahora si  $x = 0$  entonces  $u = 0$  y si  $x \rightarrow \infty$  entonces  $u \rightarrow \infty$ ; luego:

$$\int_0^{\infty} \lambda x^2 e^{-\lambda x} dx = \int_0^{\infty} u \left(\frac{u}{\lambda}\right) e^{-u} \left(\frac{1}{\lambda}\right) du = \frac{1}{\lambda^2} \int_0^{\infty} u^2 e^{-u} du,$$

usando el método de integración tabular (ver [68]). Si se elige  $v = u^2$  y  $w = e^{-u}$ , y se realiza las derivaciones e integraciones respectivas. En la tabla 1.10 se muestran los resultados de este procedimiento.

Tabla 1.10. Método de integración tabular

$j$	Signo	Derivadas $v^{(j)}$	Integrales $w^{(j)}$
0	+	$u^2$	$e^{-u}$
1	-	$2u$	$-e^{-u}$
2	+	$2$	$e^{-u}$
3	-	$0$	$-e^{-u}$

El valor de la integral es la suma de los productos con los signos apropiados de las respectivas derivaciones e integraciones que dan las diagonales descendentes. De lo obtenido en la tabla anterior se obtiene que

$$\int u^2 e^{-u} du = -u^2 e^{-u} - 2u e^{-u} - 2e^{-u} = -[(u + 1)^2 + 1] e^{-u},$$

entonces

$$\int \lambda x^2 e^{-\lambda x} dx = \frac{-1}{\lambda^2} [(\lambda x + 1)^2 + 1] e^{-\lambda x},$$

por tanto

$$\mathbb{E}(X^2) = \lim_{k \rightarrow \infty} \left( - \left[ \frac{(\lambda x + 1)^2 + 1}{\lambda^2 e^{\lambda x}} \right] \Big|_0^k \right) = \frac{-1}{\lambda^2} \lim_{k \rightarrow \infty} \left( \left[ \frac{(\lambda k + 1)^2 + 1}{e^{\lambda k}} \right] - 2 \right),$$

usando la regla de L'Hôpital

$$\begin{aligned}\mathbb{E}(X^2) &= \frac{-1}{\lambda^2} \lim_{k \rightarrow \infty} \left( \frac{(\lambda k + 1)^2 + 1}{e^{\lambda k}} \right) + \frac{2}{\lambda^2} = \frac{-1}{\lambda^2} \lim_{k \rightarrow \infty} \left( \frac{2(\lambda k + 1)\lambda}{\lambda e^{\lambda k}} \right) + \frac{2}{\lambda^2} \\ &= \frac{-1}{\lambda^2} \lim_{k \rightarrow \infty} \left( \frac{2\lambda}{\lambda e^{\lambda k}} \right) + \frac{2}{\lambda^2} = \frac{2}{\lambda^2}. \quad \square\end{aligned}$$

**DEFINICIÓN 1.22.** (Varianza). La *varianza* de una variable aleatoria  $X$  es el segundo momento central de  $X$ . Se denota por:

$$\mathbb{V}(X) := \mathbb{E}\{[X - \mathbb{E}(X)]^2\}. \quad (1.42)$$

A la raíz cuadrada de  $\mathbb{V}(X)$  se le llama *desviación estándar* de  $X$ .

**Teorema 1.7.** Propiedades de la varianza. La función  $\mathbb{V}(X)$  tiene las siguientes propiedades:

- i)  $\mathbb{V}(X) = 0$  si y solo si  $X$  es constante.
- ii)  $\mathbb{V}(cX) = c^2 \mathbb{V}(X)$ , para toda  $c \in \mathbb{R}$ .
- iii) Dadas  $X, Y$  variables aleatorias y  $a, b \in \mathbb{R}$ , entonces:

$$\mathbb{V}(aX \pm bY) = a^2 \mathbb{V}(X) + b^2 \mathbb{V}(Y) \pm 2ab \mathbb{Cov}(X, Y), \quad (1.43)$$

donde (ver [20, p. 169]):

$$\begin{aligned}\mathbb{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\end{aligned} \quad (1.44)$$

es una relación conocida como “covarianza”.

Esta función covarianza, para cualesquiera variables aleatorias  $X, Y, Z$  y  $a, b \in \mathbb{R}$ , cumple las siguientes propiedades:

- i)  $\mathbb{Cov}(X, a) = \mathbb{Cov}(a, X) = 0$ .
- ii)  $\mathbb{Cov}(X, Y) = \mathbb{Cov}(Y, X)$ . Si  $X = Y$  entonces  $\mathbb{Cov}(X, X) = \mathbb{V}(X)$ .
- iii)  $\mathbb{Cov}(X, Y) = 0$ , si las variables  $X, Y$  son independientes<sup>5</sup>.
- iv)  $\mathbb{Cov}(aX + b, Y) = a\mathbb{Cov}(X, Y) = \mathbb{Cov}(X, aY + b)$ .
- v)  $\mathbb{Cov}(X + Y, Z) = \mathbb{Cov}(X, Z) + \mathbb{Cov}(Y, Z)$ .
- vi)  $\mathbb{Cov}(X, Y + Z) = \mathbb{Cov}(X, Y) + \mathbb{Cov}(X, Z)$ .

*Demostración.* Ver detalles de la prueba en [43]. □

<sup>5</sup>El recíproco no es válido, ya que hay variables aleatorias con  $\mathbb{Cov}(X, Y) = 0$ , pero dependientes, un ejemplo se puede consultar en [20, p. 171].

**Ejemplo 1.64.** Considerando la información del ejemplo 1.35, es decir, una variable aleatoria  $X$  con distribución uniforme discreta de parámetro  $n$ , calcular  $\mathbb{V}(X)$ .

*Solución.* Usando los resultados de los ejemplo 1.51 y 1.58,

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}\{[X - \mathbb{E}(X)]^2\} = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \\ &= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 \\ &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n+1}{2} \left[ \frac{2n+1}{3} - \frac{n+1}{2} \right] \\ &= \frac{n+1}{2} \left[ \frac{4n+2-3n-3}{6} \right] = \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12}. \quad \checkmark\end{aligned}$$

**Ejemplo 1.65.** Considerando la información del ejemplo 1.37, es decir, una variable aleatoria  $X$  con distribución binomial de parámetros  $m$  y  $p$ , calcular  $\mathbb{V}(X)$ .

*Solución.* Usando los resultados de los ejemplo 1.52 y 1.59,

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}\{[X - \mathbb{E}(X)]^2\} = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \\ &= \mathbb{E}(X^2 - X) + \mathbb{E}(X) - [\mathbb{E}(X)]^2 = m(m-1)p^2 + mp - m^2p^2 \\ &= -mp^2 + mp = mp(1-p). \quad \checkmark\end{aligned}$$

**Ejemplo 1.66.** Considerando la información del ejemplo 1.39, es decir, una variable aleatoria  $X$  con distribución Poisson de parámetro  $\lambda$ , calcular  $\mathbb{V}(X)$ .

*Solución.* Usando los resultados de los ejemplo 1.53 y 1.60,

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}\{[X - \mathbb{E}(X)]^2\} = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \\ &= \mathbb{E}(X^2 - X) + \mathbb{E}(X) - [\mathbb{E}(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \quad \checkmark\end{aligned}$$

**Ejemplo 1.67.** Considerando la información del ejemplo 1.41, es decir, una variable aleatoria  $X$  con distribución geométrica de parámetro  $p$ , calcular  $\mathbb{V}(X)$ .

*Solución.* Usando los resultados de los ejemplo 1.54 y 1.61,

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}\{[X - \mathbb{E}(X)]^2\} = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \\ &= \mathbb{E}(X^2 - X) + \mathbb{E}(X) - [\mathbb{E}(X)]^2 = \frac{2(1-p)^2}{p^2} + \frac{1-p}{p} - \left(\frac{1-p}{p}\right)^2 \\ &= \frac{1-p}{p} \left( \frac{2(1-p)}{p} + 1 - \frac{1-p}{p} \right) = \frac{1-p}{p} \left( \frac{1-p}{p} + \frac{p}{p} \right) = \frac{1-p}{p^2}. \quad \checkmark\end{aligned}$$

**Ejemplo 1.68.** Considerando la información del ejemplo 1.46, es decir, una variable aleatoria  $X$  con distribución uniforme sobre el intervalo  $[a, b]$ , con  $a < b$  números reales, calcular  $\mathbb{V}(X)$ .

*Solución.* Usando los resultados de los ejemplos 1.56 y 1.62,

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}\{[X - \mathbb{E}(X)]^2\} = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \\ &= \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12} \\ &= \frac{a^2 - 2ab + b^2}{12} = \frac{(b-a)^2}{12}. \quad \checkmark \end{aligned}$$

**Ejemplo 1.69.** Considerando la información del ejemplo 1.57, es decir, una variable aleatoria  $X$  con distribución exponencial de parámetro  $\lambda$ , calcular  $\mathbb{V}(X)$ .

*Solución.* Usando el resultado del ejemplo 1.63,

$$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \quad \checkmark$$

**Ejemplo 1.70.** Según [20, p. 56], una variable aleatoria  $X$  tiene distribución Cauchy estándar si su *fdp* es:

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad -\infty < x < \infty. \quad (1.45)$$

Muestre que esta variable aleatoria  $X$  no tiene valor esperado.

*Demostración.* Por la expresión (1.32) se tiene que:

$$\int_{-\infty}^{\infty} |x|f_X(x)dx = \int_{-\infty}^{\infty} \frac{|x|}{\pi} \frac{1}{1+x^2} dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{|x|}{1+x^2} dx,$$

como  $\frac{|x|}{1+x^2}$  es una función par, entonces:

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{|x|}{1+x^2} dx &= 2 \int_0^{\infty} \frac{x}{1+x^2} dx = \lim_{C \rightarrow \infty} \int_0^C \frac{2x}{1+x^2} dx \\ &= \lim_{C \rightarrow \infty} \left( \ln(1+x^2) \Big|_0^C \right) = \infty. \end{aligned}$$

Así,

$$\int_{-\infty}^{\infty} |x|f_X(x)dx = \infty,$$

y por tanto la variable aleatoria  $X$  no tiene valor esperado. \checkmark

**DEFINICIÓN 1.23. Medidas estadísticas.** Si  $X$  es una variable aleatoria otras medidas principales son [45]:

$$\begin{aligned}\mathcal{K}_X &= \frac{\mu_2(X) - \mu_1'(X)}{\mu_1'(X)} = \mathcal{J}_X - 1 && \text{Razón de Katz} \\ CV_X &= \frac{\sqrt{\mu_2(X)}}{\mu_1'(X)} && \text{Coeficiente de variación} \\ \beta_1(X) &= \frac{\mu_3(X)}{\sqrt{\mu_2^3(X)}} && \text{Coeficiente de asimetría} \\ \beta_2(X) &= \frac{\mu_4(X)}{\mu_2^2(X)} && \text{índice de curtosis}\end{aligned}$$

donde  $\mathcal{J}_X$  denota el índice de dispersión de  $X$ . En general,

$$\beta_{r-2}(X) = \frac{\mu_r(X)}{\sqrt{\mu_2^r(X)}} \quad r = 3, 4, \dots$$

La expresión  $CV_X$  mide la dispersión de una variable aleatoria en relación con la media. El término  $\beta_1(X)$  es una medida de la asimetría de una **fdp** sobre su dispersión. Una **fdp** es asimétrica positivamente, negativamente o simétrica si  $\beta_1(X) > 0$ ,  $\beta_1(X) < 0$  o  $\beta_1(X) = 0$ , respectivamente.

De acuerdo con [19, p. 71], la curtosis,  $\beta_2(X)$ , es una medida que nos indica qué tan *puntiaguda* es la **fdp**. Una **fdp** presenta un pico relativamente alto, relativamente bajo (relativamente plana) o no exhibe un pico muy alto ni muy bajo si  $\beta_2(X) > 3$ ,  $\beta_2(X) < 3$  o  $\beta_2(X) = 3$ , respectivamente. En [20, p. 79],  $\beta_2(X)$ , se define como la **curtosis** estandarizada.

Cualquier variable aleatoria  $X$  satisface la siguiente desigualdad dada en [92],

$$\beta_2(X) - \beta_1^2(X) - 1 \geq 0.$$

## 1.9. Aproximaciones de distribuciones de probabilidad

**Teorema 1.8.** De acuerdo con [19, p. 104], si  $X$  es una variable aleatoria con distribución binomial y **fmp** dada por (1.12), tal que la relación  $p = \lambda/m$  para  $m = 1, 2, 3, \dots$ , es válida para alguna constante  $\lambda > 0$ , entonces:

$$\lim_{m \rightarrow \infty} P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x = 0, 1, 2, 3, \dots$$

*Demostración.* Ver detalles de la prueba en [19, p. 128]. ☑

**Ejemplo 1.71.** De acuerdo con [19, p. 104], un comprador de grandes cantidades de circuitos integrados ha adoptado un plan para aceptar un envío de estos que consiste en inspeccionar una muestra aleatoria de 100 circuitos provenientes del lote. Si el comprador encuentra no más de dos circuitos defectuosos en la muestra, acepta el lote; de otra forma, lo rechaza. Si se envía al comprador un lote que contiene 1% de circuitos defectuosos, ¿cuál es la probabilidad de que éste sea aceptado?

*Solución.* Sea  $X :=$  “número de circuitos defectuosos encontrados en una muestra de tamaño 100”. Luego,  $X \sim \text{Bin}(100, 0,01)$  y la probabilidad de aceptar el lote sería

$$P(X \leq 2) = \binom{100}{0} \left(\frac{1}{100}\right)^0 \left(1 - \frac{1}{100}\right)^{100-0} + \binom{100}{1} \left(\frac{1}{100}\right)^1 \left(1 - \frac{1}{100}\right)^{100-1} + \binom{100}{2} \left(\frac{1}{100}\right)^2 \left(1 - \frac{1}{100}\right)^{100-2} = 0,9206.$$

Dado que  $m = 100$  es relativamente un valor grande y  $p = 0,01$  es pequeño; la probabilidad binomial se puede aproximar mediante la distribución de Poisson, escogiendo  $\lambda = mp = 1$ :

$$P(X \leq 2) = \frac{1^0}{0!} e^{-1} + \frac{1^1}{1!} e^{-1} + \frac{1^2}{2!} e^{-1} = 2,5e^{-1} = 0,919698. \quad \checkmark$$

**Nota 1.14.** La fórmula de Stirling es una aproximación para factoriales grandes. Lleva el nombre en honor al matemático escocés del siglo XVIII, James Stirling. Esta fórmula es dada por (ver [8, p. 24]):

$$n! = n\Gamma(n) \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n = \sqrt{2\pi} e^{-n} n^{n+1/2}. \quad (1.46)$$

**Teorema 1.9.** De acuerdo con [19, p. 141], sea  $X$  una variable aleatoria con distribución binomial y  $f_{mp}$  dada por (1.12), si  $m \rightarrow \infty$  y, ni  $p$  ni  $q = 1 - p$  son muy pequeños y, además cuando:

$$p \begin{cases} < \frac{1}{2} & \text{se satisface que } mpq > 9 \text{ o } mp > 5 \\ > \frac{1}{2} & \text{se cumple que } mq > 5, \end{cases}$$

entonces:

$$X \sim N(mp, mpq).$$

*Demostración.* Ver detalles de la prueba en [19, p. 182]. ☑

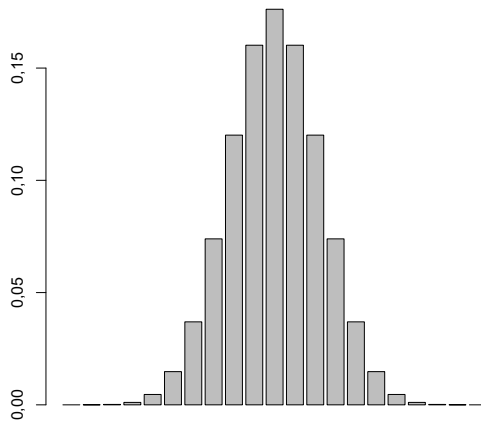


Figura 1.16. fmp binomial con  $m = 20$  y  $p = 0,5$

La corrección por continuidad es un ajuste que se realiza cuando una distribución discreta se aproxima mediante una distribución continua, en la tabla 1.11 se muestra la respectiva corrección:

Tabla 1.11. Corrección por continuidad

Probabilidad	Aproximación
$P(X = x)$	$P(x - 0,5 < X < x + 0,5)$
$P(X > x)$	$P(X > x + 0,5)$
$P(X \geq x)$	$P(X > x - 0,5)$
$P(X < x)$	$P(X < x - 0,5)$
$P(X \leq x)$	$P(X < x + 0,5)$

**Ejemplo 1.72.** De acuerdo con [19, p. 143], un partido político planea llevar a cabo una encuesta para detectar la preferencia de los votantes con respecto a los candidatos  $A$  y  $B$  que ocuparán un cargo en la administración pública. Supóngase que toma una muestra aleatoria de mil ciudadanos, ¿cuál es la probabilidad de que 550 o más de los votantes indiquen una preferencia por el candidato  $A$  si la población, con respecto a los candidatos, se encuentra igualmente dividida?

**Solución.** Sea  $X :=$  “número de ciudadanos que muestran preferencia por el candidato  $A$ ”. Luego,  $X \sim Bin(1000, 0,5)$ , pero dado que  $m = 1000$  es

relativamente un valor grande y  $p = 0,5$  no es pequeño; la probabilidad binomial se puede aproximar mediante la distribución normal, escogiendo  $\mu = mp = 500$ ,  $\sigma^2 = mpq = 250$  y usando la corrección por continuidad:

$$\begin{aligned} P(X \geq 550) &= P(X > 549,5) = P\left(\frac{X - \mu}{\sigma} > \frac{549,5 - 500}{\sqrt{250}}\right) \\ &= P(Z > 3,13) = 1 - P(Z \leq 3,13) = 1 - \Phi(3,13) = 0,0009. \end{aligned}$$

Luego, se concluye que  $A$  será el ganador en la encuesta. ☑

## 1.10. Desigualdad de Chebyshev

**Teorema 1.10.** (Desigualdad básica). *Sea  $X$  una variable aleatoria cualquiera y  $g$  una función no negativa definida sobre  $\mathbb{R}$ . Si  $g$  es par y es no decreciente sobre  $[0, \infty)$ , para cada  $a > 0$  se tiene que:*

$$P(|X| \geq a) \leq \frac{\mathbb{E}[g(X)]}{g(a)}. \tag{1.47}$$

*Si  $g$  es no decreciente sobre  $\mathbb{R}$  y  $P(|X| \geq a)$  es reemplazada por  $P(X \geq a)$ , (1.47) es cierta para todo real  $a$ .*

*Demostración.* Ver detalles de la prueba en [11, p. 98]. ☑

Un caso especial, que se conoce como la *desigualdad de Chebyshev*, es cuando  $g(x) = |x - c|^\alpha$  para  $\alpha > 0$  y  $c \in \mathbb{R}$ . En este caso, sea  $X$  variable aleatoria y  $\alpha > 0$  con  $\mathbb{E}(|X|^\alpha) < \infty$ . Entonces,  $\forall c \in \mathbb{R}$  y  $\forall \epsilon > 0$ ,

$$P(|X - c| \geq \epsilon) \leq \frac{\mathbb{E}(|X - c|^\alpha)}{\epsilon^\alpha}. \tag{1.48}$$

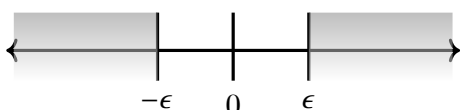
*Veamos:*

$$\begin{aligned} |X - c| &\leq |X| + |c| \leq 2 \max\{|X|, |c|\} \\ |X - c|^\alpha &\leq 2^\alpha (\max\{|X|, |c|\})^\alpha = 2^\alpha \max\{|X|^\alpha, |c|^\alpha\} \leq 2^\alpha [|X|^\alpha + |c|^\alpha], \end{aligned}$$

aplicando el operador esperanza se tiene que:

$$\mathbb{E}(|X - c|^\alpha) \leq \mathbb{E}(2^\alpha [|X|^\alpha + |c|^\alpha]) < \infty \quad (\text{hipótesis}).$$

Por otro lado,

$$\mathbb{E}(|X - c|^\alpha) = \int_{-\infty}^{\infty} |x - c|^\alpha dF(x) \underbrace{\geq}_{\text{subconj.}} \int_{\{|x-c| \geq \epsilon\}} |x - c|^\alpha dF(x)$$


Si  $|x - c|^\alpha \geq \epsilon^\alpha$  entonces<sup>6</sup>:

$$\int_{\{|x-c| \geq \epsilon\}} |x - c|^\alpha dF(x) \geq \int_{\{|x-c| \geq \epsilon\}} \epsilon^\alpha dF(x).$$

Luego,

$$\int_{\{|x-c| \geq \epsilon\}} |x - c|^\alpha dF(x) \geq \int_{\{|x-c| \geq \epsilon\}} \epsilon^\alpha dF(x) = \epsilon^\alpha \underbrace{\int_{\{|x-c| \geq \epsilon\}} dF(x)}_{P(|X-c| \geq \epsilon)}$$

$$\mathbb{E}(|X - c|^\alpha) \geq \epsilon^\alpha P(|X - c| \geq \epsilon)$$

$$P(|X - c| \geq \epsilon) \leq \frac{\mathbb{E}(|X - c|^\alpha)}{\epsilon^\alpha}. \quad \checkmark$$

**Nota 1.15.** Si  $c = \mu = \mathbb{E}(X)$  y  $\alpha = 2$  en (1.48), entonces:

$$P(|X - \mu| \geq \epsilon) \leq \frac{\mathbb{E}(|X - \mu|^2)}{\epsilon^2} = \frac{\mathbb{V}(X)}{\epsilon^2}.$$

**Ejemplo 1.73.** Sea  $X$  una variable aleatoria con  $\mathbb{E}(X) = \mu$  y  $\mathbb{V}(X) = \sigma^2$ , aplique la desigualdad de Chebyshev, para:

1.  $\epsilon = 1$

2.  $\epsilon = 10$

3.  $\epsilon = t\sigma$ .

*Solución.*

1. Para  $\epsilon = 1$  se tiene que  $P(|X - \mu| \geq 1) \leq \mathbb{V}(X)$ .

<sup>6</sup>Dadas dos funciones  $f(x)$  y  $g(x)$ , entonces:

$$\int g(x)f(x) dx \quad \begin{cases} \geq \epsilon \int f(x) dx & \text{si } g(x) \geq \epsilon \\ \leq \epsilon \int f(x) dx & \text{si } g(x) \leq \epsilon \end{cases}$$

2. Si  $\epsilon = 10$ , entonces  $P(|X - \mu| \geq 10) \leq \frac{V(X)}{10^2} = \frac{V(X)}{100}$ .

3. Si  $\epsilon = t\sigma$ , se tiene que:

$$\begin{aligned} P(|X - \mu| \geq \epsilon) &\leq \frac{V(X)}{\epsilon^2} &\Leftrightarrow P(|X - \mu| \geq t\sigma) &\leq \frac{\sigma^2}{(t\sigma)^2} &\Leftrightarrow \\ P(|X - \mu| \geq t\sigma) &\leq \frac{1}{t^2} &\Leftrightarrow 1 - P(|X - \mu| < t\sigma) &\leq \frac{1}{t^2} &\Leftrightarrow \\ -1 + P(|X - \mu| < t\sigma) &\geq \frac{-1}{t^2} &\Leftrightarrow P(|X - \mu| < t\sigma) &\geq 1 - \frac{1}{t^2}. &\checkmark \end{aligned}$$

**Ejemplo 1.74.** Sea  $X$  una variable aleatoria tal que su media es 20 y su desviación estándar es 5. Si se obtienen realizaciones de esta variable aleatoria:

1. ¿Al menos qué porcentaje de las realizaciones de la variable aleatoria se espera que estén entre 10 y 30?
2. ¿Al menos qué porcentaje de las realizaciones se espera que estén entre 0 y 40?

*Solución.*

1. Para determinar la probabilidad:

$$\begin{aligned} P(10 < X < 30) &= P(10 - \mu < X - \mu < 30 - \mu) \\ &= P(10 - 20 < X - 20 < 30 - 20) \\ &= P(-10 < X - 20 < 10) = P(|X - 20| < 10) \\ &= P(|X - 20| < \underbrace{5}_{\sigma} \underbrace{(2)}_t) \\ &\geq 1 - \frac{1}{t^2} = 1 - \frac{1}{2^2} = 0,75, \end{aligned}$$

por tanto, al menos el 75 % de las realizaciones de la variable aleatoria se espera que estén entre 10 y 30.

2. En este caso, la probabilidad solicitada queda

$$\begin{aligned} P(0 < X < 40) &= P(0 - \mu < X - \mu < 40 - \mu) \\ &= P(0 - 20 < X - 20 < 40 - 20) \\ &= P(-20 < X - 20 < 20) = P(|X - 20| < 20) \\ &= P(|X - 20| < \underbrace{5}_{\sigma} \underbrace{(4)}_t) \\ &\geq 1 - \frac{1}{t^2} = 1 - \frac{1}{4^2} = 0,9375, \end{aligned}$$

por tanto, al menos el 93,75 % de las realizaciones de la variable aleatoria se espera que estén entre 0 y 40. ✓

## 1.11. Funciones generadoras

En esta sección se estudian las tres funciones generadoras: de probabilidad, momentos y característica, las cuales corresponden a niveles crecientes de generalidad. La primera es la más restrictiva, pero también, por mucho, la más simple, ya que la teoría se reduce a hechos básicos sobre series de potencias temáticas de cálculo. La tercera es la más general y para la que la teoría es más completa y elegante, pero también requiere conocimientos básicos de análisis y variable compleja. La del medio es quizás la que se usa con más frecuencia y es suficiente para la mayoría de las distribuciones en probabilidad aplicada.

### 1.11.1. Función generadora de probabilidad

Dada una función generadora  $P(t) = \sum_k a_k t^k$ , donde los coeficientes  $a_k$  se encuentran mediante derivación y se establecen por la fórmula  $a_k = \frac{P^{(k)}(0)}{k!}$  (coeficientes de la serie de Taylor, ver apéndice A), en la práctica, puede resultar imposible la obtención de expresiones explícitas, sin embargo, en algunas ocasiones resulta más sencillo usar una función auxiliar,  $H_X(t) = t^X$ , y calcular su esperanza.

**DEFINICIÓN 1.24.** Sea  $X$  una variable aleatoria y  $0 \leq t < 1$ , la “función generadora de probabilidad (fgp)” se define como el valor esperado (cuando exista) de la variable aleatoria  $t^X$ ; esto es,

$$\underbrace{G_X(t)}_{\mathbb{E}(t^X)} = \begin{cases} \sum_x t^x P[X = x], & \text{si } X \text{ es discreta,} \\ \int_{-\infty}^{\infty} t^x f_X(x) dx, & \text{si } X \text{ es continua.} \end{cases} \quad (1.49)$$

La mayor utilidad de la fgp se presenta cuando  $X$  es una variable aleatoria discreta, luego se consideran solo las fgp para este tipo de variables.

**Teorema 1.11. Propiedades de la fgp.** Según [45, p. 58], si  $X$  es una variable aleatoria de valores enteros cuya fgp existe, entonces:

1.  $G_X(1) = 1$ .
2. La probabilidad del evento  $\{X = r\}$  se establece como la  $r$ -ésima derivada de  $G_X(t)$  evaluada en  $t = 0$  dividido por  $r!$ , es decir:

$$P(X = r) = \frac{1}{r!} \left. \frac{d^r}{dt^r} G_X(t) \right|_{t=0} = \frac{G_X^{(r)}(0)}{r!}, \quad (1.50)$$

siendo  $G_X^{(0)}(0) \equiv G_X(0)$ .

3. Si  $Y = aX + b$  con  $a$  y  $b \in \mathbb{R}$ , entonces:

$$G_Y(t) = t^b G_X(t^a).$$

4. Si  $X$  y  $Y$  son variables aleatorias independientes, entonces:

$$G_{X+Y}(t) = G_X(t)G_Y(t), \quad 0 \leq t < 1.$$

5. Si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes, entonces:

$$G_{\sum_{k=1}^n X_k}(t) = \prod_{k=1}^n G_{X_k}(t), \quad \forall 0 \leq t < 1.$$

6. El  $r$ -ésimo momento factorial de  $X$  se establece como la  $r$ -ésima derivada de  $G_X(t)$  evaluada en  $t = 1$ , es decir:

$$\mu_{[r]}(X) = \left. \frac{d^r}{dt^r} G_X(t) \right|_{t=1} = G_X^{(r)}(1). \quad (1.51)$$

7. **Unicidad.** Si  $X$  y  $Y$  son dos variables aleatorias cuyas *fgp* existen, y son iguales, entonces ellas tienen la misma *fmp*.

*Demostración.* En todas las propiedades, basta con realizar los respectivos reemplazos y se obtienen las pruebas:

1. Por la definición se tiene que:

$$G_X(t) = \sum_{k=0}^{\infty} t^k P[X = k], \quad (1.52)$$

claramente si se toma  $t = 1$  entonces  $G_X(1) = 1$ .

2. Al derivar  $r$  veces la expresión (1.52) respecto a  $t$ , se llega a:

$$\begin{aligned} G_X^{(r)}(t) &= \frac{d^r}{dt^r} G_X(t) = \sum_{k=r}^{\infty} \frac{k!}{(k-r)!} t^{k-r} P[X = k] \\ &= r! \left\{ P(X = r) + \sum_{k=r+1}^{\infty} \binom{k}{k-r} t^{k-r} P[X = k] \right\}. \end{aligned} \quad (1.53)$$

Nótese que si  $t = 0$ , esta última expresión queda:

$$G_X^{(r)}(0) = r! P(X = r).$$

Se llega a lo mismo usando el desarrollo en serie de Taylor de la función  $G_X(t)$ :

$$G_X(t) = \sum_{k=0}^{\infty} \frac{G_X^{(k)}(0)}{k!} t^k.$$

En otras palabras, la función  $G_X(t)$  se define por las  $P_X[x]$  y, a su vez, esta define la  $P(X = r)$  desde una expansión polinómica que es única.

3. Nótese que:

$$G_{aX+b}(t) = \mathbb{E}\left(t^{(aX+b)}\right) = \mathbb{E}\left(t^{aX} t^b\right) = t^b \underbrace{\mathbb{E}\left[(t^a)^X\right]}_{G_X(t^a)} = t^b G_X(t^a).$$

4. En este caso, como  $X$  y  $Y$  son independientes:

$$G_{X+Y}(t) = \mathbb{E}\left(t^{(X+Y)}\right) = \mathbb{E}\left(t^X t^Y\right) \underbrace{=}_{\text{Indep.}} \underbrace{\mathbb{E}(t^X)}_{G_X(t)} \underbrace{\mathbb{E}(t^Y)}_{G_Y(t)} = G_X(t) G_Y(t).$$

5. Similar al caso anterior.

6. Si se hace  $t = 1$  en la expresión (1.53) se llega al resultado.

7. Este resultado es un caso especial del teorema de unicidad para series de potencias. ☑

**Ejemplo 1.75.** Considerando la información del ejemplo 1.37, esto es, una variable aleatoria  $X$  con fmp binomial de parámetros  $m$  y  $p$ , determine  $G_X(t)$  y  $\mu_{[2]}(X)$ .

*Solución.* Su fgp es dada por:

$$\begin{aligned} G_X(t) &= \mathbb{E}(t^X) = \sum_{k=0}^m t^k \binom{m}{k} p^k (1-p)^{m-k} = \sum_{k=0}^m \binom{m}{k} (pt)^k (1-p)^{m-k} \\ &= [pt + (1-p)]^m, \end{aligned}$$

aquí se empleó la expresión (A.4) dada en el apéndice A. Las respectivas derivadas de esta función son:

$$\begin{aligned} \frac{d}{dt} G_X(t) &= mp [pt + (1-p)]^{m-1} \\ \frac{d^2}{dt^2} G_X(t) &= m(m-1)p^2 [pt + (1-p)]^{m-2} \\ &\vdots \\ \frac{d^r}{dt^r} G_X(t) &= \frac{m!}{(m-r)!} p^r [pt + (1-p)]^{m-r}. \end{aligned}$$

Usando la expresión (1.51) con  $r = 2$ :

$$\mu_{[2]}(X) = \left. \frac{d^2}{dt^2} G_X(t) \right|_{t=1} = \frac{m!}{(m-2)!} p^2 [p + (1-p)]^{m-2} = m(m-1)p^2,$$

la cual coincide con la expresión dada en (1.39). ✓

**Ejemplo 1.76.** Considerando la información del ejemplo 1.39, esto es, una variable aleatoria  $X$  con distribución Poisson de parámetro  $\lambda$ , determine  $G_X(t)$  y  $\mu_{[2]}(X)$ .

*Solución.* Su fgp es dada por:

$$\begin{aligned} G_X(t) &= \mathbb{E}(t^X) = \sum_{k=0}^{\infty} t^k \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{t^k \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(t\lambda)^k}{k!} \\ &= e^{-\lambda} e^{t\lambda} = \exp \{ \lambda(t-1) \}, \end{aligned}$$

aquí se usó la expresión (A.2) dada en el apéndice A. Las respectivas derivadas de esta función son:

$$\begin{aligned} \frac{d}{dt} G_X(t) &= \lambda \exp \{ \lambda(t-1) \} \\ \frac{d^2}{dt^2} G_X(t) &= \lambda^2 \exp \{ \lambda(t-1) \} \\ &\vdots \\ \frac{d^r}{dt^r} G_X(t) &= \lambda^r \exp \{ \lambda(t-1) \}. \end{aligned}$$

Usando la expresión (1.51) con  $r = 2$ :

$$\mu_{[2]}(X) = \left. \frac{d^2}{dt^2} G_X(t) \right|_{t=1} = \lambda^2,$$

la cual coincide con la expresión dada en (1.40). ✓

### 1.11.2. Función generadora de momentos

**DEFINICIÓN 1.25.** (Función generadora de momentos). Según [16, p. 87], si  $X$  es una variable aleatoria tal que  $\mathbb{E}(e^{tX})$  es finito para todo  $t \in (-\alpha, \alpha)$ , con  $\alpha$  real positivo. Entonces se define la función generadora de momentos (fgm) de  $X$ , denotada por  $m_X(\cdot)$  como:

$$m_X(t) = \mathbb{E}(e^{tX}), \quad \text{con } t \in (-\alpha, \alpha).$$

**Teorema 1.12. Propiedades de la fgm.** Si  $X$  tiene fgm  $m_X(t)$ , entonces:

1.  $m_X(0) = 1$ .
2.  $m_X(t)$  es continua en todo su dominio.
3. Si  $Y = aX + b$  con  $a$  y  $b \in \mathbb{R}$ , entonces:

$$m_Y(t) = e^{bt} m_X(at), \quad \forall t \in \mathbb{R}.$$

4. Si  $X$  y  $Y$  son variables aleatorias independientes, entonces:

$$m_{X+Y}(t) = m_X(t)m_Y(t) \quad \text{para todo } t \in \mathbb{R}.$$

5. Si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes, entonces:

$$m_{\sum_{k=1}^n X_k}(t) = \prod_{k=1}^n m_{X_k}(t), \quad \forall t \in \mathbb{R}.$$

6. El  $r$ -ésimo momento de  $X$  se establece como la  $r$ -ésima derivada de  $m_X(t)$  evaluada en  $t = 0$ , es decir:

$$\mathbb{E}(X^r) = m_X^{(r)}(0) = \left. \frac{d^r}{dt^r} m_X(t) \right|_{t=0} \quad r = 1, 2, \dots$$

7. **Unicidad.** Si  $X$  y  $Y$  son dos variables aleatorias cuyas funciones generadoras de momentos existen, y son iguales, entonces ellas tienen la misma fda.

*Demostración.* En todas las propiedades, basta con realizar los respectivos reemplazos y se obtienen las pruebas:

1. Por la definición de la fgm:

$$m_X(0) = \mathbb{E}(e^{0 \cdot X}) = \mathbb{E}(e^0) = 1.$$

2. Este resultado es consecuencia del teorema de convergencia dominada.
3. Ver [20, p. 68].
4. Nótese que:

$$m_{X+Y}(t) = \mathbb{E}(e^{t(X+Y)}) = \mathbb{E}(e^{tX} e^{tY}) \underset{\text{Indep.}}{=} \underbrace{\mathbb{E}(e^{tX})}_{m_X(t)} \underbrace{\mathbb{E}(e^{tY})}_{m_Y(t)} = m_X(t)m_Y(t).$$

5. Si  $m_{X_j}(t)$ ,  $j = 1, 2, 3, \dots, n$  son las respectivas funciones generadoras de momentos de las  $X_j$ , entonces la fgm de la suma está dada por:

$$m_S(t) = \mathbb{E}[e^{tS}] = \mathbb{E}[e^{tX_1} e^{tX_2} \dots e^{tX_n}],$$

donde

$$S = X_1 + X_2 + X_3 + \dots + X_n. \quad (1.54)$$

Si las variables aleatorias  $X_j$  son independientes, se tiene que:

$$m_S(t) = m_{X_1}(t)m_{X_2}(t) \dots m_{X_n}(t) = \prod_{j=1}^n m_{X_j}(t), \quad (1.55)$$

y si además, las  $X_j$  son idénticamente distribuidas, es decir, las  $m_{X_j}(t)$  son iguales para  $t$  en alguna vecindad de cero, entonces:

$$m_S(t) = [m_X(t)]^n. \quad (1.56)$$

6. *Caso continuo*: suponiendo que se puede diferenciar bajo el signo de integración, para  $r = 1$ , se tiene que [20, p. 62]:

$$\begin{aligned} \frac{d}{dt} m_X(t) &= \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f_X(x) dx = \int_{-\infty}^{\infty} \left( \frac{d}{dt} e^{tx} \right) f_X(x) dx \\ &= \int_{-\infty}^{\infty} (x e^{tx}) f_X(x) dx = \mathbb{E}(X e^{tX}), \end{aligned}$$

evaluando en  $\frac{d}{dt} m_X(t)$  en  $t = 0$  se tiene que:

$$\left. \frac{d}{dt} m_X(t) \right|_{t=0} = \mathbb{E}(X e^{tX}) \Big|_{t=0} = \mathbb{E}(X e^0) = \mathbb{E}(X).$$

De forma similar, para  $r = 2$ ,

$$\begin{aligned} \frac{d^2}{dt^2} m_X(t) &= \frac{d}{dt} \left[ \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f_X(x) dx \right] = \int_{-\infty}^{\infty} \frac{d}{dt} (e^{tX}) x f_X(x) dx \\ &= \int_{-\infty}^{\infty} x^2 e^{tx} f_X(x) dx = \mathbb{E}(X^2 e^{tX}), \end{aligned}$$

evaluando en  $\frac{d^2}{dt^2} m_X(t)$  en  $t = 0$  se tiene que:

$$\left. \frac{d^2}{dt^2} m_X(t) \right|_{t=0} = \mathbb{E}(X^2 e^{tX}) \Big|_{t=0} = \mathbb{E}(X^2 e^0) = \mathbb{E}(X^2),$$

etc.

7. Ver detalles de la prueba en [83, p. 106]. ☑

**Ejemplo 1.77.** Considerando la información del ejemplo 1.35, esto es, una variable aleatoria  $X$  con distribución uniforme discreta de parámetro  $n$ , determine  $m_X(t)$ .

*Solución.* La fgm está dada por:

$$m_X(t) = \mathbb{E}(e^{tX}) = \sum_{k=1}^n e^{tk} \frac{1}{n} = \frac{1}{n} \sum_{k=1}^n e^{tk}.$$

Como la suma parcial de la serie geométrica es dada por (ver [5, p. 474]):

$$\sum_{k=0}^n x^k = \frac{1 - x^{n+1}}{1 - x} \quad \text{si } x \neq 1,$$

tomando en esta expresión  $x = e^t$  con  $t \neq 0$  se tiene que:

$$\sum_{k=0}^n (e^t)^k = \frac{1 - (e^t)^{n+1}}{1 - e^t} \quad \Rightarrow \quad 1 + \sum_{k=1}^n e^{kt} = \frac{1 - e^{(n+1)t}}{1 - e^t}.$$

Luego,

$$\sum_{k=1}^n e^{kt} = \frac{1 - e^{(n+1)t}}{1 - e^t} - 1 = \frac{e^t - e^{(n+1)t}}{1 - e^t} = \frac{e^t (1 - e^{nt})}{1 - e^t}, \quad t \neq 0,$$

por tanto:

$$m_X(t) = \frac{e^t (e^{nt} - 1)}{n(e^t - 1)} \quad \text{para } t \neq 0. \quad \checkmark$$

**Ejemplo 1.78.** Considerando la información del ejemplo 1.37, esto es, una variable aleatoria  $X$  con distribución binomial de parámetros  $m$  y  $p$ , determine  $m_X(t)$ .

*Solución.* La fgm es dada por:

$$\begin{aligned} m_X(t) &= \mathbb{E}(e^{tX}) = \sum_{k=0}^m e^{tk} \binom{m}{k} p^k (1-p)^{m-k} \\ &= \sum_{k=0}^m \binom{m}{k} (pe^t)^k (1-p)^{m-k} = [pe^t + (1-p)]^m, \end{aligned}$$

aquí se usó la expresión (A.4) dada en el apéndice A. \checkmark

**Ejemplo 1.79.** Considerando la información del ejemplo 1.39, esto es, una variable aleatoria  $X$  con distribución Poisson de parámetro  $\lambda$ , determine  $m_X(t)$ .

*Solución.* La fgm es dada por:

$$m_X(t) = \mathbb{E}(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{e^{tk} \lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t \lambda)^k}{k!}$$

$$= e^{-\lambda} e^{(e^t \lambda)} = \exp \{ \lambda [\exp(t) - 1] \},$$

aquí se usó la expresión (A.2) dada en el apéndice A. ☑

**Ejemplo 1.80.** Considerando la información del ejemplo 1.41, esto es, una variable aleatoria  $X$  con distribución geométrica de parámetro  $p$ , determine  $m_X(t)$ .

*Solución.* La fgm es dada por:

$$\begin{aligned} m_X(t) &= \mathbb{E}(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} p(1-p)^k = p \underbrace{\sum_{k=0}^{\infty} [(1-p)e^t]^k}_{\substack{\text{Serie geométrica,} \\ \text{expresión (1.15)}}} \\ &= \underbrace{p \left( \frac{1}{1-(1-p)e^t} \right)}_{\text{Si } |(1-p)e^t| < 1} \end{aligned}$$

entonces si:

$$\begin{aligned} (1-p)e^t < 1 &\Leftrightarrow e^t < \frac{1}{1-p}, \quad p > 0, \\ t < \ln \left( \frac{1}{1-p} \right) &\Leftrightarrow t < -\ln(1-p), \end{aligned}$$

luego,

$$m_X(t) = \frac{p}{1 - (1-p)e^t}, \quad \text{para } t < -\ln(1-p). \quad \checkmark$$

**Ejemplo 1.81.** Considerando la información del ejemplo 1.46, esto es, la variable aleatoria  $X \sim U(a, b)$ , con  $a < b$  números reales, determine  $m_X(t)$ .

*Solución.* La fgm es dada por:

$$\begin{aligned} m_X(t) &= \mathbb{E}(e^{tX}) = \int_a^b e^{tx} \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b e^{tx} dx = \left( \frac{1}{b-a} \right) \frac{e^{tx}}{t} \Bigg|_a^b \\ &= \frac{1}{b-a} \left( \frac{e^{tb}}{t} - \frac{e^{ta}}{t} \right) = \frac{1}{(b-a)t} (e^{tb} - e^{ta}), \quad \text{para } t \neq 0. \quad \checkmark \end{aligned}$$

**Ejemplo 1.82.** Considerando la información del ejemplo 1.15, esto es, la variable aleatoria  $X \sim N(\mu, \sigma^2)$ , con  $\mu$  número real y  $\sigma^2 > 0$ , determine  $m_X(t)$ .

*Solución.* La fgm es dada por:

$$\begin{aligned} m_X(t) &= \mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} \exp\{tx\} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (x^2 - 2\mu x + \mu^2) + tx \right\} dx \end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx)\right\} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{x^2 - 2x(\mu + \sigma^2 t) + \mu^2 + (\mu + \sigma^2 t)^2 - (\mu + \sigma^2 t)^2}{2\sigma^2}\right\} dx,
\end{aligned}$$

el término que se suma y resta en el exponente es para completar cuadrados:

$$\begin{aligned}
m_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left\{[x - (\mu + \sigma^2 t)]^2 + \mu^2 - (\mu + \sigma^2 t)^2\right\}\right\} dx \\
&= \exp\left\{\frac{(\mu + \sigma^2 t)^2 - \mu^2}{2\sigma^2}\right\} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[x - (\mu + \sigma^2 t)]^2}{2\sigma^2}\right\} dx}_1 \\
&= \exp\left\{\frac{\mu^2 + 2\sigma^2 t\mu + \sigma^4 t^2 - \mu^2}{2\sigma^2}\right\} = \exp\left\{t\mu + \frac{\sigma^2 t^2}{2}\right\}. \quad \square
\end{aligned}$$

**Ejemplo 1.83.** Considerando la información del ejemplo 1.49, esto es, una variable aleatoria  $X$  con distribución gamma de parámetros  $r$  y  $\lambda$ , con  $r > 0$  y  $\lambda > 0$ , determine  $m_X(t)$ .

*Solución.* La fgm es dada por:

$$\begin{aligned}
m_X(t) &= \mathbb{E}(e^{tX}) = \int_0^{\infty} e^{tx} \frac{\lambda}{\Gamma(r)} (\lambda x)^{r-1} e^{-\lambda x} dx \\
&= \frac{\lambda^r}{\Gamma(r)} \int_0^{\infty} x^{r-1} e^{tx-\lambda x} dx = \frac{\lambda^r}{\Gamma(r)} \int_0^{\infty} x^{r-1} e^{-x(\lambda-t)} dx,
\end{aligned}$$

considerando que  $\lambda - t > 0$  y haciendo el cambio de variable  $u = (\lambda - t)x$ ,  $du = (\lambda - t)dx$  se tiene

$$\begin{aligned}
m_X(t) &= \frac{\lambda^r}{\Gamma(r)} \int_0^{\infty} \left(\frac{u}{\lambda-t}\right)^{r-1} e^{-u} \frac{du}{\lambda-t} = \frac{\lambda^r}{\Gamma(r)} \frac{1}{(\lambda-t)^r} \int_0^{\infty} u^{r-1} e^{-u} du \\
&= \frac{\lambda^r}{(\lambda-t)^r} \underbrace{\int_0^{\infty} \frac{u^{r-1}}{\Gamma(r)} e^{-u} du}_1 = \left(\frac{\lambda}{\lambda-t}\right)^r = \left(\frac{1}{1-\frac{t}{\lambda}}\right)^r \quad \text{si } t < \lambda. \quad \square
\end{aligned}$$

**Ejemplo 1.84.** Un caso particular del ejemplo 1.83 es cuando  $r = 1$ , esto es,  $X$  tiene distribución exponencial, determine  $m_X(t)$ .

*Solución.* La fgm de  $X$  es dada por:

$$m_X(t) = \left(\frac{1}{1-\frac{t}{\lambda}}\right)^r \underbrace{=}_{\text{Si } r=1} \frac{1}{1-\frac{t}{\lambda}} \quad \text{si } t < \lambda. \quad \square$$

**Ejemplo 1.85.** Un caso particular del ejemplo 1.83 es cuando  $\lambda = \frac{1}{2}$ , esto es,  $X$  tiene distribución chi-cuadrada, determine  $m_X(t)$ .

*Solución.* La fgm de  $X$  es dada por:

$$\begin{aligned} m_X(t) &= \left( \frac{1}{1 - \frac{t}{\lambda}} \right)^r \underset{\text{Si } r = \frac{k}{2}}{=} \left( \frac{1}{1 - \frac{t}{\frac{1}{2}}} \right)^{k/2} \quad \text{si } t < \lambda, \quad \lambda = \frac{1}{2}, \\ &= (1 - 2t)^{-k/2} \quad \text{si } t < \frac{1}{2}. \quad \checkmark \end{aligned}$$

**Ejemplo 1.86.** Considerando la información del ejemplo 1.82 (distribución normal), evaluar  $\mathbb{E}(X)$  y  $\mathbb{E}(X^2)$  a partir de  $m_X(t)$ .

*Solución.*

$$\begin{aligned} \mathbb{E}(X) &= \left. \frac{dm_X(t)}{dt} \right|_{t=0} = \left. \frac{d}{dt} \left( \exp \left\{ t\mu + \frac{\sigma^2 t^2}{2} \right\} \right) \right|_{t=0} \\ &= \exp \left\{ t\mu + \frac{\sigma^2 t^2}{2} \right\} \left( \mu + \frac{2\sigma^2 t}{2} \right) \Big|_{t=0} = \exp \{ (0)\mu + \sigma^2(0)^2 \} (\mu + \sigma^2(0)), \end{aligned}$$

es decir,  $\mathbb{E}(X) = \mu$  y

$$\begin{aligned} \mathbb{E}(X^2) &= \left. \frac{d^2 m_X(t)}{dt^2} \right|_{t=0} = \left. \frac{d^2}{dt^2} \left( \exp \left\{ t\mu + \frac{\sigma^2 t^2}{2} \right\} \right) \right|_{t=0} \\ &= \left. \frac{d}{dt} \left[ \frac{d}{dt} \left( \exp \left\{ t\mu + \frac{\sigma^2 t^2}{2} \right\} \right) \right] \right|_{t=0} = \left. \frac{d}{dt} \left[ \exp \left\{ t\mu + \frac{\sigma^2 t^2}{2} \right\} (\mu + \sigma^2 t) \right] \right|_{t=0} \\ &= \left[ \exp \left\{ t\mu + \frac{\sigma^2 t^2}{2} \right\} (\mu + \sigma^2 t) (\mu + \sigma^2 t) + \exp \left\{ t\mu + \frac{\sigma^2 t^2}{2} \right\} \sigma^2 \right] \Big|_{t=0} \\ &= \exp \left( (0)\mu + \frac{\sigma^2(0)^2}{2} \right) (\mu + \sigma^2(0))^2 + \exp \left( (0)\mu + \frac{\sigma^2(0)^2}{2} \right) \sigma^2 \\ &= \mu^2 + \sigma^2, \end{aligned}$$

luego,

$$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2. \quad \checkmark$$

**Ejemplo 1.87.** Considerando la información del ejemplo 1.83 (distribución gamma), evaluar  $\mathbb{E}(X)$  y  $\mathbb{E}(X^2)$  a partir de  $m_X(t)$ .

*Solución.*

$$\mathbb{E}(X) = \left. \frac{dm_X(t)}{dt} \right|_{t=0} = \left. \frac{d}{dt} \left[ \left( \frac{\lambda}{\lambda - t} \right)^r \right] \right|_{t=0} = \lambda^r \left. \frac{d}{dt} [(\lambda - t)^{-r}] \right|_{t=0}$$

$$\begin{aligned}
&= \lambda^r (-r) (\lambda - t)^{-r-1} (-1) \Big|_{t=0} = r\lambda^r (\lambda - t)^{-r-1} \Big|_{t=0} \\
&= r\lambda^r (\lambda - 0)^{-r-1} = r\lambda^r \lambda^{-r-1} = \frac{r}{\lambda}, \\
\mathbb{E}(X^2) &= \frac{d^2 m_X(t)}{dt^2} \Big|_{t=0} = \frac{d^2}{dt^2} \left[ \left( \frac{\lambda}{\lambda - t} \right)^r \right] \Big|_{t=0} = \frac{d}{dt} \left\{ \lambda^r \frac{d}{dt} [(\lambda - t)^{-r}] \right\} \Big|_{t=0} \\
&= \frac{d}{dt} \left\{ r\lambda^r (\lambda - t)^{-r-1} \right\} \Big|_{t=0} = \left\{ r\lambda^r (-r-1) (\lambda - t)^{-r-1-1} (-1) \right\} \Big|_{t=0} \\
&= r\lambda^r (r+1) \lambda^{-r-2} = \frac{r(r+1)}{\lambda^2},
\end{aligned}$$

luego,

$$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{r(r+1)}{\lambda^2} - \left( \frac{r}{\lambda} \right)^2 = \frac{r}{\lambda^2}.$$

Como casos especiales se tiene  $X \sim \text{Exp}(\lambda)$  ( $r = 1$ ),

$$\mathbb{E}(X) = \frac{1}{\lambda}, \quad \mathbb{E}(X^2) = \frac{2}{\lambda^2} \quad \text{y} \quad \mathbb{V}(X) = \frac{1}{\lambda^2},$$

y  $X \sim \chi_k^2$  ( $\lambda = \frac{1}{2}$  y  $r = \frac{k}{2}$ ),

$$\mathbb{E}(X) = \frac{\frac{k}{2}}{\frac{1}{2}} = k, \quad \mathbb{E}(X^2) = \frac{\frac{k}{2} \left( \frac{k}{2} + 1 \right)}{\left( \frac{1}{2} \right)^2} = k(k+2) \quad \text{y} \quad \mathbb{V}(X) = \frac{\frac{k}{2}}{\left( \frac{1}{2} \right)^2} = 2k. \quad \square$$

**Ejemplo 1.88.** De acuerdo con [65, p. 190], sea  $X$  una variable aleatoria con distribución normal estándar, esto es,  $X \sim N(0, 1)$ . Sea la variable aleatoria  $Y = X^2$ , encuentre la *fdm* de  $Y$ .

*Solución.* La *fgm* de  $Y$  es dada por:

$$\begin{aligned}
m_Y(t) &= \mathbb{E}(e^{tY}) = \mathbb{E}(e^{tX^2}) = \int_{-\infty}^{\infty} \exp\{tx^2\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2 + tx^2\right\} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(1-2t)x^2\right\} dx \\
&= \frac{1}{(1-2t)^{\frac{1}{2}}} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \frac{1}{(1-2t)^{\frac{1}{2}}}} \exp\left\{-\frac{1}{2\left(\frac{1}{1-2t}\right)}x^2\right\} dx}_{1} = (1-2t)^{-1/2},
\end{aligned}$$

por la unicidad de la *fgm*, esta corresponde a una distribución chi-cuadrada con  $k = 1$  grado de libertad,  $Y \sim \chi_{(1)}^2$ . □

En la tabla 1.12 se presenta un resumen de la media, la varianza y la fgm para algunas distribuciones.

Tabla 1.12. Resumen  $\mathbb{E}(X)$ ,  $\mathbb{V}(X)$  y  $m_X(t)$  para algunas distribuciones

Distribución	Parámetros	$\mathbb{E}(X)$	$\mathbb{V}(X)$	$m_X(t)$
Uniforme discreta	$n$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\frac{e^t(e^{nt}-1)}{n(e^t-1)}$
Binomial	$m$ y $p$	$mp$	$mp(1-p)$	$[pe^t + (1-p)]^m$
Poisson	$\lambda$	$\lambda$	$\lambda$	$\exp\{\lambda(e^t-1)\}$
Geométrica	$p$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	$\frac{p}{1-(1-p)e^t}$
Uniforme continua	$a$ y $b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{1}{t(b-a)}(e^{tb}-e^{ta})$
Normal	$\mu$ y $\sigma^2$	$\mu$	$\sigma^2$	$\exp\left\{t\mu + \frac{\sigma^2 t^2}{2}\right\}$
Gamma	$r$ y $\lambda$	$\frac{r}{\lambda}$	$\frac{r}{\lambda^2}$	$\left(\frac{1}{1-t/\lambda}\right)^r$
Exponencial	$\lambda$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{1}{1-t/\lambda}$
chi-cuadrada	$k$	$k$	$2k$	$(1-2t)^{-k/2}$

Fuente: adaptada de [65, p. 541].

**Ejemplo 1.89.** Sea  $X$  una variable aleatoria con fdp Weibull de parámetros  $\lambda$  y  $\alpha$ :

$$f_X(x) = \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha},$$

con  $0 < x < \infty$ ,  $\alpha > 0$  y  $\lambda > 0$ . Determine  $m_X(t)$ .

*Solución.* Una aplicación de las series de Taylor (ver apéndice A):

$$m_X(t) = \mathbb{E}(e^{tX}) = \mathbb{E}\left(\sum_{k=0}^{\infty} \frac{(tX)^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbb{E}(X^k).$$

Nótese que:

$$\mathbb{E}(X^k) = \int_0^{\infty} x^k f_X(x) dx = \lambda \alpha \int_0^{\infty} x^k x^{\alpha-1} e^{-\lambda x^\alpha} dx.$$

Haciendo el cambio de variable  $u = \lambda x^\alpha$ ,  $du = \lambda \alpha x^{\alpha-1} dx$ , se obtiene que:

$$\mathbb{E}(X^k) = \int_0^{\infty} \left(\frac{u}{\lambda}\right)^{\frac{k}{\alpha}} e^{-u} du = \frac{1}{\lambda^{\frac{k}{\alpha}}} \int_0^{\infty} u^{\frac{k}{\alpha}} e^{-u} du = \frac{1}{\lambda^{\frac{k}{\alpha}}} \Gamma\left(\frac{k}{\alpha} + 1\right).$$

Por lo tanto,

$$m_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \frac{1}{\lambda^{\frac{k}{\alpha}}} \Gamma\left(\frac{k}{\alpha} + 1\right) = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{t}{\sqrt[\alpha]{\lambda}}\right)^k \Gamma\left(\frac{k}{\alpha} + 1\right). \quad \square$$

### 1.11.3. Función característica

**DEFINICIÓN 1.26.** De acuerdo con [48, p. 224], si  $X$  es una variable aleatoria, la **función característica** de  $X$  es una función  $\psi_X(t) : \mathbb{R} \rightarrow \mathbb{C}$  ( $\mathbb{C}$  números complejos) definida por:

$$\psi_X(t) := \psi(t) = \mathbb{E}(e^{itX}) \underbrace{=} \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)] \quad \text{con } t \in \mathbb{R}.$$

7

donde  $i = \sqrt{-1}$  se llama la unidad imaginaria.

**Teorema 1.13. Propiedades de la función característica.** Algunas de las propiedades de la función característica son dadas en [55]:

1.  $|\psi_X(t)| \leq 1 \quad \forall t \in \mathbb{R}$ .
2.  $\psi_X(0) = 1$ .
3.  $\overline{\psi_X(t)} = \psi_X(-t)$  para todo  $t \in \mathbb{R}$ , en que  $\overline{\psi_X(t)}$  es el complejo conjugado<sup>8</sup> de  $\psi_X(t)$ .
4.  $\psi_X$  es uniformemente continua.
5. Si  $Y = aX + b$  con  $a$  y  $b \in \mathbb{R}$ , entonces:

$$\psi_Y(t) = e^{ibt} \psi_X(at), \quad \forall t \in \mathbb{R}.$$

6. Si  $X$  y  $Y$  son variables aleatorias independientes, entonces:

$$\psi_{X+Y}(t) = \psi_X(t)\psi_Y(t) \quad \text{para todo } t \in \mathbb{R}.$$

7. Si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes, entonces:

$$\psi_{\sum_{k=1}^n X_k}(t) = \prod_{k=1}^n \psi_{X_k}(t), \quad \forall t \in \mathbb{R}.$$

8. **Fórmula de inversión.** Sean  $F_X(x)$  y  $\psi_X(t)$  la *fda* y la función característica de la variable aleatoria  $X$ , respectivamente. Si  $a = y - h$  y  $b = y + h$  son

<sup>7</sup>Según [5, p. 448], la forma polar del número complejo  $z = x + iy$  es

$$z = r(\cos \theta + i \sen \theta) = r e^{i\theta}, \quad (1.57)$$

donde  $r = \sqrt{x^2 + y^2}$  y  $\tan \theta = \frac{y}{x}$ . Entonces  $e^{itX} = \cos(tX) + i \sen(tX)$ .

<sup>8</sup>Si  $z = x + iy$ , su **complejo conjugado** es  $\bar{z} = x - iy$ . [5, p. 445].

puntos de continuidad de la *fda*, entonces:

$$F_X(b) - F_X(a) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{it} \psi_X(t) dt. \quad (1.58)$$

9. **Unicidad.** La función característica de una variable aleatoria  $X$  determina la *fda* de  $X$ .
10. Una variable aleatoria  $X$  tiene distribución simétrica en torno de cero, es decir,  $P(X \leq x) = P(X \geq -x)$ ,  $\forall x \in \mathbb{R}$ , si y solo si  $\psi_X(t)$  es real,  $\forall t \in \mathbb{R}$ .
11. Si  $\mathbb{E}|X|^n < \infty$ , entonces  $\psi_X$  tiene  $n$  derivadas continuas y

$$\psi_X^{(k)}(t) = \int (ix)^k e^{itx} dF_X(x), \quad k = 1, 2, \dots, n.$$

En particular,  $\psi_X^{(k)}(0) = i^k \mathbb{E}(X^k)$ .

*Demostración.*

1. Por la definición de la función característica

$$\psi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx \quad \Rightarrow \quad |\psi_X(t)| = \left| \int_{-\infty}^{\infty} e^{itx} f_X(x) dx \right|.$$

Por la desigualdad triangular para integrales ([81, p. 85])

$$\left| \int_{-\infty}^{\infty} e^{itx} d[F_X(x)] \right| \leq \int_{-\infty}^{\infty} |e^{itx}| d[F_X(x)],$$

y el valor absoluto (módulo) del número complejo  $e^{itx}$  es uno<sup>9</sup>, ya que:

$$|e^{itx}| = \sqrt{\cos^2(tx) + \sen^2(tx)} = 1.$$

Aquí se usó la identidad trigonométrica  $\sen^2(x) + \cos^2(x) = 1$ . Por lo tanto,

$$|\psi_X(t)| \leq \int_{-\infty}^{\infty} |e^{itx}| d[F_X(x)] = \int_{-\infty}^{\infty} d[F_X(x)] = 1.$$

2. Claramente:

$$\psi_X(0) = \mathbb{E}(e^{i \cdot 0 \cdot X}) = \mathbb{E}(e^0) = 1.$$

3. Puesto que:

$$\begin{aligned} \psi_X(-t) &= \mathbb{E}(e^{-itX}) = \mathbb{E}[\cos(-tX)] + i\mathbb{E}[\sen(-tX)] \\ &= \mathbb{E}[\cos(tX)] - i\mathbb{E}[\sen(tX)] = \overline{\mathbb{E}(e^{itX})} = \overline{\psi_X(t)} \end{aligned}$$

<sup>9</sup>El valor absoluto de un número complejo  $z = x + iy$  es dado por  $|z| = \sqrt{x^2 + y^2}$ .

4. Ver [36, p. 220].

5. Veamos:

$$\psi_{aX+b}(t) = \mathbb{E}(e^{it(aX+b)}) = e^{itb} \mathbb{E}(e^{i(at)X}) = e^{itb} \psi_X(at).$$

6. Nótese que:

$$\psi_{X+Y}(t) = \mathbb{E}(e^{itX} e^{itY}) \underbrace{=}_{\text{Indep.}} \mathbb{E}(e^{itX}) \mathbb{E}(e^{itY}) = \psi_X(t) \psi_Y(t).$$

7. Si  $S = \sum_{k=1}^n X_k$ , entonces:

$$\begin{aligned} \psi_S(t) &= \mathbb{E}(e^{itS}) = \mathbb{E}\left(\exp\left\{it \sum_{k=1}^n X_k\right\}\right) = \mathbb{E}\left(\prod_{k=1}^n e^{itX_k}\right) \\ &= \underbrace{\prod_{k=1}^n}_{\text{Indep.}} \mathbb{E}(e^{itX_k}) = \prod_{k=1}^n \psi_{X_k}(t). \end{aligned}$$

8. Esta propiedad fue demostrada en [31]. Usando la fórmula de Euler se obtiene que (ver [5, p. 454]):

$$\text{sen}(z) = \frac{e^{iz} - e^{-iz}}{2i},$$

luego, factorizando el integrando de la expresión (1.58) queda:

$$F_X(y+h) - F_X(y-h) = \lim_{T \rightarrow \infty} \frac{1}{\pi} \int_{-T}^T \frac{\text{sen}(ht)}{t} e^{-ity} \psi_X(t) dt.$$

Para calcular esta integral impropia, se denota

$$J_T = \frac{1}{\pi} \int_{-T}^T \frac{\text{sen}(ht)}{t} e^{-ity} \psi_X(t) dt,$$

y por la definición de la función característica se tiene:

$$J_T = \frac{1}{\pi} \int_{-T}^T \int_{-\infty}^{\infty} \frac{\text{sen}(ht)}{t} e^{it(x-y)} f_X(x) dx dt.$$

Aquí el teorema de Fubini se puede aplicar para justificar el intercambio de integrales y, además, usando (1.57) se obtiene que:

$$\begin{aligned}
 J_T &= \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-T}^T \frac{\text{sen}(ht)}{t} e^{it(x-y)} f_X(x) dt dx \\
 &= \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-T}^T \frac{\text{sen}(ht)}{t} [\cos[(x-y)t] + i \text{sen}[(x-y)t]] f_X(x) dt dx;
 \end{aligned}$$

como la función coseno es par y la función seno es impar, entonces:

$$J_T = \frac{1}{\pi} \int_{-\infty}^{\infty} \int_0^T \frac{2}{t} \text{sen}(ht) \cos[(x-y)t] f_X(x) dt dx.$$

Por la identidad trigonométrica (ver [5, p. 120])

$$2 \text{sen } u \cos v = \text{sen}(u + v) + \text{sen}(u - v),$$

y las sustituciones  $u = ht$  y  $v = (x - y)t$ , se llega a:

$$\begin{aligned}
 J_T &= \frac{1}{\pi} \int_{-\infty}^{\infty} \int_0^T \frac{1}{t} [\text{sen}(ht + (x - y)t) + \text{sen}(ht - (x - y)t)] f_X(x) dt dx \\
 &= \int_{-\infty}^{\infty} \left[ \int_0^T \frac{\text{sen}[(x - y + h)t]}{\pi t} dt - \int_0^T \frac{\text{sen}[(x - y - h)t]}{\pi t} dt \right] f_X(x) dx \\
 &= \int_{-\infty}^{\infty} [k_T(x - a) - k_T(x - b)] f_X(x) dx = \int_{-\infty}^{\infty} g_T(x) f_X(x) dx,
 \end{aligned}$$

donde  $g_T(x)$  denota la expresión entre corchetes. Se sabe del análisis matemático que la integral impropia de la función seno sobre la línea real positiva es (ver [4, p. 285]):

$$\lim_{T \rightarrow \infty} \int_0^T \frac{\text{sen } w}{w} dw = \frac{\pi}{2},$$

haciendo el cambio de variable  $w = \theta t$ ,  $dw = \theta dt$ , luego para  $\theta \neq 0$ :

$$\lim_{T \rightarrow \infty} \frac{1}{\pi} \int_0^T \frac{\text{sen}(\theta t)}{t} dt = \lim_{T \rightarrow \infty} k_T(\theta) = \frac{1}{2} \text{sign}(\theta) = \frac{1}{2} \frac{|\theta|}{\theta},$$

donde  $\text{sign}(\cdot)$  denota la función signo. Tomando  $\theta = x - y \pm h$ , la convergencia aquí es uniforme con respecto a  $\theta$  donde  $|\theta| > \delta > 0$ . De este hecho se obtiene que:

$$\lim_{T \rightarrow \infty} g_T(x) = \begin{cases} 0 & \text{si } |x - y| > h \\ \frac{1}{2} & \text{si } |x - y| = h \\ 1 & \text{si } |x - y| < h \end{cases}$$

Luego, realizando la partición adecuada, se llega a

$$J_T = \int_{-\infty}^{y-h} g_T(x) f_X(x) dx + \int_{y-h}^{y+h} g_T(x) f_X(x) dx + \int_{y+h}^{\infty} g_T(x) f_X(x) dx,$$

por lo tanto,

$$\lim_{T \rightarrow \infty} J_T = \int_{y-h}^{y+h} f_X(x) dx = F_X(y+h) - F_X(y-h).$$

9. Dividiendo ambos lados de (1.58) por  $b - a = 2h$  y haciendo  $h \rightarrow 0$

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{F_X(y+h) - F_X(y-h)}{2h} &= \lim_{T \rightarrow \infty} \frac{1}{\pi} \int_{-T}^T \left[ \lim_{h \rightarrow 0} \frac{\text{sen}(ht)}{2ht} \right] e^{-ity} \psi_X(t) dt \\ F'_X(y) = f_X(y) &= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T e^{-ity} \psi_X(t) dt. \end{aligned} \quad (1.59)$$

Esta fórmula relaciona la **fdp**  $f_X(x)$  con la función característica  $\psi_X(t)$  y, permite determinar  $f_X(x)$  a partir de  $\psi_X(t)$  bajo el supuesto que  $\psi_X(t)$  es absolutamente integrable.

10. Puesto que  $z = \bar{z}$  si y solo si  $z$  es un número real. De la propiedad 3 se tiene que:

$$\overline{\psi_{-X}(t)} = \psi_{-X}(-t) = \mathbb{E}\left(e^{i(-t)(-X)}\right) = \mathbb{E}(e^{itX}) = \psi_X(t).$$

( $\Rightarrow$ ) Si  $X$  es una variable aleatoria con **fda** simétrica alrededor de cero, entonces usando la propiedad anterior, que dice que  $F_X = F_{-X}$  si y solo si  $\psi_X(t) = \psi_{-X}(t)$ , se tiene que:

$$\overline{\psi_{-X}(t)} = \psi_X(t) = \psi_{-X}(t),$$

entonces  $\psi_X(t)$  es real.

( $\Leftarrow$ ) Si  $\psi_X(t)$  es real, entonces  $\psi_{-X}(t) = \overline{\psi_{-X}(t)} = \psi_X(t)$ , luego  $X$  es una variable aleatoria con **fda** simétrica en torno de cero.

11. Ver detalles de la prueba en [43, p. 229].  $\square$

**Nota 1.16.** Si  $X$  es una variable aleatoria que tiene **fgm**  $m_X(t)$ , entonces:

$$\psi_X(t) = m_X(it). \quad (1.60)$$

Como aplicación de la nota anterior, consideremos las distribuciones dadas en la tabla 1.12, por lo que se tiene que:

Tabla 1.13. Resumen  $m_X(t)$  y  $\psi_X(t)$  para algunas distribuciones

Distribución	Parámetros	$m_X(t)$	$\psi_X(t)$
Uniforme discreta	$n$	$\frac{e^t(1-e^{nt})}{n(1-e^t)}$	$\frac{e^{it}(1-e^{int})}{n(1-e^{it})}$
Binomial	$m$ y $p$	$[pe^t + (1-p)]^m$	$[pe^{it} + (1-p)]^m$
Poisson	$\lambda$	$\exp\{\lambda(e^t - 1)\}$	$\exp\{\lambda(e^{it} - 1)\}$
Geométrica	$p$	$\frac{p}{1-(1-p)e^t}$	$\frac{p}{1-(1-p)e^{it}}$
Uniforme continua	$a$ y $b$	$\frac{1}{t(b-a)}(e^{tb} - e^{ta})$	$\frac{i}{t(b-a)}(e^{itb} - e^{ita})$
Normal	$\mu$ y $\sigma^2$	$\exp\left\{t\mu + \frac{\sigma^2 t^2}{2}\right\}$	$\exp\left\{it\mu - \frac{\sigma^2 t^2}{2}\right\}$
Gamma	$r$ y $\lambda$	$\left(\frac{\lambda}{\lambda-t}\right)^r$	$\left(\frac{\lambda}{\lambda-it}\right)^r$
Exponencial	$\lambda$	$\frac{\lambda}{\lambda-t}$	$\frac{\lambda}{\lambda-it}$
chi-cuadrada	$k$	$(1-2t)^{-k/2}$	$(1-2it)^{-k/2}$

**Ejemplo 1.90.** Según [18, p. 261], si  $X \sim \text{Cauchy}(\alpha, \beta)$  con *fdp* dada por:

$$f_X(x) = \frac{1}{\pi\beta \left[1 + \left(\frac{x-\alpha}{\beta}\right)^2\right]}, \quad x \in \mathbb{R}, \alpha \in \mathbb{R}, \beta > 0, \quad (1.61)$$

determinar  $\psi_X(t)$ .

*Solución.* Sea  $f(t) = e^{-\kappa|t|}$ ,  $\kappa > 0$ , entonces la transformada de Fourier de  $f(t)$  es dada en [69, p. 23] como:

$$\begin{aligned} \mathcal{F}[f(t)] &= \int_{-\infty}^{\infty} e^{-ixt} f(t) dt = \lim_{T \rightarrow \infty} \int_{-T}^T e^{-ixt} e^{-\kappa|t|} dt \\ &= \lim_{T \rightarrow \infty} \left[ \int_{-T}^0 e^{\kappa t} e^{-ixt} dt + \int_0^T e^{-\kappa t} e^{-ixt} dt \right]. \end{aligned}$$

En la primera integral del lado derecho hacemos la sustitución  $u = -t$  entonces  $du = -dt$ , luego:

$$\begin{aligned} \mathcal{F}[f(t)] &= \lim_{T \rightarrow \infty} \left[ - \int_T^0 e^{-\kappa u} e^{ixu} du + \int_0^T e^{-\kappa t} e^{-ixt} dt \right] \\ &= \lim_{T \rightarrow \infty} \int_0^T e^{-\kappa t} (e^{ixt} + e^{-ixt}) dt \\ &= \lim_{T \rightarrow \infty} \int_0^T e^{-\kappa t} [(\cos(xt) + i \text{sen}(xt)) + (\cos(xt) - i \text{sen}(xt))] dt \end{aligned}$$

$$= \lim_{T \rightarrow \infty} \int_0^T 2e^{-\kappa t} \cos(xt) dt = 2 \int_0^{\infty} e^{-\kappa t} \cos(xt) dt.$$

Esta última integral es la transformada de Laplace de la función coseno, evaluada en  $s = \kappa$  (ver [1, p. 1022]), luego,

$$\mathcal{F}[f(t)] = \frac{2\kappa}{\kappa^2 + x^2}.$$

Ahora, la transformada inversa de esta última expresión es:

$$\begin{aligned} f(t) &= \mathcal{F}^{-1}[\mathcal{F}[f(t)]] = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(x)e^{ixt} dx \\ e^{-\kappa|t|} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2\kappa}{\kappa^2 + x^2} e^{ixt} dx = \frac{\kappa}{\pi} \int_{-\infty}^{\infty} \frac{e^{ixt}}{\kappa^2 + x^2} dx. \end{aligned}$$

Al tomar  $\kappa = 1$ , en la integral anterior se obtiene que:

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{ixt}}{x^2 + 1} dx = e^{-|t|}, \quad (1.62)$$

esta expresión es la  $\psi_X(t)$  de la **fdp** dada en (1.45). Por consiguiente,

$$\psi_X(t) = \frac{1}{\pi\beta} \int_{-\infty}^{\infty} \frac{e^{ixt}}{1 + \left(\frac{x-\alpha}{\beta}\right)^2} dx,$$

haciendo el cambio de variable

$$z = \frac{x - \alpha}{\beta} \quad (\text{derivamos}) \quad dz = \frac{dx}{\beta}.$$

Usando la expresión (1.62) se llega a

$$\psi_X(t) = \frac{e^{i\alpha t}}{\pi} \int_{-\infty}^{\infty} \frac{e^{i\beta z t}}{1 + z^2} dz = \exp\{i\alpha t - \beta|t|\}. \quad \square$$

**Ejemplo 1.91.** Según [33, p. 118], la función característica de una variable aleatoria  $X$  es dada por:

$$\psi_X(t) = \exp\left\{-\frac{t^2}{2}\right\}, \quad (1.63)$$

determine la **fdp** de esta variable aleatoria.

*Solución.* De la expresión (1.59) se tiene que:

$$\begin{aligned}
 f_X(x) &= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \exp\{-itx\} \exp\left\{-\frac{t^2}{2}\right\} dt \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{-\frac{t^2 + 2itx}{2}\right\} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{-\frac{(t + ix)^2 - (ix)^2}{2}\right\} dt \\
 &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\} \underbrace{\left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}(t + ix)^2\right\} dt\right]}_1 \\
 &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\},
 \end{aligned}$$

esta es la fdp de la normal estándar. ☑

## 1.12. Ejercicios

### Conceptos básicos de probabilidad

1. De seis aplicantes para un trabajo ejecutivo, el señor A es graduado, extranjero y soltero; el señor B no es graduado, es extranjero y casado; el señor C es graduado, nativo y casado; el señor D no es graduado, es nativo y soltero; el señor E es graduado, nativo y casado; el señor F no es graduado, es nativo y casado. Dibuje un diagrama de Venn.
2. Sean  $\mathcal{F}_1 = \{A, A^c, \Omega, \emptyset\}$  y  $\mathcal{F}_2 = \{B, B^c, \Omega, \emptyset\}$ , su unión:

$$\{\emptyset, \Omega, A, A^c, B, B^c\}$$

no es una  $\sigma$ -álgebra. ¿por qué?

3. La tabla 1.14 presenta el número de estudiantes admitidos por Facultad en la Universidad Nacional de Colombia, sede Bogotá, en el año 2007.

Si se selecciona un estudiante aleatoriamente, entre los estudiantes admitidos en el 2007,

- a) ¿Cuál es la probabilidad de que estudie una carrera de la Facultad de Ciencias Humanas?
- b) ¿Cuál es la probabilidad de que estudie una carrera de la Facultad de Ciencias?

Tabla 1.14. Admitidos en pregrado por Facultad, 2007

Facultad	2007 - I	2007-II
Agronomía	72	86
Artes	249	233
Ciencias	397	402
Ciencias Económicas	174	169
Ciencias Humanas	498	421
Derecho, Ciencias Políticas y Sociales	171	171
Enfermería	114	82
Ingeniería	715	714
Medicina	272	235
Medicina Veterinaria y de Zootecnia	104	101
Odontología	57	57

Fuente: adaptada de [17, p. 51]

- c) ¿Cuál es la probabilidad de que estudie una carrera de la Facultad de Enfermería?
4. Considerando la información del ejercicio 3.,
- ¿Cuál es la probabilidad de que estudie una carrera de la Facultad de Medicina o de la Facultad de Odontología?
  - ¿Cuál es la probabilidad de que estudie una carrera de la Facultad de Medicina y de la Facultad de Odontología?
  - ¿Cuál es la probabilidad de que NO estudie una carrera de la Facultad de Medicina, ni una carrera de la Facultad de Odontología?
  - ¿Cuál es la probabilidad de que haya sido admitido en el segundo semestre?
  - Si el estudiante fue admitido en el periodo 2007-I, ¿cuál es la probabilidad de que estudie una carrera de la Facultad de Medicina o de la Facultad de Odontología?
  - Si el estudiante fue admitido en el periodo 2007-I, ¿cuál es la probabilidad de que estudie una carrera de la Facultad de Medicina y de la Facultad de Odontología?
  - Si el estudiante fue admitido en el periodo 2007-I, ¿cuál es la probabilidad de que NO estudie una carrera de la Facultad de Medicina, ni una carrera de la Facultad de Odontología?
5. El número de estudiantes graduados en el periodo 2007-I en la Universidad Nacional de Colombia, sede Bogotá, en la Facultad de Ciencias Humanas fue 286, en la carrera de Trabajo Social

en el año 2007 se graduaron 69, y el número de graduados en Trabajo Social en el periodo 2007-I fueron 40. Si el total de estudiantes graduados de la Facultad de Ciencias Humanas en el año 2007 fue 578 [17, p. 94], ¿cuál es la probabilidad de que un estudiante graduado en el año 2007 se haya graduado de Trabajo Social en el primer semestre del 2007?.

6. El número de estudiantes graduados en el periodo 2006-I en la Universidad Nacional de Colombia, sede Bogotá, en la Facultad de Ciencias Humanas fue 416, en la carrera de Sociología en el año 2006 se graduaron 124, y el número de graduados en Sociología en el periodo 2006-I fueron 60. Si el total de estudiantes graduados de la Facultad de Ciencias Humanas en el año 2006 fue 793 [17, p. 94], ¿cuál es la probabilidad de que un estudiante graduado en el año 2006 se haya graduado en el segundo semestre del 2006 de una carrera diferente a Sociología?.
7. Considerando la información del ejercicio 3.:
  - a) Dado que el estudiante fue admitido en el segundo semestre del año 2007, ¿cuál es la probabilidad de que haya estudiado una carrera de la Facultad de Enfermería?
  - b) Dado que el estudiante fue admitido en el segundo semestre del año 2007, ¿cuál es la probabilidad de que haya estudiado una carrera de la Facultad de Enfermería o una carrera de la Facultad de Medicina?
  - c) Dado que el estudiante se admitió a una carrera de la Facultad de Ciencias Humanas en el año 2007, ¿cuál es la probabilidad de que se haya admitido en el primer semestre del 2007?
  - d) Dado que el estudiante NO se admitió en una carrera de la Facultad de Ciencias Humanas en el año 2007, ¿cuál es la probabilidad de que se haya admitido en el primer semestre del 2007?
8. El número de estudiantes graduados en el periodo 2005-I en la Universidad Nacional de Colombia, sede Bogotá, en la Facultad de Ciencias Humanas fue 120, en la carrera de Sociología en el periodo 2005-I se graduaron 14. El total de estudiantes graduados de la Facultad de Ciencias Humanas en el año 2005 fue 825 [17, p. 94]. Dado que un estudiante de la Facultad de Ciencias Humanas se graduó en el periodo 2005-I, ¿cuál es la probabilidad de que sea de la carrera de Sociología?.

9. En un lanzamiento de una moneda 5 veces, en que C indica cara y S indica sello:
  - a) ¿Cuál es la probabilidad de que salga CCSCC?
  - b) ¿Cuál es la probabilidad de que salga SCCCC?
  - c) ¿Cuál es la probabilidad de que salga SSSCC?
10. Con base en encuestas al consumidor se sabe que la preferencia de éste con respecto a dos marcas, A y B, de un producto dado, se encuentra muy pareja [19, p. 122]. Si la opción de compra entre estas marcas es independiente, ¿cuál es la probabilidad de que entre 6 personas seleccionadas al azar, máximo 3 prefieran la marca A?
11. Con base en varios estudios una compañía ha clasificado, de acuerdo con la posibilidad de descubrir petróleo, las formaciones geológicas en tres tipos. La compañía pretende perforar un pozo en un determinado sitio, al que se le asignan las probabilidades de 0,35, 0,40, y 0,25 para los tres tipos de formaciones, respectivamente [19, p. 51]. De acuerdo con la experiencia, se sabe que el petróleo se encuentra en un 40 % de formaciones tipo I, en un 20 % de formaciones tipo II y en un 30 % de formaciones tipo III. Si la compañía no descubre petróleo en ese lugar, determínese la probabilidad de que exista una formación del tipo II.
12. Los clientes se encargan de evaluar los diseños preliminares de varios productos. En el pasado, el 95 % de los productos con mayor éxito en el mercado recibieron buenas evaluaciones, el 60 % de los productos con éxito moderado recibieron buenas evaluaciones y el 10 % de productos de escaso éxito recibieron buenas evaluaciones. Además, el 40 % de los productos ha tenido mucho éxito, el 35 % un éxito moderado y el 25 % una baja aceptación [64, p. 93].
  - a) ¿Cuál es la probabilidad de que un producto obtenga buena evaluación?
  - b) Si un nuevo diseño obtiene una buena evaluación ¿cuál es la probabilidad de que se convierta en un producto de gran éxito?
  - c) Si un producto no obtiene una buena evaluación, ¿cuál es la probabilidad de que se convierta en un producto de gran éxito?

### Tipos de variables aleatorias

13. Sea  $X$  una variable aleatoria tal que

$$f_X(x) = \begin{cases} 0,2 & \text{si } -1 < x \leq 0 \\ 0,2 + cx & \text{si } 0 < x \leq 1 \\ 0 & \text{en otro caso.} \end{cases}$$

Determinar  $c$  y obtener la **fda** de  $X$  [16, p. 93].

14. Determine el valor o los valores de  $c$ , para que las siguientes funciones sean una **fdp**

$$\begin{aligned} & \bullet f(x) = \cos(x)I_{(0,c)}(x) \\ & \bullet f(x) = cx^2I_{(-c,c)}(x) \\ & \bullet f(x) = cx^2e^{-x^3}I_{(0,\infty)}(x) \end{aligned} \quad \bullet f(x) = \begin{cases} 0 & x \leq -1 \\ -cx & -1 < x \leq 0 \\ ce^{-6x} & x > 0 \end{cases}$$

15. Muestre que (1.11) satisface las propiedades de una **fmp** (definición 1.11, ítems a) y b)).

16. Muestre que (1.12) satisface las propiedades de una **fmp** (definición 1.11, ítems a) y b)).

17. La probabilidad de que un paciente se recupere de una delicada operación de corazón es 0,9. ¿Cuál es la probabilidad de que:

- a) exactamente 5 de los próximos 7 pacientes que se sometan a esta intervención sobrevivan?
- b) a lo más 2 de los próximos 10 pacientes que se sometan a esta intervención mueran?

18. Muestre que (1.13) satisface las propiedades de función una **fmp** (definición 1.11, ítems a) y b)).

19. El número de componentes que fallan antes de cumplir 100 horas de operación se considera que proviene de una variable aleatoria con **fmp** Poisson [19, p. 124]. Si el número promedio de estas es ocho, ¿cuál es la probabilidad de que:

- a) falle un componente en 25 horas?
- b) fallen no más de dos componentes en 50 horas?
- c) fallen por lo menos diez en 125 horas?

20. Considere el lanzamiento de una moneda legal repetidas veces, ¿cuál es la probabilidad de que aparezca un sello solo hasta el décimo lanzamiento?

21. Considere el lanzamiento de un dado repetidas veces, ¿cuál es la probabilidad de que aparezca el 5 solo hasta el décimo lanza-

- miento?, ¿cuál es la probabilidad de que sea necesario lanzar el dado más de 6 veces para que aparezca un 5?
22. Muestre que (1.21) satisface las propiedades de una fdp (definición 1.11, ítems a) y b)).
23. Suponga que el tiempo de duración de una conferencia de “una hora” en el auditorio B oscila entre 45 y 80 minutos. Considérese que una distribución uniforme es apropiada para la variable aleatoria “tiempo de duración de una conferencia de una hora”,
- ¿Qué porcentaje de conferencias durará más de una hora?
  - ¿Qué porcentaje de conferencias durará máximo 50 minutos?
24. Suponga que el puntaje en un examen de ingreso a una institución A se puede considerar que sigue una distribución normal con media 350 puntos y desviación estándar 50 puntos. Evaluar la probabilidad de que el puntaje de un examen seleccionado al azar en una cohorte,
- sea menor de 220 puntos.
  - sea mayor de 450 puntos.
  - este entre 300 y 400 puntos.
  - sea menor de 250 puntos o mayor de 400 puntos.
  - Encontrar un punto  $x_0$  que tenga la propiedad que el 25 % de los exámenes tenga un puntaje por debajo de este valor.
  - Encontrar un punto  $x_0$  que tenga la propiedad que el 25 % de los exámenes tenga un puntaje por encima de este valor.
25. Continuando con el ejemplo 1.45, es decir, una variable aleatoria  $X$  con distribución hipergeométrica de parámetros  $M$ ,  $m$  y  $n$ , evaluar
- $\mu_{[2]}(X)$ .
  - $\mathbb{V}(X)$ .
26. Encontrar la  $\mathbb{E}(X)$ , la  $\mathbb{E}(X^2)$  y la  $\mathbb{V}(X)$  para las variables aleatorias dadas:
- En el ejemplo 1.10.
  - En el ejemplo 1.32.
  - En el ejercicio 1.12.
  - En el ejercicio 1.12.
27. Demuestre que si  $X \sim \text{Beta}(\alpha, \beta)$ , entonces:

$$\mathbb{E}(X) = \frac{\alpha}{\alpha + \beta} \quad \text{y} \quad \mathbb{V}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

28. Sea  $X \sim \text{Beta}(a, b)$  con  $a, b \in \mathbb{Z}^+$  tales que  $a + b = n + 1$ . Muestre que para  $0 < x < 1$  la fda se puede expresar como:

$$I(x; a, b) = \frac{\int_0^x u^{a-1}(1-u)^{b-1} du}{\int_0^1 u^{a-1}(1-u)^{b-1} du} = \sum_{k=a}^n \binom{n}{k} x^k (1-x)^{n-k}. \quad (1.64)$$

La expresión en el centro se denomina *función beta incompleta regularizada* (ver [1]).

29. Muestre que si  $X$  sigue la fdp dada en (1.29), el valor esperado y la varianza están dados por:

$$\mu_X = C^{\mu + \frac{1}{2}\sigma^2 \ln C} \quad \text{y} \quad \sigma_X^2 = C^{2\mu + \sigma^2 \ln C} (C^{\sigma^2 \ln C} - 1),$$

### Funciones generadoras

30. Sea  $X$  una variable aleatoria con distribución Laplace, esto es,

$$f_X(x) = \frac{1}{2b} \exp\left\{-\frac{|x-a|}{b}\right\}, \quad x \in \mathbb{R},$$

en que  $a$  y  $b$  son parámetros de localización y de escala, respectivamente, tal que  $a \in \mathbb{R}$  y  $b > 0$ . Mostrar que la fgm es dada por:

$$m_X(t) = \frac{e^{at}}{1 - b^2 t^2}, \quad \text{si } |t| < \frac{1}{b}.$$

31. Evaluar  $\mathbb{E}(X)$  y  $\mathbb{E}(X^2)$  usando la fgm presentadas en los ejemplos 1.77, 1.78, 1.79, 1.80 y 1.81.
32. Para las siguientes fmp o fdp, según sea el caso, encuentre la fgm y obtenga a partir de ella la  $\mathbb{E}(X)$  y  $\mathbb{V}(X)$ .

- a) La familia binomial negativa de parámetros  $n$  (fijo) y  $\theta$ ,

$$f_X(x) = \binom{n+x-1}{n-1} \theta^n (1-\theta)^x I_{\{0,1,2,\dots\}}(x), \quad \text{para } 0 < \theta < 1.$$

- b)  $f_X(x) = \frac{2x}{c^2}$ ,  $0 < x < c$ , [20, p. 80].

- c) Distribución Gumbel

$$f_X(x) = \frac{1}{b} \exp\left\{-\frac{(x-a)}{b}\right\} \exp\left\{-\exp\left(\frac{-(x-a)}{b}\right)\right\}, \quad x \in \mathbb{R}, \quad b > 0.$$

$$d) f(x) = \begin{cases} 0 & x \leq -1 \\ -cx & -1 < x \leq 0 \\ ce^{-6x} & x > 0 \end{cases}$$

33. Encontrar la función característica de  $X$  para las siguientes distribuciones:

- Binomial negativa  $(r, p)$ .
- Gamma  $(r, \lambda)$ .
- Weibull  $(\lambda, \alpha)$ .  $f_X(x) = \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha}$ , con  $0 < x < \infty$ ,  $\alpha > 0$  y  $\lambda > 0$ .

34. Sea  $X$  una variable aleatoria con fdp  $f_X(x) = \frac{1}{2}e^{-|x|}$ ,  $x$  número real [59, p. 289]. Muestre que la función característica de  $X$  es dada por  $\psi_X(t) = \frac{1}{1+t^2}$ .

35. Sea  $X$  una variable aleatoria con fdp  $f_X(x) = \left(\frac{1}{4} + \frac{x}{8}\right)I_{[0,2)}(x) + \frac{1}{4}I_{(-1,0)}(x)$  [59, p. 290]. Obtenga la función característica de  $X$  y verifique su respuesta calculando  $\psi_X(0)$ .

36. Sea  $X$  una variable aleatoria tal que  $X \sim N(\mu, \sigma^2)$  [59, p. 290]. Obtenga la función característica de  $X^2$  e identifique la fdp.

37. Usando la fórmula de Stirling dada en (1.46), demuestre que:

$$\frac{\Gamma(m + \kappa)}{\Gamma(m)} \approx \frac{\sqrt{m}}{e^\kappa} \left(1 + \frac{\kappa}{m}\right)^m (m + \kappa)^{\kappa - \frac{1}{2}}, \quad \kappa > 0. \quad (1.65)$$

Capítulo  
*dos*  
**Vectores  
aleatorios y  
distribuciones  
multivariadas**



En este capítulo se desarrolla la teoría del capítulo anterior para el caso en el que se trabaje con muchas variables aleatorias, se considera primero el caso bivariado.

**DEFINICIÓN 2.1. Vector aleatorio bidimensional.** Sea  $Z(\omega) = (X(\omega), Y(\omega))$  para cualquier  $\omega \in \Omega$ , entonces:

- $Z(\omega)$  es un punto en el plano euclidiano  $\mathbb{R}^2$ .
- $Z$  define una función de  $\Omega$  a  $\mathbb{R}^2$ .
- Dada  $\mathcal{C} = \{B_2 \subset \mathbb{R}^2\}$  la clase de todas las regiones rectangulares limitadas por las líneas  $x = a, x = b, y = c, y = d$  con  $a < b, c < d$ , tal que:

$$B_2 = \{(x, y) \in \mathbb{R}^2 : a < x < b, c < y < d\}, \quad a, b, c, d \in \mathbb{R}. \quad (2.1)$$

El mínimo  $\sigma$ -campo<sup>10</sup> que contiene a  $\mathcal{C}$  es el campo de Borel  $\mathfrak{B}^2$  en  $\mathbb{R}^2$ .

- $Z$  se llamará un vector aleatorio bidimensional si  $Z^{-1}(\mathfrak{B}^2) \subset \mathcal{C}$ .
- $Z^{-1}(\mathfrak{B}^2)$  es el  $\sigma$ -campo inducido por  $Z$ , es decir,  $Z^{-1}(\mathfrak{B}^2) = \sigma\{Z^{-1}(B_2)\}$ .

En otras palabras,  $Z$  será un vector aleatorio bidimensional si y solo si  $Z^{-1}(\mathfrak{B}^2)$  es un evento para cualquier rectángulo  $B_2$  en  $\mathbb{R}^2$  (ver [11, p. 30]).

**Ejemplo 2.1.** Sean  $X = I_A, Y = I_B$  como en (1.10) y defina  $Z = (I_A, I_B)$  como

$$Z(\omega) = \begin{cases} (0, 0) & \text{si } \omega \in A^c \cap B^c, \\ (1, 0) & \text{si } \omega \in A \cap B^c, \\ (0, 1) & \text{si } \omega \in A^c \cap B, \\ (1, 1) & \text{si } \omega \in A \cap B. \end{cases}$$

Determine la función inversa de esta aplicación.

*Solución.* Considérese la región rectangular en el plano cartesiano dada en (2.1). Sea  $\mathcal{C} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ , si el rectángulo  $B_2$  contiene solo uno de los puntos de  $\mathcal{C}$ , entonces:

$$Z^{-1}(B_2) = \begin{cases} A^c \cap B^c & \text{si } B_2 \text{ contiene solo el punto } (0, 0) \\ A \cap B^c & \text{si } B_2 \text{ contiene solo el punto } (1, 0) \\ A^c \cap B & \text{si } B_2 \text{ contiene solo el punto } (0, 1) \\ A \cap B & \text{si } B_2 \text{ contiene solo el punto } (1, 1). \end{cases}$$

<sup>10</sup>Un campo es una clase de conjuntos no vacía  $\mathcal{A}$ , cerrada bajo complementos, es decir, si  $A \in \mathcal{A}$ , entonces  $A^c \in \mathcal{A}$ , y cerrada bajo intersecciones finitas, es decir, si  $A, B \in \mathcal{A}$ , entonces  $A \cap B \in \mathcal{A}$ . Para ser un  $\sigma$ -campo  $\mathcal{A}$  debe ser cerrada bajo complementos y cerrada bajo intersecciones contables. Un campo que contiene un número infinito de conjuntos puede no ser un  $\sigma$ -campo.

Si el rectángulo  $B_2$  contiene más de un punto de  $\mathcal{C}$ , entonces:

$$Z^{-1}(B_2) = \begin{cases} \emptyset & \text{si } B_2 \text{ no contiene ningún punto de } \mathcal{C} \\ A^c \cap B^c + A \cap B^c & \text{si } B_2 \text{ contiene solo los puntos } (0,0) \text{ y } (1,0) \\ A^c \cap B^c + A^c \cap B & \text{si } B_2 \text{ contiene solo los puntos } (0,0) \text{ y } (0,1) \\ A^c \cap B^c + A \cap B & \text{si } B_2 \text{ contiene solo los puntos } (0,0) \text{ y } (1,1) \\ \vdots & \vdots \\ \Omega & \text{si } B_2 \text{ contiene los cuatro puntos de } \mathcal{C}, \end{cases}$$

en que  $C + D$  denota  $C \cup D$  siempre y cuando  $C \cap D = \emptyset$ . ☑

**DEFINICIÓN 2.2.** (Variables aleatorias independientes). Sean  $X$  y  $Y$  dos variables aleatorias reales definidas sobre el mismo espacio de probabilidad. Si para cualquier pareja de conjuntos Borel  $A$  y  $B$  de  $\mathbb{R}$  se tiene que:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B),$$

entonces  $X$  y  $Y$  se dice que son independientes.

Las variables aleatorias  $X, Y$  se denominan “independientes” cuando:

$$F(x,y) = F_X(x)F_Y(y). \tag{2.2}$$

**Nota 2.1.** Para todo  $y$ ; con  $f_Y(y) > 0$  y todo conjunto Borel  $A$  de  $\mathbb{R}$ , se tiene que:

$$P(X \in A|Y = y) = \int_A f_{X|Y}(x|y) dx.$$

## 2.1. Distribución de probabilidad conjunta

Al igual que en el caso univariado, existen variables aleatorias bidimensionales asociadas a otra función llamada, **fmp** o **fdp** conjunta. Cada par de valores representa la medición de dos características de un mismo individuo, ya sea de forma individual o conjunta, lo que puede interpretarse como las componentes de un vector.

- *Caso discreto:*

Dadas dos variables aleatorias  $X, Y$ , se llama “**fmp** conjunta” de  $X$  y  $Y$  a la función  $p(x,y) = P[X = x, Y = y]$  si:

- a)  $p(x,y) \geq 0$  para todos los valores de  $x, y$ .
- b)  $\sum_x \sum_y p(x,y) = 1$ .

Las “fmp marginales” de las variables aleatorias  $X$  y  $Y$  son dadas por:

$$\underbrace{P(X = x)}_{p_X(x)} = \sum_y p(x,y) \quad \text{y} \quad \underbrace{P(Y = y)}_{p_Y(y)} = \sum_x p(x,y). \quad (2.3)$$

La “fmp condicional” de  $Y$  dado que  $X = x$  es definida como:

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)} \quad \text{con} \quad p_X(x) > 0. \quad (2.4)$$

De manera similar se define la fmp condicional de  $X$ , dado que  $Y = y$ . Las variables aleatorias  $X, Y$  se denominan “independientes” si para todos los valores de  $x$  y  $y$ , se cumple que:

$$p(x,y) = p_X(x)p_Y(y). \quad (2.5)$$

■ *Caso continuo:*

Si  $X, Y$  son dos variables aleatorias, entonces la función  $f(x,y)$  se dice que es la “fdp conjunta” de  $X$  y  $Y$  si:

- a)  $f(x,y) \geq 0$  para todos los valores de  $x, y$ .
- b)  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dx dy = 1$ .

Las “fdp marginales” de las variables aleatorias  $X, Y$  son dadas por las integrales

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y) dy \quad \text{y} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y) dx. \quad (2.6)$$

En este caso, la “fdp condicional” de  $X$  dada  $Y = y$  es definida como:

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} \quad \text{con} \quad f_Y(y) > 0. \quad (2.7)$$

De manera análoga, se define la fdp condicional de  $Y$  dado que  $X = x$ . Por otra parte, las variables aleatorias  $X, Y$  son “estadísticamente independientes” si se satisface que:

$$f_{(X,Y)}(x,y) = f_X(x)f_Y(y). \quad (2.8)$$

**Ejemplo 2.2.** Una urna contiene tres bolas rojas y dos bolas verdes, se extrae una muestra aleatoria de tamaño 2 (con reemplazo) [16, p. 260]. Sean:

$$X = \begin{cases} 1 & \text{si la primera bola extraída es roja} \\ 0 & \text{si la primera bola extraída es verde} \end{cases}$$

$$Y = \begin{cases} 1 & \text{si la segunda bola extraída es roja} \\ 0 & \text{si la segunda bola extraída es verde} \end{cases}$$

Determine:

1) La fmp conjunta.

2)  $P(X = x|Y = k)$ ,  $k = 0, 1$ .

*Solución.* Primero se determina el espacio muestral de este experimento estadístico:

$$\Omega = \{(R_1, R_1), (R_1, R_2), (R_1, R_3), (R_2, R_1), (R_2, R_2), (R_2, R_3), (R_3, R_1), (R_3, R_2), (R_3, R_3), (R_1, V_1), (R_1, V_2), (R_2, V_1), (R_2, V_2), (R_3, V_1), (R_3, V_2), (V_1, R_1), (V_2, R_1), (V_1, R_2), (V_2, R_2), (V_1, R_3), (V_2, R_3), (V_1, V_1), (V_1, V_2), (V_2, V_1), (V_2, V_2)\}$$

Luego,

1. La fmp conjunta está dada por:

X \ Y	0	1	Total
0	4/25	6/25	2/5
1	6/25	9/25	3/5
Total	2/5	3/5	1

Nótese que las fmp marginales son idénticas, es decir, son variables aleatorias idénticamente distribuidas.

2. Las probabilidades condicionales quedan:

$$P(X = x|Y = 0) = \frac{P(X = x, Y = 0)}{P(Y = 0)} = \begin{cases} 2/5 & \text{si } x = 0 \\ 3/5 & \text{si } x = 1 \end{cases}$$

$$P(X = x|Y = 1) = \frac{P(X = x, Y = 1)}{P(Y = 1)} = \begin{cases} 2/5 & \text{si } x = 0 \\ 3/5 & \text{si } x = 1 \end{cases} \quad \checkmark$$

## 2.2. Función de distribución conjunta

La “**fda** conjunta” de dos variables aleatorias  $X, Y$  se define como:

$$F(x, y) = P[X \leq x, Y \leq y]. \quad (2.9)$$

Las “**fda** marginales” de las variables aleatorias  $X, Y$  están dadas por:

$$\underbrace{P[X \leq x]}_{F_X(x)} = \lim_{y \rightarrow \infty} F(x, y) \quad \text{y} \quad \underbrace{P[Y \leq y]}_{F_Y(y)} = \lim_{x \rightarrow \infty} F(x, y). \quad (2.10)$$

Las “**fdas** condicionales” de  $X$ , dada  $Y = y$ , y de  $Y$ , dada  $X = x$ , son definidas como:

$$F_{X|Y}(x|y) = \frac{F(x, y)}{F_Y(y)} \quad \text{y} \quad F_{Y|X}(y|x) = \frac{F(x, y)}{F_X(x)}, \quad (2.11)$$

con  $F_X(x) > 0$  y  $F_Y(y) > 0$ .

**Ejemplo 2.3.** Según [16, p. 263], si  $X$  y  $Y$  son variables aleatorias con **fdp** conjunta dada por:

$$f_{(X,Y)}(x, y) = \frac{e^{-x/y} e^{-y}}{y} \quad x > 0, y > 0.$$

Encuentre  $P(X > 1 | Y = y)$ .

*Solución.* Primero se determina la **fdp** marginal de  $Y$ , es decir,

$$f_Y(y) = \int_0^\infty f_{(X,Y)}(x, y) dx = \int_0^\infty \frac{e^{-x/y} e^{-y}}{y} dx,$$

haciendo el cambio de variable  $u = \frac{x}{y}$  entonces  $du = \frac{dx}{y}$ , luego,

$$f_Y(y) = e^{-y} \int_0^\infty e^{-u} du = e^{-y} (-e^{-u}) \Big|_0^\infty = e^{-y}.$$

Luego  $Y \sim \exp(1)$  y se tiene que:

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)} = \frac{e^{-x/y}}{y} \quad y > 0.$$

Por lo tanto,

$$P(X > 1 | Y = y) = \int_1^{\infty} f_{X|Y}(x|y) dx = \int_1^{\infty} \frac{e^{-x/y}}{y} dx = \left(-e^{-\frac{x}{y}}\right)\Big|_1^{\infty} = e^{-\frac{1}{y}}. \quad \square$$

**Nota 2.2. Regla de Bayes.** Dado que la “fdp condicional” de  $X$  dada  $Y = y$  es definida como:

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)} \quad \text{con} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y) dx > 0. \quad (2.12)$$

Además,  $f_{(X,Y)}(x,y) = f_X(x)f_{Y|X}(y|x)$ , entonces:

$$f(x|y) = \frac{f(y|x)f(x)}{f(y)} = \frac{f(y|x)f(x)}{\int_{-\infty}^{\infty} f(y|x)f(x) dx} \quad (2.13)$$

## 2.3. Valores esperados y momentos

En esta sección se estudiarán los conceptos de valor esperado y momentos para distribuciones de probabilidad conjuntas.

**DEFINICIÓN 2.3.** (Ley del estadístico inconsciente). Según [19, p. 191], si  $X$  y  $Y$  son variables aleatorias que se distribuyen conjuntamente, el valor esperado de una función de  $X$  y de  $Y$ ,  $h(x,y)$ , se define como:

$$\mathbb{E}[h(X,Y)] = \begin{cases} \sum_x \sum_y h(x,y)P(X=x, Y=y) & \text{si } X, Y \text{ son discretas,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)f_{(X,Y)}(x,y) dy dx & \text{si } X, Y \text{ son continuas.} \end{cases}$$

**Ejemplo 2.4.** Según [19, p. 192], si  $X$  y  $Y$  son variables aleatorias con fdp conjunta:

$$f(x,y) = \begin{cases} \frac{2}{3}(x+y)e^{-x} & \text{si } x > 0, 0 < y < 1 \\ 0 & \text{en otro caso.} \end{cases}$$

Determine  $\mathbb{E}[h(X,Y)]$  donde  $h(x,y) = xy$ .

**Solución.** En este caso,

$$\begin{aligned} \mathbb{E}[h(X,Y)] &= \mathbb{E}[XY] = \int_0^{\infty} \int_0^1 \frac{2}{3}xy(x+y)e^{-x} dy dx \\ &= \int_0^{\infty} \frac{2}{3} \left( x^2 \frac{y^2}{2} + x \frac{y^3}{3} \right) e^{-x} \Big|_{y=0}^{y=1} dx \end{aligned}$$

$$= \frac{1}{3} \int_0^\infty x^2 e^{-x} dx + \frac{2}{9} \int_0^\infty x e^{-x} dx,$$

usando el método tabular empleado en el ejemplo 1.63 (ver tabla 1.10):

$$\begin{aligned} \mathbb{E}[h(X,Y)] &= \frac{1}{3} [-x^2 e^{-x} - 2x e^{-x} - 2e^{-x}] \Big|_{x=0}^{x \rightarrow \infty} + \frac{2}{9} [-x e^{-x} - e^{-x}] \Big|_{x=0}^{x \rightarrow \infty} \\ &= \frac{2}{3} + \frac{2}{9} = \frac{8}{9}. \quad \square \end{aligned}$$

## 2.4. Valor esperado condicional

Sean  $X$  y  $Y$  variables aleatorias distribuidas conjuntamente, entonces el valor esperado de la variable aleatoria  $X$ , dado  $Y = y$  se define como:

$$\mathbb{E}[X|Y = y] = \begin{cases} \sum x p_{X|Y}(x|y) & \text{si } X, Y \text{ son discretas,} \\ \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx & \text{si } X, Y \text{ son continuas.} \end{cases}$$

**Teorema 2.1. Propiedades del valor esperado condicional.** *Las propiedades fundamentales de  $\mathbb{E}[Y|X = x]$  son las siguientes:*

- i) Ley de esperanza total: *El valor esperado respecto a la ley de probabilidad de  $X$  proporciona  $\mathbb{E}(Y)$ , es decir,*

$$\mathbb{E}[\mathbb{E}[Y|X = x]] = \mathbb{E}[Y]. \quad (2.14)$$

- ii) *El valor esperado condicional de una variable aleatoria que es el producto de  $Y$  por una función  $h(X)$  de la variable aleatoria  $X$  satisface:*

$$\mathbb{E}[h(X)Y|X = x] = h(x) \mathbb{E}[Y|X = x], \quad (2.15)$$

*tal que  $\mathbb{E}[h(X)Y]$  sea finita.*

*Demostración.* Ver detalles de la prueba en [83, p. 55]. □

**Ejemplo 2.5.** Según [16, p. 262], dadas  $X$  y  $Y$  variables aleatorias con *fdp* conjunta:

$$f_{(X,Y)}(x,y) = \begin{cases} x + y & \text{si } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en otro caso.} \end{cases}$$

*Determine:*

1.  $f_{X|Y}(x|y)$                       2.  $F_{X|Y}(x|y)$                       3.  $\mathbb{E}[X|Y = y]$ .

*Solución.* Primero se determina la **fdp** marginal de  $Y$ , es decir,

$$f_Y(y) = \int_0^{\infty} f_{(X,Y)}(x,y) dx = \int_0^1 (x+y) dx = \left( \frac{x^2}{2} + xy \right) \Big|_0^1 = \frac{1}{2} + y,$$

luego,

$$f_Y(y) = \begin{cases} \frac{1}{2} + y & \text{si } 0 < y < 1 \\ 0 & \text{en otro caso.} \end{cases}$$

Entonces,

1. La **fdp** condicional queda:

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)} = \begin{cases} \frac{x+y}{\frac{1}{2}+y} & \text{si } 0 < x < 1 \\ 0 & \text{en otro caso.} \end{cases}$$

2. La **fd** condicional queda

$$\begin{aligned} F_{X|Y}(x|y) &= \int_0^x f_{X|Y}(t|y) dt = \int_0^x \frac{t+y}{\frac{1}{2}+y} dt \\ &= \frac{1}{\frac{1}{2}+y} \left( \frac{t^2}{2} + ty \right) \Big|_0^x = \frac{\frac{x^2}{2} + xy}{\frac{1}{2}+y} \\ &= \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{x^2+2xy}{1+2y} & \text{si } 0 < x < 1 \\ 1 & \text{si } x \geq 1. \end{cases} \end{aligned}$$

3. El valor esperado condicional es

$$\begin{aligned} \mathbb{E}(X|Y = y) &= \int_0^1 x f_{X|Y}(x|y) dx = \int_0^1 x \frac{x+y}{\frac{1}{2}+y} dx \\ &= \frac{1}{\frac{1}{2}+y} \left( \frac{x^3}{3} + \frac{x^2}{2} y \right) \Big|_0^1 = \frac{1}{\frac{1}{2}+y} \left( \frac{1}{3} + \frac{1}{2} y \right). \quad \square \end{aligned}$$

**Ejemplo 2.6.** Según [16, p. 266], dadas  $X$  y  $Y$  variables aleatorias con **fdp** conjunta:

$$f(x,y) = \begin{cases} \frac{1}{2} y e^{-xy} & \text{si } x > 0, 0 < y < 2 \\ 0 & \text{en otro caso.} \end{cases}$$

Determine  $\mathbb{E}[\exp\{X/2\}|Y = 1]$ .

*Solución.* Primero se determina la **fdp** marginal de  $Y$ , es decir,

$$f_Y(y) = \int_0^\infty \frac{1}{2} y e^{-xy} dx = -\frac{1}{2} e^{-xy} \Big|_{x=0}^{x=\infty} = \frac{1}{2} \quad 0 < y < 2.$$

Nótese que  $Y \sim U(0, 2)$ . Luego,

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = y e^{-xy}, \quad x > 0.$$

Por lo tanto,

$$\begin{aligned} \mathbb{E}[e^{X/2}|Y = 1] &= \int_0^\infty e^{x/2} (y e^{-xy})|_{y=1} dx = \int_0^\infty e^{x/2} e^{-x} dx \\ &= -2e^{-x/2} \Big|_{x=0}^{x=\infty} = 2. \end{aligned} \quad \checkmark$$

**Teorema 2.2.** Sean  $X$  y  $Y$  variables aleatorias reales definidas sobre un espacio de probabilidad  $(\Omega, \mathcal{F}, P)$  y  $h(\cdot)$  una función tal que  $h(X)$  es una variable aleatoria. Si  $\mathbb{E}[h(X)]$  existe, entonces (ver [16, p. 268]):

$$\mathbb{E}[h(X)] = \mathbb{E}[\mathbb{E}[h(X)|Y]].$$

*Demostración.* Supóngase que  $X$  y  $Y$  son ambas variables aleatorias discretas. Entonces,

$$\begin{aligned} \mathbb{E}_Y[\underbrace{\mathbb{E}_X[h(X)|Y]}] &= \sum_y \underbrace{\mathbb{E}_X[h(X)|Y = y]} P(Y = y) \\ &= \sum_y \left[ \sum_x h(x) P(X = x|Y = y) \right] P(Y = y) \\ &= \sum_y \sum_x h(x) P(X = x, Y = y) \\ &= \sum_x h(x) \left[ \sum_y P(X = x, Y = y) \right] \\ &= \sum_x h(x) P(X = x) = \mathbb{E}[h(X)]. \end{aligned}$$

La cual es similar a la expresión (2.14). \checkmark

## 2.5. Varianza condicional

Sean  $X$  y  $Y$  variables aleatorias definidas sobre un mismo espacio de probabilidad. Si  $\mathbb{E}[Y^2] < \infty$ , la varianza condicional de  $Y$ , dada  $X = x$ , se define por:

$$\mathbb{V}[Y|X = x] = \mathbb{E}[Y^2|X = x] - (\mathbb{E}[Y|X = x])^2. \quad (2.16)$$

En general,  $\mathbb{E}[Y^k|X = x]$  para  $k = 1, 2, \dots$ , se establece como sigue:

$$\mathbb{E}[Y^k|X = x] = \begin{cases} \sum y^k P_{Y|X}(y|x) & \text{si } X, Y \text{ son discretas,} \\ \int_{-\infty}^{\infty} y^k f_{Y|X}(y|x) dy & \text{si } X, Y \text{ son continuas.} \end{cases} \quad (2.17)$$

**Teorema 2.3. Descomposición de la varianza.** Si  $X$  y  $Y$  son variables aleatorias distribuidas conjuntamente, entonces la varianza de cualquiera de ellas se puede descomponer como:

$$\mathbb{V}(Y) = \mathbb{E}[\mathbb{V}(Y|X)] + \mathbb{V}[\mathbb{E}(Y|X)]. \quad (2.18)$$

*Demostración.* Tomando esperanza en la expresión (2.16) se obtiene:

$$\begin{aligned} \mathbb{E}[\mathbb{V}(Y|X)] &= \mathbb{E}_X[\mathbb{E}_Y[Y^2|X]] - \mathbb{E}_X[(\mathbb{E}_Y[Y|X])^2] \\ &= \mathbb{E}(Y^2) - \mathbb{E}_X[\mathbb{E}_Y^2(Y|X)]. \end{aligned}$$

Por otra parte,

$$\begin{aligned} \mathbb{V}[\mathbb{E}(Y|X)] &= \mathbb{E}_X[\mathbb{E}_Y^2(Y|X)] - (\mathbb{E}_X[\mathbb{E}_Y(Y|X)])^2 \\ &= \mathbb{E}_X[\mathbb{E}_Y^2(Y|X)] - (\mathbb{E}(Y))^2. \end{aligned}$$

Sumando las dos últimas expresiones se obtiene el resultado.  $\square$

**Teorema 2.4. Descomposición de la covarianza.** Si  $X, Y$  y  $Z$  son variables aleatorias sobre el mismo espacio de probabilidad, y las covarianzas de  $X$  y  $Y$  son finitas, entonces:

$$\text{Cov}(X, Y) = \mathbb{E}(\text{Cov}(X, Y|Z)) + \text{Cov}(\mathbb{E}(X|Z), \mathbb{E}(Y|Z)). \quad (2.19)$$

*Demostración.* Esta fórmula se prueba usando la ley de esperanza total: por la expresión de la covarianza (1.44) se tiene que:

$$\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Al reescribir el lado derecho en términos de la ley de esperanza total condicionada por la variable aleatoria  $Z$  se obtiene que:

$$\text{Cov}[X, Y] = \mathbb{E}[\mathbb{E}[XY|Z]] - \mathbb{E}[\mathbb{E}[X|Z]]\mathbb{E}[\mathbb{E}[Y|Z]].$$

El término dentro de la primera esperanza del lado derecho se reescribe usando la definición de covarianza:

$$\begin{aligned} \text{Cov}[X, Y] &= \mathbb{E}[\text{Cov}[X, Y|Z] + \mathbb{E}[X|Z]\mathbb{E}[Y|Z]] - \mathbb{E}[\mathbb{E}[X|Z]]\mathbb{E}[\mathbb{E}[Y|Z]] \\ &= \mathbb{E}[\text{Cov}[X, Y|Z]] + \mathbb{E}[\mathbb{E}[X|Z]\mathbb{E}[Y|Z]] - \mathbb{E}[\mathbb{E}[X|Z]]\mathbb{E}[\mathbb{E}[Y|Z]], \end{aligned}$$

aquí se usó el hecho que el valor esperado es lineal. Finalmente, nótese que los dos últimos términos equivalen a la covarianza de la esperanzas condicionales  $\mathbb{E}[X|Z]$  y  $\mathbb{E}[Y|Z]$ :

$$\text{Cov}[X, Y] = \mathbb{E}[\text{Cov}(X, Y|Z)] + \text{Cov}[\mathbb{E}(X|Z), \mathbb{E}(Y|Z)]. \quad \checkmark$$

**Nota 2.3. Desigualdad de Cauchy-Schwarz.** Para dos variables aleatorias cualquiera  $X$  y  $Y$  con  $\mathbb{E}(X^2) < \infty$  y  $\mathbb{E}(Y^2) < \infty$  (ver [16, p. 210]),

$$|\mathbb{E}(XY)| \leq \mathbb{E}(|XY|) \leq [\mathbb{E}(|X|^2)]^{1/2} [\mathbb{E}(|Y|^2)]^{1/2}.$$

Entonces,

$$[\text{Cov}(X, Y)]^2 \leq \sigma_X^2 \sigma_Y^2$$

y se tiene la igualdad solo si  $X - \mu_X = c(Y - \mu_Y)$  para alguna constante  $c$ , es decir, si y solo si  $X$  y  $Y$  están linealmente relacionadas.

## 2.6. Correlación entre variables

Una medida de dependencia indica, en cierta medida, lo estrechamente relacionadas que están dos variables cuantitativas  $X$  y  $Y$ , con extremos en la independencia mutua y la dependencia mutua completa (si  $Y = g(X)$ , se dice que  $Y$  depende completamente de  $X$  o que están mutuamente determinadas). El coeficiente de correlación de Pearson cuantifica la dependencia lineal entre las variables  $X$  y  $Y$ , para luego poder compararlas.

**DEFINICIÓN 2.4.** De acuerdo con [20, p. 169], dadas  $X$  y  $Y$  variables aleatorias sobre una misma población; el coeficiente de correlación de Pearson se simboliza con la letra  $\rho_{X,Y}$ , siendo la expresión que nos permite calcularlo:

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} = \frac{\mathbb{E}[XY] - \mu_X \mu_Y}{\sigma_X \sigma_Y}, \quad (2.20)$$

donde:

- $\sigma_{XY}$  es la covarianza entre las variables  $X$  y  $Y$ .
- $\sigma_X$  y  $\sigma_Y$  son las desviaciones estándar de las variables  $X$  y  $Y$ , respectivamente.

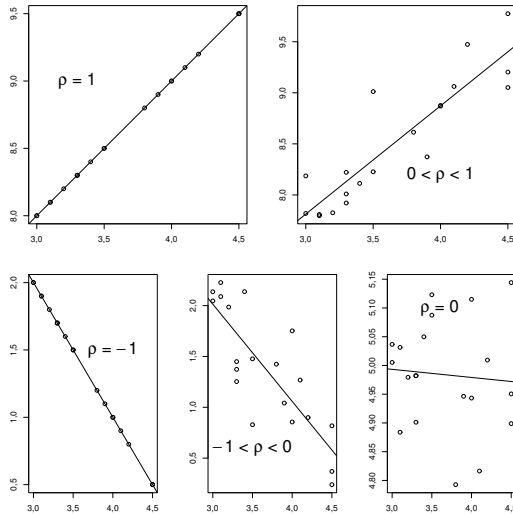


Figura 2.1. Coeficiente de correlación

**Nota 2.4.** La interpretación geométrica de las rectas presentadas en la figura 2.1 es

- Si  $\rho_{X,Y} = 1$ , existe una correlación positiva perfecta. El coeficiente indica una dependencia total entre las dos variables denominada “relación directa”: cuando una de ellas aumenta, la otra también lo hace en proporción constante.
- Si  $0 < \rho_{X,Y} < 1$ , existe una correlación positiva.
- Si  $\rho_{X,Y} = -1$ , existe una correlación negativa perfecta. El coeficiente indica una dependencia total entre las dos variables llamada “relación inversa”: cuando una de ellas aumenta, la otra disminuye en proporción constante.
- Si  $-1 < \rho_{X,Y} < 0$ , existe una correlación negativa.
- Si  $\rho_{X,Y} = 0$ , no existe relación lineal. Pero esto no necesariamente implica que las variables son independientes: pueden existir todavía relaciones no lineales entre las dos variables.

Nótese que el valor del coeficiente de correlación varía en el intervalo  $[-1, 1]$ , donde el signo indica el sentido de la relación.

**Teorema 2.5.** Sean  $X$  y  $Y$  variables aleatorias, entonces:

- 1)  $\rho_{X,Y} = \rho_{Y,X}$ . Si  $X = Y$  entonces  $\rho_{X,X} = 1$ .

- ii)  $\rho_{X,Y} = 0$ , si las variables  $X, Y$  son independientes<sup>11</sup>.
- iii)  $\rho_{aX+b,Y} = \rho_{X,aY+b} = \text{sign}(a)\rho_{X,Y}$ .
- iv)  $|\rho_{X,Y}| \leq 1$ .

*Demostración.* Ver detalles de la prueba en [20]. ✓

**Nota 2.5.** De manera análoga, se puede calcular este coeficiente sobre una muestra de tamaño  $n$ , denotado como  $r_{xy}$  así:

$$\begin{aligned}
 r_{xy} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y} \tag{2.21} \\
 &= \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2}}.
 \end{aligned}$$

**Ejemplo 2.7.** Según [16, p. 226] dadas dos variables aleatorias  $X$  y  $Y$  con fmp conjunta:

\	Y	-1	0	1
X		1/6	2/6	0
	1	1/6	2/6	0
	2	1/6	1/6	1/6

Definiendo  $\mathbf{X} = (X, Y)$ . Determine  $\mathbb{E}(\mathbf{X})$ ,  $\mathbb{V}(\mathbf{X})$  y  $\rho_{\mathbf{X}}$ .

*Solución.* Primero se determina las fmp marginales:

\	Y	-1	0	1	$P(X = x)$
X		1/6	2/6	0	1/2
	1	1/6	2/6	0	1/2
	2	1/6	1/6	1/6	1/2
	$P(Y = y)$	1/3	1/2	1/6	

Si  $\mathbf{A}$  es la matriz cuyos elementos son  $a_{ij} = p(x_i, y_j)$ , entonces:

$$\mathbb{E}(\mathbf{X}) = \mathbf{X}^T \mathbf{A} \mathbf{1}_3 = \begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 1/6 & 2/6 & 0 \\ 1/6 & 1/6 & 1/6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \frac{3}{2}$$

<sup>11</sup>El recíproco no es válido, ya que hay variables aleatorias con  $\rho_{X,Y} = 0$ , pero dependientes, un ejemplo se puede consultar en [20, p. 171].

y

$$\mathbb{E}(Y) = \mathbf{1}_2^\top \mathbf{A} \mathbf{Y} = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1/6 & 2/6 & 0 \\ 1/6 & 1/6 & 1/6 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = -\frac{1}{6}.$$

Luego,  $\mathbb{E}(\mathbf{X}) = (3/2, -1/6)$ . Ahora se halla:

$$\mathbb{E}(X^2) = 5/2 \qquad \mathbb{E}(Y^2) = 1/2,$$

por lo tanto,

$$\mathbb{V}(X) = 1/4 \qquad \mathbb{V}(Y) = 17/36.$$

Finalmente, se encuentra la covarianza de la siguiente manera:

$$\underbrace{(\mathbf{X} - \mathbb{E}(\mathbf{X}))^\top \mathbf{A} (\mathbf{Y} - \mathbb{E}(\mathbf{Y}))}_{\text{Cov}(X,Y)} = \begin{pmatrix} -1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 1/6 & 2/6 & 0 \\ 1/6 & 1/6 & 1/6 \end{pmatrix} \begin{pmatrix} -5/6 \\ 1/6 \\ 7/6 \end{pmatrix}$$

$$\text{Cov}(X,Y) = \begin{pmatrix} 0 & -1/12 & 1/12 \end{pmatrix} \begin{pmatrix} -5/6 \\ 1/6 \\ 7/6 \end{pmatrix} = \frac{1}{12}.$$

También, la covarianza se puede establecer como:

$$\underbrace{\text{Cov}(X,Y)}_{\sigma_{XY}} = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbf{X}^\top \mathbf{A} \mathbf{Y} - (\mathbf{X}^\top \mathbf{A} \mathbf{1}_3)(\mathbf{1}_2^\top \mathbf{A} \mathbf{Y})$$

$$= \mathbf{X}^\top \mathbf{A} \mathbf{Y} - \mathbf{X}^\top \mathbf{A} (\mathbf{1}_3 \mathbf{1}_2^\top) \mathbf{A} \mathbf{Y} = \mathbf{X}^\top (\mathbf{A} - \mathbf{A} \mathbf{J} \mathbf{A}) \mathbf{Y}, \quad (2.22)$$

donde  $\mathbf{J} = \mathbf{1}_3 \mathbf{1}_2^\top$  es una matriz de unos, por lo que:

$$(\mathbf{I}_2 - \mathbf{A} \mathbf{J}) \mathbf{A} = \left[ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1/6 & 2/6 & 0 \\ 1/6 & 1/6 & 1/6 \end{pmatrix} \right] \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1/6 & 2/6 & 0 \\ 1/6 & 1/6 & 1/6 \end{pmatrix}$$

$$= \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 1/6 & 2/6 & 0 \\ 1/6 & 1/6 & 1/6 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{12} & -\frac{1}{12} \\ 0 & -\frac{1}{12} & \frac{1}{12} \end{pmatrix}.$$

Por lo tanto, al sustituir en (2.22) se tiene que:

$$\text{Cov}(X,Y) = \begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{12} & -\frac{1}{12} \\ 0 & -\frac{1}{12} & \frac{1}{12} \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \frac{1}{12}.$$

Por consiguiente, la matriz de varianza covarianza queda:

$$\Sigma_X = \begin{pmatrix} \frac{1}{4} & \frac{1}{12} \\ \frac{1}{12} & \frac{1}{36} \end{pmatrix} \Rightarrow \rho_X = \frac{\frac{1}{12}}{\sqrt{\frac{1}{4} \cdot \frac{1}{36}}} = \frac{1}{\sqrt{17}},$$

aquí se estableció el coeficiente de correlación mediante (2.20). ☑

## 2.7. Distribución de suma de variables aleatorias independientes

En muchos casos es conveniente definir nuevas variables aleatorias en términos de  $X_1$  y  $X_2$ . Cuando estas variables son no negativas, la variable  $S = X_1 + X_2$  es de gran interés. Denotando las fda de  $X_1$  y  $X_2$  por  $F$  y  $G$ , respectivamente, la convolución entre las funciones  $F$  y  $G$  es un operador matemático que las convierte en una nueva función  $H$ , que representa la magnitud en la que se superponen  $F$  y una versión trasladada e invertida de  $G$ . En este caso, la fdp,  $h_S(s)$ , y fda,  $H_S(s)$ , están dadas por:

$$h_S(s) = \begin{cases} \sum_{y \leq s} P[X_1 = s - y, X_2 = y] & \text{si } X_1, X_2 \text{ son discretas,} \\ \int_0^s f_{(X_1, X_2)}(s - y, y) dy & \text{si } X_1, X_2 \text{ son continuas.} \end{cases} \quad (2.23)$$

$$H_S(s) = \begin{cases} \sum_{x+y \leq s} \sum P[X_1 = x, X_2 = y] & \text{si } X_1, X_2 \text{ son discretas,} \\ \iint_{\{x+y \leq s\}} f_{(X_1, X_2)}(x, y) dx dy & \text{si } X_1, X_2 \text{ son continuas.} \end{cases} \quad (2.24)$$

Cuando las variables  $X_1$  y  $X_2$  son independientes, entonces:

$$h_S(s) = \begin{cases} \sum_{y \leq s} P[X_1 = s - y]P[X_2 = y] & \text{si } X_1, X_2 \text{ son discretas,} \\ \int_0^s f_{X_1}(s - y)d[G_{X_2}(y)] & \text{si } X_1, X_2 \text{ son continuas.} \end{cases} \quad (2.25)$$

$$\int_0^s f_{X_1}(s - y)d[G_{X_2}(y)] \quad \text{si } X_1, X_2 \text{ son continuas.} \quad (2.26)$$

$$H_S(s) = \begin{cases} \sum_{y \leq s} F_{X_1}(s - y)P[X_2 = y] & \text{si } X_1, X_2 \text{ son discretas,} \\ \int_0^s F_{X_1}(s - y)d[G_{X_2}(y)] & \text{si } X_1, X_2 \text{ son continuas.} \end{cases} \quad (2.27)$$

$$\int_0^s F_{X_1}(s - y)d[G_{X_2}(y)] \quad \text{si } X_1, X_2 \text{ son continuas.} \quad (2.28)$$

Por la propiedad de simetría, las fda  $F$  y  $G$  se pueden intercambiar sin afectar la convolución. Las variables aleatorias positivas son de interés especial, ya que si  $f$  y  $g$  están definidas en  $(0, \infty)$ , la convolución  $f * g$  dada en (2.26)

queda:

$$(f * g)(s) = \int_0^s f_{X_1}(s-y)g_{X_2}(y) dy = \int_0^s f_{X_1}(x)g_{X_2}(s-x) dx. \quad (2.29)$$

**Teorema 2.6. Propiedades fundamentales de las convoluciones.** Sean  $f, g$  y  $h$  fdp, la convolución de ellas satisface que:

1. *Conmutatividad:*  $(f * g)(s) = (g * f)(s)$ .
2. *Asociatividad:*  $[f * (g * h)](s) = [(f * g) * h](s)$ .
3. *Distributividad:*  $[f * (g + h)](s) = (f * g)(s) + (f * h)(s)$ .
4. *Multipliación por escalar:*  $[\alpha(f * g)](s) = [(\alpha f) * g](s) = [f * (\alpha g)](s)$ .
5. *Regla de derivación:*  $\mathcal{D}[(f * g)(s)] = [\mathcal{D}(f) * g](s) = [f * \mathcal{D}(g)](s)$ , donde  $\mathcal{D} = \frac{d}{ds}$  denota el operador derivada.

*Demostración.* Queda como ejercicio para el lector. ☑

**Ejemplo 2.8.** Según [85, p. 395], dadas  $X_1, X_2$  variables aleatorias independientes Poisson con parámetros  $\lambda$  y  $\mu$  respectivamente. Sea  $S = X_1 + X_2$ , calcule  $h_S(s)$  usando:

- 1) El método de convoluciones.
- 2) La propiedad 4 de la fgm.

*Solución.*

- 1) Usando la expresión (2.25) de convolución se tiene que:

$$\begin{aligned} h_S(s) &= \sum_{x \leq s} P[X_2 = s-x]P[X_1 = x] = \sum_{x=0}^s e^{-\mu} \frac{\mu^{s-x}}{(s-x)!} e^{-\lambda} \frac{\lambda^x}{x!} \\ &= \frac{e^{-(\lambda+\mu)}}{s!} \sum_{x=0}^s \binom{s}{x} \mu^{s-x} \lambda^x = \frac{(\mu + \lambda)^s}{s!} e^{-(\lambda+\mu)}, \end{aligned} \quad (2.30)$$

luego,  $S$  se distribuye Poisson con parámetro  $\lambda + \mu$ .

- 2) Empleando la propiedad 4 de la fgm:

$$\begin{aligned} m_{X_1+X_2}(t) &= m_{X_1}(t)m_{X_2}(t) = \exp\{\lambda(\exp(t) - 1)\} \exp\{\mu(\exp(t) - 1)\} \\ &= \exp\{(\lambda + \mu)(\exp(t) - 1)\}, \end{aligned}$$

que corresponde a la fgm de una variable aleatoria con distribución Poisson de parámetro  $\lambda + \mu$ . Por lo que,

$$X_1 + X_2 \sim P(\lambda + \mu). \quad \square$$

**Ejemplo 2.9.** De acuerdo con [75, p. 66], dadas  $X_1$  y  $X_2$  variables aleatorias independientes donde  $X_1 \sim B(n_1, p)$  y  $X_2 \sim B(n_2, p)$ . ¿Cuál es la distribución de  $X_1 + X_2$ ?

*Solución.* Usando la propiedad 4 de la fgm:

$$\begin{aligned} m_{X_1+X_2}(t) &= m_{X_1}(t)m_{X_2}(t) = [pe^t + (1-p)]^{n_1} [pe^t + (1-p)]^{n_2} \\ &= [pe^t + (1-p)]^{n_1+n_2}, \end{aligned}$$

que corresponde a la fgm de una variable aleatoria con distribución binomial de parámetro  $n_1 + n_2$  y  $p$ . Entonces:

$$X_1 + X_2 \sim B(n_1 + n_2, p). \quad \checkmark$$

**Ejemplo 2.10.** De acuerdo con [75, p. 67], dadas  $X_1$  y  $X_2$  variables aleatorias independientes donde  $X_1 \sim N(\mu_1, \sigma_1^2)$  y  $X_2 \sim N(\mu_2, \sigma_2^2)$ . ¿Cuál es la distribución de  $X_1 + X_2$ ?

*Solución.* Usando la propiedad 4 de la fgm:

$$\begin{aligned} m_{X_1+X_2}(t) &= m_{X_1}(t)m_{X_2}(t) = \exp\left\{t\mu_1 + \frac{\sigma_1^2 t^2}{2}\right\} \exp\left\{t\mu_2 + \frac{\sigma_2^2 t^2}{2}\right\} \\ &= \exp\left\{t\mu_1 + t\mu_2 + \frac{\sigma_1^2 t^2}{2} + \frac{\sigma_2^2 t^2}{2}\right\} = \exp\left\{t(\mu_1 + \mu_2) + \frac{t^2}{2}(\sigma_1^2 + \sigma_2^2)\right\}, \end{aligned}$$

que corresponde a la fgm de una variable aleatoria con distribución normal de parámetros  $\mu_1 + \mu_2$  y  $\sigma_1^2 + \sigma_2^2$ . Así:

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2). \quad \checkmark$$

**Ejemplo 2.11.** Sean  $X_1$  y  $X_2$  dos variables aleatorias independientes donde  $X_1 \sim Exp(\lambda)$  y  $X_2 \sim Exp(\lambda)$ . ¿Cuál es la distribución de  $X_1 + X_2$ ?

*Solución.* Usando la propiedad 4 de la fgm:

$$m_{X_1+X_2}(t) = m_{X_1}(t)m_{X_2}(t) = \left(\frac{1}{1-\frac{t}{\lambda}}\right)\left(\frac{1}{1-\frac{t}{\lambda}}\right) = \left(\frac{1}{1-\frac{t}{\lambda}}\right)^2$$

que corresponde a la fgm de una variable aleatoria con distribución gamma de parámetros 2 y  $\lambda$ . Así:

$$X_1 + X_2 \sim gamma(2, \lambda). \quad \checkmark$$

**Ejemplo 2.12.** Sean  $X_1$  y  $X_2$  dos variables aleatorias independientes tal que  $X_1 \sim \text{gamma}(r_1, \lambda)$  y  $X_2 \sim \text{gamma}(r_2, \lambda)$ . ¿Cuál es la distribución de  $X_1 + X_2$ ?

*Solución.* Usando la propiedad 4 de la fgm:

$$m_{X_1+X_2}(t) = m_{X_1}(t)m_{X_2}(t) = \left(\frac{1}{1-\frac{t}{\lambda}}\right)^{r_1} \left(\frac{1}{1-\frac{t}{\lambda}}\right)^{r_2} = \left(\frac{1}{1-\frac{t}{\lambda}}\right)^{r_1+r_2},$$

que corresponde a la fgm de una variable aleatoria con distribución gamma de parámetros  $r_1 + r_2$  y  $\lambda$ . Así:

$$X_1 + X_2 \sim \text{gamma}(r_1 + r_2, \lambda). \quad \checkmark$$

**Ejemplo 2.13.** Según [16, p. 261], dadas  $X_1$  y  $X_2$  variables aleatorias independientes con fmp Poisson de parámetros  $\lambda$  y  $\mu$ , respectivamente. Calcular el valor esperado de  $X_1$  bajo la condición  $X_1 + X_2 = n$ , con  $n$  entero no negativo.

*Solución.* Como:

$$P(X_1 = x | S = n) = \frac{P(X_1 = x, X_2 = n - x)}{P(S = n)} = \frac{P[X_1 = x]P[X_2 = n - x]}{P(S = n)},$$

donde  $S = X_1 + X_2$ , por la expresión (2.30) se tiene que:

$$P(S = n) = \frac{(\lambda + \mu)^n}{n!} e^{-(\lambda + \mu)}.$$

Por lo tanto,

$$\begin{aligned} P(X_1 = x | X_1 + X_2 = n) &= \frac{e^{-\lambda} \frac{\lambda^x}{x!} e^{-\mu} \frac{\mu^{n-x}}{(n-x)!}}{\frac{(\lambda + \mu)^n}{n!} e^{-(\lambda + \mu)}} = \frac{n!}{(n-x)! x!} \frac{\lambda^x \mu^{n-x}}{(\lambda + \mu)^n} \\ &= \binom{n}{x} \left(\frac{\lambda}{\lambda + \mu}\right)^x \left(\frac{\mu}{\lambda + \mu}\right)^{n-x}. \end{aligned}$$

Es decir,  $X_1$  bajo la condición  $X_1 + X_2 = n$  se distribuye binomial con parámetros  $n$  y  $p = \frac{\lambda}{\lambda + \mu}$ . Por lo tanto,

$$\mathbb{E}(X_1 | X_1 + X_2 = n) = n \frac{\lambda}{\lambda + \mu}. \quad \checkmark$$

## 2.8. Transformaciones de variables aleatorias

En algunas ocasiones es necesario conocer la distribución de probabilidad de una función o transformación de una variable aleatoria. Hasta este momento,

estamos en condiciones de modelar un fenómeno en términos de una variable aleatoria  $X$  mediante su fda  $F_X(x)$ . La pregunta es si estamos además en condiciones de estudiar el comportamiento de aquellas variables aleatorias que son transformaciones de  $X$ . En esta sección se presenta una técnica que permite conocer el comportamiento probabilístico de  $U = g(X)$ .

**Teorema 2.7. Teorema del cambio de variable.** *Sea  $X$  una variable aleatoria con fdp  $f_X$  y sea  $g$  una función estrictamente monótona y diferenciable, con inversa diferenciable  $g^{-1}$ . Entonces la fdp de la variable aleatoria  $U = g(X)$  es*

$$f_U(u) = f_X[g^{-1}(u)] \left| \frac{d}{du} g^{-1}(u) \right|. \quad (2.31)$$

Se puede expresar de una manera más intuitiva como:

$$f_U(u) = f_X(y) \left| \frac{dy}{du} \right| \quad \text{con} \quad y = g^{-1}(u).$$

Si  $g$  no es biyectiva, entonces se pueden encontrar subconjuntos de  $\mathbb{R}$  en los que  $g$  es biyectiva, y aplicar el teorema separadamente en cada subconjunto.

*Demostración.* Ver detalles de la prueba en [20, p. 51]. ✓

**Ejemplo 2.14.** *Sea  $X$  una variable aleatoria con fdp normal estándar, defina  $U = e^X$ , determine la fdp de  $U$ .*

*Solución.* Considere  $u = g(x) = e^x$ , la cual es una función creciente de  $x$  y, además, que  $g(x)$  es diferenciable. Ahora:

$$u = g(x) = e^x \quad \implies \quad g^{-1}(u) = x = \ln u.$$

Luego,

$$\frac{d}{du} g^{-1}(u) = \frac{1}{u}.$$

Puesto que  $U = e^X$  satisface las condiciones del teorema 2.7, se tiene que:

$$f_U(u) = f_X[g^{-1}(u)] \left| \frac{d}{du} g^{-1}(u) \right| = \frac{1}{\sqrt{2\pi}u} \exp \left\{ -\frac{1}{2} (\ln u)^2 \right\}.$$

Por consiguiente,

$$f_U(u) = \begin{cases} \frac{1}{\sqrt{2\pi}u} \exp \left\{ -\frac{1}{2} (\ln u)^2 \right\} & \text{si } u > 0 \\ 0 & \text{en otra parte,} \end{cases}$$

es decir,  $U \sim LN(0, 1)$ . ✓

## 2.9. Normal bivariada

El análisis de nubes de datos fue realizado por primera vez por el científico británico Francis Galton (1822 - 1911), quien describió las propiedades de la distribución normal bivariada y su relación con el análisis de regresión y también introdujo el concepto de correlación, posteriormente desarrollado por Pearson y Sperman.

### 2.9.1. Distribución normal bivariada estándar

Las variables aleatorias  $X_1$  y  $X_2$  se dice que tienen distribución normal bivariada estándar con correlación  $\rho$  si y solo si:

$$\underbrace{\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}}_{12} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \quad \Rightarrow \quad \begin{cases} X_1 = Z_1 \\ X_2 = \rho Z_1 + \sqrt{1-\rho^2} Z_2, \end{cases}$$

donde  $Z_1$  y  $Z_2$  son variables normales estándar independientes. En cuanto a sus marginales, tanto  $X_1$  como  $X_2$  siguen una **fdp** normal estándar. Además,

$$(X_2|X_1 = x_1) \sim N(\rho x_1, 1 - \rho^2) \quad (2.32)$$

$$(X_1|X_2 = x_2) \sim N(\rho x_2, 1 - \rho^2).$$

Si  $\mathbf{X} = (X_1, X_2)$ , la **fdp** conjunta de  $X_1$  y  $X_2$  está dada por:

$$f_{\mathbf{X}}(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)\right\}. \quad (2.33)$$

Si  $X_1$  y  $X_2$  tienen **fdp** normal bivariada estándar, se dice que  $X_1$  y  $X_2$  son independientes si y solo si  $\rho = 0$ .

### 2.9.2. Distribución normal bivariada

Si  $\mathbf{X} = (X_1, X_2)$  es un vector aleatorio en  $\mathbb{R}^2$  tal que  $X_1 \sim N(\mu_1, \sigma_1^2)$ ,  $X_2 \sim N(\mu_2, \sigma_2^2)$  y  $\rho(X_1, X_2) = \rho$ , definiendo:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix},$$

<sup>12</sup>Aplicando descomposición de Cholesky a la matriz de correlaciones.

donde  $Z_1$  y  $Z_2$  son variables normales estándar independientes, por lo tanto:

$$X_1 = \mu_1 + \sigma_1 Z_1 \quad \text{y} \quad X_2 = \mu_2 + \sigma_2 (\rho Z_1 + \sqrt{1 - \rho^2} Z_2),$$

entonces la **fdp** conjunta de  $X_1$  y  $X_2$  tiene una distribución normal bivariada con media  $\boldsymbol{\mu}_X$  y matriz de varianza-covarianza  $\boldsymbol{\Sigma}_X$ , dada por:

$$f_X(\mathbf{x}) = \frac{1}{2\pi|\boldsymbol{\Sigma}_X|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_X)^\top \boldsymbol{\Sigma}_X^{-1}(\mathbf{x} - \boldsymbol{\mu}_X)\right\}. \quad (2.34)$$

En este caso, el determinante está dado por:

$$|\boldsymbol{\Sigma}_X| = \begin{vmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{vmatrix} = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = (1 - \rho^2)\sigma_1^2\sigma_2^2,$$

y la inversa está dada por:

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}^{-1} = \frac{(1 - \rho^2)^{-1}}{\sigma_1^2\sigma_2^2} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} = \frac{1}{1 - \rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix},$$

por lo que la forma cuadrática que aparece en el exponente sería:

$$\begin{aligned} & (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}_X^{-1}(\mathbf{x} - \boldsymbol{\mu}) \\ &= \frac{1}{1 - \rho^2} \left( \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \right)^\top \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix} \left( \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \right) \\ &= \frac{1}{1 - \rho^2} \begin{pmatrix} x_1 - \mu_1 & x_2 - \mu_2 \end{pmatrix} \begin{pmatrix} \frac{x_1 - \mu_1}{\sigma_1^2} - \frac{\rho(x_2 - \mu_2)}{\sigma_1\sigma_2} \\ \frac{x_2 - \mu_2}{\sigma_2^2} - \frac{\rho(x_1 - \mu_1)}{\sigma_1\sigma_2} \end{pmatrix} \\ &= \frac{1}{1 - \rho^2} \left[ \left( \frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left( \frac{x_1 - \mu_1}{\sigma_1} \right) \left( \frac{x_2 - \mu_2}{\sigma_2} \right) + \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right]. \end{aligned}$$

Luego, la **fdp** dada en (2.34) queda:

$$\begin{aligned} f_X(\mathbf{x}) &= \frac{1}{2\pi((1 - \rho^2)\sigma_1^2\sigma_2^2)^{\frac{1}{2}}} \times \\ & \exp\left\{-\frac{1}{2(1 - \rho^2)} \left[ \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\frac{\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right]\right\}. \end{aligned}$$

Haciendo los cambios de variable,

$$y_1 = \frac{x_1 - \mu_1}{\sigma_1} \quad \text{y} \quad y_2 = \frac{x_2 - \mu_2}{\sigma_2},$$

entonces,  $f(y_1, y_2)$  coincide con la expresión (2.33), es decir, es una **fdp** normal bivariada estándar.

### 2.9.2.1. Transformación lineal de las variables

Usando la transformación dada en (B.4) (ver apéndice B.1.1), se tiene que:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} = \mathbf{A}_\theta^\top (\mathbf{x} - \boldsymbol{\mu}). \quad (2.35)$$

Luego, el exponente de la expresión (2.34) se reescribe como:

$$(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}_X^{-1} (\mathbf{x} - \boldsymbol{\mu}) = (\mathbf{A}_\theta \mathbf{y})^\top \boldsymbol{\Sigma}_X^{-1} (\mathbf{A}_\theta \mathbf{y}) = \mathbf{y}^\top (\mathbf{A}_\theta^\top \boldsymbol{\Sigma}_X \mathbf{A}_\theta)^{-1} \mathbf{y} = \mathbf{y}^\top \boldsymbol{\Sigma}_Y^{-1} \mathbf{y},$$

ya que  $\mathbf{A}_\theta$  es una matriz ortogonal, entonces  $\mathbf{A}_\theta^\top = \mathbf{A}_\theta^{-1}$ . Luego,

$$\begin{aligned} \boldsymbol{\Sigma}_Y &= \begin{pmatrix} \sigma_{y_1}^2 & \sigma_{y_1 y_2} \\ \sigma_{y_1 y_2} & \sigma_{y_2}^2 \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1^2 \cos^2 \theta + \sigma_2^2 \sin^2 \theta + \rho \sigma_1 \sigma_2 \sin 2\theta & \rho \sigma_1 \sigma_2 \cos 2\theta - \frac{1}{2}(\sigma_1^2 - \sigma_2^2) \sin 2\theta \\ \rho \sigma_1 \sigma_2 \cos 2\theta - \frac{1}{2}(\sigma_1^2 - \sigma_2^2) \sin 2\theta & \sigma_1^2 \sin^2 \theta + \sigma_2^2 \cos^2 \theta - \rho \sigma_1 \sigma_2 \sin 2\theta \end{pmatrix}, \end{aligned}$$

por la expresión (B.6), el ángulo de rotación  $\theta$  se puede establecer como:

$$\tan(2\theta) = 2 \frac{\rho \sigma_1 \sigma_2}{\sigma_1^2 - \sigma_2^2} \quad 0 < \sigma_2 < \sigma_1, \quad (2.36)$$

nótese entonces que la covarianza entre las coordenadas del vector  $\mathbf{Y}$ , cuyo valor depende del ángulo  $\theta$ , es igual a cero. Por otra parte, usando las siguientes identidades trigonométricas (ver [5, p. 120])

$$2 \cos^2 \theta = 1 + \cos(2\theta) \quad \text{y} \quad 2 \sin^2 \theta = 1 - \cos(2\theta), \quad (2.37)$$

y a partir de la expresión (2.36) se deducen las funciones trigonométricas para el ángulo doble en términos de las varianzas iniciales,  $\sigma_1^2$  y  $\sigma_2^2$ ,

$$\sin(2\theta) = \frac{2\rho\kappa}{\sqrt{(\kappa^2 - 1)^2 + 4\rho^2\kappa^2}} \quad \text{y} \quad \cos(2\theta) = \frac{\kappa^2 - 1}{\sqrt{(\kappa^2 - 1)^2 + 4\rho^2\kappa^2}},$$

donde  $\kappa = \sigma_1/\sigma_2$ , sustituyendo en (2.37) el valor de  $\cos(2\theta)$  se tiene que:

$$\cos^2 \theta = \frac{1}{2} + \frac{(\kappa^2 - 1)/2}{\sqrt{(\kappa^2 - 1)^2 + 4\rho^2\kappa^2}} \quad \text{y} \quad \sin^2 \theta = \frac{1}{2} - \frac{(\kappa^2 - 1)/2}{\sqrt{(\kappa^2 - 1)^2 + 4\rho^2\kappa^2}}.$$

Empleando estas expresiones, los elementos de la diagonal de  $\boldsymbol{\Sigma}_Y$  se definen en términos de las varianzas iniciales; es decir,

$$\begin{aligned}
 \sigma_{y_1}^2 &= \sigma_1^2 \cos^2 \theta + \sigma_2^2 \sin^2 \theta + \rho \sigma_1 \sigma_2 \sin(2\theta) = \rho \sigma_1 \sigma_2 \sin(2\theta) \\
 &\quad + \frac{\sigma_1^2}{2} \left( 1 + \frac{\kappa^2 - 1}{\sqrt{(\kappa^2 - 1)^2 + 4\rho^2 \kappa^2}} \right) + \frac{\sigma_2^2}{2} \left( 1 - \frac{\kappa^2 - 1}{\sqrt{(\kappa^2 - 1)^2 + 4\rho^2 \kappa^2}} \right) \\
 &= \frac{\sigma_1^2 + \sigma_2^2}{2} + \frac{(\sigma_1^2 - \sigma_2^2)^2}{2\sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\rho^2 \sigma_1^2 \sigma_2^2}} + \frac{2(\rho \sigma_1 \sigma_2)^2}{\sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\rho^2 \sigma_1^2 \sigma_2^2}} \\
 &= \frac{\sigma_1^2 + \sigma_2^2}{2} + \frac{(\sigma_1^2 - \sigma_2^2)^2 + 4\rho^2 \sigma_1^2 \sigma_2^2}{2\sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\rho^2 \sigma_1^2 \sigma_2^2}} \\
 &= \frac{\sigma_1^2 + \sigma_2^2}{2} + \sqrt{\left(\frac{\sigma_1^2 - \sigma_2^2}{2}\right)^2 + \rho^2 \sigma_1^2 \sigma_2^2}. \tag{2.38}
 \end{aligned}$$

De manera análoga, el otro elemento de la diagonal queda:

$$\begin{aligned}
 \sigma_{y_2}^2 &= \sigma_1^2 \sin^2 \theta + \sigma_2^2 \cos^2 \theta - \rho \sigma_1 \sigma_2 \sin(2\theta) = -\rho \sigma_1 \sigma_2 \sin(2\theta) \\
 &\quad + \frac{\sigma_1^2}{2} \left( 1 + \frac{\kappa^2 - 1}{\sqrt{(\kappa^2 - 1)^2 + 4\rho^2 \kappa^2}} \right) + \frac{\sigma_2^2}{2} \left( 1 - \frac{\kappa^2 - 1}{\sqrt{(\kappa^2 - 1)^2 + 4\rho^2 \kappa^2}} \right) \\
 &= \frac{\sigma_1^2 + \sigma_2^2}{2} - \frac{(\sigma_1^2 - \sigma_2^2)^2}{2\sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\rho^2 \sigma_1^2 \sigma_2^2}} - \frac{2(\rho \sigma_1 \sigma_2)^2}{\sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\rho^2 \sigma_1^2 \sigma_2^2}} \\
 &= \frac{\sigma_1^2 + \sigma_2^2}{2} - \frac{(\sigma_1^2 - \sigma_2^2)^2 + 4\rho^2 \sigma_1^2 \sigma_2^2}{2\sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\rho^2 \sigma_1^2 \sigma_2^2}} \\
 &= \frac{\sigma_1^2 + \sigma_2^2}{2} - \sqrt{\left(\frac{\sigma_1^2 - \sigma_2^2}{2}\right)^2 + \rho^2 \sigma_1^2 \sigma_2^2}. \tag{2.39}
 \end{aligned}$$

Las expresiones (2.38) y (2.39) determinan las componentes de  $\Sigma_Y$ , luego,

$$\Sigma_Y = \begin{pmatrix} \sigma_{y_1}^2 & 0 \\ 0 & \sigma_{y_2}^2 \end{pmatrix} = \frac{\sigma_1^2 + \sigma_2^2}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sqrt{\left(\frac{\sigma_1^2 - \sigma_2^2}{2}\right)^2 + \rho^2 \sigma_1^2 \sigma_2^2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

nótese que  $|\Sigma_Y| = |\Sigma_X|$ . Por lo tanto, la **fdp** de  $Y$  es dada por:

$$f_Y(\mathbf{y}) = \frac{1}{2\pi \sigma_{y_1} \sigma_{y_2}} \exp \left\{ -\frac{1}{2} \left[ \left( \frac{y_1}{\sigma_{y_1}} \right)^2 + \left( \frac{y_2}{\sigma_{y_2}} \right)^2 \right] \right\}, \quad \text{con } \sigma_{y_1} > \sigma_{y_2},$$

donde  $Y$  es dado en (2.35).

**Ejemplo 2.15.** Supóngase que el vector aleatorio  $X$  es normal bivariado con vector de medias  $\mu_X$  y matriz de varianza-covarianza  $\Sigma_X$  dados por:

$$\mu_X = \begin{pmatrix} 0 \\ 5 \end{pmatrix} \quad y \quad \Sigma_X = \begin{pmatrix} 5 & (0,5)\sqrt{5}\sqrt{8} \\ (0,5)\sqrt{5}\sqrt{8} & 8 \end{pmatrix}.$$

Encuentre la gráfica de la *fdp* del vector  $X$ .

*Solución.* En la figura 2.2 se presenta la *fdp* del vector  $X$  normal bivariado:

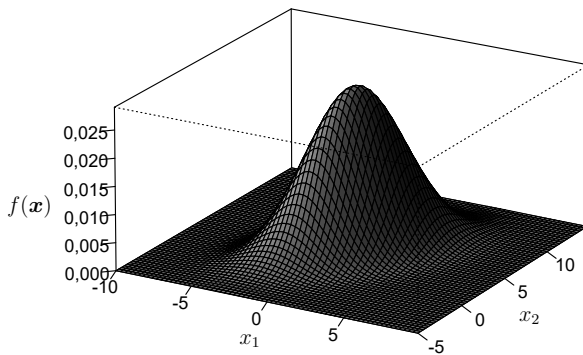


Figura 2.2. *fdp* de la normal bivariada en 3D

Las instrucciones en  $R$  para el gráfico son:

```
library(mvtnorm)
remove(list=ls())
x1 = seq(-10,10,0.3); x2 = seq(-5,15,0.3)
media = c(0,5)
Varianza=matrix(c(5,(0.5)*sqrt(5)*sqrt(8),(0.5)*sqrt(5)*sqrt(8),
8),2)
z = matrix(0,length(x1),length(x2))
for (i in 1:length(x1)){
  for (j in 1:length(x2)){
    z[i,j] = dmvnorm(c(x1[i],x2[j]),mean=media,sigma=Varianza)
  }
}
persp(x1, x2, z, col="lightgrey", theta=30, phi=20, r=50, d=0.1,
expand=0.5,ltheta=90,lphi=180,shade=0.75,ticktype="detailed",
nticks=5,xlab=" ",ylab=" ",zlab=" ")
```



## 2.10. Familia exponencial lineal

Muchas de las **fdp** utilizadas en el campo estadístico pertenecen a la familia exponencial lineal, esta es una ventaja importante con respecto a otros modelos que no pertenecen a esta familia, ventaja que es una característica cuando se trata de calcular el estadístico de una muestra aleatoria.

### 2.10.1. Familia exponencial lineal (uniparamétrica)

La familia de funciones de distribuciones de un modelo  $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$  es de la familia exponencial de un parámetro si existen:

- $c(\theta)$  y  $d(\theta)$  funciones de valor real sobre  $\Theta$ ,
- $T(x)$  y  $S(x)$  funciones de valor real sobre  $\mathbb{R}^n$  y
- Un conjunto  $A \subset \mathbb{R}^n$ , que no depende de  $\theta$ , tal que la **fdp**  $p(x; \theta)$  de  $P_\theta$  se parametriza en términos del parámetro  $\theta$ , de tal forma que:

$$p(x; \theta) = \exp [c(\theta)T(x) + d(\theta) + S(x)] I_A(x), \quad (2.40)$$

donde  $I_A(x)$  es la función indicadora del conjunto  $A$  dada en (1.10).

**Nota 2.6.** Las funciones  $c(\theta)$ ,  $d(\theta)$ ,  $T(x)$  y  $S(x)$  no son únicas.

**Ejemplo 2.16.** (Familia binomial). Sea  $X$  una variable aleatoria con distribución  $\text{Bin}(m, \theta)$ ,  $0 < \theta < 1$ . Demuestre que es un miembro de la familia exponencial.

*Solución.* Reescribiendo la expresión (1.12) se tiene que:

$$\begin{aligned} p(x; \theta) &= \binom{m}{x} \theta^x (1 - \theta)^{m-x} I_A(x) \quad \text{con } A = \{0, 1, 2, \dots, m\} \\ &= \binom{m}{x} \left( \frac{\theta}{1 - \theta} \right)^x (1 - \theta)^m I_A(x), \end{aligned}$$

usando la identidad  $a^x = e^{x \ln(a)}$  se tiene que:

$$p(x; \theta) = \exp \left\{ \ln \left\{ \binom{m}{x} \right\} + x \ln \left( \frac{\theta}{1 - \theta} \right) + m \ln(1 - \theta) \right\} I_A(x),$$

entonces  $A = \{0, 1, 2, \dots, m\}$  y las funciones son:

$$\begin{aligned} c(\theta) &= \ln \left( \frac{\theta}{1 - \theta} \right) & d(\theta) &= m \ln(1 - \theta) \\ T(x) &= x & S(x) &= \ln \left\{ \binom{m}{x} \right\}. \end{aligned} \quad \square$$

**Nota 2.7.** La familia de distribuciones obtenida por muestreo de una familia exponencial de un parámetro es en sí misma una familia exponencial de un parámetro.

Tabla 2.1. Funciones asociadas a algunas distribuciones miembros de la familia exponencial (con  $T(x) = x$ )

Distribuc.	$p(x; \theta)$	$c(\theta)$	$d(\theta)$	$S(x)$
$Bin(m, \theta)$	$\binom{m}{x} \theta^x (1 - \theta)^{m-x}$	$\ln\left(\frac{\theta}{1-\theta}\right)$	$\ln[(1 - \theta)^m]$	$\ln\binom{m}{x}$
$\mathcal{P}(\theta)$	$\frac{\theta^x}{x!} e^{-\theta}$	$\ln(\theta)$	$-\theta$	$-\ln(x!)$
$BN(r, \theta)$	$\binom{r+x-1}{x} \theta^r (1 - \theta)^x$	$\ln(1 - \theta)$	$r \ln(\theta)$	$\ln\binom{r+x-1}{x}$
$N(\theta, \sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$	$\frac{\theta}{\sigma^2}$	$-\frac{1}{2} \frac{\theta^2}{\sigma^2}$	$-\frac{\left[\frac{x^2}{\sigma^2} + \ln\left(\frac{2\pi}{\sigma^2}\right)\right]}{2}$
$G(r, \theta)$	$\frac{\theta}{\Gamma(r)} (\theta x)^{r-1} e^{-\theta x}$	$-\theta$	$r \ln(\theta)$	$\ln\left(\frac{x^{r-1}}{\Gamma(r)}\right)$
$NI(\theta, \lambda)$	$\left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} e^{-\frac{\lambda(x-\theta)^2}{2\theta^2 x}}$	$-\frac{\lambda}{2\theta^2}$	$\frac{\lambda}{\theta}$	$\frac{\ln\left(\frac{\lambda}{2\pi x^3}\right) - \frac{\lambda}{x}}{2}$

**Nota 2.8.** De acuerdo con [15, p. 179] una *fmp* o *fdp* es de tipo exponencial si

$$p(x; \theta) = \frac{a(x)}{v(\theta)} \exp\{c(\theta)T(x)\} \quad \text{con } T(x) = x, \quad (2.41)$$

donde  $a(x) \geq 0$ ,  $v(\theta) = \sum a(x) \exp\{c(\theta)x\}$  o  $v(\theta) = \int a(x) \exp\{c(\theta)x\} dx$  según sea el caso. Comparando las expresiones (2.40) y (2.41) se tiene que

$$a(x) = \exp\{S(x)\} \quad \text{y} \quad v(\theta) = \exp\{-d(\theta)\}.$$

**Nota 2.9.** Cuando el soporte de la *fdp* depende de  $\theta$ , por ejemplo, si  $X_k \sim U(0, \theta)$ , la familia de distribuciones no son de la familia exponencial.

### 2.10.2. Familia exponencial de un parámetro en forma natural

Esta clase se obtiene reparametrizando (2.40) al usar  $\eta = c(\theta)$ . Entonces,

$$\begin{aligned} p(x; \theta) &= \left\{ \exp \left[ \underbrace{c(\theta)}_{\eta} T(x) + d(\theta) + S(x) \right] \right\} I_A(x) \\ &= \left\{ \exp [\eta T(x) + d_0(\eta) + S(x)] \right\} I_A(x) \\ &= p(x; \eta). \end{aligned} \quad (2.42)$$

Note que si la función  $c(\theta)$  es una función uno a uno,  $c^{-1}(\eta) = \theta$  y

$$d(\theta) = d(c^{-1}(\eta)) = d_0(\eta).$$

**Ejemplo 2.17.** Sea  $X$  una variable aleatoria con distribución  $\text{Bin}(n, \theta)$ ,  $n \in \mathbb{N}$ ,  $0 < \theta < 1$ . Demuestre que su *fmp* se puede expresar en forma natural.

*Solución.* Considerando la información del ejemplo 2.16,

$$\begin{aligned} p(x; \theta) &= \binom{n}{x} \theta^x (1 - \theta)^{n-x} I_A(x) \quad \text{con } A = \{0, 1, 2, \dots, n\} \\ &= \exp \left\{ \underbrace{\ln \left( \frac{\theta}{1 - \theta} \right)}_{\eta = \eta(\theta)} \underbrace{x}_{T(x)} + \underbrace{n \ln(1 - \theta)}_{d_0(\eta(\theta))} + \underbrace{\ln \left\{ \binom{n}{x} \right\}}_{S(x)} \right\} I_A(x), \quad (2.43) \end{aligned}$$

entonces  $A = \{0, 1, 2, \dots, n\}$  y las funciones son:

$$\begin{aligned} \eta(\theta) &= \ln \left( \frac{\theta}{1 - \theta} \right) \Rightarrow e^\eta = \frac{\theta}{1 - \theta} \Rightarrow \theta(\eta) = \frac{e^\eta}{1 + e^\eta} \\ d_0(\eta) &= d_0(\eta(\theta)) = n \ln(1 - \theta) = n \ln \left( 1 - \frac{e^\eta}{1 + e^\eta} \right) = n \ln \left( \frac{1}{1 + e^\eta} \right) \\ &= -n \ln(1 + e^\eta) \\ T(x) &= x \qquad S(x) = \ln \left\{ \binom{n}{x} \right\}. \quad \square \end{aligned}$$

## 2.11. Probabilidad y estadística multivariada

El énfasis principal de esta sección es la extensión de la noción básica de datos multidimensionales.

**DEFINICIÓN 2.5. Vector aleatorio  $n$ -dimensional.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias reales definidas sobre el mismo espacio de probabilidad  $(\Omega, \mathcal{F}, P)$ . La  $\mathcal{F}$ - $\mathfrak{B}^n$  variable aleatoria  $X : \Omega \rightarrow \mathbb{R}^n$  definida por:

$$X(\omega) := (X_1(\omega), X_2(\omega), \dots, X_n(\omega)),$$

se llama vector aleatorio  $n$ -dimensional.

**DEFINICIÓN 2.6. Función de probabilidad conjunta.** Sea un vector aleatorio  $X = (X_1, X_2, \dots, X_n)$   $n$ -dimensional. Si las variables aleatorias  $X_k$  con  $k = 1, 2, \dots, n$  son todas discretas, se dice que el vector aleatorio es discreto. En tal caso,

se define la *fmp* de  $\mathbf{X}$  o *fmp* conjunta de las variables aleatorias  $X_1, X_2, \dots, X_n$  como:

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} P(\mathbf{X} = \mathbf{x}) & \text{si } \mathbf{x} \text{ pertenece al rango de } \mathbf{X} \\ 0 & \text{en otro caso.} \end{cases}$$

### 2.11.1. Función de distribución acumulada conjunta

**DEFINICIÓN 2.7. Función de distribución acumulada conjunta.** Sea  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  un vector aleatorio  $n$ -dimensional. La función

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n),$$

para todo  $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ , se llama *fda* conjunta de las variables aleatorias  $X_1, X_2, \dots, X_n$  o *fda* del vector aleatorio  $n$ -dimensional  $\mathbf{X}$ .

**Teorema 2.8.** Según [16, p. 183], si  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  es un vector aleatorio  $n$ -dimensional discreto. Entonces, para todo  $j = 1, 2, \dots, n$  se satisface:

$$P[\underbrace{X_j = x}_{\text{sin } x_j}] = \sum_{x_1} \sum_{x_2} \dots \sum_{x_n} P(X_1 = x_1, X_2 = x_2, \dots, X_{j-1} = x_{j-1}, \underbrace{X_j = x, X_{j+1} = x_{j+1}, \dots, X_n = x_n}_{\text{sin } x_j})$$

y se llama *distribución marginal* de la variable aleatoria  $X_j$  a la función

$$f_{X_j}(x) = \begin{cases} P(X_j = x) & \text{si } x \text{ pertenece al rango de } X_j \\ 0 & \text{en otro caso.} \end{cases}$$

*Demostración.* Ver detalles de la prueba en [16, p. 183]. □

**DEFINICIÓN 2.8.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias reales definidas sobre el mismo espacio de probabilidad. Si existe una función  $f$  no negativa e integrable tal que [16, p. 183]:

$$P[(X_1, X_2, \dots, X_n) \in C] = \int_C \dots \int f_{\mathbf{X}}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n,$$

con  $C \in \mathfrak{B}^n$ , entonces son variables conjuntamente continuas.

**Teorema 2.9.** Según [16, p. 192], si  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  es un vector aleatorio  $n$ -dimensional. Entonces, para todo  $j = 1, 2, \dots, n$  se satisface:

$$f_{X_j}(x) = \int_{x_n} \dots \int_{x_1} f(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n) dx_1 \dots dx_{j-1} dx_{j+1}, \dots, dx_n$$

*sin  $x_j$*

esta función  $f_{X_j}(x)$  se llama *fdp* marginal de la variable aleatoria  $X_j$ .

*Demostración.* Ver detalles de la prueba en [16, p. 192]. ✓

**Nota 2.10.** La *fmp* condicional  $p(\cdot | \mathbf{y})$  de  $X$  dado  $Y = \mathbf{y}$  es definida como:

$$p(\mathbf{x}|\mathbf{y}) = P(\mathbf{X} = \mathbf{x} | Y = \mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y})}{p_{\mathbf{y}}(\mathbf{y})}, \quad (2.44)$$

donde  $p$  y  $p_{\mathbf{y}}$  son las *fmp* conjunta de  $(\mathbf{X}, Y)$  y  $Y$ , respectivamente (ver [12, p. 477]).

**DEFINICIÓN 2.9.** (Distribución multinomial). De acuerdo con [20, p. 180], si  $m$  y  $n$  son enteros positivos y  $\mathbf{p} = (p_1, p_2, \dots, p_n)$  es el vector de probabilidades cuyos elementos son tales que  $0 \leq p_i \leq 1, i = 1, 2, \dots, n$ , y  $\sum_{i=1}^n p_i = 1$ . Entonces, el vector aleatorio  $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$  tiene una *fmp* multinomial con  $m$  ensayos,  $n$  categorías y vector de probabilidades  $\mathbf{p}$  si la distribución conjunta de  $\mathbf{X}$  es:

$$f(\mathbf{x}; \mathbf{p}) = \frac{m!}{x_1! x_2! x_3! \dots x_n!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \dots p_n^{x_n} = m! \prod_{i=1}^n \frac{p_i^{x_i}}{x_i!}, \quad (2.45)$$

donde el conjunto de  $\mathbf{x} = (x_1 \ x_2 \ x_3 \ \dots \ x_n)^\top$  es tal que  $x_i \in \mathbb{Z}^+$  y  $\sum_{i=1}^n x_i = m$ .

## 2.12. Valor esperado de un vector aleatorio y matriz de covarianzas

**DEFINICIÓN 2.10.** Sea  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  un vector aleatorio. La esperanza (o valor esperado) de  $\mathbf{X}$ , se define como:

$$\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{X}] = (\mathbb{E}(X_1), \mathbb{E}(X_2), \dots, \mathbb{E}(X_n)) \quad (2.46)$$

**Teorema 2.10. Propiedades del valor esperado.** Sean  $\mathbf{a}$  un vector de constantes,  $\mathbf{A}$  una matriz de constantes y  $\mathbf{X}$  un vector aleatorio, entonces la función  $\mathbb{E}(\cdot)$  tiene las siguientes propiedades:

1.  $\mathbb{E}[\mathbf{a}^\top \mathbf{X}] = \mathbf{a}^\top \mathbb{E}[\mathbf{X}]$ .
2.  $\mathbb{E}[\mathbf{A}\mathbf{X}] = \mathbf{A}\mathbb{E}[\mathbf{X}]$ .
3.  $\mathbb{E}(\mathbf{X} \pm \mathbf{Y}) = \mathbb{E}(\mathbf{X}) \pm \mathbb{E}(\mathbf{Y})$ .

*Demostración.* Solo se presenta la prueba de la propiedad 1:

$$\mathbb{E}(\mathbf{a}^\top \mathbf{X}) = \mathbb{E}\left(\sum_{j=1}^n a_j X_j\right) = \sum_{j=1}^n a_j \mathbb{E}(X_j) = \sum_{j=1}^n a_j \mu_j = \mathbf{a}^\top \boldsymbol{\mu}_X. \quad (2.47)$$

Las demás quedan como ejercicio.  $\square$

**DEFINICIÓN 2.11.** Si las entradas del vector-columna  $\mathbf{X} = [X_1 \ \dots \ X_n]^\top$  son variables aleatorias, cada una con varianza finita, entonces la matriz de covarianzas  $\boldsymbol{\Sigma}$  es la matriz cuya entrada  $(i, j)$  es la covarianza entre  $X_i$  y  $X_j$  (ver (1.44)):

$$\sigma_{ij} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = \begin{cases} \mathbb{E}(X_i^2) - \mu_i^2 = \mathbb{V}(X_i) < \infty & i = j \\ \mathbb{E}(X_i X_j) - \mu_i \mu_j^{13} & i \neq j, \end{cases}$$

donde  $\mu_i = \mathbb{E}(X_i)$  es el valor esperado de la  $i$ -ésima entrada del vector  $\mathbf{X}$ . En otras palabras, tenemos:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbb{E}[(X_1 - \mu_1)(X_1 - \mu_1)] & \mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)] & \cdots & \mathbb{E}[(X_1 - \mu_1)(X_n - \mu_n)] \\ \mathbb{E}[(X_2 - \mu_2)(X_1 - \mu_1)] & \mathbb{E}[(X_2 - \mu_2)(X_2 - \mu_2)] & \cdots & \mathbb{E}[(X_2 - \mu_2)(X_n - \mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[(X_n - \mu_n)(X_1 - \mu_1)] & \mathbb{E}[(X_n - \mu_n)(X_2 - \mu_2)] & \cdots & \mathbb{E}[(X_n - \mu_n)(X_n - \mu_n)] \end{bmatrix}. \quad (2.48)$$

La anterior definición es equivalente a la igualdad matricial:

$$\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top]. \quad (2.49)$$

**Teorema 2.11. Propiedades de la matriz de varianza covarianza.** Sea  $\mathbb{C}_{\text{ov}}(\mathbf{X}, \mathbf{Y}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])^\top]$  y  $\boldsymbol{\mu}_X = \mathbb{E}(\mathbf{X})$ , las siguientes propiedades fundamentales se satisfacen:

1.  $\boldsymbol{\Sigma}_X = \mathbb{C}_{\text{ov}}(\mathbf{X}, \mathbf{X}) = \mathbb{E}(\mathbf{X}\mathbf{X}^\top) - \boldsymbol{\mu}_X \boldsymbol{\mu}_X^\top$ .
2.  $\mathbb{V}[\mathbf{a}^\top \mathbf{X}] = \mathbf{a}^\top \mathbb{V}[\mathbf{X}] \mathbf{a}$ .
3.  $\boldsymbol{\Sigma}_X$  es una matriz semidefinida positiva.
4.  $\mathbb{V}(\mathbf{A}\mathbf{X} + \mathbf{a}) = \mathbf{A} \mathbb{V}(\mathbf{X}) \mathbf{A}^\top$ .
5.  $\mathbb{C}_{\text{ov}}(\mathbf{X}, \mathbf{Y}) = \mathbb{C}_{\text{ov}}(\mathbf{Y}, \mathbf{X})^\top$ .
6.  $\mathbb{C}_{\text{ov}}(\mathbf{X}_1 + \mathbf{X}_2, \mathbf{Y}) = \mathbb{C}_{\text{ov}}(\mathbf{X}_1, \mathbf{Y}) + \mathbb{C}_{\text{ov}}(\mathbf{X}_2, \mathbf{Y})$ .

<sup>13</sup>Por la desigualdad de Cauchy-Schwarz,  $\sigma_{ii} < \infty$  implica que  $\sigma_{ij} < \infty$ .

7. Si los vectores  $\mathbf{X}$  y  $\mathbf{Y}$  son de igual tamaño, entonces:

$$\mathbb{V}(\mathbf{X} + \mathbf{Y}) = \mathbb{V}(\mathbf{X}) + \mathbb{Cov}(\mathbf{X}, \mathbf{Y}) + \mathbb{Cov}(\mathbf{Y}, \mathbf{X}) + \mathbb{V}(\mathbf{Y}).$$

8.  $\mathbb{Cov}(\mathbf{A}\mathbf{X}, \mathbf{B}\mathbf{Y}) = \mathbf{A} \mathbb{Cov}(\mathbf{X}, \mathbf{Y}) \mathbf{B}^\top$ .

9. Si  $\mathbf{X}$  y  $\mathbf{Y}$  son independientes, entonces  $\mathbb{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{O}$ ,

donde  $\mathbf{X}$ ,  $\mathbf{X}_1$  y  $\mathbf{X}_2$  son vectores aleatorios de tamaño  $(n \times 1)$ ,  $\mathbf{Y}$  es un vector aleatorio  $(m \times 1)$ ,  $\mathbf{a}$  es  $(n \times 1)$ ,  $\mathbf{A}^\top$  y  $\mathbf{B}$  son matrices de  $(n \times m)$ .

*Demostración.* Solo se prueba la propiedad 2 y las demás quedan como ejercicio. Sea  $\mathbf{a} \in \mathbb{R}^n$  un vector columna arbitrario, considérese la variable aleatoria  $Y$  definida como:

$$Y = \mathbf{a}^\top \mathbf{X}.$$

Por la propiedad 1 del teorema 2.10

$$\mathbb{E}[Y] = \mathbb{E}[\mathbf{a}^\top \mathbf{X}] = \mathbf{a}^\top \mathbb{E}[\mathbf{X}].$$

Usando la expresión (2.49) se tiene que:

$$\begin{aligned} \mathbb{V}(Y) &= \mathbb{V}(\mathbf{a}^\top \mathbf{X}) = \mathbb{E}\{[\mathbf{a}^\top \mathbf{X} - \mathbb{E}(\mathbf{a}^\top \mathbf{X})][\mathbf{a}^\top \mathbf{X} - \mathbb{E}(\mathbf{a}^\top \mathbf{X})]^\top\} \\ &= \mathbb{E}\{\mathbf{a}^\top [\mathbf{X} - \mathbb{E}(\mathbf{X})][\mathbf{X} - \mathbb{E}(\mathbf{X})]^\top \mathbf{a}\} = \mathbf{a}^\top \boldsymbol{\Sigma}_X \mathbf{a}, \end{aligned} \quad (2.50)$$

donde  $\boldsymbol{\Sigma}_X = \mathbb{V}(\mathbf{X})$ . Nótese que la expresión (2.50) es un escalar y dado que  $\mathbb{V}(Y) \geq 0$ , entonces  $\boldsymbol{\Sigma}_X$  es una matriz semidefinida positiva.  $\checkmark$

## 2.13. Funciones generadoras multidimensionales

Al igual que en el caso univariado, las fgm y característica conjuntas, especifican la fdp conjunta y, además, a partir de estas se obtienen los momentos conjuntos de órdenes  $(k_1, k_2, \dots, k_n)$ , es decir, aquellos de la forma  $\mathbb{E}[X_1^{k_1} X_2^{k_2} \dots X_n^{k_n}]$ .

### 2.13.1. Función generadora de momentos multidimensional

**DEFINICIÓN 2.12.** (fgm conjunta o multidimensional). Según [16, p. 231] y [59, p. 270], dado un vector aleatorio  $n$  dimensional  $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$ . Si existe  $\delta > 0$  tal que  $\mathbb{E}(\exp\{\mathbf{t}^\top \mathbf{X}\})$  es finito para todo  $\mathbf{t} = (t_1, t_2, \dots, t_n)^\top \in \mathbb{R}^n$

con  $\|\mathbf{t}\| := \sqrt{t_1^2 + t_2^2 + \dots + t_n^2} < \delta$ , entonces, la **fgm conjunta** de  $\mathbf{X}$ , denotada por  $m_{\mathbf{X}}(\mathbf{t})$ , se define como sigue:

$$m_{\mathbf{X}}(\mathbf{t}) := \mathbb{E}(e^{\mathbf{t}^\top \mathbf{X}}) \quad \text{para } \|\mathbf{t}\| < \delta, \quad (2.51)$$

donde  $\delta$  es conocido como el radio de convergencia de  $m_{\mathbf{X}}(\mathbf{t})$ .

**Teorema 2.12. Propiedades de la fgm multidimensional.** Sea  $\mathbf{X}$  un vector aleatorio  $n$  dimensional con fgm  $m_{\mathbf{X}}(\mathbf{t})$ , entonces [83, p. 110]

1.  $m_{\mathbf{X}}(\mathbf{0}) = 1$ .
2. Sean  $\mathbf{A}$  una matriz de tamaño  $m \times n$  y  $\mathbf{b} \in \mathbb{R}^m$  arbitrarios,

$$m_{\mathbf{A}\mathbf{X}+\mathbf{b}}(\mathbf{t}) = e^{\mathbf{t}^\top \mathbf{b}} m_{\mathbf{X}}(\mathbf{A}^\top \mathbf{t}), \quad \forall \mathbf{t} \in \mathbb{R}^m.$$

3. Si  $\mathbf{X}$  y  $\mathbf{Y}$  son vectores aleatorios independientes, entonces:

$$m_{\mathbf{X}+\mathbf{Y}}(\mathbf{t}) = m_{\mathbf{X}}(\mathbf{t})m_{\mathbf{Y}}(\mathbf{t}) \quad \text{para todo } \mathbf{t} \in \mathbb{R}^n.$$

4. Si  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$  son vectores aleatorios independientes, todos de dimensión  $n \times 1$ , entonces:

$$m_{\sum_{k=1}^m \mathbf{X}_k}(\mathbf{t}) = \prod_{k=1}^m m_{\mathbf{X}_k}(\mathbf{t}), \quad \forall \mathbf{t} \in \mathbb{R}^n.$$

5. El vector de medias de  $\mathbf{X}$  se establece como el gradiente (ver nota B.10) de  $m_{\mathbf{X}}(\mathbf{t})$  evaluado en  $\mathbf{t} = \mathbf{0}$ , es decir,

$$\mathbb{E}(\mathbf{X}) = \nabla m_{\mathbf{X}}(\mathbf{0}) = \nabla m_{\mathbf{X}}(\mathbf{t})|_{\mathbf{t}=\mathbf{0}}. \quad (2.52)$$

6. Los momentos conjuntos de órdenes  $(k_1, k_2, \dots, k_n)$  se establecen como:

$$\mathbb{E}\left[X_1^{k_1} X_2^{k_2} \dots X_n^{k_n}\right] = \frac{\partial^k}{\partial t_1^{k_1} \partial t_2^{k_2} \dots \partial t_n^{k_n}} m_{\mathbf{X}}(\mathbf{t}) \Bigg|_{\mathbf{t}=\mathbf{0}} \quad \text{con } k = \sum_{j=1}^n k_j. \quad (2.53)$$

*Demostración.* Nótese que estas propiedades son las generalizaciones de las propiedades dadas en el teorema 1.12, solo se presenta la prueba del ítem 2.

2. Sea  $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ , luego,

$$\mathbf{t}^\top \mathbf{Y} = \mathbf{t}^\top (\mathbf{A}\mathbf{X} + \mathbf{b}) = \underbrace{\mathbf{t}^\top \mathbf{A}}_{\mathbf{c}^\top} \mathbf{X} + \mathbf{t}^\top \mathbf{b} = \mathbf{c}^\top \mathbf{X} + \mathbf{t}^\top \mathbf{b}.$$

La *fgm* de  $Y$  es dada por:

$$m_Y(\mathbf{t}) = \mathbb{E}\left(e^{\mathbf{t}^\top Y}\right) = \mathbb{E}\left(e^{\mathbf{c}^\top X + \mathbf{t}^\top b}\right) = e^{\mathbf{t}^\top b} \mathbb{E}\left(e^{\mathbf{c}^\top X}\right) = e^{\mathbf{t}^\top b} m_X(\mathbf{c}). \quad \square$$

**Ejemplo 2.18.** Supóngase, como en [59, p. 270], que la *fmp* conjunta de  $X_1, X_2$  y  $X_3$  es multinomial con parámetros  $n, p_1, p_2$  y  $p_3$ , esto es,

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3},$$

con  $x_j \in \{0, 1, 2, 3, \dots, n\}$ ,  $x_1 + x_2 + x_3 = n$  y  $p_1 + p_2 + p_3 = 1$ . Encontrar:

1. La *fgm* de  $(X_1, X_2, X_3)$ .
2. El vector de medias.
3. El momento cruzado de orden  $(1, 1, 0)$ , es decir, el valor esperado de  $X_1 X_2$ .

*Solución.* Sean  $Z = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$  y  $\mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \in \mathbb{R}^3$  con  $\|\mathbf{t}\| < \delta$ ,  $\delta > 0$ .

1. De la definición 2.12 para  $\|\mathbf{t}\| < \delta$ , se tiene que:

$$\begin{aligned} \underbrace{\mathbb{E}\left(\exp\{\mathbf{t}^\top Z\}\right)}_{m_Z(\mathbf{t})} &= \mathbb{E}\left(\exp\left\{\begin{pmatrix} t_1 & t_2 & t_3 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}\right\}\right) \\ &= \mathbb{E}\left(e^{t_1 X_1 + t_2 X_2 + t_3 X_3}\right) \\ &= \sum_{x_1} \sum_{x_2} \sum_{x_3} e^{t_1 x_1 + t_2 x_2 + t_3 x_3} \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \\ &\quad x_1 + x_2 + x_3 = n \\ &= \sum_{x_1} \sum_{x_2} \sum_{x_3} \frac{n!}{x_1!x_2!x_3!} (e^{t_1} p_1)^{x_1} (e^{t_2} p_2)^{x_2} (e^{t_3} p_3)^{x_3} \\ &\quad x_1 + x_2 + x_3 = n \\ &= (e^{t_1} p_1 + e^{t_2} p_2 + e^{t_3} p_3)^n. \end{aligned}$$

En la última expresión se empleó el teorema multinomial de Leibniz (ver [20, p. 181]).

2. Por la expresión (2.52), se tiene que:

$$\nabla m_Z(\mathbf{t}) = n(e^{t_1} p_1 + e^{t_2} p_2 + e^{t_3} p_3)^{n-1} (p_1 e^{t_1}, p_2 e^{t_2}, p_3 e^{t_3}),$$

evaluando esta expresión en  $\mathbf{t} = \mathbf{0}$  se obtiene  $\mathbb{E}(Z)$  como:

$$\mathbb{E}(Z) = \nabla m_Z(\mathbf{0}) = n(p_1, p_2, p_3).$$

3. Por la expresión (2.53), se tiene que:

$$\begin{aligned}\frac{\partial^2}{\partial t_1 \partial t_2} m_Z(\mathbf{t}) &= \frac{\partial}{\partial t_1} \left( n \left[ e^{t_1} p_1 + e^{t_2} p_2 + e^{t_3} p_3 \right]^{n-1} e^{t_2} p_2 \right) \\ &= n(n-1) e^{t_2} p_2 \left[ e^{t_1} p_1 + e^{t_2} p_2 + e^{t_3} p_3 \right]^{n-2} e^{t_1} p_1,\end{aligned}$$

evaluando esta expresión en  $\mathbf{t} = \mathbf{0}$  se obtiene  $\mathbb{E}(X_1 X_2)$  como:

$$\mathbb{E}(X_1 X_2) = \left. \frac{\partial^2}{\partial t_1 \partial t_2} m_Z(\mathbf{t}) \right|_{\mathbf{t}=\mathbf{0}} = n(n-1) p_1 p_2. \quad \checkmark$$

## 2.13.2. Función característica multidimensional

**DEFINICIÓN 2.13.** (Función característica multidimensional). *De acuerdo con [43, p. 241], si  $\mathbf{X} = (X_1, X_2, X_3, \dots, X_n)^\top$  es un vector aleatorio  $n$ -dimensional y  $\mathbf{t} = (t_1, t_2, \dots, t_n)^\top$  un vector cualquiera en  $\mathbb{R}^n$ . Entonces la **función característica** de  $\mathbf{X}$  es una función  $\Psi_{\mathbf{X}}(\mathbf{t}) : \mathbb{R}^n \rightarrow \mathbb{C}$  definida por:*

$$\Psi(\mathbf{t}) = \Psi_{\mathbf{X}}(\mathbf{t}) := \mathbb{E}(e^{i\mathbf{t}^\top \mathbf{X}}) = \mathbb{E}\left(\exp\left\{i \sum_{j=1}^n t_j X_j\right\}\right). \quad (2.54)$$

La función característica multivariada tiene propiedades análogas a las enunciadas en el teorema 1.13, las cuales se presentan a continuación.

**Teorema 2.13.** **Propiedades de la función característica multidimensional.** *Sea  $\mathbf{X}$  un vector aleatorio  $n$  dimensional entonces [83, p. 81]:*

1.  $\Psi_{\mathbf{X}}(\mathbf{0}) = 1$ .
2.  $|\Psi_{\mathbf{X}}(\mathbf{t})| \leq \Psi_{\mathbf{X}}(\mathbf{0}) \quad \forall \mathbf{t} \in \mathbb{R}^n$ .
3.  $\overline{\Psi_{\mathbf{X}}(\mathbf{t})} = \Psi_{\mathbf{X}}(-\mathbf{t})$  para todo  $\mathbf{t} \in \mathbb{R}^n$ .
4. Sean  $\mathbf{A}$  una matriz de tamaño  $m \times n$  y  $\mathbf{b} \in \mathbb{R}^m$  arbitrarios,

$$\Psi_{\mathbf{A}\mathbf{X}+\mathbf{b}}(\mathbf{t}) = e^{i\mathbf{t}^\top \mathbf{b}} \Psi_{\mathbf{X}}(\mathbf{A}^\top \mathbf{t}), \quad \forall \mathbf{t} \in \mathbb{R}^m.$$

5. Si  $\mathbf{X}$  y  $\mathbf{Y}$  son vectores aleatorios independientes, entonces:

$$\Psi_{\mathbf{X}+\mathbf{Y}}(\mathbf{t}) = \Psi_{\mathbf{X}}(\mathbf{t}) \Psi_{\mathbf{Y}}(\mathbf{t}) \quad \text{para todo } \mathbf{t} \in \mathbb{R}^n.$$

6. Si  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$  son vectores aleatorios independientes, todos de dimensión  $n \times 1$ , entonces:

$$\Psi_{\sum_{k=1}^m \mathbf{X}_k}(\mathbf{t}) = \prod_{k=1}^m \Psi_{\mathbf{X}_k}(\mathbf{t}), \quad \forall \mathbf{t} \in \mathbb{R}^n.$$

7. **Fórmula de inversión.** Sean  $F_X(x)$  y  $\Psi_X(t)$  la fda y la función característica del vector aleatorio  $X$ , respectivamente. Si  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  y  $\mathbf{b} = (b_1, b_2, \dots, b_n)$  elementos de  $\mathbb{R}^n$  son puntos de continuidad de la fda con  $\mathbf{a} < \mathbf{b}$ , entonces [18, p. 261]:

$$F_X(\mathbf{b}) - F_X(\mathbf{a}) = \lim_{T \rightarrow \infty} \frac{1}{(2\pi)^n} \int_{-T}^T \cdots \int_{-T}^T \prod_{k=1}^n \psi_k(t_k) \Psi_X(\mathbf{t}) dt_1 \cdots dt_n, \quad (2.55)$$

$$\text{donde } \psi_k(t_k) = \frac{e^{-it_k a_k} - e^{-it_k b_k}}{it_k}.$$

8. El vector de medias de  $X$  se establece como el gradiente de  $\Psi_X(\mathbf{t})$  evaluado en  $\mathbf{t} = \mathbf{0}$ , es decir,

$$i \mathbb{E}(X) = \nabla \Psi_X(\mathbf{0}) = \nabla \Psi_X(\mathbf{t})|_{\mathbf{t}=\mathbf{0}}.$$

*Demostración.* Las pruebas se realizan con las modificaciones adecuadas, como que en vez de trabajar sobre la recta real se trabaja sobre  $\mathbb{R}^n$ . Solo se presenta la prueba del ítem 4.

4. Sea  $Y = AX + b$ , entonces:

$$\mathbf{t}^\top Y = \mathbf{t}^\top (AX + b) = \underbrace{\mathbf{t}^\top A}_{\mathbf{c}^\top} X + \mathbf{t}^\top b = \mathbf{c}^\top X + \mathbf{t}^\top b.$$

Usando la expresión (2.54), la función característica de  $Y$  es dada por:

$$\Psi_Y(\mathbf{t}) = \mathbb{E}\left(e^{i(\mathbf{c}^\top X + \mathbf{t}^\top b)}\right) = e^{i\mathbf{t}^\top b} \mathbb{E}\left(e^{i\mathbf{c}^\top X}\right) = e^{i\mathbf{t}^\top b} \Psi_X(\mathbf{c}). \quad \checkmark$$

**Nota 2.11.** La función característica de cualquier distribución marginal se obtiene haciendo todos los términos “extras” iguales a cero. Por ejemplo, para las variables aleatorias  $X, Y$  y  $Z$ , se tiene que:

$$\Psi_{X,Y,Z}(t_1, t_2, t_3) = \mathbb{E}\left(\exp\{i(t_1 X + t_2 Y + t_3 Z)\}\right).$$

Para encontrar la función característica de  $X$  y  $Y$  basta tomar  $t_3 = 0$ , es decir,

$$\Psi_{X,Y}(t_1, t_2) = \mathbb{E}\left(\exp\{i(t_1 X + t_2 Y + 0 \cdot Z)\}\right) = \mathbb{E}\left(\exp\{i(t_1 X + t_2 Y)\}\right).$$

**Teorema 2.14. Cramér-Wold.** Según [55, p. 34], la distribución de un vector aleatorio  $X = (X_1, X_2, X_3, \dots, X_n)^\top$  es completamente determinada por el conjunto de las distribuciones de todas las combinaciones lineales de la forma

$$\mathbf{t}^\top X = \sum_{j=1}^n t_j X_j, \text{ donde } \mathbf{t} \in \mathbb{R}^n.$$

*Demostración.*

( $\Rightarrow$ ) Para todo  $t \in \mathbb{R}^n$  se tiene que:

$$\psi_X(t) = \mathbb{E}(e^{it^\top X}) = \mathbb{E}\left(\exp\left\{i \sum_{j=1}^n t_j X_j\right\}\right) = \psi_{t^\top X}(1).$$

$$\underbrace{\psi_X(t)}_{\downarrow F_X} = \underbrace{\psi_{t^\top X}(1)}_{\downarrow F_{t^\top X}} = \int_{\mathbb{R}} e^{i w} dF_{t^\top X}(w) \quad \text{con } w = \sum_{j=1}^n t_j X_j \quad \forall t \in \mathbb{R}^n.$$

Note que las **fdp** marginales determinan de manera única a la **fdp** conjunta.

( $\Leftarrow$ ) Para  $t \in \mathbb{R}^n$ , fijo,

$$\psi_X(ut) = \mathbb{E}\left(\exp\left\{iu \sum_{j=1}^n t_j X_j\right\}\right) = \mathbb{E}(e^{iut^\top X}) = \psi_{t^\top X}(u), \quad \forall u \in \mathbb{R},$$

entonces,

$$\underbrace{\psi_X(ut)}_{\downarrow F_X} = \underbrace{\psi_{t^\top X}(u)}_{\downarrow F_{t^\top X}} = \int_{\mathbb{R}^n} e^{iut^\top X} dF_X(x).$$

Note que la **fdp** conjunta determina las **fdp** marginales. \(\square\)

## 2.14. Distribución de suma de variables aleatorias independientes

En esta sección se generaliza la metodología dada en el apartado 2.7, esto permite encontrar la **fdp** de la variable suma,  $S = X_1 + X_2 + \dots + X_n$ .

**Teorema 2.15.** Según [65, p. 193], si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes Poisson con parámetros  $\lambda_1, \lambda_2, \dots, \lambda_n$ , entonces la suma tiene distribución Poisson con parámetro  $\lambda_1 + \lambda_2 + \dots + \lambda_n$ .

*Demostración.* Puesto que las variables aleatorias  $X_1, X_2, \dots, X_n$  son independientes, de (1.55) se tiene que:

$$m_S(t) = \prod_{i=1}^n m_{X_i}(t) = \prod_{i=1}^n e^{\lambda_i(e^t - 1)} = \exp\left\{\sum_{i=1}^n \lambda_i (e^t - 1)\right\},$$

la cual coincide con la **fgm** de una variable aleatoria Poisson con parámetro  $\sum_{i=1}^n \lambda_i$ . ✓

**Teorema 2.16.** Si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes e idénticamente distribuidas geométricas con parámetro  $q$ , entonces la suma es una variable aleatoria binomial negativa con parámetros  $p$  y  $r = n$ .

*Demostración.* Como las variables aleatorias  $X_1, X_2, \dots, X_n$  son independientes e idénticamente distribuidas (**iid**) geométricas, de (1.56) se tiene que:

$$m_S(t) = \prod_{i=1}^n m_{X_i}(t) = \left( \frac{p}{1 - qe^t} \right)^n,$$

la cual coincide con la **fgm** de una variable aleatoria binomial negativa con parámetros  $p$  y  $r = n$ . ✓

**Teorema 2.17.** Según [65, p. 193], si  $X_1, X_2, \dots, X_n$  son variables aleatorias **iid** exponencialmente con parámetro  $\beta$ , entonces la suma es una variable aleatoria gamma con parámetros  $\alpha = n$  y  $\beta$ .

*Demostración.* Puesto que las variables aleatorias  $X_1, X_2, \dots, X_n$  son **iid** exponencialmente, de (1.56) se tiene que:

$$m_S(t) = \prod_{i=1}^n m_{X_i}(t) = \left( \frac{\beta}{\beta - t} \right)^n,$$

la cual coincide con la **fgm** de una variable aleatoria gamma con parámetros  $\alpha = n$  y  $\beta$ . ✓

**Teorema 2.18.** Si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes normales con medias respectivas  $\mu_1, \mu_2, \dots, \mu_n$  y varianzas respectivas  $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ , entonces la suma es una variable aleatoria normal con media  $\mu_1 + \mu_2 + \dots + \mu_n$  y varianza  $\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2$ .

*Demostración.* Como las variables aleatorias  $X_1, X_2, \dots, X_n$  son independientes, de (1.55) se tiene que:

$$\begin{aligned} m_S(t) &= \prod_{i=1}^n m_{X_i}(t) = \prod_{i=1}^n e^{\mu_i t + \frac{1}{2} \sigma_i^2 t^2} \\ &= \exp \left\{ \sum_{i=1}^n \left( \mu_i t + \frac{1}{2} \sigma_i^2 t^2 \right) \right\} = \exp \left\{ t \sum_{i=1}^n \mu_i + \frac{1}{2} t^2 \sum_{i=1}^n \sigma_i^2 \right\}, \end{aligned}$$

la cual coincide con la **fgm** de una variable aleatoria normal con parámetros  $\mu = \sum_{i=1}^n \mu_i$  y  $\sigma^2 = \sum_{i=1}^n \sigma_i^2$ .  $\checkmark$

## 2.15. Transformaciones de dos o más variables aleatorias

En esta sección se analiza cómo determinar las **fdp** de variables (o vectores) aleatorios que son función de otras variables aleatorias, el siguiente resultado es una generalización del teorema 2.7.

**Teorema 2.19. Teorema del cambio de variable.** Sea  $X$  un vector aleatorio  $n$  dimensional con **fdp** conjunta  $f_X$  y sea  $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  una aplicación inyectiva. Supóngase que tanto  $\mathbf{g}$  y  $\mathbf{g}^{-1} : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$  son continuas. Si las derivadas parciales de  $\mathbf{g}^{-1}$  existen y son continuas y si su jacobiano,  $|\mathcal{J}|$ , es distinto de cero, entonces la **fdp** conjunta de  $\mathbf{Y} = \mathbf{g}(X)$  es:

$$f_Y(\mathbf{y}) = \begin{cases} f_X[\mathbf{g}^{-1}(\mathbf{y})] |\mathcal{J}_{\mathbf{g}^{-1}}(\mathbf{y})| & \text{si } \mathbf{y} \text{ está en el rango de } \mathbf{g} \\ 0 & \text{en otro caso,} \end{cases} \quad (2.56)$$

donde  $\mathcal{J}_{\mathbf{g}^{-1}}(\mathbf{y})$  es la matriz jacobiana definida por

$$\mathcal{J}_{\mathbf{g}^{-1}}(\mathbf{y}) = \frac{\partial}{\partial \mathbf{y}} \mathbf{g}^{-1}(\mathbf{y}) = \begin{bmatrix} \nabla h_1(\mathbf{y}) \\ \vdots \\ \nabla h_n(\mathbf{y}) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial y_1} h_1(\mathbf{y}) & \cdots & \frac{\partial}{\partial y_n} h_1(\mathbf{y}) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial y_1} h_n(\mathbf{y}) & \cdots & \frac{\partial}{\partial y_n} h_n(\mathbf{y}) \end{bmatrix}, \quad (2.57)$$

y  $\nabla h_k(\mathbf{y})$  es el vector gradiente de  $h_k(\mathbf{y}) = g_k^{-1}(\mathbf{y})$  (ver nota B.10).

*Demostración.* Ver detalles de la prueba en [20, p. 185].  $\checkmark$

**Ejemplo 2.19.** Supóngase, como en [16, p. 216], que  $X$  y  $Y$  son dos variables aleatorias independientes con **fdp** conjunta  $f$ . Sean  $U$  y  $V$  dos variables aleatorias dadas por  $U = X + Y$  y  $V = X - Y$ . Determine una **fdp** de  $U$  y una **fdp** de  $V$ .

*Solución.* En este caso,  $\mathbf{g} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  y es dada por:

$$\mathbf{g}(x, y) = (g_1(x, y), g_2(x, y)) = \left( \underbrace{x + y}_u, \underbrace{x - y}_v \right),$$

como es una transformación lineal, matricialmente queda:

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} x \\ y \end{pmatrix} = -\frac{1}{2} \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{2} \begin{pmatrix} u+v \\ u-v \end{pmatrix};$$

la transformación inversa está dada por:

$$\mathbf{g}^{-1}(x,y) = (h_1(x,y), h_2(x,y)) = \left( \frac{x+y}{2}, \frac{x-y}{2} \right),$$

y la matriz jacobiana dada en (2.57) de la transformación inversa queda:

$$\mathcal{J}_h(x,y) = \begin{bmatrix} \frac{\partial h_1(x,y)}{\partial x} & \frac{\partial h_1(x,y)}{\partial y} \\ \frac{\partial h_2(x,y)}{\partial x} & \frac{\partial h_2(x,y)}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

El jacobiano es el determinante de la matriz  $\mathcal{J}_h$ . Luego,

$$|\mathcal{J}_h(x,y)| = |\det(\mathcal{J}_h(x,y))| = \left| -\frac{1}{2} \right| = \frac{1}{2}.$$

Por lo tanto, la **fdp** conjunta de  $U$  y  $V$  está dada por (2.56); es decir,

$$f_{(U,V)}(x,y) = \frac{1}{2} f_{(X,Y)} \left( \frac{x+y}{2}, \frac{x-y}{2} \right). \quad (2.58)$$

1. Luego, la **fdp** marginal de la variable  $U$  es:

$$f_U(x) = \frac{1}{2} \int_{-\infty}^{\infty} f_{(X,Y)} \left( \frac{x+y}{2}, \frac{x-y}{2} \right) dy,$$

haciendo el siguiente cambio de variable:

$$w = \frac{x+y}{2} \quad \quad \quad dw = \frac{1}{2} dy,$$

entonces,

$$f_U(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(w, x-w) dw.$$

2. Luego, la **fdp** marginal de la variable  $V$  es:

$$f_V(y) = \frac{1}{2} \int_{-\infty}^{\infty} f_{(X,Y)} \left( \frac{x+y}{2}, \frac{x-y}{2} \right) dx,$$

haciendo el siguiente cambio de variable:

$$z = \frac{x-y}{2} \quad \quad \quad dz = \frac{1}{2} dx,$$

entonces,

$$f_V(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(z+y, z) dz. \quad \checkmark$$

**Ejemplo 2.20.** Sean  $X$  y  $Y$  dos variables aleatorias independientes con distribución exponencial de parámetro  $\lambda$ , esto es,  $X \sim \text{Exp}(\lambda)$  y  $Y \sim \text{Exp}(\lambda)$ . Sean  $U$  y  $V$  dos variables aleatorias dadas por  $U = X + Y$  y  $V = X - Y$ . Encuentre la **fdp** de  $U$ , la **fdp** de  $V$  y la **fgm** conjunta de  $U$  y  $V$ .

*Solución.* Como  $X$  y  $Y$  son variables aleatorias independientes, entonces la **fdp** conjunta es:

$$f_{(X,Y)}(x,y) = f_X(x) \cdot f_Y(y) = \begin{cases} \lambda^2 \exp\{-\lambda \cdot (x+y)\} & \text{si } x > 0, y > 0 \\ 0 & \text{en otra parte,} \end{cases}$$

como la **fdp** conjunta de  $U$  y  $V$  está dada por (2.58), al reemplazar:

$$f_{(U,V)}(x,y) = \begin{cases} \frac{1}{2} \lambda^2 \exp\left\{-\lambda \left(\frac{x+y}{2} + \frac{x-y}{2}\right)\right\} = \frac{1}{2} \lambda^2 e^{-\lambda x} & \text{si } |y| < x \\ 0 & \text{en otra parte} \end{cases}$$

1. Puesto que  $|y| < x$ , la **fdp** marginal de la variable  $U$  es:

$$f_U(x) = \frac{1}{2} \lambda^2 \int_{-x}^x e^{-\lambda x} dy = \lambda^2 x e^{-\lambda x} \quad x > 0,$$

es decir,  $U \sim \text{gamma}(2, \lambda)$ .

2. Si  $y \geq 0$ , entonces  $y < x < \infty$ , luego la **fdp** marginal de  $V$  es:

$$f_V(y) = \frac{1}{2} \lambda^2 \int_y^{\infty} e^{-\lambda x} dx = -\frac{1}{2} \lambda e^{-\lambda x} \Big|_y^{\infty} = \frac{1}{2} \lambda e^{-\lambda y} \quad y \geq 0.$$

Si  $y < 0$ , entonces  $-y < x < \infty$ , luego la **fdp** marginal de  $V$  es:

$$f_V(y) = \frac{1}{2} \lambda^2 \int_{-y}^{\infty} e^{-\lambda x} dx = -\frac{1}{2} \lambda e^{-\lambda x} \Big|_{-y}^{\infty} = \frac{1}{2} \lambda e^{\lambda y} \quad y < 0.$$

Por lo tanto,

$$f_V(y) = \begin{cases} \lambda e^{-\lambda y} & \text{si } y \geq 0 \\ \lambda e^{\lambda y} & \text{si } y < 0, \end{cases} \quad (2.59)$$

es decir,  $V \sim \text{Laplace}(0, \frac{1}{\lambda})$ .

3. Sea  $\mathbf{Z} = \begin{pmatrix} U \\ V \end{pmatrix}$  y  $\mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \in \mathbb{R}^2$  con  $\|\mathbf{t}\| = \sqrt{t_1^2 + t_2^2} < \delta$ ,  $\delta > 0$ . De la definición 2.12 se tiene que para  $\|\mathbf{t}\| < \delta$ ,

$$\begin{aligned} \underbrace{\mathbb{E}(\exp\{\mathbf{t}^\top \mathbf{Z}\})}_{m_{\mathbf{Z}}(\mathbf{t})} &= \mathbb{E}\left(\exp\left\{\begin{pmatrix} t_1 & t_2 \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix}\right\}\right) = \mathbb{E}(\exp\{t_1 U + t_2 V\}) \\ &= \mathbb{E}(\exp\{t_1[X + Y] + t_2[X - Y]\}) \\ &= \mathbb{E}(\exp\{X[t_1 + t_2] + Y[t_1 - t_2]\}) \\ &= \underbrace{\mathbb{E}(\exp\{X[t_1 + t_2]\})}_{\text{Indep.}} \underbrace{\mathbb{E}(\exp\{Y[t_1 - t_2]\})}_{m_Y(t_1 - t_2)} \\ &= \left(\frac{1}{1 - \frac{t_1 + t_2}{\lambda}}\right) \left(\frac{1}{1 - \frac{t_1 - t_2}{\lambda}}\right). \end{aligned} \tag{2.60}$$

■ Para obtener la distribución de  $U$  se toma  $\mathbf{t} = (t_1, 0)^\top$  en (2.60),

$$m_{\binom{U}{V}}\left[\binom{t_1}{0}\right] = \left(\frac{1}{1 - \frac{t_1 + 0}{\lambda}}\right) \left(\frac{1}{1 - \frac{t_1 - 0}{\lambda}}\right) = \left(\frac{1}{1 - \frac{t_1}{\lambda}}\right)^2,$$

que corresponde a la fgm de una variable aleatoria con distribución gamma de parámetros  $r = 2$  y  $\lambda$ , esto es,  $U \sim \text{gamma}(2, \lambda)$ .

■ Para obtener la distribución de  $V$  se toma  $\mathbf{t} = (0, t_2)^\top$  en (2.60),

$$\begin{aligned} m_{\binom{U}{V}}\left[\binom{0}{t_2}\right] &= \left(\frac{1}{1 - \frac{0 + t_2}{\lambda}}\right) \left(\frac{1}{1 - \frac{0 - t_2}{\lambda}}\right) = \left(\frac{1}{1 - \frac{t_2}{\lambda}}\right) \left(\frac{1}{1 + \frac{t_2}{\lambda}}\right) \\ &= \frac{1}{1 - \frac{t_2^2}{\lambda^2}} = \frac{e^{0 \times t}}{1 - \frac{t_2^2}{\lambda^2}}, \end{aligned}$$

que corresponde a la fgm de una variable aleatoria con distribución Laplace, ver ejercicio 1.12, de parámetros  $a = 0$  y  $b = \frac{1}{\lambda}$ . □

**Ejemplo 2.21.** De acuerdo con [76, p. 229], sean  $X$  y  $Y$  dos variables aleatorias independientes tales que  $X \sim N(0, 1)$  y  $Y \sim \chi_{(n)}^2$ . Sean  $U$  y  $V$  dos variables aleatorias dadas por  $U = \frac{X}{\sqrt{Y/n}}$  y  $V = Y$ . Encuentre la fdp de  $U$ .

*Solución.* Primero, se calcula la fdp conjunta de  $X$  y  $Y$ , dado que son independientes:

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\} \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}y\right)^{n/2-1} e^{-\frac{1}{2}y} \quad x \in \mathbb{R}, y > 0.$$

Ahora, para calcular la **fdp** conjunta de  $U$  y  $V$  se considera la transformación

$$\mathbf{g}(x,y) = (g_1(x,y), g_2(x,y)) = \left( \underbrace{\frac{x}{\sqrt{y/n}}}_u, \underbrace{y}_v \right),$$

la transformación inversa está dada por:

$$\mathbf{g}^{-1}(x,y) = (h_1(x,y), h_2(x,y)) = \left( x\sqrt{\frac{y}{n}}, y \right),$$

y la matriz jacobiana de la transformación inversa es

$$\mathcal{J}_h(x,y) = \begin{bmatrix} \frac{\partial h_1(x,y)}{\partial x} & \frac{\partial h_1(x,y)}{\partial y} \\ \frac{\partial h_2(x,y)}{\partial x} & \frac{\partial h_2(x,y)}{\partial y} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{y}{n}} & \frac{x}{2}\sqrt{\frac{1}{ny}} \\ 0 & 1 \end{bmatrix}.$$

El jacobiano es el determinante de la matriz  $\mathcal{J}_h$ . Luego,

$$|\mathcal{J}_h(x,y)| = |\det(\mathcal{J}_h(x,y))| = \left| \sqrt{\frac{y}{n}} \right| = \sqrt{\frac{y}{n}}.$$

Por lo tanto, la **fdp** conjunta de  $U$  y  $V$  está dada por:

$$\begin{aligned} f_{(U,V)}(u,v) &= \sqrt{\frac{v}{n}} f_{(X,Y)}\left(u\sqrt{\frac{v}{n}}, v\right) \\ &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{v}{n}} \exp\left\{-\frac{1}{2}\left(u\sqrt{\frac{v}{n}}\right)^2\right\} \frac{v^{n/2-1}}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} e^{-\frac{1}{2}v}. \end{aligned}$$

Luego, la **fdp** marginal de la variable  $U$  es:

$$f_U(u) = \frac{1}{\sqrt{2n\pi}} \left(\frac{1}{2}\right)^{n/2} \int_0^\infty \frac{v^{(n-1)/2}}{\Gamma(n/2)} \exp\left\{-\frac{1}{2}\left(\frac{u^2v}{n} + v\right)\right\} dv,$$

haciendo el siguiente cambio de variable:

$$w = \frac{v}{2} \left( \frac{u^2}{n} + 1 \right) \qquad dw = \frac{1}{2} \left( \frac{u^2}{n} + 1 \right) dv.$$

Al despejar  $v$  y  $dv$ , y sustituir en la integral:

$$f_U(u) = \frac{2}{2^{n/2}\Gamma(n/2)\sqrt{2n\pi}} \int_0^\infty \left[ 2w \left( \frac{u^2}{n} + 1 \right)^{-1} \right]^{\frac{n-1}{2}} \left( \frac{u^2}{n} + 1 \right)^{-1} e^{-w} dw$$

$$\begin{aligned}
 &= \frac{2^{\frac{n+1}{2}}}{2^{\frac{n+1}{2}} \Gamma(n/2) \sqrt{n\pi}} \int_0^\infty \left(\frac{u^2}{n} + 1\right)^{-\frac{n+1}{2}} w^{\frac{n-1}{2}} e^{-w} dw \\
 &= \frac{1}{\Gamma(n/2) \sqrt{n\pi}} \Gamma\left(\frac{n+1}{2}\right) \left(\frac{u^2}{n} + 1\right)^{-\frac{n+1}{2}} \quad u \in \mathbb{R}, \quad (2.61)
 \end{aligned}$$

una variable aleatoria  $U$  con **fdp** dada por (2.61) sigue una **fdp**  $t$ -Student con  $n$  grados de libertad y se denota  $U \sim t_{(n)}$ . Si  $n = 1$  la **fdp** dada en (2.61) coincide con la expresión (1.45), es decir, la **fdp**  $t$ -Student se convierte en la **fdp** estándar de Cauchy. Por otra parte, usando la fórmula dada en (1.65), con  $m = \frac{n}{2}$  y  $\kappa = \frac{1}{2}$ , cuando  $n \rightarrow \infty$  en (2.61) se tiene que<sup>14</sup>

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt{2} e} \left(1 + \frac{1}{n}\right)^{\frac{n}{2}} = \frac{1}{\sqrt{2}} \quad \text{y} \quad \lim_{n \rightarrow \infty} \left(1 + \frac{u^2}{n}\right)^{-\frac{n+1}{2}} = e^{-\frac{1}{2}u^2}. \quad \square$$

Luego, la expresión (2.61) converge a la **fdp** normal estándar (ver figura 2.3).

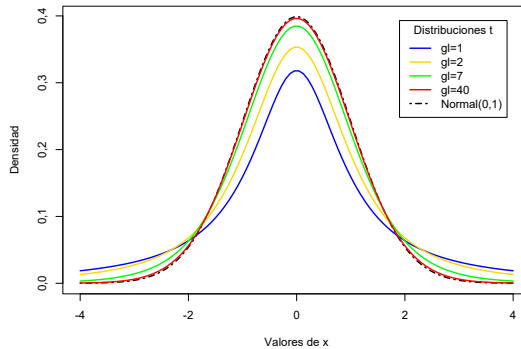


Figura 2.3. **fdp**  $t$ -Student para diferentes grados de libertad

**Ejemplo 2.22.** De acuerdo con [76, p. 231], dadas  $X$  y  $Y$  dos variables aleatorias independientes tales que  $X \sim \chi^2_{(m)}$  y  $Y \sim \chi^2_{(n)}$ . Sean  $U$  y  $V$  dos variables aleatorias dadas por  $U = \frac{X/m}{Y/n}$  y  $V = Y$ . Encuentre la **fdp** de  $U$ .

**Solución.** Primero, se calcula la **fdp** conjunta de  $X$  y  $Y$ , dado que son independientes:

$$f_{(X,Y)}(x,y) = \frac{1}{2} \frac{1}{2} \left(\frac{1}{2}x\right)^{m/2-1} \left(\frac{1}{2}y\right)^{n/2-1} e^{-\frac{1}{2}(x+y)} \quad x > 0, y > 0.$$

<sup>14</sup>De acuerdo con [5, p. 370], para todo número real  $a$ ,  $\lim_{x \rightarrow \infty} \left(1 + \frac{a}{x}\right)^x = e^a$ .

Ahora, para calcular la **fdp** conjunta de  $U$  y  $V$  se considera la transformación

$$\mathbf{g}(x,y) = (g_1(x,y), g_2(x,y)) = \left( \underbrace{\frac{nx}{my}}_u, \underbrace{y}_v \right),$$

la transformación inversa está dada por:

$$\mathbf{g}^{-1}(x,y) = (h_1(x,y), h_2(x,y)) = \left( \frac{m}{n}xy, y \right)$$

y la matriz jacobiana de la transformación inversa es

$$\mathcal{J}_h(x,y) = \begin{bmatrix} \frac{\partial h_1(x,y)}{\partial x} & \frac{\partial h_1(x,y)}{\partial y} \\ \frac{\partial h_2(x,y)}{\partial x} & \frac{\partial h_2(x,y)}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{m}{n}y & \frac{m}{n}x \\ 0 & 1 \end{bmatrix}.$$

El jacobiano es el determinante de la matriz  $\mathcal{J}_h$ . Luego,

$$|\mathcal{J}_h(x,y)| = |\det(\mathcal{J}_h(x,y))| = \left| \frac{m}{n}y \right| = \frac{m}{n}y.$$

Por lo tanto, la **fdp** conjunta de  $U$  y  $V$  está dada por:

$$\begin{aligned} f_{(U,V)}(u,v) &= \frac{m}{n}v f_{(X,Y)}\left(\frac{m}{n}uv, v\right) \\ &= \frac{m \cdot v}{n\Gamma(m/2)\Gamma(n/2)} \left(\frac{1}{2}\right)^{\frac{m+n}{2}} \left(\frac{m}{n}uv\right)^{\frac{m}{2}-1} v^{\frac{n}{2}-1} \exp\left\{-\frac{v}{2}\left(\frac{m}{n}u + 1\right)\right\}. \end{aligned}$$

Luego, la **fdp** marginal de la variable  $U$  es:

$$f_U(u) = \frac{(m/n) \cdot 2^{-\frac{m+n}{2}}}{\Gamma(m/2)\Gamma(n/2)} \int_0^\infty \left(\frac{m}{n}u\right)^{\frac{m}{2}-1} v^{\frac{m+n}{2}-1} \exp\left\{-\frac{v}{2}\left(\frac{m}{n}u + 1\right)\right\} dv,$$

haciendo el siguiente cambio de variable:

$$w = \frac{v}{2}\left(\frac{m}{n}u + 1\right) \quad dw = \frac{1}{2}\left(\frac{m}{n}u + 1\right) dv.$$

Al despejar  $v$  y  $dv$ , y sustituir en la integral:

$$\begin{aligned} f_U(u) &= \frac{u^{-1} \cdot 2^{1-\frac{m+n}{2}}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}u\right)^{\frac{m}{2}} \int_0^\infty \left[2w\left(\frac{m}{n}u + 1\right)^{-1}\right]^{\frac{m+n}{2}-1} \left(\frac{m}{n}u + 1\right)^{-1} e^{-w} dw \\ &= \frac{u^{-1}}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}u\right)^{\frac{m}{2}} \int_0^\infty \left(\frac{m}{n}u + 1\right)^{-\frac{m+n}{2}} w^{\frac{m+n}{2}-1} e^{-w} dw \end{aligned}$$

$$\begin{aligned}
 &= \frac{u^{-1}}{\Gamma(m/2)\Gamma(n/2)} \Gamma\left(\frac{m+n}{2}\right) \left(\frac{m}{n}u\right)^{\frac{m}{2}} \left(\frac{m}{n}u+1\right)^{-\frac{m+n}{2}} \\
 &= \frac{u^{-1}}{B(m/2, n/2)} \left(\frac{m}{m+n}u\right)^{m/2} \left(1 - \frac{m}{m+n}u\right)^{n/2} \quad u > 0, \quad (2.62)
 \end{aligned}$$

donde  $B(\cdot)$  es la función beta dada en (1.22), una variable aleatoria real  $U$  con fdp dada por la expresión (2.62) se dice que tiene distribución  $F$ , con  $m$  grados de libertad en el numerador y  $n$  grados de libertad en el denominador y se denota  $U \sim F_{(m,n)}$ .  $\checkmark$

**Teorema 2.20.** Propiedades de la distribución  $F$ . Según [20, p. 225]:

1. Si  $X \sim F_{(m,n)}$ , entonces la variable aleatoria  $\frac{1}{X} \sim F_{(n,m)}$ .
2. Si  $X \sim t_{(n)}$ , entonces la variable aleatoria  $X^2 \sim F_{(1,n)}$ .

*Demostración.* La prueba es directa usando el teorema del cambio de variable (teorema 2.7) y con  $g(x) = \frac{1}{x}$  y  $g(x) = x^2$ , respectivamente, en (2.31).  $\checkmark$

## 2.16. Normal multivariada

De acuerdo con [74, p. 83], si  $\mathbf{Y}$  es un vector aleatorio  $p$ -dimensional tal que tiene una distribución normal multivariada con media  $\boldsymbol{\mu}_Y$  y matriz de varianza-covarianza  $\boldsymbol{\Sigma}_Y$ , la fdp conjunta es dada por:

$$f_Y(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\boldsymbol{\Sigma}_Y|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_Y)^\top \boldsymbol{\Sigma}_Y^{-1} (\mathbf{y} - \boldsymbol{\mu}_Y)\right\}.$$

Notación:  $\mathbf{Y} \sim N_p(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$ .

De acuerdo a [74, p. 88]: Si  $\mathbf{Y} \sim N_p(\boldsymbol{\mu}_Y, \mathbf{V})$ , se puede particionar como:

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}\right),$$

usando el complemento de Schur  $(\mathbf{V}/\mathbf{V}_{ii}) = \mathbf{V}_{jj} - \mathbf{V}_{ji} \mathbf{V}_{ii}^{-1} \mathbf{V}_{ij}$ ,  $i \neq j$ . Entonces,

$$\mathbf{Y}_1 | \mathbf{Y}_2 \sim N\left(\boldsymbol{\mu}_1 + \mathbf{V}_{12} \mathbf{V}_{22}^{-1} (\mathbf{Y}_2 - \boldsymbol{\mu}_2), (\mathbf{V}_{11} - \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{V}_{21})\right). \quad (2.63)$$

**Ejemplo 2.23.** (Normal  $p$ -variada). Sea  $\mathbf{X} = (X_1, X_2, X_3, \dots, X_p)^\top$  un vector aleatorio  $p$ -dimensional con vector de medias:

$$\boldsymbol{\mu}_X = (\mu_1, \mu_2, \mu_3, \dots, \mu_p)^\top = (\mathbb{E}(X_1), \mathbb{E}(X_2), \mathbb{E}(X_3), \dots, \mathbb{E}(X_p))^\top \in \mathbb{R}^p$$

y matriz de covarianzas finita:

$$\boldsymbol{\Sigma}_X = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{pmatrix} = (\sigma_{ij})_{1 \leq i, j \leq p}$$

Como aplicación del teorema de Cramér-Wold se tiene que  $\mathbf{X} = (X_1, X_2, \dots, X_p)^\top$  tiene distribución normal  $p$ -variada, notación:  $\mathbf{X} \sim N_p(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$ , si para todo  $\mathbf{c} = (c_1, c_2, \dots, c_p)^\top \in \mathbb{R}^p$  la variable aleatoria  $\mathbf{c}^\top \mathbf{X}$  tiene distribución normal univariada. La esperanza y varianza de  $\mathbf{c}^\top \mathbf{X}$  son dadas por las expresiones (2.47) y (2.50). Por lo tanto, si  $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , entonces:

$$\mathbf{c}^\top \mathbf{X} \sim N(\mathbf{c}^\top \boldsymbol{\mu}_X, \mathbf{c}^\top \boldsymbol{\Sigma}_X \mathbf{c}). \quad (2.64)$$

**Ejemplo 2.24.** Mostrar que si  $\mathbf{X} \sim N_p(\mathbf{0}, \mathbf{I}_p)$  y  $\mathbf{Y}_{p \times 1} = \mathbf{A}_{p \times p} \mathbf{X}_{p \times 1} + \boldsymbol{\mu}_{p \times 1}$  entonces  $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \mathbf{A} \mathbf{A}^\top)$ .

*Solución.* Nótese que  $\forall \mathbf{c} \in \mathbb{R}^p$ ,

$$\mathbf{c}^\top \mathbf{Y} = \mathbf{c}^\top (\mathbf{A} \mathbf{X} + \boldsymbol{\mu}) = \underbrace{\mathbf{c}^\top \mathbf{A}}_{\mathbf{b}^\top} \mathbf{X} + \mathbf{c}^\top \boldsymbol{\mu} = \mathbf{b}^\top \mathbf{X} + \mathbf{c}^\top \boldsymbol{\mu} = \mathbf{c}^\top \boldsymbol{\mu} + \sum_{j=1}^p b_j X_j.$$

La función característica de  $\mathbf{c}^\top \mathbf{Y}$  es dada por:

$$\begin{aligned} \psi_{\mathbf{c}^\top \mathbf{Y}}(u) &= \mathbb{E} \left( e^{i u \mathbf{c}^\top \mathbf{Y}} \right) = \mathbb{E} \left[ \exp \left\{ i u \left( \mathbf{c}^\top \boldsymbol{\mu} + \sum_{j=1}^p b_j X_j \right) \right\} \right] \\ &= \mathbb{E} \left[ e^{i u \mathbf{c}^\top \boldsymbol{\mu}} \exp \left\{ i u \sum_{j=1}^p b_j X_j \right\} \right] = e^{i u \mathbf{c}^\top \boldsymbol{\mu}} \mathbb{E} \left[ \exp \left\{ i u \sum_{j=1}^p b_j X_j \right\} \right] \\ &= \exp \{ i u \mathbf{c}^\top \boldsymbol{\mu} \} \underbrace{\psi_{\sum_{j=1}^p b_j X_j}(u)}_{\text{Indep.}} = \exp \{ i u \mathbf{c}^\top \boldsymbol{\mu} \} \prod_{j=1}^p \psi_{b_j X_j}(u) \\ &= \exp \{ i u \mathbf{c}^\top \boldsymbol{\mu} \} \prod_{j=1}^p \psi_{X_j}(b_j u) \stackrel{15}{=} \exp \{ i u \mathbf{c}^\top \boldsymbol{\mu} \} \prod_{j=1}^p e^{-\frac{1}{2} b_j^2 u^2} \\ &= \exp \{ i u \mathbf{c}^\top \boldsymbol{\mu} \} \exp \left\{ -\frac{1}{2} u^2 \sum_{j=1}^p b_j^2 \right\} = \exp \left\{ i u \mathbf{c}^\top \boldsymbol{\mu} - \frac{1}{2} u^2 \mathbf{b}^\top \mathbf{b} \right\} \\ &= \exp \left\{ i u \mathbf{c}^\top \boldsymbol{\mu} - \frac{1}{2} u^2 \mathbf{c}^\top \mathbf{A} \mathbf{A}^\top \mathbf{c} \right\}. \end{aligned}$$

Entonces,

$$\mathbf{c}^\top \mathbf{Y} \sim N(\mathbf{c}^\top \boldsymbol{\mu}; \mathbf{c}^\top \mathbf{A} \mathbf{A}^\top \mathbf{c}) \quad \forall \mathbf{c} \in \mathbb{R}^p$$

Usando el teorema de Cramér-Wold y del ejemplo 2.23 se tiene que:

$$\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \mathbf{A} \mathbf{A}^\top). \quad \checkmark$$

**Teorema 2.21.** Sea  $\boldsymbol{\mu}_X \in \mathbb{R}^n$  y sea  $\boldsymbol{\Sigma}_X$  una matriz definida no negativa de tamaño  $n \times n$ . El vector aleatorio  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  tiene una distribución normal multivariante con parámetros  $\boldsymbol{\mu}_X$  y  $\boldsymbol{\Sigma}_X$  y su fgm multivariante es:

$$m_X(\mathbf{t}) = \exp \left\{ \mathbf{t}^\top \boldsymbol{\mu}_X + \frac{1}{2} \mathbf{t}^\top \boldsymbol{\Sigma}_X \mathbf{t} \right\}, \quad \mathbf{t} \in \mathbb{R}^n.$$

*Demostración.* Sea  $\mathbf{a} \in \mathbb{R}^n$  un vector columna arbitrario, considérese la variable aleatoria  $Y$  definida como  $Y = \mathbf{a}^\top \mathbf{X}$ . Por las propiedades 1 del teorema 2.10 y 2 del teorema 2.11 se tiene que:

$$\mathbb{E}[Y] = \mathbb{E}[\mathbf{a}^\top \mathbf{X}] = \mathbf{a}^\top \boldsymbol{\mu}_X \quad \text{y} \quad \mathbb{V}(Y) = \mathbf{a}^\top \boldsymbol{\Sigma}_X \mathbf{a}.$$

Por la expresión (2.64) la variable  $Y$  tiene distribución normal y, por lo tanto, su fgm es:

$$m_Y(u) = \mathbb{E}(e^{uY}) = \exp \left\{ u(\mathbf{a}^\top \boldsymbol{\mu}_X) + \frac{1}{2} u^2 (\mathbf{a}^\top \boldsymbol{\Sigma}_X \mathbf{a}) \right\}.$$

Haciendo el cambio de variable  $\mathbf{t} = u\mathbf{a}$ , se tiene que:

$$\mathbb{E}(e^{\mathbf{a}^\top \mathbf{X}}) = \exp \left\{ \mathbf{t}^\top \boldsymbol{\mu}_X + \frac{1}{2} \mathbf{t}^\top \boldsymbol{\Sigma}_X \mathbf{t} \right\}. \quad \checkmark$$

## 2.17. Ejercicios

### Probabilidad y estadística univariada

1. Usando la fgm, si  $X_1$  y  $X_2$  son dos variables aleatorias independientes donde  $X_1 \sim \chi_{(k_1)}^2$  y  $X_2 \sim \chi_{(k_2)}^2$ . ¿Cuál es la distribución de  $X_1 + X_2$ ?
2. Sea  $X$  una variable aleatoria con fdp dada por  $f_X(x) = 3e^{-3x}$ ,  $x \geq 0$ , defina  $U = \sqrt{X}$ , determine la fdp de  $U$ .

<sup>15</sup> Como  $\mathbf{X} \sim N_p(\mathbf{0}, \mathbf{I}_p)$ , se tiene que  $X_k \sim N(0, 1)$  y, por tanto,  $\psi_{X_k}(v) = e^{-\frac{v^2}{2}}$ .

3. Sean  $X$  y  $Y$  dos variables aleatorias con **fdp** conjunta dada por [16, p. 227]:

$$f_{(X,Y)}(x,y) = \begin{cases} \frac{1}{x} & \text{si } 0 < x \leq 1, 0 < y \leq x \\ 0 & \text{en otro caso.} \end{cases}$$

Defina  $\mathbf{X} = (X, Y)$ . Determine  $E(\mathbf{X})$ ,  $V(\mathbf{X})$  y  $\rho_{\mathbf{X}}$ .

### Probabilidad y estadística multivariada

4. Sean  $X_1$  y  $X_2$  variables aleatorias **iid** normales estándar, es decir,  $X_k \sim N(0, 1)$  [65, p. 214]. Encuentre la **fgm** de  $Y = X_1 X_2$ .
5. Comprobar que las funciones  $c(\theta)$ ,  $d(\theta)$ ,  $T(x)$  y  $S(x)$  presentadas en la tabla 2.1 permiten representar la familia dada como un miembro de la familia exponencial y encontrar el conjunto  $\mathcal{A}$ .
6. Encontrar funciones  $\theta(\eta)$  y  $d_0(\eta)$  para cada una de las familias de distribuciones dadas en la tabla 2.1.





### 3.1. Conceptos básicos de muestra aleatoria

**DEFINICIÓN 3.1. Muestra aleatoria.** Variables aleatorias que poseen la misma fda son llamadas **idénticamente distribuidas**. Si  $X_1, X_2, \dots, X_n$  son variables aleatorias iid, con fda  $F = F_{X_i}$ , se dice que las  $X_i$  forman una **muestra aleatoria** de tamaño  $n$  (extraída de  $F$ , o extraída de una población con distribución  $F$ ) (ver [43, p. 84]).

**DEFINICIÓN 3.2. Estadístico(a).** Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de tamaño  $n$  de una población y sea  $T(X_1, X_2, \dots, X_n)$  una función de valor real o un vector cuyo dominio incluye el espacio muestral de  $(X_1, X_2, \dots, X_n)$ , entonces la variable aleatoria o vector aleatorio  $Y = T(X_1, X_2, \dots, X_n)$  se llama un **estadístico(a)**. La distribución de probabilidad de un estadístico  $Y$  es llamada la **distribución de muestreo** de  $Y$  (ver [20, p. 211]).

**Ejemplo 3.1.** Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria y sea el vector aleatorio  $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$ , encuentre algunos estadísticos.

*Solución.* Algunos ejemplos de estadísticos son:

$$\begin{aligned}
 T_1(\mathbf{X}) &= \sum_{i=1}^n X_i & T_6(\mathbf{X}) &= \frac{T_4(\mathbf{X}) + T_5(\mathbf{X})}{2} \\
 T_2(\mathbf{X}) &= \sum_{i=1}^n X_i^2 \\
 T_3(\mathbf{X}) &= \frac{1}{n} \sum_{i=1}^n X_i & T_7(\mathbf{X}) &= \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \quad \checkmark \\
 T_4(\mathbf{X}) &= \min_{1 \leq i \leq n} \{X_i\} \\
 T_5(\mathbf{X}) &= \max_{1 \leq i \leq n} \{X_i\}
 \end{aligned}$$

**Ejemplo 3.2.** En la tabla 3.1 se presentan algunos datos ilustrativos de un experimento sobre la respuesta de los árboles jóvenes de tung a los fertilizantes. Los datos muestran las alturas medidas en centímetros, tomadas luego de 7 meses de plantados los árboles.

Tabla 3.1. Datos ilustrativos sobre altura (cm) de árboles tung luego de siete meses de plantados

80	86	70	74	82	86	78	55	82
67	67	57	68	73	75	78	51	66

Determine algunos estadísticos.

*Solución.* Los datos ordenados se presentan en la tabla 3.2.

Tabla 3.2. Datos ilustrativos sobre altura (cm) de árboles tung luego de siete meses de plantados, datos ordenados

51	55	57	66	67	67	68	70	73
$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$
74	75	78	78	80	82	82	86	86
$x_{(10)}$	$x_{(11)}$	$x_{(12)}$	$x_{(13)}$	$x_{(14)}$	$x_{(15)}$	$x_{(16)}$	$x_{(17)}$	$x_{(18)}$

Luego, las estadísticas dadas en el ejemplo 3.1 para estos datos son:

$$\begin{aligned}
 T_1(\mathbf{X}) &= \sum_{i=1}^{18} X_i = 1295 & T_6(\mathbf{X}) &= \frac{T_4(\mathbf{X})+T_5(\mathbf{X})}{2} = 68,5 \\
 T_2(\mathbf{X}) &= \sum_{i=1}^{18} X_i^2 = 94971 \\
 T_3(\mathbf{X}) &= \frac{1}{18} \sum_{i=1}^{18} X_i = 71,944 & T_7(\mathbf{X}) &= \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_{18} \end{pmatrix} = \begin{pmatrix} 80 \\ 86 \\ 70 \\ \vdots \\ 51 \\ 66 \end{pmatrix} \quad \checkmark \\
 T_4(\mathbf{X}) &= \min_{1 \leq i \leq 18} \{X_i\} = 51 \\
 T_5(\mathbf{X}) &= \max_{1 \leq i \leq 18} \{X_i\} = 86
 \end{aligned}$$

**Ejemplo 3.3.** Sean  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una variable aleatoria con distribución  $X \sim N(0, 1)$ . Entonces a medida que aumenta  $n$ , la probabilidad de que  $|X_n|$  sea mayor que, por ejemplo 1,5, es:

$$P(|X_n| > 1,5) = P(X_n > 1,5) + P(X_n < -1,5) = 0,1336,$$

la cual permanece fija y no tiende a cero.

**Ejemplo 3.4.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid  $X \sim \exp(1)$ . Sea  $Y_n = \frac{X_n}{\ln(n)}$ , muestre que a medida que aumenta  $n$ , la probabilidad de que  $|Y_n|$  sea mayor que un valor  $\zeta > 0$  tiende a cero.

*Solución.* Por el teorema de transformación (teorema 2.7):

$$P(|Y_n| \geq \zeta) = P\left(\frac{X_n}{\ln(n)} \geq \zeta\right) = P(X_n \geq \zeta \ln(n)) = P(X_n \geq \ln(n^\zeta)).$$

Puesto que  $X_n \sim \exp(1)$ , se sabe que:

$$f_{X_n}(x) = e^{-x} \quad \text{y} \quad F_{X_n}(x) = 1 - e^{-x}, \quad (3.1)$$

por tanto,

$$P(X_n \geq \ln(n^\zeta)) = \exp\{-\ln(n^\zeta)\} = \exp\{\ln(n^{-\zeta})\} = \frac{1}{n^\zeta}. \quad (3.2)$$

Si  $\zeta = 0,3$ , entonces  $P(|Y_n| > 0,3) = \frac{1}{n^{0,3}}$ , *la cual tiende a cero.*  $\square$

### 3.2. Estadísticos de orden

**DEFINICIÓN 3.3.** De acuerdo con [20, p. 226], dada una muestra aleatoria de tamaño  $n$ ,  $X_1, X_2, \dots, X_n$  de una población con *fda*,  $F$ , los estadísticos de orden son los valores de la muestra organizados en orden ascendente. Estos se denotan por  $X_{(1)}, X_{(2)}, \dots, X_{(j)}, \dots, X_{(n)}$ .

**Nota 3.1.** La mediana muestral depende del tamaño de muestra  $n$  y está dada por:

$$Me = \begin{cases} X_{(m+1)}, & \text{si } n = 2m + 1, \\ \frac{1}{2} [X_{(m)} + X_{(m+1)}], & \text{si } n = 2m. \end{cases}$$

**Ejemplo 3.5.** Según [16, p. 206], si  $X_1, X_2, \dots, X_n$  son variables aleatorias reales independientes definidas sobre  $(\Omega, \mathcal{F}, P)$  y se definen las variables aleatorias  $Y$  y  $Z$  como:

$$Y = X_{(n)} = \max_{1 \leq k \leq n} \{X_k\} \quad \text{y} \quad Z = X_{(1)} = \min_{1 \leq k \leq n} \{X_k\}.$$

Determine  $F_Y(y)$  y  $F_Z(z)$ .

*Solución.* Sean  $F_{X_k}(\cdot)$  y  $f_{X_k}(\cdot)$  la *fda* y *fdp* de las variables aleatorias  $X_k$ , es claro que:

1. Para la variable  $Y = X_{(n)}$  se tiene que:

$$\begin{aligned} Y : \Omega &\rightarrow \mathbb{R} \\ \omega &\rightarrow \max_{1 \leq k \leq n} \{X_k(\omega)\} \end{aligned}$$

luego,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X_1 \leq y, X_2 \leq y, X_3 \leq y, \dots, X_n \leq y) \\ &= \prod_{k=1}^n P(X_k \leq y) = \prod_{k=1}^n F_{X_k}(y), \end{aligned} \tag{3.3}$$

al derivar logarítmicamente (3.3) se llega a:

$$\ln [F_Y(y)] = \sum_{k=1}^n \ln [F_{X_k}(y)]$$

$$f_Y(y) = \sum_{k=1}^n \frac{f_{X_k}(y)}{F_{X_k}(y)} \prod_{k=1}^n F_{X_k}(y).$$

Si las variables aleatorias son idénticamente distribuidas entonces:

$$F_Y(y) = [F_X(y)]^n \quad \text{y} \quad f_Y(y) = n [F_X(y)]^{n-1} f_X(y) \quad (3.4)$$

2. Para la variable  $Z = X_{(1)}$  se tiene que:

$$\begin{aligned} Z : \Omega &\rightarrow \mathbb{R} \\ \omega &\rightarrow \min_{1 \leq k \leq n} \{X_k(\omega)\}, \end{aligned}$$

luego,

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = 1 - P(X_1 > z, X_2 > z, X_3 > z, \dots, X_n > z) \\ &= 1 - \prod_{k=1}^n P(X_k > z) = 1 - \prod_{k=1}^n [1 - P(X_k \leq z)] \\ &= 1 - \prod_{k=1}^n [1 - F_{X_k}(z)], \end{aligned} \quad (3.5)$$

al derivar logarítmicamente (3.5) se llega a:

$$\begin{aligned} \ln [1 - F_Z(z)] &= \sum_{k=1}^n \ln [1 - F_{X_k}(z)] \\ f_Z(z) &= \sum_{k=1}^n \frac{f_{X_k}(z)}{1 - F_{X_k}(z)} \prod_{k=1}^n [1 - F_{X_k}(z)]. \end{aligned}$$

Si las variables aleatorias son idénticamente distribuidas entonces:

$$F_Z(z) = 1 - [1 - F_X(z)]^n \quad \text{y} \quad f_Z(z) = n [1 - F_X(z)]^{n-1} f_X(z). \quad (3.6)$$

☑

**Teorema 3.1.** De acuerdo con [20, p. 227], dada una muestra aleatoria  $X_1, X_2, \dots, X_n$  de una distribución discreta con fmp  $p(x_i) = P(X = x_i) = p_i$ ,  $i = 1, 2, \dots$ , donde  $x_1 < x_2 < \dots$  son los posibles valores de  $X$  en orden ascendente. Sean:

$$P_i = \sum_{k=1}^i p_k = P(X_j \leq x_i) = F_{X_j}(x_i) \quad j = 1, 2, \dots, n,$$

donde  $P_0 = 0$  y puesto que las variables aleatorias  $X_1, X_2, \dots, X_n$  son idénticamente distribuidas  $P_i = F_{X_1}(x_i) = \dots = F_{X_j}(x_i)$ . Entonces,

$$P(X_{(j)} \leq x_i) = \sum_{k=j}^n \binom{n}{k} P_i^k (1 - P_i)^{n-k}. \tag{3.7}$$

*Demostración.* Para  $i$  fijo sea la variable aleatoria  $Y$  que cuenta el número de  $X_1, X_2, \dots, X_n$  que son menores o iguales a  $x_i$ , para cada  $X_1, X_2, \dots, X_n$  el evento  $\{X_j \leq x_i\}$  es un acierto y el evento  $\{X_j > x_i\}$  es un fracaso.

Entonces se puede considerar que  $Y$  es el número de éxitos en  $n$  ensayos independientes, es decir,  $Y \sim \text{Bin}(n, P_i)$ . Luego, el evento  $\{X_{(j)} \leq x_i\}$  es equivalente al evento  $\{Y \geq j\}$ , por lo tanto,

$$P(X_{(j)} \leq x_i) = P(Y \geq j) \tag{3.8}$$

y la expresión (3.7) representa la fda. □

**Ejemplo 3.6.** Suponga que se genera una muestra aleatoria de tamaño 20 de una variable con fmp uniforme de parámetro  $n = 10$ , y cuyos valores son 10, 20, 35, 50, 55, 70, 75, 85, 90 y 100. En la tabla 3.3 se presenta  $p(x_i)$  y  $P_i$ .

Tabla 3.3. Función de frecuencia,  $P(10)$

$i$	$x_i$	$P(X = x_i) = p_i$	$P(X \leq x_i) = P_i$
0			0
1	10	1/10	1/10
2	20	1/10	2/10
3	35	1/10	3/10
4	50	1/10	4/10
5	55	1/10	5/10
6	70	1/10	6/10
7	75	1/10	7/10
8	85	1/10	8/10
9	90	1/10	9/10
10	100	1/10	1

Determine para esta muestra e  $i = 8$  la probabilidad  $P(X_{(j)} = x_8)$ .

*Solución.* La tabla 3.4 muestra los datos generados (usando `set.seed(500)`).

Tabla 3.4. Muestra aleatoria de tamaño 20, ejemplo 3.6

90	85	100	55	90	35	70	100	90	85
35	90	85	20	85	35	70	75	75	35

De acuerdo con (3.7), la variable aleatoria  $Y$  que cuenta el número de  $X_1, X_2, \dots, X_{20}$  que son menores o iguales a  $x_8 = 85$  tiene distribución binomial de parámetros  $n = 20$  y  $P_8 = 8/10$ , esto es,  $Y \sim \text{Bin}(20, 8/10)$ .

En la tabla 3.5 se presentan los datos ordenados, en la cual se observa que  $y = 14$ , esto es, el número de resultados en los 20 ensayos que son menores o iguales a 85 fueron 14. Note que  $x_{(11)} = x_{(12)} = x_{(13)} = x_{(14)} = 85$ .

Tabla 3.5. Muestra aleatoria de tamaño 20, ejemplo 3.6, datos ordenados

20	35	35	35	35	55	70	70	75	75
$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$	$x_{(10)}$
85	85	85	85	90	90	90	90	100	100
$x_{(11)}$	$x_{(12)}$	$x_{(13)}$	$x_{(14)}$	$x_{(15)}$	$x_{(16)}$	$x_{(17)}$	$x_{(18)}$	$x_{(19)}$	$x_{(20)}$

Para  $n = 20$  e  $i = 8$ , la probabilidad solicitada  $P(Y = 14)$ , sería:

$$P(Y = 14) = \binom{20}{14} (8/10)^{14} (1 - 8/10)^{20-14} = 0,1091. \quad \checkmark$$

Las instrucciones en R para generar la muestra fueron:

```
remove(list=ls())
set.seed(500)
sam=sample(c(10, 20, 35, 50, 55, 70, 75, 85, 90, 100), 20,
           replace=TRUE)
sam[order(sam)]
```

**Ejemplo 3.7.** Considere la muestra del ejemplo 3.6, determine la probabilidad de que el mínimo, el máximo y la estadística de orden 2 sean menores o iguales que 85.

*Demostración.* Se está interesado en las siguientes probabilidades:

1. La probabilidad de que al menos 1 dato de los 20 datos generados en la muestra sea menor o igual que 85, esto es,

$$P(X_{(1)} \leq 85) = P(Y \geq 1) = 1 - P(Y < 1) = 1 - P(Y = 0) \approx 1.$$

2. La probabilidad de que todos los 20 datos generados en la muestra sean menores o iguales que 85, es decir, por la expresión (3.8),

$$P(X_{(20)} \leq 85) = P(Y \geq 20) = P(Y = 20) = 0,0115.$$

3. La probabilidad de que al menos 2 datos de los 20 datos generados en la muestra sean menores o iguales que 85, esto es,

$$P(X_{(2)} \leq 85) = P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1) \approx 1. \quad \checkmark$$

**Ejemplo 3.8.** Suponga que se genera una muestra aleatoria de tamaño 15 de una variable con distribución Poisson de parámetro  $\lambda = 10$ , esto es,  $P(10)$ . En la tabla 3.6 se presenta  $p(x_i)$  y  $P_i$ .

Tabla 3.6. Función de frecuencia,  $P(10)$

$i$	$x_i$	$P(X = x_i) = p_i$	$P(X_k \leq x_i) = P_i$
0			0,0000
1	0	0,00005	0,00005
2	1	0,0005	0,00055
3	2	0,0023	0,0027
4	3	0,0076	0,0103
5	4	0,0189	0,0293
6	5	0,0378	0,0671
7	6	0,0631	0,1301
8	7	0,0901	0,2202
9	8	0,1126	0,3328
10	9	0,1251	0,4579
11	10	0,1251	0,5830
12	11	0,1137	0,6968
13	12	0,0948	0,7916
14	13	0,0729	0,8645
15	14	0,0521	0,9165
16	15	0,0347	0,9512
17	16	0,0217	0,9730
18	17	0,0128	0,9857
...	...	...	...

Determine para esta muestra e  $i = 9$  la probabilidad  $P(X_{(j)} = x_9)$ .

*Solución.* La tabla 3.7 muestra los datos generados (usando `set.seed(1000)`).

De acuerdo con (3.7), la variable aleatoria  $Y$  que cuenta el número de  $X_1, X_2, \dots, X_{15}$  que son menores o iguales a  $x_9 = 8$  tiene distribución binomial de parámetros  $n = 15$  y  $P_9 = 0,3328$ , esto es,  $Y \sim B(15, 0,3328)$ .

Tabla 3.7. Muestra aleatoria de tamaño 15, fmp Poisson de parámetro  $\lambda = 10$

8	6	5	7	12	13	5	13
5	9	5	4	10	3	8	

En la tabla 3.8 se presentan los datos ordenados, en la cual se observa que  $y = 10$ , esto significa que el número de resultados en los 15 ensayos menores o iguales a 8 fueron 10. Note que  $x_{(9)} = x_{(10)} = 8$ .

Tabla 3.8. Muestra aleatoria de tamaño 15, fmp Poisson de parámetro  $\lambda = 10$ , datos ordenados

3	4	5	5	5	5	6	7
$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$
8	8	9	10	12	13	13	
$x_{(9)}$	$x_{(10)}$	$x_{(11)}$	$x_{(12)}$	$x_{(13)}$	$x_{(14)}$	$x_{(15)}$	

Para  $n = 15$  e  $i = 9$ , la probabilidad solicitada  $P(Y = 14)$ , sería: si se está interesado en  $P(Y = 10)$ , cuando  $n = 15$  e  $i = 9$ , se tendría:

$$P(Y = 10) = \binom{15}{10} (0,3328)^{10} (1 - 0,3328)^{15-10} = 0,0066. \quad \checkmark$$

Las instrucciones en  $R$  para generar la muestra fueron:

```
remove(list=ls())
set.seed(1000)
muestra=rpois(15,10)
muestra[order(muestra)]
```

**Nota 3.2.** En general, una vez fijado  $i$ ,  $Y \sim \text{Bin}(n, P_i)$ , esto es,

$$P(Y = j) = \binom{n}{j} P_i^j (1 - P_i)^{n-j}.$$

Para obtener la *fda* de  $X_{(j)}$ , esto es, la *fda* de la estadística de orden  $j$ ,  $P(X_{(j)} \leq x_i)$ , se emplea la expresión (3.7), por tanto:

$$P(X_{(j)} \leq x_i) = P(Y \geq j) = \sum_{k=j}^n \binom{n}{k} P_i^k (1 - P_i)^{n-k}.$$

Por ejemplo para  $j = 1$ , el mínimo,

$$\begin{aligned} P(X_{(1)} \leq x_i) &= P(Y \geq 1) = \sum_{k=1}^n \binom{n}{k} P_i^k (1 - P_i)^{n-k} \\ &= 1 - \binom{n}{0} P_i^0 (1 - P_i)^{n-0} = 1 - (1 - P_i)^n, \end{aligned}$$

la cual coincide con la expresión (3.6) y para  $j = n$ , el máximo,

$$P(X_{(n)} \leq x_i) = P(Y \geq n) = \sum_{k=n}^n \binom{n}{k} P_i^k (1 - P_i)^{n-k} = P_i^n.$$

Nótese que coincide con la expresión (3.4). Para la fmp de  $X_{(j)}$  se tiene que:

$$\begin{aligned} \underbrace{P(x_{i-1} < X_{(j)} \leq x_i)} &= \sum_{k=j}^n \binom{n}{k} P_i^k (1 - P_i)^{n-k} - \sum_{k=j}^n \binom{n}{k} P_{i-1}^k (1 - P_{i-1})^{n-k} \\ P(X_{(j)} = x_i) &= \sum_{k=j}^n \binom{n}{k} [P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k}]. \end{aligned}$$

**Teorema 3.2.** De acuerdo con [20, p. 229], si  $X_{(1)}, X_{(2)}, X_{(3)}, \dots, X_{(n)}$  son los estadísticos de orden de una muestra aleatoria  $X_1, X_2, \dots, X_n$  de una distribución continua con fdp  $f(x)$ , entonces para un  $x_i$  fijo, la fdp de  $X_{(j)}$  es:

$$f_{X_{(j)}}(x_i) = \frac{n!}{(n-j)!(j-1)!} f_X(x_i) [F_X(x_i)]^{j-1} [1 - F_X(x_i)]^{n-j}, \quad (3.9)$$

donde la fda es:

$$F_X(x_i) = \int_{-\infty}^{x_i} f_X(t) dt.$$

*Demostración.* Supóngase que la variable aleatoria  $Y$  cuenta el número de  $X_1, X_2, \dots, X_n$  que son menores (o iguales) a  $x_i$ , luego  $Y \sim B(n, F_X(x_i))$  y la expresión (3.7) queda:

$$P(X_{(j)} \leq x_i) = P(Y \geq j) = \sum_{k=j}^n \binom{n}{k} [F_X(x_i)]^k [1 - F_X(x_i)]^{n-k}. \quad (3.10)$$

Para  $j = 1$ , la fda del mínimo está dada por:

$$P(X_{(1)} \leq x_i) = P(Y \geq 1) = \sum_{k=1}^n \binom{n}{k} [F_X(x_i)]^k [1 - F_X(x_i)]^{n-k}$$

$$= 1 - \binom{n}{0} [F_X(x_i)]^0 [1 - F_X(x_i)]^{n-0} = 1 - (1 - F_X(x_i))^n,$$

la cual coincide con (3.6) y para  $j = n$ , la fda del máximo es dada por:

$$P(X_{(n)} \leq x_i) = P(Y \geq n) = \sum_{k=n}^n \binom{n}{k} [F_X(x_i)]^k [1 - F_X(x_i)]^{n-k} = [F_X(x_i)]^n.$$

Nótese que esta expresión coincide con (3.4). Para encontrar  $f_{X_{(j)}}(x_i)$ , se usa el primer teorema fundamental del cálculo (ver nota B.11),

$$\begin{aligned} f_{X_{(j)}}(x_i) &= \frac{dF_{X_{(j)}}(x_i)}{dx_i} = \frac{d}{dx_i} \sum_{k=j}^n \binom{n}{k} [F_X(x_i)]^k [1 - F_X(x_i)]^{n-k} \\ &= \sum_{k=j}^n \binom{n}{k} \left\{ k [F_X(x_i)]^{k-1} f_X(x_i) [1 - F_X(x_i)]^{n-k} \right. \\ &\quad \left. - [F_X(x_i)]^k (n-k) [1 - F_X(x_i)]^{n-k-1} f_X(x_i) \right\} \\ &= f_X(x_i) \sum_{k=j}^n \binom{n}{k} k [F_X(x_i)]^{k-1} [1 - F_X(x_i)]^{n-k} \\ &\quad - f_X(x_i) \sum_{k=j}^n \binom{n}{k} (n-k) [F_X(x_i)]^k [1 - F_X(x_i)]^{n-k-1} \\ &= f_X(x_i) \binom{n}{j} j [F_X(x_i)]^{j-1} [1 - F_X(x_i)]^{n-j} \\ &\quad + f_X(x_i) \sum_{k=j+1}^n \binom{n}{k} k [F_X(x_i)]^{k-1} [1 - F_X(x_i)]^{n-k} \\ &\quad - f_X(x_i) \sum_{k=j}^{n-1} \binom{n}{k} (n-k) [F_X(x_i)]^k [1 - F_X(x_i)]^{n-k-1} \\ &= f_X(x_i) \frac{n!}{(n-j)!(j-1)!j} [F_X(x_i)]^{j-1} [1 - F_X(x_i)]^{n-j} \\ &\quad + f_X(x_i) \sum_{k=j+1}^n \frac{n!}{(n-k)!(k-1)!} [F_X(x_i)]^{k-1} [1 - F_X(x_i)]^{n-k} \\ &\quad - f_X(x_i) \sum_{k=j}^{n-1} \frac{n!}{(n-k-1)!k!} [F_X(x_i)]^k [1 - F_X(x_i)]^{n-k-1}. \\ &\stackrel{=}{=} \frac{n!}{(n-j)!(j-1)!} f_X(x_i) [F_X(x_i)]^{j-1} [1 - F_X(x_i)]^{n-j} \\ &\quad + f_X(x_i) \sum_{l=j}^{n-1} \frac{n!}{(n-l-1)!l!} [F_X(x_i)]^l [1 - F_X(x_i)]^{n-l-1} \\ &\quad - f_X(x_i) \sum_{k=j}^{n-1} \frac{n!}{(n-k-1)!k!} [F_X(x_i)]^k [1 - F_X(x_i)]^{n-k-1}. \end{aligned}$$

El cambio de variable  $l = k - 1$  solo se hizo en la primera sumatoria. Cancelando términos se llega a:

$$f_{X_{(j)}}(x_i) = \frac{n!}{(n-j)!(j-1)!} f_X(x_i) [F_X(x_i)]^{j-1} [1 - F_X(x_i)]^{n-j}. \quad \checkmark$$

**Ejemplo 3.9.** Sean  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una distribución uniforme sobre el intervalo  $\mathcal{I}_\theta = (0, \theta)$ ,  $\theta > 0$ . ¿Cuál es la *fdp* y *fda* del máximo, esto es,  $f_{X_{(n)}}(x)$  y  $F_{X_{(n)}}(x)$ ?

*Solución.* La *fdp* de  $X_1$  es dada por:

$$p_{X_1}(x; \theta) = f_{X_1}(x) = \begin{cases} \frac{1}{\theta}, & \text{si } 0 < x < \theta \\ 0, & \text{en otro caso,} \end{cases} \quad (3.11)$$

y

$$F_{X_1}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{x}{\theta} & \text{para } 0 < x < \theta \\ 1 & \text{si } x \geq \theta, \end{cases}$$

por la expresión (3.4) se tiene que:

$$f_{X_{(n)}}(x) = \frac{n}{\theta} \left[ \frac{x}{\theta} \right]^{n-1} = n\theta^{-n} x^{n-1} \quad \text{para } 0 < x < \theta,$$

y

$$F_{X_{(n)}}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{x^n}{\theta^n} & \text{para } 0 < x < \theta \\ 1 & \text{si } x \geq \theta. \end{cases} \quad (3.12)$$

De acuerdo con [87] la variable  $X_{(n)}$  tiene *fdp* uniforme generalizada y se denota por  $X_{(n)} \sim GU(n - 1, 0, \theta)$ . \checkmark

**Teorema 3.3.** De acuerdo con [20, p. 230], si  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ , son los estadísticos de orden de una muestra aleatoria,  $X_1, X_2, \dots, X_n$  con *fdp*  $f_X(x)$  y *fda*  $F_X(x)$ . Entonces la *fdp* conjunta de  $X_{(i)}$  y  $X_{(j)}$ ,  $1 \leq i < j \leq n$ , es dada por:

$$f_{(X_{(i)}, X_{(j)})}(u, v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f_X(u) f_X(v) \times [F_X(u)]^{i-1} [F_X(v) - F_X(u)]^{j-1-i} [1 - F_X(v)]^{n-j},$$

para  $-\infty < u < v < \infty$ .

*Demostración.* Ver detalles de la prueba en [20, p. 260]. \checkmark

**Nota 3.3.** (Distribución del rango). La *fdp* conjunta del mínimo y el máximo está dada por ( $i = 1, j = n$ ):

$$\begin{aligned} f_X(x_1, x_n) &= \frac{n! f_X(x_1) f_X(x_n)}{(n-2)!} [F_X(x_1)]^0 [F_X(x_n) - F_X(x_1)]^{n-2} [1 - F_X(x_n)]^0 \\ &= n(n-1) f_X(x_1) f_X(x_n) [F_X(x_n) - F_X(x_1)]^{n-2}, \end{aligned} \quad (3.13)$$

donde  $\mathbf{X} = (X_{(1)}, X_{(n)})$ . Sean  $U$  y  $V$  las variables aleatorias rango medio y rango, es decir,

$$U = \frac{1}{2} [X_{(n)} + X_{(1)}] \quad y \quad V = X_{(n)} - X_{(1)}.$$

Este sistema de ecuaciones puede escribirse matricialmente como:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_n \end{pmatrix} \quad y \quad \begin{pmatrix} x_1 \\ x_n \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{2} \\ 1 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u - v/2 \\ u + v/2 \end{pmatrix};$$

la última transformación (inversa) se puede expresar como:

$$(h_1(x, y), h_2(x, y)) = \left(x - \frac{y}{2}, x + \frac{y}{2}\right),$$

y la matriz jacobiana dada en (2.57) de la transformación inversa queda:

$$\mathcal{J}_h(x, y) = \begin{bmatrix} 1 & -\frac{1}{2} \\ 1 & \frac{1}{2} \end{bmatrix} \quad y \quad \det [\mathcal{J}_h(x, y)] = 1.$$

Por lo tanto, la *fdp* conjunta de  $U$  y  $V$  está dada por (2.56); es decir,

$$\begin{aligned} f_{(U, V)}(u, v) &= f_{(X_{(1)}, X_{(n)})}\left(u - \frac{v}{2}, u + \frac{v}{2}\right) \\ &= n(n-1) f_X\left(u - \frac{v}{2}\right) f_X\left(u + \frac{v}{2}\right) [F_X\left(u + \frac{v}{2}\right) - F_X\left(u - \frac{v}{2}\right)]^{n-2}. \end{aligned}$$

Luego, la *fdp* marginal de la variable  $V$  es:

$$f_V(v) = \int_{-\infty}^{\infty} n(n-1) f_X\left(u - \frac{v}{2}\right) f_X\left(u + \frac{v}{2}\right) [F_X\left(u + \frac{v}{2}\right) - F_X\left(u - \frac{v}{2}\right)]^{n-2} du,$$

haciendo cambio de variable  $z = u - \frac{v}{2}$ , entonces  $dz = du$ ; luego:

$$f_V(v) = \int_{-\infty}^{\infty} n(n-1) f_X(z) f_X(z+v) [F_X(z+v) - F_X(z)]^{n-2} dz. \quad (3.14)$$

**Ejemplo 3.10.** Continuando con el ejemplo 3.9, ¿cuál es la distribución conjunta del máximo y del mínimo?

*Solución.* En este caso, por la expresión (3.13) se tiene que:

$$f_{X_{(1)}, X_{(n)}}(x_1, x_n) = \frac{n(n-1)}{\theta^2} \times \frac{(x_n - x_1)^{n-2}}{\theta^{n-2}} = \frac{n(n-1)(x_n - x_1)^{n-2}}{\theta^n},$$

para  $0 < x_1 < x_n < \theta$ . ✓

### 3.3. Órdenes de magnitud

Un orden de magnitud es una medida o magnitud de alguna cantidad, donde cada tipo incluye valores de un cociente fijo en términos del valor previo. Las referencias usadas en la elaboración de esta sección fueron [55] y [82].

#### 3.3.1. Órdenes de magnitud de sucesiones de números reales y vectores. Notación: $O(\cdot)$ y $o(\cdot)$

El orden de magnitud de un número nos permite concebir de inmediato su dimensión respecto a sus centenas, decenas, unidades, décimas, centésimas.

**Ejemplo 3.11.** La tabla 3.9 muestra algunas sucesiones de números  $a_n$ , y la figura 3.1, los respectivos gráficos que exhiben el comportamiento monótono (creciente o decreciente) de estas.

Tabla 3.9. Ejemplo de sucesiones de números  $a_n$

$a_n$	$n$								
	1	2	3	4	10	50	100	200	500
$n^2$	1	4	9	16	100	2500	10000	40000	250000
$\frac{1}{n^2}$	1	$\frac{1}{4}$	$\frac{1}{9}$	$\frac{1}{16}$	$\frac{1}{100}$	$\frac{1}{2500}$	$\frac{1}{10000}$	$\frac{1}{40000}$	$\frac{1}{250000}$
$\frac{1}{n}$	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{10}$	$\frac{1}{50}$	$\frac{1}{100}$	$\frac{1}{200}$	$\frac{1}{500}$
$1 - \frac{1}{n}$	0	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{3}{4}$	$\frac{9}{10}$	$\frac{49}{50}$	$\frac{99}{100}$	$\frac{199}{200}$	$\frac{499}{500}$
$(1 + \frac{1}{n})^n$	2	2,250	2,370	2,441	2,594	2,692	2,705	2,712	2,716
$\ln(n)$	0	0,693	1,099	1,386	2,303	3,912	4,605	5,298	6,215

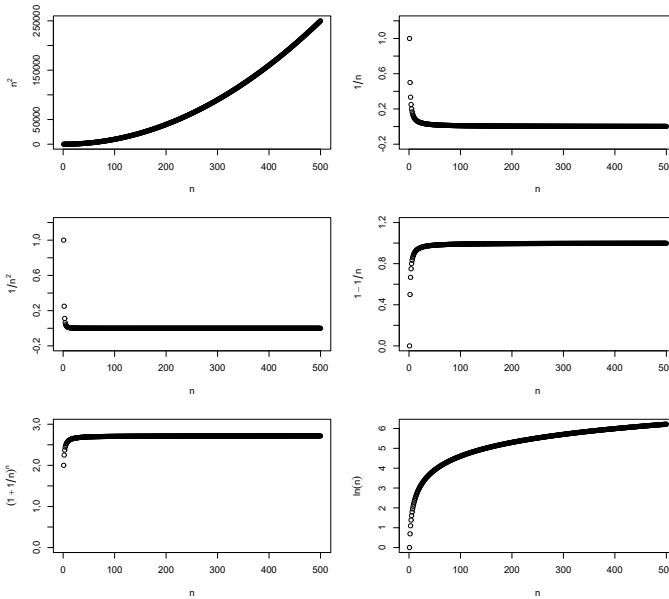


Figura 3.1. Ejemplos de sucesiones de números  $a_n$

**DEFINICIÓN 3.4.** De acuerdo con [82, p. 36], dadas  $\{a_n\}_{n \geq 1}$  y  $\{b_n\}_{n \geq 1}$  sucesiones de números reales. Se dice que:

1.  $a_n = O(b_n)$  si existe un número real  $k > 0$  y un número entero positivo  $n_0 = n_0(k)$  tal que:

$$\left| \frac{a_n}{b_n} \right| \leq k, \quad \forall n \geq n_0.$$

2.  $a_n = o(b_n)$  si para todo número real  $\varepsilon > 0$  existe un número entero positivo  $n_0 = n_0(\varepsilon)$  tal que:

$$\left| \frac{a_n}{b_n} \right| \leq \varepsilon, \quad \forall n \geq n_0.$$

Así,  $a_n = O(b_n)$  si la razón  $\left| \frac{a_n}{b_n} \right|$  está acotada para todo  $n$  suficientemente grande, y que  $a_n = o(b_n)$  si:

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0.$$

**Ejemplo 3.12.** Muestre que:

1)  $n = o(n^2)$ .

2)  $10n^2 + n = O(n^2)$ .

*Solución.* Veamos:

1)  $n = o(n^2)$ ,

$$\lim_{n \rightarrow \infty} \frac{n}{n^2} = \lim_{n \rightarrow \infty} \frac{1}{n} = 0.$$

2)  $10n^2 + n = O(n^2)$ ,

$$\frac{10n^2 + n}{n^2} = 10 + \frac{1}{n},$$

tomando  $k = 11$ , se tiene que:

$$\left| \frac{10n^2 + n}{n^2} \right| \leq 11, \forall n \geq 1. \quad \checkmark$$

**Ejemplo 3.13.** Muestre que si  $a_n = O(b_n)$ , entonces  $-a_n = O(b_n)$ .

*Solución.* Por la definición 3.4 si  $a_n = O(b_n)$ , entonces existe  $k_1 > 0$  y  $n_1(k_1)$  tal que  $\left| \frac{a_n}{b_n} \right| \leq k_1$  para todo  $n \geq n_1(k_1)$ , luego,

$$\left| \frac{a_n}{b_n} \right| \leq k_1 \Leftrightarrow -k_1 \leq \frac{a_n}{b_n} \leq k_1 \Leftrightarrow k_1 \geq \frac{-a_n}{b_n} \geq -k_1 \Leftrightarrow \left| \frac{-a_n}{b_n} \right| \leq k_1,$$

entonces existe  $k_1 > 0$  y  $n_1(k_1)$  tal que  $\left| \frac{-a_n}{b_n} \right| \leq k_1$  para todo  $n \geq n_1(k_1)$ , es decir,  $-a_n = O(b_n)$ . \checkmark

**Ejemplo 3.14.** Muestre que si  $a_n = O(b_n)$  y  $b_n = O(c_n)$ , entonces

$$a_n + b_n = O(c_n).$$

*Solución.* Por la definición 3.4 si  $a_n = O(b_n)$ , entonces existe  $k_1 > 0$  y  $n_1(k_1)$  tal que  $\left| \frac{a_n}{b_n} \right| \leq k_1$  para todo  $n \geq n_1(k_1)$  y como  $b_n = O(c_n)$ , entonces existe  $k_2 > 0$  y  $n_2(k_2)$  tal que  $\left| \frac{b_n}{c_n} \right| \leq k_2$  para todo  $n \geq n_2(k_2)$ , luego,

$$\left| \frac{a_n}{b_n} \right| \leq k_1 \Leftrightarrow |a_n| \leq k_1 |b_n| \quad \text{y} \quad \left| \frac{b_n}{c_n} \right| \leq k_2 \Leftrightarrow |b_n| \leq k_2 |c_n|,$$

tomando  $n \geq \max\{n_1, n_2\}$  y por la desigualdad triangular se tiene que:

$$\begin{aligned} & \underbrace{|a_n| + |b_n|}_{\geq |a_n + b_n|} \leq k_1 \underbrace{|b_n|}_{\leq k_2 |c_n|} + k_2 |c_n| \\ & |a_n + b_n| \leq k_1 k_2 |c_n| + k_2 |c_n| = k_2 |c_n| (k_1 + 1). \end{aligned}$$

Al dividir a ambos lados por  $|c_n|$  se obtiene que:

$$\left| \frac{a_n + b_n}{c_n} \right| \leq \underbrace{k_2(k_1 + 1)}_{\tilde{k}} \quad \forall n \geq n_0 = \max\{n_1, n_2\},$$

<sup>16</sup>Si  $a$  y  $b \in \mathbb{R}$ , entonces  $|a + b| \leq |a| + |b|$  (ver [56, p. 14]).

entonces existe  $\tilde{k} > 0$  y  $n_0(\tilde{k})$  tal que  $\left| \frac{a_n + b_n}{c_n} \right| \leq \tilde{k}$  para todo  $n \geq n_0(\tilde{k})$ , es decir,  $a_n + b_n = O(c_n)$ . ☑

**Teorema 3.4. Propiedades de los órdenes de magnitud.** Sean  $\{a_n\}_{n \geq 1}$ ,  $\{b_n\}_{n \geq 1}$ ,  $\{c_n\}_{n \geq 1}$  y  $\{d_n\}_{n \geq 1}$  sucesiones de números reales:

1. Si  $a_n = o(b_n)$ , entonces  $a_n = O(b_n)$ .
2. Si  $a_n = O(b_n)$  y  $c_n = O(d_n)$ , entonces:
  - a)  $a_n c_n = O(b_n d_n)$ .
  - b)  $a_n + c_n = O(\max\{|b_n|, |d_n|\})$ .
  - c)  $|a_n|^s = O(|b_n|^s), \forall s > 0$ .
3. Si  $a_n = o(b_n)$  y  $c_n = o(d_n)$ , entonces:
  - a)  $a_n c_n = o(b_n d_n)$ .
  - b)  $a_n + c_n = o(\max\{|b_n|, |d_n|\})$ .
  - c)  $|a_n|^s = o(|b_n|^s), \forall s > 0$ .
4. Si  $a_n = O(b_n)$  y  $c_n = o(d_n)$ , entonces  $a_n c_n = o(b_n d_n)$ .
5. Si  $a_n = O(b_n)$  y  $b_n = o(c_n)$ , entonces  $a_n = o(c_n)$ .

*Demostración.* Ver detalles de la prueba en [55]. ☑

**DEFINICIÓN 3.5.** Sean  $\{a_n\}_{n \geq 1}$  una sucesión de vectores en  $\mathbb{R}^m$ , y  $\{b_n\}_{n \geq 1}$  una sucesión de números reales, entonces:

$$a_n = O(b_n) \text{ si } \|a_n\| = O(b_n) \quad \text{y} \quad a_n = o(b_n) \text{ si } \|a_n\| = o(b_n),$$

donde  $\|a_n\|$  se define como en (B.1).

**Ejemplo 3.15.** En la tabla 3.10 se muestran algunas sucesiones de vectores en  $\mathbb{R}^3$ :

Tabla 3.10. Ejemplo de sucesiones de vectores en  $\mathbb{R}^3$

$a_n$	$n$					
	1	2	3	50	100	500
$\begin{pmatrix} \frac{1}{n} \\ \frac{1}{n^2} \\ 1 - \frac{1}{n} \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0,5 \\ 0,25 \\ 0,5 \end{pmatrix}$	$\begin{pmatrix} 0,33 \\ 0,11 \\ 0,667 \end{pmatrix}$	$\begin{pmatrix} 0,02 \\ 0,0004 \\ 0,98 \end{pmatrix}$	$\begin{pmatrix} 0,01 \\ 0,0001 \\ 0,99 \end{pmatrix}$	$\begin{pmatrix} 0,002 \\ 0,000004 \\ 0,998 \end{pmatrix}$
$\ a_n\ $	1,414	0,750	0,754	0,980	0,990	0,998
$\begin{pmatrix} \frac{1}{\sqrt{n}} \\ \frac{1}{n} \\ e^{-n} \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \\ 0,368 \end{pmatrix}$	$\begin{pmatrix} 0,707 \\ 0,5 \\ 0,135 \end{pmatrix}$	$\begin{pmatrix} 0,577 \\ 0,333 \\ 0,05 \end{pmatrix}$	$\begin{pmatrix} 0,141 \\ 0,02 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0,1 \\ 0,01 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0,045 \\ 0,002 \\ 0 \end{pmatrix}$
$\ a_n\ $	1,461	0,877	0,669	0,143	0,1	0,045

y en la figura 3.2, el respectivo gráfico.

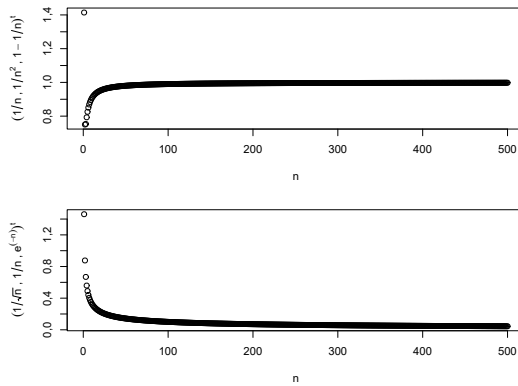


Figura 3.2. Ejemplos de sucesiones de normas de vectores en  $\mathbb{R}^3$

**Teorema 3.5.** Sean  $\{a_n\}_{n \geq 1}$  una sucesión de vectores en  $\mathbb{R}^m$  y  $\{b_n\}_{n \geq 1}$  una sucesión de números reales:

1.  $a_n = O(b_n)$  si y solo si  $a_{n_j} = O(b_n)$ , para  $j = 1, 2, \dots, m$ .
2.  $a_n = o(b_n)$  si y solo si  $a_{n_j} = o(b_n)$ , para  $j = 1, 2, \dots, m$ .

*Demostración.* Teniendo en cuenta que si  $a_n = (a_{n_1}, \dots, a_{n_m})^\top$ , entonces:

$$\frac{\|a_n\|}{|b_n|} = \frac{|(a_n^\top a_n)^{\frac{1}{2}}|}{|b_n|} = \frac{1}{(b_n^2)^{1/2}} \left( \sum_{j=1}^m a_{n_j}^2 \right)^{1/2} = \left( \frac{1}{b_n^2} \sum_{j=1}^m a_{n_j}^2 \right)^{1/2}. \quad (3.15)$$

1. De la definición 3.5 se sabe que  $a_n = O(b_n)$  si  $\|a_n\| = O(b_n)$  y por la definición 3.4 existe  $k > 0$  y  $n_0 = n_0(k)$  tal que  $\left| \frac{\|a_n\|}{b_n} \right| \leq k, \forall n \geq n_0$ .

( $\Rightarrow$ ) A partir de (3.15)

$$\frac{\|a_n\|}{|b_n|} = \frac{1}{|b_n|} \left( \sum_{j=1}^m a_{n_j}^2 \right)^{1/2} \leq k$$

nótese que:

$$\left| \frac{a_{n_j}}{b_n} \right| = \frac{(a_{n_j}^2)^{1/2}}{|b_n|} \leq \frac{1}{|b_n|} \left( \sum_{j=1}^m a_{n_j}^2 \right)^{1/2} \leq k,$$

luego, se tiene que  $a_{n_j} = O(b_n)$ .

- ( $\Leftarrow$ ) De la definición 3.4 se sabe que  $a_{n_j} = O(b_n), j = 1, 2, 3, \dots, m$ , es decir, para cada  $j$  existe  $k_j > 0$  y  $n_j = n_j(k_j)$  tal que  $\left| \frac{a_{n_j}}{b_n} \right| \leq k_j$

para todo  $n \geq n_j$ . Sea  $k = \max_{1 \leq j \leq m} \{k_j\}$  y  $n_0 = \max_{1 \leq j \leq m} \{n_j\}$ , entonces:

$$\frac{\|a_n\|}{|b_n|} = \left( \sum_{j=1}^m \frac{a_{n_j}^2}{b_n^2} \right)^{1/2} \leq \underbrace{\left( \sum_{j=1}^m k_j^2 \right)^{1/2}}_{\forall n \geq n_0} \leq m^{1/2}k,$$

luego,  $\|a_n\| = O(b_n)$ .

2. De la definición 3.5 se sabe que  $a_n = o(b_n)$  si  $\|a_n\| = o(b_n)$ , y por la definición 3.4  $\forall \varepsilon > 0, \exists n_0 = n_0(\varepsilon)$  tal que  $\frac{\|a_n\|}{|b_n|} < \varepsilon, \forall n \geq n_0$ .

( $\Rightarrow$ ) Sea  $a_n = (a_{n_1}, \dots, a_{n_m})^\top$ , se tiene que:

$$\frac{\|a_n\|}{|b_n|} < \varepsilon, \forall n \geq n_0,$$

y notando que:

$$\frac{|a_{n_j}|}{|b_n|} = \frac{(a_{n_j}^2)^{1/2}}{|b_n|} \leq \frac{1}{|b_n|} \left( \sum_{j=1}^m a_{n_j}^2 \right)^{1/2} = \frac{\|a_n\|}{|b_n|} < \varepsilon,$$

se tiene  $\forall n \geq n_0$  y para  $j = 1, 2, \dots, m$ . Por tanto,  $a_{n_j} = o(b_n)$ .

( $\Leftarrow$ ) De la definición 3.4 se sabe que  $a_{n_j} = o(b_n), j = 1, 2, 3, \dots, m$ , esto significa que  $\forall j, 1 \leq j \leq m, \forall \varepsilon_j > 0$  existe un número entero positivo  $n_j = n_j(\varepsilon_j)$  tal que  $\frac{|a_{n_j}|}{|b_n|} < \varepsilon_j, \forall n \geq n_j$ . Sea  $n_0 = \max_{1 \leq j \leq m} \{n_j\}$ , por la expresión (3.15):

$$\frac{\|a_n\|}{|b_n|} = \left( \frac{1}{b_n^2} \sum_{j=1}^m a_{n_j}^2 \right)^{1/2} = \left( \sum_{j=1}^m \frac{a_{n_j}^2}{b_n^2} \right)^{1/2} < \left( \sum_{j=1}^m \varepsilon_j^2 \right)^{1/2},$$

para todo  $n \geq n_0$ , luego  $\|a_n\| = o(b_n)$ . □

**Ejemplo 3.16.** Considere la sucesión definida por  $a_n = \frac{2}{n} + \frac{3}{n^2} - \frac{4}{n^3}$  y aproxime mediante las sucesiones:

(a)  $b_n = \frac{2}{n}$  (b)  $c_n = \frac{2}{n} + \frac{3}{n^2}$ .

*Solución.* En la tabla 3.11 se presentan las dos aproximaciones y las diferencias  $a_n - b_n$  y  $a_n - c_n$ .

Note que cuando  $n$  aumenta, las diferencias  $a_n - b_n$  y  $a_n - c_n$  se acercan a cero; sin embargo la segunda,  $a_n - c_n$ , llegó a cero más rápido que  $a_n - b_n$ . □

Tabla 3.11. Sucesión  $a_n$  aproximada por (a) y por (b)

	$n$						
	1	2	3	4	50	100	500
$a_n = \frac{2}{n} + \frac{3}{n^2} - \frac{4}{n^3}$	1	1,25	0,85185	0,625	0,04117	0,0203	0,00401
$b_n = \frac{2}{n}$	2	1	0,66667	0,5	0,04	0,02	0,004
$c_n = \frac{2}{n} + \frac{3}{n^2}$	5	1,75	1	0,6875	0,0412	0,0203	0,00401
$a_n - b_n$	-1	0,25	0,18519	0,125	0,00117	0,0003	0,00001
$a_n - c_n$	-4	-0,5	-0,14815	-0,0625	-0,00003	0	0

### 3.3.2. Órdenes de magnitud de funciones

Las diferencias en el comportamiento de las funciones para valores grandes del argumento  $x \in \mathbb{R}$  llevan a la noción de orden de magnitud. Debido a su gran importancia, esta temática es abordada en este apartado, aunque no está directamente relacionada con el concepto de la integral o de la derivada.

**DEFINICIÓN 3.6.** De acuerdo con [55, p. 19], dadas  $f$  y  $g$  funciones reales de variable real y  $x_0$  un número real. Se dice que:

1.  $f(x) = O(g(x))$  cuando  $x \rightarrow x_0$ , si existen números reales  $k > 0$  y  $\delta > 0$  tal que para todo  $x \in \mathbb{R}$  que satisface que  $|x - x_0| < \delta$ , entonces

$$\left| \frac{f(x)}{g(x)} \right| \leq k.$$

2.  $f(x) = O(g(x))$  cuando  $x \rightarrow +\infty(-\infty)$ , si existen números reales  $k > 0$  y  $M > 0$  ( $M < 0$ ) tal que para todo  $x > M$  ( $x < M$ ), entonces

$$\left| \frac{f(x)}{g(x)} \right| \leq k.$$

3.  $f(x) = o(g(x))$  cuando  $x \rightarrow x_0$ , si para todo  $\varepsilon > 0$  arbitrario existe un número real  $\delta > 0$  tal que para todo  $x \in \mathbb{R}$  que satisface que  $|x - x_0| < \delta$ , entonces

$$\left| \frac{f(x)}{g(x)} \right| < \varepsilon \quad \text{o} \quad \lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = 0.$$

4.  $f(x) = o(g(x))$  cuando  $x \rightarrow +\infty(-\infty)$ , si para todo  $\varepsilon > 0$  arbitrario existe un número real  $M > 0$  ( $M < 0$ ) tal que para todo  $x > M$  ( $x < M$ ), entonces

$$\left| \frac{f(x)}{g(x)} \right| < \varepsilon \quad \text{o} \quad \lim_{x \rightarrow \pm\infty} \frac{f(x)}{g(x)} = 0 \quad .$$

**Ejemplo 3.17.** Determine los órdenes de magnitud de las siguientes funciones:

1)  $7x^3 + 10x^2$ .

2)  $10(x - 1)^2 + 4(x - 1)$ .

*Solución.* Veamos:

1) La función  $7x^3 + 10x^2 = o(x)$ ,  $x \neq 0$ ,  $x \rightarrow 0$ , ya que

$$\lim_{x \rightarrow 0} \frac{7x^3 + 10x^2}{x} = \lim_{x \rightarrow 0} (7x^2 + 10x) = 0.$$

Entonces  $7x^3 + 10x^2 = o(x)$  cuando  $x \rightarrow 0$ .

2) La función  $10(x - 1)^2 + 4(x - 1) = o(1)$  cuando  $x \rightarrow 1$ , ya que

$$\lim_{x \rightarrow 1} 10(x - 1)^2 + 4(x - 1) = 0.$$

Entonces,

$$10(x-1)^2 + 4(x-1) = o(1) \quad \text{y} \quad 10(x-1)^2 - 4(x-1) = O(1). \quad \square$$

**Teorema 3.6.** De acuerdo con [55, p. 21], dadas  $f_1$ ,  $f_2$ ,  $g_1$ ,  $g_2$  y  $g$  funciones reales de variable real y  $x_0$  un número real:

1. Si  $f_1(x) = O(g(x))$  y  $f_2(x) = O(g(x))$  cuando  $x \rightarrow x_0(\pm\infty)$ , entonces

$$f_1(x) + f_2(x) = O(g(x)) \text{ cuando } x \rightarrow x_0(\pm\infty).$$

2. Si  $f_1(x) = o(g(x))$  y  $f_2(x) = o(g(x))$  cuando  $x \rightarrow x_0(\pm\infty)$ , entonces

$$f_1(x) + f_2(x) = o(g(x)) \text{ cuando } x \rightarrow x_0(\pm\infty).$$

3. Si  $f_1(x) = O(g_1(x))$  y  $f_2(x) = O(g_2(x))$  cuando  $x \rightarrow x_0(\pm\infty)$ , entonces

$$f_1(x)f_2(x) = O(g_1(x)g_2(x)) \text{ cuando } x \rightarrow x_0(\pm\infty).$$

4. Si  $f_1(x) = o(g_1(x))$  y  $f_2(x) = o(g_2(x))$  cuando  $x \rightarrow x_0(\pm\infty)$ , entonces

$$f_1(x)f_2(x) = o(g_1(x)g_2(x)) \text{ cuando } x \rightarrow x_0(\pm\infty).$$

*Demostración.* Ver detalles de la prueba en [55].

□

### 3.3.3. Órdenes de magnitud en probabilidad de sucesiones estocásticas. Notación: $O_p(\cdot)$ y $o_p(\cdot)$

**DEFINICIÓN 3.7.** De acuerdo con [55, p. 39], dada una sucesión  $\{X_n\}_{n \geq 1}$  de variables aleatorias definidas en un mismo espacio de probabilidad y una sucesión  $\{b_n\}_{n \geq 1}$  de números reales (o de variables aleatorias). Se dice que:

1.  $X_n = O_p(b_n)$  si para todo número real  $\eta > 0$  existe un número real positivo  $k = k(\eta)$  y un número entero positivo  $n_0 = n_0(\eta)$ , tal que:

$$P\left(\left|\frac{X_n}{b_n}\right| \geq k\right) \leq \eta, \quad \forall n \geq n_0, \quad (3.16)$$

es decir,  $\left|\frac{X_n}{b_n}\right|$  está acotada con probabilidad 1.

2.  $X_n = o_p(b_n)$  si para todo número real  $\varepsilon > 0$  y para todo número real  $\eta > 0$  existe un número entero positivo  $n_0 = n_0(\varepsilon, \eta)$ , tal que:

$$P\left(\left|\frac{X_n}{b_n}\right| \geq \varepsilon\right) \rightarrow 0 \text{ cuando } n \rightarrow \infty, \quad (3.17)$$

es decir,  $\left|\frac{X_n}{b_n}\right| \rightarrow 0$  con probabilidad 1 cuando  $n \rightarrow \infty$ .

**Nota 3.4.** Note que:

1.  $X_n = O_p(1)$  si  $\forall \eta > 0$ , existe  $0 < k = k(\eta) (< \infty)$ , tal que:

$$P(|X_n| \geq k) \leq \eta, \quad \text{para todo } n \geq n_0(\eta).$$

2.  $X_n = o_p(1)$  si para todo número real  $\varepsilon > 0$

$$P(|X_n| \geq \varepsilon) \rightarrow 0 \quad \text{cuando } n \rightarrow \infty,$$

o lo que es igual a:

$$\lim_{n \rightarrow \infty} P(|X_n| \geq \varepsilon) = 0.$$

**Ejemplo 3.18.** Considerando la información del ejemplo 3.3, sobre la muestra aleatoria de  $X_n \sim N(0, 1)$ ,  $X_n = O_p(1)$ , pero  $P(|X_n| > 1,5)$  no converge a cero<sup>17</sup>.

<sup>17</sup>Ver [www.stat.purdue.edu/~dasgupta/lst.pdf](http://www.stat.purdue.edu/~dasgupta/lst.pdf)

**Ejemplo 3.19.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid con  $\mathbb{E}(X_k) = \mu$  y  $\mathbb{V}(X_k) = \sigma^2 < \infty$ . Sea  $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ , muestre que:

$$\bar{X}_n - \mu = O_p\left(\frac{1}{\sqrt{n}}\right).$$

*Solución.* Por la desigualdad de Chebyshev se tiene que:

$$P(|Y - \mu_Y| \geq \varepsilon) \leq \frac{\mathbb{V}(Y)}{\varepsilon^2}, \quad \forall \varepsilon > 0. \quad (3.18)$$

Sea  $Y = \bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ , entonces,

$$\mu_Y = \mathbb{E}(Y) = \mathbb{E}(\bar{X}_n) = \mu \quad \text{y} \quad \mathbb{V}(Y) = \mathbb{V}(\bar{X}_n) = \frac{\sigma^2}{n},$$

luego, tomando  $\varepsilon = \frac{m}{\sqrt{n}}\sigma$  con  $m > \sqrt{n}$  en (3.18):

$$P\left(|\bar{X}_n - \mu| \geq \frac{m}{\sqrt{n}}\sigma\right) \leq \frac{\frac{\sigma^2}{n}}{\left(\frac{m}{\sqrt{n}}\sigma\right)^2} = \frac{\frac{\sigma^2}{n}}{\frac{m^2}{n}\sigma^2} = \frac{\cancel{\sigma^2}}{m^2 \cancel{\sigma^2}} = \frac{1}{m^2},$$

por lo tanto,

$$P\left(\left|\frac{\bar{X}_n - \mu}{1/\sqrt{n}}\right| \geq m\sigma\right) \leq \frac{1}{m^2},$$

sea  $\eta = \frac{1}{m^2}$ , entonces  $m = \frac{1}{\sqrt{\eta}}$ . Luego, para todo  $\eta > 0$ , se tiene que:

$$P\left(\left|\frac{\bar{X}_n - \mu}{\frac{1}{\sqrt{n}}}\right| \geq \sqrt{\frac{\sigma^2}{\eta}}\right) \leq \eta,$$

haciendo  $k = \sqrt{\frac{\sigma^2}{\eta}}$  se concluye que  $\bar{X}_n - \mu = O_p\left(\frac{1}{\sqrt{n}}\right)$ . □

**Ejemplo 3.20.** Considerando la información del ejemplo 3.4 sobre una muestra aleatoria de  $X \sim \exp(1)$ , muestre que  $X_n = o_p(\ln n)$ .

*Solución.* Por la expresión (3.2) se tiene que  $\forall \varepsilon > 0$  y  $n \geq 2$ :

$$P\left(\left|\frac{X_n}{\ln n}\right| \geq \varepsilon\right) = \frac{1}{n^\varepsilon} \quad \Rightarrow \quad \lim_{n \rightarrow \infty} P\left(\left|\frac{X_n}{\ln n}\right| \geq \varepsilon\right) = \lim_{n \rightarrow \infty} \frac{1}{n^\varepsilon} = 0.$$

Como se satisface (3.17), entonces  $X_n = o_p(\ln n)$ . □

**Ejemplo 3.21.** Sea  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias reales tales que  $\mathbb{E}(X_n^2) < \infty$ ,  $n \geq 1$ . Muestre que  $X_n = O_p\left([\mathbb{E}(X_n^2)]^{\frac{1}{2}}\right)$ .

*Solución.* Por la desigualdad de Chebyshev se tiene que:

$$P\left(\left|\frac{X_n}{\sqrt{\mathbb{E}(X_n^2)}}\right| \geq k\right) \leq \frac{1}{k^2} \mathbb{E}\left(\left|\frac{X_n}{\sqrt{\mathbb{E}(X_n^2)}}\right|^2\right) = \frac{\mathbb{E}(|X_n|^2)}{k^2 \mathbb{E}(X_n^2)} = \frac{1}{k^2} \quad \forall k > 0,$$

haciendo el cambio de variable  $\eta = \frac{1}{k^2}$ , entonces  $k = \frac{1}{\sqrt{\eta}}$ . Luego, se satisface (3.16), para todo  $\eta > 0$ , existe  $k = \frac{1}{\sqrt{\eta}} = k(\eta) > 0$  y  $n_0 = n_0(\eta)$  tal que

$$P\left(\left|\frac{X_n}{\sqrt{\mathbb{E}(X_n^2)}}\right| \geq k\right) \leq \eta \text{ para todo } n \geq n_0, \text{ es decir, } X_n = O_p\left([\mathbb{E}(X_n^2)]^{\frac{1}{2}}\right). \quad \square$$

**Ejemplo 3.22.** Sea  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias y  $\{a_n\}_{n \geq 1}$  una sucesión de números reales, tales que  $\mu_n = \mathbb{E}(X_n) = O(a_n)$  y  $\sigma_n^2 = \mathbb{V}(X_n) = O(a_n^2)$ . Muestre que  $X_n = O_p(a_n)$ .

*Solución.* Del enunciado y por la definición 3.4 se tiene que existe:

- Un  $k_1 > 0$  y  $n_0 = n_0(k_1)$  tal que  $\left|\frac{\mathbb{E}(X_n)}{a_n}\right| \leq k_1$  para todo  $n \geq n_0(k_1)$ .
- Un  $k_2 > 0$  y  $n_1 = n_1(k_2)$  tal que  $\left|\frac{\mathbb{V}(X_n)}{a_n^2}\right| \leq k_2$  para todo  $n \geq n_1(k_2)$ .

Luego, por la desigualdad de Chebyshev,

$$\begin{aligned} P\left(\left|\frac{X_n}{a_n}\right| \geq k\right) &\leq \frac{1}{k^2} \mathbb{E}\left(\left|\frac{X_n}{a_n}\right|^2\right) = \frac{\mathbb{V}\left(\frac{X_n}{a_n}\right) + \left[\mathbb{E}\left(\frac{X_n}{a_n}\right)\right]^2}{k^2} \quad \forall k > 0 \\ &= \frac{1}{k^2} \left[ \frac{\mathbb{V}(X_n)}{a_n^2} + \frac{[\mathbb{E}(X_n)]^2}{a_n^2} \right] = \frac{1}{k^2} \underbrace{\frac{\mathbb{V}(X_n)}{a_n^2}}_{\leq k_2} + \frac{1}{k^2} \underbrace{\frac{[\mathbb{E}(X_n)]^2}{a_n^2}}_{\leq k_1^2} \\ &= \frac{1}{k^2} \underbrace{\left| \frac{\mathbb{V}(X_n)}{a_n^2} \right|}_{\leq k_2} + \frac{1}{k^2} \underbrace{\left| \frac{[\mathbb{E}(X_n)]^2}{a_n^2} \right|}_{\leq k_1^2} \\ &\leq \underbrace{\frac{1}{k^2} (k_2 + k_1^2)}_{\eta} \quad \forall n \geq \max\{n_0, n_1\}. \end{aligned}$$

Entonces, para todo  $\eta > 0$  existe  $k = \sqrt{\frac{k_2 + k_1^2}{\eta}} = k(\eta, k_1, k_2) > 0$ , tal que  $P\left(\left|\frac{X_n}{a_n}\right| \geq k\right) \leq \eta$  para todo  $n \geq \max\{n_0, n_1\}$ , por tanto,  $X_n = O_p(a_n)$ .  $\square$

**Nota 3.5.** Según [52, p. 54], para cualquier sucesión  $\{c_n\}$  de números reales, se tiene que:

- a) Si  $X_n = o_p(b_n)$ , entonces  $c_n X_n = c_n o_p(b_n) = o_p(c_n b_n)$
- b) Si  $X_n = O_p(b_n)$ , entonces  $c_n X_n = c_n O_p(b_n) = O_p(c_n b_n)$ .

**Teorema 3.7.** Propiedades de los órdenes de magnitud en probabilidad. De acuerdo con [55, p. 40], dadas  $\{X_n\}_{n \geq 1}$ ,  $\{Y_n\}_{n \geq 1}$  sucesiones de variables aleatorias y  $\{a_n\}_{n \geq 1}$ ,  $\{b_n\}_{n \geq 1}$  sucesiones de números reales (o variables aleatorias):

- 1. Si  $X_n = o_p(a_n)$ , entonces  $X_n = O_p(a_n)$ .
- 2. Si  $X_n = O_p(a_n)$  y  $Y_n = O_p(b_n)$ , entonces:

- a)  $X_n Y_n = O_p(a_n b_n)$ .
- b)  $X_n + Y_n = O_p(\max\{|a_n|, |b_n|\})$ .
- c)  $|X_n|^s = O_p(|a_n|^s)$ ,  $\forall s > 0$ .

- 3. Si  $X_n = o_p(a_n)$  y  $Y_n = o_p(b_n)$ , entonces:

- a)  $X_n Y_n = o_p(a_n b_n)$ .
- b)  $X_n + Y_n = o_p(\max\{|a_n|, |b_n|\})$ .
- c)  $|X_n|^s = o_p(|a_n|^s)$ ,  $\forall s > 0$ .

- 4. Si  $X_n = O_p(a_n)$  y  $Y_n = O_p(b_n)$ , entonces  $X_n Y_n = O_p(a_n b_n)$ .

*Demostración.* Ver detalles de la prueba en [55, p. 41].  $\square$

**DEFINICIÓN 3.8.** De acuerdo con [55, p. 42], dada una sucesión  $\{\mathbf{X}_n\}_{n \geq 1} = \{(X_{n1}, \dots, X_{nm})^\top\}_{n \geq 1}$  de vectores aleatorios  $m$ -dimensionales ( $m \geq 2$ ) y una sucesión  $\{b_n\}_{n \geq 1}$  de números reales, se dice que:

$$\mathbf{X}_n = O_p(b_n) \text{ si } \|\mathbf{X}_n\| = O_p(b_n) \quad \text{y} \quad \mathbf{X}_n = o_p(b_n) \text{ si } \|\mathbf{X}_n\| = o_p(b_n),$$

donde  $\|\mathbf{X}_n\|$  se define como en (B.1).

**Teorema 3.8.** De acuerdo con [55, p. 42], dada una sucesión  $\{\mathbf{X}_n\}_{n \geq 1}$  de vectores aleatorios  $m$ -dimensionales y una sucesión  $\{b_n\}_{n \geq 1}$  de números reales, entonces:

- 1.  $\mathbf{X}_n = O_p(b_n)$  si y solo si  $X_{nj} = O_p(b_n)$ , para  $j = 1, \dots, m$ .
- 2.  $\mathbf{X}_n = o_p(b_n)$  si y solo si  $X_{nj} = o_p(b_n)$ , para  $j = 1, \dots, m$ .

*Demostración.* Se presenta la prueba del primer ítem.

( $\Leftarrow$ ) De la definición 3.8 se sabe que  $X_n = O_p(b_n)$  si  $\|X_n\| = O_p(b_n)$  y de la definición 3.7 se sabe que  $\|X_n\| = O_p(b_n)$  si para todo número real  $\eta > 0$  existe un número real positivo  $k = k(\eta)$  y un número entero positivo  $n_0 = n_0(\eta)$ , tal que:

$$P\left(\frac{\|X_n\|}{|b_n|} \geq k\right) \leq \eta, \quad \forall n \geq n_0.$$

Luego, para  $j$ ,  $1 \leq j \leq m$ , se tiene que:

$$\left\{\frac{|X_{nj}|}{|b_n|} \geq k\right\} = \left\{\frac{X_{nj}^2}{b_n^2} \geq k^2\right\} \subseteq \left\{\frac{\|X_n\|^2}{b_n^2} \geq k^2\right\} = \left\{\frac{\|X_n\|}{|b_n|} \geq k\right\}.$$

Tomando probabilidad se obtiene que:

$$P\left(\frac{|X_{nj}|}{|b_n|} \geq k\right) \leq P\left(\frac{\|X_n\|}{|b_n|} \geq k\right) \leq \eta, \quad \forall n \geq n_0,$$

por tanto,  $X_{nj} = O_p(b_n)$ .

( $\Rightarrow$ ) De la definición 3.7 se sabe que  $X_{nj} = O_p(b_n)$  si para todo número real  $\eta > 0$  existe un número real positivo  $k_j = k_j(\eta)$  y un número entero positivo  $n_j = n_j(\eta)$ , tal que:

$$P\left(\left|\frac{X_{nj}}{b_n}\right| \geq k_j\right) \leq \frac{\eta}{m}, \quad \forall n \geq n_j.$$

Sea  $k = \max\{k_1, k_2, \dots, k_m\}$ ,  $\tilde{k} = \sqrt{mk}$  y  $n_0 = \max\{n_1, n_2, \dots, n_m\}$ , entonces  $\forall n \geq n_0$ :

$$\begin{aligned} P\left(\frac{\|X_n\|}{|b_n|} \geq \tilde{k}\right) &= P\left(\left[\sum_{j=1}^m \frac{X_{nj}^2}{b_n^2}\right]^{\frac{1}{2}} \geq \sqrt{mk}\right) = P\left(\sum_{j=1}^m \frac{X_{nj}^2}{b_n^2} \geq mk^2\right) \\ &= P\left(\sum_{j=1}^m X_{nj}^2 \geq mk^2 b_n^2\right) \leq \sum_{j=1}^m P\left(X_{nj}^2 \geq \frac{mk^2 b_n^2}{m}\right) \\ &\quad \text{ver lema B.1} \\ &\quad \text{en apéndice B} \\ &\leq \sum_{j=1}^m P(X_{nj}^2 \geq k_j^2 b_n^2) = \sum_{j=1}^m \underbrace{P(|X_{nj}| \geq k_j |b_n|)}_{\leq \frac{\eta}{m}} \leq \sum_{j=1}^m \frac{\eta}{m} = \eta \end{aligned}$$

por tanto,  $P\left(\frac{\|X_n\|}{|b_n|} \geq \tilde{k}\right) \leq \eta, \quad \forall n_0 \geq \max\{n_1, n_2, \dots, n_m\}. \quad \checkmark$

## 3.4. Convergencia estocástica

### 3.4.1. Tipos de convergencia estocástica

Sean  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias y  $X$  una variable aleatoria definida en el mismo espacio de probabilidad.

**DEFINICIÓN 3.9.** De acuerdo con [11, p. 105], una sucesión  $\{X_n\}_{n \geq 1}$  de variables aleatorias reales **converge en probabilidad** para una variable aleatoria  $X$  (puede ser degenerada<sup>18</sup>) si, para todo  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0 \quad \text{es decir} \quad X_n - X = o_p(1).$$

$$\text{Notación: } X_n \xrightarrow{P} X.$$

Esto significa que la diferencia entre  $X_n$  y  $X$  es pequeña con una alta probabilidad, para  $n$  grande. No implica que  $X_n(\omega) - X(\omega)$  sea pequeña para cada  $\omega$  aún cuando  $n$  sea suficientemente grande.

**Ejemplo 3.23.** Sea  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias reales tales que  $\mathbb{E}(X_n) = \mu_n \xrightarrow{n \rightarrow \infty} c$  y  $\mathbb{V}(X_n) = \sigma_n^2 \xrightarrow{n \rightarrow \infty} 0$ , muestre que  $X_n \xrightarrow{P} c$ .

*Solución.*  $\forall \varepsilon > 0$ , se sabe (por la desigualdad de Chebyshev) que:

$$\begin{aligned} 0 \leq P(|X_n - c| \geq \varepsilon) &\leq \frac{\mathbb{E}([X_n - c]^2)}{\varepsilon^2} = \frac{\mathbb{E}([X_n - \mu_n + \mu_n - c]^2)}{\varepsilon^2} \\ &= \frac{\mathbb{E}([X_n - \mu_n]^2) + \mathbb{E}([\mu_n - c]^2) + 2[\mu_n - c]\mathbb{E}([X_n - \mu_n])}{\varepsilon^2} \\ &= \frac{\sigma_n^2}{\varepsilon^2} + \frac{[\mu_n - c]^2}{\varepsilon^2}, \end{aligned}$$

tomando límite:

$$\begin{aligned} 0 \leq \lim_{n \rightarrow \infty} P(|X_n - c| \geq \varepsilon) &\leq \frac{1}{\varepsilon^2} \lim_{n \rightarrow \infty} \sigma_n^2 + \frac{1}{\varepsilon^2} \lim_{n \rightarrow \infty} [\mu_n - c]^2 \\ &= 0 + 0 = 0, \end{aligned}$$

por tanto,  $X_n - c = o_p(1)$ , es decir,  $X_n \xrightarrow{P} c$ . ☑

<sup>18</sup>La fda de una variable aleatoria  $X$  degenerada en  $c$  está dada por:

$$F_X(x) = I_{[c, \infty)}(x) = \begin{cases} 1 & \text{si } x \geq c \\ 0 & \text{si } x < c \end{cases}$$

**DEFINICIÓN 3.10.** Según [55, p. 48], una sucesión de vectores aleatorios  $m$ -dimensionales  $\{\mathbf{X}_n\}_{n \geq 1} = \{(X_{n1}, \dots, X_{nm})^\top\}_{n \geq 1}$  **converge en probabilidad** para un vector aleatorio  $\mathbf{X}_0 = (X_{01}, \dots, X_{0m})^\top$  (puede ser degenerado), si la sucesión de variables aleatorias  $\{\|\mathbf{X}_n - \mathbf{X}_0\|\}_{n \geq 1}$  converge en probabilidad para 0.

$$\text{Notación: } \mathbf{X}_n \xrightarrow{P} \mathbf{X}_0 \quad \text{ó} \quad \|\mathbf{X}_n - \mathbf{X}_0\| = o_p(1).$$

**Teorema 3.9.** De acuerdo con [55, p. 48], dada una sucesión  $\{\mathbf{X}_n\}_{n \geq 1}$  de vectores aleatorios y  $\mathbf{X}_0$  un vector aleatorio. Entonces,

$$\mathbf{X}_n \xrightarrow{P} \mathbf{X}_0 \quad \text{si y solo si} \quad X_{nj} \xrightarrow{P} X_{0j}, \quad \text{para } j = 1, \dots, m.$$

*Demostración.* De la definición 3.10 se tiene que:

$$\begin{aligned} \mathbf{X}_n \xrightarrow{P} \mathbf{X}_0 &\iff \|\mathbf{X}_n - \mathbf{X}_0\| = o_p(1) \iff \underbrace{X_{nj} - X_{0j} = o_p(1)}_{\text{Teorema 3.8}}, \\ &\iff \underbrace{X_{nj} \xrightarrow{P} X_{0j}}_{\text{Defin. 3.9}}, \quad 1 \leq j \leq m. \quad \checkmark \end{aligned}$$

**Ejemplo 3.24.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid con  $\mathbb{E}(X_1) = \mu$ ,  $\mathbb{V}(X_1) = \sigma^2$  y  $\mathbb{E}[(X - \mu)^4] < \infty$ . Sean  $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$  y  $S_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2$ , muestre que:

$$(\bar{X}_n, S_n^2) \xrightarrow[n \rightarrow \infty]{P} (\mu, \sigma^2). \tag{3.19}$$

*Solución.* Reescribiendo  $S_n^2$  se tiene que:

$$\begin{aligned} S_n^2 &= \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2 = \frac{1}{n-1} \sum_{j=1}^n [(X_j - \mu) - (\bar{X}_n - \mu)]^2 \\ &= \frac{1}{n-1} \sum_{j=1}^n \left[ (X_j - \mu)^2 - 2(X_j - \mu)(\bar{X}_n - \mu) + (\bar{X}_n - \mu)^2 \right]. \end{aligned}$$

Distribuyendo la sumatoria se puede expresar como:

$$\begin{aligned} S_n^2 &= \frac{1}{n-1} \sum_{j=1}^n (X_j - \mu)^2 - 2 \frac{(\bar{X}_n - \mu)}{n-1} \sum_{j=1}^n (X_j - \mu) + \frac{n}{n-1} (\bar{X}_n - \mu)^2 \\ &= \frac{1}{n-1} \sum_{j=1}^n (X_j - \mu)^2 - 2 \frac{(\bar{X}_n - \mu)}{n-1} (n\bar{X}_n - n\mu) + \frac{n}{n-1} (\bar{X}_n - \mu)^2 \end{aligned}$$

$$= \frac{1}{n-1} \sum_{j=1}^n (X_j - \mu)^2 - \frac{n}{n-1} (\bar{X}_n - \mu)^2.$$

Empleando los órdenes de magnitud dados en la tabla 3.12:

Tabla 3.12. Órdenes de magnitud de números  $a_{nj}$

$j$	$a_{nj}$	$\lim_{n \rightarrow \infty} a_{nj}$	orden de magnitud
1	$\frac{1}{\sqrt{n}}$	0	$o(1)$
2	$\frac{1}{n}$	0	$o(1)$
3	$\frac{1}{n-1}$	0	$o(1)$
4	$\frac{n}{n-1}$	1	$O(1)$

- Sea  $T_n = \bar{X}_n - \mu$ , del ejemplo 3.19 se sabe que  $T_n = O_p\left(\frac{1}{\sqrt{n}}\right) = O_p(a_{n1})$ , por tanto,

$$\bar{X}_n - \mu = O_p(a_{n1}) = O_p(o(1)) \underbrace{=}_{\text{Ejercicio b)}} o_p(1).$$

De otro lado, por el teorema 3.7 se tiene que:

$$\underbrace{|\bar{X}_n - \mu|^2}_{|T_n|^2} = O_p\left(\frac{1}{n}\right) = O_p(a_{n2}) = O_p(o(1)) \underbrace{=}_{\text{Ejercicio b)}} o_p(1).$$

- Sea  $U_n = a_{n4}|T_n|^2 = \frac{n}{n-1}(\bar{X}_n - \mu)^2$ , usando el ítem a) de la nota 3.5:

$$\underbrace{a_{n4}|T_n|^2}_{U_n} = a_{n4}O_p(o(1)) = O_p(a_{n4}o(1)) = O_p(O(1)o(1))$$

$$U_n = O_p(o(1)) \underbrace{=}_{\text{Ejercicio b)}} o_p(1), \text{ es decir, } n \frac{(\bar{X}_n - \mu)^2}{n-1} = o_p(1).$$

Por otro lado, las variables aleatorias  $(X_1 - \mu)^2, (X_2 - \mu)^2, \dots, (X_n - \mu)^2$  son iid con media  $\mathbb{E}[(X_1 - \mu)^2] = \sigma^2$  y varianza:

$$\mathbb{V}[(X_1 - \mu)^2] = \mathbb{E}\left[\{(X_1 - \mu)^2\}^2\right] - [\mathbb{E}(X_1 - \mu)^2]^2 = \mathbb{E}[(X_1 - \mu)^4] - \sigma^4 < \infty.$$

- Sea  $Y_n = \frac{1}{n} \sum_{j=1}^n (X_j - \mu)^2 - \sigma^2$ , del ejemplo 3.19, se sabe que:

$$Y_n = \frac{1}{n} \sum_{j=1}^n (X_j - \mu)^2 - \sigma^2 = O_p\left(\frac{1}{\sqrt{n}}\right) = O_p(a_{n1}).$$

- Sea

$$\begin{aligned} R_n &= a_{n4}Y_n + \frac{\sigma^2}{n-1} = \frac{\cancel{n}}{n-1} \frac{1}{\cancel{n}} \sum_{j=1}^n (X_j - \mu)^2 - \frac{n}{n-1} \sigma^2 + \frac{\sigma^2}{n-1} \\ &= \frac{1}{n-1} \sum_{j=1}^n (X_j - \mu)^2 - \sigma^2, \end{aligned}$$

se tiene que:

$$\begin{aligned} R_n &= a_{n4}Y_n + \underbrace{\frac{\sigma^2}{n-1}}_{19} = a_{n4}O_p(a_{n1}) + o(1) = O_p(a_{n4} a_{n1}) + o(1) \\ &= O_p(O(1)o(1)) + o(1) = O_p(o(1)) + o(1) = o_p(1) + o(1) = o_p(1). \end{aligned}$$

Entonces,

$$\begin{aligned} S_n^2 - \sigma^2 &= \frac{1}{n-1} \sum_{j=1}^n (X_j - \mu)^2 - \frac{n}{n-1} (\bar{X}_n - \mu)^2 - \sigma^2 \\ &= R_n - U_n = o_p(1) - o_p(1) = o_p(1), \end{aligned}$$

por tanto,

$$S_n^2 - \sigma^2 = o_p(1) \Rightarrow S_n^2 \xrightarrow{P} \sigma^2,$$

concluyendo que:

$$(\bar{X}_n, S_n^2) \xrightarrow{P} (\mu, \sigma^2). \quad \checkmark$$

**Ejemplo 3.25.** Sea  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias *iid* uniformes en el intervalo  $(0,1)$  y sea  $Y_n = \max\{X_1, X_2, \dots, X_n\}$ . Para verificar numéricamente que a medida que  $n$  aumenta,  $Y_n$  se acerca a 1, se emplea el siguiente código en R:

```
remove(list=ls())
library(plotly)
muestra.n=NULL
```

---

<sup>19</sup>  $\lim_{n \rightarrow \infty} \frac{\sigma^2}{n-1} = \sigma^2 \lim_{n \rightarrow \infty} \frac{1}{n-1} = \sigma^2 \times 0 = 0$ .

```

for(i in 1:100) {
  Xn=runif(i,0,1)
  Yn=max(Xn)
  muestra.n=rbind(muestra.n,cbind(i,Yn))
}
muestra.n=data.frame(muestra.n)
p1 <- muestra.n %>%
  plot_ly(
    x = ~ i,
    y = ~ Yn,
    frame = ~ i,
    type = "scatter",
    mode = "markers",
    showlegend = FALSE
  ) %>% add_trace(y = ~1,marker=list(color="red"))
p1

```

**DEFINICIÓN 3.11.** Según [11, p. 110] y [16, p. 297], una sucesión de variables aleatorias  $\{X_n\}_{n \geq 1}$  **converge casi siempre** (o converge en casi toda parte) o fuertemente para una variable aleatoria  $X$  si  $X_n(\omega) \rightarrow X(\omega)$  para todo  $\omega$ , excepto para aquellos  $\omega$  que pertenecen a un conjunto nulo, es decir,

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

Notación:  $X_n \xrightarrow{c.s.} X$ .

**Nota 3.6.** En [20, p. 234] se presenta la definición 3.11 como sigue: una sucesión de variables aleatorias  $\{X_n\}_{n \geq 1}$  **converge casi siempre** (o converge en casi toda parte) o fuertemente para una variable aleatoria  $X$  si, para cada  $\epsilon > 0$ ,

$$P\left(\lim_{n \rightarrow \infty} |X_n - X| < \epsilon\right) = 1.$$

**Nota 3.7.** De acuerdo con [82, p. 38]:

$$X_n - X \xrightarrow{P} 0 \iff |X_n - X| = o_p(1)$$

$$X_n - X \xrightarrow{c.s.} 0 \iff |X_n - X| = o(1) \text{ casi siempre}^{20}.$$

<sup>20</sup>Para una sucesión de variables aleatorias  $\{X_n\}_{n \geq 1}$ , si para cada  $\eta > 0$  y  $\epsilon > 0$ , existe un entero positivo  $n(\eta, \epsilon)$ , tal que:

$$P\{|X_n| > \eta \text{ para algún } N \geq n\} < \epsilon, \quad n \geq n(\eta, \epsilon),$$

entonces se dice que  $X_n = o(1)$  casi siempre (ver [82, p. 38]).

**DEFINICIÓN 3.12.** Según [82, p. 44], una sucesión de variables aleatorias  $\{X_n\}_{n \geq 1}$  **converge completamente** para una variable aleatoria  $X$  (posiblemente degenerada) si para todo  $\varepsilon > 0$ ,

$$\sum_{n=1}^{\infty} P(|X_n - X| \geq \varepsilon) < \infty.$$

Notación:  $X_n \xrightarrow{C} X$ .

**Nota 3.8.** Sea  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias:

- Si  $X_n \xrightarrow{C} X$ , entonces  $X_n \xrightarrow{c.s.} X$ .
- Si  $X_n \xrightarrow{c.s.} X$ , entonces  $X_n \xrightarrow{P} X$ .

**DEFINICIÓN 3.13.** De acuerdo con [55, p. 51], una sucesión de vectores aleatorios  $\{X_n\}_{n \geq 1} = \{(X_{n1}, \dots, X_{nm})^T\}_{n \geq 1}$  **converge casi siempre** (o converge en casi toda parte) para un vector aleatorio  $X_0$  (posiblemente degenerado) si la sucesión de variables aleatorias  $\|X_n - X_0\|_{n \geq 1}$  converge casi siempre para cero. Notación:  $X_n \xrightarrow{c.s.} X_0$ .

**Nota 3.9.** Según [55, p. 52]:

$$X_n \xrightarrow{c.s.} X_0 \Leftrightarrow X_{nj} \xrightarrow{c.s.} X_{0j}, \quad \text{para } 1 \leq j \leq m.$$

**DEFINICIÓN 3.14.** De acuerdo con [11, p. 117], dadas  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias y  $X$  una variable aleatoria, y  $r$  un número real positivo tal que  $\mathbb{E}(|X_n|^r) < \infty$ , para todo  $n \geq 1$ . Se dice que  $\{X_n\}_{n \geq 1}$  **converge en media de orden  $r$**  para  $X$  si  $\mathbb{E}(|X_n - X|^r) \xrightarrow[n \rightarrow \infty]{} 0$ . Notación:  $X_n \xrightarrow{m.r.} X$ .

**Ejemplo 3.26.** De acuerdo con [55, p. 53], si  $\{X_n\}_{n \geq 1}$  es una sucesión de variables aleatorias *iid* con  $\mathbb{E}(X_1) = \mu$  y  $\mathbb{V}(X_1) = \sigma^2 < \infty$ . Entonces,

$$\mathbb{E}(|\bar{X}_n - \mu|^2) = \mathbb{V}(\bar{X}_n) = \frac{\sigma^2}{n} \xrightarrow[n \rightarrow \infty]{} 0,$$

por tanto,

$$\bar{X}_n \xrightarrow{m.2} \mu.$$

**Nota 3.10.** Según [82, p. 44] y [55, p. 53], dada una sucesión de variables aleatorias  $\{X_n\}_{n \geq 1}$  y  $r$  un número real positivo.

- Si  $X_n \xrightarrow{m.r.} X$ , entonces  $X_n \xrightarrow{P} X$ .
- Si  $\sum_{n=1}^{\infty} \mathbb{E}(|X_n - X|^r) < \infty$ , entonces  $X_n \xrightarrow{c.s.} X$ .
- Si  $X_n \xrightarrow{m.r.} X$ , entonces  $X_n \xrightarrow{m.s} X$ , para todo  $0 < s \leq r$ .

### 3.4.2. Convergencia en distribución o convergencia en ley

**DEFINICIÓN 3.15.** De acuerdo con [43, p. 233], dadas las variables aleatorias  $X, X_1, X_2, \dots$  con fda  $F, F_1, F_2, F_3, \dots$ , respectivamente. Se dice que la sucesión  $\{X_n\}_{n \geq 1}$  **converge en distribución (o en ley<sup>21</sup>)** para  $X$  si, para todo punto  $x$  de continuidad de  $F$ ,

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x).$$

Notación:  $X_n \xrightarrow{D} X$  ó  $X_n \xrightarrow{L} X$ .

**Ejemplo 3.27.** Según [14, p. 477], dada una sucesión de variables aleatorias  $\{X_n\}_{n \geq 1}$  iid con  $X_1 \sim \exp(\lambda)$ . Si  $Y_n = \min\{X_1, X_2, X_3, \dots, X_n\}$ , para todo  $n \geq 1$ . Mostrar que  $Z_n \xrightarrow{D} Y_1$  con  $Z_n = nY_n$ .

*Solución.* De estadísticas de orden, expresión (3.9), se sabe que:

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) [F_X(x)]^{j-1} [1 - F_X(x)]^{n-j};$$

tomando  $j = 1$  y usando las expresiones dadas en (3.1), se tiene que:

$$\begin{aligned} f_{X_{(1)}}(x) &= \frac{n!}{(1-1)!(n-1)!} f_X(x) [F_X(x)]^{1-1} [1 - F_X(x)]^{n-1} \\ &= n \underbrace{f_X(x)}_{\lambda e^{-\lambda x}} \underbrace{[1 - F_X(x)]^{n-1}}_{e^{-\lambda x}} = n\lambda e^{-\lambda x} (e^{-\lambda x})^{n-1} = (n\lambda) e^{-\lambda x - \lambda x n + \lambda x} \\ &= (n\lambda) e^{-n\lambda x}, \end{aligned}$$

luego  $f_{X_{(1)}}(x)$  coincide con la fdp  $\exp(n\lambda)$ , es decir,  $Y_n \sim \exp(n\lambda)$ .

La función característica de  $Y_n$  (ver tabla 1.13) está dada por:

$$\psi_{Y_n}(t) = \frac{1}{1 - \frac{it}{n\lambda}},$$

luego la función característica de  $Z_n = nY_n$  es dada por:

$$\psi_{Z_n}(t) = \psi_{nY_n}(t) = \psi_{Y_n}(nt) = \frac{1}{1 - \frac{it}{\lambda}} = \frac{1}{1 - \frac{it}{\lambda}},$$

<sup>21</sup>Según [11, p. 114], una sucesión de fda  $\{F_n(x)\}$ , no necesariamente de variables aleatorias, y si  $F_n(x) \rightarrow F(x)$  para todo punto  $x$  de continuidad de  $F$ , entonces  $\{F_n\}$  se dice que converge débilmente o en ley para  $F$ . En general,  $F(x)$  puede no ser una fda, pero si  $F_n(x)$  es una fda de una variable aleatoria  $X_n$  y  $F(x)$  es una fda de una variable aleatoria  $X$ , entonces la sucesión  $X_n$  se dice que converge en distribución o en ley o débilmente a  $X$ .

que corresponde a la distribución  $\exp(\lambda)$ . Aquí se usó la propiedad 5 de la función característica que establece que  $\psi_{aX}(t) = \psi_X(at)$ .

Aplicando límite:

$$\lim_{n \rightarrow \infty} F_{Z_n}(x) = \lim_{n \rightarrow \infty} (1 - e^{-\lambda x}) = (1 - e^{-\lambda x}) = 1 - e^{-(1 \times \lambda)x} = F_{Y_1}(x),$$

concluyendo que  $Z_n \xrightarrow{D} Y_1$ . ✓

**Teorema 3.10.** (Lema de Helly-Bray). De acuerdo con [82, p. 102], dadas  $F, F_1, F_2, F_3, \dots$  fda. Si  $\{F_n\}$  converge en distribución (o débilmente) para  $F$ , entonces

$$\int g(x) d[F_n(x)] \xrightarrow{n \rightarrow \infty} \int g(x) d[F(x)],$$

para toda función  $g$  continua y acotada ( $g : \mathbb{R} \rightarrow \mathbb{R}$ ).

*Demostración.* Ver detalles de la prueba en [82, p. 103] o [66, p. 178]. ✓

**Teorema 3.11.** De acuerdo con [55, p. 55], dadas  $X, X_1, X_2, \dots$  variables aleatorias con fda  $F, F_1, F_2, \dots$  y funciones características  $\psi, \psi_1, \psi_2, \dots$ , respectivamente. Entonces para cada  $t \in \mathbb{R}$  se tiene que:

$$X_n \xrightarrow{D} X \iff \psi_{X_n}(t) \xrightarrow{n \rightarrow \infty} \psi_X(t).$$

*Demostración.* Por la definición 3.15 se sabe que  $X_n \xrightarrow{D} X$  si

$$F_{X_n}(x) \xrightarrow{n \rightarrow \infty} F_X(x).$$

( $\Rightarrow$ ) Por la definición de función característica, definición 1.26, se tiene que para todo  $t \in \mathbb{R}$ :

$$\begin{aligned} \psi_{X_n}(t) &= \mathbb{E} \left( e^{itX_n} \right) = \mathbb{E} [\cos(tX_n)] + i \mathbb{E} [\sin(tX_n)] \\ &= \int \cos(tx) d[F_{X_n}(x)] + i \int \sin(tx) d[F_{X_n}(x)] \\ &\xrightarrow{n \rightarrow \infty} \int \cos(tx) d[F_X(x)] + i \int \sin(tx) d[F_X(x)] \\ &= \mathbb{E} [\cos(tX)] + i \mathbb{E} [\sin(tX)] = \psi_X(t). \end{aligned}$$

En el penúltimo paso, se usó Helly-Bray ya que las funciones seno y coseno son continuas y acotadas. Por tanto,  $\psi_{X_n}(t) \xrightarrow{n \rightarrow \infty} \psi_X(t)$ .

( $\Leftarrow$ ) Se tiene que para todo  $t \in \mathbb{R}$

$$\begin{array}{ccc} \underbrace{\psi_{X_n}(t)} & \xrightarrow{n \rightarrow \infty} & \underbrace{\psi_X(t)} \\ \downarrow & & \downarrow \\ F_{X_n} & & F_X \end{array}$$

Usando la propiedad 9 de función característica se concluye que  $F_{X_n}(x) \xrightarrow{n \rightarrow \infty} F_X(x)$ , luego  $X_n \xrightarrow{D} X$ .  $\square$

**Nota 3.11.** Sean  $X, X_1, X_2, \dots$  variables aleatorias reales cuyas fda son  $F, F_1, F_2, \dots$ , respectivamente.

1. Si  $X_n \xrightarrow{P} X$ , entonces  $X_n \xrightarrow{D} X$ .  
Para la demostración, consultar [16, p. 302].
2. Si  $X$  es una variable aleatoria degenerada en un punto.

$$X_n \xrightarrow{D} X \Rightarrow X_n \xrightarrow{P} X.$$

3. Si  $X_n \xrightarrow{D} X$ , entonces  $X_n = O_p(1)$ .  
Ver detalles de la prueba en [82, p. 106].

**Nota 3.12.** En general,

$$X_n \xrightarrow{D} X \not\Rightarrow X_n \xrightarrow{P} X.$$

**Ejemplo 3.28.** Sean las variables aleatorias  $X, X_1, X_2, \dots$ , iid Bernoulli de parámetro  $p = \frac{1}{2}$ , es decir,  $P(X_1 = 1) = P(X_1 = 0) = \frac{1}{2}$ . Muestre que  $X_n \not\xrightarrow{P} X$ .

*Solución.* Se tiene que  $X_n \xrightarrow{D} X$ , sin embargo, para todo  $0 < \epsilon < 1$ ,

$$\begin{aligned} P(|X_n - X| > \epsilon) &= P(X_n - X > \epsilon) + P(X_n - X < -\epsilon) \\ &= P(X_n = 1, X = 0) + P(X_n = 0, X = 1) \\ &= \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2} \quad \text{para todo } n, \end{aligned}$$

aplicando límite:

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = \lim_{n \rightarrow \infty} \frac{1}{2} \neq 0,$$

luego,  $X_n \not\xrightarrow{P} X$ .

$\square$

**DEFINICIÓN 3.16.** De acuerdo con [55, p. 59], dados los vectores aleatorios  $m$ -dimensionales  $\mathbf{X}_0 = (X_{01}, X_{02}, \dots, X_{0m})^\top$  y  $\mathbf{X}_n = (X_{n1}, X_{n2}, \dots, X_{nm})^\top$ ,  $n = 1, 2, 3, \dots$  con fda  $F_{X_0}, F_{X_1}, F_{X_2}, \dots$ , respectivamente. Se dice que la sucesión  $\{\mathbf{X}_n\}_{n \geq 1}$  converge en distribución (o en ley) para  $\mathbf{X}_0$  si, para todo punto  $\mathbf{x} = (x_1, x_2, \dots, x_m)^\top$  de continuidad de  $F_{X_0}$ ,

$$F_{X_n}(\mathbf{x}) \xrightarrow[n \rightarrow \infty]{} F_{X_0}(\mathbf{x}).$$

Notación:  $\mathbf{X}_n \xrightarrow{D} \mathbf{X}_0$ .

**Teorema 3.12.** Según [55, p. 59], dados los vectores aleatorios  $m$ -dimensionales  $\mathbf{X}_0 = (X_{01}, X_{02}, \dots, X_{0m})^\top$  y  $\mathbf{X}_n = (X_{n1}, X_{n2}, \dots, X_{nm})^\top$ ,  $n = 1, 2, \dots$  con funciones características  $\psi_{X_0}, \psi_{X_1}, \dots$ , respectivamente. Entonces,

$$\mathbf{X}_n \xrightarrow{D} \mathbf{X}_0 \iff \psi_{X_n}(\mathbf{t}) \xrightarrow[n \rightarrow \infty]{} \psi_{X_0}(\mathbf{t}) \quad \forall \mathbf{t} \in \mathbb{R}^m.$$

*Demostración.* Ver detalles de la prueba en [13]. □

**Teorema 3.13.** (Cramér-Wold). De acuerdo con [82, p. 106], dados los vectores aleatorios  $m$ -dimensionales  $\mathbf{X}_0$  y  $\mathbf{X}_n$ ,  $n = 1, 2, 3, \dots$  Entonces:

$$\mathbf{X}_n \xrightarrow{D} \mathbf{X}_0 \iff \mathbf{t}^\top \mathbf{X}_n \xrightarrow{D} \mathbf{t}^\top \mathbf{X}_0 \quad \forall \mathbf{t} \in \mathbb{R}^m.$$

*Demostración.*

( $\Rightarrow$ ) Usando el teorema 3.12, se sigue que:

$$\mathbf{X}_n \xrightarrow{D} \mathbf{X}_0 \iff \forall \mathbf{t} \in \mathbb{R}^m \quad \psi_{X_n}(\mathbf{t}) \xrightarrow[n \rightarrow \infty]{} \psi_{X_0}(\mathbf{t}),$$

luego para todo  $\mathbf{t} \in \mathbb{R}^m$ , fijo,

$$\begin{aligned} \psi_{X_n}(u\mathbf{t}) &\xrightarrow[n \rightarrow \infty]{} \psi_{X_0}(u\mathbf{t}) \\ &\downarrow \text{Definición 2.13} \\ \mathbb{E}\left(e^{i u \mathbf{t}^\top X_n}\right) &\xrightarrow[n \rightarrow \infty]{} \mathbb{E}\left(e^{i u \mathbf{t}^\top X_0}\right) \\ \psi_{\mathbf{t}^\top X_n}(u) &\xrightarrow[n \rightarrow \infty]{} \psi_{\mathbf{t}^\top X_0}(u), \end{aligned}$$

por tanto,

$$\mathbf{t}^\top \mathbf{X}_n \xrightarrow{D} \mathbf{t}^\top \mathbf{X}_0 \quad \forall \mathbf{t} \in \mathbb{R}^m.$$

( $\Leftarrow$ ) Se parte de

$$\mathbf{t}^\top \mathbf{X}_n \xrightarrow{D} \mathbf{t}^\top \mathbf{X}_0 \quad \forall \mathbf{t} \in \mathbb{R}^m.$$

Del teorema 3.11 se tiene que:

$$\begin{aligned} \psi_{\mathbf{X}_n}(\mathbf{t}) &= \mathbb{E} \left( e^{i\mathbf{t}^\top \mathbf{X}_n} \right) \\ &= \psi_{\mathbf{t}^\top \mathbf{X}_n}(1) \xrightarrow[n \rightarrow \infty]{} \psi_{\mathbf{t}^\top \mathbf{X}_0}(1) = \mathbb{E} \left( e^{i\mathbf{t}^\top \mathbf{X}_0} \right) = \psi_{\mathbf{X}_0}(\mathbf{t}), \end{aligned}$$

por tanto, por el teorema 3.12 se tiene que  $\mathbf{X}_n \xrightarrow{D} \mathbf{X}_0$ .  $\square$

**Teorema 3.14.** De acuerdo con [55, p. 61], dados los vectores aleatorios  $m$ -dimensionales  $\mathbf{X}_0$  y  $\mathbf{X}_n$  definidos en un mismo espacio de probabilidad y  $g : \mathbb{R}^m \rightarrow \mathbb{R}^s$  ( $s \geq 1$ ) una función continua, entonces:

1.  $\mathbf{X}_n \xrightarrow{c.s.} \mathbf{X}_0 \Rightarrow g(\mathbf{X}_n) \xrightarrow{c.s.} g(\mathbf{X}_0)$ .
2.  $\mathbf{X}_n \xrightarrow{P} \mathbf{X}_0 \Rightarrow g(\mathbf{X}_n) \xrightarrow{P} g(\mathbf{X}_0)$ .
3.  $\mathbf{X}_n \xrightarrow{D} \mathbf{X}_0 \Rightarrow g(\mathbf{X}_n) \xrightarrow{D} g(\mathbf{X}_0)$ .

*Demostración.* Ver detalles de la prueba en [55, p. 61] y [11, p. 109].  $\square$

**Ejemplo 3.29.** Sea  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias que converge en distribución hacia  $Z$ ,  $Z \sim N(0, 1)$ . Sea  $g : \mathbb{R} \rightarrow \mathbb{R}$  dada por  $g(x) = x^2$ , luego,

$$X_n \xrightarrow{D} Z \quad \Rightarrow \quad X_n^2 \xrightarrow{D} Z^2,$$

es decir,  $X_n^2$  converge en distribución a la distribución chi-cuadrada con un grado de libertad (ver [20, p. 53]).

**Ejemplo 3.30.** Sean  $(Z_1 \ Z_2)^\top$  y  $(X_n \ Y_n)^\top$ ,  $n = 1, 2, 3, \dots$ , vectores aleatorios bidimensionales, tal que:

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{D} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \quad \text{con} \quad \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right),$$

esto es,  $(Z_1, Z_2)$  tiene *fdp* normal bivariada estándar, con  $\rho = 0$ , luego las variables aleatorias  $Z_1$  y  $Z_2$  son *iid*  $N(0, 1)$ . Sea  $f : \mathbb{R} \times \mathbb{R} - \{0\} \rightarrow \mathbb{R}$  dada por  $f(x, y) = \frac{x}{y}$ , luego por el teorema de transformación (teorema 2.7):

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{D} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \quad \Rightarrow \quad \frac{X_n}{Y_n} \xrightarrow{D} \frac{Z_1}{Z_2},$$

es decir, el cociente  $\frac{X_n}{Y_n}$  converge en distribución a una *fdp* Cauchy estándar  $(0, 1)$  (ver [20, p. 162]).

**Nota 3.13.** Si  $X_n \xrightarrow{D} X$  y  $Y_n \xrightarrow{D} Y$ , **NO** necesariamente se tiene que:

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{D} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

**Teorema 3.15.** Sean  $\{X_n\}_{n \geq 1}$  y  $\{Y_n\}_{n \geq 1}$  sucesiones de variables aleatorias definidas en el mismo espacio de probabilidad  $(\Omega, \mathcal{F}, P)$  y  $X$  una variable aleatoria. Entonces, si

$$Y_n - X_n \xrightarrow{P} 0 \quad \text{y} \quad X_n \xrightarrow{D} X \quad \implies \quad Y_n \xrightarrow{D} X.$$

*Demostración.* Ver detalles de la prueba en [30, p. 41] o [55, p. 56]. □

El teorema 3.15 se conoce como Slutsky I.

**Teorema 3.16. Operaciones algebraicas.** De acuerdo con [55, p. 63], dadas dos sucesiones de variables aleatorias  $\{X_n\}_{n \geq 1}$  y  $\{Y_n\}_{n \geq 1}$  definidas en el mismo espacio de probabilidad  $(\Omega, \mathcal{F}, P)$  y  $X$  y  $Y$  variables aleatorias.

1. Si  $X_n \xrightarrow{c.s.} X$  y  $Y_n \xrightarrow{c.s.} Y$ , entonces:

$$a) \quad X_n + Y_n \xrightarrow{c.s.} X + Y, \quad b) \quad X_n Y_n \xrightarrow{c.s.} XY.$$

2. Si  $X_n \xrightarrow{P} X$  y  $Y_n \xrightarrow{P} Y$ , entonces:

$$a) \quad X_n + Y_n \xrightarrow{P} X + Y, \quad b) \quad X_n Y_n \xrightarrow{P} XY.$$

3. (Slutsky) Si  $X_n \xrightarrow{D} X$  y  $Y_n \xrightarrow{P} c$  donde  $c$  es una constante, entonces:

$$a) \quad X_n + Y_n \xrightarrow{D} X + c, \quad c) \quad \text{Si } c \neq 0, \quad \frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{c}.$$

$$b) \quad X_n Y_n \xrightarrow{D} cX,$$

*Demostración.*

1. Sean  $f_1 : \mathbb{R}^2 \rightarrow \mathbb{R}$  y  $f_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$  las funciones definidas para todo  $(x, y)^\top \in \mathbb{R}^2$  por:

$$f_1(x, y) = x + y \quad \text{y} \quad f_2(x, y) = xy. \quad (3.20)$$

Note que  $f_1$  y  $f_2$  son funciones continuas. Por otra parte, dado que:

$$X_n \xrightarrow{c.s.} X \quad \text{y} \quad Y_n \xrightarrow{c.s.} Y,$$

usando la nota 3.9 se tiene que:

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{c.s.} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Como  $f_1$  y  $f_2$  son funciones continuas, del teorema 3.14 se tiene que:

$$X_n + Y_n \xrightarrow{c.s.} X + Y \quad \text{y} \quad X_n Y_n \xrightarrow{c.s.} XY.$$

2. Se parte de  $X_n \xrightarrow{P} X$  y  $Y_n \xrightarrow{P} Y$ , luego por el teorema 3.9 se tiene que:

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{P} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Considerando de nuevo las funciones dadas en (3.20), por el teorema 3.14 se tiene que:

$$X_n + Y_n \xrightarrow{P} X + Y \quad \text{y} \quad X_n Y_n \xrightarrow{P} XY.$$

3. Sean  $g_1 : \mathbb{R} \rightarrow \mathbb{R}$  y  $g_2 : \mathbb{R} \rightarrow \mathbb{R}$  las funciones definidas por

$$g_1(x) = x + c \quad \text{y} \quad g_2(x) = tx \quad \forall c, t \in \mathbb{R}$$

y  $g_3$  una función tal que  $g_3 : \mathbb{R} - \{0\} \rightarrow \mathbb{R}$  definida por  $g_3(x) = \frac{1}{x}$  (note que  $g_1$ ,  $g_2$  y  $g_3$  son funciones continuas).

- a) Como  $X_n \xrightarrow{D} X$  se tiene, por el teorema 3.14 y usando la función  $g_1$ , que la variable  $W_n = X_n + c \xrightarrow{D} X + c$ . Por otro lado,

$$Y_n - c \xrightarrow{P} 0, \quad \text{sumando y restando } X_n :$$

$$\underbrace{Y_n + X_n}_{Z_n} - \underbrace{(X_n + c)}_{W_n} \xrightarrow{P} 0,$$

usando el teorema 3.15,

$$Z_n \xrightarrow{D} X + c$$

$$Y_n + X_n \xrightarrow{D} X + c.$$

- b) Como  $X_n \xrightarrow{D} X$  se tiene, por el teorema 3.14, y considerando  $g_2$ , que  $t_1 X_n \xrightarrow{D} t_1 X$  para todo  $t_1 \in \mathbb{R}$ , de igual forma, como

$Y_n \xrightarrow{P} c$  se tiene, por el teorema 3.14, que  $t_2 Y_n \xrightarrow{P} t_2 c$  para todo  $t_2 \in \mathbb{R}$ . Usando el resultado del ítem 3a:

$$t_1 X_n + t_2 Y_n \xrightarrow{D} t_1 X + ct_2 \quad \forall t = (t_1, t_2)^\top \in \mathbb{R}^2,$$

por tanto, por Cramér-Wold, teorema 3.13,

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{D} \begin{pmatrix} X \\ c \end{pmatrix}.$$

Aplicando el teorema 3.14 y usando la función  $f_2$  dada en (3.20), se tiene que:

$$X_n Y_n \xrightarrow{D} cX.$$

- c) Como  $Y_n \xrightarrow{P} c$  y  $c \neq 0$  se tiene, por el teorema 3.14 y usando la función  $g_3$ , que  $\frac{1}{Y_n} \xrightarrow{P} \frac{1}{c}$ . Ahora, aplicando el resultado del ítem 3b:

$$\begin{pmatrix} X_n \\ \frac{1}{Y_n} \end{pmatrix} \xrightarrow{D} \begin{pmatrix} X \\ \frac{1}{c} \end{pmatrix},$$

luego, considerando la función  $f_2$  dada en (3.20):

$$X_n \frac{1}{Y_n} \xrightarrow{D} X \frac{1}{c}. \quad \checkmark$$

**Ejemplo 3.31.** Supóngase, como en [77, p. 269], que  $\{X_n\}_{n \geq 1}$  es una sucesión de variables aleatorias tal que  $X_n \xrightarrow{D} X$ ,  $X \sim N(\mu, \sigma^2)$ , y sean  $a_n$ ,  $a$ ,  $b_n$  y  $b$  constantes tales que  $a_n \rightarrow a$  y  $b_n \rightarrow b$ . Muestre que:

$$a_n X_n + b_n \xrightarrow{D} Y, \quad Y \sim N(a\mu + b, a^2 \sigma^2).$$

*Solución.* De manera trivial se tiene que:

$$a_n \xrightarrow{D} a \quad \text{y} \quad b_n \xrightarrow{D} b.$$

Luego, por las partes 3b y 3a del teorema 3.16:

$$a_n X_n \xrightarrow{D} aX, \quad \text{y} \quad a_n X_n + b_n \xrightarrow{D} aX + b.$$

Puesto que  $X \sim N(\mu, \sigma^2)$  implica que  $aX + b \sim N(a\mu + b, a^2 \sigma^2)$ . Por lo tanto,

$$a_n X_n + b_n \xrightarrow{D} aX + b = Y, \quad Y \sim N(a\mu + b, a^2 \sigma^2). \quad \checkmark$$

**Teorema 3.17.** (Método delta). De acuerdo con [20, p. 243], dada una sucesión de variables aleatorias  $\{X_n\}_{n \geq 1}$  y un valor específico de  $\theta$ , tal que:

$$\sqrt{n}(X_n - \theta) \xrightarrow{D} N(0, \sigma^2).$$

Si  $g$  es una función continua que es diferenciable en  $\theta$  y  $g'(\theta) \neq 0$ . Entonces,

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{D} N(0, \sigma^2 [g'(\theta)]^2). \quad (3.21)$$

*Demostración.* Siguiendo a [82, p. 131], sean

$$Z_n = \sqrt{n}(X_n - \theta) \quad \text{y} \quad G_n = \frac{g(X_n) - g(\theta)}{X_n - \theta},$$

luego:

$$Z_n G_n = \sqrt{n} \cancel{(X_n - \theta)} \frac{g(X_n) - g(\theta)}{\cancel{X_n - \theta}} = \sqrt{n}[g(X_n) - g(\theta)].$$

Como  $Z_n \xrightarrow{D} Z$  donde  $Z \sim N(0, \sigma^2)$ , por el ítem b) de la nota 3.5, los órdenes de magnitud dados en la tabla 3.12 y la nota 3.11, se tiene que:

$$\begin{aligned} Z_n &= O_p(1) \\ \frac{1}{\sqrt{n}} \sqrt{n}(X_n - \theta) &= \frac{1}{\sqrt{n}} O_p(1) = O_p\left(\frac{1}{\sqrt{n}}\right) = o_p(1), \end{aligned}$$

luego,

$$X_n \xrightarrow{P} \theta,$$

y por el teorema 3.14:

$$g(X_n) \xrightarrow{P} g(\theta).$$

Por otro lado, la expansión de Taylor para  $g(X_n)$  en torno de  $X_n = \theta$  es:

$$g(X_n) = g(\theta) + g'(\theta)(X_n - \theta) + o(X_n - \theta),$$

al despejar se obtiene que:

$$\underbrace{g(X_n) - g(\theta)}_{o_p(1)} = g'(\theta) \underbrace{(X_n - \theta)}_{o_p(1)} + \underbrace{o(X_n - \theta)}_{o(1) \text{ cuando } X_n \rightarrow \theta},$$

así,

$$G_n = \frac{g(X_n) - g(\theta)}{X_n - \theta} \xrightarrow{P} g'(\theta).$$

Usando Slutsky, se tiene que:

$$\begin{aligned} Z_n G_n &\xrightarrow{D} g'(\theta)Z \\ \sqrt{n} [g(X_n) - g(\theta)] &\xrightarrow{D} g'(\theta)Z = W, \end{aligned}$$

donde  $W = g'(\theta)Z \sim N(0, \sigma^2 [g'(\theta)]^2)$ . ☑

### 3.4.3. Teoremas límites

**Teorema 3.18.** (Ley débil de los grandes números). Según [52, p. 49], dadas las variables aleatorias  $X_1, X_2, \dots, X_n$  *iid* con media  $\mathbb{E}(X_1) = \mu$  y varianza  $\mathbb{V}(X_1) = \sigma^2 < \infty$ . Si  $T_n = \sum_{k=1}^n X_k$ . Entonces,

$$\frac{T_n}{n} \xrightarrow{P} \mu.$$

*Demostración.* Del ejemplo 3.19 se tiene que:

$$\frac{T_n}{n} - \mu = O_p\left(\frac{1}{\sqrt{n}}\right) = o_p(1),$$

y de la definición 3.9:

$$\frac{T_n}{n} \xrightarrow{P} \mu. \quad \text{☑}$$

Para una animación de la ley débil de los grandes números, se puede usar el código en R:

```
library(animation)
oopt=ani.options(interval=0.01,nmax=ifelse(interactive(),150,2))
lambda=4
lln.ani(FUN=function(n,lambda) rpois(n,lambda),mu=lambda,cex=0.6)
```

**Teorema 3.19.** (Ley débil de los grandes números de Khintchine). De acuerdo con [82, p. 61], dada una sucesión de variables aleatorias  $\{X_k\}_{k \geq 1}$  *iid* con media  $\mathbb{E}(X_1) = \mu$ . Si  $T_n = \sum_{k=1}^n X_k$ , entonces,

$$\frac{T_n}{n} \xrightarrow{P} \mu.$$

*Demostración.* Sea  $F_{X_1}(x)$ ,  $x \in \mathbb{R}$ , la fda de  $X_1$ , y sea  $\psi_{X_1}(t)$ ,  $t \in \mathbb{R}$ , la función característica de  $X_1$ :

$$\begin{aligned}\psi_{\frac{T_n}{n}}(t) &= \mathbb{E} \left( \exp \left\{ it \frac{T_n}{n} \right\} \right) = \mathbb{E} \left( \exp \left\{ \frac{it}{n} \sum_{k=1}^n X_k \right\} \right) \\ &= \mathbb{E} \left( \prod_{k=1}^n \exp \left\{ \frac{itX_k}{n} \right\} \right) \underbrace{=}_{\text{Indep.}} \prod_{k=1}^n \mathbb{E} \left( \exp \left\{ \frac{itX_k}{n} \right\} \right) \\ &= \prod_{k=1}^n \psi_{X_k} \left( \frac{t}{n} \right) \underbrace{=}_{\text{Ident.Distr.}} \left[ \psi_{X_1} \left( \frac{t}{n} \right) \right]^n\end{aligned}$$

Usando la expresión (A.9) del apéndice A, para  $\psi_{X_1}(\tilde{t})$ , con  $\tilde{t} = \frac{t}{n}$  en torno de 0, para algún  $c$  tal que  $0 < c < \tilde{t}$ , se tiene que:

$$\begin{aligned}\psi_{X_1}(\tilde{t}) &= \psi_{X_1}(0) + (\tilde{t} - 0) \psi'_{X_1}(0) + \frac{(\tilde{t} - 0)^2}{2} \psi''_{X_1}(c) \\ &= 1 + i\mathbb{E}(X_1)(\tilde{t}) + \frac{\tilde{t}^2}{2} \psi''_{X_1}(c) = 1 + i\mu\tilde{t} + \frac{\tilde{t}^2}{2} \psi''_{X_1}(c),\end{aligned}$$

aquí se usó la propiedad 5 de la función característica. Elevando a ambos lados a la  $n$ :

$$\left[ \psi_{X_1}(\tilde{t}) \right]^n = \left[ 1 + i\mu\tilde{t} + \frac{\tilde{t}^2}{2} \psi''_{X_1}(c) \right]^n = \exp \left\{ n \ln \left[ 1 + i\mu\tilde{t} + \frac{\tilde{t}^2}{2} \psi''_{X_1}(c) \right] \right\}.$$

Por otra parte, la fórmula de Taylor para  $\ln(1+z)$  está dada por:

$$\ln(1+z) = \sum_{k=0}^{\infty} (-1)^k \frac{z^{k+1}}{k+1} \quad 0 < z < 1,$$

tomando  $z = i\mu\tilde{t} + \frac{\tilde{t}^2}{2} \psi''_{X_1}(c)$  se obtiene que:

$$\begin{aligned}\ln \left[ 1 + i\mu\tilde{t} + \frac{\tilde{t}^2}{2} \psi''_{X_1}(c) \right] &= i\mu\tilde{t} + \frac{\tilde{t}^2}{2} \psi''_{X_1}(c) \quad \text{cuando } \tilde{t} \rightarrow 0 \\ &= i\mu\tilde{t} + o(\tilde{t}) = \frac{i\mu t}{n} + o\left(\frac{t}{n}\right) \quad \text{cuando } \frac{t}{n} \rightarrow 0,\end{aligned}$$

luego, dado un  $n > 0$  fijo, cuando  $t \rightarrow 0$ :

$$\left[ \psi_{X_1} \left( \frac{t}{n} \right) \right]^n = \exp \left\{ n \left[ \frac{i\mu t}{n} + o\left(\frac{t}{n}\right) \right] \right\} = \exp \left\{ i\mu t + n \cdot o\left(\frac{t}{n}\right) \right\}.$$

Ahora, para un  $t$  fijo, tomando límite sobre  $n$ , se tiene que:

$$\begin{aligned} \lim_{n \rightarrow \infty} \psi_{\frac{T_n}{n}}(t) &= \lim_{n \rightarrow \infty} \left\{ \left[ \psi_{X_1} \left( \frac{t}{n} \right) \right]^n \right\} = \lim_{n \rightarrow \infty} \left\{ \exp \left[ i \mu t + \underbrace{n}_{O(n)} o \left( \frac{t}{n} \right) \right] \right\} \\ &= \lim_{n \rightarrow \infty} \{ \exp [i \mu t + o(t)] \} = \exp \{i \mu t\}. \end{aligned}$$

Como  $\exp \{i \mu t\}$  corresponde a la función característica de una variable aleatoria que es degenerada en el punto  $\mu$ , se tiene por el teorema 3.11 que:

$$\frac{T_n}{n} \xrightarrow{D} \mu$$

y por la nota 3.11 se tiene que:

$$\frac{T_n}{n} \xrightarrow{P} \mu. \quad \checkmark$$

**Teorema 3.20.** (Ley fuerte de los grandes números de Kolmogorov). *De acuerdo con [82, p. 67], dada una sucesión  $\{X_k\}_{k \geq 1}$  de variables aleatorias independientes con media  $\mathbb{E}(X_k) = \mu_k$  y varianza  $\mathbb{V}(X_k) = \sigma_k^2 < \infty$ . Si*

$$\bar{\mu}_n = \frac{1}{n} \sum_{k=1}^n \mu_k \quad \text{y} \quad T_n = \sum_{k=1}^n X_k, \quad n \geq 1.$$

Entonces,

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty \implies \frac{T_n}{n} - \bar{\mu}_n \xrightarrow{c.s.} 0.$$

*Demostración.* Ver detalles de la prueba en [82, p. 67]. \checkmark

**Teorema 3.21.** (Ley fuerte de los grandes números de Khintchine). *De acuerdo con [82, p. 69], dada una sucesión de variables aleatorias  $\{X_k\}_{k \geq 1}$  iid. Si*

$$T_n = \sum_{k=1}^n X_k, \text{ entonces,}$$

$$\frac{T_n}{n} \xrightarrow{c.s.} c.$$

donde  $c$  es una constante si y solo si  $\mathbb{E}(|X_1|) < \infty$  y  $c = \mathbb{E}(X_1)$ .

*Demostración.* Ver detalles de la prueba en [82, p. 69]. \checkmark

**Ejemplo 3.32.** *Sea una muestra aleatoria  $(X_1, Y_1)^\top, (X_2, Y_2)^\top, \dots, (X_n, Y_n)^\top$  de la distribución bivariada del vector aleatorio  $(X_1, Y_1)^\top$ , con vector de medias y*

matriz de varianzas-covarianzas dadas por:

$$\boldsymbol{\mu} = \mathbb{E} \begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \quad y \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

Se define el coeficiente de correlación muestral de  $r_n$  como:

$$r_n = \frac{\sum_{k=1}^n (X_k - \bar{X}_n)(Y_k - \bar{Y}_n)}{\sqrt{\sum_{k=1}^n (X_k - \bar{X}_n)^2 \sum_{k=1}^n (Y_k - \bar{Y}_n)^2}},$$

donde  $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$  y  $\bar{Y}_n = \frac{1}{n} \sum_{k=1}^n Y_k$ . Muestre que  $r_n \xrightarrow{c.s.} \rho$ .

*Solución.* De manera análoga al proceso dado en la nota 2.5, se tiene que:

$$r_n = \frac{\sum_{k=1}^n \frac{X_k Y_k}{n} - \bar{X}_n \bar{Y}_n}{\sqrt{\left(\sum_{k=1}^n \frac{X_k^2}{n} - \bar{X}_n^2\right) \left(\sum_{k=1}^n \frac{Y_k^2}{n} - \bar{Y}_n^2\right)}} = \frac{S_{XY,n}}{\sqrt{S_{XX,n}} \sqrt{S_{YY,n}}}$$

1. Por la ley fuerte de los grandes números de Khintchine, teorema 3.21,

$$\sum_{k=1}^n \frac{X_k Y_k}{n} \xrightarrow{c.s.} \rho\sigma_X\sigma_Y + \mu_X\mu_Y, \quad (3.22)$$

porque  $(X_1, Y_1)^\top, (X_2, Y_2)^\top, \dots, (X_n, Y_n)^\top$  son vectores aleatorios iid con:

$$\mathbb{E}(X_1 Y_1) = \text{Cov}(X_1 Y_1) + \mathbb{E}(X_1)\mathbb{E}(Y_1) = \rho\sigma_X\sigma_Y + \mu_X\mu_Y < \infty.$$

2. Sea  $f(x) = x^2$ , por el teorema 3.14:

- $\bar{X}_n \xrightarrow{c.s.} \mu_X \Rightarrow \bar{X}_n^2 \xrightarrow{c.s.} \mu_X^2$  porque  $X_1, X_2, \dots, X_n$  son variables aleatorias iid con  $\mathbb{E}(X_1) = \mu_X < \infty$ .
- $\bar{Y}_n \xrightarrow{c.s.} \mu_Y \Rightarrow \bar{Y}_n^2 \xrightarrow{c.s.} \mu_Y^2$  porque  $Y_1, Y_2, \dots, Y_n$  son variables aleatorias iid con  $\mathbb{E}(Y_1) = \mu_Y < \infty$ .

3. Si  $Y_k = X_k, \forall k$ , entonces  $\rho = 1$  y la expresión (3.22) queda:

- $\sum_{k=1}^n \frac{X_k^2}{n} \xrightarrow{c.s.} \mu_X^2 + \sigma_X^2$  porque  $\mathbb{E}(X_1^2) = \sigma_X^2 + \mu_X^2 < \infty$ .
- $\sum_{k=1}^n \frac{Y_k^2}{n} \xrightarrow{c.s.} \mu_Y^2 + \sigma_Y^2$  porque  $\mathbb{E}(Y_1^2) = \sigma_Y^2 + \mu_Y^2 < \infty$ .

entonces por la nota 3.9,

$$\begin{pmatrix} \sum_{k=1}^n \frac{X_k Y_k}{n} \\ \bar{X}_n \\ \bar{X}_n^2 \\ \bar{Y}_n \\ \bar{Y}_n^2 \\ \sum_{k=1}^n \frac{X_k^2}{n} \\ \sum_{k=1}^n \frac{Y_k^2}{n} \end{pmatrix} \xrightarrow{c.s.} \begin{pmatrix} \rho\sigma_X\sigma_Y + \mu_X\mu_Y \\ \mu_X \\ \mu_X^2 \\ \mu_Y \\ \mu_Y^2 \\ \mu_X^2 + \sigma_X^2 \\ \mu_Y^2 + \sigma_Y^2 \end{pmatrix}$$

y por el teorema 3.20,

$$\begin{aligned} r_n &= \frac{\sum_{k=1}^n \frac{X_k Y_k}{n} - \bar{X}_n \bar{Y}_n}{\sqrt{\sum_{k=1}^n \frac{X_k^2}{n} - \bar{X}_n^2} \sqrt{\sum_{k=1}^n \frac{Y_k^2}{n} - \bar{Y}_n^2}} \xrightarrow{c.s.} \frac{\rho\sigma_X\sigma_Y + \mu_X\mu_Y - \mu_X\mu_Y}{\sqrt{\mu_X^2 + \sigma_X^2 - \mu_X^2} \sqrt{\mu_Y^2 + \sigma_Y^2 - \mu_Y^2}} \\ &= \frac{\rho\sigma_X\sigma_Y}{\sigma_X\sigma_Y} = \rho. \quad \square \end{aligned}$$

**Teorema 3.22.** (Teorema del límite central clásico). De acuerdo con [82, p. 107], si  $\{X_k\}_{k \geq 1}$  son variables aleatorias iid con media  $\mathbb{E}(X_1) = \mu$  y varianza  $\mathbb{V}(X_1) = \sigma^2 < \infty$ . Definiendo  $T_n = \sum_{k=1}^n X_k$ , entonces,

$$Z_n = \frac{T_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} Z, \tag{3.23}$$

donde  $Z \sim N(0, 1)$ . Notación:  $Z_n \xrightarrow{D} N(0, 1)$ .

*Demostración.* Siguiendo a [82, p. 107], la función característica de  $Z_n$  es:

$$\begin{aligned} \psi_{Z_n}(t) &= \mathbb{E} \left\{ \exp [itZ_n] \right\} = \mathbb{E} \left\{ \exp \left[ it \left( \frac{T_n - n\mu}{\sigma\sqrt{n}} \right) \right] \right\} \\ &= \mathbb{E} \left\{ \exp \left[ \frac{it}{\sigma\sqrt{n}} \left( \sum_{k=1}^n X_k - n\mu \right) \right] \right\} = \mathbb{E} \left\{ \exp \left[ \frac{it}{\sqrt{n}} \sum_{k=1}^n \underbrace{\frac{(X_k - \mu)}{\sigma}}_{U_k} \right] \right\} \\ &= \mathbb{E} \left\{ \prod_{k=1}^n \exp \left[ \frac{it}{\sqrt{n}} U_k \right] \right\} \stackrel{\text{Indep.}}{=} \prod_{k=1}^n \mathbb{E} \left\{ \exp \left[ \frac{it}{\sqrt{n}} U_k \right] \right\} \stackrel{\text{Ident. Distr.}}{=} \left[ \psi_{U_1} \left( \frac{t}{\sqrt{n}} \right) \right]^n. \end{aligned}$$

$\psi_{U_k} \left( \frac{t}{\sqrt{n}} \right)$

Usando la fórmula de Taylor dada en (A.2) para la función exponencial:

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + R_3(x, 0),$$

donde  $R_3(x, 0)$  denota el residuo, se tiene que la expansión de  $\exp\left\{\frac{it}{\sqrt{n}}U_1\right\}$  en torno de 0, es dada por:

$$\exp\left\{\frac{itU_1}{\sqrt{n}}\right\} = 1 + \frac{itU_1}{\sqrt{n}} - \frac{t^2U_1^2}{2n} + R_3\left(\frac{it}{\sqrt{n}}U_1, 0\right).$$

Por otra parte, como  $U_1 = \frac{X_1 - \mu}{\sigma}$ , aplicando el operador esperanza:

$$\mathbb{E}(U_1) = 0 \quad \text{y} \quad \mathbb{E}(U_1^2) = \mathbb{V}\left(\frac{X_1 - \mu}{\sigma}\right) + \left[\cancel{\mathbb{E}\left(\frac{X_1 - \mu}{\sigma}\right)}\right]^2 = 1,$$

se tiene que:

$$\begin{aligned} \underbrace{\mathbb{E}\left[\exp\left\{\frac{itU_1}{\sqrt{n}}\right\}\right]}_{\psi_{U_1}\left(\frac{t}{\sqrt{n}}\right)} &= \mathbb{E}\left[1 + \frac{itU_1}{\sqrt{n}} - \frac{t^2U_1^2}{2n} + R_3\left(\frac{it}{\sqrt{n}}U_1, 0\right)\right] \\ \psi_{U_1}\left(\frac{t}{\sqrt{n}}\right) &= 1 - \frac{t^2}{2n} + \mathbb{E}\left(R_3\left(\frac{itU_1}{\sqrt{n}}, 0\right)\right), \end{aligned}$$

elevando a ambos lados a la  $n$  se obtiene que:

$$\psi_{Z_n}(t) = \left[\psi_{U_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n = \left\{1 - \frac{t^2}{2n} + \mathbb{E}\left(R_3\left(\frac{itU_1}{\sqrt{n}}, 0\right)\right)\right\}^n.$$

Aplicando límite cuando  $n \rightarrow \infty$  y dado que  $R_3\left(\frac{itU_1}{\sqrt{n}}, 0\right) = o\left(i^2t^2\frac{U_1^2}{n}\right)$  cuando  $iU_1\frac{t}{\sqrt{n}} \rightarrow 0$  ( $\frac{t}{\sqrt{n}} \rightarrow 0$ ),

$$\lim_{n \rightarrow \infty} \left[\psi_{U_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n = \lim_{n \rightarrow \infty} \left[\left\{1 - \frac{t^2}{2n} + \mathbb{E}\left(R_3\left(\frac{itU_1}{\sqrt{n}}, 0\right)\right)\right\}^n\right] = \exp\left\{-\frac{t^2}{2}\right\},$$

en la última expresión se usó el resultado del pie de página 14 y corresponde a la función característica de la fdp normal estándar dada en (1.63).  $\square$

Una animación del teorema central de límite, se obtiene con el código en R:

```
library(animation)
oopt=ani.options(interval=0.1,nmax=ifelse(interactive(),200,2))
```

```

op=par(mar=c(3,3,1,0.5), mgp = c(1.5, 0.5, 0), tcl = -0.3)
lambda=4
f = function(n) rpois(n, lambda)
clt.ani(FUN = f, mean = lambda, sd = lambda,type="s")
par(op)
gl=5
f = function(n) rchisq(n, gl)
clt.ani(FUN = f, mean = gl, sd = sqrt(2 * gl),type="s")
    
```

**Ejemplo 3.33.** Sean  $X_1, X_2, X_3, \dots, X_n$  variables aleatorias *iid* Bernoulli de parámetro  $p$ , luego  $\mathbb{E}(X_1) = \mu = p$  y  $\mathbb{V}(X_1) = \sigma^2 = p(1 - p)$ . Definiendo  $T_n = \sum_{k=1}^n X_k$  y  $\bar{X}_n = \frac{T_n}{n}$ . Muestre que:

$$\sqrt{n}(\bar{X}_n - p) \xrightarrow{D} N(0, p(1 - p)).$$

*Solución.* Como  $X_k$  es una variable aleatoria Bernoulli, solo toma los valores

$$X_k = \begin{cases} 1, & \text{con probabilidad } p, \\ 0, & \text{con probabilidad } 1 - p. \end{cases} \quad (3.24)$$

Por la expresión (3.23) se tiene que:

$$\begin{aligned} Z_n = \frac{T_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} Z & \Leftrightarrow \frac{\sum_{k=1}^n X_k - np}{\sqrt{n}\sqrt{p(1-p)}} \xrightarrow{D} N(0, 1) \Leftrightarrow \\ \frac{\sum_{k=1}^n X_k - np}{\sqrt{n}} \xrightarrow{D} N(0, p(1-p)) & \Leftrightarrow \frac{n(\bar{X}_n - p)}{\sqrt{n}} \xrightarrow{D} N(0, p(1-p)). \end{aligned}$$

Al racionalizar el denominador se tiene el resultado. □

**Ejemplo 3.34.** Sean  $X_1, X_2, X_3, \dots, X_n$  variables aleatorias *iid* Poisson de parámetro  $\lambda$ , luego  $\mathbb{E}(X_1) = \mu = \lambda$  y  $\mathbb{V}(X_1) = \sigma^2 = \lambda$ . Definiendo  $T_n = \sum_{k=1}^n X_k$  y  $\bar{X}_n = \frac{T_n}{n}$ . Muestre que:

$$\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{D} N(0, \lambda). \quad (3.25)$$

*Solución.* Por la expresión (3.23) se tiene que:

$$Z_n = \frac{T_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} Z \Leftrightarrow \frac{\sum_{k=1}^n X_k - n\lambda}{\sqrt{n}\sqrt{\lambda}} \xrightarrow{D} N(0, 1) \Leftrightarrow$$

$$\frac{\sum_{k=1}^n X_k - n\lambda}{\sqrt{n}} \xrightarrow{D} N(0, \lambda) \quad \Leftrightarrow \quad \frac{n(\bar{X}_n - \lambda)}{\sqrt{n}} \xrightarrow{D} N(0, \lambda).$$

Al racionalizar el denominador se tiene el resultado. En este caso, si se usa el método delta para encontrar una distribución asintótica que sea independiente del parámetro  $\lambda$ , se puede tomar  $g(y) = \sqrt{y}$  (ver [52, p. 88]). De la expresión (3.25):

$$\sqrt{n} \left( \bar{X}_n - \underbrace{\lambda}_{\theta} \right) \xrightarrow{D} N \left( 0, \underbrace{\lambda}_{\sigma^2} \right),$$

y se tiene que  $g'(y) = \frac{1}{2\sqrt{y}}$ , entonces por la expresión (3.21):

$$\sqrt{n} [g(\bar{X}_n) - g(\theta)] \xrightarrow{D} N \left( 0, \sigma^2 [g'(\theta)]^2 \right),$$

sustituyendo,

$$\begin{aligned} \sqrt{n} [g(\bar{X}_n) - g(\lambda)] &\xrightarrow{D} N \left( 0, \underbrace{\sigma^2}_{\lambda} \underbrace{[g'(\lambda)]^2}_{\left[\frac{1}{2\sqrt{\lambda}}\right]^2} \right) \\ \sqrt{n} \left[ \sqrt{\bar{X}_n} - \sqrt{\lambda} \right] &\xrightarrow{D} N \left( 0, \lambda \left[ \frac{1}{2\sqrt{\lambda}} \right]^2 \right) \\ \sqrt{n} \left[ \sqrt{\bar{X}_n} - \sqrt{\lambda} \right] &\xrightarrow{D} N \left( 0, \frac{1}{4} \right) \\ 2\sqrt{n} \left[ \sqrt{\bar{X}_n} - \sqrt{\lambda} \right] &\xrightarrow{D} N(0, 1). \quad \checkmark \end{aligned}$$

**Teorema 3.23.** (Teorema central del límite de Lindeberg). Según [55, p. 81], dada una sucesión  $\{X_n\}_{n \geq 1}$  de variables aleatorias independientes con *fd*  $F_n(x)$  tales que  $\mathbb{E}(X_n) = 0$  y  $\mathbb{V}(X_n) = \sigma_n^2 < \infty$ . Definiendo  $T_n = \sum_{i=1}^n X_i$  y  $s_n^2 = \sum_{i=1}^n \sigma_i^2$ .

Si para todo  $\epsilon > 0$ :

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n \int_{|y| \geq \epsilon s_n} y^2 d[F_i(y)] = 0,$$

entonces,

$$\frac{T_n - \mathbb{E}(T_n)}{s_n} \xrightarrow{D} N(0, 1).$$

*Demostración.* Ver detalles de la prueba en [55, p. 81]. \checkmark

**Teorema 3.24.** (Teorema central del límite de Liapunov). Según [55, p. 84], dada una sucesión  $\{X_n\}_{n \geq 1}$  de variables aleatorias independientes con  $\mathbb{E}(X_n) = \mu_n$  y  $\mathbb{V}(X_n) = \sigma_n^2 < \infty$ , en que por lo menos un  $\sigma_n^2 > 0$ , suponiendo que, para algún  $\delta > 0$ ,  $\mathbb{E}(|X_n - \mu_n|^{2+\delta}) < \infty$ ,  $n = 1, 2, 3, \dots$ . Definiendo  $T_n = \sum_{i=1}^n X_i$  y

$$s_n^2 = \sum_{i=1}^n \sigma_i^2. \text{ Si}$$

$$\lim_{n \rightarrow \infty} s_n^{-2-\delta} \sum_{i=1}^n \mathbb{E}(|X_i - \mu_i|^{2+\delta}) = 0, \tag{3.26}$$

entonces,

$$\frac{T_n - \mathbb{E}(T_n)}{s_n} \xrightarrow{D} N(0, 1).$$

*Demostración.* Ver detalles de la prueba en [55, p. 85] . □

**Ejemplo 3.35.** Supóngase, como en [59, p. 345], una sucesión  $X_1, X_2, \dots, X_n$  de variables aleatorias independientes definidas como:

$$X_n = \begin{cases} \frac{-1}{\sqrt{n}} & \text{con probabilidad } \frac{1}{2} \\ \frac{1}{\sqrt{n}} & \text{con probabilidad } \frac{1}{2} \end{cases} \quad y \quad T_n = \sum_{i=1}^n X_i.$$

Encuentre la distribución límite de  $T_n$  cuando es convenientemente normalizada.

*Solución.* Para cada una de las  $X_k$  se tiene que:

$$\begin{aligned} \mathbb{E}(X_k) &= \frac{-1}{\sqrt{k}} \times \frac{1}{2} + \frac{1}{\sqrt{k}} \times \frac{1}{2} = 0 \\ \mathbb{V}(X_k) &= \mathbb{E}(X_k^2) - [\mathbb{E}(X_k)]^2 = \mathbb{E}(X_k^2) = \left(\frac{-1}{\sqrt{k}}\right)^2 \times \frac{1}{2} + \left(\frac{1}{\sqrt{k}}\right)^2 \times \frac{1}{2} \\ &= \frac{1}{2k} + \frac{1}{2k} = \frac{1}{k}, \end{aligned}$$

luego,

$$\begin{aligned} \mathbb{E}(T_n) &= \mathbb{E}(X_1 + X_2 + X_3 + \dots + X_n) = \mathbb{E}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \mathbb{E}(X_k) = 0 \\ \mathbb{V}(T_n) &= s_n^2 = \mathbb{V}(X_1 + X_2 + X_3 + \dots + X_n) \underset{\text{indep.}}{=} \sum_{k=1}^n \mathbb{V}(X_k) = \sum_{k=1}^n \frac{1}{k}. \end{aligned}$$

Note que  $\mathbb{V}(T_n)$  diverge. Por otro lado, se va a evaluar (3.26), esto es, la condición de Liapunov,

$$\begin{aligned} \lim_{n \rightarrow \infty} s_n^{-2-\delta} \sum_{k=1}^n \mathbb{E}(|X_k - \mu_k|^{2+\delta}) &= \lim_{n \rightarrow \infty} \left( \sum_{k=1}^n \frac{1}{k} \right)^{\frac{-2-\delta}{2}} \underbrace{\sum_{k=1}^n \mathbb{E}(|X_k|^{2+\delta})}_{(a)} \\ (a) = \mathbb{E}(|X_k|^{2+\delta}) &= \left| \frac{-1}{\sqrt{k}} \right|^{2+\delta} \times \frac{1}{2} + \left| \frac{1}{\sqrt{k}} \right|^{2+\delta} \times \frac{1}{2} \\ &= \left( \frac{1}{\sqrt{k}} \right)^{2+\delta} = \frac{1}{k^{\frac{2+\delta}{2}}} = \frac{1}{k^{1+\frac{\delta}{2}}}. \end{aligned}$$

Entonces,

$$\lim_{n \rightarrow \infty} \left( \sum_{k=1}^n \frac{1}{k} \right)^{\frac{-2-\delta}{2}} \sum_{k=1}^n \mathbb{E}(|X_k|^{2+\delta}) = \lim_{n \rightarrow \infty} \frac{\overbrace{\sum_{k=1}^n \frac{1}{k^{1+\frac{\delta}{2}}}}^{\text{converge}}}{\underbrace{\left( \sum_{k=1}^n \frac{1}{k} \right)^{1+\frac{\delta}{2}}}_{\text{diverge}}} = 0.$$

Por el teorema del límite central de Liapunov:

$$\frac{T_n - \overbrace{\mathbb{E}(T_n)}^0}{s_n} \xrightarrow{D} N(0, 1) \quad \Rightarrow \quad \frac{T_n}{s_n} \xrightarrow{D} N(0, 1). \quad \checkmark$$

**Teorema 3.25.** De acuerdo con [82, p. 167], dada una muestra aleatoria  $X_1, X_2, \dots, X_n$  de una variable aleatoria con fda  $F$  y considerando que su  $q$ -ésimo cuantil,  $\xi_q$ ,  $0 < q < 1$ , está definido de forma única, es decir, para todo  $\eta > 0$ ,  $F(\xi_q - \eta) < F(\xi_q) = q < F(\xi_q + \eta)$ . Entonces, para el  $q$ -ésimo cuantil muestral<sup>22</sup>  $X_{n:k}$ , con  $k = k_q$  tal que:

$$k = k_q = \lfloor nq \rfloor + 1 \quad \text{o} \quad k = k_q = \lfloor (n + 1)q \rfloor,$$

donde la función  $\lfloor \cdot \rfloor$  indica parte entera, se tiene que:

$$X_{n:k} \xrightarrow{c.s.} \xi_q.$$

<sup>22</sup>La notación  $X_{n:k} = X_{(k)}$ , es decir, es el estadístico de orden  $k$ . Si  $\frac{k}{n}$  converge a  $q$  de alguna forma, entonces  $X_{n:k}$  es llamado el  $q$ -ésimo cuantil muestral (ver [82, p. 158]).

*Demostración.* Ver detalles de la prueba en [82, p. 167]. ☑

**Teorema 3.26.** De acuerdo con [82, p. 168], dada una muestra aleatoria  $X_1, X_2, \dots, X_n$  de una variable aleatoria con *fda*  $F$  y *fdp*  $f$ , continua en  $\xi_q$ ,  $0 < q < 1$ , y tal que  $f(\xi_q) > 0$ ; definiendo  $k$  de modo que  $k = nq + o(n^{1/2})$ . Entonces,

$$\sqrt{n}(X_{n:k} - \xi_q) \xrightarrow{D} N\left(0, \frac{q(1-q)}{f^2(\xi_q)}\right).$$

*Demostración.* Ver detalles de la prueba en [82, p. 168]. ☑

**Teorema 3.27.** De acuerdo con [82, p. 169], dada una muestra aleatoria  $X_1, X_2, \dots, X_n$  de una variable aleatoria con *fda*  $F$  y *fdp*  $f$  continua en  $\xi_{q_j}$ ,  $j = 1, 2$ ;  $0 < q_1 < q_2 < 1$ , y tal que  $f(\xi_{q_j}) > 0$ ,  $j = 1, 2$ . Si  $k_j = nq_j + o(n^{1/2})$ ,  $j = 1, 2$ . Entonces,

$$\sqrt{n} \begin{pmatrix} X_{n:k_1} - \xi_{q_1} \\ X_{n:k_2} - \xi_{q_2} \end{pmatrix} \xrightarrow{D} N_2(\mathbf{0}, \Sigma).$$

con:

$$\Sigma = \begin{pmatrix} \frac{q_1(1-q_1)}{f^2(\xi_{q_1})} & \frac{q_1(1-q_2)}{f(\xi_{q_1})f(\xi_{q_2})} \\ \frac{q_1(1-q_2)}{f(\xi_{q_1})f(\xi_{q_2})} & \frac{q_2(1-q_2)}{f^2(\xi_{q_2})} \end{pmatrix}.$$

*Demostración.* Ver detalles de la prueba en [82, p. 170]. ☑

**Teorema 3.28.** De acuerdo con [82, p. 173], dada una muestra aleatoria  $X_1, X_2, \dots, X_n$  de una variable aleatoria con *fda*  $F$ . Si  $F$  tiene un punto extremo superior en  $\xi_1$ , entonces  $X_{n:n} \xrightarrow{c.s.} \xi_1$ . Si  $F$  tiene un punto extremo inferior en  $\xi_0$ , entonces  $X_{n:1} \xrightarrow{c.s.} \xi_0$ .

*Demostración.* Ver detalles de la prueba en [82, p. 174]. ☑

## 3.5. Ejercicios

### Conceptos básicos de muestra aleatoria

1. Sean  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una distribución uniforme sobre el intervalo  $(0, \theta)$ . ¿Cuál es la *fdp* del mínimo, esto es,  $f_{X_{(1)}}(x)$ ?
2. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid con *fdp* [20, p. 259]:

$$f_X(x) = \begin{cases} \frac{a}{\theta^a} x^{a-1} & \text{si } 0 < x < \theta \\ 0 & \text{en otro caso.} \end{cases}$$

¿Cuál es la **fdp** del mínimo, esto es,  $f_{X_{(1)}}(x)$ ?, ¿cuál es la **fdp** del máximo, esto es,  $f_{X_{(n)}}(x)$ ?

3. En el ejercicio 2., ¿cuál es la **fdp** conjunta del máximo y el mínimo, esto es,  $f_{X_{(1)}, X_{(n)}}(u, v)$ ?
4. Muestre que la expresión (3.10) se puede expresar usando (1.64) como:

$$F_{X_{(j)}}(x) = \sum_{k=j}^n \binom{n}{k} [F_X(x)]^k (1 - F_X(x))^{n-k} = I(F_X(x); j, r), \quad (3.27)$$

donde  $r = n - j + 1$ .

5. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias **iid** continuamente con **fdp**  $f_X(x)$  y **fda**  $F_X(x)$ . Sean los estadísticos de orden para esta muestra aleatoria  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ . Muestre que para cualquier dos valores  $a$  y  $b$  con  $a < b$ :

$$\begin{aligned} P(a \leq X_{(j)} \leq b) &= \frac{n!}{(j-1)!(n-j)!} \int_{F_X(a)}^{F_X(b)} u^{j-1} (1-u)^{n-j} du \\ &= I(F_X(b); j, r) - I(F_X(a); j, r), \end{aligned} \quad (3.28)$$

donde  $I(y; a, b)$  es dada en (1.64) y  $r = n - j + 1$ .

### Órdenes de magnitud

6. Sean  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias y  $\{a_n\}_{n \geq 1}$ ,  $\{b_n\}_{n \geq 1}$  sucesiones de números reales [55, p. 46]. Muestre que:
  - a) Si  $a_n = O(b_n)$  y  $X_n = O_p(a_n)$ , entonces  $X_n = O_p(b_n)$ .
  - b) Si  $a_n = o(b_n)$  y  $X_n = O_p(a_n)$ , entonces  $X_n = o_p(b_n)$ .
  - c) Si  $a_n = O(b_n)$  y  $X_n = o_p(a_n)$ , entonces  $X_n = o_p(b_n)$ .
  - d) Si  $a_n = o(b_n)$  y  $X_n = o_p(a_n)$ , entonces  $X_n = o_p(b_n)$ .
7. Sean  $\{a_n\}_{n \geq 1}$  y  $\{b_n\}_{n \geq 1}$  sucesiones de números reales. Muestre que:
  - a)  $a_n = O(b_n) \iff a_n = b_n O(1) \iff \frac{a_n}{b_n} = O(1)$ .
  - b)  $a_n = o(b_n) \iff a_n = b_n o(1) \iff \frac{a_n}{b_n} = o(1)$ .

### Convergencia estocástica

8. Sea  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias tales que  $X_n \sim \exp(n)$  [16, p. 312]. Demostrar que  $X_n \xrightarrow{P} 0$ .
9. Suponga, como en [21, p. 35], que  $X_1, X_2, X_3, \dots, X_n$  son variables aleatorias **iid**, tales que  $X_k \sim B(1, p)$ . Encuentre  $c$  tal que

$$P(|\bar{X} - 0,5| < c) > 0,5 \quad \text{con } p = 0,5.$$

- a) Utilizando el teorema central de límite.
  - b) Aplicando Chebyshev.
10. Una compañía compra grandes cantidades de componentes electrónicos. La decisión para rechazar o aceptar un lote de componentes se toma con base en una muestra aleatoria de 100 unidades [19, p. 123]. Si el lote se rechaza al encontrar tres o más unidades defectuosas en la muestra:
- a) ¿Cuál es la probabilidad de rechazar un lote si este contiene un 1 % de componentes defectuosos?
  - b) ¿Cuál es la probabilidad de rechazar un lote si este contiene un 8 % de unidades defectuosas?
11. En un estudio se encontró que la probabilidad de morir por causa de una cierta vacuna es 0,00002 [19, p. 124]. Si se administra la vacuna a 100 000 personas y se supone que ellas constituyen un conjunto independiente de ensayos:
- a) ¿Cuál es la probabilidad de que mueran no más de dos personas a causa de la vacuna?
  - b) ¿Cuál es la probabilidad de que muera más de una persona a causa de la vacuna?







Supóngase que  $X$  es una variable aleatoria con fmp o fdp  $f_X(x; \theta)$ , pero con  $\theta$  desconocido. En este capítulo se aborda la pregunta ¿cómo obtener una estimación puntual para el parámetro  $\theta$ ?, y se presentan algunos métodos para estimar dicho parámetro. El procedimiento de la estimación puntual busca, a partir de una muestra  $X_1, \dots, X_n$ , un estadístico  $T(X_1, \dots, X_n)$  que mejor estime el parámetro  $\theta$ , es decir, una función de los datos muestrales; esto es, una fórmula que depende de los datos obtenidos de una muestra, con la cual se obtiene el valor exacto del parámetro  $\theta$ .

Existen varios métodos que nos van a permitir obtener los estimadores puntuales. Este capítulo inicia presentando tres métodos para obtener la estimación puntual del parámetro  $\theta$ : principio de sustitución, método de máxima verosimilitud y método bayesiano.

**DEFINICIÓN 4.1.** (Estimador). *Un **estimador de**  $g(\theta)$  es cualquier estadístico, digamos  $\delta(\mathbf{X})$ , cuyos valores son usados para estimar  $g(\theta)$ , donde  $g(\cdot)$  es alguna función del parámetro  $\theta$  (ver [65, p. 273]).*

Se espera que  $\delta(\mathbf{X})$  tienda a estar cerca de  $g(\theta)$  (que es desconocido), aunque este requerimiento no sea parte de la definición de estimador. Sea  $\delta(\mathbf{x})$  un valor de  $\delta(\mathbf{X})$  para el valor (vector) observado  $\mathbf{x}$  de la variable aleatoria (vector aleatorio)  $\mathbf{X}$ . A  $\delta(\mathbf{x})$  se le llama la *estimativa* de  $g(\theta)$ , es decir, la estimativa será un número.

## 4.1. Principio de sustitución

En esta sección se presenta un resultado relevante de la convergencia de  $F_n$  a  $F$ , la definición de estimadores mediante el principio de sustitución el cual considera dos métodos:

- Sustitución de frecuencias.
- Método de momentos.

### 4.1.1. Estimación por sustitución de frecuencias

Suponga que se desea estimar una función continua  $q(\theta)$ . Si  $p_1, \dots, p_k$  son funciones continuas de  $\theta$ , se puede expresar  $q(\theta)$  como una función continua de  $p_1, \dots, p_k$ :

$$q(\theta) = h(p_1(\theta), \dots, p_k(\theta)),$$

con  $h$  definida y continua sobre:

$$J_k = \left\{ (p_1, \dots, p_k) : p_i \geq 0, \sum_{i=1}^k p_i = 1 \right\}.$$

El *estimador por sustitución de frecuencias* de  $q(\theta)$  se define como:

$$T(\mathbf{X}) = T \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = h \begin{pmatrix} \frac{N_1}{N} \\ \frac{N_2}{N} \\ \vdots \\ \frac{N_k}{N} \end{pmatrix}.$$

**Ejemplo 4.1.** De acuerdo con [12, p. 103], al muestrear de una población en equilibrio genético respecto a un gen con dos alelos  $A$  y  $a$ . Las posibilidades para los genes de los progenitores se muestran en el siguiente cuadro de Punnett:

Alelos	$A$	$a$
$A$	$AA$	$Aa$
$a$	$aA$	$aa$

Nótese que hay tres diferentes genotipos identificables,  $AA$ ,  $Aa$  y  $aa$ , los cuales ocurren de acuerdo con las proporciones de Hardy-Weinberg (ver [48, p. 301]):

$j$	$AA$	$Aa$	$aa$
$p_j$	$\theta^2$	$2\theta(1-\theta)$	$(1-\theta)^2$

con  $0 < \theta < 1$ . Encuentre algunos estimadores por frecuencias para  $\theta$ .

*Solución.* Si se estiman las proporciones  $p_1$ ,  $p_2$  y  $p_3$  usando las proporciones muestrales de cada genotipo, es decir,

$$\hat{p}_j = \frac{n_j}{n}, \quad j = 1, 2, 3,$$

donde  $n_j$  es el número de individuos del tipo  $j$  en una muestra de tamaño  $n = n_1 + n_2 + n_3$ . Luego,

- Para  $j = 1$  se obtiene el estimador:

$$p_1 = \theta^2 \quad \rightarrow \quad \theta = \sqrt{p_1} \quad \rightarrow \quad \hat{\theta}_1 = \sqrt{\hat{p}_1} \quad \rightarrow \quad \hat{\theta}_1 = \sqrt{\frac{n_1}{n}}.$$

- Si  $j = 2$  se tiene que:

$$p_2 = 2\theta(1-\theta) \quad \rightarrow \quad \frac{p_2}{2} = \theta - \theta^2 \quad \rightarrow \quad \theta^2 + (-\theta) + \frac{p_2}{2} = 0,$$

aplicando cuadrática (ver nota B.7):

$$\theta = \frac{1 \pm \sqrt{1 - 2p_2}}{2} \quad \rightarrow \quad \hat{\theta}_2 = \frac{1 \pm \sqrt{1 - 2\hat{p}_2}}{2}$$

$$\hat{\theta}_2 = \frac{1 \pm \sqrt{1 - 2\frac{n_2}{n}}}{2}.$$

- Para  $j = 3$  se obtiene el estimador:

$$\theta = 1 - \sqrt{p_3} \quad \rightarrow \quad \hat{\theta}_3 = 1 - \sqrt{\hat{p}_3} \quad \rightarrow \quad \hat{\theta}_3 = 1 - \sqrt{\frac{n_3}{n}}.$$

- Un cuarto estimador posible sería  $\hat{\theta}_4 = \frac{n_1}{n} + \frac{n_2}{2n}$ , veamos:

$$\hat{\theta}_4 = \underbrace{\frac{n_1}{n}}_{\hat{p}_1} + \underbrace{\frac{n_2}{2n}}_{\frac{\hat{p}_2}{2}} \quad \rightarrow \quad p_1 + \frac{p_2}{2} = \theta^2 + \theta(1 - \theta) = \theta. \quad (4.1)$$

Usando (4.1), un estimador por sustitución de frecuencias del *odds ratio* o razón de chances,  $\frac{\theta}{1-\theta}$ , es dado por:

$$\frac{\hat{\theta}}{1 - \hat{\theta}} = \frac{\frac{n_1}{n} + \frac{n_2}{2n}}{1 - \left(\frac{n_1}{n} + \frac{n_2}{2n}\right)} = \frac{2n_1 + n_2}{2n - 2n_1 - n_2} = \frac{2n_1 + n_2}{2n_3 + n_2}, \quad (4.2)$$

la última expresión se obtuvo reemplazando el tamaño de muestra.  $\checkmark$

### 4.1.2. Estimación por método de momentos

Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid, con la misma fdp o con fmp  $p(x; \boldsymbol{\theta})$  con  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)^\top$  vector de dimensión  $k$ . Suponga que  $m_1(\boldsymbol{\theta}), m_2(\boldsymbol{\theta}), \dots, m_k(\boldsymbol{\theta})$  son los primeros  $k$  momentos de la población de la cual se está muestreando. Entonces,

$$m_j(\boldsymbol{\theta}) = \mathbb{E}_\theta(X_1^j) \quad 1 \leq j \leq k.$$

Sea  $\hat{m}_j$  el  $j$ -ésimo momento muestral dado por (ver [20, p. 312]):

$$\hat{m}_j = \frac{1}{n} \sum_{i=1}^n X_i^j \quad 1 \leq j \leq k.$$

Suponga que se desea estimar una función continua  $q(\boldsymbol{\theta})$  y que se puede expresar como una función que depende de  $m_1(\boldsymbol{\theta}), m_2(\boldsymbol{\theta}), \dots, m_k(\boldsymbol{\theta})$ , es

decir,

$$q(\boldsymbol{\theta}) = g \begin{bmatrix} m_1(\boldsymbol{\theta}) \\ m_2(\boldsymbol{\theta}) \\ \vdots \\ m_k(\boldsymbol{\theta}) \end{bmatrix}.$$

El método de los momentos consiste en estimar  $q(\boldsymbol{\theta})$  usando:

$$T(\mathbf{X}) = g \begin{bmatrix} \hat{m}_1 \\ \hat{m}_2 \\ \vdots \\ \hat{m}_k \end{bmatrix}.$$

Según [20, p. 313], el estimador por el método de los momentos (**EMM**) de  $\boldsymbol{\theta}$ ,  $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)^\top$  se obtiene al resolver el sistema de ecuaciones para  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)^\top$  en términos de  $(\hat{m}_1, \hat{m}_2, \dots, \hat{m}_k)^\top$ :

$$\hat{m}_j = m_j(\boldsymbol{\theta}) \quad \text{para} \quad j = 1, 2, \dots, k.$$

**Ejemplo 4.2.** Sea  $q(\boldsymbol{\theta}) = \mathbb{V}(X) = m_2(\boldsymbol{\theta}) - m_1^2(\boldsymbol{\theta})$ . El **EMM** de  $q(\boldsymbol{\theta})$  es:

$$T(\mathbf{X}) = \hat{m}_2 - \hat{m}_1^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left( \frac{1}{n} \sum_{i=1}^n X_i \right)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2.$$

**Ejemplo 4.3.** Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de la distribución exponencial de parámetro  $\lambda$ ,  $X_1 \sim \exp(\lambda)$ , encuentre diferentes estimadores para  $\lambda$  usando el método de los momentos.

*Solución.* Por la expresión (1.27) se tiene que:

$$f(x; \lambda) = \lambda e^{-\lambda x}, \quad 0 < x < \infty, \quad \lambda > 0,$$

donde,

$$\mathbb{E}(X_1) = \frac{1}{\lambda} \quad \text{y} \quad \mathbb{V}(X_1) = \frac{1}{\lambda^2}.$$

Luego, algunos estimadores para  $\lambda$  son:

1. Tomando  $q_1(\lambda) = m_1 = \frac{1}{\lambda}$ , entonces:

$$\lambda = \frac{1}{m_1} \quad \rightarrow \quad \hat{\lambda}_1 = \left( \frac{1}{n} \sum_{i=1}^n X_i \right)^{-1} = \frac{1}{\bar{X}}.$$

2. Tomando  $q_2(\lambda) = m_2 - m_1^2 = \frac{1}{\lambda^2}$ , entonces:

$$\lambda^2 = \frac{1}{m_2 - m_1^2} \rightarrow \lambda = \frac{1}{\sqrt{m_2 - m_1^2}} \rightarrow \hat{\lambda}_2 = \frac{1}{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2}}.$$

Usando  $\hat{\lambda}_1$ , un estimador, por ejemplo, de  $P_\lambda(X_1 \geq 1)$  es dado por:

$$p = P_\lambda(X_1 \geq 1) = \int_1^\infty \lambda e^{-\lambda x} dx = \lambda \int_1^\infty e^{-\lambda x} dx.$$

Esta integral se evalúa mediante cambio de variable, sea  $u = \lambda x$  entonces  $du = \lambda dx$ , ahora si  $x = 1$  entonces  $u = \lambda$  y si  $x \rightarrow \infty$  entonces  $u \rightarrow \infty$ ; luego:

$$p = \int_\lambda^\infty e^{-u} du = \lim_{a \rightarrow \infty} [-e^{-u}]_\lambda^a = - \lim_{a \rightarrow \infty} [e^{-u}]_\lambda^a = - \lim_{a \rightarrow \infty} [e^{-a} - e^{-\lambda}] = e^{-\lambda},$$

entonces,

$$\hat{p} = e^{-\hat{\lambda}} = e^{-1/\hat{X}}. \quad \checkmark$$

**Ejemplo 4.4.** De acuerdo con [20, p. 313], si  $X_1, X_2, \dots, X_n$  son variables aleatorias *iid* binomiales con parámetros  $m$  y  $p$  desconocidos,  $B(m, p)$ , determine la estimación de los parámetros.

*Solución.* Puesto que:

$$P(X_1 = x; m, p) = \binom{m}{x} p^x (1-p)^{m-x} \quad x = 0, 1, 2, \dots, m.$$

Se desean estimadores puntuales para los parámetros,  $m$  y  $p$ . Se sabe que:

$$\mathbb{E}(X_1) = mp \quad \text{y} \quad \mathbb{V}(X_1) = mp(1-p).$$

Tomando  $\boldsymbol{\theta} = (m, p)^\top$ , se tiene que:

$$\begin{aligned} m_1(\boldsymbol{\theta}) &= mp & m_2(\boldsymbol{\theta}) - m_1^2(\boldsymbol{\theta}) &= mp(1-p) \\ m_2(\boldsymbol{\theta}) &= mp(1-p) + m_1^2(\boldsymbol{\theta}) = mp(1-p) + m^2 p^2. \end{aligned}$$

Luego, el sistema de ecuaciones para  $\boldsymbol{\theta} = (m, p)^\top$  en términos de  $(\hat{m}_1, \hat{m}_2)^\top$  es dado por:

$$\hat{m}_1 = \hat{m}\hat{p} \quad (4.3)$$

$$\hat{m}_2 = \hat{m}\hat{p}(1-\hat{p}) + \hat{m}^2 \hat{p}^2, \quad (4.4)$$

de (4.3) se tiene que:

$$\hat{m}_1 = \hat{m}\hat{p} \rightarrow \frac{1}{n} \sum_{i=1}^n X_i = \hat{m}\hat{p} \rightarrow \hat{p} = \frac{\bar{X}}{\hat{m}}, \quad (4.5)$$

reemplazando (4.5) en (4.4) se obtiene que:

$$\begin{aligned} \hat{m}_2 &= \hat{m}\hat{p}(1 - \hat{p}) + \hat{m}^2\hat{p}^2 = \hat{m}\hat{p}(1 - \hat{p}) + \bar{X}^2 \\ \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 &= \cancel{\hat{m}} \frac{\bar{X}}{\cancel{\hat{m}}} \left[ 1 - \frac{\bar{X}}{\hat{m}} \right] = \bar{X} \left[ 1 - \frac{\bar{X}}{\hat{m}} \right] \\ 1 - \frac{\bar{X}}{\hat{m}} &= \frac{S_n^2}{\bar{X}} \rightarrow \frac{\bar{X}}{\hat{m}} = 1 - \frac{S_n^2}{\bar{X}} \rightarrow \hat{m} = \frac{\bar{X}^2}{\bar{X} - S_n^2}. \end{aligned}$$

Luego, se pueden obtener valores negativos para  $\hat{m}$  y  $\hat{p}$ . Por tanto, el rango del estimador no coincide con el rango del parámetro.  $\square$

En la tabla 4.1 se presentan los EMM de los parámetros de algunas distribuciones.

Tabla 4.1. EMM de los parámetros de algunas distribuciones

Distribución	Parámetro(s)	Estimador(es) <sup>a</sup>
Uniforme discreta	$n$	$\hat{n} = 2\bar{X} - 1$
Binomial	$m$ y $p$	$\hat{m} = \left\lfloor \frac{\bar{X}^2}{\bar{X} - S_n^2} \right\rfloor$ y $\hat{p} = \frac{\bar{X} - S_n^2}{\bar{X}}$
Poisson	$\lambda$	$\hat{\lambda} = \bar{X}$
Geométrica	$p$	$\hat{p} = \frac{1}{1 + \bar{X}}$
Uniforme continua	$a$ y $b$	$\hat{a} = \bar{X} - \sqrt{3}S_n$ y $\hat{b} = \bar{X} + \sqrt{3}S_n$
Normal	$\mu$ y $\sigma^2$	$\hat{\mu} = \bar{X}$ y $\hat{\sigma}^2 = S_n^2$
Gamma	$r$ y $\lambda$	$\hat{r} = \frac{\bar{X}^2}{S_n^2}$ y $\hat{\sigma}^2 = \frac{\bar{X}}{S_n^2}$
Exponencial	$\lambda$	$\hat{\lambda} = \frac{1}{\bar{X}}$
chi-cuadrada	$k$	$\hat{k} = \bar{X}$

<sup>a</sup> Donde  $S_n^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2$ .

## 4.2. Estimación por máxima verosimilitud

El método de máxima verosimilitud es el más usado para encontrar estimadores. Este consiste en encontrar el valor del parámetro o vector de parámetros,  $\theta$ , más probable, dados los datos observados.

**DEFINICIÓN 4.2.** Sea  $p(x; \theta)$  la *fmp* o *fdp* conjunta de una muestra aleatoria  $X = (X_1, X_2, \dots, X_n)^\top$ , con  $x = (x_1, x_2, \dots, x_n)^\top$  y  $\theta = (\theta_1, \theta_2, \dots, \theta_k)^\top$ . Entonces, dado que se observa  $X = x$ , la función de  $\theta$  definida por

$$L(\theta; x) = p(x; \theta), \quad (4.6)$$

se llama la **función de verosimilitud** (ver [20, p. 290]).

**Ejemplo 4.5.** De acuerdo con [21, p. 48], si  $X$  es una variable aleatoria con distribución hipergeométrica de parámetros  $N\theta = 5\theta$ ,  $N = 5$  y  $n = 3$ , entonces,

$$p(x; \theta) = \begin{cases} \frac{\binom{N\theta}{x} \binom{N-N\theta}{n-x}}{\binom{N}{n}} & \text{si } \max\{0, n - N(1 - \theta)\} \leq x \leq \min\{N\theta, n\} \\ 0 & \text{en otro caso} \end{cases} \quad (4.7)$$

Si se sabe que  $\theta$  toma los valores del conjunto  $\Theta = \{0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1\}$ , determine el valor de  $\theta$  que maximiza la expresión (4.6) si se observa  $x = 1$ .

*Solución.* En este caso,

Valor de $\theta$	Rango de $x$ [ $\max\{0, 5\theta - 2\}$ , $\min\{5\theta, 3\}$ ]
0	[0,0]
$\frac{1}{5}$	[0,1]
$\frac{2}{5}$	[0,2]
$\frac{3}{5}$	[1,3]
$\frac{4}{5}$	[2,3]
1	[3,3]

Como  $x = 1$ , se descartan los valores de  $\theta$  iguales a 0,  $\frac{4}{5}$  y a 1, ya que  $x = 1$  no está en el rango de los posibles valores de  $x$ . Dado que  $x = 1$  los posibles valores de  $\theta$  son  $\frac{1}{5}$ ,  $\frac{2}{5}$  y  $\frac{3}{5}$ . Para encontrar, de estos tres, el valor más probable de  $\theta$  dado que se observó  $x = 1$  se sustituye en (4.7) y se obtiene que:

$$p(1; \theta) = L(\theta; x = 1) = \frac{5}{4}\theta(1 - \theta)(4 - 5\theta),$$

evaluando la verosimilitud para cada uno de los valores de  $\theta$ :

$\theta$	$L(\theta; x = 1)$
$\frac{1}{5}$	$\frac{3}{5}$
$\frac{2}{5}$	$\frac{3}{5}$
$\frac{3}{5}$	$\frac{3}{10}$

Luego, la estimación de máxima verosimilitud de  $\theta$  es  $\hat{\theta} = \frac{1}{5}$  o  $\hat{\theta} = \frac{2}{5}$ . ✓

**DEFINICIÓN 4.3.** (Principio de verosimilitud). Si  $x$  y  $y$  son dos puntos muestrales tales que  $L(\theta; x)$  es proporcional a  $L(\theta; y)$ , es decir, existe una constante  $k$ , tal que

$$L(\theta; x) = kL(\theta; y)$$

para todo  $\theta$ . Luego, las conclusiones que se extraen de  $x$  y  $y$  sobre  $\theta$  son equivalentes, en particular si  $k = 1$  las conclusiones que se extraen de  $x$  y  $y$  sobre  $\theta$  son idénticas (ver [20, p. 291]).

**Nota 4.1.** Del principio de verosimilitud se observa que:

- Si dos puntos muestrales,  $x$  y  $y$ , tienen la misma función de verosimilitud, es decir,  $k = 1$ , entonces  $x$  y  $y$  contienen la misma información acerca de  $\theta$ .
- La función de verosimilitud se usa para comparar si un valor candidato para  $\theta$ , digamos  $\theta_1$ , es más **factible** que otro valor candidato para  $\theta$ , digamos  $\theta_2$ . Por ejemplo, si  $L(\theta_1; x) = 2L(\theta_2; x)$  se tiene que  $\theta_1$  es dos veces más factible que  $\theta_2$ .
- Considerando que el vector de parámetros  $\theta$  es un valor fijo (aunque sea desconocido), se debe usar palabras como factible o plausible o verosímil, en vez de la palabra probable, ya que  $L(\theta; x)$  a pesar de ser una **fdp** o una **fmp** como función de  $x$ , puede no ser una **fdp** o **fmp** como función de  $\theta$ .

Sean  $X_1, X_2, \dots, X_n$  variables aleatorias **iid** de una población con **fdp** o con **fmp**  $p(x; \theta)$ , donde  $x = (x_1, x_2, \dots, x_n)^\top$  y  $\theta = (\theta_1, \theta_2, \theta_3, \dots, \theta_k)^\top$ . La función de verosimilitud es dada por:

$$L(\theta; x) = \prod_{i=1}^n p(x_i; \theta_1, \dots, \theta_k). \quad (4.8)$$

**DEFINICIÓN 4.4.** Para cada punto muestral  $x$ , sea  $\hat{\theta}(x)$  un valor del parámetro tal que  $L(\theta; x)$  alcanza su máximo como una función de  $\theta$ , mientras  $x$  permanece fijo. Un estimador vía máxima verosimilitud (**EMV**) del parámetro  $\theta$  basado en una

muestra aleatoria  $X$  es dado por  $\hat{\theta}(X)$  (ver [20, p. 316]). En otras palabras si  $X = x$ , se puede buscar  $\hat{\theta}(X)$  tal que satisfice:

$$\begin{aligned} L(\hat{\theta}(x); x) &= p(x; \hat{\theta}(x)) = \text{máx}\{p(x; \theta) : \theta \in \Theta\} \\ &= \text{máx}\{L(\theta; x) : \theta \in \Theta\}. \end{aligned}$$

Si existe  $\hat{\theta}(x)$ , cualquier función  $q(\theta)$  se puede estimar por  $q(\hat{\theta})$ .

**Nota 4.2.**

- El **EMV** da una estimación puntual para el parámetro que según la muestra observada es más factible (o plausible o verosimil).
- Por construcción, el rango del **EMV** coincide con el rango del parámetro.
- Solo tiene sentido calcular el **EMV** en modelos paramétricos regulares.
- El **EMV** puede o no existir o no ser único.
- En general hay dos inconvenientes asociados con el problema de encontrar el máximo de una función:
  1. Encontrar el máximo y verificar que efectivamente lo es.
  2. La sensibilidad numérica del estimador encontrado: ¿qué tanto cambia el estimador ante pequeños cambios en los datos?

Suponga que  $k = 1$ , es decir,  $\theta = \theta$  y sean  $\theta_1, \dots, \theta_k$  posibles candidatos de  $\theta$  que cumplen con que

$$\left. \frac{\partial L(\theta; x)}{\partial \theta} \right|_{\theta=\theta_i} = 0.$$

¿Por qué posibles candidatos? Porque ser un valor en el que la primera derivada se anula es una condición necesaria para ser máximo pero no suficiente (ver apéndice B.2.1).

**Ejemplo 4.6.** Determine un **EMV** para la proporción  $\theta$  del ejemplo 4.1.

*Solución.* La ley de Hardy-Weinberg es una distribución trinomial (caso especial de (2.45)) con probabilidades  $p_1 = \theta^2, p_2 = 2\theta(1-\theta), p_3 = (1-\theta)^2$ , considerando  $n_k = x_k$ , la función de verosimilitud queda:

$$\begin{aligned} L(\theta; x) &= \frac{n!}{x_1!x_2!x_3!} \theta^{2x_1} (2\theta(1-\theta))^{x_2} (1-\theta)^{2x_3} \\ \ell(\theta; x) &= \ln(n!) - \sum_{i=1}^3 \ln(x_i!) + 2x_1 \ln \theta + x_2 \ln(2\theta(1-\theta)) \\ &\quad + 2x_3 \ln(1-\theta), \end{aligned}$$

derivando con respecto a  $\theta$  se tiene que:

$$\frac{\partial \ell(\theta; \mathbf{x})}{\partial \theta} = \frac{2x_1 + x_2}{\theta} - \frac{2x_3 + x_2}{1 - \theta},$$

igualando esta última expresión a cero se obtiene:

$$\frac{2x_1 + x_2}{\hat{\theta}} - \frac{2x_3 + x_2}{1 - \hat{\theta}} = 0 \quad \Rightarrow \quad \frac{\hat{\theta}}{1 - \hat{\theta}} = \frac{2x_1 + x_2}{2x_3 + x_2}. \quad (4.9)$$

La expresión (4.9) coincide con la razón de chances encontrada en el ejemplo 4.1, expresión (4.2). Despejando  $\theta$  se llega a:

$$\hat{\theta} = \frac{2x_1 + x_2}{2n}.$$

Este **EMV** es igual a la estimación dada en (4.1) del ejemplo 4.1.  $\square$

**Ejemplo 4.7.** De acuerdo con [20, p. 316], dadas las variables aleatorias  $X_1, X_2, \dots, X_n$  *iid*  $N(\theta, 1)$ . Encuentre un **EMV** para  $\theta$ .

*Solución.* La función de verosimilitud es dada por:

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x_i - \theta)^2\right\} = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2\right\},$$

derivando se tiene que:

$$\begin{aligned} \frac{\partial L(\theta; \mathbf{x})}{\partial \theta} &= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2\right\} \left(-\frac{\partial}{\partial \theta} \sum_{i=1}^n (x_i - \theta)(-1)\right) \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2\right\} \left(\sum_{i=1}^n (x_i - \theta)\right), \end{aligned}$$

igualando a cero se obtiene:

$$\underbrace{\frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \hat{\theta})^2\right\}}_{>0} \left(\sum_{i=1}^n (x_i - \hat{\theta})\right) = 0$$

$$\sum_{i=1}^n x_i - n\hat{\theta} = 0 \quad \Rightarrow \quad \hat{\theta} = \bar{x}.$$

Luego,  $\hat{\theta} = \bar{x}$  es un candidato a ser **EMV**. Para verificarlo se toma la segunda derivada de  $L(\theta; \mathbf{x})$  con respecto a  $\theta$  y se evalúa en  $\theta = \bar{x}$ :

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} L(\theta; \mathbf{x}) &= \frac{\partial}{\partial \theta} \left[ \frac{1}{(2\pi)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2 \right\} \left( \sum_{i=1}^n x_i - n\theta \right) \right] \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2 \right\} \left[ \sum_{i=1}^n x_i - n\theta \right]^2 + \\ &\quad \frac{1}{(2\pi)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2 \right\} (-n) \\ \frac{\partial^2}{\partial \theta^2} L(\theta; \mathbf{x}) \Big|_{\theta=\bar{x}} &= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 \right\} \left[ \left( \sum_{i=1}^n x_i - n\bar{x} \right)^2 - n \right] \\ &= \frac{-n}{(2\pi)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 \right\} < 0. \end{aligned}$$

Entonces,  $\bar{x}$  es un extremo en el interior y es máximo.  $\square$

En la mayoría de los casos, es más fácil trabajar con el logaritmo natural de  $L(\theta; \mathbf{x})$ ,  $\ell(\theta; \mathbf{x}) = \ln [L(\theta; \mathbf{x})]$ , conocido como la **log-verosimilitud**. Se puede usar  $\ell(\theta; \mathbf{x})$  porque la función logaritmo es estrictamente creciente en  $(0, \infty)$ , lo cual implica que el extremo de  $L(\theta; \mathbf{x})$  y  $\ell(\theta; \mathbf{x})$  coinciden.

**Ejemplo 4.8.** Sean  $X_1, X_2, \dots, X_n$ , variables aleatorias *iid* Bernoulli( $\theta$ ), encuentre un **EMV** para  $\theta$ .

*Solución.* En este caso,

$$\begin{aligned} L(\theta; \mathbf{x}) &= \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} \\ \ell(\theta; \mathbf{x}) &= \left( \sum_{i=1}^n x_i \right) \ln \theta + \left( n - \sum_{i=1}^n x_i \right) \ln(1 - \theta). \end{aligned}$$

Si  $0 < \sum_{i=1}^n x_i < n$ , entonces,

$$\begin{aligned} \frac{\partial[\ell(\theta; \mathbf{x})]}{\partial \theta} &= \frac{\sum_{i=1}^n x_i}{\theta} + (-1) \frac{\left( n - \sum_{i=1}^n x_i \right)}{1 - \theta} = \frac{n\bar{x}}{\theta} - \frac{n - n\bar{x}}{1 - \theta} \\ &= n \left( \frac{\bar{x}}{\theta} - \frac{1 - \bar{x}}{1 - \theta} \right) = n \left( \frac{\bar{x} - \bar{x}\theta - \theta + \bar{x}\theta}{\theta - \theta^2} \right) = n \left( \frac{\bar{x} - \theta}{\theta - \theta^2} \right), \end{aligned}$$

igualando a cero se tiene que:

$$n \left( \frac{\bar{x} - \hat{\theta}}{\hat{\theta} - \hat{\theta}^2} \right) = 0 \quad \rightarrow \quad \bar{x} - \hat{\theta} = 0 \quad \rightarrow \quad \hat{\theta} = \bar{x}.$$

Luego  $\hat{\theta} = \bar{x}$  es un candidato para ser **EMV**. Para verificarlo se toma la segunda derivada de  $\ell(\theta; \mathbf{x})$  con respecto a  $\theta$  y se evalúa en  $\theta = \bar{x}$ :

$$\begin{aligned} \frac{\partial^2[\ell(\theta; \mathbf{x})]}{\partial \theta^2} &= \frac{\partial}{\partial \theta} \left[ n \left( \frac{\bar{x} - \theta}{\theta - \theta^2} \right) \right] = n \frac{(-1)\theta(1-\theta) - (\bar{x} - \theta)(1-2\theta)}{\theta^2(1-\theta)^2} \\ &= -n \frac{\theta(1-\theta) + (\bar{x} - \theta)(1-2\theta)}{\theta^2(1-\theta)^2} \\ \left. \frac{\partial^2[\ell(\theta; \mathbf{x})]}{\partial \theta^2} \right|_{\theta=\bar{x}} &= -n \frac{\bar{x}(1-\bar{x}) + (\bar{x} - \bar{x})(1-2\bar{x})}{\bar{x}^2(1-\bar{x})^2} = \frac{-n}{\bar{x}(1-\bar{x})} < 0, \end{aligned}$$

puesto que  $0 < \bar{x} < 1$ , se tiene entonces que  $\hat{\theta} = \bar{x}$  es un máximo.  $\square$

**Ejemplo 4.9.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias *iid* binomial negativa (**BN**) de parámetros  $r$  y  $\theta$ , con *fmp* dada por (1.16). Encuentre los parámetros mediante **EMM** y **EMV**.

*Solución.* Para el **EMM** se tiene que:

$$\hat{r} \frac{1 - \hat{\theta}}{\hat{\theta}} = \bar{X} \quad \text{y} \quad \hat{r} \frac{1 - \hat{\theta}}{\hat{\theta}^2} = S_n^2.$$

Luego,

$$\hat{\theta} = \frac{\bar{X}}{S_n^2} = \frac{1}{\mathcal{J}_X} \quad \text{y} \quad \hat{r} = \frac{\bar{X}^2}{S_n^2 - \bar{X}} = \frac{\bar{X}}{\mathcal{J}_X - 1}, \quad (4.10)$$

donde  $\mathcal{J}_X$  es el índice de dispersión muestral de  $X$ . Para el **EMV** se usa la expresión (1.18) y se obtiene que la función de verosimilitud es:

$$L(r, \beta) = \prod_{k=0}^n \left[ \binom{r+x_k-1}{x_k} \left( \frac{1}{1+\beta} \right)^r \left( \frac{\beta}{1+\beta} \right)^{x_k} \right] \quad \text{con} \quad \beta = \frac{1-\theta}{\theta}.$$

Al tomar logaritmo natural se obtiene:

$$\ln \{L(r, \beta)\} = \sum_{k=0}^n \left[ \ln \binom{r+x_k-1}{x_k} - r \ln(1+\beta) + x_k \ln \left( \frac{\beta}{1+\beta} \right) \right].$$

Puesto que  $r > 0$  ( $r \in \mathbb{R}$ ), se puede emplear la expresión (1.19), se tiene que la función de log-verosimilitud se expresa como:

$$\ln \{L(r, \beta)\} = \sum_{k=1}^n \left[ x_k \ln \beta - x_k \ln(1 + \beta) - \ln(x_k!) + \sum_{m=0}^{x_k-1} \ln(m + r) \right] - \sum_{k=0}^n r \ln(1 + \beta).$$

Las derivadas parciales de  $\ln \{L(r, \beta)\}$  con respecto a  $\beta$  y a  $r$  son:

$$\begin{aligned} \frac{\partial}{\partial \beta} \ln \{L(r, \beta)\} &= \left( \frac{1}{\beta} - \frac{1}{1 + \beta} \right) \sum_{k=1}^n x_k - \sum_{k=0}^n \widehat{r} \frac{1}{1 + \widehat{\beta}} \\ \frac{\partial}{\partial r} \ln \{L(r, \beta)\} &= \sum_{k=1}^n \left[ \sum_{m=0}^{x_k-1} \frac{1}{m + \widehat{r}} \right] - \sum_{k=0}^n \ln(1 + \widehat{\beta}). \end{aligned}$$

Las condiciones necesarias para alcanzar el máximo se obtienen igualando a cero las anteriores derivadas parciales respecto de los parámetros, es decir,

$$\widehat{r} \widehat{\beta} = \frac{1}{n} \sum_{k=1}^n x_k = \bar{X} \quad \text{y} \quad n \ln(1 + \widehat{\beta}) = \sum_{k=1}^n \left[ \sum_{m=0}^{x_k-1} \frac{1}{m + \widehat{r}} \right],$$

si se despeja  $\widehat{\beta}$  de la primera expresión y se reemplaza en la segunda, el valor de  $\widehat{r}$  se determina como la solución de la ecuación

$$n \ln \left( 1 + \frac{\bar{X}}{\widehat{r}} \right) = \sum_{k=1}^n \left[ \sum_{m=0}^{x_k-1} \frac{1}{m + \widehat{r}} \right]. \quad (4.11)$$

La ecuación (4.11) se puede resolver numéricamente mediante el método de Newton-Raphson (ver [56, p. 366]), luego se tiene que:

$$r_{n+1} = r_n - \frac{f(r_n)}{f'(r_n)}, \quad r_n \neq 0, \quad (4.12)$$

donde

$$\begin{aligned} f(r) &= \ln \left( 1 + \frac{\bar{X}}{r} \right) - \frac{1}{n} \sum_{k=1}^n \left[ \sum_{m=0}^{x_k-1} \frac{1}{m + \widehat{r}} \right] \quad \text{y} \\ f'(r) &= -\frac{\bar{X}}{r^2 + \bar{X}r} + \frac{1}{n} \sum_{k=1}^n \left[ \sum_{m=0}^{x_k-1} \frac{1}{(m + \widehat{r})^2} \right]. \end{aligned}$$

Aquí se puede usar como valor inicial  $r_0$  el estimador de  $r$  establecido en (4.10) por el método de los momentos. Por lo tanto, la estimación de  $\theta$ , se obtiene como:

$$\hat{\theta} = \frac{1}{1 + \hat{\beta}} = \frac{\hat{r}}{\bar{X} + \hat{r}}. \quad (4.13)$$

Un ejemplo numérico se encuentra en el apéndice B.4.2. ☑

**Ejemplo 4.10.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid  $G(\theta)$ , distribución geométrica, encuentre un **EMV** para  $\theta$ .

*Solución.* En este caso,

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \theta(1-\theta)^{x_i} = \theta^n (1-\theta)^{\sum_{i=1}^n x_i}$$

$$\ell(\theta; \mathbf{x}) = n \ln(\theta) + \ln(1-\theta) \sum_{i=1}^n x_i,$$

derivando  $\ell(\theta; \mathbf{x})$  con respecto a  $\theta$  se tiene que:

$$\frac{\partial[\ell(\theta; \mathbf{x})]}{\partial \theta} = \frac{n}{\theta} - \frac{1}{1-\theta} \sum_{i=1}^n x_i,$$

igualando a cero:

$$\frac{n}{\hat{\theta}} - \frac{1}{1-\hat{\theta}} \sum_{i=1}^n x_i = 0 \quad \rightarrow \quad \frac{n(1-\hat{\theta}) - n\hat{\theta}\bar{x}}{\hat{\theta}(1-\hat{\theta})} = 0$$

$$n[(1-\hat{\theta}) - \hat{\theta}\bar{x}] = 0 \quad \rightarrow \quad \hat{\theta}(1+\bar{x}) = 1 \quad \rightarrow \quad \hat{\theta} = \frac{1}{1+\bar{x}}.$$

Luego,  $\hat{\theta} = \frac{1}{1+\bar{x}}$  es un candidato para ser **EMV**. Para verificarlo se toma la segunda derivada de  $\ell(\theta; \mathbf{x})$  con respecto a  $\theta$  y se evalúa en  $\theta = \frac{1}{1+\bar{x}}$ :

$$\frac{\partial^2[\ell(\theta; \mathbf{x})]}{\partial \theta^2} = \frac{\partial}{\partial \theta} \left[ \frac{n}{\theta} - \frac{1}{1-\theta} \sum_{i=1}^n x_i \right] = -\frac{n}{\theta^2} - \frac{n\bar{x}}{(1-\theta)^2}$$

$$\frac{\partial^2[\ell(\theta; \mathbf{x})]}{\partial \theta^2} \Bigg|_{\theta=\frac{1}{1+\bar{x}}} = -\frac{n}{\left(\frac{1}{1+\bar{x}}\right)^2} - \frac{n\bar{x}}{\left(1-\frac{1}{1+\bar{x}}\right)^2} < 0,$$

como  $\bar{x} \geq 0$ , se tiene entonces que  $\hat{\theta} = \frac{1}{1+\bar{x}}$  es un máximo. ☑

**Ejemplo 4.11.** Sean  $X_1, X_2, \dots, X_n$ , variables aleatorias *iid*  $P(\theta)$ , encuentre un *EMV* para  $\theta$ .

*Solución.* En este caso:

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!} = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$$

$$\ell(\theta; \mathbf{x}) = -n\theta + \ln(\theta) \sum_{i=1}^n x_i - \ln\left(\prod_{i=1}^n x_i!\right),$$

derivando  $\ell(\theta; \mathbf{x})$  con respecto a  $\theta$  se tiene que:

$$\frac{\partial[\ell(\theta; \mathbf{x})]}{\partial\theta} = -n + \frac{1}{\theta} \sum_{i=1}^n x_i = -n + \frac{n\bar{x}}{\theta},$$

igualando a cero:

$$-n + \frac{n\bar{x}}{\hat{\theta}} = 0 \quad \rightarrow \quad \frac{\bar{x}}{\hat{\theta}} = 1 \quad \rightarrow \quad \hat{\theta} = \bar{x}. \quad (4.14)$$

Luego  $\hat{\theta} = \bar{x}$  es un candidato para ser *EMV*. Para verificarlo se toma la segunda derivada de  $\ell(\theta; \mathbf{x})$  con respecto a  $\theta$  y se evalúa en  $\theta = \bar{x}$ :

$$\frac{\partial^2[\ell(\theta; \mathbf{x})]}{\partial\theta^2} = \frac{\partial}{\partial\theta} [-n + n\bar{x}\theta^{-1}] = -\frac{n\bar{x}}{\theta^2}$$

$$\left. \frac{\partial^2[\ell(\theta; \mathbf{x})]}{\partial\theta^2} \right|_{\theta=\bar{x}} = -\frac{n\bar{x}}{\bar{x}^2} = -\frac{n}{\bar{x}} < 0,$$

puesto que  $\bar{x} \geq 0$ , se tiene entonces que  $\hat{\theta} = \bar{x}$  es un máximo. ✓

**Ejemplo 4.12.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias *iid* con *fmp* uniforme discreta sobre  $1, 2, \dots, \theta$ , es decir,  $P(X = x) = \frac{1}{\theta} I_{\{x \in \{1, 2, 3, \dots, \theta\}\}}$ , encuentre un *EMV* para  $\theta$ .

*Solución.* La función de verosimilitud es dada por:

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \frac{1}{\theta} I_{\{x_i \in \{1, 2, 3, \dots, \theta\}\}} = \frac{1}{\theta^n} \prod_{i=1}^n I_{\{x_i \in \{1, 2, 3, \dots, \theta\}\}}$$

$$\underbrace{=}_{23} \frac{1}{\theta^n} I\left(\bigcap_{i=1}^n \{x_i \in \{1, 2, 3, \dots, \theta\}\}\right) = \frac{1}{\theta^n} \min_{1 \leq i \leq n} \{I_{\{x_i \in \{1, 2, 3, \dots, \theta\}\}}\}$$

Luego<sup>24</sup>,

$$L(\theta; \mathbf{x}) = \theta^{-n} I_{\{\max\{x_1, x_2, \dots, x_n\} \in \{1, 2, 3, \dots, \theta\}\}} = \theta^{-n} I_{\{1, 2, 3, \dots, \theta\}}(x_{(n)}),$$

la cual no es diferenciable con respecto a  $\theta$ . Los valores candidatos posibles para  $\theta$  son  $1, 2, 3, \dots, \max\{x_1, x_2, \dots, x_n\}$ ,

$$L(\theta; \mathbf{x}) = \begin{cases} \frac{1}{1^n} I_{\{1\}}(x_{(n)}) & \text{si } \theta = 1 \\ \frac{1}{2^n} I_{\{1, 2\}}(x_{(n)}) & \text{si } \theta = 2 \\ \frac{1}{3^n} I_{\{1, 2, 3\}}(x_{(n)}) & \text{si } \theta = 3 \\ \vdots & \\ \frac{1}{(x_{(n)} - 1)^n} I_{\{1, 2, 3, \dots, x_{(n)} - 1\}}(x_{(n)}) & \text{si } \theta = x_{(n)} - 1 \\ \frac{1}{(x_{(n)})^n} I_{\{1, 2, 3, \dots, x_{(n)}\}}(x_{(n)}) & \text{si } \theta = x_{(n)} \end{cases}$$

$$= \begin{cases} 0 & \text{si } \theta = 1 \\ 0 & \text{si } \theta = 2 \\ 0 & \text{si } \theta = 3 \\ \vdots & \\ 0 & \text{si } \theta = x_{(n)} - 1 \\ \frac{1}{(x_{(n)})^n} & \text{si } \theta = x_{(n)} \end{cases}$$

y, por tanto,  $L(\theta; \mathbf{x}) = 0$  para  $\theta = 1, 2, 3, \dots, x_{(n)} - 1$  y  $L(\theta; \mathbf{x}) = \frac{1}{\theta^n}$  para  $\theta = x_{(n)}$ . En este sentido, el **EMV** es dado por  $\hat{\theta} = x_{(n)}$ .  $\square$

<sup>23</sup> Considere la función indicadora  $I_A$  dada en (1.10). Sea  $A_1, A_2, A_3, \dots, A_n$  una colección de subconjuntos de  $\Omega$  y denotemos  $I(A_k)$  sus funciones indicadoras. Se puede mostrar que (ver [11, p. 22]):

$$I\left(\bigcap_{j=1}^n A_j\right) = \prod_{j=1}^n I(A_j) = \min\{I(A_1), I(A_2), I(A_3), \dots, I(A_n)\}.$$

<sup>24</sup> El principio de inclusión-exclusión para funciones indicadoras está dado por:

$$I\left(\bigcup_{j=1}^n A_j\right) = \sum_{i=1}^n I(A_i) - \sum_{i,j: 1 \leq i < j \leq n} I(A_i \cap A_j) + \sum_{i,j,k: 1 \leq i < j < k \leq n} I(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} I(A_1 \cap \dots \cap A_n) = \max\{I(A_1), I(A_2), I(A_3), \dots, I(A_n)\}.$$

**Ejemplo 4.13.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias *iid* uniforme sobre el intervalo  $\mathcal{F}_\theta = (0, \theta)$  con  $\theta > 0$  desconocido, encuentre un *EMV* para  $\theta$ .

*Solución.* La función de verosimilitud es:

$$\begin{aligned} L(\theta; \mathbf{x}) &= \prod_{i=1}^n \frac{1}{\theta} I_{\{x_i \in \mathcal{F}_\theta\}} = \frac{1}{\theta^n} \prod_{i=1}^n I_{\{x_i \in \mathcal{F}_\theta\}} = \frac{1}{\theta^n} I\left(\bigcap_{i=1}^n \{x_i \in \mathcal{F}_\theta\}\right) \\ &= \frac{1}{\theta^n} I(0 < \max\{x_1, x_2, \dots, x_n\} \leq \theta) = \frac{1}{\theta^n} I_{\mathcal{F}_\theta}(x_{(n)}) \\ &= \frac{1}{\theta^n} I_{[x_{(n)}, \infty)}(\theta). \end{aligned}$$

Dado que la función indicadora en el intervalo  $[x_{(n)}, \infty)$  está dada por:

$$I_{[x_{(n)}, \infty)}(\theta) = \begin{cases} 1, & \text{si } \theta \in [x_{(n)}, \infty) \\ 0, & \text{si } \theta \in (0, x_{(n)}), \end{cases}$$

entonces,

$$L(\theta; \mathbf{x}) = \begin{cases} \theta^{-n}, & \text{si } \theta \geq x_{(n)} \\ 0, & \text{si } \theta < x_{(n)} \end{cases}$$

Nótese que  $L(\theta; \mathbf{x})$  es estrictamente decreciente en  $(x_{(n)}, \infty)$  y es 0 en  $(0, x_{(n)})$ , ver figura 4.1, luego, un máximo único de  $L(\theta; \mathbf{x})$  es  $x_{(n)}$ , el cual es un punto de discontinuidad de  $L(\theta; \mathbf{x})$ .

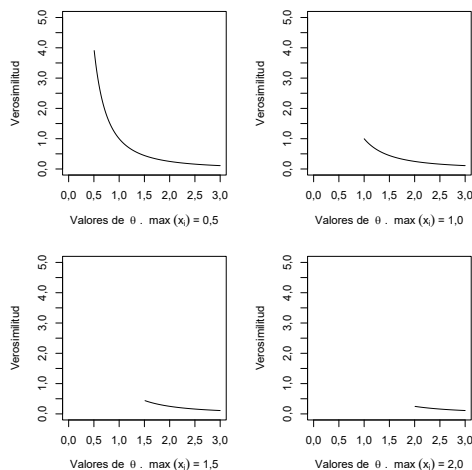


Figura 4.1. Funciones de verosimilitud para la *fdp* uniforme en  $\mathcal{F}_\theta$ , considerando diferentes valores de  $x_{(n)}$

Por otra parte,  $L(\theta; \mathbf{x})$  no es diferenciable en  $x_{(n)}$ :

$$\frac{dL(\theta; \mathbf{x})}{d\theta} = \begin{cases} -n\theta^{-n-1}, & \text{si } \theta \geq x_{(n)} \\ 0, & \text{si } \theta < x_{(n)}. \end{cases}$$

y, por ende, obtener un **EMV** no es viable. Ahora, considere el caso donde  $\mathcal{F}_\theta^* = (\theta - \frac{1}{2}, \theta + \frac{1}{2})$  con  $\theta \in \mathbb{R}$ . Luego, la **fdp** de  $X_1$  es dada por:

$$f_{X_1}(x) = \begin{cases} 1, & \text{si } \theta - \frac{1}{2} < x < \theta + \frac{1}{2} \\ 0, & \text{en otro caso.} \end{cases}$$

Por lo tanto, la función de verosimilitud es [24, p. 97]:

$$L(\theta; \mathbf{x}) = I_{(x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2})}(\theta),$$

donde  $x_{(1)} = \min\{x_1, \dots, x_n\}$ . Nuevamente, el método de usar la función de verosimilitud no es aplicable. Sin embargo, se deduce que cualquier estadística  $T(X)$  que satisface

$$x_{(n)} - \frac{1}{2} \leq T(x) \leq x_{(1)} + \frac{1}{2},$$

es un **EMV** de  $\theta$ .

Este ejemplo indica que los **EMV** pueden no ser únicos y tampoco ser razonables. ✓

**Ejemplo 4.14.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias **iid**  $Exp(\theta)$ , encuentre un **EMV** para  $\theta$ .

*Solución.* La función de verosimilitud es dada por:

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^n x_i}$$

$$\ell(\theta; \mathbf{x}) = n \ln(\theta) - \theta \sum_{i=1}^n x_i = n \ln(\theta) - n\theta \bar{x},$$

derivando se tiene que:

$$\frac{\partial \ell(\theta; \mathbf{x})}{\partial \theta} = \frac{n}{\theta} - n\bar{x},$$

igualando a cero:

$$n\hat{\theta}^{-1} - n\bar{x} = 0 \quad \Rightarrow \quad \hat{\theta} = \bar{x}^{-1}. \quad (4.15)$$

Luego,  $\hat{\theta} = \frac{1}{\bar{x}}$  es un candidato para ser **EMV**. Para verificarlo se toma la segunda derivada de  $\ell(\theta; \mathbf{x})$  con respecto a  $\theta$  y se evalúa en  $\theta = \frac{1}{\bar{x}}$ :

$$\begin{aligned} \frac{\partial^2[\ell(\theta; \mathbf{x})]}{\partial \theta^2} &= \frac{\partial}{\partial \theta} \left[ \frac{n}{\theta} - n\bar{x} \right] = -\frac{n}{\theta^2} \\ -\frac{n}{\theta^2} \Big|_{\theta=\frac{1}{\bar{x}}} &= -n \left( \frac{1}{\bar{x}} \right)^{-2} < 0, \end{aligned}$$

por tanto  $\hat{\theta} = \frac{1}{\bar{x}}$  es un máximo. ✓

La tabla 4.2 muestra los **EMV** de los parámetros de algunas distribuciones.

Tabla 4.2. **EMV** de los parámetros de algunas distribuciones

Distribución	Parámetro(s)	Estimador(es)
Uniforme discreta	$\theta$	$\hat{\theta} = X_{(n)}$
Binomial	$m$ y $p$	$\hat{m} = \left\lfloor \frac{\bar{X}^2}{\bar{X} - S_n^2} \right\rfloor$ y $\hat{p} = \frac{\bar{X} - S_n^2}{\bar{X}}$
Poisson	$\lambda$	$\hat{\lambda} = \bar{X}$
Geométrica	$p$	$\hat{p} = \frac{1}{1 + \bar{X}}$
Normal	$\mu$ y $\sigma^2$	$\hat{\mu} = \bar{X}$ y $\hat{\sigma}^2 = S_n^2$
Gamma	$r$ y $\lambda$	$\hat{r} = \frac{\bar{X}^2}{S_n^2}$ y $\hat{\sigma}^2 = \frac{\bar{X}}{S_n^2}$
Exponencial	$\lambda$	$\hat{\lambda} = \frac{1}{\bar{X}}$

**Ejemplo 4.15.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias *iid*  $N(\theta, \sigma^2)$  con  $\theta$  y  $\sigma^2$  desconocidos. Encuentre estimadores para  $\theta$  y  $\sigma^2$ .

*Solución.* En este caso,

$$\begin{aligned} L(\theta, \sigma^2; \mathbf{x}) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2 \right\} \\ \ell(\theta, \sigma^2; \mathbf{x}) &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2, \end{aligned}$$

derivando con respecto a  $\theta$  y con respecto a  $\sigma^2$  se tiene que:

$$\frac{\partial \ell(\theta, \sigma^2; \mathbf{x})}{\partial \theta} = -\frac{1}{2\sigma^2} (-2) \sum_{i=1}^n (x_i - \theta) = \frac{n}{\sigma^2} (\bar{x} - \theta) \tag{4.16}$$

$$\frac{\partial \ell(\theta, \sigma^2; \mathbf{x})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \theta)^2, \quad (4.17)$$

igualando a cero, de (4.16) y de (4.17):

$$\begin{aligned} \frac{n}{\sigma^2} (\bar{x} - \hat{\theta}) &= 0 \quad \rightarrow \quad \hat{\theta} = \bar{x} \\ -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^n (x_i - \hat{\theta})^2 &= 0 \quad \rightarrow \quad -n\hat{\sigma}^2 + \sum_{i=1}^n (x_i - \hat{\theta})^2 = 0 \\ &\rightarrow \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \geq 0. \end{aligned}$$

Para verificar que esta solución es un método global, ver apéndice B.2.1. Primero se calculan las segundas derivadas parciales

$$\begin{aligned} \frac{\partial^2 \ell(\theta, \sigma^2; \mathbf{x})}{\partial \theta^2} &= -\frac{n}{\sigma^2} \\ \frac{\partial^2 \ell(\theta, \sigma^2; \mathbf{x})}{\partial (\sigma^2)^2} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \theta)^2 \end{aligned}$$

y las respectivas derivadas mixtas o cruzadas son:

$$\frac{\partial^2 \ell(\theta, \sigma^2; \mathbf{x})}{\partial \theta \partial \sigma^2} = \frac{\partial}{\partial \sigma^2} \left( \frac{\partial \ell(\theta, \sigma^2; \mathbf{x})}{\partial \theta} \right) = -\frac{n}{\sigma^4} (\bar{x} - \theta).$$

Luego, la matriz de segundas derivadas (hessiana, ver apéndice B.2.1) queda:

$$\mathbf{H}_\ell(\theta, \sigma^2; \mathbf{x}) = \begin{pmatrix} \frac{\partial^2 \ell(\theta, \sigma^2; \mathbf{x})}{\partial \theta^2} & \frac{\partial^2 \ell(\theta, \sigma^2; \mathbf{x})}{\partial \theta \partial \sigma^2} \\ \frac{\partial^2 \ell(\theta, \sigma^2; \mathbf{x})}{\partial \theta \partial \sigma^2} & \frac{\partial^2 \ell(\theta, \sigma^2; \mathbf{x})}{\partial (\sigma^2)^2} \end{pmatrix} = \begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{n(\bar{x} - \theta)}{\sigma^4} \\ -\frac{n(\bar{x} - \theta)}{\sigma^4} & \frac{n}{2\sigma^4} - \sum_{i=1}^n \frac{(x_i - \theta)^2}{\sigma^6} \end{pmatrix}.$$

Puesto que  $\hat{\sigma}^2 > 0$ , y evaluando en  $\hat{\theta} = \bar{x}$  y  $\sigma^2 = \hat{\sigma}^2$ :

$$\mathbf{H}_\ell(\bar{x}, \hat{\sigma}^2; \mathbf{x}) = \begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{n}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6} \sum_{i=1}^n (x_i - \bar{x})^2 \end{pmatrix} = \begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4} \end{pmatrix}, \quad (4.18)$$

como  $\mathbf{H}_\ell(\bar{x}, \hat{\sigma}^2; \mathbf{x})$  es una matriz diagonal, sus valores propios son los elementos de la diagonal, como ambos son negativos se tiene que  $(\bar{x}, \hat{\sigma}^2)$  es un

máximo local de  $\ell(\bar{x}, \hat{\sigma}^2; \mathbf{x})$ . Por lo tanto, los **EMV** para  $(\theta, \sigma^2)$  son:

$$\hat{\theta} = \bar{x} \quad \text{y} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \quad \checkmark$$

### 4.3. Estimación bayesiana

El enfoque bayesiano a la estadística es diferente al procedimiento clásico a la estadística.

- En el enfoque clásico, el parámetro,  $\theta$ , se considera una cantidad **desconocida pero fija** y para tener información de  $\theta$  se elige una muestra aleatoria  $X_1, X_2, \dots, X_n$  de una población indexada por  $\theta$ , luego basados en los valores observados se obtiene la información de  $\theta$ .
- En el enfoque bayesiano,  $\theta$  (el parámetro de la distribución) es considerado una realización de una variable aleatoria  $\theta$  cuya distribución es conocida. La distribución de  $\theta$  (la variable aleatoria) se puede expresar por una **fdp** que se llamará la **distribución a priori**. Esta distribución es **subjetiva** y se basa en las “creencias” que el investigador tenga sobre  $\theta$  antes de observar los datos, es decir, la distribución *a priori* se propone **antes** de observar los datos. Después se selecciona una muestra aleatoria  $X_1, X_2, \dots, X_n$  de una población indexada por  $\theta$  y la distribución *a priori* se actualiza con los datos obtenidos en esta muestra. A la **fdp** actualizada se le llama función de **distribución a posteriori**. Cabe aclarar que la distribución *a posteriori* es una distribución condicional a la muestra observada.

#### 4.3.1. Distribuciones *a priori* y *a posteriori*

Bajo el enfoque clásico se usa la notación  $p(\mathbf{x}; \theta)$ , pero en la versión bayesiana, el parámetro  $\theta$  es una cantidad aleatoria con distribución inicial  $p(\mathbf{x} | \theta)$ ,

$$p(\mathbf{x} | \theta) = P(\mathbf{X} = \mathbf{x} | \vartheta = \theta) = \frac{P(\mathbf{X} = \mathbf{x}, \vartheta = \theta)}{P(\vartheta = \theta)}, \quad (4.19)$$

con  $\vartheta$  la variable aleatoria asociada a  $\theta$ .  $P(\vartheta = \theta) = p(\theta; \lambda)$  es la **distribución a priori** y usualmente se le denota por  $\pi(\theta; \lambda)$ , con  $\lambda$  siendo otro parámetro, algunas veces llamado hiperparámetro. Entonces de (4.19):

$$P(\mathbf{X} = \mathbf{x}, \vartheta = \theta) = \pi(\theta; \lambda)p(\mathbf{x} | \theta)$$

$$p(x, \theta) = \pi(\theta; \lambda)p(x | \theta). \tag{4.20}$$

Si tanto  $X$  como  $\vartheta$  son variables aleatorias continuas (discretas, respectivamente), entonces  $\begin{pmatrix} \vartheta \\ X \end{pmatrix}$  es un vector aleatorio continuo (discreto, respectivamente) con fdp (fmp, respectivamente)<sup>25</sup>.

La **distribución a posteriori** es dada por la distribución condicional de  $\vartheta$  dada  $X = x$ ,

$$P(\vartheta = \theta | X = x) = \frac{P(X = x, \vartheta = \theta)}{P(X = x)} = \frac{\pi(\theta; \lambda)p(x | \theta)}{p(x)}. \tag{4.21}$$

A  $P(\vartheta = \theta | X = x) = p(\theta|x)$  se le denota por  $\pi(\theta|x)$  y es una fmp o una fdp, de acuerdo a si la distribución a priori era una fmp o una fdp. Retomando (4.21):

$$\underbrace{\frac{\pi(\theta; \lambda)p(x|\theta)}{p(x)}}_{\pi(\theta|x)} = \begin{cases} \frac{\pi(\theta; \lambda)p(x|\theta)}{\sum_{t=1}^{\infty} \pi(\theta_t)p(x|\theta_t)} & \text{si } \vartheta \text{ es una variable} \\ & \text{aleatoria discreta} \\ \frac{\pi(\theta; \lambda)p(x|\theta)}{\int_{-\infty}^{\infty} \pi(t)p(x|t)dt} & \text{si } \vartheta \text{ es una variable} \\ & \text{aleatoria continua.} \end{cases}$$

**Ejemplo 4.16.** (Ensayos binomiales). *Supóngase, como en [12, p. 14], que  $X_1, \dots, X_n$  son funciones indicadoras de  $n$  ensayos Bernoulli con probabilidad de éxito  $\vartheta$ ,  $0 < \vartheta < 1$ . En la escogencia de la distribución a priori para  $\vartheta$ , se debe cumplir que se concentre en el intervalo  $(0, 1)$ , una candidata es la distribución beta con parámetros  $\alpha$  y  $\beta$ , con fdp  $B(\alpha, \beta)$  dada en (1.22). Determine  $\pi(\theta|x)$ .*

*Solución.*

- Distribución modelo: la fmp condicional en  $\vartheta = \theta$  es:

$$p(x | \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}.$$

- Distribución a priori: note que el hiperparámetro  $\lambda$ , mencionado en (4.19), corresponde al vector de parámetros  $\lambda = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ . Entonces,

$$\pi(\theta; \lambda) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}. \tag{4.22}$$

<sup>25</sup>Cuando  $\vartheta$  es una variable aleatoria continua y  $X$  es una variable aleatoria discreta, la distribución conjunta no es ni discreta ni continua (ver [12, p. 13]).

- Distribución *a posteriori*: se considera que  $\vartheta$  tiene **fdp a priori**  $\pi(\theta)$ , luego:

$$\pi(\theta|x) = \frac{\pi(\theta)\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}}{\int_0^1 \pi(t)t^{\sum_{i=1}^n x_i} (1-t)^{n-\sum_{i=1}^n x_i} dt}$$

Reemplazando en esta última expresión la *a priori* dada en (4.22):

$$\begin{aligned} \pi(\theta|x) &= \frac{\pi(\theta)\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}}{\int_0^1 \pi(t)t^{\sum_{i=1}^n x_i} (1-t)^{n-\sum_{i=1}^n x_i} dt} \\ &= \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}}{\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}t^{\alpha-1}(1-t)^{\beta-1}t^{\sum_{i=1}^n x_i} (1-t)^{n-\sum_{i=1}^n x_i} dt} \\ &= \frac{\theta^{\alpha+\sum_{i=1}^n x_i-1} (1-\theta)^{\beta+n-\sum_{i=1}^n x_i-1}}{\int_0^1 t^{\alpha+\sum_{i=1}^n x_i-1} (1-t)^{\beta+n-\sum_{i=1}^n x_i-1} dt} \end{aligned}$$

Tomando  $\tilde{\alpha} = \alpha + \sum_{i=1}^n x_i$  y  $\tilde{\beta} = \beta + n - \sum_{i=1}^n x_i$ , se tiene que:

$$\pi(\theta|x) = \frac{\overbrace{\frac{\Gamma(\tilde{\alpha} + \tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})}\theta^{\tilde{\alpha}-1}(1-\theta)^{\tilde{\beta}-1}}^{B(\tilde{\alpha}, \tilde{\beta})}}{\underbrace{\int_0^1 \frac{\Gamma(\tilde{\alpha} + \tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})}t^{\tilde{\alpha}-1}(1-t)^{\tilde{\beta}-1}dt}_1} = B(\tilde{\alpha}, \tilde{\beta}).$$

Cabe notar que  $\pi(\theta|x)$  depende de los datos solo mediante  $\sum_{i=1}^n x_i$ .

En la figura 4.2 se presentan distribución *a priori*, verosimilitud y **fdp a posteriori** para  $n = 10$ ,  $\sum_{i=1}^{10} x_i = 3$ ,  $B(1,5)$ ,  $B(6,6)$  y  $B(6,0,2)$ . En la figura 4.3 se presentan distribución *a priori*, verosimilitud y **fdp a posteriori** para  $n = 10$ ,  $\sum_{i=1}^{10} x_i = 9$ ,  $B(1,5)$ ,  $B(6,6)$  y  $B(6,0,2)$ .

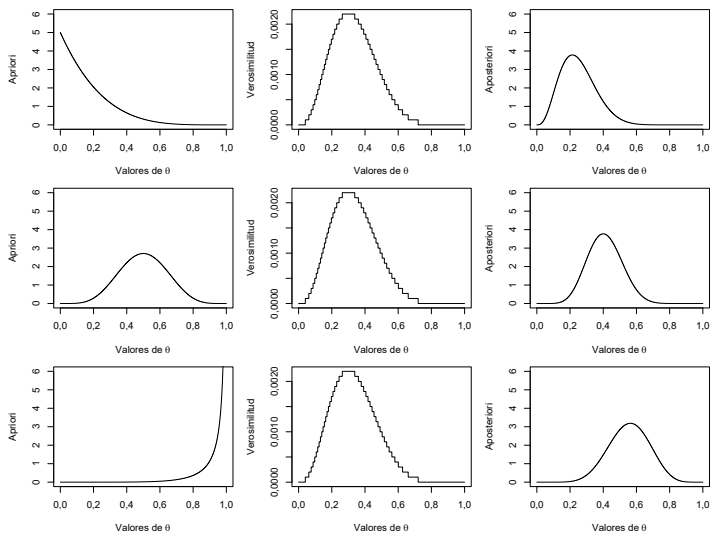


Figura 4.2. Distribución *a priori*, verosimilitud y fdp *a posteriori* para  $n = 10$ ,  $\sum_{i=1}^{10} x_i = 3$ ,  $B(1,5)$  (1.<sup>a</sup> fila),  $B(6,6)$  (2.<sup>a</sup> fila) y  $B(6,0,2)$  (3.<sup>a</sup> fila)

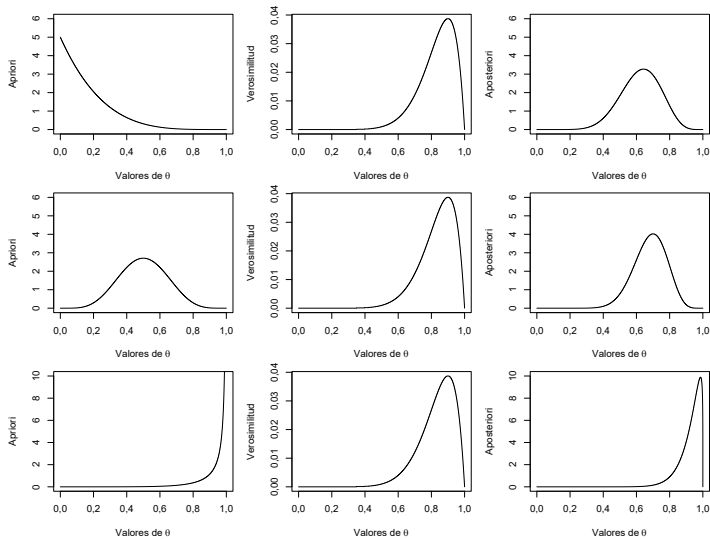


Figura 4.3. Distribución *a priori*, verosimilitud y fdp *a posteriori* para  $n = 10$ ,  $\sum_{i=1}^{10} x_i = 9$ ,  $B(1,5)$  (1.<sup>a</sup> fila),  $B(6,6)$  (2.<sup>a</sup> fila) y  $B(6,0,2)$  (3.<sup>a</sup> fila)

Las instrucciones en *R* para las figuras son:

```
remove(list=ls())
x1<-seq(0,1,length=10000)
x2<-seq(0,1,length=100000)
alpha=1; beta=5; n=10; x=3; alpha2=6; beta2=6; beta3=0.2
y1=round(dbeta(x1,alpha,beta),digits=4)
y2=round((x2^x)*(1-x2)^(n-x),digits=4)
y3=round(dbeta(x1,alpha+x,beta+n-x),digits=4)
y4=round(dbeta(x1,alpha2,beta2),digits=4)
y5=round(dbeta(x1,alpha2+x,beta2+n-x),digits=4)
y6=round(dbeta(x1,alpha2,beta3),digits=4)
y7=round(dbeta(x1,alpha2+x,beta3+n-x),digits=4)
par(mfrow=c(3,3),mar=c(4,4,1,1))
plot(y1~x1,ylim=c(0,6),xlim=c(0,1),type="s",lty=1,
      xlab=expression("Valores de"~theta),ylab="a priori")
plot(y2~x2,xlim=c(0,1),type="l",lty=1,
      xlab=expression("Valores de"~theta),ylab="Verosimilitud")
plot(y3~x1,ylim=c(0,6),xlim=c(0,1),type="s",lty=1,
      xlab=expression("Valores de"~theta),ylab="a posteriori")
plot(y4~x1,ylim=c(0,6),xlim=c(0,1),type="s",lty=1,
      xlab=expression("Valores de"~theta),ylab="a priori")
plot(y2~x2,xlim=c(0,1),type="l",lty=1,
      xlab=expression("Valores de"~theta),ylab="Verosimilitud")
plot(y5~x1,ylim=c(0,6),xlim=c(0,1),type="s",lty=1,
      xlab=expression("Valores de"~theta),ylab="a posteriori")
plot(y6~x1,ylim=c(0,6),xlim=c(0,1),type="s",lty=1,
      xlab=expression("Valores de"~theta),ylab="a priori")
plot(y2~x2,xlim=c(0,1),type="l",lty=1,
      xlab=expression("Valores de"~theta),ylab="Verosimilitud")
plot(y7~x1,ylim=c(0,6),xlim=c(0,1),type="s",lty=1,
      xlab=expression("Valores de"~theta),ylab="a posteriori")
```

Los gráficos por fila (distribución *a priori*, verosimilitud y *fdp a posteriori*) en las figuras 4.2 y 4.3, para una proporción (ensayos Bernoulli) y usando como *a priori* la distribución beta, se pueden simplificar en un solo gráfico usando la librería de R LearnBayes:

```
library(LearnBayes)
prior=c(1,5); datos=c(3,7)
triplot(prior,datos) # En datos: 3 éxitos y 7 fracasos
prior=c(6,6); datos=c(3,7); triplot(prior,datos)
prior=c(6,0.2); datos=c(3,7); triplot(prior,datos)
prior=c(1,5); datos=c(9,1)
triplot(prior,datos) # En datos: 9 éxitos y 1 fracaso
prior=c(6,6); datos=c(9,1); triplot(prior,datos)
prior=c(6,0.2); datos=c(9,1); triplot(prior,datos)
```



**Ejemplo 4.17.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias *iid* Poisson de parámetro  $\theta$ ,  $\mathcal{P}(\theta)$ , y sea  $\theta \sim \text{gamma}(a, b)$ . Encontrar la distribución a posteriori de  $\theta$  dado  $\mathbf{X} = \mathbf{x}$ .

*Solución.*

- Distribución modelo:

$$p(\mathbf{x} | \theta) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!} = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} = \frac{e^{-n\theta} \theta^{n\bar{x}}}{\prod_{i=1}^n x_i!}.$$

- Distribución *a priori*:

$$\pi(\theta; \lambda) = \frac{b}{\Gamma(a)} (b\theta)^{a-1} e^{-b\theta} = \frac{b^a \theta^{a-1}}{\Gamma(a)} e^{-b\theta}, \quad \text{con } \lambda = \begin{pmatrix} a \\ b \end{pmatrix}.$$

- Distribución *a posteriori*:

$$\begin{aligned} \pi(\theta | \mathbf{x}) &= \frac{\pi(\theta) \frac{e^{-n\theta} \theta^{n\bar{x}}}{\prod_{i=1}^n x_i!}}{\int_0^\infty \pi(t) \frac{e^{-nt} t^{n\bar{x}}}{\prod_{i=1}^n x_i!} dt} = \frac{\pi(\theta) e^{-n\theta} \theta^{n\bar{x}}}{\int_0^\infty \pi(t) e^{-nt} t^{n\bar{x}} dt} \\ &= \frac{\frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} e^{-n\theta} \theta^{n\bar{x}}}{\int_0^\infty \frac{b^a}{\Gamma(a)} t^{a-1} e^{-bt} e^{-nt} t^{n\bar{x}} dt} = \frac{\theta^{a-1+n\bar{x}} e^{-b\theta-n\theta}}{\int_0^\infty t^{a-1+n\bar{x}} e^{-bt-nt} dt} \\ &= \frac{\theta^{a+n\bar{x}-1} e^{-(b+n)\theta}}{\int_0^\infty t^{a+n\bar{x}-1} e^{-(b+n)t} dt}, \end{aligned}$$

haciendo el cambio de variable  $u = (b+n)t$ ,  $du = (b+n)dt$  se tiene:

$$\begin{aligned} \pi(\theta | \mathbf{x}) &= \frac{\theta^{a-1+n\bar{x}} e^{-(b+n)\theta}}{\int_0^\infty \left(\frac{u}{b+n}\right)^{a+n\bar{x}-1} e^{-u} \frac{du}{b+n}} = \frac{\theta^{a+n\bar{x}-1} e^{-(b+n)\theta}}{\left(\frac{1}{b+n}\right)^{a+n\bar{x}} \underbrace{\int_0^\infty u^{a+n\bar{x}-1} e^{-u} du}_{\Gamma(a+n\bar{x})}} \\ &= \frac{(b+n)^{a+n\bar{x}}}{\Gamma(a+n\bar{x})} \theta^{a+n\bar{x}-1} e^{-(b+n)\theta}. \end{aligned}$$

Sea  $\tilde{a} = a + n\bar{x}$  y  $\tilde{b} = b + n$ . Entonces,

$$\begin{aligned} \pi(\theta|x) &= \frac{(\tilde{b})^{\tilde{a}}}{\Gamma(\tilde{a})} \theta^{\tilde{a}-1} e^{-\theta\tilde{b}} = \text{gamma}(\tilde{a}, \tilde{b}) \\ &= G(a + n\bar{x}, b + n). \end{aligned}$$

Cabe notar que  $\pi(\theta|x)$  depende de los datos solo mediante  $\sum_{i=1}^n x_i$ .  $\checkmark$

**Ejemplo 4.18.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid exponencial de parámetro  $\vartheta$ ,  $\text{Exp}(\vartheta)$ , y sea  $\vartheta \sim \text{gamma}(a, b)$ . Encontrar la distribución a posteriori de  $\vartheta$  dado  $\mathbf{X} = \mathbf{x}$ .

*Solución.*

- Distribución modelo:

$$p(\mathbf{x} | \theta) = \prod_{i=1}^n \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^n x_i}.$$

- Distribución a priori

$$\pi(\theta; \lambda) = \frac{b}{\Gamma(a)} (b\theta)^{a-1} e^{-b\theta} = \frac{b^a \theta^{a-1}}{\Gamma(a)} e^{-b\theta}, \quad \text{con} \quad \lambda = \begin{pmatrix} a \\ b \end{pmatrix}.$$

- Distribución a posteriori:

$$\begin{aligned} \pi(\theta|x) &= \frac{\pi(\theta)\theta^n e^{-\theta \sum_{i=1}^n x_i}}{\int_0^\infty \pi(t)t^n e^{-t \sum_{i=1}^n x_i} dt} = \frac{\frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \theta^n e^{-\theta \sum_{i=1}^n x_i}}{\int_0^\infty \frac{b^a}{\Gamma(a)} t^{a-1} e^{-bt} t^n e^{-t \sum_{i=1}^n x_i} dt} \\ &= \frac{\theta^{a+n-1} e^{-b\theta - \theta \sum_{i=1}^n x_i}}{\int_0^\infty t^{a+n-1} e^{-bt - t \sum_{i=1}^n x_i} dt} = \frac{\theta^{a+n-1} e^{-(b+n\bar{x})\theta}}{\int_0^\infty t^{a+n-1} e^{-(b+n\bar{x})t} dt}, \end{aligned}$$

haciendo el cambio de variable  $u = (b + n\bar{x})t$ ,  $du = (b + n\bar{x})dt$  se tiene:

$$\pi(\theta|x) = \frac{\theta^{a+n-1} e^{-(b+n\bar{x})\theta}}{\int_0^\infty \left(\frac{u}{b+n\bar{x}}\right)^{a+n-1} e^{-u} \frac{du}{b+n\bar{x}}}$$

$$\begin{aligned}
 &= \frac{\theta^{a+n-1} e^{-(b+n\bar{x})\theta}}{\left(\frac{1}{b+n\bar{x}}\right)^{a+n} \underbrace{\int_0^\infty u^{a+n-1} e^{-u} du}_{\Gamma(a+n)}} \\
 &= \frac{(b+n\bar{x})^{a+n}}{\Gamma(a+n)} \theta^{a+n-1} e^{-(b+n\bar{x})\theta}.
 \end{aligned}$$

Sea  $\tilde{a} = a + n$  y  $\tilde{b} = b + n\bar{x}$ . Entonces,

$$\pi(\theta|x) = \frac{(\tilde{b})^{\tilde{a}}}{\Gamma(\tilde{a})} \theta^{\tilde{a}-1} e^{-\theta\tilde{b}} = \text{gamma}(\tilde{a}, \tilde{b}) = G(a+n, b+n\bar{x}).$$

Cabe notar que  $\pi(\theta|x)$  depende de los datos solo mediante  $\sum_{i=1}^n x_i$ . ☑

### 4.3.2. Estimación puntual bayesiana

**DEFINICIÓN 4.5.** Sea  $\mathcal{F}$  que denota la clase de las *fmp* o las *fdp*  $p(x|\theta)$ , indexada por  $\theta$ . Una clase  $\Pi$  de distribuciones a priori es una **familia conjugada** para  $\mathcal{F}$  si la distribución a posteriori está en la clase  $\Pi$  para todas las  $p(x|\theta) \in \mathcal{F}$ , todas las a priori en  $\Pi$ , y todas las  $x$ .

En la tabla 4.3 se presentan las distribuciones conjugadas para las distribuciones consideradas en los ejemplos previos.

Tabla 4.3. Algunas distribuciones conjugadas

Modelo $p(x \theta)$	A priori $\pi(\theta)$	A posteriori $\pi(\theta x)$
Binomial $Bin(n, \theta)$	Beta $Be(\alpha, \beta)$	Beta $Be(\alpha + n\bar{x}, \beta + n - n\bar{x})$
Poisson $\mathcal{P}(\theta)$	Gamma $G(\alpha, \beta)$	Gamma $G(\alpha + n\bar{x}, \beta + n)$
Exponencial $\exp\{\theta\}$	Gamma $G(\alpha, \beta)$	Gamma $G(\alpha + n, \beta + n\bar{x})$
Normal $N(\theta, \sigma^2)$	Normal $N(\vartheta, \tau^2)$	Normal <sup>a</sup> $N\left(\omega_n \vartheta + (1 - \omega_n)\bar{x}, (1 - \omega_n)\frac{\sigma^2}{n}\right)$

<sup>a</sup> Donde  $\omega_n = \frac{\sigma^2/n}{\sigma^2/n + \tau^2}$ .

La primera línea es interpretada como la familia beta es conjugada para la familia binomial, ya que si se inicia con una distribución a priori beta, se tendrá una distribución a posteriori beta.

**DEFINICIÓN 4.6.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid con fdp  $p(x | \theta)$ . Sea  $\pi(\theta; \lambda)$  la distribución a priori de  $\theta$ . La media de la distribución a posteriori es usada como el estimador puntual bayesiano del parámetro, es decir,  $\mathbb{E}(\theta|x)$ .

**Ejemplo 4.19.** Considerando la información del ejemplo 4.16, se encontró que:

$$\pi(\theta|x) = \frac{\Gamma(\tilde{\alpha} + \tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})} \theta^{\tilde{\alpha}-1} (1-\theta)^{\tilde{\beta}-1},$$

con  $\tilde{\alpha} = \alpha + \sum_{i=1}^n x_i$  y  $\tilde{\beta} = \beta + n - \sum_{i=1}^n x_i$ . Determine un estimador de  $\theta$ .

*Solución.* Como la esperanza de la distribución  $Beta(\alpha, \beta)$  es  $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$ , entonces:

$$\begin{aligned} \hat{\theta} = \mathbb{E}(\theta|x) &= \frac{\tilde{\alpha}}{\tilde{\alpha} + \tilde{\beta}} = \frac{\alpha + \sum_{i=1}^n x_i}{\alpha + \sum_{i=1}^n x_i + \beta + n - \sum_{i=1}^n x_i} = \frac{\alpha + \sum_{i=1}^n x_i}{\alpha + \beta + n} \\ &= \frac{\alpha + n\bar{x}_n}{\alpha + \beta + n}. \end{aligned} \quad \checkmark$$

**Ejemplo 4.20.** Considerando la información del ejemplo 4.17, se encontró que:

$$\pi(\theta|x) = \frac{(\tilde{b})^{\tilde{a}}}{\Gamma(\tilde{a})} \theta^{\tilde{a}-1} e^{-\theta\tilde{b}},$$

con  $\tilde{a} = a + n\bar{x}$  y  $\tilde{b} = b + n$ . Determine un estimador de  $\theta$ .

*Solución.* En el ejemplo (1.83) se estableció que la esperanza de la distribución  $gamma(r, \lambda)$  era  $\mathbb{E}(X) = \frac{r}{\lambda}$ , entonces,

$$\hat{\theta} = \mathbb{E}(\theta|x) = \frac{\tilde{a}}{\tilde{b}} = \frac{a + n\bar{x}}{b + n}. \quad \checkmark$$

**Ejemplo 4.21.** Considerando la información del ejemplo 4.18, se encontró que:

$$\pi(\theta|x) = \frac{(\tilde{b})^{\tilde{a}}}{\Gamma(\tilde{a})} \theta^{\tilde{a}-1} e^{-\theta\tilde{b}},$$

con  $\tilde{a} = a + n$  y  $\tilde{b} = b + n\bar{x}$ . Determine un estimador de  $\theta$ .

*Solución.* En el ejemplo (1.83) se estableció que la esperanza de la distribución  $gamma(r, \lambda)$  era  $\mathbb{E}(X) = \frac{r}{\lambda}$ , entonces,

$$\hat{\theta} = \mathbb{E}(\theta|x) = \frac{\tilde{a}}{\tilde{b}} = \frac{a + n}{b + n\bar{x}}. \quad \checkmark$$

## 4.4. Estimación por el método de mínimos cuadrados

Suponga que  $Y_1, Y_2, Y_3, \dots, Y_n$  es una muestra aleatoria y considérese el problema de encontrar el estimador  $\theta$  que minimiza la norma euclidiana al cuadrado

$$\|Y - \theta \mathbf{1}_n\|^2 = (\mathbf{Y} - \theta \mathbf{1}_n)^\top (\mathbf{Y} - \theta \mathbf{1}_n), \quad (4.23)$$

donde  $\mathbf{Y} = (Y_1, Y_2, Y_3, \dots, Y_n)^\top$ , la expresión (4.23) mide la distancia cuadrática total entre el estimador  $\theta$  y los valores muestrales  $Y_i$ . Luego,

$$\frac{d}{d\theta} \|Y - \theta \mathbf{1}_n\|^2 = \frac{d}{d\theta} (\mathbf{Y}^\top \mathbf{Y} - 2\theta \mathbf{1}_n^\top \mathbf{Y} + \underbrace{\theta^2 \mathbf{1}_n^\top \mathbf{1}_n}_{2n\theta}) = 2(n\theta - \mathbf{1}_n^\top \mathbf{Y}),$$

igualando esta última ecuación a cero se obtiene que la expresión (4.23) se minimiza cuando  $\theta = \bar{Y}$ .

Puesto que  $\theta$  es un estimador de  $\mathbb{E}(Y)$ , entonces se puede interpretar como una predicción de cada valor  $Y_i$ , en consecuencia, para cada  $i$  la distancia  $Y_i - \theta$  denota un error de predicción. Luego, la expresión (4.23) representa la suma de los cuadrados de los errores de predicción y el estimador  $\theta$  que minimiza esta expresión se denomina estimador de mínimos cuadrados.

En esta sección se presentan los métodos de estimación por mínimos cuadrados ordinarios (MCO) y por mínimos cuadrados ponderados (MCP).

### 4.4.1. Mínimos cuadrados ordinarios

Suponga que las variables aleatorias  $Y_1, Y_2, Y_3, \dots, Y_n$ , se pueden expresar de la forma:

$$Y_i = g_i(\theta_1, \theta_2, \dots, \theta_k) + \varepsilon_i, \quad 1 \leq i \leq n,$$

con  $g_i$  funciones conocidas y  $\theta_1, \theta_2, \dots, \theta_k$ , parámetros de interés, de valor desconocido. Suponga que  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)^\top$ , el vector de parámetros, puede variar libremente sobre un conjunto  $\mathcal{A} \subset \mathbb{R}^r$ . También, suponga que los  $\varepsilon_i$  satisfacen, al menos en forma aproximada, las siguientes restricciones:

- $\mathbb{E}(\varepsilon_i) = 0, 1 \leq i \leq n.$
- $\mathbb{V}(\varepsilon_i) = \sigma^2 > 0, 1 \leq i \leq n.$
- $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0^{26}, 1 \leq i < j \leq n.$

<sup>26</sup>Ver expresión (1.44)

Cabe notar que  $\sigma^2$  es, también, un parámetro. Sea  $\mathbf{g}(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), g_2(\boldsymbol{\theta}), \dots, g_n(\boldsymbol{\theta}))^\top$  y  $\mathbf{Y} = (Y_1, Y_2, Y_3, \dots, Y_n)^\top$ , el valor esperado de  $\mathbf{Y}$  es:

$$\mathbb{E} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} g_1(\boldsymbol{\theta}) \\ g_2(\boldsymbol{\theta}) \\ \vdots \\ g_n(\boldsymbol{\theta}) \end{bmatrix}$$

$$\mathbb{E}(\mathbf{Y}) = \mathbf{g}(\boldsymbol{\theta})$$

Una vez se observa  $\mathbf{Y} = \mathbf{y}$ , con  $\mathbf{y} = (y_1, y_2, y_3, \dots, y_n)^\top$ , el estimador obtenido por **MCO** resulta de minimizar

$$\sum_{i=1}^n [y_i - g_i(\theta_1, \dots, \theta_k)]^2 = [\mathbf{y} - \mathbf{g}(\boldsymbol{\theta})]^\top [\mathbf{y} - \mathbf{g}(\boldsymbol{\theta})] = \|\mathbf{y} - \mathbf{g}(\boldsymbol{\theta})\|^2,$$

con rango de  $\boldsymbol{\theta}$  sobre  $A$ . Si las  $g_i$ ,  $i = 1, 2, \dots, n$  son diferenciables y el rango de  $(g_1, \dots, g_n)$  es cerrado,  $\hat{\boldsymbol{\theta}}$  está siempre definido.

**Ejemplo 4.22.** (Regresión lineal simple). *Un caso especial es cuando*<sup>27</sup>:

$$Y_i = g_i(\boldsymbol{\theta}) + \varepsilon_i = a + bx_i + \varepsilon_i, \quad 1 \leq i \leq n,$$

con  $\boldsymbol{\theta} = (a, b)^\top$ , el vector de parámetros, y  $x_i$ ,  $i = 1, 2, 3, \dots, n$ , valores conocidos (variables explicativas). Determine la estimación de  $\boldsymbol{\theta}$ .

*Solución.* En este caso

$$g_i(\boldsymbol{\theta}) = a + bx_i, \quad i = 1, 2, 3, \dots, n. \quad (4.24)$$

Para minimizar:

$$Q = \sum_{i=1}^n [y_i - g_i(\theta_1, \theta_2)]^2 = \sum_{i=1}^n [y_i - a - bx_i]^2,$$

se deriva  $Q$  con respecto a  $\boldsymbol{\theta}$ ,  $\nabla Q(\boldsymbol{\theta}) = \frac{\partial Q}{\partial \boldsymbol{\theta}}$  (ver nota B.10), y se iguala a cero,  $\nabla Q(\boldsymbol{\theta}) = \mathbf{0}$ :

$$\begin{pmatrix} \frac{\partial Q}{\partial a} \\ \frac{\partial Q}{\partial b} \end{pmatrix} = \begin{pmatrix} (-1) \sum_{i=1}^n 2(y_i - a - bx_i) \\ \sum_{i=1}^n 2(y_i - a - bx_i)(-x_i) \end{pmatrix} = -2 \begin{pmatrix} \sum_{i=1}^n y_i - na - b \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i - b \sum_{i=1}^n x_i^2 \end{pmatrix}$$

<sup>27</sup>Lineal en los parámetros significa que ningún parámetro en el modelo aparece como un exponente o es multiplicado (o dividido) por cualquier otro parámetro.

$$\begin{pmatrix} \frac{\partial Q}{\partial a} \\ \frac{\partial Q}{\partial b} \end{pmatrix} = -2 \begin{pmatrix} n\bar{y} - na - bn\bar{x} \\ \sum_{i=1}^n x_i y_i - an\bar{x} - b \sum_{i=1}^n x_i^2 \end{pmatrix}, \quad (4.25)$$

igualando (4.25) al vector nulo y reescribiendo se tiene que:

$$\begin{aligned} & \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{pmatrix} - \begin{pmatrix} n\hat{a} + \hat{b}n\bar{x} \\ \hat{a}n\bar{x} + \hat{b} \sum_{i=1}^n x_i^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ & n \begin{pmatrix} \hat{a} + \hat{b}\bar{x} \\ \hat{a}\bar{x} + \frac{1}{n}\hat{b} \sum_{i=1}^n x_i^2 \end{pmatrix} = n \begin{pmatrix} \bar{y} \\ \frac{1}{n} \sum_{i=1}^n x_i y_i \end{pmatrix} \\ & \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \frac{1}{n} \sum_{i=1}^n x_i^2 \end{pmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \frac{1}{n} \sum_{i=1}^n x_i y_i \end{pmatrix}. \end{aligned}$$

La solución de este sistema lineal de ecuaciones se realiza mediante la matriz inversa y se obtiene:

$$\begin{aligned} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} &= \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \frac{1}{n} \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \bar{y} \\ \frac{1}{n} \sum_{i=1}^n x_i y_i \end{pmatrix} \\ &= \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2} \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \begin{pmatrix} \bar{y} \\ \frac{1}{n} \sum_{i=1}^n x_i y_i \end{pmatrix} \\ &= \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2} \begin{pmatrix} \frac{\bar{y}}{n} \sum_{i=1}^n x_i^2 - \frac{\bar{x}}{n} \sum_{i=1}^n x_i y_i \\ -\bar{x}\bar{y} + \frac{1}{n} \sum_{i=1}^n x_i y_i \end{pmatrix} \\ &= \frac{1}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \begin{pmatrix} \bar{y} \sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \end{pmatrix}. \quad (4.26) \end{aligned}$$

Usando el coeficiente de correlación muestral dado en (2.21):

$$r_{xy} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y},$$

al reemplazar en la expresión del coeficiente  $\hat{b}$  y simplificar, se obtiene:

$$\hat{b} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x^2} = \frac{(n-1)r_{xy}s_x s_y}{(n-1)s_x^2} = r_{xy} \cdot \frac{s_y}{s_x}.$$

La segunda derivada de  $Q$ , es decir, la derivada de (4.25), es la matriz hessiana:

$$H_Q(a, b) = \begin{pmatrix} \frac{\partial^2 Q}{\partial a^2} & \frac{\partial^2 Q}{\partial b \partial a} \\ \frac{\partial^2 Q}{\partial a \partial b} & \frac{\partial^2 Q}{\partial b^2} \end{pmatrix} = \begin{pmatrix} 2n & 2n\bar{x} \\ 2n\bar{x} & 2 \sum_{i=1}^n x_i^2 \end{pmatrix} = 2n \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \frac{1}{n} \sum_{i=1}^n x_i^2 \end{pmatrix},$$

nótese que  $\det(H_Q(a, b)) = 4n^2 S_n^2$  y  $\text{tr}(H_Q(a, b)) = 2n(1 + \frac{1}{n} \sum_{i=1}^n x_i^2)$ , luego ambos son positivos y por ende  $H_Q(a, b)$  es definida positiva (apéndice B.1.2), es decir,  $(\hat{a}, \hat{b})^\top$  es un mínimo.  $\checkmark$

Para una ilustración del método de mínimos cuadrados en el caso de regresión lineal simple, se puede usar el código en R:

```
remove(list=ls())
library(animation)
least.squares(n=25)
```

**Ejemplo 4.23.** Considere los datos presentados en la tabla 4.4, donde  $y_i$  corresponde al valor pagado (en millones) y  $x_i$  al número de reclamaciones en el  $i$ -ésimo ramo. Encuentre las estimativas de  $a$  y  $b$ .

Tabla 4.4. Ejemplo ilustrativo sobre un modelo de regresión. N.º de reclamaciones y valor pagado para diferentes ramos

Código ramo	Reclamaciones	Valor pagado
1	65	350
2	35	250
3	20	130
4	10	75
5	15	60
6	5	20
7	30	120

*Solución.* Si fuera adecuado el modelo (4.24) para este conjunto de datos, de la información dada:

$$\bar{x} = 25,71428571$$

$$\bar{y} = 143,57142857$$

$$\sum_{i=1}^7 x_i^2 = 7100 \quad \sum_{i=1}^7 y_i^2 = 225925 \quad \sum_{i=1}^7 x_i y_i = 39450,$$

luego, a partir de (4.26) las estimaciones son:

$$\begin{aligned} \hat{a} &= \frac{\bar{y} \sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \\ &= \frac{(143,57142857)(7100) - 25,71428571(39450)}{7100 - (7)(25,7142^2)} = 1,99422 \\ \hat{b} &= \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{39450 - (7)(25,7142)(143,57142857)}{7100 - (7)(25,7142^2)} \\ &= 5,50578. \end{aligned}$$

En la figura 4.4 se presenta un gráfico de dispersión para los datos del ejemplo 4.23, agregando la línea estimada por MCO.

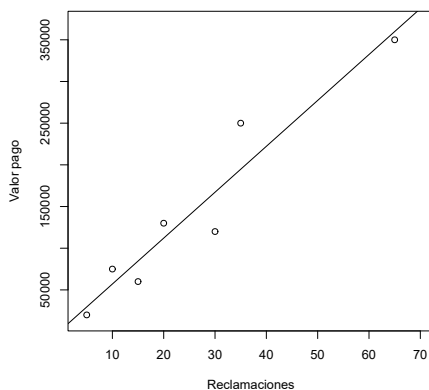


Figura 4.4. Gráfico de dispersión para los datos del ejemplo 4.23 y línea ajustada

Las instrucciones en *R* para la estimación y el gráfico se presentan a continuación:

```
remove(list=ls())
Reclamaciones=c(65,35,20,10,15,5,30)
Valor_Pago=c(350,250,130,75,60,20,120)
```

```
ajuste1=lm(Valor_Pago~Reclamaciones)
par(mfrow=c(1,1))
plot(Valor_Pago~Reclamaciones,ylim=c(15,370),xlim=c(4,70))
abline(coef(ajuste1))
```



**Ejemplo 4.24.** (Regresión lineal múltiple). *Otro caso especial es cuando:*

$$Y_i = g_i(\boldsymbol{\theta}) + \varepsilon_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad 1 \leq i \leq n, \quad (4.27)$$

con  $\boldsymbol{\theta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$ , el vector de parámetros, y  $x_{ij}$ ,  $i = 1, 2, 3, \dots, n$  y  $j = 1, 2, 3, \dots, p$ , valores conocidos (variables explicativas). Matricialmente el modelo (4.27) se expresa como:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\mathbf{y} = \mathbf{X} \boldsymbol{\theta} + \boldsymbol{\varepsilon}.$$

Encuentre la estimación por MCO de  $\boldsymbol{\theta}$ , para ello, supóngase que:

1.  $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$ .
2.  $\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top] = \sigma^2 I_n$ .
3. La matriz  $\mathbf{X}$  es no estocástica, es decir, consta de números fijos.
4. El rango de  $\mathbf{X}$  es  $\rho(\mathbf{X}) = p$ .
5.  $\boldsymbol{\varepsilon}$  tiene una distribución normal multivariada, es decir:

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I_n).$$

*Solución.* En este caso:

$$g_i(\boldsymbol{\theta}) = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_p x_{ip}, \quad i = 1, 2, 3, \dots, n.$$

El objetivo es minimizar

$$Q = \|\mathbf{y} - \mathbf{g}(\boldsymbol{\theta})\|^2 = [\mathbf{y} - \mathbf{g}(\boldsymbol{\theta})]^\top [\mathbf{y} - \mathbf{g}(\boldsymbol{\theta})] = (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\theta});$$

para ello, se deriva  $Q$  respecto a  $\boldsymbol{\theta}$ ,  $\nabla Q(\boldsymbol{\theta}) = \frac{\partial Q}{\partial \boldsymbol{\theta}}$ , y se iguala al vector nulo,  $\nabla Q(\boldsymbol{\theta}) = \mathbf{0}$ .

Si se define  $\frac{\partial}{\partial \boldsymbol{\theta}} = \left[ \left( \frac{\partial}{\partial \beta_i} \right) \right]$ , entonces:

a) para un vector  $\mathbf{a}$  de tamaño adecuado:

$$\frac{\partial(\boldsymbol{\theta}^\top \mathbf{a})}{\partial \boldsymbol{\theta}} = \mathbf{a} \quad \text{y} \quad \frac{\partial(\mathbf{a}^\top \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = (\mathbf{a}^\top)^\top = \mathbf{a} \quad (4.28)$$

b) para una matriz  $\mathbf{A}$ , simétrica, de tamaño adecuado:

$$\frac{\partial(\boldsymbol{\theta}^\top \mathbf{A} \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{A} \boldsymbol{\theta} + (\boldsymbol{\theta}^\top \mathbf{A})^\top = 2\mathbf{A} \boldsymbol{\theta}. \quad (4.29)$$

Usando las expresiones (4.28) y (4.29):

$$\begin{aligned} \nabla Q(\boldsymbol{\theta}) &= \frac{\partial Q}{\partial \boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}} (\mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{X} \boldsymbol{\theta} - \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{y} + \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\theta}) \\ &= -(\mathbf{y}^\top \mathbf{X})^\top - \mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X} \boldsymbol{\theta} \\ &= -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X} \boldsymbol{\theta}, \end{aligned} \quad (4.30)$$

igualando al vector nulo,

$$\begin{aligned} \mathbf{0} &= \frac{\partial Q}{\partial \boldsymbol{\theta}} = -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\theta}} \\ \mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\theta}} &= \mathbf{X}^\top \mathbf{y}. \end{aligned} \quad (4.31)$$

Si  $\mathbf{X}^\top \mathbf{X}$  es no singular,

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (4.32)$$

La segunda derivada de  $Q$ , es decir, la derivada de (4.30), está dada por:

$$\frac{\partial^2 Q}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = 2(\mathbf{X}^\top \mathbf{X}).$$

Puesto que  $\mathbf{X}^\top \mathbf{X}$  es de rango columna completo, esta matriz de segundas derivadas (hessiana) es definida positiva, es decir,  $\hat{\boldsymbol{\theta}}$  es un mínimo. Si  $\mathbf{X}$  no es de rango columna completo,  $\mathbf{X}^\top \mathbf{X}$  es singular, ver apéndice B.1.3.  $\square$

**Ejemplo 4.25.** Considere los datos ilustrativos presentados en la tabla 4.5 sobre el número de reclamaciones, valor pagado (en millones) y número de expuestos año póliza<sup>28</sup>, para diferentes ramos.

<sup>28</sup>Esta variable se establece como la suma de días que las personas estuvieron aseguradas en la vigencia anual de una póliza dividido por 365 días, es decir,

$$x_{1i} = \frac{\text{Total de días que las personas de la } i\text{-ésima póliza estuvieron aseguradas}}{365 \text{ días}}.$$

Tabla 4.5. Ejemplo ilustrativo sobre un modelo de regresión múltiple. N.º de reclamaciones, valor pagado para diferentes ramos y N.º de expuestos año póliza

N.º de expuestos año póliza	Reclamaciones	Valor pagado
1200,35	65	350
800,70	35	250
300,65	20	130
170,10	10	75
300,40	15	60
120,15	5	20
600,24	30	120

Si el modelo (4.27) fuera adecuado, el modelo propuesto sería:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

donde las variables aleatorias  $\varepsilon_i$  son independientes con distribución normal de media cero y varianza  $\sigma^2$ . Encuentre las estimaciones por MCO de los parámetros del modelo (4.27).

*Solución.* Las respectivas matrices  $\mathbf{y}$ ,  $\mathbf{X}$ ,  $\boldsymbol{\beta}$  y  $\boldsymbol{\varepsilon}$  son dadas por:

$$\begin{pmatrix} 350 \\ 250 \\ 130 \\ 75 \\ 60 \\ 20 \\ 120 \end{pmatrix} = \begin{pmatrix} 1 & 1200,35 & 65 \\ 1 & 800,70 & 35 \\ 1 & 300,65 & 20 \\ 1 & 170,10 & 10 \\ 1 & 300,40 & 15 \\ 1 & 120,15 & 5 \\ 1 & 600,24 & 30 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \end{pmatrix}$$

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Los cálculos para encontrar el estimador de los parámetros,  $\hat{\boldsymbol{\beta}}$ , por MCO son:

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 7 & 3492,59 & 180 \\ 3492,59 & 2666249,285 & 136875,2 \\ 180 & 136875,2 & 7100 \end{pmatrix}, \quad \mathbf{X}^\top \mathbf{y} = \begin{pmatrix} 1005 \\ 764595,3 \\ 39450 \end{pmatrix}$$

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \frac{1}{10^5} \begin{pmatrix} 41335 & -34 & -400 \\ -34 & 3,6 & -70 \\ -400 & -70 & 1370 \end{pmatrix},$$

luego,

$$\hat{\beta} = \frac{1}{1000} \begin{pmatrix} 633,43 \\ 147,27 \\ 2701,164 \end{pmatrix}.$$

En la figura 4.5 se presenta una matriz de diagrama de dispersión (*scatter plot*) para las variables Número de expuestos año póliza (*Insured*), Reclamaciones y Valor\_Pago.

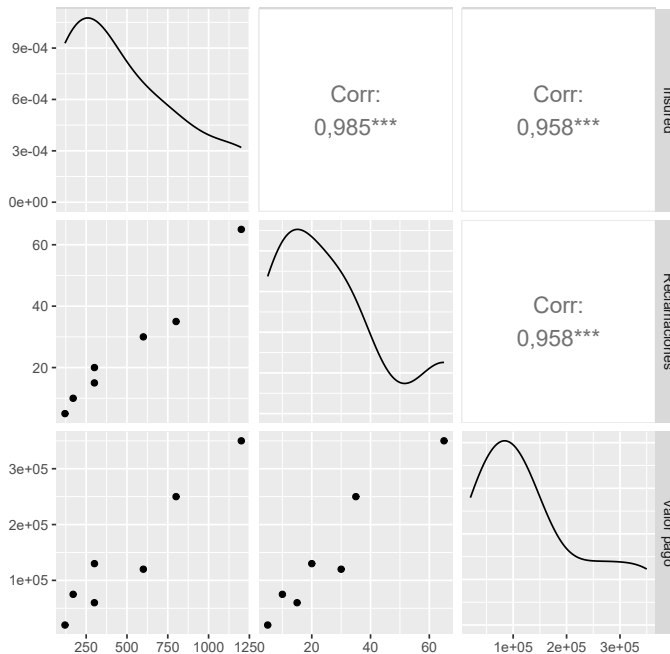


Figura 4.5. Matriz de *scatter plot* para los datos del ejemplo 4.25

En la figura 4.6 se presenta un gráfico de dispersión, y en la figura 4.7 se presenta un gráfico de dispersión agregando un plano estimado por MCO.

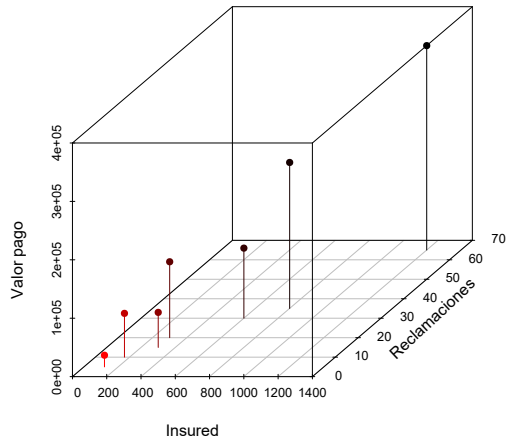


Figura 4.6. Gráfico de dispersión para los datos del ejemplo 4.25 en 3D

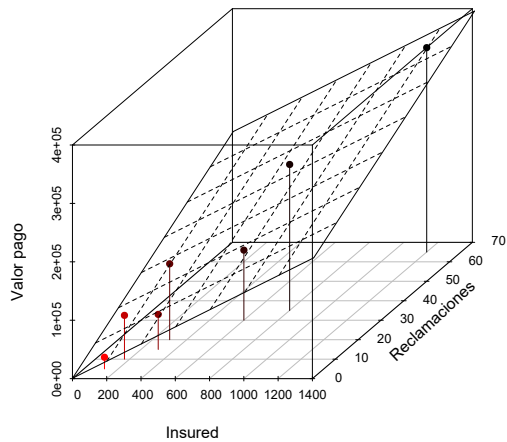


Figura 4.7. Gráfico de dispersión para los datos del ejemplo 4.25 en 3D y plano ajustado

Las instrucciones en *R* para la estimación y el gráfico se presentan a continuación:

```
remove(list=ls())
library(scatterplot3d)
library(car)
```

```

Insured=c(1200.35,800.70,300.65,170.10,300.40,120.15,600.24)
Reclamaciones=c(65,35,20,10,15,5,30)
Valor_Pago=c(350,250,130,75,60,20,120)
ajuste2=lm(Valor_Pago~Insured+Reclamaciones)
ejemplo.regm=data.frame(Insured,Reclamaciones,Valor_Pago)
pairs(ejemplo.regm, pch=19)
figura3d <-scatterplot3d(Insured,Reclamaciones,Valor_Pago,
  xlim=c(100,1250),ylim=c(4,70),zlim=c(15,370),
  angle=30,pch=16, highlight.3d=TRUE, type="h")
library(rgl)
figura3d\$plane3d(ajuste2, lty.box = "solid")
scatter3d(Valor_Pago~Insured+Reclamaciones,
  data=ejemplo.regm, fit="linear")

```

☑

## 4.4.2. Mínimos cuadrados ponderados

Supóngase que las variables aleatorias  $Y_1, Y_2, Y_3, \dots, Y_n$ , se pueden expresar de la forma

$$Y_i = g_i(\theta_1, \theta_2, \dots, \theta_k) + \varepsilon_i, \quad 1 \leq i \leq n,$$

donde  $g_i$  son funciones conocidas y  $\theta_1, \theta_2, \dots, \theta_k$  son parámetros de interés, de valor desconocido. Considere que varía libremente el vector de parámetros  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)^\top$  sobre un conjunto  $A \subset \mathbb{R}^r$ . También, suponga que los  $\varepsilon_i$  satisfacen, al menos en forma aproximada, las siguientes restricciones:

- El valor esperado del vector de perturbaciones es:

$$\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0} \quad (4.33)$$

- La matriz de varianza-covarianza del vector de perturbaciones es:

$$\boldsymbol{\Sigma} = [\sigma_{ij}] = \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^t] = \begin{cases} w_i\sigma^2, & \text{si } i = j, \\ 0, & \text{si } i \neq j. \end{cases} \quad (4.34)$$

Cabe notar que  $\sigma^2$  es, también, un parámetro desconocido, pero los  $w_i$  se consideran pesos conocidos. Este tipo de modelo está enmarcado en los llamados modelos *heterocedásticos*. Para afrontar este problema se proponen las variables aleatorias  $Z_1, Z_2, Z_3, \dots, Z_n$ , como:

$$Z_i = \frac{Y_i}{\sqrt{w_i}} = \frac{g_i(\boldsymbol{\theta})}{\sqrt{w_i}} + \frac{\varepsilon_i}{\sqrt{w_i}}, \quad (4.35)$$

entonces,

$$\mathbb{V}\left(\frac{\varepsilon_i}{\sqrt{w_i}}\right) = \frac{w_i \sigma^2}{w_i} = \sigma^2.$$

Tomando  $g_i^*(\boldsymbol{\theta}) = \frac{g_i(\boldsymbol{\theta})}{\sqrt{w_i}}$  y  $\varepsilon_i^* = \frac{\varepsilon_i}{\sqrt{w_i}}$ , se tiene que:

$$Z_i = g_i^*(\boldsymbol{\theta}) + \varepsilon_i^* \quad 1 \leq i \leq n.$$

Igual que en el método de **MCO**, sea  $\mathbf{g}(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), g_2(\boldsymbol{\theta}), \dots, g_n(\boldsymbol{\theta}))^\top$  y  $\mathbf{Y} = (Y_1, Y_2, Y_3, \dots, Y_n)^\top$ , el valor esperado de  $\mathbf{Y}$  es:

$$\mathbb{E}\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} g_1(\boldsymbol{\theta}) \\ g_2(\boldsymbol{\theta}) \\ \vdots \\ g_n(\boldsymbol{\theta}) \end{bmatrix}$$

$$\mathbb{E}(\mathbf{Y}) = \mathbf{g}(\boldsymbol{\theta}).$$

Una vez se observa  $\mathbf{Y} = \mathbf{y}$ , con  $\mathbf{y} = (y_1, y_2, y_3, \dots, y_n)^\top$ , se tiene  $\mathbf{Z} = \mathbf{z}$ , con  $\mathbf{z} = (z_1, z_2, z_3, \dots, z_n)^\top$ . El estimador obtenido por **MCP** resulta de minimizar

$$\sum_{i=1}^n [z_i - g_i^*(\boldsymbol{\theta})]^2 = \sum_{i=1}^n \frac{1}{w_i} [y_i - g_i(\boldsymbol{\theta})]^2 = [\mathbf{y} - \mathbf{g}(\boldsymbol{\theta})]^\top \mathbf{W}^{-1} [\mathbf{y} - \mathbf{g}(\boldsymbol{\theta})],$$

con  $\mathbf{W} = \text{diag}\{w_1, w_2, w_3, \dots, w_n\}$ <sup>29</sup> y rango de  $\boldsymbol{\theta}$  sobre  $A$ .

**Ejemplo 4.26.** Considérese el modelo de regresión:

$$Y_i = g_i(\boldsymbol{\theta}) + \varepsilon_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad 1 \leq i \leq n, \quad (4.36)$$

con  $\boldsymbol{\theta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$ , el vector de parámetros,  $x_{ij}$ ,  $i = 1, 2, 3, \dots, n$  y  $j = 1, 2, 3, \dots, p$ , valores conocidos (variables explicativas) y suponga que:

- $\mathbb{E}(\varepsilon_i) = 0$ ,  $1 \leq i \leq n$ .
- $\mathbb{V}(\varepsilon_i) = w_i \sigma^2 > 0$ ,  $1 \leq i \leq n$ .
- $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ ,  $1 \leq i < j \leq n$ .

Encuentre las estimaciones de los parámetros del modelo (4.36) por **MCP**.

*Solución.* En este caso:

$$g_i^*(\boldsymbol{\theta}) = \frac{1}{\sqrt{w_i}} (\beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_p x_{ip}), \quad i = 1, 2, 3, \dots, n.$$

<sup>29</sup>La notación  $\mathbf{C} = [c_{ij}] = \text{diag}\{c_1, c_2, c_3, \dots, c_a\}$  indica la matriz diagonal cuyos elementos  $c_{ii} = c_i$  y  $c_{ij} = 0$  para toda  $i \neq j$ .

Considerando (4.35) se tiene que:

$$\mathbf{Z} = \begin{pmatrix} \frac{Y_1}{\sqrt{w_1}} \\ \frac{Y_2}{\sqrt{w_2}} \\ \vdots \\ \frac{Y_n}{\sqrt{w_n}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{w_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{w_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{w_n}} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \mathbf{W}^{-\frac{1}{2}} \mathbf{Y}, \quad (4.37)$$

$$\begin{aligned} \mathbf{g}^*(\boldsymbol{\theta}) &= \begin{pmatrix} \frac{x_{11}}{\sqrt{w_1}} & \frac{x_{12}}{\sqrt{w_1}} & \cdots & \frac{x_{1p}}{\sqrt{w_1}} \\ \frac{x_{21}}{\sqrt{w_2}} & \frac{x_{22}}{\sqrt{w_2}} & \cdots & \frac{x_{2p}}{\sqrt{w_2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_{n1}}{\sqrt{w_n}} & \frac{x_{n2}}{\sqrt{w_n}} & \cdots & \frac{x_{np}}{\sqrt{w_n}} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sqrt{w_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{w_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{w_n}} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} \\ &= \mathbf{W}^{-\frac{1}{2}} \mathbf{X} \boldsymbol{\theta} \end{aligned} \quad (4.38)$$

$$\boldsymbol{\varepsilon}^* = \begin{pmatrix} \varepsilon_1^* \\ \varepsilon_2^* \\ \vdots \\ \varepsilon_n^* \end{pmatrix} = \mathbf{W}^{-\frac{1}{2}} \boldsymbol{\varepsilon}. \quad (4.39)$$

Para minimizar:

$$\begin{aligned} Q &= [\mathbf{z} - \mathbf{g}^*(\boldsymbol{\theta})]^\top [\mathbf{z} - \mathbf{g}^*(\boldsymbol{\theta})] \\ &= [\mathbf{W}^{-\frac{1}{2}} \mathbf{y} - \mathbf{W}^{-\frac{1}{2}} \mathbf{X} \boldsymbol{\theta}]^\top [\mathbf{W}^{-\frac{1}{2}} \mathbf{y} - \mathbf{W}^{-\frac{1}{2}} \mathbf{X} \boldsymbol{\theta}] \\ &= (\mathbf{y} - \mathbf{X} \boldsymbol{\theta})^\top \mathbf{W}^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\theta}), \end{aligned}$$

se deriva  $Q$  respecto a  $\boldsymbol{\theta}$ ,  $\nabla Q(\boldsymbol{\theta}) = \frac{\partial Q}{\partial \boldsymbol{\theta}}$ , y se iguala al vector nulo,  $\nabla Q(\boldsymbol{\theta}) = \mathbf{0}$ . Usando (4.28) y (4.29):

$$\begin{aligned} \nabla Q(\boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}} \left( \mathbf{y}^\top \mathbf{W}^{-1} \mathbf{y} - \mathbf{y}^\top \mathbf{W}^{-1} \mathbf{X} \boldsymbol{\theta} - \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{W}^{-1} \mathbf{y} + \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{W}^{-1} \mathbf{X} \boldsymbol{\theta} \right) \\ &= \frac{\partial}{\partial \boldsymbol{\theta}} \left( \mathbf{y}^\top \mathbf{W}^{-1} \mathbf{y} - 2\boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{W}^{-1} \mathbf{y} + \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{W}^{-1} \mathbf{X} \boldsymbol{\theta} \right) \\ &= -2\mathbf{X}^\top \mathbf{W}^{-1} \mathbf{y} + 2\mathbf{X}^\top \mathbf{W}^{-1} \mathbf{X} \boldsymbol{\theta}, \end{aligned}$$

igualando a  $\mathbf{0}$ ,

$$\begin{aligned}\mathbf{0} &= \frac{\partial Q}{\partial \boldsymbol{\theta}} = -2\mathbf{X}^\top \mathbf{W}^{-1} \mathbf{y} + 2\mathbf{X}^\top \mathbf{W}^{-1} \mathbf{X} \hat{\boldsymbol{\theta}} \\ \mathbf{X}^\top \mathbf{W}^{-1} \mathbf{X} \hat{\boldsymbol{\theta}} &= \mathbf{X}^\top \mathbf{W}^{-1} \mathbf{y}.\end{aligned}\quad (4.40)$$

Si  $\mathbf{X}^\top \mathbf{W}^{-1} \mathbf{X}$  es no singular,

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^\top \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{-1} \mathbf{y}.\quad (4.41)$$

Note que si  $\mathbf{W} = \mathbf{I}_n$  esta estimación coincide con (4.32).  $\square$

**Ejemplo 4.27.** Considerando la información del ejemplo 4.26, sea  $p = 2$ ,  $x_{i1} = 1$  y  $x_{i2} = x_i$ , para  $i = 1, \dots, n$ , entonces  $Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ ,  $i = 1, \dots, n$ . Determine los parámetros por el método ponderado.

*Solución.* Primero se crean los vectores dados en (4.37), (4.38) y (4.39), luego, usando la expresión (4.41) se tiene que:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}^\top \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{-1} \mathbf{y} \\ &= \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{pmatrix} \frac{1}{w_1} & 0 & \dots & 0 \\ 0 & \frac{1}{w_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{w_n} \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \times \\ &\quad \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} \frac{1}{w_1} & 0 & \dots & 0 \\ 0 & \frac{1}{w_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{w_n} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n \frac{1}{w_i} & \sum_{i=1}^n \frac{x_i}{w_i} \\ \sum_{i=1}^n \frac{x_i}{w_i} & \sum_{i=1}^n \frac{x_i^2}{w_i} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n \frac{y_i}{w_i} \\ \sum_{i=1}^n \frac{x_i y_i}{w_i} \end{pmatrix} \\ &= \frac{1}{\left( \sum_{i=1}^n \frac{1}{w_i} \right) \left( \sum_{i=1}^n \frac{x_i^2}{w_i} \right) - \left( \sum_{i=1}^n \frac{x_i}{w_i} \right)^2} \begin{pmatrix} \sum_{i=1}^n \frac{x_i^2}{w_i} & - \sum_{i=1}^n \frac{x_i}{w_i} \\ - \sum_{i=1}^n \frac{x_i}{w_i} & \sum_{i=1}^n \frac{1}{w_i} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n \frac{y_i}{w_i} \\ \sum_{i=1}^n \frac{x_i y_i}{w_i} \end{pmatrix} \\ &= \frac{1}{\left( \sum_{i=1}^n \frac{1}{w_i} \right) \left( \sum_{i=1}^n \frac{x_i^2}{w_i} \right) - \left( \sum_{i=1}^n \frac{x_i}{w_i} \right)^2} \left[ \begin{pmatrix} \sum_{i=1}^n \frac{x_i^2}{w_i} \\ \sum_{i=1}^n \frac{1}{w_i} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n \frac{y_i}{w_i} \\ \sum_{i=1}^n \frac{x_i y_i}{w_i} \end{pmatrix} - \begin{pmatrix} \sum_{i=1}^n \frac{x_i}{w_i} \\ \sum_{i=1}^n \frac{x_i y_i}{w_i} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n \frac{x_i y_i}{w_i} \\ \sum_{i=1}^n \frac{y_i}{w_i} \end{pmatrix} \right],\end{aligned}$$

dividiendo tanto numerador como denominador por  $\left(\sum_{i=1}^n \frac{1}{w_i}\right)^2$  y denotando  $u_i = \left(\sum_{j=1}^n \frac{1}{w_j}\right)^{-1} \frac{1}{w_i}$  se tiene que:

$$\hat{\beta} = \frac{1}{\sum_{i=1}^n u_i x_i^2 - \left(\sum_{i=1}^n u_i x_i\right)^2} \begin{bmatrix} \left(\sum_{i=1}^n u_i x_i^2\right)\left(\sum_{i=1}^n u_i y_i\right) - \left(\sum_{i=1}^n u_i x_i\right)\left(\sum_{i=1}^n u_i x_i y_i\right) \\ \left(\sum_{i=1}^n u_i x_i y_i\right) - \left(\sum_{i=1}^n u_i x_i\right)\left(\sum_{i=1}^n u_i y_i\right) \end{bmatrix},$$

nótese que si  $w_i = 1$ ,  $\forall i$  entonces  $u_i = \frac{1}{n}$  y se obtienen las estimaciones del modelo vía **MCO** dadas en (4.26).  $\square$

## 4.5. Conceptos básicos para estimación minimax

En la teoría de decisiones, la metodología estándar de evaluación-elección (toma de decisión) se facilita construyendo una matriz de decisiones. En esta matriz se recogen los resultados  $x_{ij}$  del decisor para cada *alternativa*  $a_i$  (representadas por las filas de la matriz) con respecto a cada *situación o estado de la naturaleza*  $e_j$  (ubicado por las columnas).

El criterio minimax dado en [79] consiste en comparar el resultado de una alternativa bajo una situación (probable o incierta) con todos los demás resultados, independientemente de la situación bajo la cual ocurren. Cuando la situación corresponda a un estado de la naturaleza, este no es controlable por el decisor, por lo que el resultado de una alternativa solo debe ser comparado con los resultados de las demás alternativas bajo el mismo estado de la naturaleza.

**Ejemplo 4.28.** (Concurso de canción inédita). *Supóngase, como en [34, p. 23], que en un festival de música hay siete jurados para evaluar cuatro canciones inéditas de los participantes, A, B, C y D. Cada juez valora las canciones conforme sus preferencias y atribuye 4 puntos para la primera que escoge, 3 puntos para la segunda, etc. ¿Cuál es la canción ganadora?, es decir, la que suma más puntos. Si un participante es descalificado durante o después del festival, ¿esta descalificación afecta la premiación?*

**Solución.** En la tabla 4.6 se presentan los puntajes otorgados.

Por el criterio establecido, la canción *C* es la ganadora.

Los compositores de la canción *D* no debieron entrar en el concurso porque ya habían participado con esa canción en otro festival, lo que iba en

Tabla 4.6. Puntajes de las 4 canciones en el festival de música

Canción \ Voto	Voto							Total
	$J_1$	$J_2$	$J_3$	$J_4$	$J_5$	$J_6$	$J_7$	
A	4	1	2	4	1	2	4	18
B	3	4	1	3	4	1	3	19
C	2	3	4	2	3	4	2	20
D	1	2	3	1	2	3	1	13

Fuente: adaptada de [34, p. 23].

contra del reglamento del concurso. Los organizadores del evento aceptan que hubo un error y descalifican a la canción  $D$ , manteniendo la premiación que fue divulgada, ya que  $D$  fue la de menor puntaje y no influye en los resultados. Sin embargo, los compositores de la canción  $A$  protestan, ya que si la canción  $D$  se hubiera descalificado antes de la votación, habría apenas tres canciones para escoger. En ese caso cada juez daría 3 puntos a la primera escogida, dos puntos a la segunda y un punto a la última. Como se esperaba, los jurados mantuvieron su orden de preferencia, por lo que los resultados fueron:

Tabla 4.7. Puntajes de las 3 canciones en el festival de música

Canción \ Voto	Voto							Total
	$J_1$	$J_2$	$J_3$	$J_4$	$J_5$	$J_6$	$J_7$	
A	3	1	2	3	1	2	3	15
B	2	3	1	2	3	1	2	14
C	1	2	3	1	2	3	1	13

Fuente: adaptada de [34, p. 24].

Así, la canción ganadora es la  $A$ , con lo cual los compositores de la canción  $A$  quedan satisfechos. En este caso, la premiación cambia. ☑

**Ejemplo 4.29.** Según [34, p. 16], una asociación de padres de familia está organizando un evento para recoger fondos para el colegio. Hay dos alternativas ( $a_i$ ) entre las cuales el comité organizador debe seleccionar: un día de deportes ( $a_1$ ) o un día de campo ( $a_2$ ). El ingreso (\$) dependerá del estado de la naturaleza (clima), el cual por simplicidad se considera que será seco ( $e_1$ ) o lluvioso ( $e_2$ ). El presidente del comité quiere evitar arriesgarse y resalta que si se realiza el día de deportes se garantiza una entrada de \$850 000, mientras que si se realiza el día de campo el ingreso será de \$750 000, por lo que propone que se organice el día de deportes. Los ingresos dependiendo de si llueve o no se presentan en la tabla 4.8.

Sin embargo, antes de la votación, el tesorero recuerda que es posible tomar un seguro por si llueve. La póliza paga \$500 000 si llueve y la prima es de \$100 000.

Tabla 4.8. Ingresos (\$) del evento de acuerdo al clima

Evento	Clima lluvioso	Clima seco
Día de deportes	\$850 000	\$1 200 000
Día de campo	\$750 000	\$1 500 000

Fuente: adaptada de [34, p. 16].

Con este nuevo panorama, los ingresos tomando la póliza contra lluvia, dependiendo si llueve o no, se presentan en la tabla 4.9.

Tabla 4.9. Ingresos (\$) del evento según el clima si se compra una póliza contra la lluvia

Evento	Clima lluvioso	Clima seco
Día de deportes	\$1 250 000	\$1 100 000
Día de campo	\$1 150 000	\$1 400 000

Fuente: adaptada de [34, p. 16].

El presidente del comité resalta que el día de campo, comprando el seguro, garantiza una entrada de \$1 150 000, mientras que el día de deportes garantiza solo \$1 100 000, por tanto él propone que se organice un día de campo. Sin embargo, la secretaria interviene y comenta que no hay diferencia entre tomar o no tomar el seguro contra lluvia, ya que en ambos casos el día de deportes da \$100 000 más si el día es lluvioso y \$300 000 menos si el día es seco, por lo cual ella considera que no se tome la decisión ni a partir de la tabla 4.8, ni a partir de la tabla 4.9. La secretaria pide tener en cuenta la **pérdida** o lo que “lamentarían” o “dejarían de ganar” al tomar una decisión. Lo que “dejarían de ganar” se presenta en la tabla 4.10.

Tabla 4.10. Lo que dejarían de ganar

Evento	Clima		Lo máximo que dejarían de ganar
	Lluvioso	Seco	
Día de deportes	\$ 0	\$300 000	\$300 000
Día de campo	\$100 000	\$ 0	\$100 000

Fuente: adaptada de [34, p. 17].

El presidente del comité, una vez más, analiza las alternativas y recuerda a los miembros del comité que la tabla 4.10 muestra lo que pueden dejar de ganar y no los ingresos proponiendo organizar un día de campo, ya que sería el escenario donde menos se lamentarían, garantizando que la pérdida no excedería los \$100 000. Pero el tesorero pregunta si deberían considerar o no la compra del seguro, ya que en realidad son 4 las opciones entre las cuales debería elegir el comité: día de deportes

comprando el seguro, día de deportes sin comprar el seguro, día de campo comprando el seguro y día de campo sin comprar el seguro. Para el clima lluvioso y seco, la ganancia más alta es \$ 1 250 000 y \$ 1 500 000, respectivamente. Las cuatro opciones mostrando los ingresos y las pérdidas (restando en cada una de las columnas el máximo valor a los demás elementos) se presentan en la tabla 4.11.

Tabla 4.11. Ingresos y pérdidas (entre paréntesis) para las 4 opciones

Evento	Clima		Lo máximo que dejarían de ganar
	Lluvioso	Seco	
Día de deportes sin seguro	\$850 000 (\$ 400 000)	\$ 1 200 000 (\$ 300 000)	\$ 400 000
Día de deportes con seguro	\$ 1 250 000 (\$ 0)	\$ 1 100 000 (\$ 400 000)	\$ 400 000
Día de campo sin seguro	\$ 750 000 (\$ 500 000)	\$ 1 500 000 (\$ 0)	\$ 500 000
Día de campo con seguro	\$ 1 150 000 (\$ 100 000)	\$ 1 400 000 (\$ 100 000)	\$ 100 000

Fuente: adaptada de [34, p. 17].

Finalmente, observando la tabla 4.11, el presidente del comité sugiere que se escoja la opción que garantiza que la pérdida no sea más de \$100 000, esto es organizar un día de campo y comprar el seguro. Así, el comité elige la opción del día de campo comprando el seguro (el comité usó el criterio *minimax*).

**DEFINICIÓN 4.7.** Una *función de decisión* (o una regla de decisión)  $\delta$  es una función (medible) definida de  $\mathbb{R}^n$  a  $\mathbb{R}$ . El valor de  $\delta$  evaluado en  $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$ , es decir,  $\delta(\mathbf{x})$  se llama una *decisión*<sup>30</sup> (ver [76, p. 309]).

**DEFINICIÓN 4.8.** Para estimar  $\theta$  con base en  $X_1, X_2, \dots, X_n$  y usando una función de decisión  $\delta$ , se establece una *función de pérdida*<sup>31</sup> como una función no negativa en los argumentos  $\theta$  y  $\delta(\mathbf{x})$ , la cual expresa la pérdida incurrida cuando  $\theta$  es estimada por  $\delta(\mathbf{x})$  (ver [76, p. 309]).

**Ejemplo 4.30.** Algunas funciones de pérdida son:

- $l[\theta; \delta(\mathbf{x})] = [\theta - \delta(\mathbf{x})]^2$ , función de pérdida cuadrática.
- $l[\theta; \delta(\mathbf{x})] = w(\theta)[\theta - \delta(\mathbf{x})]^2$ , función de pérdida cuadrática ponderada.

<sup>30</sup>Aunque escoger y decidir no es lo mismo, ya que escoger se puede hacer y se supone decidir que requiere pensar cuidadosamente, en teoría de la decisión no necesariamente se hace distinción entre escoger y decidir (ver [34, p. 27]).

<sup>31</sup>Se expresará por  $l$ , sigla de loss.

- $l[\theta; \delta(x)] = |\theta - \delta(x)|$ , *función de pérdida absoluta.*
- $l[\theta; \delta(x)] = \begin{cases} [\theta - \delta(x)]^2 & \text{si } |\theta - \delta(x)| \leq k \\ 2k|\theta - \delta(x)| - k^2 & \text{si } |\theta - \delta(x)| \geq k \end{cases}$ , *dada en [40].*
- $l[\theta; \delta(x)] = |\theta - \delta(x)|^p$ ,  $1 < p < 2$ .

**DEFINICIÓN 4.9.** Una *función de riesgo* correspondiente a la función de pérdida  $l[\theta; \delta(x)]$  es denotada por  $R[\theta; \delta(x)]$  y definida por:

$$R[\theta; \delta(x)] = \mathbb{E}_\theta \{l[\theta; \delta(x)]\}.$$

Es decir, la función de riesgo correspondiente a una función de decisión,  $\delta$ , es el promedio de pérdida cometido al usar dicha función de decisión (ver [76, p. 309]).

**Nota 4.3.**

- Dadas dos funciones de decisión  $\delta$  y  $\delta^*$ . Si para todo  $\theta \in \Theta$ :

$$R[\theta; \delta(x)] = R[\theta; \delta^*(x)],$$

entonces se dice que las funciones de riesgo son *equivalentes*.

- Bajo estimación puntual,  $\delta(x)$  corresponde a un estimado de  $\theta$ , donde su calidad se mira con respecto a la función de pérdida seleccionada.

**DEFINICIÓN 4.10.** El estimador  $\delta(x)$  de  $\theta$  se dice *admisibles* si no hay otro estimador de  $\theta$  tal que  $R[\theta; \delta^*(x)] \leq R[\theta; \delta(x)]$  para todo  $\theta \in \Theta$ , donde la desigualdad es estricta al menos para un  $\theta$  (ver [76, p. 310]).

**Ejemplo 4.31.** Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria y defínase

$$\delta(X) = a + bT(X), \tag{4.42}$$

donde  $T(X)$  es una estadística tal que  $\mathbb{E}_\theta [T(X)] = \theta$ . Encuentre la función de riesgo correspondiente a la función de pérdida cuadrática.

*Solución.* En este caso,  $l[\theta; \delta(X)] = [\theta - \delta(X)]^2$ , entonces:

$$\begin{aligned} R[\theta; \delta(X)] &= \mathbb{E}_\theta \{l[\theta; \delta(X)]\} = \mathbb{E}_\theta \{[\theta - \delta(X)]^2\} \\ &= \mathbb{E}_\theta \{[\theta - (a + bT(X))]^2\} = \mathbb{E}_\theta \{[(\theta - a) - bT(X)]^2\} \\ &= \mathbb{E}_\theta \{[(\theta - a) - bT(X) + b\mathbb{E}_\theta [T(X)] - b\mathbb{E}_\theta [T(X)]]^2\} \\ &= \mathbb{E}_\theta \{[(\theta - a - b\mathbb{E}_\theta [T(X)]) - b(T(X) - \mathbb{E}_\theta [T(X)])]^2\} \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}_\theta \left[ ((1-b)\theta - a)^2 \right] + b^2 \mathbb{E}_\theta \left[ (T(\mathbf{X}) - \mathbb{E}_\theta[T(\mathbf{X})])^2 \right] \\
&\quad - 2 \mathbb{E}_\theta \left[ ((1-b)\theta - a)(T(\mathbf{X}) - \mathbb{E}_\theta[T(\mathbf{X})]) \right] \\
&= b^2 \mathbb{V}_\theta [T(\mathbf{X})] + [(1-b)\theta - a]^2.
\end{aligned} \tag{4.43}$$

Esta función de riesgo es cuadrática. ✓

**DEFINICIÓN 4.11.** Un estimador  $\delta^M(x)$  de  $\theta$ , que minimiza el máximo riesgo, y satisface

$$\inf_{\delta} \sup_{\theta} R[\theta; \delta(x)] = \sup_{\theta} R[\theta; \delta^M(x)],$$

se llama un **estimador minimax** (ver [53, p. 309]).

**Teorema 4.1.** Suponga que hay una *fdp* a priori  $\pi(\theta; \lambda)$  sobre  $\Theta$ , tal que  $\delta$  es un estimador bayesiano del parámetro  $\theta$ . Si el riesgo  $R[\theta; \delta(x)]$  es constante para todo  $\theta \in \Theta$ , entonces  $\delta$  es minimax.

*Demostración.* Ver detalles de la prueba en [53, p. 311]. ✓

**Ejemplo 4.32.** Considerando la información del ejemplo 4.16 sobre ensayos Bernoulli, determine el estimador minimax usando la función de pérdida cuadrática.

*Solución.* En el ejemplo 4.19 se encontró que el estimador puntual bayesiano para  $\theta$  era:

$$\hat{\theta} = \frac{\alpha + n\bar{x}_n}{\alpha + \beta + n} = \frac{\alpha}{\alpha + \beta + n} + \frac{n}{\alpha + \beta + n} \bar{x}_n = a + b\bar{x}_n,$$

con  $\alpha + \beta \neq 0$ . Además, se sabe que si  $X_i \sim \text{Bernoulli}(\theta)$  y se toma una muestra aleatoria de tamaño  $n$ , se tiene que  $\sum_{i=1}^n X_i \sim \text{Binomial}(n, \theta)$ , luego

$$\mathbb{E}_\theta \left( \sum_{i=1}^n X_i \right) = n\theta \quad \text{y} \quad \mathbb{V}_\theta \left( \sum_{i=1}^n X_i \right) = n\theta(1 - \theta).$$

Sea en (4.42),  $T(\mathbf{X}) = \bar{X}_n$ , y como  $l[\theta; \delta(x)] = [\theta - \delta(x)]^2$ , entonces por (4.43):

$$\begin{aligned}
R[\theta; \delta(\mathbf{X})] &= b^2 \frac{\theta(1 - \theta)}{n} + [(1-b)\theta - a]^2 \\
&= b^2 \frac{\theta(1 - \theta)}{n} + (1-b)^2 \theta^2 - 2a(1-b)\theta + a^2.
\end{aligned}$$

Agrupando términos:

$$R[\theta; \delta(\mathbf{X})] = \left[ (1-b)^2 - \frac{b^2}{n} \right] \theta^2 + \left[ \frac{b^2}{n} - 2a(1-b) \right] \theta + a^2. \quad (4.44)$$

Ahora, se establecen los coeficientes de  $R[\theta; \delta(\mathbf{X})]$  para que sea constante, es decir, no dependa de  $\theta$  y así  $\delta^*$  (un caso particular de  $\delta$ ) sea un estimador minimax. Luego, los coeficientes que acompañan a  $\theta^2$  y a  $\theta$  en la expresión (4.44) deben ser iguales a cero, esto es,

$$\begin{aligned} (1-b)^2 - \frac{b^2}{n} = 0 & \quad \text{y} \quad \frac{b^2}{n} - 2a(1-b) = 0 \\ (1-b)^2 = \frac{b^2}{n} & \quad \text{y} \quad 2a(1-b) = \frac{b^2}{n}, \end{aligned}$$

usando el método de igualación y dado que  $b \neq 1$ :

$$(1-b)^2 = 2a(1-b) \quad \Rightarrow \quad 1-b = 2a,$$

al sustituir  $a$  y  $b$  se obtiene:

$$\begin{aligned} 1 - \frac{n}{\alpha + \beta + n} &= 2 \frac{\alpha}{\alpha + \beta + n} \\ \alpha + \beta &= 2\alpha, \end{aligned}$$

luego  $\alpha = \beta$  y sustituyendo en la primera expresión, se tiene que:

$$\begin{aligned} \left( 1 - \frac{n}{\alpha + \beta + n} \right)^2 &= \frac{1}{n} \left( \frac{n}{\alpha + \beta + n} \right)^2 \\ (\alpha + \beta)^2 &= n \\ (2\alpha)^2 &= n \\ \alpha &= \frac{\sqrt{n}}{2}. \end{aligned}$$

Puesto que  $\delta(\mathbf{X}) = a + b\bar{X}_n$ , reemplazando  $a$  y  $b$  se llega a:

$$\begin{aligned} \delta^*(\mathbf{X}) &= \frac{\alpha + n\bar{X}_n}{\alpha + \beta + n} = \frac{\frac{\sqrt{n}}{2} + n\bar{X}_n}{\frac{\sqrt{n}}{2} + \frac{\sqrt{n}}{2} + n} = \frac{\cancel{\sqrt{n}} \left( \frac{1}{2} + \sqrt{n}\bar{X}_n \right)}{\cancel{\sqrt{n}}(1 + \sqrt{n})} \\ &= \frac{1 + 2\sqrt{n}\bar{X}_n}{2(1 + \sqrt{n})}, \end{aligned}$$

y por la expresión (4.44):

$$R[\theta; \delta^*(\mathbf{X})] = \frac{\alpha^2}{(\alpha + \beta + n)^2} = \frac{n}{[2(n + \sqrt{n})]^2} = \frac{1}{[2(1 + \sqrt{n})]^2},$$

la cual no depende de  $\theta$  y, por el teorema 4.1,  $\delta^*(\mathbf{X})$  es minimax.  $\square$

## 4.6. Ejercicios

1. Una variable aleatoria  $X$ , con distribución de Gumbel tiene como fda a [62, p. 138]:

$$F_X(x) = \exp \left\{ - \exp \left\{ - \frac{x - \theta_1}{\theta_2} \right\} \right\}.$$

Las componentes del parámetro  $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$  son tales que  $\theta_1 \in \mathbb{R}$ ,  $\theta_2 > 0$ . Explore la forma de estimar puntualmente el parámetro  $\boldsymbol{\theta}$ , teniendo en cuenta que  $\mathbb{E}(X) = \theta_1 + \gamma\theta_2$ , siendo la constante de Euler,  $\gamma \approx 0,577216$  y, además,  $\mathbb{V}(X) = \frac{\pi^2\theta_2^2}{6}$ .

2. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid tal que la fdp de  $X_1$  es dada por [21, p. 43]:

$$f_{X_1}(x) = \frac{2(\theta - x)}{\theta^2} I_{(0, \theta)}(x).$$

Halle los estimadores de  $\theta$  y  $\mathbb{V}(X)$  por el método de los momentos.

3. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid tal que la fdp de  $X_1$  es beta de parámetros  $\alpha$  y  $\beta$  [21, p. 43], es decir,

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} I_{(0,1)}(x),$$

donde  $\alpha > 0$  y  $\beta > 0$ . Si  $\alpha + \beta = 5$ , encuentre los estimadores de  $\alpha$  y  $\beta$  por el método de los momentos.

4. Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de la distribución normal inversa  $NI(\theta, \lambda)$ , con  $\lambda$  y  $\theta$  desconocidos [20, p. 303],

$$f(x; \theta) = \left( \frac{\lambda}{2\pi x^3} \right)^{\frac{1}{2}} e^{-\frac{\lambda(x-\theta)^2}{2\theta^2 x}} \quad \text{para } x > 0.$$

Encuentre los **EMV** de  $\theta$  y  $\lambda$ .

5. Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de la **fmp** [20, p. 366]:

$$f(x; \theta) = \left(\frac{\theta}{2}\right)^{|x|} (1-\theta)^{1-|x|}, \quad x = \{-1, 0, 1\}; \quad 0 \leq \theta \leq 1.$$

Encuentre el **EMV** para  $\theta$ .

6. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias **iid**. Para cada una de las siguientes **fdp**, encontrar el **EMV**.

- a) Para el parámetro  $\theta$  desconocido:

$$p(x; \theta) = \theta x^{-2}, \quad 0 < \theta \leq x < \infty.$$

- b) Para el parámetro  $\mu$  considerando que  $\lambda$  es conocido:

$$p(x; \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}} \quad \text{para } x > 0.$$

- c) Para el parámetro  $\theta$  desconocido [65, p. 367]:

$$f(x; \theta) = \exp\{-(x-\theta) - e^{-(x-\theta)}\} \quad \text{con } -\infty < \theta < \infty.$$

- d) Para el parámetro  $\alpha$  considerando que  $\beta$  es conocido:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \left(\frac{x}{\beta}\right)^\alpha & \text{si } 0 \leq x \leq \beta \\ 1 & \text{si } x > \beta. \end{cases}$$

7. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias **iid** normales con media  $\vartheta$  y varianza  $\sigma^2$ , y sea  $\vartheta \sim N(\mu, \tau^2)$ , se considera que  $\mu$ ,  $\tau^2$  y  $\sigma^2$  son conocidos. Encontrar la distribución *a posteriori* de  $\vartheta$  dado  $\mathbf{X} = \mathbf{x}$  y el estimador de Bayes.

8. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias **iid** normales con media  $\mu$  y varianza  $\vartheta$ , y sea  $\frac{1}{\vartheta} \sim \chi_{\frac{1}{2}}^2$ , se considera que  $\mu$  y  $k$  son conocidos. Encontrar la distribución *a posteriori* de  $\frac{1}{\vartheta}$  dado  $\mathbf{X} = \mathbf{x}$  y el estimador de Bayes.

9. En el ejemplo 5.17, mostrar que  $\hat{\theta} = 1 + \bar{X}$  es un máximo.

10. Sea el modelo  $y_{ij} = \mu + i\alpha + e_{ij}$ , con  $i = 1, \dots, a$  y  $j = 1, \dots, r$ . Encontrar la estimación por MCO de  $\begin{pmatrix} \hat{\mu} \\ \hat{\alpha} \end{pmatrix}$ .
11. ¿Cómo quedan las estimaciones de  $a$  y  $b$  del ejemplo 4.22 si en vez de usar  $x_i$  y  $y_i$ , se usan rangos? (Suponga que no hay empates).
12. Considere para el modelo  $y_i = a + bx_i + \epsilon_i$ ,  $i = 1, 2, 3$ , el siguiente conjunto de datos:

$i$	1	2	3
$x_i$	2	-6	7
$y_i$	3	4	6

Encuentre las estimaciones por MCO de  $a$  y de  $b$ , suponiendo que:

- a)  $\mathbb{E}(\epsilon_i) = 0$  y  $\mathbb{V}(\epsilon_i) = \sigma^2$ ,  $i = 1, 2, 3$ .
- b)  $\mathbb{E}(\epsilon_i) = 0$  y  $\mathbb{V}(\epsilon_1) = \sigma^2$ ,  $\mathbb{V}(\epsilon_2) = 4\sigma^2$  y  $\mathbb{V}(\epsilon_3) = 9\sigma^2$ .
13. Una compañía de seguros desea determinar el grado de relación que existe entre el ingreso familiar  $x$  y el monto de seguro de vida  $Y$  del jefe de familia (en miles de dólares) [19, p. 492]. En la tabla 4.12 se presentan los datos para una muestra aleatoria de 18 familias. Si fuera

Tabla 4.12. Ingreso y monto del seguro de vida del jefe de familia

Ingreso (en miles)	Seguro de vida	Ingreso (en miles)	Seguro de vida	Ingreso (en miles)	Seguro de vida
45	70	47	90	15	40
20	50	30	55	35	65
40	60	25	55	40	75
40	50	20	35	55	105
50	110	60	120	15	30
30	40	35	65	45	80

Fuente: adaptada de [19, p. 492].

adecuado el modelo (4.24) para este conjunto de datos, obtener las estimativas de  $a$  y  $b$ . Realizar un gráfico de dispersión con los datos de ingreso y seguro de vida, presentando la recta ajustada.

14. Una compañía desea predecir el costo unitario de fabricación,  $Y$ , como una función de la tasa de producción,  $x_1$ , y de los costos,  $x_2$ . En la tabla 4.13 se presentan los datos de 12 meses.

Tabla 4.13. Datos sobre el valor unitario, tasa de producción y costos

Valor unitario	Tasa de producción	Costos	Valor unitario	Tasa de producción	Costos
14,39	85	80	16,70	80	95
16	80	105	21,20	65	115
24,60	50	125	22,25	60	130
19,90	70	116	14,9	90	90
15,10	95	94	16,50	100	110
15,90	100	115	17,40	80	115

Si el modelo (4.27) fuera adecuado para este conjunto de datos, obtener una estimación para  $\beta$ . Realizar un gráfico de dispersión, presentando la recta ajustada.

15. ¿Cómo quedan las estimaciones de  $\beta_1$  y  $\beta_2$  del ejemplo 4.27 si en vez de usar  $x_i$  y  $y_i$ , se usan rangos? (Suponga que no hay empates).
16. Sea  $p = 2$ ,  $x_{i1} = 1$  y  $x_{i2} = x_i$ , para  $i = 1, \dots, n$ , entonces,

$$Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \quad i = 1, \dots, n.$$

Si  $\mathbb{V}(\varepsilon_i) = \sigma^2(1 - h_{ii})$ , donde  $h_{ii}$  es el elemento  $i$ -ésimo de la diagonal de la matriz hat,  $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ , ¿cómo quedan las estimaciones de  $\beta_1$  y  $\beta_2$ ?

17. En el ejemplo 4.8, encontrar el EMV cuando  $\sum_{i=1}^n x_i = 0$  o cuando  $\sum_{i=1}^n x_i = n$ , respectivamente.





En el capítulo anterior se estudiaron algunos métodos para establecer estimadores de un parámetro desconocido  $\theta$ . Dado que se pueden considerar candidatas para ser empleadas como estimadores de los parámetros desconocidos de una fdp, una o varias estadísticas, entonces se deben estudiar algunas características con el fin de escoger el mejor estimador posible. Pero, ¿qué características hacen que un estimador sea bueno? Hay varias respuestas a esta pregunta. En este capítulo se establecen algunas propiedades óptimas que deben tener los estimadores puntuales.

## 5.1. Suficiencia

Una vez adoptado un modelo estadístico, se debe separar la información irrelevante que está presente en los datos, y que puede entorpecer el entendimiento de la situación.

La idea de suficiencia es reducir los datos con “estadísticos” cuyo uso implique *no* perder información sobre el(los) parámetro(s) de interés. En otras palabras, una *estadística suficiente* para un parámetro  $\theta$  es una estadística que capture, de alguna forma, *toda* la información contenida en la muestra acerca de  $\theta$  (ver [20, p. 272]).

**DEFINICIÓN 5.1.** Una estadística  $T(\mathbf{X})$  es llamada **suficiente** para un parámetro  $\theta$ , si y solo si, la distribución condicional de  $\mathbf{X}$  dado  $T(\mathbf{X}) = t$  no involucra a  $\theta$  (ver [12, p. 42]).

**Ejemplo 5.1.** Según [12, p. 42], si  $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$  con  $X_i$  son funciones indicadoras de un conjunto de  $n$  ensayos Bernoulli con probabilidad de éxito  $\theta$ , demostrar que  $T(\mathbf{x}) = \sum_{i=1}^n X_i$  es una estadística suficiente para  $\theta$ .

*Solución.* Sean

$$X_i = \begin{cases} 0 & \text{si el ensayo resulta en fracaso} \\ 1 & \text{si el ensayo resulta en éxito} \end{cases}$$

y la estadística

$$Y = T(\mathbf{x}) = \sum_{i=1}^n X_i,$$

la cual denota el número total de ensayos que resultaron en éxito. Entonces,

$$P(Y = y) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}.$$

Por otra parte, la fmp conjunta está dada por:

$$\begin{aligned} P(\mathbf{X} = \mathbf{x}, Y = y) &= \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} \\ &= \theta^y (1 - \theta)^{n-y}. \end{aligned}$$

Usando la expresión (2.44) se tiene que:

$$p(x|y) = \frac{\theta^y (1 - \theta)^{n-y}}{\binom{n}{y} \theta^y (1 - \theta)^{n-y}} = \frac{1}{\binom{n}{y}}.$$

Por lo tanto,  $P(\mathbf{X} = \mathbf{x} | T(\mathbf{x}) = y)$  no depende de  $\theta$ , el parámetro, así:

$$T(\mathbf{x}) = \sum_{i=1}^n X_i$$

es una estadística suficiente para  $\theta$ . □

**Ejemplo 5.2.** De acuerdo con [53, p. 33], dadas  $X_1$  y  $X_2$  variables aleatorias independientes con fmp Poisson y esperanza común  $\lambda$ , mostrar que  $X_1 + X_2$  es una estadística suficiente para  $\lambda$ .

*Solución.* Su fmp conjunta es dada por:

$$P_\lambda(X_1 = x_1, X_2 = x_2) = \frac{\lambda^{x_1+x_2} e^{-2\lambda}}{x_1! x_2!}$$

y la fmp de  $S = X_1 + X_2$  es Poisson con esperanza  $2\lambda$  (ver ejemplo 2.8). Usando (2.44) la distribución condicional de  $X_1$  dado  $S = t$  está dada por:

$$\begin{aligned} P(X_1 = x_1 | S = t) &= \frac{P(X_1 = x_1, S = t)}{P(S = t)} = \frac{P(X_1 = x_1, X_2 = t - x_1)}{P(S = t)} \\ &= \frac{\frac{\lambda^{x_1} e^{-\lambda}}{x_1! (t-x_1)!}}{\frac{e^{-2\lambda} (2\lambda)^t}{t!}} = \frac{\frac{\lambda^{x_1} e^{-\lambda}}{x_1! (t-x_1)!}}{\frac{2^t e^{-2\lambda} \lambda^t}{t!}} = \frac{t!}{x_1! (t-x_1)!} \times \frac{1}{2^t} = \binom{t}{x_1} \frac{1}{2^t}, \end{aligned}$$

esto es,  $X_1$ , bajo la condición  $S = t$ , tiene una distribución binomial de parámetros  $t$  y  $\frac{1}{2}$ , luego,  $P(X_1 = x_1 | S = t)$  no depende de  $\lambda$ , por ende,

$$T \left[ \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \right] = S = X_1 + X_2$$

es una estadística suficiente para  $\lambda$ .

Generalizando, si se tiene una muestra aleatoria con fmp Poisson de parámetro  $\lambda$ , entonces  $T(\mathbf{X}) = \sum_{i=1}^n X_i$  es una estadística suficiente para  $\lambda$ .  $\square$

**Teorema 5.1. Criterio de factorización de Fisher-Neyman.** Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria, una estadística  $T(\mathbf{x})$  con rango  $\mathcal{T}$  es suficiente para  $\theta$  si y solo si existe una función  $g(t, \theta)$  definida para  $t$  en  $\mathcal{T}$  y  $\theta \in \Theta$  y una función  $h$  definida sobre  $\mathbb{R}^n$  tal que [12, p. 43]:

$$p(\mathbf{x}; \theta) = g(T(\mathbf{x}); \theta) h(\mathbf{x}) \quad \forall \mathbf{x} \in \mathbb{R}^n, \theta \in \Theta. \quad (5.1)$$

*Demostración.* Se prueba para el caso discreto: sean  $x_1, x_2, \dots$  un conjunto de posibles realizaciones del vector aleatorio  $\mathbf{X}$  y sea  $T(x_j) = t_j$ . Entonces  $T(\mathbf{X})$  es discreta y

$$\sum_{i=1}^{\infty} P_{\theta}[T(\mathbf{X}) = t_i] = 1 \quad \text{para cada } \theta.$$

( $\Leftarrow$ ) Se necesita probar que  $P_{\theta}[\mathbf{X} = \mathbf{x}_j | T(\mathbf{X}) = t_i]$  es independiente de  $\theta$  sobre cada uno de los conjuntos

$$S_i = \{\theta : P_{\theta}(T(\mathbf{X}) = t_i) > 0\} \quad i = 1, 2, 3, \dots$$

Si la expresión (5.1) es cierta, entonces,

$$P_{\theta}(T(\mathbf{X}) = t_i) \quad \text{y} \quad P_{\theta}(\mathbf{X} = \mathbf{x}_j, T(\mathbf{X}) = t_i)$$

se pueden expresar como:

$$\begin{aligned} P_{\theta}(T(\mathbf{X}) = t_i) &= \sum_{\{\mathbf{x}: T(\mathbf{x})=t_i\}} p(\mathbf{x}; \theta) \\ &= \sum_{\{\mathbf{x}: T(\mathbf{x})=t_i\}} g(T(\mathbf{x}); \theta) h(\mathbf{x}) \\ &= \sum_{\{\mathbf{x}: T(\mathbf{x})=t_i\}} g(t_i; \theta) h(\mathbf{x}) \\ &= g(t_i; \theta) \sum_{\{\mathbf{x}: T(\mathbf{x})=t_i\}} h(\mathbf{x}). \end{aligned}$$

Por otra parte, la fmp conjunta es dada por:

$$\underbrace{P_{\theta}(\mathbf{X} = \mathbf{x}_j, T(\mathbf{X}) = t_i)}_{p(\mathbf{x}_j; \theta)} = \begin{cases} g(t_i; \theta)h(\mathbf{x}_j) & \text{si } T(\mathbf{x}_j) = t_i \\ 0 & \text{si } T(\mathbf{x}_j) \neq t_i. \end{cases}$$

Luego,  $P_\theta [X = x_j | T(X) = t_i] = p(x_j; \theta)$  es:

$$p(x_j; \theta) = \begin{cases} \frac{g(t_i; \theta)h(x_j)}{\sum_{\{x; T(x)=t_i\}} h(x)} & \text{si } T(x_j) = t_i \\ 0 & \text{si } T(x_j) \neq t_i \end{cases}$$

$$= \begin{cases} \frac{h(x_j)}{\sum_{\{x; T(x)=t_i\}} h(x)} & \text{si } T(x_j) = t_i \\ 0 & \text{si } T(x_j) \neq t_i, \end{cases}$$

como  $P_\theta [X = x_j | T(X) = t_i]$  no depende de  $\theta$ , entonces  $T(X)$  es una estadística suficiente.

( $\Rightarrow$ ) Se tiene que  $T(X)$  es una estadística suficiente. Se definen  $g(t_i; \theta)$  y  $h(x)$  como sigue:

$$g(t_i; \theta) = P_\theta(T(X) = t_i)$$

$$h(x) = P(X = x | T = T(x)) \left( \begin{array}{l} h(x) \text{ no depende de } \theta \text{ porque} \\ T(X) \text{ es una estadística} \\ \text{suficiente} \end{array} \right).$$

Entonces,

$$\begin{aligned} p(x; \theta) &= P_\theta(X = x) = P_\theta(X = x, T = T(x)) \\ &= P_\theta(X = x | T = T(x)) P_\theta(T = T(x)) \\ &= h(x)g(T(x); \theta). \end{aligned} \quad \square$$

**Ejemplo 5.3.** De acuerdo con [20, p. 274], si  $X_1, X_2, \dots, X_n$  son variables aleatorias iid tal que  $X_1 \sim N(\theta, \sigma^2)$ , con  $\theta$  parámetro desconocido y  $\sigma^2$  conocida. Encontrar una estadística suficiente para  $\theta$ .

*Solución.* En este caso, la fdp conjunta de  $X_1, X_2, \dots, X_n$  es dada por:

$$\begin{aligned} p(x_1, \dots, x_n; \theta) &= \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\left\{\frac{-1}{2\sigma^2}(x_i - \theta)^2\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta - \bar{x} + \bar{x})^2\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \times \end{aligned}$$

$$\begin{aligned}
& \exp \left\{ \frac{-1}{2\sigma^2} \left[ \sum_{i=1}^n (x_i - \bar{x})^2 + 2(\bar{x} - \theta) \underbrace{\sum_{i=1}^n (x_i - \bar{x})}_0 + \sum_{i=1}^n (\bar{x} - \theta)^2 \right] \right\} \\
&= \underbrace{(2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ \frac{-1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right\}}_{h(x)} \underbrace{\exp \left\{ \frac{-n}{2\sigma^2} (\bar{x} - \theta)^2 \right\}}_{g(\bar{x};\theta)} \\
&= h(x)g(\bar{x};\theta).
\end{aligned}$$

Luego, la estadística  $T(x) = \bar{x}$  es suficiente para  $\theta$ . ✓

**Ejemplo 5.4.** De acuerdo con [20, p. 277], si  $X_1, X_2, \dots, X_n$  son variables aleatorias *iid* con *fmp* uniforme discreta sobre  $1, 2, \dots, \theta$ . Encuentre una estadística suficiente para  $\theta$ .

*Solución.* En este caso, la *fmp* conjunta de  $X_1, X_2, \dots, X_n$  es:

$$P_\theta(X = x) = \begin{cases} \theta^{-n} & x_i \in \{1, 2, 3, \dots, \theta\} \text{ para } i = 1, 2, 3, \dots, n \\ 0 & \text{en otro caso,} \end{cases}$$

note que el soporte de la distribución incluye al parámetro de interés  $\theta$ . Para afrontar esta situación se reescribe la distribución de probabilidad  $P_\theta(X = x)$  como el producto de una función  $h(x)$ , que no depende del parámetro  $\theta$ , dada por:

$$h(x) = I_{\{1, 2, 3, \dots\}}(x_i) = \begin{cases} 1 & x_i \in \{1, 2, 3, \dots\} \text{ para } i = 1, 2, 3, \dots, n \\ 0 & \text{en otro caso,} \end{cases}$$

y una función  $g(t; \theta)$ , que depende de  $T(x) = \max_{1 \leq i \leq n} \{x_i\} = x_{(n)}$ , dada por:

$$g(t; \theta) = \theta^{-n} I_{\{t \leq \theta\}}(t) = \begin{cases} \theta^{-n} & t \leq \theta \\ 0 & \text{en otro caso.} \end{cases}$$

Cabe notar que la restricción  $t \leq \theta$  se satisface automáticamente por todos los valores de  $\theta$ , afrontando así el problema de que el soporte de la distribución dependa del parámetro (ver [24, p. 20]). Luego,

$$h(x)g(t; \theta) = I_{\{1, 2, 3, \dots\}}(x_i) \theta^{-n} I_{\{t \leq \theta\}}(t)$$

$$\begin{aligned}
&= \begin{cases} \theta^{-n} & \text{si } \{x_i \in \{1, 2, 3, \dots\} \text{ para } i = 1, 2, 3, \dots, n\} \text{ y } \{t \leq \theta\} \\ 0 & \text{en otro caso} \end{cases} \\
&= \begin{cases} \theta^{-n} & \text{si } x_i \in \{1, 2, 3, \dots, \theta\} \text{ para } i = 1, 2, 3, \dots, n \\ 0 & \text{en otro caso} \end{cases} \\
&= P_\theta(\mathbf{X} = \mathbf{x}),
\end{aligned}$$

entonces  $T(x) = \max_{1 \leq i \leq n} \{x_i\} = x_{(n)}$  es una estadística suficiente para  $\theta$ .  $\square$

Cuando es necesario estimar más de un parámetro, se tendrá un vector de parámetros, digamos  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_s)^\top$ . En tal caso, es usual que un vector  $T(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X}), \dots, T_r(\mathbf{X}))^\top$  sea el estadístico suficiente para  $\boldsymbol{\theta}$ . Es frecuente que  $s = r$ .

**Ejemplo 5.5.** De acuerdo con [12, p. 44], dadas  $X_1, X_2, \dots, X_n$  variables aleatorias iid normales con media  $\mu$  y varianza  $\sigma^2$ , desconocidas. Encontrar una estadística suficiente para  $\boldsymbol{\theta} = (\mu, \sigma^2)^\top$ .

*Solución.* En este caso, la fdp conjunta de  $X_1, X_2, \dots, X_n$  es dada por:

$$\begin{aligned}
p(x_1, \dots, x_n; \boldsymbol{\theta}) &= \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\left\{\frac{-1}{2\sigma^2}(x_i - \mu)^2\right\} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^n (x_i^2 - 2x_i\mu + \mu^2)\right\} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^n (x_i^2 - 2\mu x_i)\right\} \exp\left\{-\frac{n\mu^2}{2\sigma^2}\right\} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{n\mu^2}{2\sigma^2}\right\} \exp\left\{\frac{-1}{2\sigma^2} \left[\sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i\right]\right\} \\
&= h(\mathbf{x})g(T(\mathbf{x}); \boldsymbol{\theta}).
\end{aligned}$$

Luego, la estadística  $T(\mathbf{x}) = \left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2\right)^\top$  es suficiente para  $\boldsymbol{\theta} = (\mu, \sigma^2)^\top$ .  $\square$

**Nota 5.1.** Según [20, p. 280], una muestra completa,  $\mathbf{X}$ , en sí es una estadística suficiente, ya que en particular:

$$p(\mathbf{x}; \theta) = \underbrace{g(T(\mathbf{x}); \theta)}_{T(\mathbf{x})=\mathbf{x}} \underbrace{h(\mathbf{x})}_1 .$$

Luego, cualquier función uno a uno de una estadística suficiente es una estadística suficiente.

**Nota 5.2.** Si existe una estadística suficiente  $T(\mathbf{X})$  para el parámetro  $\theta$ , la distribución a posteriori dependerá de los datos solo mediante  $T(\mathbf{X})$ . Más aún, la distribución a posteriori es igual a la distribución condicional de  $\vartheta$  dado  $T(\mathbf{X})$ .

**Ejemplo 5.6.** En el ejemplo 5.1 sobre ensayos Bernoulli, se estableció que la estadística  $T(\mathbf{X}) = \sum_{i=1}^n X_i$  era suficiente para  $\theta =$  probabilidad de éxito y en el ejemplo 4.16 se obtuvo que la distribución a posteriori depende de los datos solo mediante  $\sum_{i=1}^n X_i$ .

## 5.2. Estadística suficiente minimal

Una estadística suficiente es de mayor interés cuando es minimal, en otras palabras, cuando da la mayor reducción de los datos posibles sin perder información del parámetro (ver [51]).

**DEFINICIÓN 5.2.** Según [20, p. 280], una estadística suficiente  $T(\mathbf{X})$  se dice que es una **estadística suficiente minimal** si para cualquier otra estadística suficiente  $T'(\mathbf{X})$ , se tiene que  $T(\mathbf{x})$  es una función de  $T'(\mathbf{x})$ .

**Nota 5.3.** En otras palabras, una estadística suficiente minimal  $T(\mathbf{X})$  es necesariamente una función de cualquier otra estadística suficiente que se pueda construir (ver [24, p. 23]).

**Nota 5.4.** Según [20, p. 280], decir que  $T(\mathbf{x})$  es una función de  $T'(\mathbf{x})$  significa que si  $T'(\mathbf{x}) = T'(\mathbf{y})$  entonces  $T(\mathbf{x}) = T(\mathbf{y})$ .

**Teorema 5.2. Criterio para suficiencia minimal.** De acuerdo con [20, p. 281], dada  $p(\mathbf{x}; \theta)$  una *fmp* o una *fdp* de una muestra  $\mathbf{X}$ . Supóngase que existe una función  $T(\mathbf{x})$  tal que, para cada dos puntos muestrales,  $\mathbf{x}$  y  $\mathbf{y}$ , la razón

$$\frac{p(\mathbf{x}; \theta)}{p(\mathbf{y}; \theta)} \tag{5.2}$$

es constante como una función de  $\theta$  si y solo si  $T(\mathbf{x}) = T(\mathbf{y})$ . Entonces  $T(\mathbf{X})$  es una estadística suficiente minimal para  $\theta$ .

*Demostración.* Ver detalles de la prueba en [20, p. 281]. ☑

**Ejemplo 5.7.** Considerando la información del ejemplo 5.1, sobre ensayos Bernoulli, pruebe que  $T(\mathbf{X}) = \sum_{i=1}^n X_i$  es una estadística suficiente minimal para  $\theta$ .

*Solución.* Se sabe que  $T(\mathbf{X}) = \sum_{i=1}^n X_i$  tiene distribución  $B(n, \theta)$ . Por otra parte, en el ejemplo 5.1, se concluyó que  $T(\mathbf{X})$  es una estadística suficiente para  $\theta$ . Evaluando el cociente dado en (5.2):

$$\frac{p(\mathbf{x}; \theta)}{p(\mathbf{y}; \theta)} = \frac{\theta^{t_x} (1 - \theta)^{n - t_x}}{\theta^{t_y} (1 - \theta)^{n - t_y}} = \theta^{t_x - t_y} (1 - \theta)^{n - t_x - (n - t_y)} = \left( \frac{\theta}{1 - \theta} \right)^{t_x - t_y},$$

se concluye que no depende de  $\theta$  si  $t_x - t_y = 0$ , esto es, si  $\sum_{i=1}^n X_i = \sum_{j=1}^n Y_j$ . Así,

$T(\mathbf{X}) = \sum_{i=1}^n X_i$  es una estadística suficiente minimal para  $\theta$ . ☑

**Ejemplo 5.8.** Considerando la información del ejemplo 5.2, pruebe que la estadística  $T(\mathbf{X}) = \sum_{i=1}^n X_i$  es suficiente minimal para  $\lambda$ .

*Solución.* Se sabe que  $T(\mathbf{X}) = \sum_{i=1}^n X_i$  es una estadística suficiente para  $\lambda$ .

Además, se sabe que  $T(\mathbf{X}) = \sum_{i=1}^n X_i$  tiene distribución  $P(n\lambda)$ . Evaluando el cociente dado en (5.2):

$$\frac{p(\mathbf{x}; \lambda)}{p(\mathbf{y}; \lambda)} = \frac{\prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}}{\prod_{i=1}^n \frac{e^{-\lambda} \lambda^{y_i}}{y_i!}} = \frac{\frac{e^{-n\lambda} \lambda^{t_x}}{\prod_{i=1}^n x_i!}}{\frac{e^{-n\lambda} \lambda^{t_y}}{\prod_{i=1}^n y_i!}} = \left( \prod_{i=1}^n \frac{y_i!}{x_i!} \right) \lambda^{t_x - t_y}$$

se concluye que no depende de  $\lambda$  si  $t_x - t_y = 0$ , esto es, si  $\sum_{i=1}^n X_i = \sum_{j=1}^n Y_j$ . Así,

$T(\mathbf{X}) = \sum_{i=1}^n X_i$  es una estadística suficiente minimal para  $\lambda$ . ☑

**Nota 5.5.** Una estadística suficiente minimal no es única.

**DEFINICIÓN 5.3.** Según [20, p. 282], una estadística  $S(\mathbf{X})$  cuya distribución no depende del parámetro  $\theta$  se llama una **estadística ancilar**<sup>32</sup>.

**Ejemplo 5.9.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid normales con media  $\mu$  y varianza 1. Muestre que  $(n - 1)S^2$  es una estadística ancilar para  $\mu$ .

*Solución.* Sea

$$S^2 = \frac{1}{n - 1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Notando que

$$\begin{aligned} (n - 1)S^2 &= \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) = \left( \sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) \\ &= \left( \begin{pmatrix} X_1 & \dots & X_n \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} - \frac{1}{n} \begin{pmatrix} X_1 & \dots & X_n \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \right) \\ &= \left( \mathbf{X}^\top \mathbf{I}_n \mathbf{X} - \frac{1}{n} \mathbf{X}^\top \underbrace{\mathbf{1}_n \mathbf{1}_n^\top}_{\mathbf{J}_n} \mathbf{X} \right) = \mathbf{X}^\top \underbrace{\left( \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \right)}_{\mathbf{A}} \mathbf{X} = \mathbf{X}^\top \mathbf{A} \mathbf{X}, \end{aligned}$$

es una forma cuadrática en  $\mathbf{X}$ . Se tiene que la distribución<sup>33</sup> de  $(n - 1)S^2$  es

$$(n - 1)S^2 \sim \chi^2(r, \lambda),$$

con

$$r = \text{rango}(\mathbf{A}) = \text{rango}\left(\mathbf{I}_n - \frac{1}{n} \mathbf{J}_n\right) = \text{tr}\left(\mathbf{I}_n - \frac{1}{n} \mathbf{J}_n\right) = n - 1,$$

aquí se usó que la traza de una matriz idempotente es igual al rango de la matriz (ver [38, p. 133,135]), y

$$\begin{aligned} \lambda &= \frac{1}{2} \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} = \frac{1}{2} \boldsymbol{\mu}^\top \mathbf{1}_n^\top \left( \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \right) \mathbf{1}_n \boldsymbol{\mu} \\ &= \frac{\mu^2}{2} \left( \underbrace{\mathbf{1}_n^\top \mathbf{1}_n}_n - \frac{1}{n} \underbrace{\mathbf{1}_n^\top \mathbf{1}_n}_n \underbrace{\mathbf{1}_n^\top \mathbf{1}_n}_n \right) = 0. \end{aligned}$$

<sup>32</sup>Para una revisión amplia de la utilidad de las estadísticas ancilares, puede consultar [35].  
<sup>33</sup> Si  $Y \sim N(\boldsymbol{\mu}, \mathbf{V})$  y  $\mathbf{Q} = \mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ , entonces  $\mathbf{Q} \sim \chi^2(r, \lambda)$  con  $r = \text{rango}(\mathbf{A})$  y parámetro de no centralidad,  $\lambda = \frac{1}{2} \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}$ , si y solo si  $\mathbf{A} \mathbf{V}$  es idempotente (ver [39, p. 51]).

Por tanto,  $(n-1)S^2 \sim \chi_{n-1}^2$  no depende de  $\mu$  indicando que  $(n-1)S^2$  es una estadística ancilar para  $\mu$ .  $\square$

### 5.3. Completez

**DEFINICIÓN 5.4.** (Completez). Sea  $p(t; \theta)$  una *fdp* o una *fmp* para un estadístico  $T(\mathbf{X})$ . La familia de distribuciones de probabilidad se llama **completa** si para todo  $\theta$ ,

$$\mathbb{E}_\theta \{k[T(\mathbf{X})]\} = 0, \quad \text{implica que} \quad P_\theta \{k[T(\mathbf{X})] = 0\} = 1,$$

donde  $k(\cdot)$  es una función cualquiera. Equivalentemente,  $T(\mathbf{X})$  es llamada una estadística completa (ver [20, p.285]).

**Nota 5.6.** Una estadística  $T(\mathbf{X})$  es llamada completa si una condición necesaria para tener que  $\mathbb{E}_\theta \{k[T(\mathbf{X})]\} = 0$ , para todo  $\theta$ , es que  $k[T(\mathbf{X})] = 0$  excepto posiblemente en un conjunto de medida cero. La propiedad de completez garantiza la unicidad en los resultados de algunos procesos estadísticos basados en  $T(\mathbf{X})$ , ver [24, p. 30].

**Nota 5.7.** En [51] se aclara que una estadística  $T(\mathbf{X})$  es completa si no contiene información ancilar o, de igual forma, si toda la información ancilar es independiente de  $T(\mathbf{X})$ .

**Ejemplo 5.10.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias *iid* Bernoulli de parámetro  $\theta$ ,  $0 < \theta < 1$ . Encuentre una estadística completa para  $\theta$ .

*Solución.* Del ejemplo 5.1 se tiene que  $T(\mathbf{x}) = \sum_{i=1}^n X_i$  es una estadística suficiente para  $\theta$ . Para chequear si  $T(\mathbf{X})$  es una estadística completa suponga que la función  $k(t)$  es tal que satisface:

$$\mathbb{E}_\theta \{k[T(\mathbf{X})]\} = 0, \quad \forall \theta, \quad \text{donde} \quad T \sim \text{Bin}(n, \theta).$$

Entonces,

$$\begin{aligned} \mathbb{E}_\theta \{k[T(\mathbf{X})]\} &= \sum_{t=0}^n k(t) \binom{n}{t} \theta^t (1-\theta)^{n-t} \\ &= (1-\theta)^n \sum_{t=0}^n k(t) \binom{n}{t} \underbrace{\left(\frac{\theta}{1-\theta}\right)^t}_{\gamma} = (1-\theta)^n \sum_{t=0}^n k(t) \binom{n}{t} \gamma^t = 0. \end{aligned}$$

La suma corresponde a un polinomio de grado  $n$  en  $\gamma$ ,  $0 < \gamma < \infty$ , que es igual a cero, implicando que los coeficientes  $k(t) = 0$  para  $t = 0, 1, 2, \dots, n$ , entonces  $T(\mathbf{x}) = \sum_{i=1}^n X_i$  es una estadística completa.  $\checkmark$

**Ejemplo 5.11.** De acuerdo con [20, p. 286], dadas  $X_1, X_2, \dots, X_n$  variables aleatorias *iid* de una distribución uniforme sobre  $(0, \theta)$ , con  $\theta$  un parámetro positivo pero desconocido. Encontrar una estadística completa para  $\theta$ .

*Solución.* En este caso, la **fdp** de  $X_1$  es dada por:

$$p(x_1; \theta) = \begin{cases} \frac{1}{\theta}, & \text{si } 0 < x_1 < \theta \\ 0, & \text{en otro caso} \end{cases}$$

y la **fdp** conjunta de  $X_1, X_2, \dots, X_n$  es dada por:

$$p(\mathbf{x}; \theta) = \begin{cases} \frac{1}{\theta^n}, & \text{si } 0 < x_i < \theta \text{ para } i = 1, \dots, n \\ 0, & \text{en otro caso.} \end{cases}$$

La restricción  $x_i \in (0, \theta)$ , para  $i = 1, \dots, n$ , se puede reescribir como

$$x_i \in (0, +\infty), \quad \text{y} \quad \max_{1 \leq i \leq n} \{x_i\} \leq \theta.$$

Entonces<sup>34</sup>,

$$\begin{aligned} p(\mathbf{x}; \theta) &= \begin{cases} \frac{1}{\theta^n} & \text{si } x_i \in (0, +\infty) \text{ para } i = 1, \dots, n, \text{ y } \max_{1 \leq i \leq n} \{x_i\} \leq \theta \\ 0 & \text{en otro caso} \end{cases} \\ &= \frac{1}{\theta^n} I(\{x_i \in (0, +\infty), i = 1, \dots, n\}) I\left(\max_{1 \leq i \leq n} \{x_i\} \leq \theta\right), \end{aligned}$$

usando el teorema 5.1, criterio de factorización de Fisher-Neyman, para  $t = T(\mathbf{x}) = \max_{1 \leq i \leq n} \{x_i\} = x_{(n)}$ , y

$$g(t; \theta) = \frac{1}{\theta^n} I(t \leq \theta) \quad \text{y} \quad h(\mathbf{x}) = I(x_i \in (0, +\infty), i = 1, \dots, n).$$

Luego,  $t = x_{(n)}$  es suficiente para  $\theta$ . Para chequear si  $T(\mathbf{X})$  es una estadística completa suponga que la función  $k(t)$  es tal que satisface

$$\mathbb{E}_\theta \{k[T(\mathbf{X})]\} = 0, \quad \forall \theta,$$

<sup>34</sup>Ver pie de página 23 del capítulo 4.

y usando la expresión (3.11) obtenida en el ejemplo 3.9, se tiene que:

$$f_{X_{(n)}}(t) = n\theta^{-n}t^{n-1}, \quad \text{si } 0 < t < \theta,$$

por lo tanto,

$$\mathbb{E}_\theta(k(T(X))) = \int_0^\theta k(t)n\theta^{-n}t^{n-1}dt = 0, \quad (5.3)$$

luego, si  $k(\cdot)$  es continua, derivando con respecto a  $\theta$ :

$$\begin{aligned} 0 &= \frac{d}{d\theta} [\mathbb{E}_\theta(k(T(X)))] = n \frac{d}{d\theta} \left[ \theta^{-n} \int_0^\theta k(t)t^{n-1}dt \right] \\ &= n \left( -n\theta^{-n-1} \int_0^\theta k(t)t^{n-1}dt + \theta^{-n} \underbrace{\frac{d}{d\theta} \int_0^\theta k(t)t^{n-1}dt}_{k(\theta)\theta^{n-1}} \right). \end{aligned}$$

Usando la expresión (5.3):

$$\begin{aligned} n \left( \theta^{-n} \theta^{n-1} k(\theta) - \underbrace{\frac{1}{\theta} \int_0^\theta n\theta^{-n}t^{n-1}k(t)dt}_{\mathbb{E}_\theta\{k[T(X)]\}=0} \right) &= 0 \quad (5.4) \\ n\theta^{-1}k(\theta) &= 0. \end{aligned}$$

El paso (5.4) se debe al primer teorema fundamental del cálculo (ver nota B.11). Como  $\theta > 0$  y para  $n \neq 0$  se tiene que  $k(\theta) = 0$  y es cierto para todo  $\theta > 0$ , entonces  $T(X)$  es una *estadística completa*.  $\square$

## 5.4. Error cuadrático medio e insesgamiento

Una buena propiedad para un estimador es que su valor promedio coincida con el parámetro a estimar. Esta propiedad se precisa en este apartado.

**DEFINICIÓN 5.5.** *El error cuadrático medio (ECM) de un estimador  $W$  de un parámetro  $\theta$  es la función de  $\theta$  definida por [20, p. 330]:*

$$\mathbb{E}_\theta [(W - \theta)^2].$$

**Nota 5.8.** El **ECM** mide la diferencia al cuadrado promedio entre el estimador  $W$  y el parámetro  $\theta$ .

$$\begin{aligned}
 \mathbb{E}_\theta [(W - \theta)^2] &= \mathbb{E}_\theta [(W - \mathbb{E}_\theta(W) + \mathbb{E}_\theta(W) - \theta)^2] \\
 &= \mathbb{E}_\theta [\{ [W - \mathbb{E}_\theta(W)] + [\mathbb{E}_\theta(W) - \theta] \}^2] \\
 &= \mathbb{E}_\theta [\{W - \mathbb{E}_\theta(W)\}^2 + (\mathbb{E}_\theta(W) - \theta)^2 \\
 &\quad + 2\{W - \mathbb{E}_\theta(W)\}(\mathbb{E}_\theta(W) - \theta)] \\
 &= \mathbb{E}_\theta [\{W - \mathbb{E}_\theta(W)\}^2] + \mathbb{E}_\theta [(\mathbb{E}_\theta(W) - \theta)^2] \\
 &\quad + 2\mathbb{E}_\theta [\{W - \mathbb{E}_\theta(W)\} \{\mathbb{E}_\theta(W) - \theta\}] \\
 &= \mathbb{E}_\theta [\{W - \mathbb{E}_\theta(W)\}^2] + \mathbb{E}_\theta [\{\mathbb{E}_\theta(W) - \theta\}^2] \\
 &\quad + 2[\mathbb{E}_\theta(W) - \mathbb{E}_\theta(W)] [\mathbb{E}_\theta(W) - \theta] \\
 &= \underbrace{\mathbb{E}_\theta [(W - \mathbb{E}_\theta(W))^2]}_{\mathbb{V}_\theta(W)} + \underbrace{\mathbb{E}_\theta [(\mathbb{E}_\theta(W) - \theta)^2]}_{B_\theta(W)} \\
 &= \mathbb{V}_\theta(W) + B_\theta^2(W)
 \end{aligned}$$

**DEFINICIÓN 5.6.** El **sesgo** de un estimador puntual  $W$  de un parámetro  $\theta$  es la diferencia entre el valor esperado de  $W$  y  $\theta$  (ver [20, p. 330]):

$$B_\theta(W) = \mathbb{E}_\theta(W) - \theta$$

Un estimador cuyo sesgo (en  $\theta$ ) es idénticamente igual a cero, es llamado **insesgado** y satisface  $\mathbb{E}_\theta(W) = \theta$ , para todo  $\theta$ .

**Ejemplo 5.12.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias **iid**  $Exp(\theta)$ . Del ejemplo 4.14 se sabe que el **EMV** es  $\hat{\theta} = \frac{1}{\bar{X}}$ . Determine si es insesgado.

*Solución.* Puesto que,  $Y = \sum_{i=1}^n X_i \sim \Gamma(n, \theta)$ , por la expresión (1.25):

$$f_Y(y) = \frac{\theta}{\Gamma(n)} (\theta y)^{n-1} e^{-\theta y}; \quad y > 0.$$

Luego,

$$\mathbb{E}_\theta \left( \frac{1}{Y} \right) = \int_0^\infty \frac{1}{y} f_Y(y) dy = \int_0^\infty \frac{\theta^2}{\Gamma(n)} (\theta y)^{n-2} e^{-\theta y} dy,$$

haciendo el cambio de variable  $u = \theta y$ ,  $du = \theta dy$  se tiene que:

$$\begin{aligned}\mathbb{E}_\theta \left( \frac{1}{\bar{Y}} \right) &= \int_0^\infty \frac{\theta^2}{\Gamma(n)} u^{n-2} e^{-u} \frac{du}{\theta} = \frac{\theta}{\Gamma(n)} \underbrace{\int_0^\infty u^{n-2} e^{-u} du}_{\Gamma(n-1)} \\ &= \theta \frac{\Gamma(n-1)}{\Gamma(n)} = \frac{\theta}{n-1}, \quad n > 1.\end{aligned}$$

Por lo tanto,

$$\mathbb{E}_\theta(\hat{\theta}) = \mathbb{E}_\theta \left( \frac{1}{\bar{X}} \right) = \mathbb{E}_\theta \left( \frac{n}{\bar{Y}} \right) = n \mathbb{E}_\theta \left( \frac{1}{\bar{Y}} \right) = \frac{n}{n-1} \theta \neq \theta,$$

es decir, es un estimador sesgado para  $\theta$ . ☑

**Ejemplo 5.13.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias *iid*  $N(\theta, \sigma^2)$  con  $\theta$  y  $\sigma^2$  desconocidos. Del ejemplo 4.15 se sabe que los estimadores máximos verosímiles para  $\theta$  y  $\sigma^2$  son  $\hat{\theta} = \bar{x}$  y  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ , respectivamente. Determine si son insesgados.

*Solución.* Los sesgos de estos estimadores puntuales son dados por:

$$\begin{aligned}\mathbb{E}_\theta(\bar{X}) - \theta &= \mathbb{E}_\theta \left( \frac{1}{n} \sum_{i=1}^n X_i \right) - \theta = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\theta(X_i) - \theta \\ &= \frac{n}{n} \theta - \theta = 0,\end{aligned}$$

luego  $\bar{X}$  es un estimador insesgado para  $\theta$ .

$$\begin{aligned}\mathbb{E}_\theta \left( \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right) - \sigma^2 &= \frac{1}{n} \mathbb{E}_\theta \left( \sum_{i=1}^n (X_i - \theta + \theta - \bar{X})^2 \right) - \sigma^2 \\ &= \frac{1}{n} \mathbb{E}_\theta \left( \sum_{i=1}^n [(X_i - \theta)^2 + 2(X_i - \theta)(\theta - \bar{X}) + (\theta - \bar{X})^2] \right) - \sigma^2 \\ &= \frac{1}{n} \left[ \sum_{i=1}^n \mathbb{E}_\theta [(X_i - \theta)^2] + 2 \sum_{i=1}^n \mathbb{E}_\theta [(X_i - \theta)(\theta - \bar{X})] + \right. \\ &\quad \left. \sum_{i=1}^n \mathbb{E}_\theta [(\theta - \bar{X})^2] \right] - \sigma^2 \\ &\quad \underbrace{\hspace{10em}}_{\mathbb{V}_\theta(\bar{X})}\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} \left( n\mathbb{V}_\theta(X_1) + 2\mathbb{E}_\theta \left[ (\theta - \bar{X}) \left( \sum_{i=1}^n X_i - n\theta \right) \right] + n\mathbb{V}_\theta(\bar{X}) \right) - \sigma^2 \\
&= \frac{1}{n} \left( n\mathbb{V}_\theta(X_1) - 2n\mathbb{E}_\theta [(\bar{X} - \theta)^2] + n\mathbb{V}_\theta(\bar{X}) \right) - \sigma^2 \\
&= \mathbb{V}_\theta(X_1) - \mathbb{V}_\theta(\bar{X}) - \sigma^2 \\
&= -\frac{1}{n^2} \mathbb{V}_\theta \left[ \sum_{i=1}^n X_i \right] \quad (\text{como las } X\text{'s son independientes}) \\
&= -\frac{n}{n^2} \mathbb{V}_\theta(X_1) = -\frac{\sigma^2}{n} \neq 0,
\end{aligned}$$

luego  $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  es un estimador sesgado para  $\sigma^2$ . □

## 5.5. Estimador insesgado de mínima varianza (UMVUE)

En las secciones previas se han mencionado algunas propiedades para dicho estimador como el insesgamiento y la propiedad de que tenga varianza pequeña. Dichos estimadores están centrados en torno al valor  $\theta$  y, por lo tanto, varían poco alrededor de esta cantidad. En general, nos interesa buscar estimadores insesgados que tengan la varianza más pequeña posible.

**Teorema 5.3.** (Rao-Blackwell). *De acuerdo con [20, p. 342], si  $W(\mathbf{X})$  es cualquier estimador insesgado de una función  $q(\theta)$ , y  $T(\mathbf{X})$  es una estadística suficiente para  $\theta$ , definiendo  $\phi[T(\mathbf{X})] = \mathbb{E}_W[W(\mathbf{X})|T(\mathbf{X})]$ . Entonces, para todo  $\theta$ :*

$$\mathbb{E}_\theta\{\phi[T(\mathbf{X})]\} = q(\theta) \quad \text{y} \quad \mathbb{V}_\theta\{\phi[T(\mathbf{X})]\} \leq \mathbb{V}_\theta[W(\mathbf{X})],$$

es decir,  $\phi(T(\mathbf{X}))$  es un estimador insesgado uniformemente mejor de  $q(\theta)$ .

Uniformemente mejor significa que  $\phi[T(\mathbf{X})]$  siempre es de menor (o igual) varianza que el estimador  $W(\mathbf{X})$ .

*Demostración.* Del enunciado del teorema se tiene que:

- $T(\mathbf{X})$  es una estadística suficiente para  $\theta$ .
- $\phi[T(\mathbf{X})] = \mathbb{E}_W[W(\mathbf{X})|T(\mathbf{X})]$ .
- $\mathbb{E}_\theta[W(\mathbf{X})] = q(\theta)$  porque  $W(\mathbf{X})$  es un estimador insesgado de  $q(\theta)$  que depende solo de la muestra (no depende de  $\theta$ ).

Para demostrar el teorema se verificará que  $\phi(T)$  sea un estimador insesgado y que sea uniformemente mejor:

1. Para el insesgamiento, usando la esperanza condicional, expresión (2.14):

$$q(\theta) = \mathbb{E}_\theta[W(\mathbf{X})] = \mathbb{E}_\theta \{ \mathbb{E}_W [W(\mathbf{X})|T(\mathbf{X})] \} = \mathbb{E}_\theta \{ \phi[T(\mathbf{X})] \},$$

por tanto,  $\phi[T(\mathbf{X})]$  es insesgado para  $q(\theta)$ .

2. Para la varianza, usando la expresión (2.18):

$$\begin{aligned} \mathbb{V}_\theta[W(\mathbf{X})] &= \mathbb{V}_\theta \{ \mathbb{E}_W [W(\mathbf{X})|T(\mathbf{X})] \} + \mathbb{E}_\theta \{ \mathbb{V}_W [W(\mathbf{X})|T(\mathbf{X})] \} \\ &= \mathbb{V}_\theta \{ \phi[T(\mathbf{X})] \} + \mathbb{E}_\theta \left\{ \underbrace{\mathbb{V}_W [W(\mathbf{X})|T(\mathbf{X})]}_{\geq 0} \right\} \\ &\geq \mathbb{V}_\theta \{ \phi[T(\mathbf{X})] \}. \end{aligned}$$

Por tanto,  $\phi[T(\mathbf{X})]$  es uniformemente mejor que  $W(\mathbf{X})$  ya que su varianza es menor o igual.  $\square$

Este teorema nos indica que, el valor esperado condicional de un estimador insesgado dada una estadística suficiente es otro estimador insesgado de menor varianza. De este se resalta (ver [65, p. 324]) que si el estimador insesgado  $W(\mathbf{X})$  es una función solo de la estadística suficiente  $T(\mathbf{X})$ , entonces la estadística derivada  $\phi(T(\mathbf{X}))$  es idéntica a  $W(\mathbf{X})$ .

**Ejemplo 5.14.** *Considérese, como en [65, p. 322], que  $T(\mathbf{X}) = \sum_{i=1}^n X_i$  es una estadística suficiente para  $\theta$ , la probabilidad de éxito de los ensayos binomiales (ver ejemplo 5.1). Probar que  $X_1$  es un estimador insesgado de  $\theta$ .*

*Solución.* Sea  $W(\mathbf{X}) = X_1$ , entonces,

$$\begin{aligned} \phi(T(\mathbf{X})) &= \mathbb{E}_W [W(\mathbf{X})|T(\mathbf{X})] = \mathbb{E}_W \left[ X_1 \left| \sum_{i=1}^n X_i = y \right. \right] \quad 0 \leq y \leq n \\ &= 1 \times P \left[ X_1 = 1 \left| \sum_{i=1}^n X_i = y \right. \right] + 0 \times P \left[ X_1 = 0 \left| \sum_{i=1}^n X_i = y \right. \right]. \end{aligned}$$

Dado que  $X_i$  es una función indicadora del ensayo Bernoulli, los valores que puede tomar son 0 o 1, con:

$$\begin{aligned} P \left[ X_1 = 1 \left| \sum_{i=1}^n X_i = y \right. \right] &= \frac{P \left[ X_1 = 1, \sum_{i=1}^n X_i = y \right]}{P \left[ \sum_{i=1}^n X_i = y \right]} \\ &= \frac{P \left[ X_1 = 1, \sum_{i=2}^n X_i = y - 1 \right]}{P \left[ \sum_{i=1}^n X_i = y \right]} \\ &= \frac{\theta^1 (1 - \theta)^{1-1} \binom{n-1}{y-1} \theta^{y-1} (1 - \theta)^{n-1-(y-1)}}{\binom{n}{y} \theta^y (1 - \theta)^{n-y}} \\ &= \frac{\theta \frac{(n-1)!}{(n-1-(y-1)!) (y-1)!} \theta^{y-1} (1 - \theta)^{n-y}}{\frac{n!}{(n-y)! y!} \theta^y (1 - \theta)^{n-y}} = \frac{y}{n} \end{aligned}$$

y

$$\begin{aligned} P \left[ X_1 = 0 \left| \sum_{i=1}^n X_i = y \right. \right] &= \frac{P \left[ X_1 = 0, \sum_{i=1}^n X_i = y \right]}{P \left[ \sum_{i=1}^n X_i = y \right]} \\ &= \frac{P \left[ X_1 = 0, \sum_{i=2}^n X_i = y \right]}{P \left[ \sum_{i=1}^n X_i = y \right]} = \frac{\theta^0 (1 - \theta)^{1-0} \binom{n-1}{y} \theta^y (1 - \theta)^{n-1-y}}{\binom{n}{y} \theta^y (1 - \theta)^{n-y}} \\ &= \frac{(1 - \theta) \frac{(n-1)!}{y! (n-1-y)!} (1 - \theta)^{n-1}}{\frac{n!}{y! (n-y)!}} = \frac{n - y}{n}, \end{aligned}$$

entonces,

$$\phi(T(\mathbf{X})) = \mathbb{E}_W \left[ X_1 \left| \sum_{i=1}^n X_i = y \right. \right] = 1 \times \frac{y}{n} + 0 \times \frac{n - y}{n} = \frac{y}{n} = \bar{X}.$$

Note que las respectivas varianzas son:

$$\begin{aligned} \mathbb{V}_\theta[W(\mathbf{X})] &= \mathbb{V}_\theta[X_1] = \theta(1 - \theta) \\ \mathbb{V}_\theta[\phi(T(\mathbf{X}))] &= \mathbb{V}_\theta \left[ \frac{Y}{n} \right] = \frac{1}{n^2} \mathbb{V}_\theta[Y] = \frac{1}{n^2} n\theta(1 - \theta) = \frac{\theta(1 - \theta)}{n}, \end{aligned}$$

por tanto,  $\mathbb{V}_\theta[W(\mathbf{X})] \geq \mathbb{V}_\theta[\phi(T(\mathbf{X}))]$ . Así  $\phi(T(\mathbf{X})) = \frac{Y}{n}$  es un estimador insesgado de  $\theta$  de menor varianza que  $W(\mathbf{X}) = X_1$ .  $\square$

Aunque el procedimiento Rao-Blackwell mejora un estimador insesgado a través de una estadística suficiente, surge la pregunta: ¿cómo saber si  $\phi[T(\mathbf{X})]$  es el mejor estimador?

**Teorema 5.4.** Si  $W(\mathbf{X})$  es un mejor estimador insesgado de  $q(\theta)$ , entonces  $W(\mathbf{X})$  es único (ver [20, p. 343]).

*Demostración.* Suponga que  $W'(\mathbf{X})$  es otro mejor estimador insesgado, y considere el estimador ponderado

$$\tilde{W}(\mathbf{X}) = \frac{1}{2}[W(\mathbf{X}) + W'(\mathbf{X})],$$

entonces, por el teorema 1.6, la esperanza de  $\tilde{W}(\mathbf{X})$  es:

$$\mathbb{E}_\theta[\tilde{W}(\mathbf{X})] = \frac{1}{2}\mathbb{E}_\theta[W(\mathbf{X}) + W'(\mathbf{X})] = \frac{1}{2}[q(\theta) + q(\theta)] = q(\theta) \quad (5.5)$$

y la varianza es dada por la expresión (1.43):

$$\begin{aligned} \mathbb{V}_\theta(\tilde{W}(\mathbf{X})) &= \mathbb{V}_\theta \left[ \frac{1}{2}(W(\mathbf{X}) + W'(\mathbf{X})) \right] = \frac{1}{4}\mathbb{V}_\theta [W(\mathbf{X}) + W'(\mathbf{X})] \\ &= \frac{1}{4} [\mathbb{V}_\theta[W(\mathbf{X})] + \mathbb{V}_\theta[W'(\mathbf{X})] + 2\text{Cov}_\theta[W(\mathbf{X}), W'(\mathbf{X})]]. \end{aligned} \quad (5.6)$$

Usando la desigualdad de Cauchy-Schwarz (nota 2.3) se tiene que:

$$\begin{aligned} \mathbb{V}_\theta[\tilde{W}(\mathbf{X})] &= \frac{1}{4} [\mathbb{V}_\theta[W(\mathbf{X})] + \mathbb{V}_\theta[W'(\mathbf{X})] + 2\text{Cov}_\theta[W(\mathbf{X}), W'(\mathbf{X})]] \\ &\leq \frac{1}{4}\mathbb{V}_\theta[W(\mathbf{X})] + \frac{1}{4}\mathbb{V}_\theta[W'(\mathbf{X})] + \frac{1}{2}\{\mathbb{V}_\theta[W(\mathbf{X})]\mathbb{V}_\theta[W'(\mathbf{X})]\}^{\frac{1}{2}} \\ &= \frac{1}{4}\mathbb{V}_\theta[W(\mathbf{X})] + \frac{1}{4}\mathbb{V}_\theta[W(\mathbf{X})] + \frac{1}{2}\left(\{\mathbb{V}_\theta[W(\mathbf{X})]\}^2\right)^{1/2}, \end{aligned}$$

porque  $\mathbb{V}_\theta[W(\mathbf{X})] = \mathbb{V}_\theta[W'(\mathbf{X})]$ , ya que ambos son mejores estimadores, así que:

$$\mathbb{V}_\theta[\tilde{W}(\mathbf{X})] \leq \mathbb{V}_\theta[W(\mathbf{X})],$$

lo cual es contradictorio porque  $W(\mathbf{X})$  es mejor estimador insesgado, al menos que:

$$\mathbb{V}_\theta[\tilde{W}(\mathbf{X})] = \mathbb{V}_\theta[W(\mathbf{X})] \quad \text{para todo } \theta. \quad (5.7)$$

Dado que la desigualdad proviene de utilizar Cauchy-Schwarz, se sabe que la igualdad ocurre si

$$W'(\mathbf{X}) = a(\theta)W(\mathbf{X}) + b(\theta).$$

El valor esperado y la varianza de  $W'(\mathbf{X})$  son:

$$\begin{aligned}\mathbb{E}_\theta[W'(\mathbf{X})] &= a(\theta)\mathbb{E}_\theta[W(\mathbf{X})] + b(\theta) \\ \mathbb{V}_\theta[W'(\mathbf{X})] &= [a(\theta)]^2\mathbb{V}_\theta[W(\mathbf{X})],\end{aligned}\tag{5.8}$$

por otra parte, usando la propiedad **iv)** de covarianzas:

$$\begin{aligned}\mathbb{C}_{\text{ov}_\theta}[W(\mathbf{X}), W'(\mathbf{X})] &= \mathbb{C}_{\text{ov}_\theta}[W(\mathbf{X}), a(\theta)W(\mathbf{X}) + b(\theta)] \\ &= a(\theta) \underbrace{\mathbb{C}_{\text{ov}_\theta}[W(\mathbf{X}), W(\mathbf{X})]}_{\mathbb{V}_\theta[W(\mathbf{X})]}.\end{aligned}$$

Despejando  $\mathbb{C}_{\text{ov}_\theta}[W(\mathbf{X}), W'(\mathbf{X})]$  de (5.6) y usando (5.7) se tiene que:

$$\begin{aligned}\mathbb{C}_{\text{ov}_\theta}[W(\mathbf{X}), W'(\mathbf{X})] &= 2\left\{\mathbb{V}_\theta[\tilde{W}(\mathbf{X})] - \frac{1}{4}\mathbb{V}_\theta[W(\mathbf{X})] - \frac{1}{4}\underbrace{\mathbb{V}_\theta[W'(\mathbf{X})]}\right\} \\ &= 2\left\{\mathbb{V}_\theta[\tilde{W}(\mathbf{X})] - \frac{1}{4}\mathbb{V}_\theta[W(\mathbf{X})] - \frac{1}{4}\mathbb{V}_\theta[W(\mathbf{X})]\right\} \\ &= 2\mathbb{V}_\theta[\tilde{W}(\mathbf{X})] - \mathbb{V}_\theta[W(\mathbf{X})] = \mathbb{V}_\theta[W(\mathbf{X})],\end{aligned}$$

por tanto, igualando las expresiones de covarianza,  $a(\theta) = 1$ , y reemplazando en (5.8):

$$\mathbb{E}_\theta[W'(\mathbf{X})] = \mathbb{E}_\theta[W(\mathbf{X})] + b(\theta),$$

luego  $b(\theta) = 0$  porque  $W(\mathbf{X})$  y  $W'(\mathbf{X})$  son estimadores insesgados de  $q(\theta)$ , concluyendo que:

$$W'(\mathbf{X}) = \underbrace{a(\theta)}_1 W(\mathbf{X}) + \underbrace{b(\theta)}_0 = W(\mathbf{X}).$$

Por lo que se concluye que  $W(\mathbf{X})$  es única. □

El siguiente teorema complementa el teorema de Rao-Blackwell al brindar condiciones para obtener un estimador insesgado de mínima varianza (**UMVUE**) (por sus siglas en inglés *uniformly minimum-variance unbiased estimator*).

**Teorema 5.5.** (Lehmann-Scheffé). Sea  $T$  un estimador insesgado de valor real de la función paramétrica  $\tau(\theta)$  donde el parámetro desconocido  $\theta \in \Theta \subseteq \mathbb{R}^k$ . Sea  $U$  una estadística suficiente completa para  $\theta$ . Entonces, la estadística  $W = \mathbb{E}_\theta[T|U]$  es el único *UMVUE* de  $\tau(\theta)$  (ver [67, p. 372]).

*Demostración.* Ver detalles de la prueba en [76, p. 312]. ☑

## 5.6. Cota inferior de Cramér-Rao

**DEFINICIÓN 5.7.** Sea  $X$  una variable aleatoria con *fdp*  $f(x; \theta)$ , para todo  $\theta \in \Theta$ ,

$$I(\theta) = \mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \ln f(x; \theta) \right)^2 \right] = \mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \ell(\theta; x) \right)^2 \right] \quad (5.9)$$

se le llama *información de Fisher* (ver [20, p. 338]).

A continuación se estudia un resultado interesante que establece que no es posible hacer que la varianza de un estimador insesgado sea arbitrariamente pequeña.

**Teorema 5.6.** (Desigualdad de Cramér-Rao). De acuerdo con [20, p. 335], si  $X_1, \dots, X_n$  es una muestra aleatoria con *fdp* conjunta  $f(x; \theta)$  y el estimador  $W(\mathbf{X}) = W(X_1, \dots, X_n)$  satisface que

$$\frac{d}{d\theta} (\mathbb{E}_\theta [W(\mathbf{X})]) = \int \cdots \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} [W(\mathbf{X})f(x; \theta)] dx$$

y  $\mathbb{V}_\theta [W(\mathbf{X})] < \infty$ , entonces,

$$\mathbb{V}_\theta [W(\mathbf{X})] \geq \frac{\left\{ \frac{d}{d\theta} \mathbb{E}_\theta [W(\mathbf{X})] \right\}^2}{\mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \ell(\theta; x) \right)^2 \right]} = \frac{\left\{ \frac{d}{d\theta} \mathbb{E}_\theta [W(\mathbf{X})] \right\}^2}{I_n(\theta)}, \quad (5.10)$$

donde  $I_n(\theta)$  es la información de Fisher dada en (5.9).

*Demostración.* Se parte de la desigualdad de Cauchy-Schwarz

$$[\text{Cov}(Z, V)]^2 \leq \sigma_Z^2 \sigma_V^2 \quad \implies \quad \sigma_Z^2 \geq \frac{[\text{Cov}(Z, V)]^2}{\sigma_V^2}.$$

Sean  $Z = W(\mathbf{X})$  y  $V = \frac{\partial}{\partial \theta} \ell(\theta; \mathbf{x})$ , así que las expresiones anteriores quedan:

$$\mathbb{V}_\theta(W(\mathbf{X})) \geq \frac{\left[ \text{Cov}_\theta \left( W(\mathbf{X}), \frac{\partial}{\partial \theta} \ell(\theta; \mathbf{x}) \right) \right]^2}{\mathbb{V}_\theta \left[ \frac{\partial}{\partial \theta} \ell(\theta; \mathbf{x}) \right]}. \quad (5.11)$$

Por la expresión de la covarianza dada en (1.44) se tiene que:

$$\begin{aligned} \text{Cov}_\theta \left[ W(\mathbf{X}), \frac{\partial}{\partial \theta} \ell(\theta; \mathbf{x}) \right] &= \underbrace{\mathbb{E}_\theta \left( W(\mathbf{X}) \left[ \frac{\partial}{\partial \theta} \ell(\theta; \mathbf{x}) \right] \right)}_{(b)} \\ &\quad - \underbrace{\mathbb{E}_\theta [W(\mathbf{X})] \mathbb{E}_\theta \left[ \frac{\partial}{\partial \theta} \ell(\theta; \mathbf{x}) \right]}_{(a)}. \end{aligned} \quad (5.12)$$

Pero,

$$\begin{aligned} (a) \mathbb{E} \left( \frac{\partial \ell(\theta; \mathbf{x})}{\partial \theta} \right) &= \mathbb{E} \left( \frac{\partial}{\partial \theta} \ln f(\mathbf{x}; \theta) \right) = \mathbb{E} \left( \frac{1}{f(\mathbf{x}; \theta)} \frac{\partial f(\mathbf{x}; \theta)}{\partial \theta} \right) \\ &= \int \cdots \int_{\mathbb{R}^n} \left[ \frac{1}{f(\mathbf{x}; \theta)} \frac{\partial}{\partial \theta} f(\mathbf{x}; \theta) \right] f(\mathbf{x}; \theta) d\mathbf{x} \\ &= \frac{\partial}{\partial \theta} \left[ \int \cdots \int_{\mathbb{R}^n} f(\mathbf{x}; \theta) dx_1 \dots dx_n \right] = \frac{\partial}{\partial \theta} (1) = 0 \\ \mathbb{E} \left( \frac{\partial \ell(\theta; \mathbf{x})}{\partial \theta} \right) &= 0, \end{aligned} \quad (5.13)$$

por otra parte,

$$\begin{aligned} \frac{d}{d\theta} \mathbb{E}_\theta [W(\mathbf{X})] &= \int \cdots \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} (W(\mathbf{X}) f(\mathbf{x}; \theta)) d\mathbf{x} \\ &= \int \cdots \int_{\mathbb{R}^n} \left[ W(\mathbf{X}) \frac{\partial}{\partial \theta} (f(\mathbf{x}; \theta)) \right] d\mathbf{x} \\ &= \int \cdots \int_{\mathbb{R}^n} \left[ W(\mathbf{X}) \frac{f(\mathbf{x}; \theta)}{f(\mathbf{x}; \theta)} \frac{\partial}{\partial \theta} (f(\mathbf{x}; \theta)) \right] d\mathbf{x} \\ &= \int \cdots \int_{\mathbb{R}^n} \frac{\frac{\partial}{\partial \theta} f(\mathbf{x}; \theta)}{f(\mathbf{x}; \theta)} W(\mathbf{X}) f(\mathbf{x}; \theta) d\mathbf{x} \end{aligned}$$

$$= \int \cdots \int_{\mathbb{R}^n} \left[ \frac{\partial}{\partial \theta} \ln f(x; \theta) \right] W(\mathbf{X}) f(x; \theta) dx.$$

Luego,

$$\frac{d}{d\theta} \mathbb{E}_\theta [W(\mathbf{X})] = \mathbb{E}_\theta \left[ \frac{\partial}{\partial \theta} \ln f(x; \theta) W(\mathbf{X}) \right] = \mathbb{E}_\theta \left[ \underbrace{\left( \frac{\partial}{\partial \theta} \ell(\theta, x) \right) W(\mathbf{X})}_{(b)} \right], \quad (5.14)$$

al reemplazar (5.13) y (5.14) en (5.12), se tiene que:

$$\text{Cov}_\theta \left( W(\mathbf{X}), \frac{\partial}{\partial \theta} \ell(\theta, x) \right) = \frac{d}{d\theta} \mathbb{E}_\theta (W(\mathbf{X})). \quad (5.15)$$

De otro lado,

$$\begin{aligned} \mathbb{V}_\theta \left( \frac{\partial}{\partial \theta} \ell(\theta, x) \right) &= \mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \ell(\theta, x) \right)^2 \right] - \left[ \mathbb{E}_\theta \left( \frac{\partial}{\partial \theta} \ell(\theta, x) \right) \right]^2 \\ &= \mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \ell(\theta, x) \right)^2 \right]. \end{aligned} \quad (5.16)$$

Reemplazando (5.15) y (5.16) en (5.11) se tiene que:

$$\mathbb{V}_\theta (W(\mathbf{X})) \geq \frac{\left\{ \frac{d}{d\theta} \mathbb{E}_\theta [W(\mathbf{X})] \right\}^2}{\mathbb{E}_\theta \left\{ \left[ \frac{\partial}{\partial \theta} \ell(\theta, x) \right]^2 \right\}} = \frac{\left\{ \frac{d}{d\theta} \mathbb{E}_\theta [W(\mathbf{X})] \right\}^2}{I_n(\theta)}. \quad \square$$

**Nota 5.9.** La expresión  $\frac{\left\{ \frac{\partial}{\partial \theta} \mathbb{E}_\theta [W(\mathbf{X})] \right\}^2}{\mathbb{E}_\theta \left\{ \left[ \frac{\partial}{\partial \theta} \ell(\theta, x) \right]^2 \right\}}$  se conoce como la *cota de Cramér-Rao*.

**Nota 5.10.** Si las condiciones de regularidad se satisfacen, ver apéndice B.4.1, y  $W(\mathbf{X})$  es un estimador insesgado de  $\theta$ , entonces, por la expresión (B.10) se tiene que:

$$\begin{aligned} \mathbb{V}_\theta (W(\mathbf{X})) &\geq \frac{\left\{ \frac{\partial}{\partial \theta} \mathbb{E}_\theta [W(\mathbf{X})] \right\}^2}{\mathbb{E}_\theta \left\{ \left[ \frac{\partial}{\partial \theta} \ell(\theta, x) \right]^2 \right\}} = \frac{\left\{ \frac{\partial}{\partial \theta} \mathbb{E}_\theta [W(\mathbf{X})] \right\}^2}{I_n(\theta)} = \frac{\left\{ \frac{\partial}{\partial \theta} \right\}^2}{I_n(\theta)} \\ &\geq \frac{1}{I_n(\theta)}. \end{aligned}$$

**Nota 5.11.** Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una población con *fdp*  $f(x; \theta)$ ,  $\theta \in \Theta$ . Considerando que las condiciones de regularidad se satisfacen,

$$\begin{aligned} \mathbb{V}_\theta(W(\mathbf{X})) &\geq \frac{\left\{ \frac{\partial}{\partial \theta} \mathbb{E}_\theta [W(\mathbf{X})] \right\}^2}{\mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \ell(\theta, \mathbf{x}) \right)^2 \right]} = \frac{\left\{ \frac{\partial}{\partial \theta} \mathbb{E}_\theta [W(\mathbf{X})] \right\}^2}{\mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \ln \left\{ \prod_{i=1}^n f(x_i; \theta) \right\} \right)^2 \right]} \\ &= \frac{\left\{ \frac{\partial}{\partial \theta} \mathbb{E}_\theta [W(\mathbf{X})] \right\}^2}{\mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \sum_{i=1}^n \ln [f(x_i; \theta)] \right)^2 \right]} = \frac{\left\{ \frac{\partial}{\partial \theta} \mathbb{E}_\theta [W(\mathbf{X})] \right\}^2}{\mathbb{E}_\theta \left[ \left( \sum_{i=1}^n \frac{\partial}{\partial \theta} \ell(\theta; x_i) \right)^2 \right]}, \end{aligned} \tag{5.17}$$

reescribiendo el denominador de (5.17):

$$\begin{aligned} \mathbb{E}_\theta \left[ \left( \sum_{i=1}^n \frac{\partial}{\partial \theta} \ell(\theta; x_i) \right)^2 \right] &= \mathbb{E}_\theta \left[ \sum_{i=1}^n \left( \frac{\partial}{\partial \theta} \ln [f(x_i; \theta)] \right)^2 \right. \\ &\quad \left. + \sum_{i \neq j}^n \sum_{i \neq j}^n \left( \frac{\partial}{\partial \theta} \ln [f(x_i; \theta)] \right) \left( \frac{\partial}{\partial \theta} \ln [f(x_j; \theta)] \right) \right] \\ &\stackrel{\text{Indep.}}{=} \sum_{i=1}^n \mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \ln [f(x_i; \theta)] \right)^2 \right] + \\ &\quad \sum_{i \neq j}^n \sum_{i \neq j}^n \underbrace{\mathbb{E}_\theta \left[ \frac{\partial}{\partial \theta} \ln [f(x_i; \theta)] \right]}_0 \underbrace{\mathbb{E}_\theta \left[ \frac{\partial}{\partial \theta} \ln [f(x_j; \theta)] \right]}_0 \\ &= \sum_{i=1}^n \mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \ln [f(x_i; \theta)] \right)^2 \right] \\ &\stackrel{\text{Ident.}}{=} n \mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \ln [f(x_1; \theta)] \right)^2 \right] = n \mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \ell(\theta; x_1) \right)^2 \right]. \end{aligned}$$

Luego,

$$\mathbb{E}_\theta \left[ \left( \sum_{i=1}^n \frac{\partial}{\partial \theta} \ell(\theta; x_i) \right)^2 \right] = n I_1(\theta), \tag{5.18}$$

reemplazando (5.18) en (5.17) se tiene que:

$$\mathbb{V}_\theta(W(\mathbf{X})) \geq \frac{\left[\frac{\partial}{\partial\theta}\mathbb{E}_\theta(W(\mathbf{X}))\right]^2}{nI_1(\theta)}.$$

Así, si las condiciones de regularidad se cumplen, la cota de Cramér-Rao es dada por:

$$\frac{\left[\frac{\partial}{\partial\theta}\mathbb{E}_\theta(W(\mathbf{X}))\right]^2}{nI_1(\theta)}.$$

**Ejemplo 5.15.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid Bernoulli de parámetro  $\theta$ ,  $\theta \in (0, 1)$ , encuentre la cota de Cramér-Rao.

*Solución.* En este caso:

$$p_{X_1}(x; \theta) = \theta^x(1 - \theta)^{1-x}$$

con

$$\begin{aligned} \mathbb{E}(X_1) &= \theta, & \mathbb{V}(X_1) &= \mathbb{E}(X_1^2) - (\mathbb{E}(X_1))^2 = \theta(1 - \theta) \\ \mathbb{E}(X_1^2) &= \theta(1 - \theta) + \theta^2 = \theta. \end{aligned}$$

La log-verosimilitud es dada por:

$$\begin{aligned} \ln [p(x_1; \theta)] &= x_1 \ln(\theta) + (1 - x_1) \ln(1 - \theta) \\ \frac{\partial}{\partial\theta} (\ln [p(x_1; \theta)]) &= \frac{x_1}{\theta} - \frac{1 - x_1}{1 - \theta} \\ \left\{ \frac{\partial}{\partial\theta} (\ln [p(x_1; \theta)]) \right\}^2 &= \frac{x_1^2}{\theta^2} + \frac{(1 - x_1)^2}{(1 - \theta)^2} - \frac{2x_1(1 - x_1)}{\theta(1 - \theta)}, \end{aligned}$$

luego, el valor esperado queda:

$$\begin{aligned} \underbrace{\mathbb{E}_\theta \left[ \left\{ \frac{\partial}{\partial\theta} (\ln [p(x_1; \theta)]) \right\}^2 \right]}_{I_1(\theta)} &= \mathbb{E}_\theta \left( \frac{X_1^2}{\theta^2} \right) + \mathbb{E}_\theta \left[ \frac{(1 - X_1)^2}{(1 - \theta)^2} \right] - \mathbb{E}_\theta \left[ \frac{2X_1(1 - X_1)}{\theta(1 - \theta)} \right] \\ &= \frac{1}{\theta^2} \mathbb{E}_\theta (X_1^2) + \frac{1}{(1 - \theta)^2} \left[ \mathbb{E}_\theta (1 - 2X_1 + X_1^2) \right] \\ &\quad - \frac{2}{\theta(1 - \theta)} \mathbb{E}_\theta (X_1 - X_1^2) \\ &= \frac{\theta}{\theta^2} + \frac{1 - 2\theta + \theta}{(1 - \theta)^2} - \frac{2(\theta - \theta)}{\theta(1 - \theta)} \end{aligned}$$

$$= \frac{1}{\theta} + \frac{1-\theta}{(1-\theta)^2} = \frac{(1-\theta) + \theta}{\theta(1-\theta)} = \frac{1}{\theta(1-\theta)}.$$

Entonces,

$$\mathbb{V}_\theta[W(\mathbf{X})] \geq \frac{\left[\frac{\partial}{\partial\theta}\mathbb{E}_\theta(W(\mathbf{X}))\right]^2}{\frac{n}{\theta(1-\theta)}} = \frac{\theta(1-\theta)}{n} \left[\frac{\partial}{\partial\theta}\mathbb{E}_\theta(W(\mathbf{X}))\right]^2.$$

Sea  $W(\mathbf{X})$  el **EMV** de  $\theta$ , es decir,  $W(\mathbf{X}) = \bar{X}$ , ver ejemplo 4.8:

$$\begin{aligned}\mathbb{E}_\theta(W(\mathbf{X})) &= \mathbb{E}_\theta(\bar{X}) = \frac{1}{n}\mathbb{E}_\theta\left(\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\theta(X_i) = \frac{1}{n}\sum_{i=1}^n \theta \\ &= \theta \quad \text{por tanto } W(\mathbf{X}) \text{ es un estimador insesgado.}\end{aligned}$$

$$\begin{aligned}\mathbb{V}_\theta[W(\mathbf{X})] &= \mathbb{V}_\theta[\bar{X}] = \frac{1}{n^2}\mathbb{V}_\theta\left(\sum_{i=1}^n X_i\right) \underbrace{=}_{\text{Indep.}} \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}_\theta(X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^n \theta(1-\theta) = \frac{\theta(1-\theta)}{n}.\end{aligned}$$

Como la cota de Cramér-Rao

$$\begin{aligned}\frac{\left[\frac{\partial}{\partial\theta}\mathbb{E}_\theta(W(\mathbf{X}))\right]^2}{nI_1(\theta)} &= \frac{1}{\frac{n}{\theta(1-\theta)}} \left[\frac{\partial}{\partial\theta}\mathbb{E}_\theta(W(\mathbf{X}))\right]^2 = \frac{\theta(1-\theta)}{n} \left[\frac{\partial}{\partial\theta}\theta\right]^2 \\ &= \frac{\theta(1-\theta)}{n}\end{aligned}$$

coincide con la  $\mathbb{V}_\theta[\bar{X}]$ , es decir, la  $\mathbb{V}_\theta[\bar{X}]$  alcanza la cota de Cramér-Rao, entonces  $\bar{X}$  es un **UMVUE** para  $\theta$ .  $\checkmark$

**Ejemplo 5.16.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias *iid* Poisson de parámetro  $\theta$ ,  $\theta > 0$ , encuentre la cota de Cramér-Rao.

*Solución.* La **fmp** está dada por la expresión (1.13):

$$p_{X_1}(x_1; \theta) = \frac{e^{-\theta}\theta^{x_1}}{x_1!} \quad x_1 = 0, 1, 2, 3, \dots,$$

con

$$\begin{aligned}\mathbb{E}(X_1) &= \theta, & \mathbb{V}(X_1) &= \theta \\ \mathbb{E}(X_1^2) &= \theta + \theta^2 = \theta(1 + \theta).\end{aligned}$$

La log-verosimilitud es dada por:

$$\begin{aligned}\ln [p(x_1; \theta)] &= -\theta + x_1 \ln(\theta) - \ln(x_1!) \\ \frac{\partial}{\partial \theta} (\ln [p(x_1; \theta)]) &= -1 + \frac{x_1}{\theta} \\ \left\{ \frac{\partial}{\partial \theta} (\ln [p(x_1; \theta)]) \right\}^2 &= \frac{x_1^2}{\theta^2} - \frac{2x_1}{\theta} + 1,\end{aligned}$$

luego, el valor esperado queda:

$$\begin{aligned}\underbrace{\mathbb{E}_\theta \left[ \left\{ \frac{\partial}{\partial \theta} (\ln [p(x_1; \theta)]) \right\}^2 \right]}_{I_1(\theta)} &= \mathbb{E}_\theta \left[ \frac{X_1^2}{\theta^2} \right] - \mathbb{E}_\theta \left[ \frac{2X_1}{\theta} \right] + 1 \\ &= \frac{1}{\theta^2} \mathbb{E}_\theta [X_1^2] - \frac{2}{\theta} \mathbb{E}_\theta (X_1) + 1 \\ &= \frac{\theta(1+\theta)}{\theta^2} - \frac{2\theta}{\theta} + 1 = \frac{1+\theta}{\theta} - 1 = \frac{1}{\theta}.\end{aligned}$$

Entonces,

$$\mathbb{V}_\theta [W(\mathbf{X})] \geq \frac{\left[ \frac{\partial}{\partial \theta} \mathbb{E}_\theta (W(\mathbf{X})) \right]^2}{nI_1(\theta)} = \frac{\left[ \frac{\partial}{\partial \theta} \mathbb{E}_\theta (W(\mathbf{X})) \right]^2}{n \frac{1}{\theta}}.$$

Sea  $W(\mathbf{X})$  el **EMV** de  $\theta$ , dado en (4.14), es decir,  $W(\mathbf{X}) = \bar{X}$ , ver ejemplo 4.11:

$$\mathbb{E}_\theta (W(\mathbf{X})) = \mathbb{E}_\theta (\bar{X}) = \frac{1}{n} \mathbb{E}_\theta \left( \sum_{i=1}^n X_i \right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\theta (X_i) = \frac{1}{n} \sum_{i=1}^n \theta = \theta,$$

por tanto,  $W(\mathbf{X})$  es un estimador insesgado

$$\begin{aligned}\mathbb{V}_\theta [W(\mathbf{X})] &= \mathbb{V}_\theta [\bar{X}] = \frac{1}{n^2} \mathbb{V}_\theta \left( \sum_{i=1}^n X_i \right) \underbrace{=}_{\text{Indep.}} \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}_\theta (X_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n \theta = \frac{\theta}{n}.\end{aligned}$$

Como la cota de Cramér-Rao

$$\frac{\left[ \frac{\partial}{\partial \theta} \mathbb{E}_\theta (W(\mathbf{X})) \right]^2}{nI_1(\theta)} = \frac{1}{n} \left[ \frac{\partial}{\partial \theta} \mathbb{E}_\theta (W(\mathbf{X})) \right]^2 = \frac{\theta}{n} \left[ \frac{\partial}{\partial \theta} \theta \right]^2 = \frac{\theta}{n}$$

coincide con la  $\mathbb{V}_\theta[\bar{X}]$ , es decir, la  $\mathbb{V}_\theta[\bar{X}]$  alcanza la cota de Cramér-Rao, entonces  $\bar{X}$  es un **UMVUE** para  $\theta$ .  $\checkmark$

**Ejemplo 5.17.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias *iid* geométrica de parámetro  $\frac{1}{\theta}$ ,  $\theta > 1$ , encuentre la cota de Cramér-Rao.

*Solución.* La **fmp** está dada por la expresión (1.14):

$$p(x_1; \theta) = \frac{1}{\theta} \left(1 - \frac{1}{\theta}\right)^{x_1} \quad x_1 = 0, 1, 2, 3, \dots,$$

usando los correspondientes valores de  $\mathbb{E}(X)$  y  $\mathbb{V}(X)$  dados en la tabla 1.12:

$$\begin{aligned} \mathbb{E}(X_1) &= \frac{1 - \frac{1}{\theta}}{\frac{1}{\theta}} = \theta - 1, & \mathbb{V}(X_1) &= \frac{1 - \frac{1}{\theta}}{\left(\frac{1}{\theta}\right)^2} = \frac{\theta - 1}{\frac{1}{\theta^2}} = \theta(\theta - 1) \\ \mathbb{E}(X_1^2) &= \theta(\theta - 1) + (\theta - 1)^2 = (\theta - 1)(\theta + \theta - 1) = (\theta - 1)(2\theta - 1). \end{aligned}$$

La log-verosimilitud es dada por:

$$\begin{aligned} \ln [p(x_1; \theta)] &= -\ln(\theta) + x_1 \ln \left[1 - \frac{1}{\theta}\right] \\ \frac{\partial}{\partial \theta} (\ln [p(x_1; \theta)]) &= \frac{-1}{\theta} + \frac{x_1}{1 - \frac{1}{\theta}} \left[\frac{1}{\theta^2}\right] = \frac{-1}{\theta} + \frac{x_1}{\theta(\theta - 1)} \\ \left\{ \frac{\partial}{\partial \theta} (\ln [p(x_1; \theta)]) \right\}^2 &= \frac{1}{\theta^2} - \frac{2x_1}{\theta^2(\theta - 1)} + \frac{x_1^2}{\theta^2(\theta - 1)^2}, \end{aligned}$$

luego, el valor esperado queda:

$$\begin{aligned} \mathbb{E}_\theta \left[ \underbrace{\left\{ \frac{\partial}{\partial \theta} (\ln [p(x_1; \theta)]) \right\}^2}_{I_1(\theta)} \right] &= \frac{1}{\theta^2} - \mathbb{E} \left[ \frac{2X_1}{\theta^2(\theta - 1)} \right] + \mathbb{E} \left[ \frac{X_1^2}{\theta^2(\theta - 1)^2} \right] \\ &= \frac{1}{\theta^2} - \frac{2(\theta - 1)}{\theta^2(\theta - 1)} + \frac{(\theta - 1)(2\theta - 1)}{\theta^2(\theta - 1)^2} \\ &= \frac{-1}{\theta^2} + \frac{2\theta - 1}{\theta^2(\theta - 1)} = \frac{-\theta + 1 + 2\theta - 1}{\theta^2(\theta - 1)} = \frac{1}{\theta(\theta - 1)}. \end{aligned}$$

Entonces,

$$\mathbb{V}_\theta [W(\mathbf{X})] \geq \frac{\left[ \frac{\partial}{\partial \theta} \mathbb{E}_\theta (W(\mathbf{X})) \right]^2}{n I_1(\theta)} = \frac{\left[ \frac{\partial}{\partial \theta} \mathbb{E}_\theta (W(\mathbf{X})) \right]^2}{n \frac{1}{\theta(1-\theta)}}.$$

Sea  $W(\mathbf{X})$  el **EMV** de  $\theta$ , es decir,

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \frac{1}{\theta} \left(1 - \frac{1}{\theta}\right)^{x_i} = \frac{1}{\theta^n} \left(1 - \frac{1}{\theta}\right)^{\sum_{i=1}^n x_i}$$

$$\ell(\theta; \mathbf{x}) = -n \ln(\theta) + \left(\sum_{i=1}^n x_i\right) \ln \left[1 - \frac{1}{\theta}\right] = -n \ln(\theta) + n\bar{x} \ln \left[1 - \frac{1}{\theta}\right]$$

$$\frac{\partial \ell(\theta; \mathbf{x})}{\partial \theta} = \frac{-n}{\theta} + \frac{n\bar{x}}{1 - \frac{1}{\theta}} \left[\frac{1}{\theta^2}\right] = \frac{-n}{\theta} + \frac{n\bar{x}}{\theta(\theta - 1)},$$

igualando a cero:

$$\frac{-n}{\hat{\theta}} + \frac{n\bar{x}}{\hat{\theta}(\hat{\theta} - 1)} = 0 \quad \Leftrightarrow \quad \frac{-(\hat{\theta} - 1)}{\hat{\theta}(\hat{\theta} - 1)} + \frac{\bar{x}}{\hat{\theta}(\hat{\theta} - 1)} = 0 \quad \Leftrightarrow$$

$$-\hat{\theta} + 1 + \bar{x} = 0 \quad \Leftrightarrow \quad \hat{\theta} = 1 + \bar{x},$$

luego,  $W(\mathbf{X}) = 1 + \bar{X}$  y el valor esperado queda:

$$\begin{aligned} \mathbb{E}_\theta(W(\mathbf{X})) &= \mathbb{E}_\theta(1 + \bar{X}) = 1 + \frac{1}{n} \mathbb{E}_\theta\left(\sum_{i=1}^n X_i\right) = 1 + \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\theta(X_i) \\ &= 1 + \frac{1}{n} \sum_{i=1}^n (\theta - 1) = 1 + \theta - 1 = \theta, \end{aligned}$$

por tanto,  $W(\mathbf{X})$  es un estimador insesgado

$$\begin{aligned} \mathbb{V}_\theta[W(\mathbf{X})] &= \mathbb{V}_\theta[1 + \bar{X}] = \mathbb{V}_\theta[\bar{X}] = \frac{1}{n^2} \mathbb{V}_\theta\left(\sum_{i=1}^n X_i\right) \\ &= \underbrace{\frac{1}{n^2} \sum_{i=1}^n \mathbb{V}_\theta(X_i)}_{\text{Indep.}} = \frac{1}{n^2} \sum_{i=1}^n \theta(1 - \theta) = \frac{\theta(1 - \theta)}{n}. \end{aligned}$$

Como la cota de Cramér-Rao

$$\begin{aligned} \frac{\left[\frac{\partial}{\partial \theta} \mathbb{E}_\theta(W(\mathbf{X}))\right]^2}{nI_1(\theta)} &= \frac{1}{\frac{n}{\theta(1-\theta)}} \left[\frac{\partial}{\partial \theta} \mathbb{E}_\theta(W(\mathbf{X}))\right]^2 = \frac{\theta(1-\theta)}{n} \left[\frac{\partial}{\partial \theta} \theta\right]^2 \\ &= \frac{\theta(1-\theta)}{n} \end{aligned}$$

coincide con la  $\mathbb{V}_\theta[\bar{X} + 1]$ , es decir, la  $\mathbb{V}_\theta[\bar{X} + 1]$  alcanza la cota de Cramér-Rao, entonces  $\bar{X} + 1$  es un **UMVUE** para  $\theta$ .  $\checkmark$

**Ejemplo 5.18.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias *iid*  $N(\theta, 1)$ , encuentre la cota de Cramér-Rao.

*Solución.* En este caso,

$$f(x_1; \theta) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(x_1 - \theta)^2 \right\}$$

con

$$\begin{aligned} \mathbb{E}(X_1) &= \theta, & \mathbb{V}(X_1) &= 1 \\ \mathbb{E}(X_1^2) &= 1 + \theta^2. \end{aligned}$$

La log-verosimilitud es dada por:

$$\begin{aligned} \ln [f(x_1; \theta)] &= -\ln(\sqrt{2\pi}) - \frac{1}{2}(x_1 - \theta)^2 \\ \frac{\partial}{\partial \theta} (\ln [f(x_1; \theta)]) &= -\frac{2}{2}(x_1 - \theta)(-1) = x_1 - \theta \\ \left\{ \frac{\partial}{\partial \theta} (\ln [f(x_1; \theta)]) \right\}^2 &= (x_1 - \theta)^2, \end{aligned}$$

luego, el valor esperado queda:

$$\mathbb{E}_\theta \left[ \left\{ \frac{\partial}{\partial \theta} (\ln [f(x_1; \theta)]) \right\}^2 \right] = \mathbb{E} \left[ (x_1 - \theta)^2 \right] = \mathbb{V}(X_1) = 1 = I_1(\theta).$$

Entonces,

$$\mathbb{V}_\theta [W(\mathbf{X})] \geq \frac{\left[ \frac{\partial}{\partial \theta} \mathbb{E}_\theta (W(\mathbf{X})) \right]^2}{n I_1(\theta)} = \frac{\left[ \frac{\partial}{\partial \theta} \mathbb{E}_\theta (W(\mathbf{X})) \right]^2}{n}.$$

Sea  $W(\mathbf{X})$  el EMV de  $\theta$ , es decir,  $W(\mathbf{X}) = \bar{X}$ , ver ejemplo 4.7:

$$\begin{aligned} \mathbb{E}_\theta (W(\mathbf{X})) &= \mathbb{E}_\theta (\bar{X}) = \frac{1}{n} \mathbb{E}_\theta \left( \sum_{i=1}^n X_i \right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\theta (X_i) = \frac{1}{n} \sum_{i=1}^n \theta \\ &= \theta \quad \text{por tanto } W(\mathbf{X}) \text{ es un estimador insesgado,} \end{aligned}$$

$$\begin{aligned} \mathbb{V}_\theta [W(\mathbf{X})] &= \mathbb{V}_\theta [\bar{X}] = \frac{1}{n^2} \mathbb{V}_\theta \left( \sum_{i=1}^n X_i \right) \\ &= \underbrace{\frac{1}{n^2} \sum_{i=1}^n \mathbb{V}_\theta (X_i)}_{\text{Indep.}} = \frac{1}{n^2} \sum_{i=1}^n 1 = \frac{1}{n}. \end{aligned}$$

Como la cota de Cramér-Rao

$$\frac{\left[\frac{\partial}{\partial\theta}\mathbb{E}_\theta(W(\mathbf{X}))\right]^2}{nI_1(\theta)} = \frac{1}{n \times 1} \left[\frac{\partial}{\partial\theta}\mathbb{E}_\theta(W(\mathbf{X}))\right]^2 = \frac{1}{n} \left[\frac{\partial}{\partial\theta}\theta\right]^2 = \frac{1}{n}$$

coincide con la  $\mathbb{V}_\theta[\bar{X}]$ , es decir, la  $\mathbb{V}_\theta[\bar{X}]$  alcanza la cota de Cramér-Rao, entonces  $\bar{X}$  es un **UMVUE** para  $\theta$ .  $\checkmark$

**Ejemplo 5.19.** De acuerdo con [65, p. 318], si  $X_1, X_2, \dots, X_n$  son variables aleatorias *iid* tal que  $X_1 \sim \text{Exp}(\theta)$ , encontrar la cota de Cramér-Rao.

*Solución.* En este caso,  $f_X(x) = \theta e^{-\theta x} I_{(0, \infty)}(x)$ . Considerando que las condiciones de regularidad se satisfacen, usando (5.10) se tiene que:

$$\mathbb{V}_\theta [W(\mathbf{X})] \geq \frac{\left\{\frac{\partial}{\partial\theta}\mathbb{E}_\theta [W(\mathbf{X})]\right\}^2}{\mathbb{E}_\theta \left[\left(\frac{\partial}{\partial\theta}\ell(\theta; \mathbf{x})\right)^2\right]},$$

donde la log-verosimilitud es dada por:

$$\begin{aligned} \ln(f_{X_1}(x_1)) &= \ln(\theta e^{-\theta x_1}) = \ln(\theta) - \theta x_1 \\ \frac{\partial}{\partial\theta} [\ln(f_{X_1}(x_1))] &= \frac{1}{\theta} - x_1 \\ \left\{\frac{\partial}{\partial\theta} [\ln(f_{X_1}(x_1))]\right\}^2 &= \left\{\frac{1}{\theta} - x_1\right\}^2, \end{aligned}$$

luego, el valor esperado queda:

$$\mathbb{E}_\theta \left[\left\{\frac{\partial}{\partial\theta} [\ln(f_{X_1}(x_1))]\right\}^2\right] = \mathbb{E}_\theta \left[\left\{\frac{1}{\theta} - X_1\right\}^2\right] = \mathbb{V}(X_1) = \frac{1}{\theta^2},$$

por tanto,  $I_n(\theta) = n\mathbb{V}(X_1) = \frac{n}{\theta^2}$  y

$$\mathbb{V}_\theta [W(\mathbf{X})] \geq \frac{\left\{\frac{\partial}{\partial\theta}\mathbb{E}_\theta [W(\mathbf{X})]\right\}^2}{\frac{n}{\theta^2}}.$$

Si  $W(\mathbf{X})$  es un estimador insesgado de  $\theta$ , se tiene que:

$$\mathbb{V}_\theta [W(\mathbf{X})] \geq \frac{1}{\frac{n}{\theta^2}} = \frac{\theta^2}{n},$$

luego la cota de Cramér-Rao para la varianza de un estimador insesgado de  $\theta$  es dada por  $\frac{\theta^2}{n}$ . ☑

**Ejemplo 5.20.** (Teorema de Cramér-Rao no es aplicable). Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid con fdp uniforme sobre el intervalo  $(0, \theta)$ , muestre que  $p(x; \theta)$  incumple la condición de regularidad C6.

*Solución.* Puesto que la fdp es dada por (3.11):

$$\frac{\partial}{\partial \theta} \int_0^\theta W(x)p(x; \theta) dx = \frac{\partial}{\partial \theta} \int_0^\theta W(x) \frac{1}{\theta} dx,$$

aplicando la regla de Leibniz (ver nota B.12):

$$\begin{aligned} \frac{\partial}{\partial \theta} \int_0^\theta W(x)p(x; \theta) dx &= W(\theta) \frac{1}{\theta} \frac{\partial}{\partial \theta} \theta - W(0) \frac{1}{\theta} \frac{\partial}{\partial \theta} 0 + \int_0^\theta \frac{\partial}{\partial \theta} \left( W(x) \frac{1}{\theta} \right) dx \\ &= \frac{W(\theta)}{\theta} + \int_0^\theta \frac{\partial}{\partial \theta} \frac{W(x)}{\theta} dx. \end{aligned}$$

Entonces,

$$\frac{\partial}{\partial \theta} \int_0^\theta W(x)p(x; \theta) dx \neq \int_0^\theta \frac{\partial}{\partial \theta} [W(x)p(x; \theta)] dx.$$

Salvo que  $\frac{W(\theta)}{\theta} = 0, \forall \theta$ . Luego no aplica el teorema de Cramér-Rao. ☑

**Nota 5.12.** En general, si el soporte de la fdp depende del parámetro, el teorema de Cramér-Rao no es aplicable.

## 5.7. Consistencia

Esta propiedad impone la convergencia en probabilidad del estimador al parámetro a estimar cuando el tamaño de la muestra es muy grande.

**DEFINICIÓN 5.8.** Según [82, p. 201] y [52, p. 51], una sucesión  $\{\hat{\theta}_n\}_{n \geq 1}$  de estimadores (del parámetro  $\theta$ ) es **(débilmente) consistente** si

$$\hat{\theta}_n - \theta \xrightarrow{P} 0.$$

De otro lado, una sucesión  $\{\hat{\theta}_n\}_{n \geq 1}$  de estimadores (del parámetro  $\theta$ ) es **fuertemente consistente** si

$$\hat{\theta}_n - \theta \xrightarrow{c.s.} 0.$$

**Nota 5.13.** Según [82, p. 201], una sucesión de estimadores  $\{\theta_n\}_{n \geq 1}$  del parámetro  $\theta$  es **asintóticamente insesgada** si:

$$\mathbb{E}(\hat{\theta}_n) \xrightarrow{n \rightarrow \infty} \theta \qquad \lim_{n \rightarrow \infty} \mathbb{E}(\hat{\theta}_n) = \theta.$$

**Proposición 5.1. Criterio para consistencia.** De acuerdo con [20, p. 469], si  $\{\hat{\theta}_n\}_{n \geq 1}$  es una sucesión de estimadores (del parámetro  $\theta$ ), tal que:

1.  $\lim_{n \rightarrow \infty} \mathbb{E}(\hat{\theta}_n) = \theta,$
2.  $\lim_{n \rightarrow \infty} \mathbb{V}(\hat{\theta}_n) = 0,$

entonces  $\hat{\theta}_n$  es consistente.

*Demostración.* Replicar el procedimiento de solución del ejemplo 3.23 haciendo los cambios convenientes. ☑

**Ejemplo 5.21.** Sea  $(X_1, Y_1)^\top, (X_2, Y_2)^\top, \dots, (X_n, Y_n)^\top$  una muestra aleatoria de la distribución del vector aleatorio  $(X_1, Y_1)^\top$  con distribución bivariada, con vector de medias y matriz de varianzas-covarianzas dadas por:

$$\boldsymbol{\mu} = \mathbb{E} \begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \text{y} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Sea  $r_n$  un estimador de  $\rho$  definido por:

$$r_n = \frac{1}{n} \sum_{i=1}^n X_i Y_i.$$

1. Verifique si  $r_n$  es consistente.
2. Suponiendo que la distribución del vector aleatorio  $(X_1, Y_1)^\top$  es normal bivariada, encuentre la distribución asintótica de  $r_n$  cuando es normalizado.

*Solución.* Del ejemplo 3.32 se tiene que  $r_n \xrightarrow{c.s.} \rho.$

1. Por la definición 5.8 se tiene que  $r_n$  es consistente.
2. Se tiene que  $\mathbb{E}(X_1 Y_1) = \text{Cov}(X_1, Y_1) + \mathbb{E}(X_1)\mathbb{E}(Y_1) = \rho$  y usando las expresiones (2.14) y (2.18)

$$\begin{aligned} \mathbb{V}(X_1 Y_1) &= \mathbb{E}(X_1^2 Y_1^2) - \mathbb{E}^2(X_1 Y_1) = \mathbb{E}[\mathbb{E}(X_1^2 Y_1^2 | X_1)] - \rho^2 \\ &= \mathbb{E}[X_1^2 \mathbb{E}(Y_1^2 | X_1)] - \rho^2. \end{aligned}$$

Por la expresión (2.32) se tiene que:  $Y_1 | X_1 \sim N(\rho X_1; 1 - \rho^2)$ , luego,

$$\mathbb{E}(Y_1 | X_1) = \rho X_1 \qquad \text{y}$$

$$\mathbb{E}(Y_1^2|X_1) = \mathbb{V}(Y_1|X_1) + \mathbb{E}^2(Y_1|X_1) = 1 - \rho^2 + \rho^2 X_1^2,$$

por tanto,

$$\begin{aligned} \mathbb{V}(X_1 Y_1) &= \mathbb{E}[X_1^2 \mathbb{E}(Y_1^2|X_1)] - \rho^2 \\ &= \mathbb{E}\left[X_1^2 \left(1 - \rho^2 + \rho^2 X_1^2\right)\right] - \rho^2 \\ &= \mathbb{E}[X_1^2] - \mathbb{E}[X_1^2 \rho^2] + \mathbb{E}[X_1^4 \rho^2] - \rho^2 \\ &= 1 - \rho^2 \times 1 + \underbrace{\rho^2 \mathbb{E}[X_1^4]}_3 - \rho^2 = 1 + \rho^2 < \infty, \end{aligned}$$

aplicando el teorema central del límite (clásico), teorema 3.22,

$$\begin{aligned} &\frac{\sum_{i=1}^n X_i Y_i - n\rho}{\sqrt{1 + \rho^2} \sqrt{n}} \xrightarrow{D} Z \\ &\frac{\frac{1}{n} \sum_{i=1}^n X_i Y_i - \rho}{\frac{\sqrt{1 + \rho^2} \sqrt{n}}{n}} \xrightarrow{D} Z \\ &\sqrt{1 + \rho^2} \times \sqrt{n} \frac{\left(\frac{1}{n} \sum_{i=1}^n X_i Y_i - \rho\right)}{\sqrt{1 + \rho^2}} \xrightarrow{D} \sqrt{1 + \rho^2} Z \\ &\qquad \qquad \qquad \sqrt{n}(r_n - \rho) \xrightarrow{D} \sqrt{1 + \rho^2} Z \end{aligned}$$

donde  $Z \sim N(0, 1)$ , luego  $\sqrt{1 + \rho^2} Z \sim N(0, 1 + \rho^2)$ , así

$$\sqrt{n}(r_n - \rho) \xrightarrow{D} N(0, 1 + \rho^2). \quad \checkmark$$

**Ejemplo 5.22.** Suponga que  $X_1, X_2, \dots, X_n$  son variables aleatorias iid, tal que  $X_1 \sim \exp(\theta)$ . Encuentre una forma de la mediana muestral que sea un estimador consistente de  $\theta$  y su distribución asintótica cuando es convenientemente normalizada.

*Solución.* Se procede a encontrar el estimador de la mediana muestral

$$\begin{aligned} F(\xi_{0,5} | \theta) = 0,5 &\Leftrightarrow 1 - e^{-\theta \xi_{0,5}} = 0,5 &\Leftrightarrow \\ e^{-\theta \xi_{0,5}} = 0,5 &\Leftrightarrow -\theta \xi_{0,5} = -\ln(2) &\Leftrightarrow \\ \theta = \frac{\ln(2)}{\xi_{0,5}} &\Leftrightarrow \hat{\theta} = \frac{\ln(2)}{\hat{\xi}_{0,5}} \end{aligned}$$

donde  $\hat{\xi}_{0,5}$  es la mediana muestral. Del teorema 3.25 se tiene que:

$$X_{n:k} \xrightarrow{c.s.} \xi_q \quad \text{luego} \quad \hat{\xi}_{0,5} \xrightarrow{c.s.} \xi_{0,5}.$$

Sea  $g$  la función dada por  $g(z) = \frac{\ln(2)}{z}$ , por el teorema 3.14 se sigue que:

$$g(\hat{\xi}_{0,5}) \xrightarrow{c.s.} g(\xi_{0,5}) \\ \frac{\ln(2)}{\hat{\xi}_{0,5}} \xrightarrow{c.s.} \frac{\ln(2)}{\xi_{0,5}} = \theta,$$

luego  $\frac{\ln(2)}{\hat{\xi}_{0,5}}$  es un estimador consistente para  $\theta$ . Para la distribución de  $\frac{\ln(2)}{\hat{\xi}_{0,5}}$  se usa el teorema 3.26,

$$\begin{aligned} \sqrt{n}(X_{n:k} - \xi_q) &\xrightarrow{D} N\left(0, \frac{q(1-q)}{f^2(\xi_q)}\right) && \text{con } q = \frac{1}{2} \\ \sqrt{n}(\hat{\xi}_{n:0,5} - \xi_{0,5}) &\xrightarrow{D} N\left(0, \frac{1}{4f^2(\xi_{0,5})}\right) \\ \sqrt{n}\left[\hat{\xi}_{n:0,5} - \frac{\ln(2)}{\theta}\right] &\xrightarrow{D} N\left(0, \frac{1}{[2\theta e^{-\theta\xi_{0,5}}]^2}\right) && \text{pero } \theta\xi_{0,5} = \ln(2) \\ \sqrt{n}\left[\hat{\xi}_{n:0,5} - \frac{\ln(2)}{\theta}\right] &\xrightarrow{D} N\left(0, \frac{1}{[2\theta e^{-\ln(2)}]^2}\right) \\ \sqrt{n}\left[\hat{\xi}_{n:0,5} - \frac{\ln(2)}{\theta}\right] &\xrightarrow{D} N\left(0, \frac{1}{\theta^2}\right). \end{aligned}$$

Como el interés está en la distribución de  $\hat{\theta}$ , usando el método delta, con  $g(z) = \frac{\ln(2)}{z}$ ,  $g'(z) = \frac{-\ln(2)}{z^2}$ , la expresión (3.21) queda:

$$\begin{aligned} \sqrt{n}[g(X_n) - g(\theta)] &\xrightarrow{D} N\left(0, \sigma^2[g'(\theta)]^2\right) \\ \sqrt{n}\left[g(\hat{\xi}_{n:0,5}) - g\left(\frac{\ln(2)}{\theta}\right)\right] &\xrightarrow{D} N\left(0, \frac{1}{\theta^2} \times \left[\frac{-\ln(2)}{\left(\frac{\ln(2)}{\theta}\right)^2}\right]^2\right) \\ \sqrt{n}\left[\frac{\ln(2)}{\hat{\xi}_{n:0,5}} - \frac{\ln(2)}{\theta}\right] &\xrightarrow{D} N\left(0, \frac{\theta^2}{[\ln(2)]^2}\right) \\ \sqrt{n}[\hat{\theta} - \theta] &\xrightarrow{D} N\left(0, \frac{\theta^2}{[\ln(2)]^2}\right). \quad \square \end{aligned}$$

**DEFINICIÓN 5.9.** Según [62, p. 123], un procedimiento o método estadístico se denomina **robusto** si su desempeño es imperturbable a ligeras discordancias del modelo original o de los supuestos considerados con la información acopiada. Particularmente, un estimador  $T(\mathbf{X})$  basado en una muestra aleatoria  $X_1, X_2, \dots, X_n$ , de una población con *fdp*  $f_X(x; \theta)$ , recibe la denominación de **estimador robusto**, si su desempeño permanece inalterado ante discrepancias con el modelo original.

**Nota 5.14.** En la definición 5.8 y en la nota 5.13 se presentaron los conceptos de **consistencia** y estimador **asintóticamente insesgado**.

## 5.8. Eficiencia

Un estimador  $\hat{\theta}_1$  es más eficiente o más preciso que otro estimador  $\hat{\theta}_2$ , si la varianza del primero es menor que la del segundo, es decir, si se verifica que  $\mathbb{V}(\hat{\theta}_1) \leq \mathbb{V}(\hat{\theta}_2)$ .

**Nota 5.15.** Según [82, p. 228], el concepto de **eficiencia relativa asintótica** (*ARE*, por sus siglas en inglés *asymptotic relative efficiency*) para el caso de estimadores que se distribuyen asintóticamente normal corresponde a la razón de las respectivas varianzas asintóticas, esto es, sean  $\{\hat{\theta}_n^{(1)}\}_{n \geq 1}$  y  $\{\hat{\theta}_n^{(2)}\}_{n \geq 1}$  sucesiones de estimadores tales que:

$$\sqrt{n} \left( \hat{\theta}_n^{(k)} - \theta \right) \xrightarrow{D} Z_k, \quad Z_k \sim N(0, \sigma_k^2), \quad k = 1, 2.$$

La **eficiencia relativa asintótica** de  $\hat{\theta}_n^{(2)}$  con respecto a  $\hat{\theta}_n^{(1)}$  está dada por:

$$ARE \left( \hat{\theta}_n^{(2)} | \hat{\theta}_n^{(1)} \right) = \frac{\sigma_1^2}{\sigma_2^2}.$$

Si  $ARE \left( \hat{\theta}_n^{(2)} | \hat{\theta}_n^{(1)} \right) < 1$ , entonces  $\hat{\theta}_n^{(1)}$  es más eficiente que  $\hat{\theta}_n^{(2)}$ .

**Ejemplo 5.23.** Suponga que  $X_1, X_2, \dots, X_n$  son variables aleatorias *iid*, tal que  $X_1 \sim \text{Exp}(\theta)$ . Encuentre la eficiencia relativa asintótica del estimador encontrado en el ejemplo 5.22 con relación al *EMV* de  $\theta$ .

**Solución.** Por (4.15) el *EMV* es dado por, ver ejemplo 4.14:

$$\hat{\theta}_{MV} = \frac{1}{\bar{X}}.$$

Del teorema del límite central clásico se tiene que:

$$\begin{aligned} \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} &\xrightarrow{D} N(0,1) \\ \frac{n\bar{X} - n\frac{1}{\theta}}{\frac{1}{\theta}\sqrt{n}} &\xrightarrow{D} N(0,1) \\ \sqrt{n}\left(\bar{X} - \frac{1}{\theta}\right) &\xrightarrow{D} N\left(0, \frac{1}{\theta^2}\right). \end{aligned}$$

Usando el método delta con  $g(z) = \frac{1}{z}$ ,  $z \neq 0$ ,  $g'(z) = -\frac{1}{z^2}$ , la expresión (3.21) queda:

$$\begin{aligned} \sqrt{n}\left[g(\bar{x}) - g\left(\frac{1}{\theta}\right)\right] &\xrightarrow{D} N\left(0, \frac{[g'(\frac{1}{\theta})]^2}{\theta^2}\right) \\ \sqrt{n}\left[\frac{1}{\bar{x}} - \frac{1}{\theta}\right] &\xrightarrow{D} N\left(0, \frac{1}{\theta^2}\left[\frac{1}{\theta^2}\right]^2\right) \\ \sqrt{n}[\hat{\theta}_{MV} - \theta] &\xrightarrow{D} N(0, \theta^2). \end{aligned}$$

Así, la eficiencia relativa asintótica del estimador  $\hat{\theta}$  encontrado en el ejemplo 5.22, con respecto a  $\hat{\theta}_{MV}$  es dada por:

$$ARE(\hat{\theta}|\hat{\theta}_{MV}) = \frac{\theta^2}{[\ln(2)]^2} = 0,4804$$

como  $ARE(\hat{\theta}|\hat{\theta}_{MV}) < 1$ ,  $\hat{\theta}_{MV}$  es más eficiente que  $\hat{\theta}$ , es decir, el EMV de  $\theta$  es dos veces más eficiente que el estimador  $\hat{\theta} = \frac{\ln(2)}{\hat{\xi}_{0,5}}$ .  $\square$

## 5.9. Invarianza para estimadores de máxima verosimilitud

Suponga que una fdp es indexada por un parámetro  $\theta$ , pero el interés está en encontrar un estimador para alguna función de  $\theta$ , digamos  $q(\theta)$ .

**Teorema 5.7. Principio de invarianza del EMV.** De acuerdo con [20, p. 320], si  $\hat{\theta}$  es el EMV de  $\theta$  y  $q(\theta)$  es una función cualquiera del parámetro, entonces el EMV de  $q(\theta)$  es  $q(\hat{\theta})$ .

*Demostración.* Supóngase primero que  $q(\theta)$  es una función inyectiva, entonces la función inversa  $q^{-1}(\theta)$  existe y está bien definida. Sea  $\gamma = q(\theta)$ , luego  $q^{-1}(\gamma) = \theta$  y la función de verosimilitud asociada a  $q(\theta)$  escrita en términos de  $\gamma$ , está dada por:

$$L^*(\gamma; \mathbf{x}) = \prod_{i=1}^n p(x_i; \underbrace{q^{-1}(\gamma)}_{\theta}) = L(\underbrace{q^{-1}(\gamma)}_{\theta}; \mathbf{x}) \quad y$$

$$\sup_{\gamma} L^*(\gamma; \mathbf{x}) = \sup_{\gamma} L(q^{-1}(\gamma); \mathbf{x}) = \sup_{\theta} L(\theta; \mathbf{x}).$$

Luego, el máximo de  $L^*(\gamma; \mathbf{x})$  es alcanzado en  $\gamma = q(\theta) = q(\hat{\theta})$  y entonces el **EMV** de  $q(\theta)$  es  $q(\hat{\theta})$ .

Ahora, considérese que  $q(\theta)$  no es una función inyectiva, es decir, para un valor  $\gamma$  pueden haber más de un valor de  $\theta$  que satisface  $q(\theta) = \gamma$ . En este caso, se usa la función de verosimilitud inducida  $L^*$ , dada por:

$$L^*(\gamma; \mathbf{x}) = \sup_{\{\theta: q(\theta)=\gamma\}} L(\theta; \mathbf{x}). \quad (5.19)$$

El valor  $\hat{\gamma}$  que maximiza  $L^*(\gamma; \mathbf{x})$  se llama el **EMV** para  $q(\theta)$  y se tiene que el máximo de  $L^*$  y  $L$  coinciden. Supóngase que  $\hat{\gamma}$  es el valor que maximiza a  $L^*(\gamma; \mathbf{x})$ , luego se debe mostrar que  $L^*(\hat{\gamma}; \mathbf{x}) = L^*(q(\hat{\theta}); \mathbf{x})$ . Se tiene que:

$$\begin{aligned} L^*(\hat{\gamma}; \mathbf{x}) &= \sup_{\gamma} L^*(\gamma; \mathbf{x}) \quad (\text{porque } \hat{\gamma} \text{ es el EMV}) \\ &= \sup_{\gamma} \left[ \sup_{\{\theta: q(\theta)=\gamma\}} L(\theta; \mathbf{x}) \right] \quad (\text{expresión (5.19)}) \\ &= \sup_{\theta} L(\theta; \mathbf{x}) \quad \left( \begin{array}{l} \text{porque la maximización iterada es igual a la} \\ \text{maximización no condicionada sobre } \theta \end{array} \right) \\ &= L(\hat{\theta}; \mathbf{x}) \quad (\text{definición de } \hat{\theta}). \end{aligned}$$

Por otra parte,

$$\begin{aligned} L(\hat{\theta}; \mathbf{x}) &= \sup_{\{\theta: q(\theta)=q(\hat{\theta})\}} L(\theta; \mathbf{x}) \quad (\text{porque } \hat{\theta} \text{ es el EMV}) \\ &= L^*(q(\hat{\theta}); \mathbf{x}) \quad (\text{expresión (5.19)}). \end{aligned}$$

Entonces,

$$L^*(\hat{\gamma}; \mathbf{x}) = L^*(q(\hat{\theta}); \mathbf{x}).$$

Así que  $q(\hat{\theta})$  es el **EMV** de  $q(\theta)$ . ☑

**Ejemplo 5.24.** Sean  $X_1, X_2, \dots, X_n$ , variables aleatorias *iid* Bernoulli( $\theta$ ), y considérese las funciones no inyectivas  $q_1(\theta) = \theta^2$  y  $q_2(\theta) = \theta(1 - \theta)$ , encuentre un *EMV* para  $q_1(\theta)$  y  $q_2(\theta)$ .

*Solución.* En el ejemplo 4.8 se concluyó que  $\hat{\theta}_{MV} = \bar{X}$  y por el teorema 5.7 se tiene que los *EMV* para  $q_1(\theta)$  y  $q_2(\theta)$  son dados por:

$$q_1(\hat{\theta}_{MV}) = \bar{X}^2 \quad \text{y} \quad q_2(\hat{\theta}_{MV}) = \bar{X}(1 - \bar{X}). \quad \checkmark$$

## 5.10. Familia exponencial

En la familia exponencial de un parámetro, expresión (2.40), de acuerdo a [71] la función  $T(\mathbf{x})$  es una estadística suficiente para  $\theta$ , ya que  $p(\mathbf{x}; \theta)$  se puede expresar como:

$$\begin{aligned} p(\mathbf{x}; \theta) &= [\exp \{c(\theta)T(\mathbf{x}) + d(\theta) + S(\mathbf{x})\}] I_A(\mathbf{x}) \\ &= \underbrace{\exp [c(\theta)T(\mathbf{x}) + d(\theta)]}_{g(T(\mathbf{x}); \theta)} \times \underbrace{\exp [S(\mathbf{x})]}_{h(\mathbf{x})} \times I_A(\mathbf{x}), \end{aligned}$$

es decir, cumple el criterio de factorización de Fisher-Neyman (teorema 5.1). Por otra parte, si  $p(\mathbf{x}; \theta)$  es una *fdp* conjunta, entonces:

$$\begin{aligned} \underbrace{\int_A \cdots \int_A p(\mathbf{x}; \theta) \, d\mathbf{x}}_1 &= \int_A \cdots \int_A \exp \{c(\theta)T(\mathbf{x}) + d(\theta) + S(\mathbf{x})\} \, d\mathbf{x} \\ &= e^{d(\theta)} \int_A \cdots \int_A \exp \{c(\theta)T(\mathbf{x}) + S(\mathbf{x})\} \, d\mathbf{x} \\ e^{-d(\theta)} &= \int_A \cdots \int_A \exp \{c(\theta)T(\mathbf{x}) + S(\mathbf{x})\} \, d\mathbf{x}. \end{aligned}$$

Luego,

$$p(\mathbf{x}; \theta) = \frac{\exp \{c(\theta)T(\mathbf{x}) + S(\mathbf{x})\}}{\int_A \cdots \int_A \exp \{c(\theta)T(\mathbf{x}) + S(\mathbf{x})\} \, d\mathbf{x}} I_A(\mathbf{x}).$$

En general, si  $\mathbf{X} = \text{vec}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m)^{35}$  es un vector compuesto de  $nm$  variables aleatorias de  $m$  vectores aleatorios,  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m$ , *iid*  $P_\theta$ , donde  $P_\theta$  se distribuye según (2.40). Si  $\mathbf{x} = \text{vec}(x_1, x_2, \dots, x_m)$  es un conjunto de

<sup>35</sup>Operador *vec*: este operador crea un vector columna a partir de una matriz  $\mathbf{A}$  apilando las columnas de  $\mathbf{A}$  una debajo de otra. Para una matriz  $\mathbf{A} = [a_{ij}]$  de orden  $r \times s$ , se tiene (ver

posibles realizaciones del vector aleatorio  $\mathbf{X}$ , entonces la **fdp** conjunta es dada por:

$$\underbrace{\prod_{i=1}^m p_{x_i}(\mathbf{x}_i; \theta)} = \prod_{i=1}^m [\exp \{c(\theta)T(x_i) + d(\theta) + S(x_i)\}] I_A(x_i)$$

$$p_{\mathbf{X}}(\mathbf{x}; \theta) = \exp \left\{ c(\theta) \sum_{i=1}^m T(x_i) + md(\theta) + \sum_{i=1}^m S(x_i) \right\} \times I_{A^{(m)}}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\},$$

donde  $A^{(m)} = \{\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} : \mathbf{x}_i \in A; 1 \leq i \leq m\}$ . Entonces,

$$c^{(m)}(\theta) = c(\theta) \qquad d^{(m)}(\theta) = md(\theta)$$

$$T^{(m)}(\mathbf{x}) = \sum_{i=1}^m T(x_i) \qquad S^{(m)}(\mathbf{x}) = \sum_{i=1}^m S(x_i).$$

Nuevamente, se satisface el criterio de factorización de Fisher-Neyman (teorema 5.1) y, por ende,  $T^{(m)}(\mathbf{x})$  es una estadística suficiente para  $\theta$ .

**Ejemplo 5.25.** Sea  $\mathbf{X} = (X_1, X_2, \dots, X_r)^\top$  un vector de orden  $r$  de variables aleatorias **iid**  $N(\theta, \sigma^2)$ , con  $\sigma^2$  conocido. Hallar una estadística suficiente para  $\theta$ .

*Solución.* En este caso,  $n = 1$  y  $m = r$ , luego  $nm = r$ . La **fdp** conjunta

$$f_{\mathbf{X}}(\mathbf{x}; \theta) = \prod_{i=1}^r \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2}(x_i - \theta)^2 \right\}$$

$$= (2\pi\sigma^2)^{-\frac{r}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^r (x_i^2 - 2x_i\theta + \theta^2) \right\},$$

usando la identidad  $a^x = e^{x \ln(a)}$  se tiene que:

$$f_{\mathbf{X}}(\mathbf{x}; \theta) = \exp \left\{ -\frac{r}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^r x_i^2 + \frac{2\theta}{2\sigma^2} \sum_{i=1}^r x_i - \frac{r\theta^2}{2\sigma^2} \right\}$$

---

[29, p. 412]:

$$\text{vec}(\mathbf{A}) = \begin{pmatrix} \mathbf{A}_{\cdot 1} \\ \mathbf{A}_{\cdot 2} \\ \vdots \\ \mathbf{A}_{\cdot s} \end{pmatrix} \qquad \text{con} \qquad \mathbf{A}_{\cdot j} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{rj} \end{pmatrix}.$$

$$= \exp \left\{ \underbrace{\frac{\theta}{\sigma^2} \sum_{i=1}^r x_i}_{c^{(r)}(\theta)} \underbrace{+ r \left( -\frac{\theta^2}{2\sigma^2} \right)}_{d^{(r)}(\theta)} + \underbrace{\left( -\sum_{i=1}^r \left[ \frac{1}{2} \ln(2\pi\sigma^2) + \frac{x_i^2}{2\sigma^2} \right] \right)}_{S^{(r)}(x)} \right\}.$$

Entonces,

$$\begin{aligned} c^{(r)}(\theta) &= c(\theta) = \frac{\theta}{\sigma^2}, & d^{(r)}(\theta) &= rd(\theta) = r \times \left( -\frac{\theta^2}{2\sigma^2} \right), \\ T^{(r)}(\mathbf{x}) &= \sum_{i=1}^r T(x_i) = \sum_{i=1}^r x_i, \\ S^{(r)}(\mathbf{x}) &= \sum_{i=1}^r S(x_i) = \sum_{i=1}^r \left( -\left[ \frac{1}{2} \ln(2\pi\sigma^2) + \frac{x_i^2}{2\sigma^2} \right] \right). \quad \square \end{aligned}$$

### 5.10.1. Estimación de máxima verosimilitud en la familia exponencial

El método de *máxima verosimilitud* de Fisher, presentado en la sección 4.2, para la estimación puntual del parámetro desconocido  $\theta$  de una muestra dada, consiste en elegir ese valor  $\theta$  que maximiza la verosimilitud de la muestra. Puesto que

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n p(x_i; \theta) \quad \text{y} \quad \ell(\theta; \mathbf{x}) = \ln [L(\theta; \mathbf{x})] = \sum_{i=1}^n \ln [p(x_i; \theta)]$$

son máximas para el mismo valor de  $\theta$ , es más sencillo maximizar  $\ell(\theta; \mathbf{x})$ . Si el rango de la distribución es independiente del parámetro,  $\theta$ , el EMV de  $\theta$ , será una solución (si la hubiera) de la *ecuación de verosimilitud*

$$\frac{\partial \ell(\theta; \mathbf{x})}{\partial \theta} = \frac{L'(\theta; \mathbf{x})}{L(\theta; \mathbf{x})} = \sum_{i=1}^n \frac{p'(x_i; \theta)}{p(x_i; \theta)} = 0. \quad (5.20)$$

Supóngase que  $p(x_i; \theta)$  se distribuye según (2.40), entonces como ha demostrado [41]:

$$\begin{aligned} \frac{\partial \ell(\theta; \mathbf{x})}{\partial \theta} &= \sum_{i=1}^n \frac{\partial}{\partial \theta} [c(\theta)T(x_i) + d(\theta) + S(x_i)] \\ &= \sum_{i=1}^n [c'(\theta)T(x_i) + d'(\theta)] = c'(\theta) \sum_{i=1}^n T(x_i) + nd'(\theta), \quad (5.21) \end{aligned}$$

igualando a cero se obtiene una solución  $\hat{\theta}$  para cada muestra de cualquier tamaño. Luego,

$$-\frac{d'(\hat{\theta})}{c'(\hat{\theta})} = \frac{1}{n} \sum_{i=1}^n T(x_i).$$

Para verificar que este valor es un máximo, nótese que el valor esperado de (5.21) usando la expresión (5.13) está dado por:

$$\mathbb{E}_\theta \left[ \frac{\partial \ell(\theta; \mathbf{x})}{\partial \theta} \right] = 0 \quad \Rightarrow \quad c'(\theta) \sum_{i=1}^n \mathbb{E}_\theta [T(x_i)] + nd'(\theta) = 0$$

y la varianza por la expresión (5.16) queda:

$$\mathbb{V}_\theta \left[ \frac{\partial \ell(\theta; \mathbf{x})}{\partial \theta} \right] = \mathbb{E}_\theta \left[ \left( \frac{\partial \ell(\theta; \mathbf{x})}{\partial \theta} \right)^2 \right] = I_n(\theta) = [c'(\theta)]^2 \sum_{i=1}^n \mathbb{V}_\theta [T(x_i)].$$

Por otra parte, derivando dos veces respecto a  $\theta$  la función  $\ell(\theta; \mathbf{x})$ :

$$\begin{aligned} \frac{\partial^2 \ell(\theta; \mathbf{x})}{\partial \theta^2} &= \frac{\partial}{\partial \theta} \left( \frac{\partial \ln [p(\theta; \mathbf{x})]}{\partial \theta} \right) = \frac{\partial}{\partial \theta} \left( \frac{1}{p(\mathbf{x}; \theta)} \frac{\partial p(\mathbf{x}; \theta)}{\partial \theta} \right) \\ &= \frac{1}{p(\mathbf{x}; \theta)} \frac{\partial^2 p(\mathbf{x}; \theta)}{\partial \theta^2} - \left( \frac{1}{p(\mathbf{x}; \theta)} \frac{\partial p(\mathbf{x}; \theta)}{\partial \theta} \right)^2 \\ &= \frac{1}{p(\mathbf{x}; \theta)} \frac{\partial^2 p(\mathbf{x}; \theta)}{\partial \theta^2} - \left( \frac{\partial \ell(\theta; \mathbf{x})}{\partial \theta} \right)^2, \end{aligned}$$

tomando valor esperado a esta última expresión:

$$\mathbb{E}_\theta \left[ \frac{\partial^2 \ell(\theta; \mathbf{x})}{\partial \theta^2} \right] = \mathbb{E}_\theta \left[ \frac{1}{p(\mathbf{x}; \theta)} \frac{\partial^2 p(\mathbf{x}; \theta)}{\partial \theta^2} \right] - \mathbb{E}_\theta \left[ \left( \frac{\partial \ell(\theta; \mathbf{x})}{\partial \theta} \right)^2 \right] = -I_n(\theta).$$

Al derivar respecto a  $\theta$  la expresión (5.21) se obtiene:

$$\begin{aligned} \frac{\partial^2 \ell(\theta; \mathbf{x})}{\partial \theta^2} &= c''(\theta) \sum_{i=1}^n T(x_i) + nd''(\theta) \\ &= c''(\theta) \left( -n \frac{d'(\theta)}{c'(\theta)} \right) + nd''(\theta) \end{aligned}$$

y tomando valor esperado

$$\mathbb{E}_\theta \left[ \frac{\partial^2 \ell(\theta; \mathbf{x})}{\partial \theta^2} \right] = -n \left[ \frac{c''(\theta)}{c'(\theta)} d'(\theta) - d''(\theta) \right] = -n I_1(\theta) = -I_n(\theta) < 0,$$

ya que  $I_n(\theta)$  es positivo para todo  $\theta$  porque es la varianza dada en (5.16). En particular, esto se satisface para  $\hat{\theta}$  y, por ende, se tiene un máximo.

## 5.10.2. Familia exponencial de un parámetro en forma natural

Realizando la reparametrización  $\eta = c(\theta)$  se obtiene la familia exponencial de un parámetro en forma natural dada en (2.42).

**Ejemplo 5.26.** Suponga que  $\mathbf{X} = (X_1, X_2, \dots, X_r)^\top$  es un vector de orden  $r$  de variables aleatorias iid Bin( $n, \theta$ ). Determine una estadística suficiente para  $\theta$ .

*Solución.* La distribución conjunta se establece a partir de la expresión (2.43):

$$\begin{aligned} p_{\mathbf{X}}(\mathbf{x}; \theta) &= \prod_{i=1}^r \left[ \exp \left\{ \ln \binom{n}{x_i} + x_i \ln \left( \frac{\theta}{1-\theta} \right) + n \ln(1-\theta) \right\} I_{A^*}(x_i) \right] \\ &= \exp \left( \sum_{i=1}^r \left[ \ln \binom{n}{x_i} + x_i \ln \left( \frac{\theta}{1-\theta} \right) + n \ln(1-\theta) \right] \right) I_{A^*}\{x_1, x_2, \dots, x_r\} \\ &= \exp \left( \underbrace{\ln \left( \frac{\theta}{1-\theta} \right)}_{\eta(\theta)} \underbrace{\sum_{i=1}^r x_i}_{T(\mathbf{x})} + \underbrace{nr \ln(1-\theta)}_{d_0(\eta(\theta))} + \underbrace{\sum_{i=1}^r \ln \binom{n}{x_i}}_{S(\mathbf{x})} \right) I_{A^*}\{x_1, x_2, \dots, x_r\}, \end{aligned}$$

entonces,

$$\begin{aligned} \eta(\theta) &= \ln \left( \frac{\theta}{1-\theta} \right) \Rightarrow e^\eta = \frac{\theta}{1-\theta} \Rightarrow \theta(\eta) = \frac{e^\eta}{1+e^\eta} \\ d_0(\eta) &= d_0(\eta(\theta)) = nr \ln(1-\theta) = nr \ln \left( 1 - \frac{e^\eta}{1+e^\eta} \right) = nr \ln \left( \frac{1}{1+e^\eta} \right) \\ &= -nr \ln(1+e^\eta) \\ T(\mathbf{x}) &= \sum_{i=1}^r x_i \qquad S(\mathbf{x}) = \sum_{i=1}^r \ln \binom{n}{x_i}. \quad \square \end{aligned}$$

Además, a partir de (2.42) se tienen resultados útiles para evaluar la esperanza y la varianza de  $\mathbf{X}$ . Antes de enunciar el teorema al respecto, sea  $H$  la colección de todos los  $\eta$  tales que  $d_0(\eta)$  es finito.

**Teorema 5.8.** De acuerdo con [12, p. 52], si  $\mathbf{X}$  se distribuye según (2.42) y  $\eta$  es un punto interior de  $H$ , la función generadora de momentos de  $T(\mathbf{X})$  existe y es dada por:

$$m_{T(\mathbf{X})}(s) = \mathbb{E}(e^{sT(\mathbf{X})}) = e^{d_0(\eta) - d_0(s+\eta)} \quad (5.22)$$

para  $s$  en una vecindad de cero.

*Demostración.* Prueba para el caso continuo: se tiene de la definición 1.25 y de la expresión (2.42) que:

$$\begin{aligned}
 m_{T(\mathbf{X})}(s) &= \mathbb{E} \left( e^{sT(\mathbf{X})} \right) = \int \cdots \int_A e^{sT(\mathbf{x})} f(\mathbf{x}; \eta) \, d\mathbf{x} \\
 &= \int \cdots \int_A e^{sT(\mathbf{x})} e^{\eta T(\mathbf{x}) + d_0(\eta) + S(\mathbf{x})} \, d\mathbf{x} \\
 &= \int \cdots \int_A e^{(s+\eta)T(\mathbf{x}) + d_0(\eta) + S(\mathbf{x})} \, d\mathbf{x} \\
 &= \int \cdots \int_A e^{(s+\eta)T(\mathbf{x}) + d_0(\eta) + S(\mathbf{x}) + d_0(\eta+s) - d_0(\eta+s)} \, d\mathbf{x} \\
 &= e^{d_0(\eta) - d_0(\eta+s)} \int \cdots \int_A e^{(s+\eta)T(\mathbf{x}) + d_0(\eta+s) + S(\mathbf{x})} \, d\mathbf{x} \\
 &= e^{d_0(\eta) - d_0(\eta+s)} \underbrace{\int \cdots \int_A f(\mathbf{x}; \eta + s) \, d\mathbf{x}}_1 \\
 m_{T(\mathbf{X})}(s) &= e^{d_0(\eta) - d_0(\eta+s)}. \quad \square
 \end{aligned}$$

Como utilidad del teorema 5.8, se tiene el siguiente resultado:

**Corolario 5.1.** Si  $\mathbf{X}$  se distribuye según (2.42) y  $T(\mathbf{X})$  es una estadística, entonces,

$$\mathbb{E}(T(\mathbf{X})) = -d'_0(\eta) \quad y \quad \mathbb{V}(T(\mathbf{X})) = -d''_0(\eta). \quad (5.23)$$

*Demostración.* Usando la propiedad 6 del teorema 1.12,

$$\begin{aligned}
 \mathbb{E}(T(\mathbf{X})) &= \frac{\partial}{\partial s} m_{T(\mathbf{X})}(s) \Big|_{s=0} = \frac{\partial}{\partial s} \left\{ e^{d_0(\eta) - d_0(\eta+s)} \right\} \Big|_{s=0} \\
 &= e^{d_0(\eta) - d_0(\eta+s)} \left[ -d'_0(\eta + s) \right] \Big|_{s=0} \quad \text{con } d'_0(\eta + s) = \frac{\partial}{\partial s} d_0(\eta + s) \\
 &= e^{d_0(\eta) - d_0(\eta)} \left[ -d'_0(\eta) \right] \\
 &= -d'_0(\eta).
 \end{aligned}$$

Por otra parte,

$$\mathbb{E} \left( [T(\mathbf{X})]^2 \right) = \frac{\partial^2}{\partial s^2} m_{T(\mathbf{X})}(s) \Big|_{s=0}$$

$$\begin{aligned}
&= \frac{\partial}{\partial s} \left\{ e^{d_0(\eta) - d_0(\eta+s)} [-d'_0(\eta+s)] \right\} \Big|_{s=0} \\
&= e^{d_0(\eta) - d_0(\eta+s)} \left\{ -d'_0(\eta+s) [-d'_0(\eta+s)] + [-d''_0(\eta+s)] \right\} \Big|_{s=0} \\
&= e^{d_0(\eta) - d_0(\eta)} [d'_0(\eta)]^2 + e^{d_0(\eta) - d_0(\eta)} [-d''_0(\eta)] \\
&= [d'_0(\eta)]^2 - d''_0(\eta).
\end{aligned}$$

Entonces,

$$\begin{aligned}
\mathbb{V}(T(\mathbf{X})) &= \mathbb{E} \left( [T(\mathbf{X})]^2 \right) - (\mathbb{E}[T(\mathbf{X})])^2 \\
&= [d'_0(\eta)]^2 - d''_0(\eta) - \cancel{[-d'_0(\eta)]^2} \\
&= -d''_0(\eta). \quad \square
\end{aligned}$$

**Ejemplo 5.27.** Sea  $X$  una variable aleatoria con *fmp*  $\text{Bin}(n, \theta)$ ,  $0 < \theta < 1$ . Encuentre  $\mathbb{E}(T(\mathbf{X}))$  y  $\mathbb{V}(T(\mathbf{X}))$ .

*Solución.* Por el ejemplo 2.17, se tiene que:

$$p(x; \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} I_A(x) \quad \text{con } A = \{0, 1, 2, \dots, n\}$$

luego,  $T(x) = x$  y  $d_0(\eta) = -n \ln(1 + e^\eta)$ . Por las expresiones (5.23) se obtiene que:

$$\begin{aligned}
\mathbb{E}(T(\mathbf{X})) &= -d'_0(\eta) = -\frac{d}{d\eta} [-n \ln(1 + e^\eta)] = n \underbrace{\frac{e^\eta}{1 + e^\eta}}_{\theta} = n\theta \\
\mathbb{V}(T(\mathbf{X})) &= -d''_0(\eta) = -\frac{d}{d\eta} \left[ -n \frac{e^\eta}{1 + e^\eta} \right] = n \left[ \frac{e^\eta(1 + e^\eta) - e^\eta e^\eta}{(1 + e^\eta)^2} \right] \\
&= n \left[ \frac{e^\eta}{(1 + e^\eta)^2} \right] = n \left[ \underbrace{\frac{e^\eta}{(1 + e^\eta)}}_{\theta} \underbrace{\frac{1}{(1 + e^\eta)}}_{1-\theta} \right] = n\theta(1 - \theta). \quad \square
\end{aligned}$$

**Ejemplo 5.28.** Supóngase, como en [12, p. 53], que  $X_1, X_2, \dots, X_n$  es una muestra aleatoria de una población con *fdp* de Rayleigh:

$$f(x; \theta) = \left( \frac{x}{\theta^2} \right) \exp \left\{ -\frac{x^2}{2\theta^2} \right\} \quad \text{para } x > 0, \theta > 0.$$

Encuentre  $\mathbb{E}(T(\mathbf{X}))$  y  $\mathbb{V}(T(\mathbf{X}))$ .

*Solución.* Veamos que esta **fdp** es un miembro de la familia exponencial:

$$\begin{aligned}
 p(\mathbf{x}; \theta) &= \prod_{i=1}^n \left( \frac{x_i}{\theta^2} \exp \left\{ -\frac{x_i^2}{2\theta^2} \right\} \right) = \left( \prod_{i=1}^n \frac{x_i}{\theta^2} \right) \exp \left\{ -\sum_{i=1}^n \frac{x_i^2}{2\theta^2} \right\} \\
 &= \exp \left\{ \ln \left( \prod_{i=1}^n \frac{x_i}{\theta^2} \right) - \frac{1}{2\theta^2} \sum_{i=1}^n x_i^2 \right\} \\
 &= \exp \left\{ \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \ln(\theta^2) - \frac{1}{2\theta^2} \sum_{i=1}^n x_i^2 \right\} \\
 &= \exp \left\{ \underbrace{-\frac{1}{2\theta^2} \sum_{i=1}^n x_i^2}_{\eta(\theta)} + \underbrace{(-n \ln \theta^2)}_{d_0(\eta(\theta))} + \underbrace{\sum_{i=1}^n \ln x_i}_{S(\mathbf{x})} \right\},
 \end{aligned}$$

luego,

$$\begin{aligned}
 \eta(\theta) &= -\frac{1}{2\theta^2} \quad \Rightarrow \quad \theta^2 = \frac{-1}{2\eta}, \\
 d_0(\eta) &= d_0(\eta(\theta)) = -n \ln(\theta^2) = -n \ln\left(\frac{-1}{2\eta}\right) = -n \ln\left(\frac{1}{-2\eta}\right) \\
 &= -n [\ln(1) - \ln(-2\eta)] = n \ln(-2\eta) \\
 T(\mathbf{x}) &= \sum_{i=1}^n x_i^2 \qquad \qquad \qquad S(\mathbf{x}) = \sum_{i=1}^n \ln x_i
 \end{aligned}$$

Por tanto se puede concluir que:

- $\sum_{i=1}^n X_i^2$  es una estadística suficiente para  $\theta$ .
- La esperanza y varianza de  $T(\mathbf{X})$  son dadas por las expresiones (5.23):

$$\begin{aligned}
 \mathbb{E} \left( \sum_{i=1}^n X_i^2 \right) &= -\frac{d}{d\eta} [n \ln(-2\eta)] = (-1) \frac{n}{-2\eta} (-2) = \frac{-n}{\eta} = \frac{-n}{-\frac{1}{2\theta^2}} \\
 &= 2n\theta^2
 \end{aligned}$$

$$\begin{aligned}
 \mathbb{V} \left( \sum_{i=1}^n X_i^2 \right) &= -\frac{d^2}{d\eta^2} (n \ln(-2\eta)) = -\frac{d}{d\eta} \left( \frac{n}{\eta} \right) = -(-n\eta^{-2}) \\
 &= \frac{n}{\eta^2} = \frac{n}{\left(-\frac{1}{2\theta^2}\right)^2} = 4n\theta^4.
 \end{aligned}$$

□

### 5.10.3. Familia exponencial de $k$ parámetros

Una familia de distribuciones  $\{P_\theta; \theta \in \Theta\}$  con  $\Theta \subset \mathbb{R}^k$  se dice que es una familia exponencial de  $k$  parámetros, si existen funciones de valor real  $c_1, \dots, c_k$  y  $D$  con argumento  $\theta$ , funciones de valor real  $T_1, \dots, T_k$ ,  $S$  definidas en  $\mathbb{R}^p$  y un conjunto  $A \subset \mathbb{R}^p$  tal que la **fdp** (**fmp**)  $P_\theta$  se puede expresar como [12, p. 53]:

$$p(x, \theta) = \left[ \exp \left\{ \sum_{i=1}^k c_i(\theta) T_i(x) + D(\theta) + S(x) \right\} \right] I_A(x) \quad (5.24)$$

$$= \left[ \exp \{ [C(\theta)]^\top T(x) + D(\theta) + S(x) \} \right] I_A(x), \quad (5.25)$$

donde  $\theta = (\theta_1, \theta_2, \dots, \theta_k)^\top$  y la función  $C(\theta) = (c_1(\theta), c_2(\theta), \dots, c_k(\theta))^\top$ .

**Nota 5.16.** *El vector aleatorio*

$$T(x) = \left( T_1(x) \quad \dots \quad T_k(x) \right)^\top \quad (5.26)$$

*es una estadística suficiente.*

**Nota 5.17.** *De acuerdo con [20, p. 288], si  $X_1, X_2, \dots, X_m$  son variables aleatorias **iid** de una familia exponencial de  $k$  parámetros con **fdp** o con **fmp** de la forma:*

$$p(x; \theta) = \exp \left\{ \sum_{j=1}^k c_j(\theta) T_j(x) + D(\theta) + S(x) \right\} I_A(x).$$

*Entonces la estadística*

$$T^{(m)}(\mathbf{X}) = \left( \sum_{i=1}^m T_1(X_i) \quad \sum_{i=1}^m T_2(X_i) \quad \dots \quad \sum_{i=1}^m T_k(X_i) \right)^\top$$

*es completa si  $\{C(\theta); \theta \in \Theta\}$  contiene un conjunto abierto<sup>36</sup> en  $\mathbb{R}^k$ .*

**Ejemplo 5.29.** *Sean  $X_1, X_2, \dots, X_n$  variables aleatorias **iid** normales con media  $\mu$  y varianza  $\sigma^2$ , desconocidas. Encuentre una estadística suficiente y completa para  $\theta = (\mu, \sigma^2)^\top$ .*

<sup>36</sup>Si  $(a_1, b_1), (a_2, b_2), \dots, (a_k, b_k)$  son  $k$  intervalos abiertos, al conjunto  $(a_1, b_1) \times (a_2, b_2) \times \dots \times (a_k, b_k) = \{(x_1, x_2, \dots, x_k) : a_i < x_i < b_i, 1 \leq i \leq k\}$  se dice es un rectángulo  $k$  abierto (ver [12, p. 449]).

*Solución.* Considerando la información del ejemplo 5.5, se tiene que:

$$\begin{aligned}
 p(x_1, \dots, x_n; \theta) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ \frac{-1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} \\
 &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{n\mu^2}{2\sigma^2} \right\} \exp \left\{ \frac{-1}{2\sigma^2} \left[ \sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i \right] \right\} \\
 &= \exp \left\{ \frac{-1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i - \frac{n}{2} \ln(2\pi\sigma^2) - \frac{n\mu^2}{2\sigma^2} \right\} \\
 &= \exp \left\{ \underbrace{\left( \frac{\mu}{\sigma^2} \quad \frac{-1}{2\sigma^2} \right)}_{C(\theta)} \underbrace{\begin{pmatrix} \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i^2 \end{pmatrix}}_{T(x)} + \underbrace{\left( -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{n\mu^2}{2\sigma^2} \right)}_{D(\theta)} \right\}, \quad (5.27)
 \end{aligned}$$

luego,  $S(x) = 0$ , y

$$c_1(\theta) = \frac{\mu}{\sigma^2}, \quad c_2(\theta) = \frac{-1}{2\sigma^2}, \quad T_1(x) = \sum_{i=1}^n x_i, \quad T_2(x) = \sum_{i=1}^n x_i^2,$$

entonces  $T(x) = \begin{pmatrix} \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i^2 \end{pmatrix}$  es una estadística suficiente y completa para  $\theta$ .  $\square$

#### 5.10.4. Estimación de máxima verosimilitud en la familia exponencial de $k$ parámetros

Supóngase que para  $k > 1$  y  $p = k$ , la función  $p(x_i; \theta)$  se distribuye según (5.24), si  $C(\theta)$  es diferenciable con respecto a  $\theta$ , entonces como han demostrado en [46]:

$$\frac{\partial \ell(\theta; x)}{\partial \theta} = [\mathcal{J}_C(\theta)]^\top [T(X)] + \nabla D(\theta), \quad (5.28)$$

donde  $\nabla D(\theta)$  es el vector gradiente de  $D(\theta)$  (ver nota B.10) y  $\mathcal{J}_C(\theta)$  es la matriz jacobiana dada en (2.57). Para verificar que este valor es un máximo, usando (5.13) se evalúa el valor esperado de (5.28), es decir

$$\mathbb{E}_\theta \left[ \frac{\partial \ell(\theta; x)}{\partial \theta} \right] = \mathbf{0} \quad \Rightarrow \quad [\mathcal{J}_C(\theta)]^\top \mathbb{E}_\theta [T(X)] + \nabla D(\theta) = \mathbf{0}.$$

Luego,

$$\nabla D(\boldsymbol{\theta}) = -[\mathcal{J}_C(\boldsymbol{\theta})]^\top \mathbb{E}_\theta[T(\mathbf{X})] \Rightarrow \mathbb{E}_\theta[T(\mathbf{X})] = -[\mathcal{J}_C^{-1}(\boldsymbol{\theta})]^\top \nabla D(\boldsymbol{\theta}).$$

Reemplazando en (5.28) la derivada de la función  $\ell(\boldsymbol{\theta}; \mathbf{x})$  se expresa como:

$$\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \boldsymbol{\theta}} = [\mathcal{J}_C(\boldsymbol{\theta})]^\top [T(\mathbf{X}) - \mathbb{E}_\theta[T(\mathbf{X})]]$$

y la varianza de (5.28) por la propiedad 2 del teorema 2.11 queda como:

$$\begin{aligned} \mathbb{V}_\theta \left[ \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \boldsymbol{\theta}} \right] &= \mathbb{E}_\theta \left[ \left( \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \boldsymbol{\theta}} \right) \left( \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \boldsymbol{\theta}} \right)^\top \right] \\ &= [\mathcal{J}_C(\boldsymbol{\theta})]^\top \mathbb{V}_\theta[T(\mathbf{X})] [\mathcal{J}_C(\boldsymbol{\theta})] = \mathcal{F}(\boldsymbol{\theta}), \end{aligned} \quad (5.29)$$

donde  $\mathcal{F}(\boldsymbol{\theta})$  es una matriz definida positiva de tamaño  $k \times k$  que se denomina *matriz de información de Fisher*, el elemento  $(i, j)$  de esta matriz es:

$$\mathcal{F}(\boldsymbol{\theta})_{ij} = -\mathbb{E}_\theta \left[ \frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \theta_i \partial \theta_j} \right]. \quad (5.30)$$

Por otra parte, derivando dos veces respecto a  $\boldsymbol{\theta}$  la función  $\ell(\boldsymbol{\theta}; \mathbf{x})$

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= [T(\mathbf{X}) - \mathbb{E}_\theta[T(\mathbf{X})]]^\top * \begin{bmatrix} \mathcal{J}_{\nabla c_1}(\boldsymbol{\theta}) & \mathcal{J}_{\nabla c_2}(\boldsymbol{\theta}) & \dots & \mathcal{J}_{\nabla c_k}(\boldsymbol{\theta}) \end{bmatrix} \\ &\quad - [\mathcal{J}_C(\boldsymbol{\theta})]^\top \mathbb{V}_\theta[T(\mathbf{X})] [\mathcal{J}_C(\boldsymbol{\theta})], \end{aligned}$$

donde el producto  $*$  fue dado en [58]. Tomando valor esperado a esta última expresión:

$$\mathbb{E}_\theta \left[ \frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right] = -\mathbb{E}_\theta \left[ \left( \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \boldsymbol{\theta}} \right) \left( \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \boldsymbol{\theta}} \right)^\top \right] = -\mathbb{V}_\theta \left[ \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{x})}{\partial \boldsymbol{\theta}} \right]$$

es una matriz definida negativa ya que la matriz dada en la expresión (5.29) es definida positiva para todo  $\boldsymbol{\theta}$ . En particular, esto se satisface para  $\hat{\boldsymbol{\theta}}$  y, por ende, se tiene un máximo.

**Ejemplo 5.30.** Considerando la información del ejemplo 5.29, verifique que la estadística suficiente y completa encontrada para  $\boldsymbol{\theta} = (\mu, \sigma^2)^\top$  es un máximo.

*Solución.* En el ejemplo 5.29 se obtuvo que:

$$c_1(\boldsymbol{\theta}) = \frac{\mu}{\sigma^2}, \quad c_2(\boldsymbol{\theta}) = \frac{-1}{2\sigma^2}, \quad D(\boldsymbol{\theta}) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{n\mu^2}{2\sigma^2},$$

y  $T(x) = \left( \sum_{i=1}^n x_i \quad \sum_{i=1}^n x_i^2 \right)^T$  es una estadística suficiente y completa para  $\theta$ .

En este caso, la matriz jacobiana dada en (2.57) queda:

$$\mathcal{J}_C(\theta) = \frac{\partial}{\partial \theta} C(\theta) = \begin{bmatrix} \nabla c_1(\theta) \\ \nabla c_2(\theta) \end{bmatrix} = \begin{bmatrix} \frac{\partial c_1(\theta)}{\partial \mu} & \frac{\partial c_1(\theta)}{\partial \sigma^2} \\ \frac{\partial c_2(\theta)}{\partial \mu} & \frac{\partial c_2(\theta)}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} 1 & -\frac{\mu}{\sigma^4} \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}.$$

Nótese que:

$$\nabla D(\theta) = \begin{bmatrix} \frac{\partial D(\theta)}{\partial \mu} \\ \frac{\partial D(\theta)}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} -\frac{n\mu}{\sigma^2} \\ n\mu - \frac{n\mu^2}{2\sigma^2} + \frac{n\mu^2}{2\sigma^4} \end{bmatrix},$$

y, por lo tanto,

$$\begin{aligned} \mathbb{E}_\theta[T(X)] &= -[\mathcal{J}_C^{-1}(\theta)]^T \nabla D(\theta) = [\mathcal{J}_C^T(\theta)]^{-1} (-\nabla D(\theta)) \\ &= \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ -\frac{\mu}{\sigma^4} & \frac{1}{2\sigma^4} \end{bmatrix}^{-1} \begin{bmatrix} -\frac{n\mu}{\sigma^2} \\ n\mu - \frac{n\mu^2}{2\sigma^2} + \frac{n\mu^2}{2\sigma^4} \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 \\ 2\mu\sigma^2 & 2\sigma^4 \end{bmatrix} \begin{bmatrix} -\frac{n\mu}{\sigma^2} \\ n\mu - \frac{n\mu^2}{2\sigma^2} + \frac{n\mu^2}{2\sigma^4} \end{bmatrix} \\ &= \begin{bmatrix} n\mu \\ n\sigma^2 + n\mu^2 \end{bmatrix}. \end{aligned}$$

Además, la matriz de información de Fisher es definida positiva ya que:

$$\begin{aligned} \mathcal{I}(\theta) &= -\mathbb{E}_\theta[H_\ell(\theta)] \\ &= \begin{bmatrix} -\mathbb{E}_\theta \left[ \frac{\partial^2 \ell(\theta; x)}{\partial \mu^2} \right] & -\mathbb{E}_\theta \left[ \frac{\partial^2 \ell(\theta; x)}{\partial \sigma^2 \partial \mu} \right] \\ -\mathbb{E}_\theta \left[ \frac{\partial^2 \ell(\theta; x)}{\partial \sigma^2 \partial \mu} \right] & -\mathbb{E}_\theta \left[ \frac{\partial^2 \ell(\theta; x)}{\partial (\sigma^2)^2} \right] \end{bmatrix} = \begin{bmatrix} n & 0 \\ 0 & n \end{bmatrix} \frac{1}{2\sigma^4}, \end{aligned}$$

sus elementos son los negativos de la matriz hessiana dada en (4.18).  $\square$

## 5.11. Ejercicios

1. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid tal que  $X_1 \sim N(\theta, 1)$  [65, p. 366]. Encuentre:

- Si existe, un estimador insesgado para  $\theta^2$  cuando  $n = 1$ .
- Si existe, un estimador insesgado para  $P(X > 0)$ .

- c) ¿Cuál es el **EMV** para  $P(X > 0)$ ?  
 d) Si existe, un **UMVUE** para a)  $\theta^2$  y b)  $P(X > 0)$ .

2. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias **iid** tal que [65, p. 365]:

$$f_{X_1}(x) = \theta x^{\theta-1} I_{(0,1)}(x), \quad \theta > 0.$$

Encuentre:

- a) Un **EMV** para  $\mu = \frac{\theta}{1+\theta}$ .  
 b) Una estadística suficiente y verifique completez. ¿ $\sum_{i=1}^n X_i$  es una estadística suficiente?  
 c) Una función de  $\theta$  para la cual exista un estimador insesgado cuya varianza coincida con la cota de Cramér-Rao.  
 d) Un **UMVUE** para a)  $\theta$ , b)  $\frac{1}{\theta}$  y c)  $\mu = \frac{\theta}{1+\theta}$ .

3. Sea  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ . Muestre que [20, p. 363]:

- a) Para cualquier estimador de la forma  $aS^2$ , con  $a$  constante,

$$\text{ECM}(aS^2) = \mathbb{E} \left( [aS^2 - \sigma^2]^2 \right) = a^2 \mathbb{V}(S^2) + (a-1)^2 \sigma^4.$$

- b) Si  $\kappa = \beta_2(X)$  es la curtosis estandarizada (ver definición 1.23),

$$\mathbb{V}(S^2) = \frac{1}{n} \left( \kappa - \frac{n-3}{n-1} \right) \sigma^4.$$

- c) Si no se considera normalidad, el **ECM**( $aS^2$ ) se minimiza en

$$a = \frac{n-1}{(n+1) + \frac{(\kappa-3)(n-1)}{n}}.$$

4. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias **iid** Gumbel, cuya **fdp** es [65, p. 367]:

$$f(x; \theta) = \exp \left\{ -(x - \theta) - e^{-(x-\theta)} \right\}$$

con  $-\infty < \theta < \infty$ . Encuentre

- a) Una estadística suficiente y completa.  
 b) La cota de Cramér-Rao para estimadores insesgados de  $\theta$ .

5. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid tal que [20, p. 362]

$$f_X(x) = \frac{\ln(\theta)}{\theta - 1} \theta^x, \quad 0 < x < 1 \text{ y } \theta > 1.$$

¿Existe una función de  $\theta$ , digamos  $g(\theta)$ , para la cual se encuentre un estimador insesgado cuya varianza alcance la cota de Cramér-Rao? Si existe, encuéntrela. Si no existe, muestre por qué no existe.

6. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid con fdp uniforme  $U(\theta, 2\theta)$ , con  $0 < \theta < \infty$  [76, p. 287]. Defínanse:

$$Y_1 = \frac{n+1}{2n+1} X_{(n)} \quad \text{y} \quad Y_2 = \frac{n+1}{5n+4} [2X_{(n)} + X_{(1)}].$$

Muestre que  $Y_1$  y  $Y_2$  son estimadores insesgados de  $\theta$ .

7. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid beta, con parámetro  $\alpha$  conocido, cuya fdp es dada por (1.22):

$$p(x; \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad 0 < x < 1, \quad \beta > 0.$$

Encontrar una estadística suficiente para el parámetro de escala  $\beta$ .

8. Suponga, como en [20, p. 300], que  $X_1, X_2, \dots, X_n$  son variables aleatorias iid gamma, con parámetros de forma  $r = \alpha$  y de escala  $\lambda = \frac{1}{\theta}$ , cuya fdp es dada por (1.25):

$$p(x; \theta) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-\frac{x}{\theta}} \quad x > 0.$$

Encontrar una estadística suficiente:

- a) Para el parámetro  $\theta$  suponiendo que  $\alpha$  es conocido.
  - b) Para ambos parámetros  $(\alpha, \theta)$ .
9. Usando las funciones  $d_0(\eta)$ , encontradas en el ejercicio 2.17, hallar  $\mathbb{E}(T(X))$  y  $\mathbb{V}(T(X))$  para cada una de las familias:
- a) Poisson de parámetro  $\theta$ .
  - b) Binomial negativa de parámetros  $r$  (conocido) y  $\theta$ .
  - c) Normal  $N(\theta, \sigma^2)$ ,  $\sigma^2$  conocido.
  - d) Gamma  $(\lambda, \theta)$ ,  $\lambda$  conocido (parámetro de forma).
  - e) Normal inversa  $NI(\theta, \lambda)$ ,  $\lambda$  conocido (parámetro de forma).







En el capítulo 4 se estudiaron algunos métodos de estimación puntual. En este capítulo, en vez de obtener una estimación puntual para el parámetro  $\theta$  se buscará un intervalo que incluya al parámetro con una alta probabilidad. Esta probabilidad, denominada nivel de confianza del intervalo, se denota por  $(1 - \alpha)$  y es fijada por el investigador. Dentro de este capítulo, se presentan tres métodos que permiten construir intervalos de confianza, entre ellos, el método de la cantidad pivotal.

## 6.1. Conceptos básicos de intervalos de confianza

**Ejemplo 6.1.** Sean  $X_1, X_2, \dots, X_{50}$  una muestra aleatoria de una variable con distribución Poisson de parámetro  $\theta$ . Del ejemplo 4.11 se sabe que  $\bar{X} = \frac{1}{50} \sum_{i=1}^{50} X_i$  es un estimador de máxima verosimilitud para  $\theta$ . Así, por ejemplo, al extraer una muestra se tiene una estimativa (puntual) de  $\theta$  igual a  $\bar{x} = 7,72$ . Suponga que se extrae otra muestra y la estimativa (puntual) da  $\bar{x} = 7,08$ , y así sucesivamente, por ejemplo, se extraen 100 muestras, entonces se tendrán 100 estimativas (puntuales) para  $\theta$ . ¿Será que alguna estimativa puntual es igual a  $\theta$ ? A continuación se presenta un algoritmo en R que muestra las animaciones de la situación planteada.

```
library(animation)
remove(list=ls())
media.n=NULL
lambda=8
alpha=0.10
n=50
n.sim=100
for(i in 1:n.sim) {
  sam=rpois(n, lambda)
  LI=mean(sam)-sqrt(lambda)*qnorm(1-alpha/2)/sqrt(n)
  LU=mean(sam)+sqrt(lambda)*qnorm(1-alpha/2)/sqrt(n)
  media.n=rbind(media.n,cbind(i,mean(sam),LI,LU))
}
rg = range(c(media.n[,3], media.n[,4]))
xax = pretty(1:n.sim)
oopt = ani.options(interval = .4)
for (i in 1:n.sim) {
  plot(1:n.sim, ylim = rg, type = "n", xlab = "Muestra",
       ylab = "Estimativa (puntual) - promedio",xaxt = "n")
  abline(h = 0, lty = 2)
  axis(1, xax[xax <= i])
  points(1:i, media.n[1:i,2],pch=16)
```

```

    ani.pause()
  }
ani.options(oopt)
cvr = media.n[,3] < lambda & media.n[,4] > lambda
oopt = ani.options(interval = .1)
for (i in 1:n.sim) {
  plot(1:n.sim, ylim = rg, type = "n", xlab = "Muestra",
      ylab = "Intervalo de confianza",xaxt = "n")
  abline(h = 0, lty = 2)
  axis(1, xax[xax <= i])
  arrows(1:i, media.n[1:i,3], 1:i, media.n[1:i,4],
      length = par("din")[1]/n.sim * 0.5,
      angle = 90, code = 3, col = cvr[1:i] + 1)
  points(1:i, media.n[1:i,2], col = cvr[1:i] + 1)
  legend("topright", legend = format(c(i - sum(cvr[1:i]),
      sum(cvr[1:i])),width = nchar(n.sim)),bty = "n.sim", ncol = 2)
  legend("topleft",
      legend = paste("Tasa de cobertura:",
      format(round(mean(cvr[1:i]), 3),nsmall = 3)), bty = "n.sim")
  ani.pause()
}
ani.options(oopt)

```

**Ejemplo 6.2.** Considere que se seleccionan 100 muestras de una distribución normal estándar, es decir, media cero y varianza 1, de tamaño 50 cada una. A continuación se presenta el código en R, dicha simulación muestra las estimativas por intervalo para la media (parámetro de interés).

```

library(animation)
oopt = ani.options(interval = 0.2,nmax=100)
conf.int(level = 0.95, size = 50, cl = c("red", "gray"))

```

**DEFINICIÓN 6.1.** Una estimación por intervalo de un parámetro de valor real  $\theta$  es cualquier par de funciones  $L(x)$  y  $U(x)$  de una muestra que satisface  $L(x) \leq U(x)$  para todo  $x$ . Si  $X = x$  es observado, la inferencia  $L(x) \leq \theta \leq U(x)$  está hecha. Al intervalo aleatorio  $[L(X), U(X)]$  se denomina un **estimador por intervalo** [20, p. 417].

**Nota 6.1.** Intervalos a una cola (a un lado)

- Si  $L(x) = -\infty$ , entonces se tiene un intervalo unilateral  $(-\infty, U(x)]$ .
- Si  $U(x) = +\infty$  entonces se tiene un intervalo unilateral  $[L(x), +\infty)$ .

**Ejemplo 6.3.** De acuerdo con [21, p. 93], si  $X$  es una variable aleatoria con distribución uniforme continua en  $(0, \theta)$ , es decir,  $X \sim U(0, \theta)$ . Encuentre un estimador por intervalo para  $\theta$ .

*Solución.* Usando el teorema de transformación de variables aleatorias continuas (teorema 2.7),  $Y = \frac{X}{\theta}$ , tiene distribución uniforme continua en  $(0, 1)$ , es decir,  $Y \sim U(0, 1)$ . Sea  $\alpha$  tal que  $0 < \alpha < 1$ , entonces:

$$\begin{aligned} P\left(\underbrace{\frac{\alpha}{2} \leq \frac{X}{\theta} \leq 1 - \frac{\alpha}{2}}_Y\right) &= P\left(Y \leq 1 - \frac{\alpha}{2}\right) - P\left(Y < \frac{\alpha}{2}\right) \\ &= \left(1 - \frac{\alpha}{2}\right) - \frac{\alpha}{2} = 1 - \alpha, \end{aligned}$$

luego,

$$\begin{aligned} 1 - \alpha &= P\left(\frac{\alpha}{2} \leq \frac{X}{\theta} \leq 1 - \frac{\alpha}{2}\right) = P\left(\frac{1}{1 - \frac{\alpha}{2}} \leq \frac{\theta}{X} \leq \frac{1}{\frac{\alpha}{2}}\right) \\ &= P\left(\frac{X}{1 - \frac{\alpha}{2}} \leq \theta \leq \frac{X}{\frac{\alpha}{2}}\right), \end{aligned}$$

el intervalo cerrado  $\left[\frac{2X}{2-\alpha}, \frac{2X}{\alpha}\right]$  se denomina intervalo de confianza de  $\theta$  al  $(1 - \alpha)100\%$ .  $\checkmark$

**DEFINICIÓN 6.2.** Para un estimador por intervalo  $[L(\mathbf{X}), U(\mathbf{X})]$  de un parámetro  $\theta$ , la **probabilidad de cobertura** de  $[L(\mathbf{X}), U(\mathbf{X})]$  es la probabilidad de que el intervalo aleatorio  $[L(\mathbf{X}), U(\mathbf{X})]$  incluya el verdadero valor de  $\theta$ , es decir,  $P(\theta \in [L(\mathbf{X}), U(\mathbf{X})] | \theta)$  [20, p. 418].

**Ejemplo 6.4.** De acuerdo con [20, p. 418], si  $X_1, X_2, X_3, X_4$  son variables aleatorias iid  $X_1 \sim N(\mu, 1)$  y se considera la estadística  $T(\mathbf{X}) = \bar{X}$ . Encuentre la probabilidad de que el intervalo  $[L(\mathbf{X}), U(\mathbf{X})] = [\bar{X} - 1, \bar{X} + 1]$  contenga a  $\mu$ .

*Solución.* La probabilidad de que el intervalo  $[\bar{X} - 1, \bar{X} + 1]$  contenga a  $\mu$  es:

$$\begin{aligned} P(\mu \in [\bar{X} - 1, \bar{X} + 1] | \mu) &= P(\bar{X} - 1 \leq \mu \leq \bar{X} + 1) \\ &= P(-1 - \bar{X} \leq -\mu \leq 1 - \bar{X}) \\ &= P(-1 \leq \bar{X} - \mu \leq 1) \\ &= P\left(\frac{-1}{\frac{1}{\sqrt{4}}} \leq \frac{\bar{X} - \mu}{\sqrt{\mathbb{V}(\bar{X})}} \leq \frac{1}{\frac{1}{\sqrt{4}}}\right) \\ &= P(-2 \leq Z \leq 2) = 0,9544. \end{aligned}$$

Luego se tiene más de 0,95 de probabilidad de cubrir el parámetro desconocido,  $\mu$ , con el estimador por intervalo dado por  $[\bar{X} - 1, \bar{X} + 1]$ .  $\checkmark$

**DEFINICIÓN 6.3.** Según [20, p. 418], para una estimación por intervalo de un parámetro  $\theta$ , el **coeficiente de confianza** (nivel de confianza) de  $[L(\mathbf{X}), U(\mathbf{X})]$  es el ínfimo (ver nota B.8) de la probabilidad de cobertura, es decir,

$$\inf_{\theta} \left\{ \underbrace{P_{\theta}(\theta \in [L(\mathbf{X}), U(\mathbf{X})])}_{\text{Probabilidad de cobertura}} \middle| \theta \right\}.$$

**Ejemplo 6.5.** De acuerdo con [20, p. 419], si  $X_1, X_2, \dots, X_n$  son variables aleatorias *iid*  $X_1 \sim U(0, \theta)$  y se considera la estadística “máximo muestral”, es decir,  $Y = T(\mathbf{X}) = X_{(n)}$ . Encontrar la probabilidad de cobertura de los siguientes estimadores por intervalo para  $\theta$ :

1.  $[Y + a, Y + b]$      $0 \leq a < b$ ;     $a$  y  $b$  constantes.
2.  $[cY, dY]$      $1 \leq c < d$ ;     $c$  y  $d$  constantes.

*Solución.*

1. En este caso, considerando que  $\theta > b$ , la probabilidad de cobertura es:

$$\begin{aligned} P_{\theta}(\theta \in [Y + a, Y + b]) &= P_{\theta}(Y + a \leq \theta \leq Y + b) = P_{\theta}(a \leq \theta - Y \leq b) \\ &= P_{\theta}(-b \leq Y - \theta \leq -a) = P_{\theta}(\theta - b \leq Y \leq \theta - a) \\ &\stackrel{\theta > 0}{=} \underbrace{P_{\theta}}_{\theta > 0} \left( 1 - \frac{b}{\theta} \leq \underbrace{\frac{Y}{\theta}}_T \leq 1 - \frac{a}{\theta} \right) \\ &= P_{\theta} \left( T \leq 1 - \frac{a}{\theta} \right) - P_{\theta} \left( T < 1 - \frac{b}{\theta} \right) \\ &= \left( 1 - \frac{a}{\theta} \right)^n - \left( 1 - \frac{b}{\theta} \right)^n, \end{aligned}$$

que depende de  $\theta$ , es decir, la probabilidad de cobertura del intervalo  $[Y + a, Y + b]$  depende de  $\theta$ , ver figura 6.1, y, por tanto, el coeficiente de confianza es:

$$\begin{aligned} \inf_{\theta} \left\{ \underbrace{P_{\theta}(\theta \in [L(\mathbf{X}), U(\mathbf{X})])}_{\text{Probabilidad de cobertura}} \middle| \theta \right\} &= \inf_{\theta} \left\{ P_{\theta}(\theta \in [Y + a, Y + b]) \middle| \theta \right\} \\ &= \inf_{\theta} \left\{ \left( 1 - \frac{a}{\theta} \right)^n - \left( 1 - \frac{b}{\theta} \right)^n \right\}. \end{aligned}$$

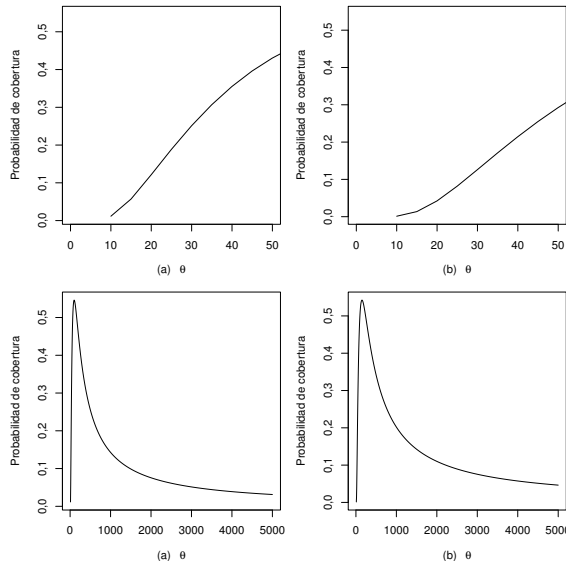


Figura 6.1. Probabilidad de cobertura versus  $\theta$ , intervalo de la forma  $[Y + a, Y + b]$ ,  $a = 2, b = 10$ , (a)  $n = 20$  y (b)  $n = 30$ , fdp uniforme  $(0, \theta)$

Tomando límite cuando  $\theta \rightarrow \infty$  se tiene que:

$$\lim_{\theta \rightarrow \infty} \left[ \left(1 - \frac{a}{\theta}\right)^n - \left(1 - \frac{b}{\theta}\right)^n \right] = 0,$$

luego el coeficiente de confianza de  $[Y + a, Y + b]$  es 0.

2. Para el segundo intervalo, la probabilidad de cobertura es:

$$\begin{aligned} P_{\theta}(\theta \in [cY, dY] | \theta) &= P_{\theta}(cY \leq \theta \leq dY) \underset{Y > 0}{=} P_{\theta}\left(c \leq \frac{\theta}{Y} \leq d\right) \\ &= P_{\theta}\left(\frac{1}{d} \leq \frac{Y}{\theta} \leq \frac{1}{c}\right) = P_{\theta}\left(\frac{1}{d} \leq T \leq \frac{1}{c}\right) \\ &= P_{\theta}\left(T \leq \frac{1}{c}\right) - P_{\theta}\left(T < \frac{1}{d}\right) = \left[\frac{1}{c}\right]^n - \left[\frac{1}{d}\right]^n, \end{aligned}$$

aquí se usó la expresión (3.12), nótese que la probabilidad de cobertura del intervalo  $[cY, dY]$  no depende de  $\theta$  y, por tanto, el coeficiente de confianza es:

$$\inf_{\theta} \left\{ \underbrace{P_{\theta}(\theta \in [L(X), U(X)] | \theta)}_{\text{Probabilidad de cobertura}} \right\} = \inf_{\theta} \left\{ P_{\theta}(\theta \in [cY, dY] | \theta) \right\}$$

$$= \inf_{\theta} \left\{ \left( \frac{1}{c} \right)^n - \left( \frac{1}{d} \right)^n \right\} = \left( \frac{1}{c} \right)^n - \left( \frac{1}{d} \right)^n.$$

Así, el coeficiente de confianza del intervalo  $[cY, dY]$  es  $c^{-n} - d^{-n}$ .  $\square$

## 6.2. Intervalos de confianza basados en una única estadística

Se pueden construir muchos intervalos de confianza en términos de un parámetro unidimensional  $\theta$  y una estadística unidimensional  $T(\mathbf{X})$ , que depende de un vector de observaciones  $\mathbf{X}$ , como en el ejemplo 6.5.

**DEFINICIÓN 6.4.** Una fda  $F_X(x)$  es *estocásticamente mayor* que una fda  $F_Y(y)$  si  $F_X(t) \leq F_Y(t)$  para todo  $t$  y  $F_X(t) < F_Y(t)$  para algún  $t$  (ver [20, p. 44]).

**DEFINICIÓN 6.5.** Una familia de fda  $\{F_X(x; \theta), \theta \in \Theta\}$  es *estocásticamente decreciente* en  $\theta$  si  $\theta_1 < \theta_2$ , esto implica que  $F_X(x; \theta_1) \leq F_X(x; \theta_2)$  para todo  $x$ , es decir,  $F_X(x; \theta_1)$  es estocásticamente mayor que  $F_X(x; \theta_2)$  (ver [20, p. 134]).

**DEFINICIÓN 6.6.** Una familia de fda  $\{F_X(x; \theta), \theta \in \Theta\}$  es *estocásticamente creciente* en  $\theta$  si  $\theta_1 < \theta_2$ , lo que implica que  $F_X(x; \theta_1) \geq F_X(x; \theta_2)$  para todo  $x$ , es decir,  $F_X(x; \theta_2)$  es estocásticamente mayor que  $F_X(x; \theta_1)$  (ver [20, p. 134]).

**Ejemplo 6.6.** La familia de fda  $\{F_X(x; \theta), \theta \in \Theta\}$ , donde  $X \sim N(\mu, \sigma^2)$  es estocásticamente creciente en  $\mu$  para  $\sigma^2$  fijo.

*Solución.* En este caso, la fdp y fda están dadas por:

$$\varphi_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\} \quad x \in \mathbb{R}.$$

$$\Phi_X(x; \mu, \sigma) = \int_{-\infty}^x \varphi_X(u; \mu, \sigma) du = \Phi_Z \left( \frac{x - \mu}{\sigma} \right) \quad Z \sim N(0, 1).$$

Sea  $\mu_1 < \mu_2$ , suponga que:

$$2x < \mu_1 + \mu_2 \quad \Leftrightarrow \quad x - \mu_1 < \mu_2 - x \quad \Leftrightarrow$$

$$\left( \frac{x - \mu_1}{\sigma} \right)^2 < \left( -\frac{x - \mu_2}{\sigma} \right)^2 \quad \Leftrightarrow \quad -\frac{1}{2} \left( \frac{x - \mu_1}{\sigma} \right)^2 > -\frac{1}{2} \left( \frac{x - \mu_2}{\sigma} \right)^2,$$

puesto que la función exponencial es una función creciente:

$$\exp \left\{ -\frac{1}{2} \left( \frac{x - \mu_1}{\sigma} \right)^2 \right\} > \exp \left\{ -\frac{1}{2} \left( \frac{x - \mu_2}{\sigma} \right)^2 \right\}$$

$$\varphi_X(x; \mu_1, \sigma) > \varphi_X(x; \mu_2, \sigma)$$

$$\Phi_X(x; \mu_1, \sigma) > \Phi_X(x; \mu_2, \sigma).$$

La última desigualdad es consecuencia de las propiedades de comparación de la integral impropia. Ver figura 6.2.

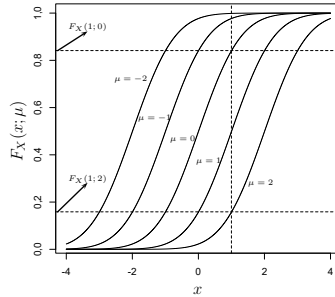


Figura 6.2. Familia estocásticamente creciente. fda  $X \sim N(\mu, 1)$  considerando diferentes valores de  $\mu$

✓

**Ejemplo 6.7.** La familia de fda  $\{F_X(x; \theta), \theta \in \Theta\}$ , donde  $X \sim Exp(\lambda)$  es estocásticamente decreciente en  $\lambda$ .

*Solución.* En este caso, para  $x > 0$  se tiene que:

$$f_X(x|\lambda) = \lambda e^{-\lambda x}, \quad \text{y} \quad F_X(x|\lambda) = 1 - e^{-\lambda x}.$$

Si  $\lambda_1 < \lambda_2$ , como  $x > 0$ ,

$$x\lambda_1 < x\lambda_2 \quad \Leftrightarrow \quad -x\lambda_1 > -x\lambda_2,$$

y como la función exponencial es estrictamente creciente:

$$e^{-x\lambda_1} > e^{-x\lambda_2} \quad \Leftrightarrow \quad 1 - e^{-x\lambda_1} < 1 - e^{-x\lambda_2}$$

$$F_X(x|\lambda_1) < F_X(x|\lambda_2).$$

Esto se ilustra en la figura 6.3.

**DEFINICIÓN 6.7.** Según [20, p. 432], una familia de fda  $\{G_T(t; \theta), \theta \in \Theta\}$  es estocásticamente creciente (decreciente) en  $\theta$  si para cada  $t \in \mathcal{T}$  el espacio muestral de la estadística  $T$  para  $\theta$ ,  $G_T(t; \theta)$  es una función decreciente (creciente) de  $\theta$  (ver figuras 6.4 y 6.5).

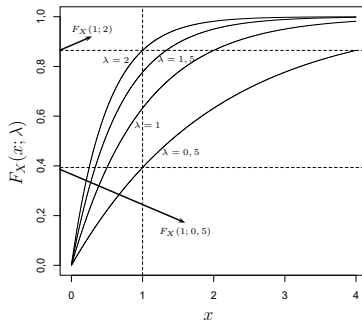


Figura 6.3. Familia estocásticamente decreciente. fda  $X \sim Exp(\lambda)$  considerando diferentes valores de  $\lambda$  ☑

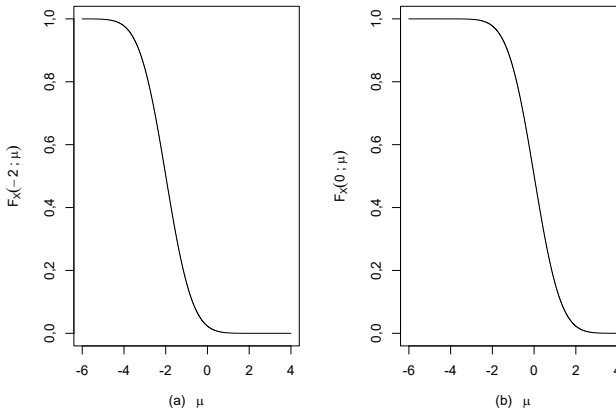


Figura 6.4. fda versus  $\mu$ ,  $X \sim N(\mu, 1)$ , para  $x = -2$  en (a) y  $x = 0$  en (b). Familia estocásticamente creciente

**Teorema 6.1.** De acuerdo con [20, p. 432], dada una estadística  $T(X)$  con fda continua,  $G_T(t; \theta)$ . Sean  $\alpha_1 > 0$  y  $\alpha_2 > 0$  valores fijos tales que  $\alpha_1 + \alpha_2 = \alpha$ , con  $0 < \alpha < 1$ . Suponga que para cada  $t \in \mathcal{T}$  las funciones  $\theta_L(t)$  y  $\theta_U(t)$  pueden ser definidas como sigue:

1. Si  $G_T(t; \theta)$  es una fda creciente de  $\theta$  para cada  $t$  (ver figura 6.6), se definen  $\theta_L(t)$  y  $\theta_U(t)$  por:

$$P[T(X) \leq t \mid \theta_L(t)] = G_T(t; \theta_L(t)) = \alpha_1 \tag{6.1}$$

$$P[T(X) \geq t \mid \theta_U(t)] = 1 - G_T(t; \theta_U(t)) = \alpha_2. \tag{6.2}$$

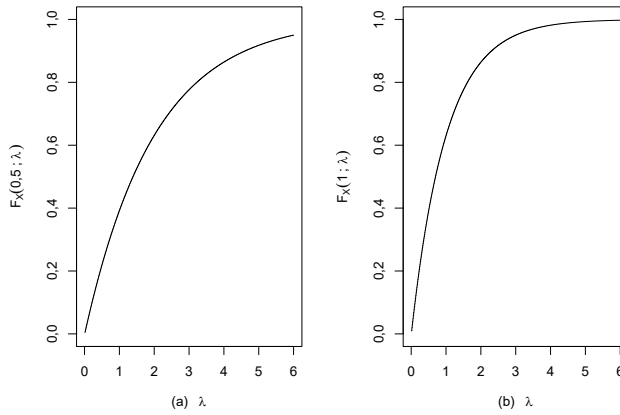


Figura 6.5. fda versus  $\lambda$ ,  $X \sim Exp(\lambda)$ , para  $x = 0,5$  en (a) y  $x = 1$  en (b). Familia estocásticamente decreciente

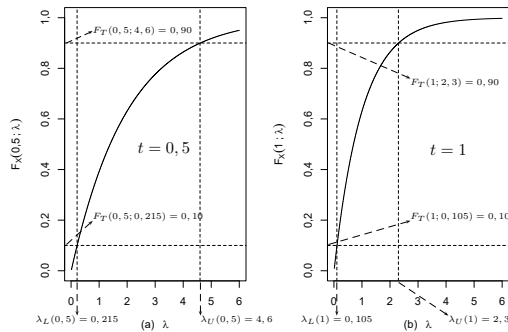


Figura 6.6.  $\lambda_L(t)$  y  $\lambda_U(t)$  de una fda  $T \sim Exp(\lambda)$  para  $\alpha_1 = \alpha_2 = 0,10$ , con  $t = 0,5$  en (a) y  $t = 1$  en (b) Familia estocásticamente decreciente

2. Si  $G_T(t; \theta)$  es una fda decreciente de  $\theta$  para cada  $t$  (ver figura 6.7), se definen  $\theta_L(t)$  y  $\theta_U(t)$  por:

$$P[T(X) \geq t \mid \theta_L(t)] = 1 - G_T(t; \theta_L(t)) = \alpha_2 \quad (6.3)$$

$$P[T(X) \leq t \mid \theta_U(t)] = G_T(t; \theta_U(t)) = \alpha_1. \quad (6.4)$$

Entonces, el intervalo aleatorio  $[\theta_L(T), \theta_U(T)]$  es un intervalo de confianza de tamaño  $1 - \alpha$  para  $\theta$ .

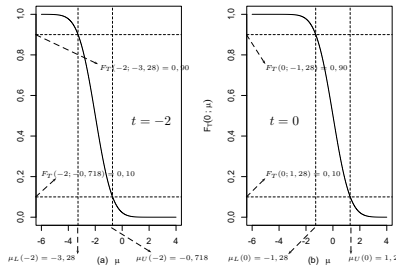


Figura 6.7.  $\mu_L(t)$  y  $\mu_U(t)$  de una fda  $T \sim N(\mu, 1)$  para  $\alpha_1 = \alpha_2 = 0, 10$ , con  $t = -2$  en (a) y  $t = 0$  en (b). Familia estocásticamente creciente

*Demostración.* En este caso, se tiene que:

$$0 < \alpha_1 + \alpha_2 < 1 \quad \implies \quad \alpha_1 < 1 - \alpha_2.$$

Sea la región de no rechazo (ver figura 6.8):

$$\{t : \alpha_1 \leq G_T(t; \theta_0) \leq 1 - \alpha_2\}. \tag{6.5}$$

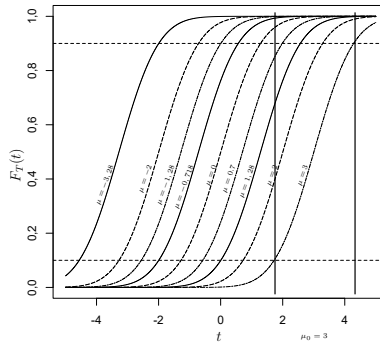


Figura 6.8. Región de no-rechazo para una fda  $X \sim N(\mu, 1)$  para  $\alpha_1 = \alpha_2 = 0, 10$

1. Como para cada  $t$ ,  $G_T(t; \theta)$  es una fda continua y creciente de  $\theta$ , entonces su inversa es continua y creciente (ver [5, p. 180]). Usando las expresiones (6.1) y (6.2), se tiene que:

$$\begin{aligned} \{\theta : \alpha_1 \leq G_T(t; \theta) \leq 1 - \alpha_2\} &= \{\theta : G_T^{-1}(\alpha_1) \leq r(\theta) \leq G_T^{-1}(1 - \alpha_2)\} \\ &= \{\theta : \theta_L(T) \leq \theta \leq \theta_U(T)\}. \end{aligned}$$

Note que los valores para  $\theta_L(t)$  y  $\theta_U(t)$  son únicos.

2. Como para cada  $t$ ,  $G_T(t; \theta)$  es una fda continua y decreciente de  $\theta$ , entonces su inversa es continua y decreciente (ver [5, p. 181]). Usando las expresiones (6.3) y (6.4), se tiene que:

$$\begin{aligned} \{\theta : \alpha_1 \leq G_T(t; \theta) \leq 1 - \alpha_2\} &= \{\theta : G_T^{-1}(1 - \alpha_2) \leq r(\theta) \leq G_T^{-1}(\alpha_1)\} \\ &= \{\theta : \theta_L(T) \leq \theta \leq \theta_U(T)\}. \end{aligned}$$

Note que los valores para  $\theta_L(t)$  y  $\theta_U(t)$  son únicos, ver figura 6.9.

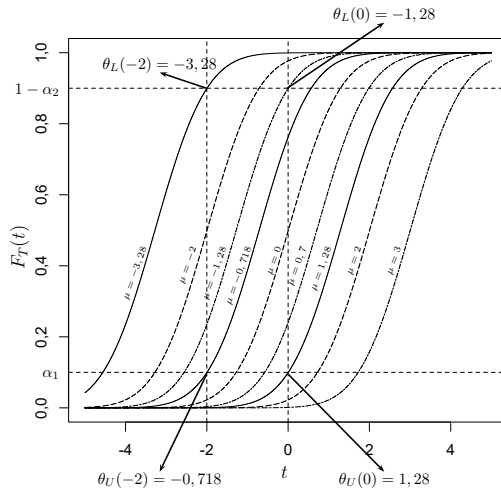


Figura 6.9.  $\theta_L(t)$  y  $\theta_U(t)$  para una fda  $X \sim N(\mu, 1)$  para  $\alpha_1 = \alpha_2 = 0,10$  ☑

**Nota 6.2.** Las ecuaciones (6.3) y (6.4) se pueden expresar como:

$$\int_{-\infty}^t g_T(u; \theta_U(t)) du = \alpha_1, \tag{6.6}$$

$$\int_t^{\infty} g_T(u; \theta_L(t)) du = \alpha_2. \tag{6.7}$$

**Nota 6.3.** Sin información adicional, es común seleccionar  $\alpha_1 = \alpha_2 = \alpha/2$ .

**Ejemplo 6.8.** De acuerdo con [20, p. 433], si  $X_1, X_2, \dots, X_n$  son variables aleatorias iid exponenciales de parámetro  $\lambda = 1$  y truncadas por la izquierda en  $\theta$ , entonces,

$$f_{X_1}(x; \theta) = \frac{e^{-x}}{\int_{\theta}^{\infty} e^{-x} dx} = e^{-(x-\theta)} I_{[\theta, +\infty)}(x).$$

Determine un intervalo de confianza para  $\theta$  de tamaño  $1 - \alpha_1 - \alpha_2$ .

*Solución.* En este caso, la fda de  $X_1$  es  $F_{X_1}(x; \theta) = 1 - e^{-(x-\theta)}$ . Puesto que  $Y = T(\mathbf{X}) = \min\{X_1, \dots, X_n\}$  es una estadística suficiente para  $\theta$ , por la expresión (3.6) la variable aleatoria  $Y$  tiene fdp dada por:

$$\begin{aligned} f_Y(y; \theta) &= ne^{-(y-\theta)} \left[ 1 - \left( 1 - e^{-(y-\theta)} \right) \right]^{n-1} I_{[\theta, +\infty)}(y) \\ &= ne^{-(y-\theta)} \left[ e^{-(y-\theta)} \right]^{n-1} I_{[\theta, +\infty)}(y) = ne^{-n(y-\theta)} I_{[\theta, +\infty)}(y), \end{aligned} \quad (6.8)$$

luego,

$$F_Y(y; \theta) = \int_{\theta}^y ne^{-n(u-\theta)} du = -e^{-n(u-\theta)} \Big|_{\theta}^y = 1 - e^{-n(y-\theta)}. \quad (6.9)$$

Puesto que  $0 < \alpha_1 + \alpha_2 < 1$ , la región se establece a partir de (6.5) como:

$$\begin{aligned} \alpha_1 \leq F_Y(y; \theta) \leq 1 - \alpha_2 &\Leftrightarrow \alpha_1 \leq 1 - e^{-n(y-\theta)} \leq 1 - \alpha_2 \\ \alpha_1 - 1 \leq -e^{-n(y-\theta)} \leq -\alpha_2 &\Leftrightarrow \alpha_2 \leq e^{-n(y-\theta)} \leq 1 - \alpha_1, \end{aligned}$$

puesto que la función logaritmo natural es una función creciente:

$$\begin{aligned} \ln(\alpha_2) \leq -n(y-\theta) \leq \ln(1-\alpha_1) \\ ny + \ln(\alpha_2) \leq n\theta \leq ny + \ln(1-\alpha_1). \end{aligned}$$

Por lo tanto, dividiendo por  $n$  y como  $Y = X_{(1)}$ , el conjunto de confianza

$$C(Y) = \left\{ \theta : X_{(1)} + \frac{1}{n} \ln(\alpha_2) \leq \theta \leq X_{(1)} + \frac{1}{n} \ln(1-\alpha_1) \right\} \quad (6.10)$$

es un intervalo de confianza de tamaño  $1 - (\alpha_1 + \alpha_2)$  para  $\theta$ .  $\square$

**Teorema 6.2.** De acuerdo con [20, p. 434], dada una estadística discreta  $T(\mathbf{X})$  con fda  $G_T(t; \theta) = P(T \leq t; \theta)$ . Sea  $\alpha_1 + \alpha_2 = \alpha$  con  $0 < \alpha < 1$  valores fijos. Suponga que para cada  $t \in \mathcal{T}$ ,  $\theta_L(t)$  y  $\theta_U(t)$  se pueden definir como:

1. Si  $G_T(t; \theta)$  es una fda decreciente de  $\theta$  para cada  $t$  (ver figura 6.10), se define  $\theta_L(t)$  y  $\theta_U(t)$  por:

$$P(T \geq t; \theta_L(t)) = \alpha_2 \quad y \quad P(T \leq t; \theta_U(t)) = \alpha_1.$$

2. Si  $G_T(t; \theta)$  es una fda creciente de  $\theta$  para cada  $t$  (ver figura 6.11), se define  $\theta_L(t)$  y  $\theta_U(t)$  por:

$$P(T \leq t; \theta_L(t)) = \alpha_2 \quad y \quad P(T \geq t; \theta_U(t)) = \alpha_1.$$

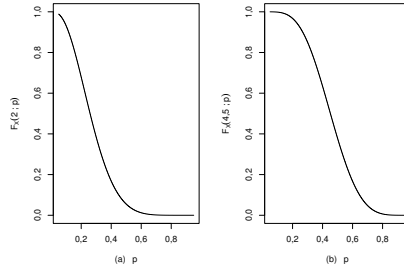


Figura 6.10. fda versus  $p$ ,  $X \sim B(10,p)$ , para  $x = 2$  en (a) y  $x = 4,5$  en (b).  
Familia estocásticamente creciente

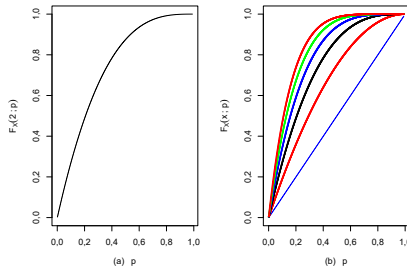


Figura 6.11. fda versus  $p$ ,  $X \sim G(p)$ , para  $x = 2$  en (a) y  $x = 6,5$  en (b).  
Familia estocásticamente decreciente

Entonces el intervalo aleatorio  $[\theta_L(t), \theta_U(t)]$  es un intervalo de confianza de tamaño  $1 - \alpha$  para  $\theta$ .

*Demostración.* Se prueba solo el ítem 1. Puesto que  $G_T(T; \theta)$  es estocásticamente mayor que una variable aleatoria con fda uniforme  $(0,1)$ , es decir,  $P_\theta(G_T(T; \theta) \leq x) \leq x$ , se tiene que (ver figura 6.12):

$$P(T \leq t; \theta_U(t)) = \alpha_1 \quad \text{y} \quad P(T \geq t; \theta_L(t)) = \alpha_2,$$

luego, la región de no rechazo es el conjunto

$$\{\theta : P(T \leq t; \theta) \leq \alpha_1 \quad \text{y} \quad P(T \geq t; \theta) \leq \alpha_2\},$$

que cumple con tener el nivel de confianza  $1 - \alpha$ .

Como  $G_T(t; \theta) = P(T \leq t; \theta)$  es una fda decreciente de  $\theta$  para cada  $t$  implica que  $P(T \geq t; \theta)$  es una función no-decreciente de  $\theta$  para cada  $t$ . Así,

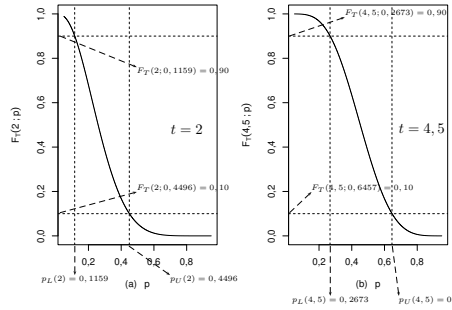


Figura 6.12. *fda* versus  $p$ ,  $X \sim B(10, p)$ , para  $x = 2$  en (a) y  $x = 4, 5$  en (b).  
Familia estocásticamente creciente

$$\begin{aligned} \theta > \theta_U(t) &\implies P(T \leq t; \theta) < \alpha_1 \\ \theta < \theta_L(t) &\implies P(T \geq t; \theta) < \alpha_2 \end{aligned}$$

y por tanto,

$$\{\theta : P(T \leq t; \theta) \leq \alpha_1 \text{ y } P(T \geq t; \theta) \leq \alpha_2\} = \{\theta : \theta_L(T) \leq \theta \leq \theta_U(T)\}. \quad \square$$

**Nota 6.4.** El algoritmo en R para obtener las gráficas mostradas en la figura 6.12 es:

```
remove(list=ls())
library(plotly)
library(matrixcalc)
x=as.matrix(seq(0, 10, 0.001))
media=as.matrix(c(0.05,0.1,0.1159,0.15,0.2,0.25,0.2673,0.3,0.35,
                 0.4,0.4496,0.45,0.5,0.55,0.6,0.6457,0.65,0.7,0.75,0.8,
                 0.85,0.9,0.95))
y=NULL
for(i in 1:nrow(media)){
y=cbind(y,round(pbinom(x,10,media[i,1]),digits=5))
}
puntos=c(3,4,1,5,1,1,3,1,4,5,3,4,1,5,1,3,4)
Y=vec(y)
X=c(rep(x,nrow(media)))
prob=media%x%rep(1,length(x))
distrib=data.frame(prob,X,Y)
p2 <- distrib %>%
  plot_ly(
    x = ~ X,
    y = ~ Y,
```

```

frame = ~ prob,
type = 'scatter',
mode = "lines",
showlegend = FALSE
)
p2%>% add_trace(y = ~0.10,line=list(color="red")) %>%
add_trace(x = ~2,line=list(color="green"))
#
p2%>% add_trace(y = ~0.90,line=list(color="blue")) %>%
add_trace(x = ~2,line=list(color="green"))
#
p2%>% add_trace(y = ~0.10,line=list(color="red")) %>%
add_trace(x = ~4,line=list(color="green"))
#
p2%>% add_trace(y = ~0.90,line=list(color="blue")) %>%
add_trace(x = ~4,line=list(color="green"))

```

**Nota 6.5.** En la figura 6.13 se presenta una región de no-rechazo para una fda binomial con parámetros  $n = 10$  y  $\theta_0 = p_0 = 0,75$ ,  $\alpha_1 = \alpha_2 = 0,10$ .

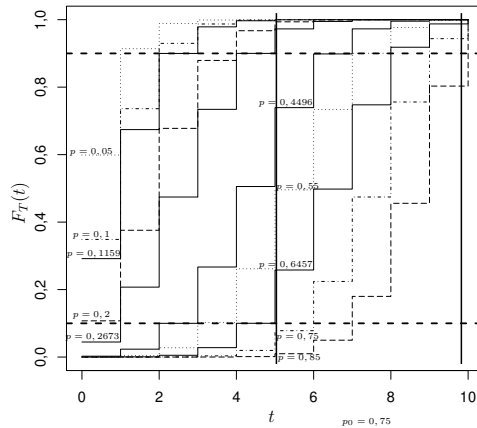


Figura 6.13. Región de no-rechazo para una fda binomial de parámetros  $n = 10$  y  $\theta_0 = p_0 = 0,75$  para  $\alpha_1 = \alpha_2 = 0,10$

**Ejemplo 6.9.** De acuerdo con [20, p. 434], dadas  $X_1, X_2, \dots, X_n$  variables aleatorias *iid* tal que  $X_1 \sim \mathcal{P}(\lambda)$ . Determine un intervalo de confianza para  $\lambda$  de tamaño  $1 - \alpha$ .

**Solución.** Se sabe que  $Y = T(X) = \sum_{i=1}^n X_i$  es una estadística suficiente para  $\lambda$ , además,  $Y \sim \mathcal{P}(n\lambda)$ . Si  $Y = y_0$  es observado, se debe hallar  $\lambda_L$  y  $\lambda_U$  tal que:

$$P(Y \geq y_0 | \lambda_L(y_0)) = \frac{\alpha}{2} \quad (6.11)$$

$$P(Y \leq y_0 | \lambda_U(y_0)) = \frac{\alpha}{2}. \quad (6.12)$$

Usando la relación entre distribución Poisson y distribución gamma<sup>37</sup> la ecuación (6.11) se puede expresar como:

$$P(Y \geq y_0 | \lambda_L(y_0)) = P(Z \leq z) = \frac{\alpha}{2},$$

donde  $Y \sim \mathcal{P}(n\lambda_L(y_0))$  y  $Z \sim \text{gamma}(\gamma, \kappa)$ , con  $\kappa z = n\lambda_L(y_0)$ . Haciendo  $\gamma = y_0$  y  $\kappa = \frac{1}{2}$ ,  $Z \sim \chi_{2y_0}^2$ . Despejando  $z$ , se tiene  $z = 2n\lambda_L(y_0)$ , luego,

$$P(Y \geq y_0 | \lambda_L(y_0)) = P(Z \leq z) = P(Z \leq 2n\lambda_L(y_0)) = \frac{\alpha}{2}$$

$$P(Y \geq y_0 | \lambda_L(y_0)) = P\left(Z \leq \chi_{2y_0, \frac{\alpha}{2}}^2\right) = \frac{\alpha}{2},$$

por tanto,

$$2n\lambda_L(y_0) = \chi_{2y_0, \frac{\alpha}{2}}^2 \implies \lambda_L(y_0) = \frac{1}{2n} \chi_{2y_0, \frac{\alpha}{2}}^2.$$

Por otro lado, la ecuación (6.12) se puede expresar como:

$$P(Y \leq y_0 | \lambda_U(y_0)) = \alpha_1 = \frac{\alpha}{2}$$

$$1 - P(Y > y_0 | \lambda_U(y_0)) = \alpha_2 = \frac{\alpha}{2}$$

$$P(Y \geq (y_0 + 1) | \lambda_U(y_0)) = P(Z \leq z) = 1 - \frac{\alpha}{2},$$

donde  $Y \sim \mathcal{P}(n\lambda_U(y_0))$  y  $Z \sim \text{gamma}(\gamma, \kappa)$ , con  $\kappa z = n\lambda_U(y_0)$ . Sea  $\gamma = y_0 + 1$  y  $\kappa = \frac{1}{2}$ ,  $Z \sim \chi_{2(y_0+1)}^2$ . Despejando  $z$ , se tiene  $z = 2n\lambda_U(y_0)$ , luego,

$$P(Y \geq (y_0 + 1) | \lambda_U(y_0)) = P(Z \leq z) = P(Z \leq 2n\lambda_U(y_0)) = 1 - \frac{\alpha}{2}$$

$$P(Y \geq (y_0 + 1) | \lambda_U(y_0)) = P\left(Z \leq \chi_{2(y_0+1), 1-\frac{\alpha}{2}}^2\right) = 1 - \frac{\alpha}{2},$$

por tanto,

$$2n\lambda_U(y_0) = \chi_{2(y_0+1), 1-\frac{\alpha}{2}}^2 \implies \lambda_U(y_0) = \frac{1}{2n} \chi_{2(y_0+1), 1-\frac{\alpha}{2}}^2.$$

<sup>37</sup>Sea  $Z$  una variable aleatoria cuya distribución es gamma( $\gamma, \kappa$ ) con fdp dada por (1.25). Si  $\gamma$  es un entero, entonces para cualquier  $z$  [20, p. 100]:

$$P(Z \leq z) = P(Y \geq \gamma),$$

donde  $Y$  es una variable aleatoria con distribución Poisson ( $\kappa z$ ), ver expresión (1.26).

De esta forma, el intervalo de confianza para  $\lambda$  al  $1 - \alpha$  es dado por:

$$\left\{ \lambda : \frac{1}{2n} \chi_{2y_0, \frac{\alpha}{2}}^2 \leq \lambda \leq \frac{1}{2n} \chi_{2(y_0+1), 1-\frac{\alpha}{2}}^2 \right\},$$

donde  $P(Z \leq \chi_{r, \alpha}^2) = \alpha$  con  $Z \sim \chi_r^2$ . Cabe aclarar que si  $y_0 = 0$  se define  $\chi_{0, \frac{\alpha}{2}}^2 = 0$ . ✓

### 6.3. Intervalos de confianza basados en la función de verosimilitud relativa (FVR)

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria con **fdp** conjunta  $P(\mathbf{x}; \theta)$  la cual depende de un parámetro desconocido  $\theta$ . Sabemos que la estimación de máxima verosimilitud  $\hat{\theta}$  es el valor de  $\theta$  que maximiza  $L(\theta; \mathbf{x})$ , es decir,

$$L(\theta; \mathbf{x}) < L(\hat{\theta}; \mathbf{x}) \quad \forall \theta \in \Theta.$$

Este valor, como se indicó en la nota 4.1, es el “más factible” de  $\theta$  en el sentido de que maximiza la verosimilitud de lo que se ha observado. Las factibilidades relativas de otros valores de  $\theta$  pueden examinarse comparándolos con  $\hat{\theta}$ . Los valores de  $\theta$  tales que  $L(\theta; \mathbf{x})$  es casi tan grande como  $L(\hat{\theta}; \mathbf{x})$  son bastante factibles en el sentido de que explican los datos casi también como lo hace  $\hat{\theta}$ . Los valores de  $\theta$  para los cuales  $L(\theta; \mathbf{x})$  es mucho menor que  $L(\hat{\theta}; \mathbf{x})$  son inverosímiles porque hacen que lo que se ha observado sea mucho menos verosímil que  $\hat{\theta}$ .

**DEFINICIÓN 6.8.** La **FVR** de  $\theta$  se define como la razón entre la función  $L(\theta; \mathbf{x})$  y su máximo  $L(\hat{\theta}; \mathbf{x})$ , es decir (ver [48]):

$$\lambda(\theta; \mathbf{x}) = \frac{L(\theta; \mathbf{x})}{L(\hat{\theta}; \mathbf{x})} < 1. \quad (6.13)$$

Además, el conjunto de valores de  $\theta$  para los que  $\lambda(\theta; \mathbf{x}) \geq \kappa(\alpha)$  se llama una *región de verosimilitud del  $100(1 - \alpha)\%$  para  $\theta$* .

Por lo general, la región de verosimilitud constituye un intervalo de valores reales, el cual se denomina intervalo de verosimilitud del  $100(1 - \alpha)\%$  para  $\theta$ ; luego

$$P_{\theta}(\{\theta : \lambda(\theta; \mathbf{x}) \geq \kappa(\alpha)\}) = 1 - \alpha, \quad (6.14)$$

donde  $\kappa(\alpha)$  es una función decreciente en  $\alpha$  tal que  $0 < \kappa(\alpha) < 1$ .

Usualmente, el estimador puntual  $\hat{\theta}$  existe y es único, y se puede establecer la **FVR** dada en (6.13). De manera más general,  $\lambda(\theta; \mathbf{x})$  puede definirse como la razón entre  $L(\theta; \mathbf{x})$  y el supremo (ver nota B.8) sobre todos los valores de los parámetros, es decir,

$$\lambda(\theta; \mathbf{x}) = \frac{L(\theta; \mathbf{x})}{\sup_{\theta \in \Theta} L(\theta; \mathbf{x})}. \quad (6.15)$$

Dado que este supremo es finito, la **FVR** existe y se puede usar para clasificar los valores de los parámetros de acuerdo con sus factibilidades incluso cuando el estimador  $\hat{\theta}$  no exista.

**Ejemplo 6.10.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid

$$f_{X_1}(x_1; \theta) = e^{-(x_1 - \theta)} I_{[\theta, +\infty)}(x_1),$$

donde  $\theta \in \mathbb{R}$ . Usando la **FVR** determine un intervalo de confianza de tamaño  $1 - \alpha$ .

*Solución.* La función de verosimilitud es dada por:

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n e^{-(x_i - \theta)} I_{[\theta, +\infty)}(x_i).$$

Usando la nota a pie de página 23, se tiene que:

$$\begin{aligned} L(\theta; \mathbf{x}) &= \exp \left\{ - \sum_{i=1}^n (x_i - \theta) \right\} \underbrace{\prod_{i=1}^n I_{[\theta, +\infty)}(x_i)}_{I\left(\bigcap_{i=1}^n \{x_i \in [\theta, +\infty)\}\right)} \\ &= \exp \left\{ - \sum_{i=1}^n (x_i - \theta) \right\} I_{[\theta, +\infty)}(x_{(1)}) = \exp \{n(\theta - \bar{x})\} I_{[\theta, +\infty)}(x_{(1)}) \\ &= \exp \{n(\theta - \bar{x})\} I(\theta \leq x_{(1)} < \infty) = \exp \{n(\theta - \bar{x})\} I_{(-\infty, x_{(1)})}(\theta). \end{aligned}$$

Usando la definición de la función indicadora en el intervalo  $(-\infty, x_{(1)})$  se obtiene

$$L(\theta; \mathbf{x}) = \begin{cases} \exp \{n(\theta - \bar{x})\} & \text{si } \theta \leq x_{(1)} \\ 0 & \text{si } \theta > x_{(1)}, \end{cases}$$

nótese que  $L(\theta; \mathbf{x})$  no es diferenciable en  $\theta = x_{(1)}$ , sin embargo, por el ejemplo 6.8, se sabe que  $L(\theta; \mathbf{x})$  alcanza su máximo en  $\hat{\theta} = x_{(1)}$ , luego,

$$L(\hat{\theta}; \mathbf{x}) = \exp \{n(x_{(1)} - \bar{x})\}.$$

Por lo tanto, la FVR se establece por (6.13) y queda

$$\begin{aligned} \lambda(\theta; \mathbf{x}) &= \frac{L(\theta; \mathbf{x})}{L(\hat{\theta}; \mathbf{x})} = \begin{cases} \frac{\exp\{n(\theta - \bar{x})\}}{\exp\{n(x_{(1)} - \bar{x})\}} & \text{si } \theta \leq x_{(1)} \\ 0 & \text{si } \theta > x_{(1)} \end{cases} \\ &= \begin{cases} e^{n(\theta - x_{(1)})} & \text{si } \theta \leq x_{(1)} \\ 0 & \text{si } \theta > x_{(1)}. \end{cases} \end{aligned}$$

Para  $\theta \leq x_{(1)}$ , usando (6.14) se establece el intervalo como:

$$\begin{aligned} 1 - \alpha &= P_{\theta}(\lambda(\theta; \mathbf{x}) \geq \kappa(\alpha)) = P_{\theta}(\exp\{n(\theta - x_{(1)})\} \geq \kappa(\alpha)) \\ &= P_{\theta}(n(\theta - x_{(1)}) \geq \ln(\kappa(\alpha))) = P_{\theta}(n\theta - \ln(\kappa(\alpha)) \geq nx_{(1)}) \\ &= P_{\theta}\left(\theta \geq x_{(1)} + \frac{\ln(\kappa(\alpha))}{n}\right). \end{aligned} \quad (6.16)$$

Usando la expresión (6.9) con  $Y = X_{(1)}$  se tiene que:

$$F_Y\left(\theta - \frac{\ln(\kappa(\alpha))}{n}; \theta\right) = 1 - e^{-n\left(\theta - \frac{\ln(\kappa(\alpha))}{n} - \theta\right)} = 1 - \kappa(\alpha),$$

igualando con (6.16), se tiene que  $\kappa(\alpha) = \alpha$  y el conjunto de confianza

$$C(X) = \left\{ \theta : X_{(1)} + \frac{1}{n} \ln(\alpha) \leq \theta \leq X_{(1)} \right\} \quad (6.17)$$

es un intervalo de confianza de tamaño  $1 - \alpha$  para  $\theta$ , el cual coincide con el intervalo de confianza dado para  $\theta$  en [42]. Además, cuando  $\alpha_1 = 0$  y  $\alpha_2 = \alpha$  en (6.10), dicho conjunto de confianza coincide con (6.17).  $\square$

**Ejemplo 6.11.** Considerando la información del ejemplo 5.3 sobre una muestra aleatoria de  $X_1 \sim N(\theta, \sigma^2)$ , con  $\sigma^2$  conocida. Determine, usando la FVR, un intervalo de confianza para  $\theta$  de tamaño  $1 - \alpha$ .

*Solución.* En este caso la función de verosimilitud está dada por:

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \theta)^2}{2\sigma^2}}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ \frac{-1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2 \right\}.$$

Por otra parte, se sabe que  $\hat{\theta} = \bar{x}$ , ver ejemplo 4.7. Dado que:

$$X_1 \sim N(\theta, 1) \quad \Rightarrow \quad \bar{X} \sim N(\theta, \sigma^2/n).$$

Luego, la FVR es dada por (6.13):

$$\begin{aligned} \lambda(\theta; \mathbf{x}) &= \frac{L(\theta; \mathbf{x})}{L(\hat{\theta}; \mathbf{x})} = \frac{(2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2 \right\}}{(2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right\}} \\ &= \exp \left\{ \frac{-1}{2\sigma^2} \left( \sum_{i=1}^n (x_i - \theta)^2 - \sum_{i=1}^n (x_i - \bar{x})^2 \right) \right\} \\ &= \exp \left\{ \frac{-1}{2\sigma^2} \left( \sum_{i=1}^n x_i^2 - 2\theta n\bar{x} + n\theta^2 - \sum_{i=1}^n x_i^2 + 2n\bar{x}^2 - n\bar{x}^2 \right) \right\} \\ &= \exp \left\{ \frac{-n}{2\sigma^2} (\theta^2 - 2\theta\bar{x} + \bar{x}^2) \right\} = \exp \left\{ \frac{-n}{2\sigma^2} (\bar{x} - \theta)^2 \right\}. \end{aligned}$$

Luego, usando (6.14) se establece el intervalo como:

$$\begin{aligned} 1 - \alpha &= P_\theta (\lambda(\theta; \mathbf{x}) \geq \kappa(\alpha)) = P_\theta \left( \exp \left\{ \frac{-n}{2\sigma^2} (\bar{X} - \theta)^2 \right\} \geq \kappa(\alpha) \right) \\ &= P_\theta \left( \frac{-n}{2\sigma^2} (\bar{X} - \theta)^2 \geq \ln(\kappa(\alpha)) \right) \\ &= P_\theta \left( \left( \frac{\bar{X} - \theta}{\sigma/\sqrt{n}} \right)^2 \leq -2 \ln(\kappa(\alpha)) \right) \\ &= P_\theta \left( \left| \frac{\bar{X} - \theta}{\sigma/\sqrt{n}} \right| \leq \sqrt{-2 \ln(\kappa(\alpha))} \right) \\ &= P_\theta \left( -\sqrt{-2 \ln(\kappa(\alpha))} \leq \frac{\bar{X} - \theta}{\sigma/\sqrt{n}} \leq \sqrt{-2 \ln(\kappa(\alpha))} \right). \end{aligned} \quad (6.18)$$

Empleando la fda normal estándar:

$$\begin{aligned} 1 - \alpha &= \Phi \left( \sqrt{-2 \ln(\kappa(\alpha))} \right) - \left[ 1 - \Phi \left( \sqrt{-2 \ln(\kappa(\alpha))} \right) \right] \\ &= 2 \left[ 1 - \Phi \left( \sqrt{-2 \ln(\kappa(\alpha))} \right) \right] = \alpha. \end{aligned}$$

Despejando, se obtiene el cuantil

$$\begin{aligned}\sqrt{-2 \ln(\kappa(\alpha))} &= \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \Leftrightarrow -2 \ln(\kappa(\alpha)) = \left(z_{(1-\frac{\alpha}{2})}\right)^2 \Leftrightarrow \\ \ln(\kappa(\alpha)) &= \frac{-1}{2} z_{(1-\frac{\alpha}{2})}^2 \Leftrightarrow \kappa(\alpha) = \exp\left\{-\frac{1}{2} z_{(1-\frac{\alpha}{2})}^2\right\}.\end{aligned}\quad (6.19)$$

Reemplazando (6.19) en (6.18) se obtiene:

$$\begin{aligned}1 - \alpha &= P_{\theta}\left(-\sigma \sqrt{\frac{-2 \ln(\kappa(\alpha))}{n}} \leq \theta - \bar{X} \leq \sigma \sqrt{\frac{-2 \ln(\kappa(\alpha))}{n}}\right) \\ &= P_{\theta}\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{(1-\frac{\alpha}{2})} \leq \theta \leq \bar{X} + \frac{\sigma}{\sqrt{n}} z_{(1-\frac{\alpha}{2})}\right).\end{aligned}$$

Por lo tanto, el conjunto de confianza

$$C(\mathbf{X}) = \left\{\theta : \bar{X} - \frac{\sigma}{\sqrt{n}} z_{(1-\frac{\alpha}{2})} \leq \theta \leq \bar{X} + \frac{\sigma}{\sqrt{n}} z_{(1-\frac{\alpha}{2})}\right\} \quad (6.20)$$

es un intervalo de confianza de tamaño  $1 - \alpha$  para  $\theta$ . Nótese que los extremos inferior y superior del intervalo de confianza son:

$$L(\mathbf{X}) = \bar{X} - \frac{\sigma}{\sqrt{n}} z_{(1-\frac{\alpha}{2})} \quad \text{y} \quad U(\mathbf{X}) = \bar{X} + \frac{\sigma}{\sqrt{n}} z_{(1-\frac{\alpha}{2})}. \quad \checkmark$$

**Nota 6.6.** Supóngase que  $p(x_i; \theta)$  se distribuye según (2.40), luego la FVR es:

$$\begin{aligned}\lambda(\theta; \mathbf{x}) &= \frac{\exp\left\{c(\theta) \sum_{i=1}^n T(x_i) + nd(\theta) + \sum_{i=1}^n S(x_i)\right\}}{\exp\left\{c(\hat{\theta}) \sum_{i=1}^n T(x_i) + nd(\hat{\theta}) + \sum_{i=1}^n S(x_i)\right\}} \\ &= \exp\left\{(c(\theta) - c(\hat{\theta})) \sum_{i=1}^n T(x_i) + n(d(\theta) - d(\hat{\theta}))\right\},\end{aligned}$$

y el intervalo de verosimilitud del  $100(1 - \alpha)\%$  para  $\theta$  está dado por:

$$\begin{aligned}1 - \alpha &= P_{\theta}\left(\exp\left\{(c(\theta) - c(\hat{\theta})) \sum_{i=1}^n T(x_i) + n(d(\theta) - d(\hat{\theta}))\right\} \geq \kappa(\alpha)\right) \\ &= P_{\theta}\left((c(\theta) - c(\hat{\theta})) \sum_{i=1}^n T(x_i) + n(d(\theta) - d(\hat{\theta})) \geq \ln(\kappa(\alpha))\right).\end{aligned}\quad (6.21)$$

**Ejemplo 6.12.** Considerando la información del ejemplo 5.3 sobre una muestra aleatoria de  $X_1 \sim N(\theta, \sigma^2)$ , con  $\sigma^2$  conocida. Determine, utilizando el método descrito en la nota 6.6, un intervalo de confianza para  $\theta$ .

*Solución.* En el ejemplo 5.25 se encontró que:

$$c(\theta) = \frac{\theta}{\sigma^2}, \quad d(\theta) = -\frac{\theta^2}{2\sigma^2} \quad \text{y} \quad \sum_{i=1}^n T(x_i) = \sum_{i=1}^n x_i,$$

puesto que  $\hat{\theta} = \bar{x}$ , ver ejemplo 4.7, entonces,

$$c(\bar{x}) = \frac{\bar{x}}{\sigma^2}, \quad d(\bar{x}) = -\frac{\bar{x}^2}{2\sigma^2} \quad \text{y} \quad \sum_{i=1}^n T(x_i) = n\bar{x},$$

sustituyendo en (6.21) se obtiene que:

$$\begin{aligned} 1 - \alpha &= P_\theta \left( \left( \frac{\theta}{\sigma^2} - \frac{\bar{X}}{\sigma^2} \right) n\bar{X} + n \left( -\frac{\theta^2}{2\sigma^2} + \frac{\bar{X}^2}{2\sigma^2} \right) \geq \ln(\kappa(\alpha)) \right) \\ &= P_\theta \left( \frac{-n}{2} \left( \frac{\bar{X} - \theta}{\sigma} \right)^2 \geq \ln(\kappa(\alpha)) \right) = P_\theta \left( \left( \frac{\bar{X} - \theta}{\sigma/\sqrt{n}} \right)^2 \leq -2 \ln(\kappa(\alpha)) \right), \end{aligned}$$

nótese que este intervalo coincide con el obtenido en (6.18) y, por consiguiente, el intervalo es idéntico a (6.20).  $\square$

## 6.4. Método pivotal

Un método útil para encontrar los puntos extremos finales de los intervalos de confianza es el método pivotal, que se basa en encontrar una cantidad pivotal.

**DEFINICIÓN 6.9.** De acuerdo con [20, p. 427], una variable aleatoria real  $Q(\mathbf{X}, \theta)$  es una cantidad pivotal (o pivote) si la distribución de  $Q(\mathbf{X}, \theta)$  es independiente de todos los parámetros  $\theta$ . Así, si  $\mathbf{X} \sim F_X(x)$ , entonces  $Q(\mathbf{X}, \theta)$  tiene la misma distribución para todos los valores de  $\theta$ .

**Ejemplo 6.13.** Según [62, p. 150], si  $X_1, X_2, \dots, X_n$  son variables aleatorias iid tal que  $X_1 \sim N(\theta, \sigma^2)$ , con  $\sigma^2$  conocida. Una variable pivote para  $\theta$  es:

$$Q(\mathbf{X}, \theta) = \frac{\sqrt{n}(\bar{X} - \theta)}{\sigma} \tag{6.22}$$

ya que

$$\frac{\sqrt{n}(\bar{X} - \theta)}{\sigma} \sim N(0, 1).$$

**Ejemplo 6.14.** De acuerdo con [20, p. 430], si  $X_1, X_2, \dots, X_n$  son variables aleatorias iid tal que  $X_1 \sim N(\mu, \sigma^2)$ , con  $\sigma^2$  desconocida. Del ejemplo B.6 del apéndice B.3 se tiene que:

$$Q(X, \sigma^2) = \frac{n-1}{\sigma^2} S^2$$

es una variable pivote, porque

$$\frac{n-1}{\sigma^2} S^2 \sim \chi_{(n-1)}^2. \quad (6.23)$$

**Ejemplo 6.15.** Según [62, p. 150], si  $X_1, X_2, \dots, X_n$  son variables aleatorias iid tal que  $X_1 \sim N(\theta, \sigma^2)$ . Definiendo  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  y  $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$  la media muestral y la varianza muestral, respectivamente, entonces:

$$Q(X, \theta) = \frac{\sqrt{n}(\bar{X}_n - \theta)}{S_n} \quad (6.24)$$

es una variable pivote para  $\theta$ , ya que, ver ejemplo 2.21 y apéndice B.3,

$$Q(X, \theta) = \frac{\frac{\sqrt{n}(\bar{X}_n - \theta)}{\sigma}}{\sqrt{\frac{(n-1)S_n^2}{(n-1)\sigma^2}}} = \frac{\sqrt{n}(\bar{X}_n - \theta)}{S_n} \sim t_{(n-1)}.$$

**Nota 6.7.** Tras establecer la variable pivote y su distribución, la técnica de construcción de intervalos de confianza por el método del pivote consiste en reescribir

$$P_\theta(a < Q(X, \theta) < b) = 1 - \alpha \quad (6.25)$$

como

$$P_\theta(T_n^{(1)} < r(\theta) < T_n^{(2)}) = 1 - \alpha.$$

El intervalo aleatorio  $(T_n^{(1)}, T_n^{(2)})$  será un intervalo de confianza para  $r(\theta)$  de tamaño  $100(1 - \alpha)\%$  (ver [62, p. 151]).

**Ejemplo 6.16.** Considerando la información del ejemplo 5.3 sobre una muestra aleatoria de  $X_1 \sim N(\theta, \sigma^2)$ , con  $\sigma^2$  conocida, determine un intervalo de confianza para  $\theta$  de tamaño  $100(1 - \alpha)\%$ .

*Solución.* En el ejemplo 6.13 se estableció que:

$$Q(X, \theta) = \frac{\sqrt{n}(\bar{X}_n - \theta)}{\sigma} \sim N(0, 1),$$

luego, la expresión (6.25) queda:

$$1 - \alpha = P_\theta \left( \underbrace{a < \frac{\sqrt{n}(\bar{X}_n - \theta)}{\sigma} < b}_{Q(X, \theta)} \right) = \int_{-\infty}^b d[F_{Q_X}(x)] - \int_{-\infty}^a d[F_{Q_X}(x)],$$

por otro lado,

$$\begin{aligned} 1 - \alpha &= P_\theta(a < Q(X, \theta) < b) = P_\theta\left(\frac{\sigma a}{\sqrt{n}} < \bar{X}_n - \theta < \frac{\sigma b}{\sqrt{n}}\right) \\ &= P_\theta\left(\frac{\sigma a}{\sqrt{n}} - \bar{X}_n < -\theta < \frac{\sigma b}{\sqrt{n}} - \bar{X}_n\right) = P_\theta\left(\underbrace{\bar{X}_n - \frac{\sigma b}{\sqrt{n}}}_{T_n^{(1)}} < \underbrace{\theta}_{r(\theta)} < \underbrace{\bar{X}_n - \frac{\sigma a}{\sqrt{n}}}_{T_n^{(2)}}\right), \end{aligned}$$

luego un intervalo de confianza del  $100(1 - \alpha)\%$  para  $\theta$  es dado por:

$$IC_{(1-\alpha)}(\theta) = \left( \bar{X}_n - b \frac{\sigma}{\sqrt{n}}, \bar{X}_n - a \frac{\sigma}{\sqrt{n}} \right), \quad (6.26)$$

la longitud  $L_I$  de este intervalo es proporcional a  $\sigma$  ya que:

$$L_I = \left( \bar{X}_n - \frac{\sigma a}{\sqrt{n}} \right) - \left( \bar{X}_n - \frac{\sigma b}{\sqrt{n}} \right) = \frac{\sigma b}{\sqrt{n}} - \frac{\sigma a}{\sqrt{n}} = \frac{(b - a)}{\sqrt{n}} \sigma.$$

Para minimizar  $L_I$  sujeto a la restricción:

$$\int_{-\infty}^b d[F_{Q_X}(x)] - \int_{-\infty}^a d[F_{Q_X}(x)] = 1 - \alpha.$$

En [21, p. 95] se presentan dos formas para encontrar la minimización de  $L_I$ , a continuación, se presentará la solución por multiplicadores de Lagrange<sup>38</sup>.

*Veamos:* se quiere minimizar  $h(a, b) = \frac{\sigma}{\sqrt{n}}(b - a)$  sujeto a la restricción

$$g(a, b) = \int_{-\infty}^b d[F_{Q_X}(x)] - \int_{-\infty}^a d[F_{Q_X}(x)] - (1 - \alpha) = 0.$$

<sup>38</sup>Ver nota B.13.

Se introduce el multiplicador de Lagrange,  $\lambda$ , para formar la función auxiliar:

$$\begin{aligned} H(a,b,\lambda) &= h(a,b) - \lambda g(a,b) \\ &= \frac{\sigma}{\sqrt{n}}(b-a) - \lambda \left( \int_{-\infty}^b d[F_{Q_X}(x)] - \int_{-\infty}^a d[F_{Q_X}(x)] - (1-\alpha) \right), \end{aligned}$$

se deriva  $H(a,b,\lambda)$  con respecto a  $a, b, \lambda$ , para ello se usa el primer teorema fundamental del cálculo (ver nota B.11):

$$\begin{aligned} \frac{\partial H(a,b,\lambda)}{\partial a} &= -\frac{\sigma}{\sqrt{n}} + \lambda \frac{\partial}{\partial a} \int_{-\infty}^a d[F_{Q_X}(x)] = -\frac{\sigma}{\sqrt{n}} + \lambda f_{Q_X}(a) \\ \frac{\partial H(a,b,\lambda)}{\partial b} &= \frac{\sigma}{\sqrt{n}} - \lambda \frac{\partial}{\partial b} \int_{-\infty}^b d[F_{Q_X}(x)] = \frac{\sigma}{\sqrt{n}} - \lambda f_{Q_X}(b) \\ \frac{\partial H(a,b,\lambda)}{\partial \lambda} &= (1-\alpha) - \int_{-\infty}^b d[F_{Q_X}(x)] + \int_{-\infty}^a d[F_{Q_X}(x)], \end{aligned}$$

igualando estas expresiones a cero:

$$-\frac{\sigma}{\sqrt{n}} + \lambda f_{Q_X}(a) = 0 \quad \Leftrightarrow \quad \frac{\sigma}{\sqrt{n}} = \lambda f_{Q_X}(a) \quad (6.27)$$

$$\frac{\sigma}{\sqrt{n}} - \lambda f_{Q_X}(b) = 0 \quad \Leftrightarrow \quad \frac{\sigma}{\sqrt{n}} = \lambda f_{Q_X}(b) \quad (6.28)$$

$$(1-\alpha) - \underbrace{\int_{-\infty}^b d[F_{Q_X}(x)]}_{\Phi(b)} + \underbrace{\int_{-\infty}^a d[F_{Q_X}(x)]}_{\Phi(a)} = 0, \quad (6.29)$$

siendo  $\Phi(z) = P(Z < z)$ , la fda evaluada en  $z$ , cuando  $Z \sim N(0,1)$ . Igualando (6.27) y (6.28):

$$\lambda f_{Q_X}(a) = \lambda f_{Q_X}(b) \quad \Leftrightarrow \quad f_{Q_X}(a) = f_{Q_X}(b)$$

y dado que la fdp normal estándar es simétrica alrededor de cero, esto ocurre cuando  $a = b$  o  $a = -b$ , pero  $a = b$  implicaría que en la expresión (6.29) el término  $1 - \alpha = 0$ , lo cual no es admisible. Luego, la solución es  $a = -b$ , reemplazando en (6.29):

$$\begin{aligned} (1-\alpha) - \Phi(b) + \Phi(a) = 0 \quad \Leftrightarrow \quad \Phi(b) - \underbrace{\Phi(-b)}_{1-\Phi(b)} = 1-\alpha \quad \Leftrightarrow \\ 2\Phi(b) - 1 = 1-\alpha \quad \Leftrightarrow \quad \Phi(b) = \frac{2-\alpha}{2} = 1 - \frac{\alpha}{2}, \end{aligned}$$

luego  $b = z_{(1-\frac{\alpha}{2})}$ , y por tanto  $a = -z_{(1-\frac{\alpha}{2})}$ . Así, el intervalo de confianza del  $100(1-\alpha)\%$  para  $\theta$  dado en (6.26) de mínima longitud es:

$$\begin{aligned} IC_{(1-\alpha)}(\theta) &= \left( \bar{X}_n - z_{(1-\frac{\alpha}{2})} \frac{\sigma}{\sqrt{n}}, \bar{X}_n - \left( -z_{(1-\frac{\alpha}{2})} \right) \frac{\sigma}{\sqrt{n}} \right) \\ &= \left( \bar{X}_n - z_{(1-\frac{\alpha}{2})} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{(1-\frac{\alpha}{2})} \frac{\sigma}{\sqrt{n}} \right). \end{aligned} \quad (6.30)$$

Nótese que este intervalo coincide con el obtenido en (6.20). Reescribiendo la expresión (6.25) se tiene que:

$$P_{\theta} \left( |\bar{X}_n - \theta| < z_{(1-\frac{\alpha}{2})} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha. \quad \square$$

**Nota 6.8.** No siempre es fácil encontrar intervalos de confianza de longitud mínima, por ejemplo en el intervalo de confianza para  $\sigma^2$ , cuando la muestra de tamaño  $n$  proviene de  $X_1 \sim N(\theta, \sigma^2)$ , se tiene de (B.9) que:

$$Q(X, \sigma) = \frac{n-1}{\sigma^2} S_n^2 \sim \chi_{(n-1)}^2,$$

luego, la expresión (6.25) queda:

$$\begin{aligned} 1 - \alpha &= P_{\theta} \left( a < \underbrace{\frac{n-1}{\sigma^2} S_n^2}_{Q(X, \sigma)} < b \right) = P_{\theta} \left( \frac{a}{(n-1)S_n^2} < \frac{1}{\sigma^2} < \frac{b}{(n-1)S_n^2} \right) \\ &= P_{\theta} \left( \frac{(n-1)S_n^2}{b} < \sigma^2 < \frac{(n-1)S_n^2}{a} \right), \end{aligned}$$

por tanto un intervalo de confianza del  $100(1-\alpha)\%$  para  $\sigma^2$  es dado por:

$$IC_{(1-\alpha)}(\sigma^2) = \left( \frac{(n-1)S_n^2}{b}, \frac{(n-1)S_n^2}{a} \right),$$

la longitud  $L_I$  de este intervalo es proporcional a  $S_n^2$  ya que:

$$L_I = \frac{(n-1)S_n^2}{a} - \frac{(n-1)S_n^2}{b} = (n-1) \left( \frac{1}{a} - \frac{1}{b} \right) S_n^2.$$

Para encontrar  $a$  y  $b$  tal que la longitud  $L_I$  sea mínima es necesario emplear métodos numéricos. En este caso, es usual tomar como extremos  $a = \chi_{\frac{\alpha}{2}, n-1}^2$  y  $b = \chi_{1-\frac{\alpha}{2}, n-1}^2$ , con  $P(Y < \chi_{\delta, v}^2) = \delta$ ,  $Y \sim \chi_v^2$ , aunque estos no generan el intervalo de menor

longitud, es decir,

$$IC_{(1-\alpha)}(\sigma^2) = \left( \frac{(n-1)S_n^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2}, \frac{(n-1)S_n^2}{\chi_{\frac{\alpha}{2}, n-1}^2} \right). \quad (6.31)$$

**Ejemplo 6.17.** (Experimento de dietas en ratas). En la tabla 6.1 se presentan los pesos, en gramos, de dos grupos de ratas con diferentes dietas después de 64 días de aplicadas.

Tabla 6.1. Pesos en gramos de ratas bajo dos dietas

Dieta A:	278	245	269	275	280	281	284	278	280	277
Dieta B:	478	496	472	628	495	490	625	600		

Encuentre para cada una de las dietas:

1. El estimador puntual para la media y la varianza.
2. El estimador por intervalo para la media con un nivel de confianza del 90 %, suponiendo que las varianzas eran  $\sigma_A^2 = 180$  y  $\sigma_B^2 = 3000$ .
3. El estimador por intervalo para la media con un nivel de confianza del 95 %, usando el estimador de la varianza.
4. El estimador por intervalo para la varianza con un nivel de confianza del 95 %.

*Solución.* De la información dada en la tabla 6.1 se tiene que:

$$n_A = 10 \quad \bar{x}_A = \sum_{i=1}^{10} \frac{x_i}{10} = 274,7 \quad s_A^2 = \sum_{i=1}^{10} \frac{(x_i - \bar{x}_A)^2}{9} = 124,9$$

$$n_B = 8 \quad \bar{x}_B = \sum_{i=1}^8 \frac{x_i}{8} = 535,5 \quad s_B^2 = \sum_{i=1}^8 \frac{(x_i - \bar{x}_B)^2}{7} = 4762,2857$$

1. Los estimadores puntuales para la media y la varianza son

$$\hat{\mu}_A = \bar{x}_A = 274,7 \quad \hat{\sigma}_A^2 = s_A^2 = 124,9$$

$$\hat{\mu}_B = \bar{x}_B = 535,5 \quad \hat{\sigma}_B^2 = s_B^2 = 4762,2857$$

2. Los intervalos de confianza para la media con un nivel de confianza del 90 %, están dados por la expresión (6.30):

- Dieta A:

$$IC_{0,90}(\mu_A) = \left( 274,7 - z_{0,95} \sqrt{\frac{180}{10}}; 274,7 + z_{0,95} \sqrt{\frac{180}{10}} \right)$$

$$\begin{aligned}
 &= (274,7 - 1,645 \times \sqrt{18}; 274,7 + 1,645 \times \sqrt{18}) \\
 &= (267,7209; 281,6791).
 \end{aligned}$$

Luego, el promedio de los pesos en gramos de las ratas a los 64 días de aplicada la dieta A está entre 267,72 y 281,68 gramos con un nivel de confianza del 90 %.

■ Dieta B:

$$\begin{aligned}
 IC_{0,90}(\mu_B) &= \left( 535,5 - z_{0,95} \sqrt{\frac{3000}{8}}; 535,5 + z_{0,95} \sqrt{\frac{3000}{8}} \right) \\
 &= \left( 535,5 - 1,645 \times \sqrt{375}; 535,5 + 1,645 \times \sqrt{375} \right) \\
 &= (503,6447; 567,3553).
 \end{aligned}$$

Concluyendo que, el promedio de los pesos en gramos de las ratas a los 64 días de aplicada la dieta B está entre 503,64 y 567,36 gramos con un nivel de confianza del 90 %.

3. Los intervalos de confianza para la media con un nivel de confianza del 95 %, suponiendo que las varianzas son desconocidas, en este caso la expresión (6.30) se modifica (ver ejercicio 2) y queda:

■ Dieta A:

$$\begin{aligned}
 IC_{0,95}(\mu_A) &= \left( 274,7 - t_{0,975;9} \sqrt{\frac{124,9}{10}}; 274,7 + t_{0,975;9} \sqrt{\frac{124,9}{10}} \right) \\
 &= \left( 274,7 - 2,262 \times \frac{11,1759}{\sqrt{10}}; 274,7 + 2,262 \times \frac{11,1759}{\sqrt{10}} \right) \\
 &= (266,7058; 282,6942).
 \end{aligned}$$

Por tanto, el promedio de los pesos en gramos de las ratas a los 64 días de aplicada la dieta A está entre 266,71 y 282,69 gramos con un nivel de confianza del 95 %.

■ Dieta B:

$$\begin{aligned}
 IC_{0,95}(\mu_B) &= \left( 535,5 - t_{0,975;7} \sqrt{\frac{4762,286}{8}}; 535,5 + t_{0,975;7} \sqrt{\frac{4762,286}{8}} \right) \\
 &= \left( 535,5 - 2,365 \times \frac{69,0093}{\sqrt{8}}; 535,5 + 2,365 \times \frac{69,0093}{\sqrt{8}} \right) \\
 &= (477,7976; 593,2024).
 \end{aligned}$$

Entonces, el promedio de los pesos en gramos de las ratas a los 64 días de aplicada la dieta B está entre 477,80 y 593,20 gramos con un nivel de confianza del 95 %.

4. Los intervalos de confianza para la varianza con un nivel de confianza del 95 %, están dados por la expresión (6.31):

- Dieta A:

$$\begin{aligned} IC_{0,95}(\sigma_A^2) &= \left( \frac{(10-1)S_{An}^2}{\chi_{1-\frac{0,05}{2}, 10-1}^2}; \frac{(10-1)S_{An}^2}{\chi_{\frac{0,05}{2}, 10-1}^2} \right) \\ &= \left( \frac{(10-1)(124,9)}{19,03}; \frac{(10-1)(124,9)}{2,70} \right) \\ &= (59,07; 416,33) : \end{aligned}$$

Entonces, la varianza de los pesos en gramos de las ratas a los 64 días de aplicada la dieta A está entre 59,07 y 416,33 con un nivel de confianza del 95 %.

- Dieta B:

$$\begin{aligned} IC_{0,95}(\sigma_B^2) &= \left( \frac{(8-1)S_{Bn}^2}{\chi_{1-\frac{0,05}{2}, 8-1}^2}; \frac{(8-1)S_{Bn}^2}{\chi_{\frac{0,05}{2}, 8-1}^2} \right) \\ &= \left( \frac{(8-1)(4762,29)}{16,02}; \frac{(8-1)(4762,29)}{1,69} \right) \\ &= (2\ 080,90; 19\ 725,44). \end{aligned}$$

Entonces, la varianza de los pesos en gramos de las ratas a los 64 días de aplicada la dieta B está entre 2 080,90 y 19 725,44 con un nivel de confianza del 95 %.  $\square$

**Ejemplo 6.18.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias *iid* Bernoulli de parámetro  $p$ . Encuentre el intervalo de confianza del  $100(1 - \alpha)\%$  para  $p$ .

*Solución.* Como  $X_k$  es una variable aleatoria Bernoulli solo toma los valores dados en (3.24). Del ejemplo 3.33 se tiene que:

$$\begin{aligned} \sqrt{n} (\bar{X}_n - p) &\xrightarrow{D} N(0, p(1-p)) \\ \frac{\sqrt{n} (\bar{X}_n - p)}{\sqrt{p(1-p)}} &\xrightarrow{D} N(0, 1). \end{aligned}$$

Así, para  $n$  suficientemente grande, notando  $\bar{X}_n = \hat{p}_n$ , el intervalo de confianza del  $100(1 - \alpha)\%$  para  $p$  es dado por, ver ejemplo 6.16,

$$IC_{(1-\alpha)}(p) = \left( \bar{X}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) \\ = \left( \hat{p}_n - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} \right). \quad (6.32)$$

Este se llama intervalo de confianza de Wald para  $p$ . ☑

**Ejemplo 6.19.** Según [62, p. 169], dadas  $X_1, X_2, \dots, X_n$  variables aleatorias iid  $N(\mu_X, \sigma_X^2)$  y  $Y_1, Y_2, \dots, Y_m$  variables aleatorias iid  $N(\mu_Y, \sigma_Y^2)$ , tal que las dos muestras aleatorias son independientes entre sí, y  $\mu_X$  y  $\mu_Y$  son desconocidas. Encuentre un intervalo de confianza del  $100(1 - \alpha)\%$  para el cociente  $\frac{\sigma_Y^2}{\sigma_X^2}$ .

*Solución.* Sean  $S_{X_n}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$  y  $S_{Y_m}^2 = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y}_m)^2$ . Del ejemplo B.6 se tiene que:

$$\frac{n-1}{\sigma_X^2} S_{X_n}^2 \sim \chi_{(n-1)}^2 \quad \text{y} \quad \frac{m-1}{\sigma_Y^2} S_{Y_m}^2 \sim \chi_{(m-1)}^2.$$

Dada la independencia entre todas las variables  $X$  y  $Y$ ,  $\frac{n-1}{\sigma_X^2} S_{X_n}^2$  y  $\frac{m-1}{\sigma_Y^2} S_{Y_m}^2$  también son independientes y, por el ejemplo 2.22, el cociente:

$$\frac{\frac{n-1}{\sigma_X^2} S_{X_n}^2}{\frac{m-1}{\sigma_Y^2} S_{Y_m}^2} \sim F_{n-1, m-1} \quad \implies \quad \frac{\sigma_Y^2 S_{X_n}^2}{\sigma_X^2 S_{Y_m}^2} \sim F_{n-1, m-1},$$

luego, la variable pivote sería:

$$Q_{X,Y} = \frac{\sigma_Y^2 S_{X_n}^2}{\sigma_X^2 S_{Y_m}^2}.$$

Para encontrar un intervalo de confianza para el cociente  $\frac{\sigma_Y^2}{\sigma_X^2}$ , se parte de

$$1 - \alpha = P_\theta \left( a < \underbrace{\frac{\sigma_Y^2 S_{X_n}^2}{\sigma_X^2 S_{Y_m}^2}}_{Q_{X,Y}} < b \right) = P_\theta \left( \frac{aS_{Y_m}^2}{S_{X_n}^2} < \frac{\sigma_Y^2}{\sigma_X^2} < \frac{bS_{Y_m}^2}{S_{X_n}^2} \right),$$

por tanto un intervalo de confianza del  $100(1 - \alpha) \%$  para  $\frac{\sigma_Y^2}{\sigma_X^2}$  es dado por:

$$IC_{(1-\alpha)} \left( \frac{\sigma_Y^2}{\sigma_X^2} \right) = \left( \frac{aS_{Ym}^2}{S_{Xn}^2}, \frac{bS_{Ym}^2}{S_{Xn}^2} \right),$$

la longitud  $L_I$  de este intervalo es proporcional a la razón  $S_{Ym}^2/S_{Xn}^2$  ya que

$$L_I = \frac{bS_{Ym}^2}{S_{Xn}^2} - \frac{aS_{Ym}^2}{S_{Xn}^2} = (b - a) \frac{S_{Ym}^2}{S_{Xn}^2}.$$

Para encontrar  $a$  y  $b$  tal que la longitud  $L_I$  sea mínima es necesario emplear métodos numéricos. En este caso, integrando la **fdp** dada en (2.62) correspondiente a  $U \sim F_{n-1, m-1}$ , se encuentra  $P(U < F_{\delta, n-1, m-1}) = \delta$ , pero por la propiedad 1 del teorema 2.20 se tiene que  $F_{\frac{\alpha}{2}, n-1, m-1} = (F_{1-\frac{\alpha}{2}, m-1, n-1})^{-1}$ . Luego, es usual tomar  $a = F_{\frac{\alpha}{2}, n-1, m-1}$  y  $b = F_{1-\frac{\alpha}{2}, n-1, m-1}$ , aunque no generan el intervalo de menor longitud, es decir,

$$IC_{(1-\alpha)} \left( \frac{\sigma_Y^2}{\sigma_X^2} \right) = \left( F_{\frac{\alpha}{2}, n-1, m-1} \frac{S_{Ym}^2}{S_{Xn}^2}, F_{1-\frac{\alpha}{2}, n-1, m-1} \frac{S_{Ym}^2}{S_{Xn}^2} \right), \quad (6.33)$$

como sugerencia se colocará la varianza más grande en el numerador.  $\checkmark$

## 6.5. Algunos casos particulares

### 6.5.1. Intervalo de confianza para una proporción

Considerando la información del ejemplo 6.18, sean  $X_1, X_2, \dots, X_n$  variables aleatorias e **iid**  $Ber(p)$ , una alternativa para el intervalo de confianza para  $p$  dado en (6.32) es propuesta por [2] (ver [21, p. 103]),

$$IC_{(1-\alpha)}(p) = \left( \tilde{\theta} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{n+4}}, \tilde{\theta} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{n+4}} \right), \quad (6.34)$$

$$\text{con } \tilde{\theta} = \frac{\sum_{i=1}^n X_i + 2}{n+4}.$$

**Ejemplo 6.20.** Según [50, p. 80], una firma productora de cigarrillos asegura que su marca  $A$  de cigarrillos sobrepasa en ventas a su marca  $B$  en 8%. Si se encuentra

que 42 de 200 fumadores prefieren la marca  $A$  y 18 de 150 fumadores la  $B$ , calcule para cada marca de cigarrillos:

1. La estimación puntual de la proporción de preferencia de los fumadores.
2. La estimación por intervalo de la proporción de preferencia de los fumadores con un 98% de nivel de confianza.

*Solución.* Se tiene del enunciado que:

Marca	$n_k$	$x_k = \sum_{i=1}^{n_k} x_i$
A	200	42
B	150	18

1. En este caso,

$$\hat{p}_A = \frac{x_A}{n_A} = \frac{42}{200} = 0,21 \quad \hat{p}_B = \frac{x_B}{n_B} = \frac{18}{150} = 0,12.$$

Por tanto, la proporción de fumadores que prefieren la marca de cigarrillos  $A$  es 21% y la proporción de fumadores que prefieren la marca de cigarrillos  $B$  es 12%.

2. Los intervalos de confianza para la proporción establecidos por (6.34). Luego, se tiene del enunciado que:

Marca	$A$	$B$
$\tilde{\theta}_k = \frac{x_k + 2}{n_k + 4}$	$\frac{42 + 2}{200 + 4} = 0,2157$	$\frac{18 + 2}{150 + 4} = 0,1299$

Estas nuevas proporciones son superiores a las iniciales.

- Marca  $A$ :

$$\begin{aligned} IC_{0,98}(p_A) &= \left( \tilde{\theta}_A - z_{0,99} \sqrt{\frac{\tilde{\theta}_A(1 - \tilde{\theta}_A)}{n_A + 4}}, \tilde{\theta}_A + z_{0,99} \sqrt{\frac{\tilde{\theta}_A(1 - \tilde{\theta}_A)}{n_A + 4}} \right) \\ &= \left( \frac{11}{51} - \frac{2,325}{\sqrt{204}} \sqrt{\frac{11}{51} \times \frac{40}{51}}, \frac{11}{51} + \frac{2,325}{\sqrt{204}} \sqrt{\frac{11}{51} \times \frac{40}{51}} \right) \\ &= (0,1487; 0,2826). \end{aligned}$$

Entonces, la proporción de fumadores que eligen la marca de cigarrillos  $A$  está entre 14,9% y el 28,3% con una confianza del 98%.

- Marca  $B$ :

$$\begin{aligned}
 & IC_{0,98}(p_B) \\
 &= \left( \tilde{\theta}_B - z_{0,99} \sqrt{\frac{\tilde{\theta}_B(1-\tilde{\theta}_B)}{n_B+4}}, \tilde{\theta}_B + z_{0,99} \sqrt{\frac{\tilde{\theta}_B(1-\tilde{\theta}_B)}{n_B+4}} \right) \\
 &= \left( \frac{10}{77} - \frac{2,325}{\sqrt{154}} \sqrt{\frac{10}{77} \times \frac{67}{77}}, \frac{10}{77} + \frac{2,325}{\sqrt{154}} \sqrt{\frac{10}{77} \times \frac{67}{77}} \right) \\
 &= (0,0669; 0,1929).
 \end{aligned}$$

Luego, la proporción de fumadores que eligen la marca de cigarrillos  $B$  está entre 6,7% y el 19,3% con una confianza del 98%.  $\checkmark$

## 6.5.2. Intervalo de confianza para diferencia de medias de dos poblaciones independientes

Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid tal que  $X_1 \sim N(\mu_X, \sigma_X^2)$  y  $Y_1, Y_2, \dots, Y_m$  variables aleatorias iid tal que  $Y_1 \sim N(\mu_Y, \sigma_Y^2)$ , donde las muestras aleatorias son independientes entre sí. Nos interesa estimar la diferencia de sus medias poblacionales. Se consideran tres casos: varianzas conocidas, varianzas desconocidas pero iguales a nivel poblacional y varianzas desconocidas pero distintas a nivel poblacional.

### 6.5.2.1. Caso de varianzas conocidas

Cuando las varianzas poblacionales se conocen se tiene que:

$$\begin{aligned}
 \bar{X}_n &\sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right) & \text{y} & & \bar{Y}_m &\sim N\left(\mu_Y, \frac{\sigma_Y^2}{m}\right) \\
 \bar{X}_n - \bar{Y}_m &\sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right), & & & & (6.35)
 \end{aligned}$$

luego la variable pivotal puede ser dada por:

$$\frac{(\bar{X}_n - \bar{Y}_m) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0,1).$$

Usando el resultado del ejemplo 6.16 se tiene que:

$$\begin{aligned}
 IC_{(1-\alpha)}(\mu_X - \mu_Y) &= (\bar{X}_n - \bar{Y}_m - z_{1-\frac{\alpha}{2}} \sigma_{\bar{X}-\bar{Y}}; \bar{X}_n - \bar{Y}_m + z_{1-\frac{\alpha}{2}} \sigma_{\bar{X}-\bar{Y}}) \\
 &= \left( \bar{X}_n - \bar{Y}_m - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}; \bar{X}_n - \bar{Y}_m + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \right). \quad (6.36)
 \end{aligned}$$

**Ejemplo 6.21.** Retomando el ejemplo 6.17, sobre las dietas en ratas. Encontrar el estimador por intervalo para la diferencia media de los pesos (en gramos) de las dos dietas al 90 % de nivel de confianza, suponga que  $\sigma_A^2 = 180$  y  $\sigma_B^2 = 3000$ .

*Solución.* Se sabe que:

$$\begin{aligned}
 \hat{\mu}_A = \bar{x}_A &= 274,7 & n_A &= 10 \\
 \hat{\mu}_B = \bar{x}_B &= 535,5 & n_B &= 8,
 \end{aligned}$$

luego, el intervalo de confianza para la diferencia de medias con un nivel de confianza del 90 %, está dado por la expresión (6.36):

$$\begin{aligned}
 &IC_{0,9}(\mu_A - \mu_B) \\
 &= \left( 274,7 - 535,5 - 1,645 \sqrt{\frac{180}{10} + \frac{3000}{8}}; \right. \\
 &\qquad \qquad \qquad \left. 274,7 - 535,5 + 1,645 \sqrt{\frac{180}{10} + \frac{3000}{8}} \right) \\
 &= (-293,4; -228,2)
 \end{aligned}$$

o

$$IC_{0,9}(\mu_B - \mu_A) = (228,2; 293,4).$$

Por tanto, la diferencia media en el peso (en gramos),  $\mu_B - \mu_A$ , de las ratas a los 64 días de aplicadas las dietas A y B está entre 228,2 y 293,4 gramos con un nivel de confianza del 90 %.  $\square$

### 6.5.2.2. Caso de varianzas desconocidas que se pueden considerar iguales a nivel poblacional

Cuando las varianzas poblacionales no se conocen, pero se puede suponer que  $\sigma_X^2 = \sigma_Y^2 = \sigma^2$  se tiene que:

$$\bar{X}_n \sim N\left(\mu_X, \frac{\sigma^2}{n}\right) \quad \text{y} \quad \bar{Y}_m \sim N\left(\mu_Y, \frac{\sigma^2}{m}\right).$$

Luego, la expresión (6.35) queda:

$$\bar{X}_n - \bar{Y}_m \sim N\left(\mu_X - \mu_Y, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}\right).$$

En este caso, la variable pivotal es dada por:

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1).$$

Por la expresión (6.23) se tiene que:

$$\frac{n-1}{\sigma^2} S_{X_n}^2 \sim \chi_{(n-1)}^2 \quad \text{y} \quad \frac{m-1}{\sigma^2} S_{Y_m}^2 \sim \chi_{(m-1)}^2,$$

donde  $S_{X_n}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  y  $S_{Y_m}^2 = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2$ . Como las dos muestras son independientes, y considerando el caso particular del ejemplo 2.12, cuando  $\lambda = \frac{1}{2}$ ,

$$\left. \begin{aligned} V &\sim \text{gamma}\left(r_v, \frac{1}{2}\right) \\ W &\sim \text{gamma}\left(r_w, \frac{1}{2}\right) \end{aligned} \right\} V + W \sim \text{gamma}\left(r_v + r_w, \frac{1}{2}\right)$$

$$V \sim \chi_{2r_v}^2, \quad W \sim \chi_{2r_w}^2 \quad \text{y} \quad V + W \sim \chi_{2r_v+2r_w}^2,$$

luego,

$$\frac{n-1}{\sigma^2} S_{X_n}^2 + \frac{m-1}{\sigma^2} S_{Y_m}^2 \sim \chi_{n-1+m-1}^2, \quad (6.37)$$

por tanto la variable pivotal, siguiendo el ejemplo 6.15, es:

$$Q_{X,Y} = \frac{\frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}}}{\sqrt{\frac{\frac{n-1}{\sigma^2} S_{X_n}^2 + \frac{m-1}{\sigma^2} S_{Y_m}^2}{n-1+m-1}}} = \frac{\frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}}}{\underbrace{\frac{1}{\sigma} \sqrt{\frac{(n-1)S_{X_n}^2 + (m-1)S_{Y_m}^2}{n+m-2}}}_{S_p}}$$

$$= \frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} = T \sim t_{n+m-2}.$$

El estimador de la varianza  $\sigma^2$  es el promedio ponderado

$$S_p^2 = pS_{X_n}^2 + (1-p)S_{Y_m}^2 \quad \text{con} \quad p = \frac{n-1}{n+m-2}, \quad (6.38)$$

conocido como varianza ponderada o *pooled*. Para encontrar un intervalo de confianza para  $\mu_X - \mu_Y$ , se parte de:

$$\begin{aligned} 1 - \alpha &= P(a < Q_{X,Y} < b) = P(Q_{X,Y} < b) - P(Q_{X,Y} \leq a) \\ &= P(T < b) - P(T \leq a). \end{aligned}$$

Sustituyendo  $Q_{X,Y}$  se tiene que:

$$\begin{aligned} 1 - \alpha &= P_\theta \left( a < \frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} < b \right) \\ &= P \left( aS_p \sqrt{\frac{1}{n} + \frac{1}{m}} - (\bar{X}_n - \bar{Y}_m) < -(\mu_X - \mu_Y) < bS_p \sqrt{\frac{1}{n} + \frac{1}{m}} - (\bar{X}_n - \bar{Y}_m) \right) \\ &= P \left( \bar{X}_n - \bar{Y}_m - bS_p \sqrt{\frac{1}{n} + \frac{1}{m}} < \mu_X - \mu_Y < \bar{X}_n - \bar{Y}_m - aS_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right). \end{aligned}$$

El intervalo de confianza del  $100(1 - \alpha)\%$  para  $\mu_X - \mu_Y$  es dado por:

$$IC_{(1-\alpha)}(\mu_X - \mu_Y) = \left( \bar{X}_n - \bar{Y}_m - bS_p \sqrt{\frac{1}{n} + \frac{1}{m}}, \bar{X}_n - \bar{Y}_m - aS_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right),$$

la longitud  $L_I$  de este intervalo es proporcional a  $S_p$ , ya que:

$$\begin{aligned} L_I &= \bar{X}_n - \bar{Y}_m - aS_p \sqrt{\frac{1}{n} + \frac{1}{m}} - \left[ \bar{X}_n - \bar{Y}_m - bS_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right] \\ &= bS_p \sqrt{\frac{1}{n} + \frac{1}{m}} - aS_p \sqrt{\frac{1}{n} + \frac{1}{m}} = (b - a) \sqrt{\frac{1}{n} + \frac{1}{m}} S_p. \end{aligned}$$

Usando multiplicadores de Lagrange, de forma similar al ejemplo 6.16, se tiene el sistema de ecuaciones

$$\begin{aligned} -S_p \sqrt{\frac{1}{n} + \frac{1}{m}} - \lambda g_T(a) &= 0 \quad \Leftrightarrow \quad S_p \sqrt{\frac{1}{n} + \frac{1}{m}} = -\lambda g_T(a) \\ S_p \sqrt{\frac{1}{n} + \frac{1}{m}} + \lambda g_T(b) &= 0 \quad \Leftrightarrow \quad S_p \sqrt{\frac{1}{n} + \frac{1}{m}} = -\lambda g_T(b) \\ \underbrace{\int_{-\infty}^b dG_T(t)}_{P(T < b)} - \underbrace{\int_{-\infty}^a dG_T(t)}_{P(T \leq a)} - (1 - \alpha) &= 0, \quad (6.39) \end{aligned}$$

luego,

$$-\lambda g_T(a) = -\lambda g_T(b) \quad \Leftrightarrow \quad g_T(a) = g_T(b),$$

puesto que la **fdp**  $t$ -Student dada en (2.61) es simétrica alrededor de cero, cuando  $a = b$  o  $a = -b$ , no obstante,  $a = b$  implicaría que  $1 - \alpha = 0$ , lo cual no es admisible, entonces, la solución es  $a = -b$ . Reemplazando en (6.39):

$$\begin{aligned} P(T < b) - P(T < a) - (1 - \alpha) &= 0 \\ P(T < b) - P(T < -b) &= 1 - \alpha \\ &\underbrace{\hspace{10em}}_{1 - P(T < b)} \\ 2P(T < b) - 1 &= 1 - \alpha \\ P(T < b) &= \frac{2 - \alpha}{2} = 1 - \frac{\alpha}{2}, \end{aligned}$$

luego  $b = t_{n+m-2, 1 - \frac{\alpha}{2}}$  y, por tanto,  $a = -t_{n+m-2, 1 - \frac{\alpha}{2}}$ . Así el intervalo de confianza del  $100(1 - \alpha)\%$  para  $\mu_X - \mu_Y$  de mínima longitud es dado por:

$$\begin{aligned} IC_{(1-\alpha)}(\mu_X - \mu_Y) &= \left( \bar{X}_n - \bar{Y}_m - bS_p \sqrt{\frac{1}{n} + \frac{1}{m}}, \bar{X}_n - \bar{Y}_m - aS_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right) \\ &= \left( \bar{X}_n - \bar{Y}_m - t'_v S_p \sqrt{\frac{1}{n} + \frac{1}{m}}, \bar{X}_n - \bar{Y}_m + t'_v S_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right), \end{aligned}$$

donde  $t'_v = t_{n+m-2, 1 - \frac{\alpha}{2}}$ .

**Ejemplo 6.22.** En un estudio realizado en el municipio de Sasaima para evaluar la respuesta de la pitahaya a dos tipos de fertilización [72], se midió, entre otras variables, el peso de los frutos en dos cosechas del año 1989. Se seleccionaron algunos datos para el ejemplo, que se presentan en la tabla 6.2.

Tabla 6.2. Pesos promedios del fruto de pitahaya en gramos, dos cosechas del año 1989

Cosecha A				Cosecha B			
145,4	114,1	134,4	155,5	226,3	152,7	179,5	192,1
150,2	153,7	229,6	162,5	178,5	142,4	185,9	177,1
120,2	83	202,3		217,1	158,9	192,6	146,9
				217,1			

Fuente: adaptada de [72].

Encontrar el intervalo de confianza al nivel 95% para la diferencia media de los pesos promedio del fruto de pitahaya en las dos cosechas consideradas.

**Solución.** A continuación se muestra el procedimiento:

1. Los promedios y las varianzas estimadas para cada cosecha son:

$$\begin{array}{lll} \hat{\mu}_A = \bar{x}_A = 150,0818 & n_A = 11 & s_A^2 = 1620,198 \\ \hat{\mu}_B = \bar{x}_B = 182,0846 & n_B = 13 & s_B^2 = 740,8981 \end{array}$$

2. Se verifica si se pueden considerar las varianzas poblacionales iguales (de los pesos promedio de los frutos), para ello se construye un intervalo de confianza para el cociente de las varianzas, ver expresión (6.33), y se verifica si incluye el 1, de ser así se dice que los datos apoyan el supuesto de varianzas poblacionales iguales. En este caso, se utilizará un nivel de confianza del 90 %.

$$\begin{aligned} IC_{0,90} \left( \frac{\sigma_A^2}{\sigma_B^2} \right) &= \left( F_{\frac{0,10}{2}, 13-1, 11-1} \frac{S_A^2}{S_B^2}; F_{1-\frac{0,10}{2}, 13-1, 11-1} \frac{S_A^2}{S_B^2} \right) \\ &= \left( 0,3632 \times \frac{1620,198}{740,8981}; 2,913 \times \frac{1620,198}{740,8981} \right) \\ &= (0,79; 6,37). \end{aligned}$$

Como el intervalo incluye el 1, se considera que las varianzas poblacionales son iguales.

3. Usando la expresión (6.38) se calcula la varianza ponderada:

$$\begin{aligned} S_p^2 &= \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2} \\ &= \frac{(11 - 1)(1620,198) + (13 - 1)(740,8981)}{11 + 13 - 2} = 1140,58, \end{aligned}$$

luego, los grados de libertad son  $n_A + n_B - 2 = 11 + 13 - 2 = 22$ .

4. El intervalo de confianza para la diferencia de medias es:

$$\begin{aligned} &\left( \bar{X}_A - \bar{X}_B - t_{22, 1-\frac{0,05}{2}} S_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}; \right. \\ &\quad \left. \bar{X}_A - \bar{X}_B + t_{22, 1-\frac{0,05}{2}} S_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \right) = \\ &\left( 150 - 182 - 2,074(33,77) \sqrt{\frac{1}{11} + \frac{1}{13}}; \right. \\ &\quad \left. 150 - 182 + 2,074(33,77) \sqrt{\frac{1}{11} + \frac{1}{13}} \right) \\ &= (-60,7; -3,3) \end{aligned}$$

o

$$(3,3; 60,7).$$

Entonces el intervalo de confianza al nivel 95 % para la diferencia de medias,  $\mu_B - \mu_A$ , de los pesos promedio del fruto de pitahaya en las dos cosechas consideradas es entre 3,3 y 60,7 gramos.  $\checkmark$

### 6.5.2.3. Caso de varianzas desconocidas y distintas a nivel poblacional

Este problema es conocido como el *problema de Behrens–Fisher* (ver [70] o [49]). Varias aproximaciones se han realizado para solucionarlo, pero las más conocidas son las propuestas por [90] y [91]:

$$\bar{X}_n \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right) \quad \text{y} \quad \bar{Y}_m \sim N\left(\mu_Y, \frac{\sigma_Y^2}{m}\right).$$

Por la expresión (6.35) se tiene que:

$$\bar{X}_n - \bar{Y}_m \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right),$$

Luego,

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0, 1).$$

Por la expresión (6.23) se tiene que:

$$\frac{n-1}{\sigma_X^2} S_{X_n}^2 \sim \chi_{(n-1)}^2 \quad \text{y} \quad \frac{m-1}{\sigma_Y^2} S_{Y_m}^2 \sim \chi_{(m-1)}^2,$$

donde  $S_{X_n}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  y  $S_{Y_m}^2 = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2$ . Como las dos muestras son independientes, a partir de (6.37) se llega a

$$\frac{n-1}{\sigma_X^2} S_{X_n}^2 + \frac{m-1}{\sigma_Y^2} S_{Y_m}^2 \sim \chi_{(n-1+m-1)}^2$$

luego, siguiendo el ejemplo 6.15, la variable pivotal es:

$$Q_{X,Y} = \frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \bigg/ \sqrt{\frac{\frac{n-1}{\sigma_X^2} S_{X_n}^2 + \frac{m-1}{\sigma_Y^2} S_{Y_m}^2}{n-1+m-1}}$$

que depende tanto de  $\sigma_X$  como de  $\sigma_Y$ . Este problema fue tratado recientemente en [27]. La fórmula dada en [90] para aproximar los grados de libertad (*g.l.* =  $\nu_1$ ) de la distribución *t*-Student es:

$$\nu_1 \approx (\gamma + 1)^2 \times \left( \frac{1}{n-1} + \frac{\gamma^2}{m-1} \right)^{-1}, \quad (6.40)$$

donde  $\gamma$  es un factor que es función del cociente de varianzas:

$$\gamma = \frac{n S_Y^2}{m S_X^2}. \quad (6.41)$$

En [91, p. 32] se modifican los grados de libertad (*g.l.* =  $\nu_2$ ) de la distribución *t*-Student por

$$\nu_2 \approx (\gamma + 1)^2 \times \left( \frac{1}{n+1} + \frac{\gamma^2}{m+1} \right)^{-1} - 2. \quad (6.42)$$

De acuerdo con [80] estos grados de libertad satisfacen que:

$$\min\{n-1, m-1\} \leq \nu_1 \leq \nu_2 \leq m+n-2$$

y, los valores tabulados de la distribución *t*-Student cumplen que:

$$t_{(m+n-2, 1-\alpha)} \leq t_{(\nu_2, 1-\alpha)} \leq t_{(\nu_1, 1-\alpha)} \leq t_{(\min\{n-1, m-1\}, 1-\alpha)}, \quad 0 < \alpha \leq 0,5.$$

Si se desea un intervalo de confianza conservador, se usaría  $\nu_1$ , pero dado que  $\nu_1$  no es un entero, [80] propone que se interpole el valor de  $t_{(\nu_1, 1-\alpha)}$ , por ejemplo, como sigue:

$$t_{(\nu_1, 1-\alpha)} = (\nu_1 - \llbracket \nu_1 \rrbracket) t_{(\llbracket \nu_1 \rrbracket + 1, 1-\alpha)} + (1 - (\nu_1 - \llbracket \nu_1 \rrbracket)) t_{(\llbracket \nu_1 \rrbracket, 1-\alpha)}, \quad (6.43)$$

donde  $\llbracket x \rrbracket = \max\{k \in \mathbb{Z} \mid k \leq x\}$  y el intervalo de confianza del  $100(1-\alpha)\%$  para  $\mu_X - \mu_Y$  es dado por:

$$IC_{(1-\alpha)}(\mu_X - \mu_Y) = \left( \bar{X}_n - \bar{Y}_m - t' \sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}; \bar{X}_n - \bar{Y}_m + t' \sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}} \right), \quad (6.44)$$

donde  $t' = t_{(\nu_1, 1-\frac{\alpha}{2})}$ .

**Ejemplo 6.23.** Considerando la información del ejemplo 6.17, sobre las dietas en ratas. Encontrar el estimador por intervalo para la diferencia de medias de los pesos (en gramos) de las dos dietas al 90% de nivel de confianza.

*Solución.* A continuación se muestra el procedimiento:

1. Se sabe por los datos obtenidos en la muestra que:

$$\begin{aligned}\hat{\mu}_A = \bar{x}_A &= 274,7 & n_A &= 10 & s_A^2 &= 124,9 \\ \hat{\mu}_B = \bar{x}_B &= 535,5 & n_B &= 8 & s_B^2 &= 4762,2857\end{aligned}$$

2. Se verifica si se pueden considerar las varianzas poblacionales iguales (de los pesos promedio de los frutos), para ello se construye un intervalo de confianza para el cociente de las varianzas, ver expresión (6.33), y se mira si incluye el 1, en cuyo caso se dice que los datos apoyan el supuesto de varianzas poblacionales iguales. En este caso, se utilizará un nivel de confianza del 90 %.

$$\begin{aligned}IC_{0,90} \left( \frac{\sigma_B^2}{\sigma_A^2} \right) &= \left( F_{\frac{0,10}{2}, 10-1, 8-1} \frac{S_B^2}{S_A^2}; F_{1-\frac{0,10}{2}, 10-1, 8-1} \frac{S_B^2}{S_A^2} \right) \\ &= \left( 0,304 \times \frac{4762,2857}{124,9}; 3,667 \times \frac{4762,2857}{124,9} \right) \\ &= (11,5796; 140,1872).\end{aligned}$$

Como el intervalo no incluye el 1, se considera que las varianzas poblacionales no son iguales.

3. Los grados de libertad se hallan con las expresiones (6.40) y (6.42):

$$\begin{aligned}\gamma &= \frac{10}{8} \times \frac{4762,2857}{124,9} = 47,661 \\ v_1 &= (47,661)^2 \times \left[ \frac{1}{10-1} + \frac{(47,661)^2}{8-1} \right]^{-1} = 7,294325 \\ v_2 &= (47,661)^2 \times \left[ \frac{1}{10+1} + \frac{(47,661)^2}{8+1} \right]^{-1} - 2 = 7,378224.\end{aligned}$$

Luego,  $\llbracket v_1 \rrbracket = 7$  y usando la expresión (6.43) con  $\alpha = 0,05$ ,

$$\begin{aligned}t_{(v_1, 0,95)} &= 0,294325 \times t_{(8, 0,95)} + 0,705675 \times t_{(7, 0,95)} \\ &= 0,294325 \times 1,860 + 0,705675 \times 1,895 = 1,8847,\end{aligned}$$

4. El intervalo de confianza para la diferencia de medias es:

$$\begin{aligned}
 IC_{0,90}(\mu_A - \mu_B) &= \left( \bar{X}_A - \bar{X}_B - t_{v_1, 1-\frac{0,10}{2}} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}; \right. \\
 &\quad \left. \bar{X}_A - \bar{X}_B + t_{v_1, 1-\frac{0,10}{2}} \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} \right) \\
 &= \left( 274,7 - 535,5 - 1,8847 \sqrt{\frac{124,9}{10} + \frac{4762,2857}{8}}; \right. \\
 &\quad \left. 274,7 - 535,5 + 1,8847 \sqrt{\frac{124,9}{10} + \frac{4762,2857}{8}} \right) \\
 &= (-307,2637; -214,3362)
 \end{aligned}$$

o

$$IC_{0,90}(\mu_B - \mu_A) = (214,3362; 307,2637).$$

Por tanto, la diferencia de medias en el peso (en gramos),  $\mu_B - \mu_A$ , de las ratas a los 64 días de aplicadas las dietas A y B está entre 214,1 y 307,5 gramos con un nivel de confianza del 90%.  $\square$

**Nota 6.9.** En el ejemplo 2.21 se estableció que a medida que los grados de libertad de la distribución *t*-Student aumentan, la *fdp* *t*-Student se aproxima a la *fdp* normal estándar. Cuando el tamaño de muestra es mayor a 30, los valores tabulados de la *t*-Student son similares a los valores tabulados de la distribución normal estándar.

**Nota 6.10.** Las instrucciones en R para una simulación sobre el concepto de intervalos de confianza, cuando los supuestos no se cumplen y su afectación en la tasa de cobertura, son dadas por [3, p. 9].

```

library(animation)
remove(list=ls())
tstatistic=function(x,y) # función que calcula el estadístico t
{
  m1=length(x); n1=length(y)
  sp=sqrt(((m1-1)*sd(x)^2+(n1-1)*sd(y)^2)/(m1+n1-2))
  difmedia=mean(x)-mean(y)
  LInf=(mean(x)-mean(y))-qt(1-level/2,n1+m1-2)*
    (sp*sqrt(1/m1+1/n1))
  LSup=(mean(x)-mean(y))+qt(1-level/2,n1+m1-2)*
    (sp*sqrt(1/m1+1/n1))
  t=(mean(x)-mean(y))/(sp*sqrt(1/m1+1/n1))
  intervalo=c(difmedia,LInf,LSup,t)
}

```

```

return(intervalo)
}
# valores para la simulación como: alpha, tamaños de muestra y
# número de intervalos
level=.1; m1=10; n1=10
n.sim=100 # número de simulaciones
n.reject=0 # para contar los rechazos
simul=NULL # considera diferentes escenarios
for (i in 1:n.sim) {
#x=rnorm(m1,mean=0,sd=1) # simulates xs from population 1
#y=rnorm(n1,mean=0,sd=1) # simulates ys from population 2
#x=rnorm(m1,mean=0,sd=1)
#y=rnorm(n1,mean=0,sd=10)
x=rt(m1,df=4)
y=rt(n1,df=4)
#x=rexp(m1,rate=1)
#y=rexp(n1,rate=1)
#x=rnorm(m1,mean=10,sd=2)
#y=rexp(n1,rate=1/10)
simul=rbind(simul,tstatistic(x,y)) # calcula el estadístico t
# para cada simulación
if (abs(simul[i,4])>qt(1-level/2,n1+m1-2))
n.reject=n.reject+1 # Rechaza si |t| excede el valor crítico
}
n.reject # imprime el número de rechazos
# instrucciones para el gráfico - modificación de
# la función conf.int
y0=simul[,2]; y1=simul[,3]
rg = range(c(y0, y1))
cvr = y0 < 0 & y1 > 0
xax = pretty(1:n.sim)
oopt = ani.options(interval = .1)
for (i in 1:n.sim) {
plot(1:n.sim, ylim = rg, type = "n", xlab = "Muestra",
ylab = "Intervalo de confianza",xaxt = "n")
abline(h = 0, lty = 2)
axis(1, xax[xax <= i])
arrows(1:i, y0[1:i], 1:i, y1[1:i],
length = par("din")[1]/n.sim * 0.5,
angle = 90, code = 3, col = cvr[1:i] + 1)
points(1:i, simul[1:i,1], col = cvr[1:i] + 1)
legend("topright", legend = format(c(i - sum(cvr[1:i]),
sum(cvr[1:i])),width = nchar(n.sim)),bty="n.sim",ncol=2)
legend("topleft", legend=paste("Tasa de cobertura:",
format(round(mean(cvr[1:i]),3),nsmall=3)),bty="n.sim")

```

```

    ani.pause()
  }
ani.options(oopt)

```

### 6.5.3. Intervalo de confianza para diferencia de medias de dos poblaciones no independientes

Sean  $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}, \begin{pmatrix} X_3 \\ Y_3 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$  vectores aleatorios iid tal que

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} \sim N_2 \left[ \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix} \right],$$

luego,

$$\begin{aligned} \mathbb{E}(X_1) &= \mu_X & \text{y} & & \mathbb{E}(Y_1) &= \mu_Y \\ \mathbb{V}(X_1) &= \sigma_X^2 & \text{y} & & \mathbb{V}(Y_1) &= \sigma_Y^2 & \text{y} & & \text{Cov}(X_1, Y_1) &= \rho\sigma_X\sigma_Y \\ \mathbb{E}(Y_1 - X_1) &= \mu_Y - \mu_X & \text{y} & & \mathbb{V}(Y_1 - X_1) &= \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y. \end{aligned}$$

Si  $D_i = Y_i - X_i$  y  $\bar{D}_n = \frac{1}{n} \sum_{i=1}^n D_i$ , entonces:

$$\begin{aligned} D_i &\sim N\left(\mu_Y - \mu_X, \underbrace{\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y}_{\sigma_D^2}\right) \\ \bar{D}_n &\sim N\left(\mu_Y - \mu_X, \frac{\sigma_D^2}{n}\right). \end{aligned}$$

En este caso, la variable pivotal es dada por:

$$\frac{\bar{D}_n - (\mu_Y - \mu_X)}{\frac{\sigma_D}{\sqrt{n}}} \sim N(0, 1).$$

Del ejemplo B.6 se tiene que:

$$\frac{n-1}{\sigma_D^2} S_{D,n}^2 \sim \chi_{(n-1)}^2 \quad \text{con} \quad S_{D,n}^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2,$$

siguiendo el ejemplo 6.15, la variable pivotal es:

$$Q_D = \frac{\bar{D}_n - (\mu_Y - \mu_X)}{\sqrt{\frac{\sigma_D^2}{n}}} = \frac{\sqrt{n} [\bar{D}_n - (\mu_Y - \mu_X)]}{S_{D,n}} \sim t_{(n-1)}.$$

Para encontrar un intervalo de confianza para  $\mu_Y - \mu_X$ , de forma similar al ejemplo 6.16, se encuentra que el intervalo de confianza del  $100(1 - \alpha)\%$  para  $\mu_Y - \mu_X$  es dado por:

$$IC_{(1-\alpha)}(\mu_Y - \mu_X) = \left( \bar{D}_n - t_{n-1, 1-\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}, \bar{D}_n + t_{n-1, 1-\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}} \right). \quad (6.45)$$

**Ejemplo 6.24.** De acuerdo con [84], respecto a un estudio en la Universidad de São Paulo donde se evaluaba el efecto de ingerir sodio en la presión arterial de pacientes hipertensos, en la tabla 6.3 se presentan los datos.

Tabla 6.3. Ingestión de sodio en la presión arterial sistólica (mmHg)

Paciente	Bajo teor de sodio	Alto teor de sodio	$D_i$ (Alto-Bajo)
1	138	143	5
2	147	154	7
3	146	147	1
4	154	147	-7
5	142	157	15
6	156	158	2
7	134	164	30
8	146	156	10
9	143	151	8
10	175	182	7
11	156	151	-5
12	117	116	-1
13	157	154	-3
14	143	149	6
15	127	126	-1
16	134	138	4
17	112	115	3
18	144	159	15
19	117	124	7
20	128	125	-3

Fuente: adaptada de [84, p. 13].

Suponiendo que la distribución de las diferencias es normal, calcular un intervalo de confianza al 95% para la diferencia media de la presión arterial sistólica entre las dos dietas.

*Solución.* A partir de los datos de la tabla 6.3, se tiene que:

$$\bar{D} = 5 \quad \text{y} \quad s_{D,20}^2 = 70,84211.$$

Usando la expresión (6.45) y el valor tabulado en la tabla C.2:

$$\begin{aligned} IC_{0,95}(\mu_Y - \mu_X) &= \left( \bar{D} - t_{20-1, 1-\frac{0,05}{2}} \frac{S_D}{\sqrt{n}}; \bar{D}_n + t_{20-1, 1-\frac{0,05}{2}} \frac{S_D}{\sqrt{n}} \right) \\ &= \left( 5 - 2,093 \sqrt{\frac{70,84211}{20}}; 5 + 2,093 \sqrt{\frac{70,84211}{20}} \right) \\ &= (1,06; 8,94). \end{aligned}$$

Así, la diferencia media de la presión arterial sistólica entre las dos dietas, alto teor de sodio menos bajo teor de sodio, está entre 1,06 y 8,94 mm Hg con un nivel de confianza del 95%.  $\checkmark$

### 6.5.4. Intervalo de confianza para diferencia de proporciones de dos poblaciones independientes

Supóngase que  $X_{11}, X_{12}, \dots, X_{1i}, \dots, X_{1,n_1}$  son variables aleatorias iid tal que  $X_{1i} \sim Ber(p_1)$  y  $X_{21}, X_{22}, \dots, X_{2i}, \dots, X_{2,n_2}$  son variables aleatorias iid tal que  $X_{2i} \sim Ber(p_2)$ , donde las muestras aleatorias son independientes entre sí. El intervalo de confianza del  $100(1-\alpha)\%$  para  $p_1 - p_2$ , recomendado para este caso es dado por [2] (ver [21, p. 111]),

$$\left( \tilde{p}_1 - \tilde{p}_2 - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2}}, \right. \\ \left. \tilde{p}_1 - \tilde{p}_2 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2}} \right),$$

donde  $\tilde{p}_k = \frac{1}{n_k+2} \left( 1 + \sum_{i=1}^{n_k} X_{ki} \right)$  para  $k = 1, 2$ .

**Ejemplo 6.25.** Considerando la información del ejemplo 6.20, calcular un intervalo de confianza del 95% para la diferencia entre los porcentajes de preferencia por las marcas de cigarrillo.

*Solución.* Se sabe que:

Marca	$n_k$	$x_k = \sum_{i=1}^{n_k} x_i$	$\tilde{p}_k = \frac{1+x_k}{n_k+2}$
A	200	42	$\frac{1+42}{202} = 0,2129$
B	150	18	$\frac{1+18}{152} = 0,125$

Luego,

$$\begin{aligned}
 IC_{(1-\alpha)}(p_A - p_B) &= \left( \tilde{p}_A - \tilde{p}_B - z_{1-\frac{0,05}{2}} \sqrt{\frac{\tilde{p}_A(1-\tilde{p}_A)}{n_A} + \frac{\tilde{p}_B(1-\tilde{p}_B)}{n_B}}, \right. \\
 &\quad \left. \tilde{p}_A - \tilde{p}_B + z_{1-\frac{0,05}{2}} \sqrt{\frac{\tilde{p}_A(1-\tilde{p}_A)}{n_A} + \frac{\tilde{p}_B(1-\tilde{p}_B)}{n_B}} \right) \\
 &= \left( 0,2129 - 0,125 - 1,96 \sqrt{\frac{0,2129(1-0,2129)}{200} + \frac{0,125(1-0,125)}{150}}, \right. \\
 &\quad \left. 0,2129 - 0,125 + 1,96 \sqrt{\frac{0,2129(1-0,2129)}{200} + \frac{0,125(1-0,125)}{150}} \right) \\
 &= (0,0103, 0,1655).
 \end{aligned}$$

Por tanto, un intervalo de confianza del 95 % para la diferencia entre los porcentajes de preferencia por las marcas de cigarrillo, marca A menos marca B, está entre 1 % y 16,5 %.  $\checkmark$

## 6.6. Estimación bayesiana por intervalo

De manera similar al intervalo de confianza clásico, en el análisis bayesiano se puede calcular un intervalo bayesiano del  $100(1-\alpha)\%$  empleando la distribución *a posteriori*.

**Teorema 6.3.** De acuerdo con [57, p. 3], si  $X_1, X_2, \dots, X_n$  son variables aleatorias *iid* normales con media  $\mu$  y varianza conocida  $\sigma^2$ , y la distribución a priori de la media poblacional es una distribución normal con media conocida  $\vartheta$  y varianza conocida  $\tau^2$ . Entonces la distribución a posteriori de la media poblacional es una distribución normal con media  $\tilde{\mu}_n$  y varianza  $\tilde{\sigma}_n^2$ , donde,

$$\tilde{\mu}_n = \frac{\sigma^2/n}{\sigma^2/n + \tau^2} \vartheta + \frac{\tau^2}{\sigma^2/n + \tau^2} \bar{x} \quad y \quad \tilde{\sigma}_n^2 = \frac{\tau^2}{\sigma^2/n + \tau^2} \frac{\sigma^2}{n}. \quad (6.46)$$

*Demostración.* Si la muestra aleatoria es  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ , entonces,

- Distribución *a priori*:

$$f(\mu; \vartheta, \tau) = \pi(\mu) = \frac{1}{\sqrt{2\pi\tau}} \exp \left\{ -\frac{1}{2} \left( \frac{\mu - \vartheta}{\tau} \right)^2 \right\}.$$

- Distribución modelo:

$$f(x_k | \mu) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{1}{2} \left( \frac{x_k - \mu}{\sigma} \right)^2 \right\}, \quad k = 1, 2, \dots, n.$$

La fdp de la muestra es:

$$f(\mathbf{x} | \mu) = \prod_{k=1}^n f(x_k | \mu) = \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp \left\{ -\frac{1}{2} \sum_{k=1}^n \left( \frac{x_k - \mu}{\sigma} \right)^2 \right\}.$$

- Distribución conjunta:

$$\begin{aligned} f(\mathbf{x}; \mu) &= f(\mathbf{x} | \mu) \pi(\mu) \\ &= \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp \left\{ -\frac{1}{2} \sum_{k=1}^n \left( \frac{x_k - \mu}{\sigma} \right)^2 \right\} \frac{1}{\sqrt{2\pi\tau}} \exp \left\{ -\frac{1}{2} \left( \frac{\mu - \vartheta}{\tau} \right)^2 \right\} \\ &= \frac{1}{(2\pi)^{\frac{n+1}{2}} \sigma^n \tau} \exp \left\{ -\frac{1}{2} \left[ \sum_{k=1}^n \left( \frac{x_k - \mu}{\sigma} \right)^2 + \left( \frac{\mu - \vartheta}{\tau} \right)^2 \right] \right\} \\ &= \frac{1}{(2\pi)^{\frac{n+1}{2}} \sigma^n \tau} \exp \left\{ -\frac{1}{2} \left[ \sum_{k=1}^n \frac{x_k^2 - 2\mu x_k + \mu^2}{\sigma^2} + \frac{\mu^2 - 2\mu\vartheta + \vartheta^2}{\tau^2} \right] \right\} \\ &= \frac{\exp \left\{ -\frac{1}{2} \left[ \sum_{k=1}^n \left( \frac{x_k}{\sigma} \right)^2 + \left( \frac{\vartheta}{\tau} \right)^2 \right] \right\}}{(2\pi)^{\frac{n+1}{2}} \sigma^n \tau} \exp \left\{ -\frac{\mu}{2} \left[ \sum_{k=1}^n \frac{\mu - 2x_k}{\sigma^2} + \frac{\mu - 2\vartheta}{\tau^2} \right] \right\} \\ &= \frac{\exp \left\{ -\frac{1}{2} \left[ \sum_{k=1}^n \left( \frac{x_k}{\sigma} \right)^2 + \left( \frac{\vartheta}{\tau} \right)^2 \right] \right\}}{(2\pi)^{\frac{n+1}{2}} \sigma^n \tau} \exp \left\{ -\frac{\mu}{2} \left[ \frac{n\mu - 2n\bar{x}}{\sigma^2} + \frac{\mu - 2\vartheta}{\tau^2} \right] \right\}, \end{aligned}$$

agrupando términos:

$$f(\mathbf{x}; \mu) = \frac{\exp \left\{ -\frac{1}{2} \left[ \sum_{k=1}^n \left( \frac{x_k}{\sigma} \right)^2 + \left( \frac{\vartheta}{\tau} \right)^2 \right] \right\}}{(2\pi)^{\frac{n+1}{2}} \sigma^n \tau} \exp \left\{ -\frac{1}{2} [\alpha_n \mu^2 - 2\beta_n \mu] \right\}.$$

donde  $\alpha_n = \frac{n}{\sigma^2} + \frac{1}{\tau^2}$  y  $\beta_n = \frac{n\bar{x}}{\sigma^2} + \frac{\vartheta}{\tau^2}$ .

■ Distribución marginal:

$$\begin{aligned}
 f(x) &= \int_{-\infty}^{\infty} f(x | \mu) \pi(\mu) d\mu \\
 &= \int_{-\infty}^{\infty} \frac{\exp \left\{ -\frac{1}{2} \left[ \sum_{k=1}^n \left( \frac{x_k}{\sigma} \right)^2 + \left( \frac{\vartheta}{\tau} \right)^2 \right] \right\}}{(2\pi)^{\frac{n+1}{2}} \sigma^n \tau} \exp \left\{ -\frac{1}{2} [\alpha_n \mu^2 - 2\beta_n \mu] \right\} d\mu \\
 &= \frac{\exp \left\{ -\frac{1}{2} \left[ \sum_{k=1}^n \left( \frac{x_k}{\sigma} \right)^2 + \left( \frac{\vartheta}{\tau} \right)^2 \right] \right\}}{(2\pi)^{\frac{n+1}{2}} \sigma^n \tau} \int_{-\infty}^{\infty} \exp \left\{ -\frac{\alpha_n}{2} \left[ \mu^2 - 2\frac{\beta_n}{\alpha_n} \mu \right] \right\} d\mu \\
 &= \frac{\exp \left\{ -\frac{1}{2} \left[ \sum_{k=1}^n \left( \frac{x_k}{\sigma} \right)^2 + \left( \frac{\vartheta}{\tau} \right)^2 \right] \right\}}{(2\pi)^{\frac{n+1}{2}} \sigma^n \tau} \int_{-\infty}^{\infty} \exp \left\{ \frac{\alpha_n}{2} \left[ \left( \frac{\beta_n}{\alpha_n} \right)^2 - \left( \mu - \frac{\beta_n}{\alpha_n} \right)^2 \right] \right\} d\mu \\
 &= \frac{\exp \left\{ -\frac{1}{2} \left[ \sum_{k=1}^n \left( \frac{x_k}{\sigma} \right)^2 + \left( \frac{\vartheta}{\tau} \right)^2 - \frac{\beta_n^2}{\alpha_n} \right] \right\}}{(2\pi)^{\frac{n+1}{2}} \sigma^n \tau} \int_{-\infty}^{\infty} \exp \left\{ -\frac{\alpha_n}{2} \left( \mu - \frac{\beta_n}{\alpha_n} \right)^2 \right\} d\mu \\
 &= \frac{\exp \left\{ -\frac{1}{2} \left[ \sum_{k=1}^n \left( \frac{x_k}{\sigma} \right)^2 + \left( \frac{\vartheta}{\tau} \right)^2 - \frac{\beta_n^2}{\alpha_n} \right] \right\}}{(2\pi)^{\frac{n}{2}} \sqrt{\alpha_n} \sigma^n \tau}.
 \end{aligned}$$

■ Distribución *a posteriori*:

$$\begin{aligned}
 \pi(\mu|x) &= \frac{f(x; \mu)}{f(x)} \\
 &= \frac{\exp \left\{ -\frac{1}{2} \left[ \sum_{k=1}^n \left( \frac{x_k}{\sigma} \right)^2 + \left( \frac{\vartheta}{\tau} \right)^2 \right] \right\} \exp \left\{ -\frac{1}{2} [\alpha_n \mu^2 - 2\beta_n \mu] \right\}}{(2\pi)^{\frac{n+1}{2}} \sigma^n \tau \frac{\exp \left\{ -\frac{1}{2} \left[ \sum_{k=1}^n \left( \frac{x_k}{\sigma} \right)^2 + \left( \frac{\vartheta}{\tau} \right)^2 - \frac{\beta_n^2}{\alpha_n} \right] \right\}}{(2\pi)^{\frac{n}{2}} \sqrt{\alpha_n} \sigma^n \tau} \\
 &= \frac{\sqrt{\alpha_n}}{(2\pi)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \frac{\beta_n^2}{\alpha_n} \right\} \exp \left\{ -\frac{1}{2} [\alpha_n \mu^2 - 2\beta_n \mu] \right\} \\
 &= \frac{1}{\sqrt{2\pi \alpha_n^{-1}}} \exp \left\{ -\frac{\alpha_n}{2} \left( \mu - \frac{\beta_n}{\alpha_n} \right)^2 \right\}, \tag{6.47}
 \end{aligned}$$

es decir,

$$\begin{aligned}
 \tilde{\mu}_n &= \frac{\beta_n}{\alpha_n} = \frac{\frac{n\bar{x}}{\sigma^2} + \frac{\vartheta}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} = \frac{\sigma^2}{\sigma^2 + n\tau^2} \vartheta + \frac{n\tau^2}{\sigma^2 + n\tau^2} \bar{x} \\
 \tilde{\sigma}_n^2 &= \frac{1}{\alpha_n} = \left( \frac{n}{\sigma^2} + \frac{1}{\tau^2} \right)^{-1} = \left( \frac{\sigma^2 + n\tau^2}{\sigma^2 \tau^2} \right)^{-1} = \frac{\tau^2}{\sigma^2 + n\tau^2} \sigma^2,
 \end{aligned}$$

si  $\omega_n = \frac{\sigma^2/n}{\sigma^2/n + \tau^2}$ , entonces,

$$\mu | x \sim N\left(\omega_n \vartheta + (1 - \omega_n)\bar{x}, (1 - \omega_n)\frac{\sigma^2}{n}\right). \quad \checkmark$$

Nótese que  $\tilde{\mu}_n$  es el promedio ponderado de la media muestral ( $\bar{x}$ ) y la media *a priori* ( $\vartheta$ ). Como ambas ponderaciones están entre 0 y 1 y suman 1, la media *a posteriori*  $\tilde{\mu}_n$  siempre se encuentra entre  $\bar{x}$  y  $\vartheta$ . Esto significa que tanto  $\bar{x}$  como  $\vartheta$  influyen en la estimación *a posteriori* de  $\tilde{\mu}_n$ . Por otra parte, el valor  $\tilde{\sigma}_n$  es menor que la desviación estándar *a priori* ( $\tau$ ) y que la desviación estándar de  $\bar{x}$ , es decir,  $\sigma/\sqrt{n}$ . Esto sugiere que la estimación *a posteriori*  $\tilde{\sigma}_n$  es más precisa que la *a priori* y que los datos muestrales.

**Teorema 6.4.** De acuerdo con [57, p. 26], si  $X_1, X_2, \dots, X_n$  son variables aleatorias *iid* normales con media conocida  $\mu$  y varianza desconocida  $\sigma^2$ , y la distribución *a priori* de  $\sigma^2$  es una  $\chi^2$ -inversa escalada con parámetros conocidos ( $\kappa; \kappa\sigma_0^2$ ). Entonces la distribución *a posteriori* de la variable:

$$\frac{\kappa\sigma_0^2 + nS_x^2}{\sigma^2} \sim \chi^2_{(v_1)} \quad \text{con} \quad S_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2,$$

donde  $v_1 = \kappa + n$ .

*Demostración.* Si la muestra aleatoria es  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ , entonces:

- Distribución modelo:

$$f(x_i | \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2\right\}, \quad i = 1, 2, \dots, n.$$

La **fdp** de la muestra es:

$$\begin{aligned} f(\mathbf{x} | \sigma^2) &= \prod_{i=1}^n f(x_i | \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{n}{2}\frac{S_x^2}{\sigma^2}\right\}. \end{aligned}$$

- La **fdp** *a priori* es la distribución gamma inversa, cuya **fdp** se define sobre el apoyo  $x > 0$ :

$$\pi(\sigma^2) = \frac{\beta}{\Gamma(\alpha)} \left(\frac{\beta}{\sigma^2}\right)^{\alpha-1} \exp\left\{-\frac{\beta}{\sigma^2}\right\},$$

con parámetro de forma y parámetro de escala  $(\alpha, \beta)$ . Si  $X$  tiene una distribución de gamma inversa con  $\alpha = k/2$ ,  $\beta = 1/2$ , entonces se dice que  $X$  tiene una distribución chi-cuadrado inversa, i.e.,  $\chi_k^{-2}$ . Una parametrización apropiada es como una distribución chi-cuadrado inversa escalada con parámetros  $\kappa$  y  $\kappa\sigma_0^2$  grados de libertad<sup>39</sup>; es decir, la distribución *a priori* de  $\sigma^2$  se toma como la distribución de  $\kappa\sigma_0^2/X$ , donde  $X$  es una  $\chi_k^2$  variable aleatoria. Usamos la notación conveniente pero no estándar,  $\sigma^2 \sim \chi_{(\kappa, \kappa\sigma_0^2)}^{-2}$ .

- Distribución conjunta:

$$\begin{aligned} f(x; \sigma^2) &= f(x | \sigma^2) \pi(\sigma^2) \\ &= \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ -\frac{n}{2} \frac{S_x^2}{\sigma^2} \right\} \cdot \frac{2}{\kappa\sigma_0^2} \left( \frac{\kappa\sigma_0^2}{2\sigma^2} \right)^{\frac{\kappa}{2}+1} \exp \left\{ -\frac{\kappa\sigma_0^2}{2\sigma^2} \right\} \\ &= \frac{1}{\pi^{\frac{n}{2}}} \cdot \frac{2}{\Gamma(\frac{\kappa}{2})} \frac{(\kappa\sigma_0^2)^{\frac{\kappa}{2}}}{(2\sigma^2)^{\frac{n+\kappa}{2}+1}} \exp \left\{ -\frac{1}{2\sigma^2} (\kappa\sigma_0^2 + nS_x^2) \right\}. \end{aligned}$$

- Distribución marginal:

$$\begin{aligned} f(x) &= \int_0^\infty f(x | \sigma^2) \pi(\sigma^2) d\sigma^2 \\ &= \int_0^\infty \frac{1}{\pi^{\frac{n}{2}}} \cdot \frac{2}{\Gamma(\frac{\kappa}{2})} \frac{(\kappa\sigma_0^2)^{\frac{\kappa}{2}}}{(2\sigma^2)^{\frac{n+\kappa}{2}+1}} \exp \left\{ -\frac{1}{2\sigma^2} (\kappa\sigma_0^2 + nS_x^2) \right\} d\sigma^2, \end{aligned}$$

haciendo el cambio de variable:

$$u = \frac{1}{2\sigma^2} (\kappa\sigma_0^2 + nS_x^2) \quad \Rightarrow \quad du = -\frac{1}{2\sigma^4} (\kappa\sigma_0^2 + nS_x^2) d\sigma^2,$$

<sup>39</sup>Si una variable aleatoria  $X$  se distribuye como una distribución chi-cuadrado, entonces  $Y = \frac{1}{X}$  se distribuye como una chi-cuadrado inversa (usando el teorema 2.7):

$$f(y|k) = \frac{\left(\frac{1}{2}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} y^{-\left(\frac{k}{2}+1\right)} \exp\left\{-\frac{1}{2y}\right\}.$$

A su vez, la distribución chi-cuadrado inversa escalada  $\chi_{(k, \kappa\sigma^2)}^{-2}$  tiene como [fdp](#)

$$f(x|k, \sigma) = \frac{\left(\frac{\kappa\sigma^2}{2}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} x^{-\left(\frac{k}{2}+1\right)} \exp\left\{-\frac{\kappa\sigma^2}{2x}\right\}. \quad (6.48)$$

Si una variable aleatoria  $X$  se distribuye como una distribución chi-cuadrado inversa escalada por  $\kappa\sigma^2$ , entonces  $Y = \frac{\kappa\sigma^2}{X}$  se distribuye como una chi-cuadrado.

sustituyendo en la integral y usando la expresión (1.23):

$$\begin{aligned}
 f(x) &= \frac{1}{\pi^{\frac{n}{2}}} \cdot \frac{(\kappa\sigma_0^2)^{\frac{\kappa}{2}}}{\Gamma(\frac{\kappa}{2})} \int_0^\infty \frac{u^{\frac{n+\kappa}{2}-1}}{(\kappa\sigma_0^2 + nS_x^2)^{\frac{n+\kappa}{2}}} \exp\{-u\} du \\
 &= \frac{1}{\pi^{\frac{n}{2}}} \frac{(\kappa\sigma_0^2)^{\frac{\kappa}{2}}}{(\kappa\sigma_0^2 + nS_x^2)^{\frac{n+\kappa}{2}}} \frac{\Gamma(\frac{n+\kappa}{2})}{\Gamma(\frac{\kappa}{2})} \\
 &= \frac{\pi^{-\frac{n}{2}} \Gamma(\frac{n}{2})}{B(\kappa/2, n/2)} \left( \frac{\kappa\sigma_0^2}{\kappa\sigma_0^2 + nS_x^2} \right)^{\kappa/2} \left( \frac{1}{\kappa\sigma_0^2 + nS_x^2} \right)^{n/2}.
 \end{aligned}$$

- La *fdp a posteriori* resultante para  $\sigma^2$  es:

$$\begin{aligned}
 \pi(\sigma^2 | x) &= \frac{f(x; \sigma^2)}{f(x)} \\
 &= \frac{\frac{1}{\pi^{\frac{n}{2}}} \cdot \frac{2}{\Gamma(\frac{\kappa}{2})} \frac{(\kappa\sigma_0^2)^{\frac{\kappa}{2}}}{(2\sigma^2)^{\frac{n+\kappa}{2}+1}} \exp\left\{-\frac{\kappa\sigma_0^2 + nS_x^2}{2\sigma^2}\right\}}{\frac{1}{\pi^{\frac{n}{2}}} \frac{(\kappa\sigma_0^2)^{\frac{\kappa}{2}}}{(\kappa\sigma_0^2 + nS_x^2)^{\frac{n+\kappa}{2}}} \frac{\Gamma(\frac{n+\kappa}{2})}{\Gamma(\frac{\kappa}{2})}} \\
 &= \left( \frac{\kappa\sigma_0^2 + nS_x^2}{2} \right)^{\frac{n+\kappa}{2}} \frac{(\sigma^2)^{-((n+\kappa)/2+1)}}{\Gamma(\frac{n+\kappa}{2})} \exp\left\{-\frac{\kappa\sigma_0^2 + nS_x^2}{2\sigma^2}\right\}.
 \end{aligned}$$

Al comparar con la *fdp* dada en (6.48) se tiene que:

$$\sigma^2 | x \sim \chi_{(\nu_1, \nu_1 S_1^2)}^{-2},$$

donde

$$\nu_1 = \kappa + n, \quad \text{y} \quad \nu_1 S_1^2 = \kappa\sigma_0^2 + nS_x^2,$$

luego la distribución *a posteriori* de  $\sigma^2$  condicionada por la observación de la muestra, es una  $\chi^2$ -inversa escalada con  $\nu_1$  grados de libertad iguales a la suma de los grados de libertad anteriores y el tamaño de la muestra, y escalada por  $\nu_1 S_1^2$  igual al promedio ponderado de grados de libertad de las escalas anteriores y de datos; o equivalentemente,

$$\frac{\nu_1 S_1^2}{\sigma^2} \Big| x \sim \chi_{(\nu_1)}^2.$$

Nótese que el valor

$$S_1^2 = \frac{\kappa}{\kappa + n} \sigma_0^2 + \frac{n}{\kappa + n} S_x^2$$

es el promedio ponderado de la varianza muestral ( $S_x^2$ ) y la varianza *a priori* ( $\sigma_0^2$ ). Los pesos son adecuados pues valores grandes de  $\kappa$  implican un conocimiento bastante preciso de  $\sigma^2$  antes del experimento y valores grandes de  $n$  corresponden naturalmente a mucho más conocimiento de la muestra.  $\checkmark$

**DEFINICIÓN 6.10.** De acuerdo con [78, p. 59], si  $\pi(\theta | x)$  es cualquier distribución a posteriori de  $\theta$  después de observar  $x$ , el intervalo  $\theta_L < \theta < \theta_U$  se denomina intervalo de confianza bayesiano del  $100(1 - \alpha)\%$  para  $\theta$  si:

$$\int_{\theta_L}^{\theta_U} \pi(\theta | x) d\theta = 1 - \alpha.$$

O equivalentemente

$$\int_{-\infty}^{\theta_L} \pi(\theta | x) d\theta = \int_{\theta_U}^{\infty} \pi(\theta | x) d\theta = \frac{\alpha}{2}. \quad (6.49)$$

**Ejemplo 6.26.** Supóngase, como en [89, p. 716], que  $X \sim \text{Bin}(2, p)$  y la distribución a priori de  $p$  es uniforme en el intervalo  $(0, 1)$ . Encuentre el intervalo de confianza bayesiano del 95% para  $p$ , si se observa  $x = 0$ .

*Solución.* En este caso:

- Distribución modelo: dado que  $X$  es binomial entonces,

$$f(x|p) = \text{Bin}(x; 2, p) = \binom{2}{x} p^x q^{2-x}, \quad x = 0, 1, 2.$$

- La distribución *a priori* de  $p$  es uniforme en el intervalo  $(0, 1)$ , es decir,

$$\pi(p) = 1, \quad \text{para } 0 < p < 1.$$

- La distribución marginal de  $x$  se puede calcular a partir de (1.22) como:

$$\begin{aligned} g(x) &= \int_0^1 f(x|p)\pi(p) dp = \binom{2}{x} \int_0^1 p^x (1-p)^{2-x} dp \\ &= \frac{\Gamma(3)}{\Gamma(x+1)\Gamma(3-x)} \frac{\Gamma(x+1)\Gamma(3-x)}{\Gamma(4)} = \frac{\Gamma(3)}{\Gamma(4)} = \frac{1}{3}. \end{aligned}$$

En la integral anterior se expresó tanto la combinatoria como la función beta en términos de la función gamma.

- La distribución *a posteriori* de  $p$ , dado  $x$ , es:

$$\pi(p|x) = \frac{\binom{2}{x} p^x q^{2-x}}{1/3} = \frac{\Gamma(4)p^x(1-p)^{2-x}}{\Gamma(x+1)\Gamma(3-x)}, \quad 0 < p < 1.$$

Es decir, esta distribución *a posteriori* sigue una **fdp** beta con parámetros  $\alpha = x + 1$  y  $\beta = 3 - x$ .

Por lo tanto, si se observa  $x = 0$ , la distribución *a posteriori* de  $p$  es una distribución beta con parámetros  $(1, 3)$ , es decir,  $\pi(p|0) = 3(1-p)^2$ , para  $0 < p < 1$ . Por consiguiente, si se desea un intervalo de confianza de 95 %, se necesita encontrar  $\theta_L$  y  $\theta_U$  en la expresión (6.49), tales que:

$$0,025 = \int_0^{\theta_L} 3(1-p)^2 dp = 1 - (1 - \theta_L)^3$$

y

$$0,025 = \int_{\theta_U}^1 3(1-p)^2 dp = (1 - \theta_U)^3.$$

Las soluciones de estas ecuaciones dan como resultado:

$$\theta_L = 0,0084 \quad \text{y} \quad \theta_U = 0,7076.$$

Por lo tanto, la probabilidad de que  $p \in (0,0084; 0,7076)$  es de 95 %.  $\square$

**Nota 6.11.** Para una población normal y el caso normal mostrado en el teorema 6.3, dado que la media a posteriori  $\tilde{\mu}_n$  es el estimador de Bayes de la media poblacional  $\mu$ , el teorema del límite central bayesiano que aparece en [10, p. 224] indica que, en condiciones de regularidad, la **fdp** a posteriori se aproxima por la **fdp** normal y, por lo tanto, el intervalo de confianza bayesiano para  $\mu$  de  $100(1 - \alpha)$  % está dado por:

$$IB_{(1-\alpha)}(\mu) = (\tilde{\mu}_n - z_{1-\alpha/2}\tilde{\sigma}_n, \tilde{\mu}_n + z_{1-\alpha/2}\tilde{\sigma}_n), \quad (6.50)$$

donde  $\tilde{\mu}_n$  y  $\tilde{\sigma}_n$  son dadas en (6.46), el cual se centra en la media a posteriori y contiene  $100(1 - \alpha)$  % de la probabilidad a posteriori. Los límites también se pueden determinar a partir de (6.49) usando la **fdp** a posteriori dada en (6.47).

Al igual que con el teorema del límite central clásico, la aproximación mejora a medida que aumenta el tamaño de muestra.

**Ejemplo 6.27.** Según [89, p. 716], una empresa de equipo eléctrico fabrica bombillas con una duración distribuida aproximadamente normal y una desviación estándar de 100 horas. La experiencia a priori nos conduce a creer que  $\mu$  es un valor de una variable aleatoria normal con una media  $\vartheta = 800$  horas y una desviación

estándar  $\tau = 10$  horas. Si una muestra aleatoria de 25 bombillas tiene una duración promedio de 780 horas, calcule un intervalo bayesiano de 95 % para  $\mu$ .

*Solución.* De acuerdo con el teorema 6.3, para  $n = 25$ , la distribución *a posteriori* de la media también es una distribución normal con media

$$\tilde{\mu}_{25} = \frac{(100)^2(800) + (25)(10)^2(780)}{(25)(10)^2 + (100)^2} = 796$$

y varianza

$$\tilde{\sigma}_{25}^2 = \frac{(10)^2(100)^2}{(25)(10)^2 + (100)^2} = 80.$$

El intervalo bayesiano de 95 % para  $\mu$  es dado por la expresión (6.50):

$$IB_{0,95}(\mu) = (796 - 1,96\sqrt{80}; 796 + 1,96\sqrt{80}),$$

o

$$IB_{0,95}(\mu) = (778,5; 813,5).$$

Por lo tanto, estamos 95 % seguros de que  $\mu$  estará entre 778,5 y 813,5. Por otro lado, si se desconoce la información *a priori* acerca de  $\mu$ , se emplea (6.30) para construir el intervalo de confianza clásico de 95 %.

$$IC_{0,95}(\mu) = \left( 780 - (1,96)\frac{100}{\sqrt{25}}; 780 + (1,96)\frac{100}{\sqrt{25}} \right),$$

o  $IC_{0,95}(\mu) = (740,8; 819,2)$ , el cual se ve que es más amplio que el intervalo bayesiano correspondiente.  $\checkmark$

## 6.7. Tamaño de muestra simple bajo normalidad

Encontrar el tamaño de muestra es muy relevante ya que si es pequeño puede ser poco significativa y si es muy grande se pierden recursos.

### 6.7.1. Tamaño de muestra para estimar la media con varianza conocida

El tamaño de muestra más simple bajo el supuesto de normalidad se obtiene de la longitud media máxima deseada, denotada por  $E$ , para el intervalo de

confianza para una media, cuando la varianza es conocida, es decir,

$$\begin{aligned}
 IC_{(1-\alpha)}(\mu) &= \left( \bar{X}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) \\
 L_I &= \bar{X}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} - \left( \bar{X}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = 2z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \\
 \frac{L_I}{2} &= \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}} = E.
 \end{aligned} \tag{6.51}$$

Despejando  $n$  de (6.51) se tiene que:

$$\begin{aligned}
 \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}} &= E \Rightarrow z_{1-\frac{\alpha}{2}} \sigma = E\sqrt{n} \\
 \frac{z_{1-\frac{\alpha}{2}} \sigma}{E} &= \sqrt{n} \Rightarrow n = \left( \frac{z_{1-\frac{\alpha}{2}} \sigma}{E} \right)^2,
 \end{aligned}$$

luego,

$$n = \left\lceil \left[ \frac{\sigma^2}{E^2} z_{1-\frac{\alpha}{2}}^2 \right] \right\rceil. \tag{6.52}$$

Si el muestreo es sin reemplazo, se introduce el factor de corrección por población finita, en este caso (6.51) queda

$$E = \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}},$$

donde  $N$  es el tamaño de la población, despejando  $n$  se obtiene

$$\begin{aligned}
 E &= \frac{z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{N-1}} \sqrt{\frac{N-n}{n}} \Rightarrow \frac{\sqrt{N-1} E}{z_{1-\frac{\alpha}{2}} \sigma} = \sqrt{\frac{N}{n} - 1} \\
 n &= \left\lceil N \left[ \frac{(N-1)E^2}{\sigma^2} z_{1-\frac{\alpha}{2}}^{-2} + 1 \right]^{-1} \right\rceil.
 \end{aligned} \tag{6.53}$$

**Ejemplo 6.28.** De acuerdo con [19, p. 298], una tienda de donas está interesada en estimar su volumen de ventas diarias. Supóngase que el valor de la desviación estándar es 50. Si el volumen de ventas diarias se encuentra aproximado por una distribución normal, ¿cuál debe ser el tamaño de la muestra para que con una probabilidad de 0,95 la media muestral se encuentre a no más de 20 del verdadero volumen de ventas promedio?

*Solución.* De la información dada:

$$\begin{aligned}\sigma &= 50 & E &= 20 \\ 1 - \alpha &= 0,95 & \alpha &= 0,05 \\ z_{1-\frac{\alpha}{2}} &= z_{0,975} = 1,96 \\ n &= \frac{1,96^2}{20^2} \times 50^2 = 24,01.\end{aligned}$$

Concluyendo que se necesitan al menos 25 días de muestra. ☑

## 6.7.2. Tamaño de muestra para la estimación de una proporción

En el caso de la estimación de una proporción, se tendría del ejemplo 6.18, que la longitud media máxima deseada para el intervalo de confianza para una proporción dado en (6.32) es:

$$\begin{aligned}L_I &= \hat{p}_n + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} - \left[ \hat{p}_n - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} \right] \\ &= 2 \frac{z_{1-\frac{\alpha}{2}} \sqrt{\hat{p}_n(1-\hat{p}_n)}}{\sqrt{n}} \\ \frac{L_I}{2} &= \frac{z_{1-\frac{\alpha}{2}} \sqrt{\hat{p}_n(1-\hat{p}_n)}}{\sqrt{n}} = E.\end{aligned}\quad (6.54)$$

Despejando  $n$  de (6.54) se tiene que:

$$\begin{aligned}\frac{z_{1-\frac{\alpha}{2}} \sqrt{\hat{p}_n(1-\hat{p}_n)}}{\sqrt{n}} &= E \quad \Rightarrow \quad z_{1-\frac{\alpha}{2}} \sqrt{\hat{p}_n(1-\hat{p}_n)} = E\sqrt{n} \\ \frac{z_{1-\frac{\alpha}{2}} \sqrt{\hat{p}_n(1-\hat{p}_n)}}{E} &= \sqrt{n} \quad \Rightarrow \quad n = \left( \frac{z_{1-\frac{\alpha}{2}} \sqrt{\hat{p}_n(1-\hat{p}_n)}}{E} \right)^2 \\ n &= \left\lceil \left[ \frac{\hat{p}_n(1-\hat{p}_n)}{E^2} z_{1-\frac{\alpha}{2}}^2 \right] \right\rceil.\end{aligned}\quad (6.55)$$

Si no se conoce  $p$  poblacional ni se tiene una estimación previa del mismo, se toma  $p = 0,5$ , ya que una vez fijos los valores de  $\alpha$  y  $E$ , el valor máximo de  $n$  se tendrá cuando  $p(1-p)$  sea máximo y esto ocurre si  $p = 0,5$ .

Si el muestreo es sin reemplazo, se introduce el factor de corrección por población finita, en este caso (6.54) queda:

$$E = \frac{z_{1-\frac{\alpha}{2}} \sqrt{\hat{p}_n(1-\hat{p}_n)}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}},$$

donde  $N$  es el tamaño de la población, despejando  $n$  se obtiene que:

$$E = \frac{z_{1-\frac{\alpha}{2}} \sqrt{\hat{p}_n(1-\hat{p}_n)}}{\sqrt{N-1}} \sqrt{\frac{N-n}{n}} \Rightarrow \frac{\sqrt{N-1} E}{z_{1-\frac{\alpha}{2}} \sqrt{\hat{p}_n(1-\hat{p}_n)}} = \sqrt{\frac{N}{n} - 1}$$

$$n = \left\lceil N \left[ \frac{(N-1)E^2}{\hat{p}_n(1-\hat{p}_n)} z_{1-\frac{\alpha}{2}}^{-2} + 1 \right]^{-1} \right\rceil. \quad (6.56)$$

**Ejemplo 6.29.** De acuerdo con [19, p. 301], las compañías de auditoría generalmente seleccionan una muestra aleatoria de los clientes de un banco y verifican los balances contables reportados por el banco. Si una compañía de este tipo se encuentra interesada en estimar la proporción de cuentas para las cuales existe una discrepancia entre el cliente y el banco, ¿cuántas cuentas deberán seleccionarse de manera tal que con una confiabilidad del 99% la proporción muestral se encuentre a no más de 0,02 unidades de la proporción real?

*Solución.* De la información dada:

$$1 - \alpha = 0,99 \quad \alpha = 0,01$$

$$z_{1-\frac{\alpha}{2}} = z_{0,995} = 2,575$$

$$E = 0,02,$$

como  $p$  es desconocida entonces se considera  $p = 0,50$ , luego,

$$n = \frac{2,575^2}{0,02^2} \times 0,50(1 - 0,50) = 4144,14.$$

Por consiguiente, es necesario tomar una muestra de 4145 cuentas.  $\square$

## 6.8. Ejercicios

1. Considere a  $\theta_1$  como una constante conocida, encuentre un intervalo de confianza para  $\theta_2$ , basado en una muestra aleatoria  $X_1, X_2, \dots, X_n$ ,

de una población con **fdp** Pareto [62, p. 184], es decir,

$$f_X(x) = \frac{\theta_2 \theta_1^{\theta_2}}{x^{\theta_2+1}} \quad x > \theta_1.$$

2. Bajo el escenario presentado en el ejemplo 6.15, y usando la cantidad pivotal presentada en el mismo, mostrar que el intervalo de confianza del  $100(1 - \alpha)\%$  para  $\theta$  de mínima longitud es dado por:

$$IC_{(1-\alpha)}(\theta) = \left( \bar{X}_n - \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}; n-1}, \bar{X}_n + \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}; n-1} \right).$$

3. Los pesos en onzas de una muestra de 15 paquetes de cereal son:

16,13	16,02	15,90	15,83	16	15,80	16,02	16,03
15,99	15,83	15,91	16,04	16	16,03	16,15	

Calcular intervalos de confianza a los niveles de (a) 80%, (b) 90%, (c) 95% y (d) 99%.

- a) Para el promedio del peso en onzas de los paquetes de cereal.  
 b) Para la varianza del peso en onzas de los paquetes de cereal.  
 c) Suponiendo que la desviación estándar poblacional es 0,06, realice de nuevo el ítem 3a.
4. Para el ejercicio 3c, ¿cuál debe ser el tamaño de la muestra para que con una probabilidad de 0,95 la media muestral esté a no más de 0,02 onzas del verdadero valor del peso de los paquetes de cereal?
5. En una encuesta realizada en la localidad de Los Mártires sobre instalación de gas natural y teléfono a 400 viviendas, se encontró que 234 tenían gas natural y 333 teléfono. Encuentre la estimación por intervalo para la proporción de viviendas que tienen instalación de:
- a) Gas natural a los niveles de confianza del 95% y 99%.  
 b) Teléfono a los niveles de confianza del 95% y 99%.
6. Si no se conocen datos de la localidad de Los Mártires sobre instalación de gas natural, ¿cuántas viviendas se deben seleccionar para que con una confiabilidad del 99% la proporción de viviendas con instalación de gas natural esté a no más de 0,05 unidades de la proporción real?

7. Se usan dos máquinas para llenar botellas de plástico con un volumen neto de 16 onzas. Supóngase que el volumen de llenado sigue una **fdp** normal. El departamento de control de calidad opina que ambas máquinas llenan el mismo volumen neto, sin importar si este volumen es 16 onzas o no [64, p. 385]. En un experimento se toma una muestra aleatoria de 10 productos de cada máquina y los resultados son:

$$\begin{aligned}\bar{x}_1 &= 16,0132 & s_1^2 &= 0,030480 \\ \bar{x}_2 &= 16,0070 & s_2^2 &= 0,017670.\end{aligned}$$

¿Existen evidencias para considerar que las dos máquinas llenan el mismo volumen neto? Utilice un intervalo de confianza al 90 % para justificar su respuesta.

8. Los contenidos de elementos esenciales en tomates frescos y enlatados se determinaron mediante el método de espectro-fotometría de absorción atómica. El contenido de cobre en tomates frescos en comparación con el que los mismos tomates registraron después de ser enlatados se muestra en la tabla 6.4. Encuentre el intervalo de confianza del 90 % para la diferencia en el contenido promedio de cobre de tomates frescos y enlatados, suponiendo que la distribución de las diferencias es normal.

Tabla 6.4. Contenido de cobre en tomates

Par	Tomates frescos	Tomates enlatados
1	0,066	0,085
2	0,079	0,088
3	0,069	0,091
4	0,076	0,096
5	0,071	0,093
6	0,087	0,095
7	0,071	0,079
8	0,073	0,078
9	0,067	0,065
10	0,062	0,068

9. Una muestra de 180 tuercas de una cierta máquina probó que 15 eran defectuosas, mientras que otra muestra de 110 tuercas de otra máquina dió 14 defectuosas. ¿Existen evidencias para creer que la proporción de tuercas defectuosas difiere entre las máquinas? Usar un intervalo de confianza del 95 % para obtener la conclusión.





El planteamiento de cualquier estudio estadístico tradicionalmente tiene por finalidad comprobar si las premisas que se tienen sobre la población en estudio se pueden aceptar como correctos o se deben considerar falsos. Este capítulo tiene como finalidad presentar los conceptos y aplicaciones de la teoría de las pruebas de hipótesis, la cual está relacionada a la de los intervalos de confianza, por ello, se darán las respectivas interpretaciones para distinguir claramente entre los dos ámbitos.

## 7.1. Elementos básicos de pruebas de hipótesis

**DEFINICIÓN 7.1.** Según [20, p. 373], una hipótesis es una aseveración o conjetura acerca de un parámetro poblacional.

**DEFINICIÓN 7.2.** Según [20, p. 373], las dos hipótesis complementarias de un problema de prueba de hipótesis son llamadas **la hipótesis nula** y **la hipótesis alterna**. Se denotan por  $H_0$  y  $H_1$  (o  $H_a$ ), respectivamente. Si  $\theta$  denota un parámetro poblacional sobre el cual se desea plantear una hipótesis, se tiene que  $H_0$  y  $H_1$  (o  $H_a$ ) son dadas por:

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_0^c,$$

donde  $\Theta_0$  es un subconjunto del espacio de parámetros  $\Theta$ .

**DEFINICIÓN 7.3.** Según [20, p. 374], un procedimiento de prueba de hipótesis o una prueba de hipótesis es una regla que especifica:

1. Para cuáles valores muestrales la decisión es aceptar (no rechazar)  $H_0$  como verdadera.
2. Para cuáles valores muestrales  $H_0$  es rechazada y  $H_1$  es aceptada (no rechazada) como verdadera.

El subconjunto del espacio muestral para el cual  $H_0$  será rechazada se llama **región de rechazo o región crítica**. El complemento de la región de rechazo se llama **región de aceptación** (no rechazo).

**Nota 7.1.** Hay dos estructuras posibles para  $\Theta_0$ : Si  $\Theta_0$  consiste de solo un punto, se dice que la hipótesis nula es simple. Cuando  $\Theta_0$  contiene más de un punto, se dice que la hipótesis nula es compuesta (ver [12, p. 215]).

## 7.2. Evaluación de tests

La eficacia de un test es medida por la frecuencia con que se hacen juzgamientos correctos cuando se usa el test [12, p. 216]. Hay dos tipos de error que se pueden cometer cuando se utiliza un test:

1. Que se rechace la hipótesis  $H_0$  cuando en realidad se debería aceptar.
2. Que se acepte (no se rechace) la hipótesis  $H_0$  cuando en realidad se debería rechazar.

Note que no se pueden cometer los dos tipos de error simultáneamente. Al error presentado en el ítem 1 se le llama *error tipo I* y al error presentado en el ítem 2 se le llama *error tipo II*. Ver tabla 7.1.

Tabla 7.1. Casos posibles para probar una hipótesis estadística

Decisión \ Realidad	$H_0$ es verdadera	$H_1$ es verdadera
	Aceptar $H_0$	No hay error
Rechazar $H_0$	<i>Error tipo I</i>	No hay error

Fuente: adaptada de [89, p. 323].

**DEFINICIÓN 7.4.** Según [20, p. 383], la *función de potencia* de una prueba de hipótesis con región de rechazo  $R$  es una función de  $\theta$  definida por:

$$\beta(\theta) = P_\theta(X \in R). \quad (7.1)$$

**Nota 7.2.** Una función de potencia *ideal* sería la que tome el valor 0 para todo  $\theta \in \Theta_0$  y el valor de 1 para todo  $\theta \in \Theta_0^c$ .

**Ejemplo 7.1.** De acuerdo con [20, p. 383], dada  $X$  una variable aleatoria tal que  $X \sim \text{Bin}(5, \theta)$ . Se define el test  $H_0 : \theta \leq \frac{1}{2}$  vs.  $H_1 : \theta > \frac{1}{2}$ . Determine la función potencia si se considera un test que rechaza  $H_0$  si y solo se obtienen solo éxitos, es decir,  $X = 5$ .

*Solución.* En este caso, la función de potencia es dada por (ver figura 7.1):

$$\beta(\theta) = P_\theta(X \in R) = P_\theta(X = 5) = \binom{5}{5} \theta^5 (1 - \theta)^{5-5} = \theta^5.$$

Note que si  $\theta \leq \frac{1}{2}$ , es decir, si la hipótesis nula es verdadera, la función de potencia es menor a 0,03125, es decir,  $\beta(\theta) \leq 0,03125$ .  $\square$

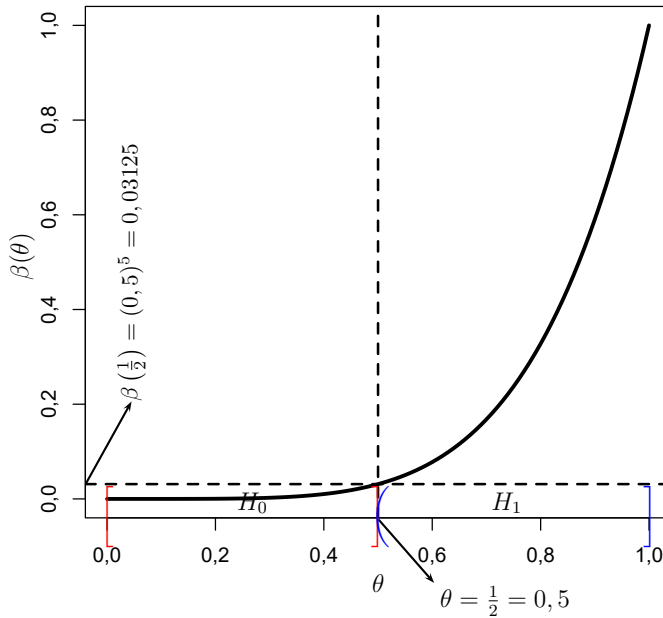


Figura 7.1. Función de potencia  $\beta(\theta)$ , ejemplo 7.1

**DEFINICIÓN 7.5.** Según [20, p. 385], un test con función de potencia  $\beta(\theta)$  es un test de tamaño  $\alpha$ <sup>40</sup>, con  $0 \leq \alpha \leq 1$ , si  $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$ .

**Ejemplo 7.2.** Considerando la información del ejemplo 7.1,

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \beta(1/2) = 0,03125,$$

luego el test que rechaza  $H_0$  si solo se obtienen éxitos es un test de tamaño 0,03125.

**DEFINICIÓN 7.6.** Según [20, p. 385], un test con función de potencia  $\beta(\theta)$  es un test al nivel  $\alpha$ , con  $0 \leq \alpha \leq 1$ , si  $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$ .

**DEFINICIÓN 7.7.** La **potencia** de un test es la probabilidad de rechazar la hipótesis nula cuando la hipótesis alternativa es verdadera [12, p. 217].

En otras palabras, la potencia es una función de  $\theta$  sobre  $\Theta_0^c$ . Note que la potencia y la probabilidad de cometer error tipo I están contenidas en la función de potencia. Ver figura 7.2. En general, se denota a la probabilidad de cometer error tipo II por  $\beta$ , luego la potencia será  $1 - \beta$ . Ver tabla 7.2.

<sup>40</sup>El valor  $\alpha$  se llama nivel de significancia [54, p. 57].

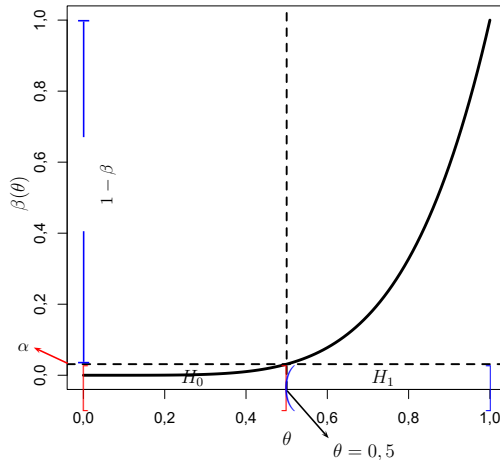


Figura 7.2. Función de potencia  $\beta(\theta)$

Tabla 7.2. Decisiones en la comprobación de hipótesis

Decisión \ Realidad	$H_0$ es verdadera	$H_0$ es falsa
	Aceptar $H_0$	No hay error $1 - \alpha$
Rechazar $H_0$	Error tipo I $\alpha$	No hay error $1 - \beta$

Fuente: adaptada de [64, p. 197].

**DEFINICIÓN 7.8.** Según [20, p. 387], un test con función de potencia  $\beta(\theta)$  es *insesgado* si  $\beta(\theta') \geq \beta(\theta'')$  para cada  $\theta' \in \Theta_0^c$  y  $\theta'' \in \Theta_0$ .

**DEFINICIÓN 7.9.** El *p-valor* es  $\alpha(T(\mathbf{X}))$  [12, p. 222].

En otras palabras, el *p-valor* o tamaño observado o probabilidad de significancia del test es un estadístico que se define como el nivel de significancia más pequeño  $\alpha$ , al cual un experimentador usando el estadístico  $T(\mathbf{X})$  rechazaría la hipótesis nula, basándose en los resultados observados  $x$  [12, p. 221]. Por tanto, para un  $\alpha$  dado por el investigador, si  $\alpha < p$ -valor la hipótesis nula no se rechaza y si  $\alpha > p$ -valor la hipótesis nula se rechaza.

## 7.3. Test uniformemente más potente

**DEFINICIÓN 7.10.** De acuerdo con [20, p. 388], dada  $\mathcal{C}$  una clase de tests<sup>41</sup> para probar  $H_0 : \theta \in \Theta_0$  vs.  $H_1 : \theta \in \Theta_0^c$ . Un test en la clase  $\mathcal{C}$ , con función de potencia  $\beta(\theta)$ , es un test **uniformemente**<sup>42</sup> **más potente (UMP)** si  $\beta(\theta) \geq \beta'(\theta)$  para cada  $\theta \in \Theta_0^c$  y cada  $\beta'(\theta)$  que sea una función potencia de un test en la clase  $\mathcal{C}$ .

**Nota 7.3.** Un test es **insesgado uniformemente más potente (UMP)** si es uniformemente más potente (UMP) en la clase de todos los test insesgados. De hecho un test UMP es siempre insesgado (ver [76, p. 349]).

**Nota 7.4.** Según [76, p. 329], si  $\Theta_0^c$  consiste solo de un punto, un test UMP se dice que es un test más potente (MP).

### 7.3.1. Test de razón de verosimilitud

**DEFINICIÓN 7.11.** Según [20, p. 375], el estadístico llamado test de razón de verosimilitud (LRT) para probar

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_0^c$$

es dado de forma similar, como en la expresión (6.15), así:

$$\lambda(x) = \frac{\sup_{\Theta_0} L(\theta; x)}{\sup_{\Theta} L(\theta; x)}. \quad (7.2)$$

Un LRT es cualquier test tal que dada una cota (ver nota B.8),  $c \in (0, 1)$ , su región de rechazo es de la forma

$$\{x : \lambda(x) \leq c\}. \quad (7.3)$$

Sean dos EMV de  $\theta$ , digamos  $\hat{\theta}(< \infty)$  y  $\hat{\theta}_0(< \infty)$ , donde  $\hat{\theta}$  se obtiene sin ninguna restricción sobre el espacio de parámetros y  $\hat{\theta}_0$  se obtiene haciendo una maximización restringida, considerando que el espacio de parámetros es  $\Theta_0$ , es decir,  $\hat{\theta}_0$  es el valor de  $\theta \in \Theta_0$  que maximiza  $L(\theta; x)$ . Entonces el estadístico LRT es

$$\lambda(x) = \frac{L(\hat{\theta}_0; x)}{L(\hat{\theta}; x)}.$$

Por tanto, para establecer el LRT es necesario encontrar  $L(\hat{\theta}_0; x)$  y  $L(\hat{\theta}; x)$ .

<sup>41</sup>La clase  $\mathcal{C}$  es la clase de todos los tests de nivel  $\alpha$ , es decir,  $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$ .

<sup>42</sup>El adverbio *uniformemente* se refiere a todos los valores alternativos de  $\theta$  (ver [65, p.422]).

**Ejemplo 7.3.** De acuerdo con [20, p. 375], dadas  $X_1, X_2, \dots, X_n$  variables aleatorias iid tal que  $X_1 \sim N(\theta, 1)$ . Se consideran las hipótesis

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0,$$

donde  $\theta_0$  es una constante preestablecida. Determine:

1. La región de rechazo.
2. La función de potencia.

**Solución.** En este caso,  $\Theta_0 = \{\theta_0\}$  y

1. Para determinar la región de rechazo:

$$\begin{aligned} \sup_{\Theta_0} L(\theta|\mathbf{x}) &= L(\theta_0|\mathbf{x}) = \prod_{i=1}^n f(x_i; \theta_0) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \theta_0)^2}{2}} \\ &= (2\pi)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta_0)^2\right\}. \end{aligned}$$

Por otro lado  $\sup_{\Theta} L(\theta|\mathbf{x})$  se tiene cuando  $\theta = \bar{x}$ , ver ejemplo 4.7. En este caso, el **LRT** es dado por (7.2):

$$\begin{aligned} \lambda(\mathbf{x}) &= \frac{\sup_{\Theta_0} L(\theta; \mathbf{x})}{\sup_{\Theta} L(\theta; \mathbf{x})} = \frac{(2\pi)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta_0)^2\right\}}{(2\pi)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2\right\}} \\ &= \exp\left\{-\frac{1}{2} \left(\sum_{i=1}^n (x_i - \theta_0)^2 - \sum_{i=1}^n (x_i - \bar{x})^2\right)\right\} \\ &= \exp\left\{-\frac{1}{2} \left(\sum_{i=1}^n x_i^2 - 2\theta_0 n \bar{x} + \theta_0^2 n - \sum_{i=1}^n x_i^2 + 2n \bar{x}^2 - n \bar{x}^2\right)\right\} \\ &= \exp\left\{-\frac{n}{2} (\theta_0^2 - 2\theta_0 \bar{x} + \bar{x}^2)\right\} = \exp\left\{-\frac{n}{2} (\bar{x} - \theta_0)^2\right\}. \end{aligned}$$

Luego, para  $0 < c < 1$ , la región de rechazo dada en (7.3) es:

$$\begin{aligned} \{\mathbf{x} : \lambda(\mathbf{x}) \leq c\} &= \left\{\mathbf{x} : \exp\left\{-\frac{n}{2} (\bar{x} - \theta_0)^2\right\} \leq c\right\} \quad \text{tomando logaritmo} \\ &= \left\{\mathbf{x} : \frac{(\bar{x} - \theta_0)^2}{-2/n} \leq \ln(c)\right\} = \left\{\mathbf{x} : (\bar{x} - \theta_0)^2 \geq \frac{-2}{n} \ln(c)\right\} \\ &= \left\{\mathbf{x} : |\bar{x} - \theta_0| \geq \sqrt{\frac{-2 \ln(c)}{n}}\right\} \end{aligned} \quad (7.4)$$

y corresponde a un test **LRT** para probar la hipótesis  $H_0 : \theta = \theta_0$ .

2. Su correspondiente función de potencia es dada por (7.1):

$$\begin{aligned}
 \beta(\theta) &= P_{\theta}(X \in R) = P_{\theta}\left(|\bar{X} - \theta_0| \geq \sqrt{\frac{-2 \ln(c)}{n}}\right) \\
 &= P_{\theta}\left(\bar{X} - \theta_0 \leq -\sqrt{\frac{-2 \ln(c)}{n}}\right) + P_{\theta}\left(\bar{X} - \theta_0 \geq \sqrt{\frac{-2 \ln(c)}{n}}\right) \\
 &= P_{\theta}\left(\bar{X} \leq \theta_0 - \sqrt{\frac{-2 \ln(c)}{n}}\right) + P_{\theta}\left(\bar{X} \geq \theta_0 + \sqrt{\frac{-2 \ln(c)}{n}}\right) \\
 &= P_{\theta}\left(\bar{X} - \theta \leq -\sqrt{\frac{2 \ln(c)}{-n}} + \theta_0 - \theta\right) + P_{\theta}\left(\bar{X} - \theta \geq \sqrt{\frac{2 \ln(c)}{-n}} + \theta_0 - \theta\right) \\
 &= P_{\theta}\left(\frac{\bar{X} - \theta}{\frac{1}{\sqrt{n}}} \leq \frac{-\sqrt{\frac{2 \ln(c)}{-n}} + \theta_0 - \theta}{\frac{1}{\sqrt{n}}}\right) + P_{\theta}\left(\frac{\bar{X} - \theta}{\frac{1}{\sqrt{n}}} \geq \frac{\sqrt{\frac{2 \ln(c)}{-n}} + \theta_0 - \theta}{\frac{1}{\sqrt{n}}}\right) \\
 &= \left[1 - \Phi\left(\sqrt{-2 \ln(c)} - \frac{\theta_0 - \theta}{1/\sqrt{n}}\right)\right] + \left[1 - \Phi\left(\sqrt{-2 \ln(c)} + \frac{\theta_0 - \theta}{1/\sqrt{n}}\right)\right],
 \end{aligned}$$

donde  $\Phi(\cdot)$  es la fda de la variable aleatoria  $Z \sim N(0, 1)$ . Bajo  $H_1$  verdadera, se tiene que:

$$\beta(\theta) = 2 - \Phi\left(\sqrt{-2 \ln(c)} - \frac{\theta_0 - \theta}{1/\sqrt{n}}\right) - \Phi\left(\sqrt{-2 \ln(c)} + \frac{\theta_0 - \theta}{1/\sqrt{n}}\right). \quad (7.5)$$

La figura 7.3 es la representación de  $\beta(\theta)$ .

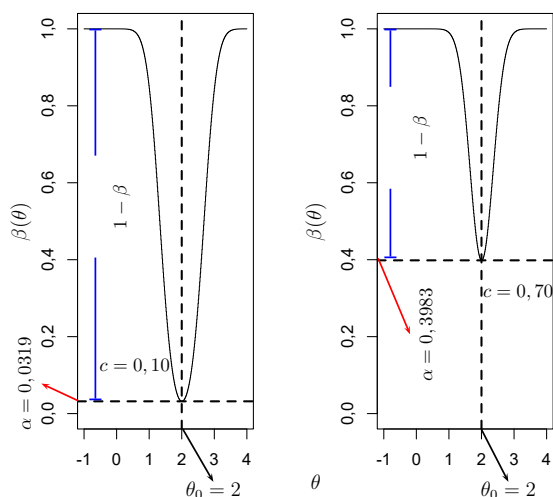


Figura 7.3. Función de potencia  $\beta(\theta)$  para  $\theta_0 = 2$ ,  $n = 10$  y considerando dos valores para  $c$ ,  $c = 0,10$  y  $c = 0,70$ , ejemplo 2

Para analizar el comportamiento de esta función se usan los criterios de las derivadas. Como la derivada de la **fda** es la **fdp**, entonces la derivada respecto a  $\theta$  es:

$$\frac{\partial}{\partial \theta} \beta(\theta) = -\sqrt{n} \varphi \left[ \sqrt{-2 \ln(c)} - \frac{\theta_0 - \theta}{1/\sqrt{n}} \right] + \sqrt{n} \varphi \left[ \sqrt{-2 \ln(c)} + \frac{\theta_0 - \theta}{1/\sqrt{n}} \right],$$

donde  $\varphi(\cdot)$  es la **fdp** de la variable aleatoria  $Z \sim N(0, 1)$ , igualando a cero para obtener el punto crítico

$$\begin{aligned} \cancel{\sqrt{n}} \varphi \left( \sqrt{-2 \ln(c)} + \frac{\theta_0 - \theta}{1/\sqrt{n}} \right) &= \cancel{\sqrt{n}} \varphi \left( \sqrt{-2 \ln(c)} - \frac{\theta_0 - \theta}{1/\sqrt{n}} \right) \\ \frac{\exp \left\{ -\frac{1}{2} \left( \sqrt{-2 \ln(c)} + \frac{\theta_0 - \theta}{1/\sqrt{n}} \right)^2 \right\}}{\cancel{\sqrt{2\pi}}} &= \frac{\exp \left\{ -\frac{1}{2} \left( \sqrt{-2 \ln(c)} - \frac{\theta_0 - \theta}{1/\sqrt{n}} \right)^2 \right\}}{\cancel{\sqrt{2\pi}}}, \end{aligned}$$

tomando logaritmo natural a ambos lados se obtiene:

$$\begin{aligned} -\frac{1}{2} \left( \sqrt{-2 \ln(c)} + \frac{\theta_0 - \theta}{1/\sqrt{n}} \right)^2 &= -\frac{1}{2} \left( \sqrt{-2 \ln(c)} - \frac{\theta_0 - \theta}{1/\sqrt{n}} \right)^2 \\ \frac{2}{2} \sqrt{-2 \ln(c)} \left( 2 \frac{\theta - \theta_0}{1/\sqrt{n}} \right) &= 0 \quad \rightarrow \quad \theta = \theta_0. \end{aligned}$$

Puesto que la derivada de  $\varphi(\cdot)$  satisface que:

$$\frac{\partial}{\partial y} \varphi(y) = -y \varphi(y) \frac{\partial y}{\partial \theta},$$

se verifica que  $\beta(\theta)$  tiene un mínimo en  $\theta = \theta_0$ . Por otra parte, la función potencia será creciente cuando su derivada tenga signo positivo, es decir,

$$\begin{aligned} \varphi \left( \sqrt{-2 \ln(c)} + \frac{\theta_0 - \theta}{1/\sqrt{n}} \right) &> \varphi \left( \sqrt{-2 \ln(c)} - \frac{\theta_0 - \theta}{1/\sqrt{n}} \right) \\ -\frac{1}{2} \left( \sqrt{-2 \ln(c)} + \frac{\theta_0 - \theta}{1/\sqrt{n}} \right)^2 &> -\frac{1}{2} \left( \sqrt{-2 \ln(c)} - \frac{\theta_0 - \theta}{1/\sqrt{n}} \right)^2 \\ \frac{2}{2} \sqrt{-2 \ln(c)} \left( 2 \frac{\theta - \theta_0}{1/\sqrt{n}} \right) &> 0 \quad \rightarrow \quad \theta > \theta_0. \end{aligned}$$

Por este motivo el test es inesgado. ☑

**Ejemplo 7.4.** De acuerdo con [20, p. 376], si  $X_1, X_2, \dots, X_n$  son variables aleatorias *iid* con fdp:

$$f_{X_1}(x_1; \theta) = e^{-(x_1 - \theta)} I_{[\theta, +\infty)}(x_1),$$

donde  $-\infty < \theta < \infty$ . Se consideran las hipótesis:

$$H_0 : \theta \leq \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0,$$

donde  $\theta_0$  es una constante preestablecida. Determine la región de rechazo.

*Solución.* En el ejemplo 6.10 se estableció la función de verosimilitud,

$$L(\theta|x) = \exp \{n(\theta - \bar{x})\} I_{(-\infty, x_{(1)}]}(\theta) = \begin{cases} \exp \{n(\theta - \bar{x})\} & \text{si } \theta \leq x_{(1)} \\ 0 & \text{si } \theta > x_{(1)}. \end{cases}$$

Para encontrar el  $\sup_{\Theta} L(\theta|x)$  se tiene que:

$$\sup_{\Theta} L(\theta|x) = \sup_{\Theta} \left[ \exp \{n(\theta - \bar{x})\} I_{(-\infty, x_{(1)}]}(\theta) \right].$$

Notando que  $L(\theta|x)$  toma su valor máximo cuando  $\theta = x_{(1)}$ , ver figura 7.4, se tiene que:

$$\sup_{\Theta} L(\theta|x) = L(x_{(1)}|x) = \exp \{n(x_{(1)} - \bar{x})\}.$$

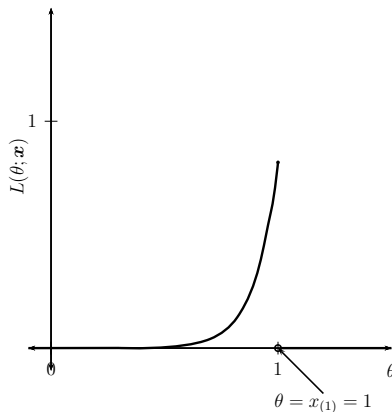


Figura 7.4.  $\theta$  vs.  $L(\theta|x)$  cuando  $x_{(1)} = 1$ ,  $\bar{x} = 1,02$  y  $n = 10$ , ejemplo 7.4

Por otro lado, para la región parametral  $\Theta_0 = (-\infty, \theta_0]$  se tiene:

$$\sup_{\Theta_0} L(\theta|\mathbf{x}) = \sup_{\Theta_0} \left[ \exp \{n(\theta - \bar{x})\} I_{(-\infty, x_{(1)}]}(\theta) \right].$$

Puesto que se desea maximizar  $L(\theta|\mathbf{x})$  sobre  $\theta \leq \theta_0$ , es decir, la hipótesis  $H_0$ . Si  $\theta_0 < x_{(1)}$  entonces, ver figura 7.5,

$$\sup_{\Theta_0} L(\theta|\mathbf{x}) = L(\theta_0|\mathbf{x}) = \exp \{n(\theta_0 - \bar{x})\}.$$

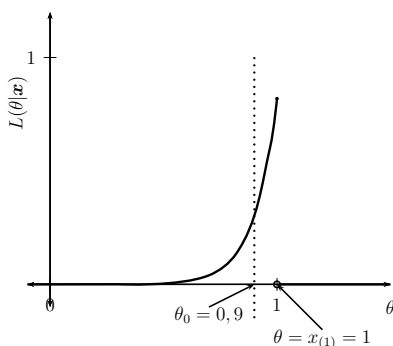


Figura 7.5.  $\theta$  vs.  $L(\theta|\mathbf{x})$  cuando  $x_{(1)} = 1$ ,  $\bar{x} = 1,02$  y  $n = 10$ .  $\theta_0 = 0,9$ , ejemplo 7.4

Si  $\theta_0 \geq x_{(1)}$  entonces, ver figura 7.6,

$$\sup_{\Theta_0} L(\theta|\mathbf{x}) = L(x_{(1)}|\mathbf{x}) = \exp \{n(x_{(1)} - \bar{x})\}.$$

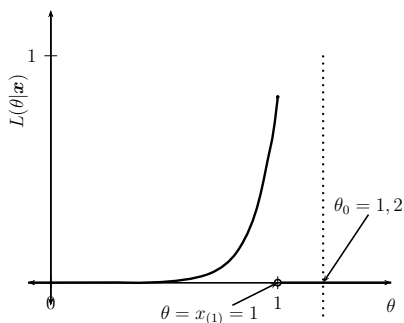


Figura 7.6.  $\theta$  vs.  $L(\theta|\mathbf{x})$  cuando  $x_{(1)} = 1$ ,  $\bar{x} = 1,02$  y  $n = 10$ .  $\theta_0 = 1,2$ . Ejemplo 7.4

Luego, el **LRT** es dado por (7.2):

$$\begin{aligned} \lambda(\mathbf{x}) &= \frac{\sup_{\Theta_0} L(\theta; \mathbf{x})}{\sup_{\Theta} L(\theta; \mathbf{x})} = \begin{cases} \frac{\exp\{n(\theta_0 - \bar{x})\}}{\exp\{n(x_{(1)} - \bar{x})\}} & \text{si } \theta_0 < x_{(1)} \\ \frac{\exp\{n(x_{(1)} - \bar{x})\}}{\exp\{n(x_{(1)} - \bar{x})\}} & \text{si } \theta_0 \geq x_{(1)} \end{cases} \\ &= \begin{cases} e^{n(\theta_0 - x_{(1)})} & \text{si } \theta_0 < x_{(1)} \\ 1 & \text{si } \theta_0 \geq x_{(1)}. \end{cases} \end{aligned}$$

Para  $0 < c < 1$ , la región de rechazo dada en (7.3) es:

$$\begin{aligned} \{\mathbf{x} : \lambda(\mathbf{x}) \leq c\} &= \{\mathbf{x} : e^{n(\theta_0 - x_{(1)})} \leq c \text{ si } \theta_0 \leq x_{(1)}\} \\ &= \{\mathbf{x} : n\theta_0 - nx_{(1)} \leq \ln(c) \text{ si } \theta_0 \leq x_{(1)}\} \\ &= \{\mathbf{x} : n\theta_0 - \ln(c) \leq nx_{(1)} \text{ si } \theta_0 \leq x_{(1)}\} \\ &= \left\{ \mathbf{x} : x_{(1)} \geq \theta_0 - \frac{\ln(c)}{n} \text{ si } \theta_0 \leq x_{(1)} \right\}, \quad (7.6) \end{aligned}$$

y corresponde a un test **LRT** para probar la hipótesis  $H_0 : \theta \leq \theta_0$ .  $\square$

**DEFINICIÓN 7.12.** Según [20, p. 386], un **LRT** de tamaño  $\alpha$  es construido escogiendo un  $c$  tal que:

$$\sup_{\theta \in \Theta_0} P_{\theta_0}(\lambda(X) \leq c) = \alpha.$$

**Ejemplo 7.5.** Considerando la información del ejemplo 7.3 y usando un **LRT** que rechaza  $H_0 : \theta = \theta_0$ , determine  $c$  [20, p. 386].

*Solución.* La región de rechazo se encontró en (7.4), luego,

$$\alpha = \sup_{\theta \in \Theta_0} P_{\theta} \left\{ |\bar{X} - \theta_0| \geq \sqrt{\frac{-2 \ln(c)}{n}} \right\} = P_{\theta_0} \left\{ |\bar{X} - \theta_0| \geq \sqrt{\frac{-2 \ln(c)}{n}} \right\}.$$

Esta expresión corresponde a la función de potencia evaluada en  $\theta_0$ , luego de la expresión (7.5) se tiene que:

$$\begin{aligned} \beta(\theta) &= 2 - \Phi\left(\sqrt{-2 \ln(c)} - \frac{\theta_0 - \theta}{1/\sqrt{n}}\right) - \Phi\left(\sqrt{-2 \ln(c)} + \frac{\theta_0 - \theta}{1/\sqrt{n}}\right) \\ \beta(\theta_0) &= 2 - \Phi\left(\sqrt{-2 \ln(c)} - \frac{\theta_0 - \theta_0}{X/\sqrt{n}}\right) - \Phi\left(\sqrt{-2 \ln(c)} + \frac{\theta_0 - \theta_0}{X/\sqrt{n}}\right) \\ \beta(\theta_0) &= 2 \left[ 1 - \Phi\left(\sqrt{-2 \ln(c)}\right) \right] = \alpha. \end{aligned}$$

Despejando, se obtiene el cuantil

$$\begin{aligned}\sqrt{-2 \ln(c)} &= \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \quad \Rightarrow \quad -2 \ln(c) = \left( z_{1-\frac{\alpha}{2}} \right)^2 \quad (7.7) \\ \Rightarrow \quad \ln(c) &= \frac{-1}{2} z_{1-\frac{\alpha}{2}}^2 \quad \Rightarrow \quad c = \exp \left\{ -\frac{1}{2} z_{1-\frac{\alpha}{2}}^2 \right\},\end{aligned}$$

reemplazando (7.7) en (7.4) se tiene que:

$$\begin{aligned}\{x : \lambda(x) \leq c\} &= \left\{ x : |\bar{x} - \theta_0| \geq \sqrt{\frac{-2 \ln(c)}{n}} \right\} \\ &= \left\{ x : |\bar{x} - \theta_0| \geq \sqrt{\frac{z_{1-\frac{\alpha}{2}}^2}{n}} \right\} = \left\{ x : |\bar{x} - \theta_0| \geq \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right\} \\ &= \left\{ x : \bar{x} - \theta_0 \geq \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \quad \text{o} \quad \bar{x} - \theta_0 \leq -\frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right\} \\ &= \left\{ x : \bar{x} \geq \theta_0 + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \quad \text{o} \quad \bar{x} \leq \theta_0 - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right\}. \quad \square\end{aligned}$$

**Ejemplo 7.6.** Considerando la información del ejemplo 7.4, se encontró que la región de rechazo está dada por (7.6), la cual corresponde a un test *LRT* para probar la hipótesis  $H_0 : \theta \leq \theta_0$ . Determinar  $c$ .

*Solución.* Luego,

$$\sup_{\theta \in \Theta_0} P_\theta \left( X_{(1)} \geq \theta_0 - \frac{\ln(c)}{n} \right) = \sup_{\theta \leq \theta_0} P_\theta \left( X_{(1)} \geq \theta_0 - \frac{\ln(c)}{n} \right) = \alpha.$$

Por la expresión (6.9) se tiene que:

$$F_{X_{(1)}}(x; \theta) = P_\theta (X_{(1)} < x) = 1 - e^{-n(x-\theta)}.$$

Entonces,

$$\begin{aligned}\alpha &= \sup_{\theta \leq \theta_0} P_\theta \left( X_{(1)} \geq \theta_0 - \frac{\ln(c)}{n} \right) = \sup_{\theta \leq \theta_0} \left[ 1 - P_\theta \left( X_{(1)} < \theta_0 - \frac{\ln(c)}{n} \right) \right] \\ &= \sup_{\theta \leq \theta_0} \left[ 1 - \left\{ 1 - \exp \left( -n \left[ \theta_0 - \frac{\ln(c)}{n} - \theta \right] \right) \right\} \right] \\ &= \sup_{\theta \leq \theta_0} [\exp\{\ln(c) - n(\theta_0 - \theta)\}] = \exp\{\ln(c)\} \sup_{\theta \leq \theta_0} [\exp\{-n(\theta_0 - \theta)\}] \\ &= c \sup_{\theta \leq \theta_0} [\exp\{-n(\theta_0 - \theta)\}],\end{aligned}$$

nótese que la función  $\exp\{-n(\theta_0 - \theta)\}$  toma su máximo cuando  $\theta = \theta_0$ , luego se tiene que  $c = \alpha$ , y así:

$$\begin{aligned} \{x : \lambda(x) \leq c\} &= \left\{x : x_{(1)} \geq \theta_0 - \frac{\ln(c)}{n} \text{ si } \theta_0 \leq x_{(1)}\right\} \\ &= \left\{x : x_{(1)} \geq \theta_0 - \frac{\ln(\alpha)}{n} \text{ si } \theta_0 \leq x_{(1)}\right\}. \quad \checkmark \end{aligned}$$

**Teorema 7.1.** Según [20, p. 377], si  $T(\mathbf{X})$  es una estadística suficiente para  $\theta$  y  $\lambda^*(t)$  y  $\lambda(x)$  son los LRT basados en  $T$  y  $\mathbf{X}$ , respectivamente, entonces

$$\lambda^*(T(x)) = \lambda(x)$$

para cada  $x$  en el espacio muestral.

*Demostración.* El LRT está dado por (7.2) y por el teorema de factorización [20, p. 377]:

$$\lambda(x) = \frac{\sup_{\Theta_0} L(\theta|x)}{\sup_{\Theta} L(\theta|x)} = \frac{\sup_{\Theta_0} f(x; \theta)}{\sup_{\Theta} f(x; \theta)} \stackrel{43}{=} \frac{\sup_{\Theta_0} g(T(x); \theta)h(x)}{\sup_{\Theta} g(T(x); \theta)h(x)},$$

como la función  $h(\cdot)$  no depende de  $\theta$  se tiene que:

$$\begin{aligned} \lambda(x) &= \frac{\cancel{h(x)} \sup_{\Theta_0} g(T(x); \theta)}{\cancel{h(x)} \sup_{\Theta} g(T(x); \theta)} = \frac{\sup_{\Theta_0} g(T(x); \theta)}{\sup_{\Theta} g(T(x); \theta)} \\ &= \frac{\sup_{\Theta_0} P_{\theta}(T(\mathbf{X}) = t_i)}{\sup_{\Theta} P_{\theta}(T(\mathbf{X}) = t_i)} = \frac{\sup_{\Theta_0} L^*(\theta|T(\mathbf{X}))}{\sup_{\Theta} L^*(\theta|T(\mathbf{X}))} \\ &= \lambda^*(T(x)). \quad \checkmark \end{aligned}$$

### 7.3.2. Lema de Neyman-Pearson

**Teorema 7.2.** (Lema de Neyman-Pearson). De acuerdo con [20, p. 388], dada  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una población con fdp  $p_X(x; \theta)$ . Considere las hipótesis nula y alterna simples

<sup>43</sup>Teorema de factorización,  $T$  es suficiente.

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1, \quad (7.8)$$

en el nivel  $\alpha$  ( $0 < \alpha < 1$ ) y sea  $\delta_k(x)$  la **función de test** definida como sigue:

$$\begin{aligned} \delta_k(x) &\stackrel{44}{=} \begin{cases} 1 & \text{si } p(x; \theta_1) > kp(x; \theta_0) \\ 0 & \text{si } p(x; \theta_1) < kp(x; \theta_0) \end{cases} \\ &= \begin{cases} 1 & \text{si } p(x; \theta_1) - kp(x; \theta_0) > 0 \\ 0 & \text{si } p(x; \theta_1) - kp(x; \theta_0) < 0, \end{cases} \end{aligned}$$

con  $k > 0$ . Aquí  $\delta_k(x)$  es una función indicadora de la región de rechazo del test. Entonces:

1. (Condición suficiente para ser un test más potente). Si  $\delta_k(x)$  es un test (una función de test) de tamaño  $\alpha$ , entonces  $\delta_k(x)$  es **UMP** en la clase  $\mathcal{C}$ .
2. (Condición necesaria para ser un test más potente). Si existe  $\delta_k(x)$ , un test (una función de test) de tamaño  $\alpha$ , entonces cada test **UMP** de nivel  $\alpha$  es un test de tamaño  $\alpha$  excepto en un conjunto  $A$  tal que:

$$P_{\theta_0}(X \in A) = P_{\theta_1}(X \in A) = 0.$$

*Demostración.* Se prueba el teorema para el caso en que  $p(x; \theta_1)$  y  $p(x; \theta_0)$  son **fdp** de variables aleatorias continuas.

1. Como  $\delta_k(x)$  es un test de tamaño  $\alpha$  se tiene que:

$$\alpha = \sup_{\theta \in \Theta_0} \beta(\theta) = \sup_{\theta \in \Theta_0} P_{\theta}(X \in \mathbb{R}) \stackrel{\Theta_0 = \{\theta_0\}}{=} P_{\theta_0}(X \in \mathbb{R}),$$

por la definición de función de potencia (definición 7.4):

$$\alpha = P_{\theta_0}(\delta_k(X) = 1) = \beta(\theta_0) = \mathbb{E}_{\theta_0}[\delta_k(X)].$$

Sea  $\delta'_k(x)$  una función de test para cualquier otro test de nivel  $\alpha$  y sean  $\beta(\theta)$  y  $\beta'(\theta)$ <sup>45</sup> las funciones de potencia asociadas a  $\delta_k(x)$  y  $\delta'_k(x)$ , respectivamente. Como también,  $\delta'_k(x) = 0$  ó  $\delta'_k(x) = 1$  se tiene que:

<sup>44</sup> Hay autores como [76, p. 330] que definen la función de test  $\delta_k(x) = \gamma$ ,  $0 < \gamma < 1$ , para el caso  $p(x; \theta_1) = kp(x; \theta_0)$ .

<sup>45</sup> Como  $\beta'(\theta)$  es de nivel  $\alpha$ , se tiene que  $\sup_{\theta \in \Theta_0} \beta'(\theta) \leq \alpha$ , es decir,  $\beta'(\theta_0) \leq \alpha$ .

$$\delta_k(x) - \delta'_k(x) = \begin{cases} -1 & \text{si } \delta_k(x) = 0 \text{ y } \delta'_k(x) = 1 \\ 0 & \text{si } \delta_k(x) = 1 \text{ y } \delta'_k(x) = 1 \\ 1 & \text{si } \delta_k(x) = 1 \text{ y } \delta'_k(x) = 0 \\ 0 & \text{si } \delta_k(x) = 0 \text{ y } \delta'_k(x) = 0, \end{cases}$$

luego,

$$\begin{aligned} 0 &\leq [\delta_k(x) - \delta'_k(x)] [p(x; \theta_1) - kp(x; \theta_0)] \\ &= \begin{cases} -1 \times \underbrace{[p(x; \theta_1) - kp(x; \theta_0)]}_{<0} & \text{si } \delta_k(x) = 0 \text{ y } \delta'_k(x) = 1 \\ 0 \times [p(x; \theta_1) - kp(x; \theta_0)] & \text{si } \delta_k(x) = 1 \text{ y } \delta'_k(x) = 1 \\ 1 \times \underbrace{[p(x; \theta_1) - kp(x; \theta_0)]}_{>0} & \text{si } \delta_k(x) = 1 \text{ y } \delta'_k(x) = 0 \\ 0 \times [p(x; \theta_1) - kp(x; \theta_0)] & \text{si } \delta_k(x) = 0 \text{ y } \delta'_k(x) = 0 \end{cases} \end{aligned}$$

para cada  $x$ . Integrando se tiene que:

$$\int \cdots \int [\delta_k(x) - \delta'_k(x)] [p(x; \theta_1) - kp(x; \theta_0)] dx \geq 0. \quad (7.9)$$

Desarrollando el producto término a término,

$$\begin{aligned} &\underbrace{\int \cdots \int \delta_k(x)p(x; \theta_1)dx}_{\beta(\theta_1)} - \underbrace{\int \cdots \int k\delta_k(x)p(x; \theta_0)dx}_{k\beta(\theta_0)} \\ &- \underbrace{\int \cdots \int \delta'_k(x)p(x; \theta_1)dx}_{\beta'(\theta_1)} + \underbrace{\int \cdots \int k\delta'_k(x)p(x; \theta_0)dx}_{k\beta'(\theta_0)} \geq 0, \end{aligned}$$

reagrupando términos se obtiene:

$$\beta(\theta_1) - \beta'(\theta_1) - k \underbrace{\left[ \underbrace{\beta(\theta_0)}_{\alpha} - \underbrace{\beta'(\theta_0)}_{\leq \alpha} \right]}_{\geq 0} \geq 0,$$

entonces,

$$0 \leq \beta(\theta_1) - \beta'(\theta_1) - k[\beta(\theta_0) - \beta'(\theta_0)] \leq \beta(\theta_1) - \beta'(\theta_1) \quad (7.10)$$

$$\beta'(\theta_1) \leq \beta(\theta_1),$$

por tanto,  $\delta_k(x)$  es más potente que  $\delta'_k(x)$ . Como  $\alpha$  era arbitrario,  $\delta_k(x)$  es un test **UMP**.

2. Sea  $\delta'_k(x)$  una función de test para cualquier test de nivel  $\alpha$  **UMP**, entonces

$$\beta'(\theta_1) = \beta(\theta_1),$$

y de (7.10) se tiene que:

$$\begin{aligned} 0 &\leq \underbrace{\beta(\theta_1) - \beta'(\theta_1)}_0 - k \left[ \underbrace{\beta(\theta_0) - \beta'(\theta_0)}_\alpha \right] \\ 0 &\leq -k \left[ \alpha - \beta'(\theta_0) \right] \quad \text{introduciendo el signo dentro del paréntesis} \\ 0 &\leq \underbrace{k}_{>0} \left[ \beta'(\theta_0) - \alpha \right] \implies 0 \leq \underbrace{\beta'(\theta_0) - \alpha}_{\leq \alpha} \\ 0 &\leq \beta'(\theta_0) - \alpha \leq 0 \end{aligned}$$

Por el teorema del emparejado  $\beta'(\theta_0) = \alpha$ . Por tanto,  $\delta'_k(x)$  es un test de tamaño  $\alpha$ . Retomando (7.9):

$$\begin{aligned} &\int \cdots \int [\varphi(x) - \varphi'(x)] [p(x; \theta_1) - kp(x; \theta_0)] dx \\ &= \underbrace{\beta(\theta_1) - \beta'(\theta_1)}_0 - k \underbrace{[\beta(\theta_0) - \beta'(\theta_0)]}_0 = 0, \end{aligned}$$

excepto si al caso en conjunto  $A$  con medida cero, es decir,

$$P_{\theta_0}(\mathbf{X} \in A) = P_{\theta_1}(\mathbf{X} \in A) = 0. \quad \square$$

**Nota 7.5.** *Considérese, como en [20, p. 389], las hipótesis nula y alterna simples dadas en (7.8), si  $T(\mathbf{X})$  es una estadística suficiente para  $\theta$  y  $g_T(t; \theta)$  es la fdp o fmp de  $T$  correspondiente a  $\theta$ , del teorema 7.1 se tiene que los **LRT** basados en  $T$  y  $x$  son iguales, por tanto cualquier test basado en  $T$  con región de rechazo  $S$  (un subconjunto del espacio muestral de  $T$ ) es un test **UMP** de nivel  $\alpha$  si cumple con que:*

$$t \in \begin{cases} S & \text{si } g_T(t; \theta_1) > k g_T(t; \theta_0) \\ S^c & \text{si } g_T(t; \theta_1) < k g_T(t; \theta_0), \end{cases}$$

para algún  $k > 0$  donde  $\alpha = P_{\theta_0}(T \in S)$ .

**Ejemplo 7.7.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid normales con media desconocida  $\theta$  y varianza conocida  $\sigma^2$ . Considere las hipótesis  $H_0 : \theta = \theta_0$  vs.  $H_1 : \theta = \theta_1$  con  $\theta_0 > \theta_1$ . Determine:

1. El test UMP para rechazar  $H_0$ .
- 2.Cuál es el tamaño de muestra para rechazar  $H_0$ , si se desea tener una probabilidad máxima de error tipo I de 10% y una probabilidad máxima de error tipo II de 20%, suponga que  $\theta_0 - \theta_1 = \sigma$ .

*Solución.*

1. Del ejemplo 5.3 se tiene que  $T(\mathbf{X}) = \bar{X}$ , la media muestral, es una estadística suficiente para  $\theta$ . Usando la nota 7.5,  $\bar{x} = t \in S$  si

$$g_T(\bar{x}; \theta_1) > k g_T(\bar{x}; \theta_0)$$

$$\exp \left\{ \frac{-n}{2\sigma^2} (\bar{x} - \theta_1)^2 \right\} > k \exp \left\{ \frac{-n}{2\sigma^2} (\bar{x} - \theta_0)^2 \right\},$$

tomando logaritmo natural a ambos lados se tiene que:

$$\begin{aligned} \frac{-n}{2\sigma^2} (\bar{x} - \theta_1)^2 &> \ln \left( k \exp \left\{ \frac{-n}{2\sigma^2} (\bar{x} - \theta_0)^2 \right\} \right) \\ \frac{-n}{2\sigma^2} (\bar{x} - \theta_1)^2 &> \ln k - \frac{n}{2\sigma^2} (\bar{x} - \theta_0)^2 \\ -(\bar{x} - \theta_1)^2 &> \frac{2\sigma^2}{n} \ln k - (\bar{x} - \theta_0)^2 \\ (\bar{x} - \theta_0)^2 - (\bar{x} - \theta_1)^2 &> \frac{2\sigma^2}{n} \ln k \\ \bar{x}^2 - 2\bar{x}\theta_0 + \theta_0^2 - \bar{x}^2 + 2\bar{x}\theta_1 - \theta_1^2 &> \frac{2\sigma^2}{n} \ln k \\ 2\bar{x}(\theta_1 - \theta_0) &> \frac{2\sigma^2}{n} \ln k + (\theta_1^2 - \theta_0^2) \\ &< 0 \\ \bar{x} &< \frac{\frac{2\sigma^2}{n} \ln k + (\theta_1 - \theta_0)(\theta_1 + \theta_0)}{2(\theta_1 - \theta_0)} \\ \bar{x} &< \frac{\theta_0 + \theta_1}{2} - \frac{\ln k}{\theta_0 - \theta_1} \frac{\sigma^2}{n}, \end{aligned}$$

esto corresponde a la región crítica óptima, por tanto,

$$\bar{x} \in S \quad \text{si} \quad \bar{x} < \frac{\theta_0 + \theta_1}{2} - \frac{\ln k}{\theta_0 - \theta_1} \frac{\sigma^2}{n} \quad \text{y}$$

$$\bar{x} \in S^c \quad \text{si} \quad \bar{x} > \frac{\theta_0 + \theta_1}{2} - \frac{\ln k}{\theta_0 - \theta_1} \frac{\sigma^2}{n},$$

su respectiva función de potencia es dada por (7.1):

$$\begin{aligned} \beta(\theta) &= P_\theta(\bar{X} \in S) = P_\theta\left(\bar{X} < \underbrace{\frac{\theta_0 + \theta_1}{2} - \frac{\ln k}{\theta_0 - \theta_1} \frac{\sigma^2}{n}}_d\right) = P_\theta(\bar{X} < d) \\ &= P_\theta\left(\underbrace{\frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}}}_Z < \frac{d - \theta}{\frac{\sigma}{\sqrt{n}}}\right) = \Phi\left(\frac{d - \theta}{\frac{\sigma}{\sqrt{n}}}\right), \end{aligned} \quad (7.11)$$

donde  $\Phi(\cdot)$  es la fda de la variable aleatoria  $Z \sim N(0, 1)$ . Luego,

$$\begin{aligned} \alpha &= \sup_{\theta \in \Theta_0} \beta(\theta) = \sup_{\theta \in \Theta_0} P_\theta(\bar{X} \in S) \underbrace{=}_{\Theta_0 = \{\theta_0\}} P_{\theta_0}(\bar{X} \in S) = \beta(\theta_0) \\ &= \Phi\left(\frac{d - \theta_0}{\frac{\sigma}{\sqrt{n}}}\right) = \Phi\left[-\left(\frac{\theta_0 - \theta_1}{2 \frac{\sigma}{\sqrt{n}}} + \frac{\ln k}{\theta_0 - \theta_1} \frac{\sigma}{\sqrt{n}}\right)\right], \end{aligned}$$

despejando, se obtiene el cuantil:

$$\Phi^{-1}(\alpha) = \frac{d - \theta_0}{\frac{\sigma}{\sqrt{n}}} < 0 \quad \implies \quad d = \frac{\sigma z_\alpha}{\sqrt{n}} + \theta_0.$$

Por tanto, el test **UMP** rechaza  $H_0$  si  $\bar{X} < \frac{\sigma z_\alpha}{\sqrt{n}} + \theta_0$ , es decir, el test **UMP** rechaza  $H_0$  si

$$\frac{\bar{X} - \theta_0}{\frac{\sigma}{\sqrt{n}}} < -z_{1-\alpha}.$$

Si las hipótesis consideradas fueran  $H_0 : \theta = \theta_0$  vs.  $H_1 : \theta = \theta_1$ , donde  $\theta_0 < \theta_1$ , el test **UMP** rechaza  $H_0$  si  $\bar{X} > \frac{\sigma z_{1-\alpha}}{\sqrt{n}} + \theta_0$ , es decir, el test **UMP** rechaza  $H_0$  si

$$\frac{\bar{X} - \theta_0}{\frac{\sigma}{\sqrt{n}}} > z_{1-\alpha}.$$

Si las hipótesis consideradas fueran  $H_0 : \theta = \theta_0$  vs.  $H_1 : \theta \neq \theta_0$ , el test insesgado **UMP** rechaza  $H_0$  si

$$\left| \frac{\bar{X} - \theta_0}{\frac{\sigma}{\sqrt{n}}} \right| > z_{1-\frac{\alpha}{2}}.$$

2. De la información dada se tiene que:

$$\beta(\theta_0) = 0,1 \quad \text{y} \quad \beta(\theta_1) = 0,8.$$

Por la expresión (7.11) se obtiene que:

$$\frac{d - \theta_0}{\frac{\sigma}{\sqrt{n}}} = \Phi^{-1}(0,1) \quad \text{y} \quad \frac{d - \theta_1}{\frac{\sigma}{\sqrt{n}}} = \Phi^{-1}(0,8),$$

al restar estas expresiones se llega a:

$$\frac{\theta_1 - \theta_0}{\sigma} \sqrt{n} = \Phi^{-1}(0,1) - \Phi^{-1}(0,8),$$

despejando  $n$  y dado que  $\theta_0 - \theta_1 = \sigma$

$$n = \left[ \frac{\Phi^{-1}(0,8) - \Phi^{-1}(0,1)}{\theta_0 - \theta_1} \right]^2 \sigma^2 = \left[ \frac{2,1232}{\sigma} \right]^2 \sigma^2 = 4,5080,$$

puesto que  $n$  debe ser entero, entonces  $n = 5$ . ☑

**Ejemplo 7.8.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid Bernoulli de parámetro  $\theta$ . Considere las hipótesis:

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1 \quad \text{con} \quad 0 < \theta_1 < \theta_0 < 1.$$

Determine el test UMP para rechazar  $H_0$ .

*Solución.* Del ejemplo 5.1 se tiene que  $T(X) = \sum_{i=1}^n X_i$  es una estadística suficiente para  $\theta$ . Usando la nota 7.5,  $\sum_{i=1}^n x_i = t \in S$  si

$$\begin{aligned} & g_T \left( \sum_{i=1}^n x_i; \theta_1 \right) > k g_T \left( \sum_{i=1}^n x_i; \theta_0 \right) \\ \exp \left\{ \ln \left( \frac{\theta_1}{1 - \theta_1} \right) \sum_{i=1}^n x_i + n \ln(1 - \theta_1) \right\} & > k \exp \left\{ \ln \left( \frac{\theta_0}{1 - \theta_0} \right) \sum_{i=1}^n x_i + n \ln(1 - \theta_0) \right\} \\ \ln \left( \frac{\theta_1}{1 - \theta_1} \right) \sum_{i=1}^n x_i + n \ln(1 - \theta_1) & > \ln(k) + \ln \left( \frac{\theta_0}{1 - \theta_0} \right) \sum_{i=1}^n x_i + n \ln(1 - \theta_0). \end{aligned}$$

Dividiendo a ambos lados de la desigualdad por  $n$  y reagrupando

$$\bar{x} \left[ \ln \left( \frac{\theta_1}{1 - \theta_1} \right) - \ln \left( \frac{\theta_0}{1 - \theta_0} \right) \right] > \frac{\ln(k)}{n} + [\ln(1 - \theta_0) - \ln(1 - \theta_1)]$$

$$\bar{x} \left[ \underbrace{\ln \left( \frac{\theta_1}{\theta_0} \right)}_{<1} \times \underbrace{\frac{1-\theta_0}{1-\theta_1}}_{<1} \right] > \frac{\ln(k)}{n} + \ln \left( \frac{1-\theta_0}{1-\theta_1} \right),$$

como el logaritmo de un número menor a uno es negativo, se tiene que:

$$\bar{x} < \frac{\ln(k) + n \ln \left( \frac{1-\theta_0}{1-\theta_1} \right)}{n \ln \left( \frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)} \right)},$$

esto corresponde a la región crítica óptima, por tanto,

$$\begin{aligned} \bar{x} \in S & \quad \text{si} \quad \bar{x} < \frac{\ln(k) + n \ln \left( \frac{1-\theta_0}{1-\theta_1} \right)}{n \ln \left( \frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)} \right)} \quad \text{y} \\ \bar{x} \in S^c & \quad \text{si} \quad \bar{x} > \frac{\ln(k) + n \ln \left( \frac{1-\theta_0}{1-\theta_1} \right)}{n \ln \left( \frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)} \right)}, \end{aligned}$$

con

$$\alpha = P_{\theta_0}(\bar{X} \in S) = P_{\theta_0} \left( \bar{X} < \underbrace{\frac{\ln(k) + n \ln \left( \frac{1-\theta_0}{1-\theta_1} \right)}{n \ln \left( \frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)} \right)}}_d \right) = P_{\theta_0}(\bar{X} < d).$$

Si  $n$  es suficientemente grande, utilizando el ejemplo 3.33:

$$\alpha \approx P_{\theta_0} \left( \underbrace{\frac{\bar{X} - \theta_0}{\frac{\sqrt{\theta_0(1-\theta_0)}}{\sqrt{n}}}}_Z < \frac{d - \theta_0}{\frac{\sqrt{\theta_0(1-\theta_0)}}{\sqrt{n}}} \right) = \Phi \left( \frac{d - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \right),$$

despejando, se obtiene el cuantil:

$$\Phi^{-1}(\alpha) = \frac{d - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \quad \Rightarrow \quad d = z_\alpha \sqrt{\frac{\theta_0(1-\theta_0)}{n}} + \theta_0.$$

Por tanto, para un tamaño de muestra ( $n$ ) suficientemente grande, el test **UMP** rechaza  $H_0$  si  $\bar{X} < z_\alpha \sqrt{\frac{\theta_0(1-\theta_0)}{n}} + \theta_0$ , es decir, el test **UMP** rechaza  $H_0$

si

$$\frac{\bar{X} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} < -z_{1-\alpha}.$$

Si las hipótesis consideradas fueran  $H_0 : \theta = \theta_0$  vs.  $H_1 : \theta = \theta_1$ , donde  $0 < \theta_0 < \theta_1 < 1$ , para  $n$  suficientemente grande, el test **UMP** rechaza  $H_0$  si  $\bar{X} > z_\alpha \sqrt{\frac{\theta_0(1-\theta_0)}{n}} + \theta_0$ , es decir, el test **UMP** rechaza  $H_0$  si

$$\frac{\bar{X} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} > z_{1-\alpha}.$$

Si las hipótesis consideradas fueran  $H_0 : \theta = \theta_0$  vs.  $H_1 : \theta \neq \theta_0$ , para  $n$  suficientemente grande, el test insesgado **UMP** rechaza  $H_0$  si

$$\left| \frac{\bar{X} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \right| > z_{1-\frac{\alpha}{2}}. \quad \square$$

Para ilustrar el caso donde no existe un test **UMP** si no se restringe a los test insesgados se presenta el siguiente ejemplo.

**Ejemplo 7.9.** Sea  $X$  una variable aleatoria con distribución binomial de parámetros  $n$  y  $\theta$  con  $0 < \theta < 1$ . Se quiere probar las hipótesis

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_a : \theta \neq \theta_0$$

a un nivel  $\alpha$ . Encuentre el test **UMP** para rechazar  $H_0$ .

*Solución.* Sean las hipótesis

$$H'_0 : \theta = \theta_0 \quad \text{vs.} \quad H'_a : \theta = \theta_1,$$

luego,  $\Theta_0 = \{\theta_0\}$  y  $\Theta = \{\theta_0, \theta_1\}$ . El **LRT** para probar  $H'_0$  vs.  $H'_a$  es:

$$\lambda(x) = \frac{\sup_{\Theta_0} L(\theta; x)}{\sup_{\Theta} L(\theta; x)} = \frac{\sup_{\Theta_0} \binom{n}{x} \theta^x (1-\theta)^{n-x}}{\sup_{\Theta} \binom{n}{x} \theta^x (1-\theta)^{n-x}} = \frac{\theta_0^x (1-\theta_0)^{n-x}}{\sup_{\Theta} \theta^x (1-\theta)^{n-x}}.$$

Si  $\sup_{\Theta} \theta^x (1 - \theta)^{n-x}$  se tiene cuando  $\theta = \theta_0$ , entonces  $\lambda(x) = 1$ , por otro lado, si  $\sup_{\Theta} \theta^x (1 - \theta)^{n-x}$  se tiene cuando  $\theta = \theta_1$ , entonces

$$\lambda(x) = \frac{\theta_0^x (1 - \theta_0)^{n-x}}{\theta_1^x (1 - \theta_1)^{n-x}} = \left(\frac{\theta_0}{\theta_1}\right)^x \left(\frac{1 - \theta_0}{1 - \theta_1}\right)^{n-x} = \left(\frac{\theta_0(1 - \theta_1)}{\theta_1(1 - \theta_0)}\right)^x \left(\frac{1 - \theta_0}{1 - \theta_1}\right)^n.$$

Luego, para  $0 \leq c \leq 1$ , la región de rechazo es de la forma

$$\begin{aligned} \{x : \lambda(x) \leq c\} &= \left\{x : \left(\frac{\theta_0(1 - \theta_1)}{\theta_1(1 - \theta_0)}\right)^x \left(\frac{1 - \theta_0}{1 - \theta_1}\right)^n \leq c\right\} \\ &= \left\{x : x \ln \left[\frac{\theta_0(1 - \theta_1)}{\theta_1(1 - \theta_0)}\right] + n \ln \left[\frac{1 - \theta_0}{1 - \theta_1}\right] \leq \ln(c)\right\} \\ &= \left\{x : x \ln \left[\frac{\theta_0(1 - \theta_1)}{\theta_1(1 - \theta_0)}\right] \leq \ln(c) - n \ln \left[\frac{1 - \theta_0}{1 - \theta_1}\right]\right\} \\ &= \begin{cases} \left\{x : x \leq \frac{\ln(c) - n \ln \left[\frac{1 - \theta_0}{1 - \theta_1}\right]}{\ln \left[\frac{\theta_0(1 - \theta_1)}{\theta_1(1 - \theta_0)}\right]}\right\} & \text{si } \theta_0 > \theta_1 \\ \left\{x : x \geq \frac{\ln(c) - n \ln \left[\frac{1 - \theta_0}{1 - \theta_1}\right]}{\ln \left[\frac{\theta_0(1 - \theta_1)}{\theta_1(1 - \theta_0)}\right]}\right\} & \text{si } \theta_0 < \theta_1, \end{cases} \end{aligned}$$

por tanto, si  $\theta_0 > \theta_1$ , el test más potente rechaza para valores pequeños de  $x$ ; pero si  $\theta_0 < \theta_1$ , el test más potente rechaza para valores grandes de  $x$ , luego el test depende de  $\theta_1$  y no existe un test **UMP**.  $\square$

**Nota 7.6.** En general, no es posible encontrar un test **UMP** de tamaño  $\alpha$ , por ejemplo, para el problema  $H_0 : \theta = \theta_0$  vs.  $H_1 : \theta \neq \theta_0$ . Para encontrar un test con óptimas propiedades se restringe a una clase más pequeña de test que son los test insesgados, ver definición 7.8 [76, p. 349]. Una clase de distribuciones en las cuales si existe un test **UMP** de tamaño  $\alpha$  es la familia de distribuciones que tiene la propiedad de la razón de verosimilitud monótona (**RVM**).

**DEFINICIÓN 7.13.** Según [20, p. 391], la familia uniparamétrica de *fdp* o de *fmp*  $\{p(x; \theta) : \theta \in \Theta\}$  para una variable aleatoria univariada  $X$  con parámetro de valor real  $\theta$  se dice que tiene una **RVM** si para cada  $\theta_2 > \theta_1$ , la razón  $\frac{p(x; \theta_2)}{p(x; \theta_1)}$  es función monótona (no creciente o no decreciente) de  $x$  sobre  $\{x : p(x; \theta_1) > 0 \text{ ó } p(x; \theta_2) > 0\}$ . Note que  $\frac{c}{0}$  se define como  $\infty$  si  $c > 0$ .

**Ejemplo 7.10.** Suponga que  $X_1, X_2, \dots, X_n$  son variables aleatorias *iid* tal que  $X_1 \sim N(\mu, 1)$  y sean  $\mu_1$  y  $\mu_2$  tal que  $0 < \mu_1 < \mu_2 < \infty$ . Determine la **RVM**.

*Solución.* Se tiene que:

$$p(\mathbf{x}; \mu_1) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \mu_1)^2\right\} \quad \text{y}$$

$$p(\mathbf{x}; \mu_2) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i - \mu_2)^2\right\},$$

luego,

$$\begin{aligned} \frac{p(\mathbf{x}; \mu_2)}{p(\mathbf{x}; \mu_1)} &= \exp\left\{-\frac{1}{2} \sum_{i=1}^n [(x_i - \mu_2)^2 - (x_i - \mu_1)^2]\right\} \\ &= \exp\left\{-\frac{1}{2} \sum_{i=1}^n (x_i^2 - 2\mu_2 x_i + \mu_2^2 - x_i^2 + 2\mu_1 x_i - \mu_1^2)\right\} \\ &= \exp\left\{-\frac{1}{2} \left(-2\mu_2 \sum_{i=1}^n x_i + n\mu_2^2 + 2\mu_1 \sum_{i=1}^n x_i - n\mu_1^2\right)\right\} \\ &= \underbrace{\exp\left\{-\frac{1}{2} n (\mu_2^2 - \mu_1^2)\right\}}_{K_0 > 0} \exp\left\{\underbrace{(\mu_2 - \mu_1) \sum_{i=1}^n x_i}_{K_1 > 0}\right\} \\ &= K_0 \exp\left\{K_1 \sum_{i=1}^n x_i\right\} = K_0 \left(e^{K_1}\right)^{\sum_{i=1}^n x_i}. \end{aligned}$$

La función  $(e^{K_1})^{\sum_{i=1}^n x_i}$  es creciente (estrictamente) en  $\sum_{i=1}^n x_i$ . Así,  $N(\mu, 1)$  es

**RVM** para  $T(\mathbf{x}) = \sum_{i=1}^n x_i$ . ☑

**Ejemplo 7.11.** Sean  $X_1, X_2, \dots, X_n$  variables aleatorias *iid* tal que

$$p(x_1; \theta) = \theta e^{-\theta x_1} I_{(0, +\infty)}(x_1), \quad \theta > 0.$$

Sean  $0 < \theta_1 < \theta_2 < \infty$ , determine la **RVM**.

*Solución.* Se tiene que:

$$p(\mathbf{x}; \theta_1) = \theta_1^n \exp\left\{-\theta_1 \sum_{i=1}^n x_i\right\} \prod_{i=1}^n I_{(0, +\infty)}(x_i)$$

$$p(\mathbf{x}; \theta_2) = \theta_2^n \exp\left\{-\theta_2 \sum_{i=1}^n x_i\right\} \prod_{i=1}^n I_{(0, +\infty)}(x_i),$$

luego,

$$\begin{aligned} \frac{p(\mathbf{x}; \theta_2)}{p(\mathbf{x}; \theta_1)} &= \underbrace{\left(\frac{\theta_2}{\theta_1}\right)^n}_{K_0 > 0} \exp \left\{ - \underbrace{(\theta_2 - \theta_1)}_{K_1 > 0} \sum_{i=1}^n x_i \right\} \prod_{i=1}^n I_{(0, +\infty)}(x_i) \\ &= K_0 \left( e^{-K_1} \right)^{\sum_{i=1}^n x_i} \prod_{i=1}^n I_{(0, +\infty)}(x_i). \end{aligned}$$

La función  $(e^{-K_1})^{\sum_{i=1}^n x_i}$  es decreciente (estrictamente) en  $\sum_{i=1}^n x_i$ , por tanto,

$p(\mathbf{x}; \theta)$  es **RVM** en  $\sum_{i=1}^n x_i$ . ☑

**Teorema 7.3.** (Karlin-Rubin). *Considérese, como en [20, p. 391], la prueba  $H_0 : \theta \leq \theta_0$  vs.  $H_1 : \theta > \theta_0$ . Suponga que  $T$  es una estadística suficiente para  $\theta$  y que la familia de fdp o de fmp  $\{g(t; \theta) : \theta \in \Theta\}$  de  $T$  tiene una **RVM**. Entonces, para cualquier  $t_0$  el test rechaza  $H_0$  si y solo si  $T > t_0$  es un test **UMP** de nivel  $\alpha$ , donde  $\alpha = P_{\theta_0}(T > t_0)$ .*

*Demostración.* Ver detalles de la prueba en [20, p. 391]. ☑

**Nota 7.7.** Según [12, p. 228], en la familia exponencial de un parámetro

$$p(\mathbf{x}; \theta) = \{ \exp [c(\theta)T(\mathbf{x}) + d(\theta) + S(\mathbf{x})] \} I_A(\mathbf{x}), \quad (7.12)$$

donde  $I_A(\mathbf{x})$  es la función indicadora del conjunto  $A$ , se sabe que la estadística  $T(\mathbf{X})$  es una estadística suficiente para  $\theta$ . Si  $\theta_2 > \theta_1$ , entonces la razón

$$\frac{p(\mathbf{x}; \theta_2)}{p(\mathbf{x}; \theta_1)} = \exp \{ [c(\theta_2) - c(\theta_1)]T(\mathbf{x}) + [d(\theta_2) - d(\theta_1)] \}.$$

Luego, si  $c(\theta)$  es creciente en  $\theta \in \Theta$ , entonces la familia  $p(\mathbf{x}; \theta)$  tiene la propiedad de **RVM** en  $T(\mathbf{x})$ .

**Ejemplo 7.12.** Sean  $X_1, X_2, X_3, \dots, X_n$  variables aleatorias **iid** tales que  $X_1 \sim N(\theta, \sigma^2)$  con  $\sigma^2$  fija. Considere las hipótesis  $H_0 : \theta \leq \theta_0$  vs.  $H_1 : \theta > \theta_0$ . Determine un test **UMP** de nivel  $\alpha$ .

*Solución.* Como la distribución normal pertenece a la familia exponencial se sabe que  $\bar{X}$  es una estadística suficiente para  $\theta$  y, por la nota 7.7, que la distribución normal tiene la propiedad de **RVM** en  $\bar{x}$ . Usando el teorema de

Karlin-Rubin, teorema 7.3, se tiene que el test rechaza  $H_0 : \theta \leq \theta_0$  si y solo si  $\bar{X} > t_0$  es un test UMP de nivel  $\alpha$ , es decir,

$$\alpha = P_{\theta_0}(\bar{X} > t_0) = P_{\theta_0}\left(\underbrace{\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}}}_Z > \frac{t_0 - \theta_0}{\sigma/\sqrt{n}}\right).$$

Por la regla del complemento:

$$P_{\theta_0}\left(Z < \frac{t_0 - \theta_0}{\sigma/\sqrt{n}}\right) = 1 - \alpha$$

$$\frac{t_0 - \theta_0}{\sigma/\sqrt{n}} = z_{1-\alpha} \quad \Rightarrow \quad t_0 = \frac{z_{1-\alpha}\sigma}{\sqrt{n}} + \theta_0,$$

luego, el test rechaza  $H_0$  si

$$\bar{X} > \frac{z_{1-\alpha}\sigma}{\sqrt{n}} + \theta_0,$$

es un test UMP de nivel  $\alpha$ . □

**Ejemplo 7.13.** De acuerdo con [12, p. 258], dadas  $X_1, X_2, X_3, \dots, X_n$  variables aleatorias iid tal que  $X_1 \sim N(\mu, \theta^2)$ . Considerando las hipótesis

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_a : \mu \neq \mu_0,$$

donde  $\mu_0$  es una constante preestablecida. Determine un test UMP de nivel  $\alpha$ .

*Solución.* En este caso,  $\Theta_0 = \{(\mu_0, \theta_0^2)\}$  para  $\mu_0$  dado. Para el  $\sup_{\Theta_0} L(\theta; \mathbf{x}, \mu_0)$ :

$$\begin{aligned} \frac{\partial \ell(\theta; \mathbf{x}, \mu_0)}{\partial \theta_0^2} &= \frac{\partial \left( -\frac{n}{2} \ln(2\pi\theta_0^2) - \sum_{i=1}^n \frac{(x_i - \mu_0)^2}{2\theta_0^2} \right)}{\partial \theta_0^2} \\ &= \frac{\partial \left( -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\theta_0^2) - \frac{1}{2\theta_0^2} \sum_{i=1}^n (x_i - \mu_0)^2 \right)}{\partial \theta_0^2} \\ &= -\frac{n}{2\theta_0^2} + \frac{1}{2\theta_0^4} \sum_{i=1}^n (x_i - \mu_0)^2, \end{aligned}$$

para determinar el punto crítico se iguala a cero y se obtiene:

$$-\frac{n}{2\hat{\theta}_0^2} + \frac{1}{2\hat{\theta}_0^4} \sum_{i=1}^n (x_i - \mu_0)^2 = 0$$

$$-n\hat{\theta}_0^2 + \sum_{i=1}^n (x_i - \mu_0)^2 = 0 \quad \Rightarrow \quad \hat{\theta}_0^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2.$$

Por otro lado,  $\sup_{\Theta} L(\theta^2; \mathbf{x})$  se tiene cuando  $\hat{\theta}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ , ver ejemplo 4.15. Luego el LRT es dado por:

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta^2; \mathbf{x})}{\sup_{\Theta} L(\theta^2; \mathbf{x})} = \frac{(\mathbf{2}\pi)^{-\frac{n}{2}} (\hat{\theta}_0^2)^{-\frac{n}{2}} \exp \left\{ \frac{-1}{2\hat{\theta}_0^2} n \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2 \right\}}{(\mathbf{2}\pi)^{-\frac{n}{2}} (\hat{\theta}^2)^{-\frac{n}{2}} \exp \left\{ \frac{-1}{2\hat{\theta}^2} n \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right\}}$$

$$= \left( \frac{\hat{\theta}_0^2}{\hat{\theta}^2} \right)^{-\frac{n}{2}} \exp \left\{ \frac{-n}{2} + \frac{n}{2} \right\} = \left( \frac{\hat{\theta}_0^2}{\hat{\theta}^2} \right)^{-\frac{n}{2}}.$$

Luego, para  $0 \leq c \leq 1$ , la región de rechazo es de la forma

$$\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\} = \left\{ \mathbf{x} : \left( \frac{\hat{\theta}_0^2}{\hat{\theta}^2} \right)^{-\frac{n}{2}} \leq c \right\} = \left\{ \mathbf{x} : \frac{-n}{2} \ln \left( \frac{\hat{\theta}_0^2}{\hat{\theta}^2} \right) \leq \ln(c) \right\}$$

$$= \left\{ \mathbf{x} : \ln \left( \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right) \geq \frac{-2}{n} \ln(c) \right\}$$

$$= \left\{ \mathbf{x} : \ln \left( \frac{\sum_{i=1}^n [(x_i - \bar{x}) + (\bar{x} - \mu_0)]^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \geq \frac{-2}{n} \ln(c) \right\}.$$

Nótese que:

$$\sum_{i=1}^n [(x_i - \bar{x}) + (\bar{x} - \mu_0)]^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$+ 2 \sum_{i=1}^n (x_i - \bar{x})(\bar{x} - \mu_0) + n(\bar{x} - \mu_0)^2,$$

luego,

$$\begin{aligned}
 \{x : \lambda(x) \leq c\} &= \left\{ x : \ln \left( \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \geq \frac{-2}{n} \ln(c) \right\} \\
 &= \left\{ x : \ln \left( 1 + \frac{(\bar{x} - \mu_0)^2}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right) \geq \frac{-2}{n} \ln(c) \right\} \\
 &= \left\{ x : \ln \left( 1 + \frac{(\bar{x} - \mu_0)^2}{\hat{\theta}^2} \right) \geq \frac{-2}{n} \ln(c) \right\} \\
 &= \left\{ x : 1 + \frac{(\bar{x} - \mu_0)^2}{\hat{\theta}^2} \geq \exp \left( \frac{-2}{n} \ln(c) \right) \right\} \\
 &= \left\{ x : \frac{(\bar{x} - \mu_0)^2}{\hat{\theta}^2} \geq \exp \left( \frac{-2}{n} \ln(c) \right) - 1 \right\},
 \end{aligned}$$

corresponde a un test **LRT** para probar la hipótesis  $H_0 : \mu = \mu_0$ . Además,

$$\begin{aligned}
 \alpha &= P_{\theta_0} \left( \frac{(\bar{X} - \mu_0)^2}{\hat{\theta}^2} \geq \exp \left[ \frac{-2}{n} \ln(c) \right] - 1 \right) \\
 &= P_{\theta_0} \left( \left| \frac{(\bar{X} - \mu_0)}{\hat{\theta}} \right| \geq \underbrace{\sqrt{\exp \left[ \frac{-2}{n} \ln(c) \right] - 1}}_d \right) \\
 &= P_{\theta_0} \left( \left| \frac{(\bar{X} - \mu_0)}{\hat{\theta}} \right| \geq d \right) = P_{\theta_0} \left( \left| \frac{(\bar{X} - \mu_0)}{\sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}} \right| \geq d \right) \\
 &= P_{\theta_0} \left( \left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sqrt{(n-1) \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}} \right| \geq d \right) = P_{\theta_0} \left( \left| \frac{\bar{X} - \mu_0}{S_n / \sqrt{n}} \right| \geq \sqrt{n-1} d \right),
 \end{aligned}$$

bajo  $H_0$  se tiene que:

$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{S_n} \sim t_{n-1}$$

y entonces

$$\alpha = P_{\theta_0} \left( \left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{S_n} \right| \geq \sqrt{n-1} d \right) = P_{\theta_0} \left( |T_n| \geq \sqrt{n-1} d \right)$$

$$\begin{aligned}
 &= P_{\theta_0} \left( T_n \leq -\sqrt{n-1}d \right) + P_{\theta_0} \left( T_n \geq \sqrt{n-1}d \right) \\
 &\quad \underbrace{\hspace{10em}}_{P_{\theta_0} \left( T_n \geq \sqrt{n-1}d \right)} \quad \underbrace{\hspace{10em}}_{1-P_{\theta_0} \left( T_n < \sqrt{n-1}d \right)} \\
 &\quad \underbrace{\hspace{10em}}_{1-P_{\theta_0} \left( T_n < \sqrt{n-1}d \right)}
 \end{aligned}$$

Aquí se usó que la **fdp**  $t$ -Student es simétrica con respecto al eje  $x = 0$ , lo cual implica en términos de la **fd** que el área a la izquierda de  $-x$  es igual al área a la derecha de  $x$ . Por lo tanto,

$$\begin{aligned}
 2 \left[ 1 - P_{\theta_0} \left( T_n < \sqrt{n-1}d \right) \right] &= \alpha \quad \Rightarrow \quad 1 - P_{\theta_0} \left( T_n < \sqrt{n-1}d \right) = \alpha/2 \\
 P_{\theta_0} \left( T_n < \sqrt{n-1}d \right) &= 1 - \frac{\alpha}{2} \quad \Rightarrow \quad t_{n-1, 1-\frac{\alpha}{2}} = \sqrt{n-1}d,
 \end{aligned}$$

es decir,

$$d = \frac{t_{n-1, 1-\frac{\alpha}{2}}}{\sqrt{n-1}}.$$

Por tanto, el test **UMP** rechaza  $H_0 : \mu = \mu_0$  si

$$\begin{aligned}
 |T_n| \geq \sqrt{n-1}d &\quad \Rightarrow \quad |T_n| \geq \sqrt{n-1} \frac{t_{n-1, 1-\frac{\alpha}{2}}}{\sqrt{n-1}} \\
 |T_n| &\geq t_{n-1, 1-\frac{\alpha}{2}}. \quad \checkmark
 \end{aligned}$$

Para encontrar la distribución de  $\lambda(x)$  se usa el teorema 7.4, que brinda la distribución asintótica de la razón de verosimilitud generalizada.

**Teorema 7.4.** Según [20, p. 490], dadas  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una **fdp** o **fmp**  $f(x; \theta)$ . Bajo condiciones de regularidad, si  $\theta \in \Theta_0$  entonces la distribución de la estadística  $-2 \ln [\lambda(x)]$  converge a la distribución chi-cuadrado cuando el tamaño de muestra tiende a infinito. Los grados de libertad de la distribución límite son la diferencia entre el número de parámetros libres dados por  $\theta \in \Theta_0$  y el número de parámetros libres dados por  $\theta \in \Theta$ . Así, se rechaza  $H_0$  si y solo si

$$-2 \ln (\lambda(x)) \geq \chi_{v, 1-\alpha}^2,$$

a un nivel de significancia del  $\alpha$ , es decir, para cada  $\theta \in \Theta_0$ :

$$\lim_{n \rightarrow \infty} P_{\theta} \{x : \lambda(x) \leq c\} = \alpha$$

*Demostración.* Ver detalles de la prueba en [20, p. 490]. \checkmark

## 7.4. Algunos casos particulares

Antes de pasar a algunos casos particulares, considerando que el *valor P* es el nivel de significancia más pequeño que conduce al rechazo es de la hipótesis nula,  $H_0$ , para su cálculo se debe tener en cuenta la hipótesis alterna. Por ejemplo, si  $z_c$  es el estadístico calculado a partir de la información en la muestra, es decir, es el estadístico de contraste<sup>46</sup>, el valor  $P$  es:

$$P - \text{valor} = \begin{cases} 2[1 - \Phi(|z_c|)] & \text{para } H_1 : \theta \neq \theta_0 \\ 1 - \Phi(z_c) & \text{para } H_1 : \theta > \theta_0 \\ \Phi(z_c) & \text{para } H_1 : \theta < \theta_0 \end{cases}$$

**Ejemplo 7.14.** Considerando la información del ejemplo 6.17 sobre dos grupos de ratas con diferentes dietas, se tenía que:

Dieta	$n_k$	$\hat{\mu}_k = \bar{x}_k$	$\hat{\sigma}_k^2 = s_k^2$
A	10	274,7	124,9
B	8	535,5	4762,2857

1. ¿Existe alguna razón para creer, a un nivel de significancia del 0,05, que el peso en gramos luego de 64 días de aplicada la dieta A es mayor de 265 gramos? (suponga que la varianza es conocida  $\sigma_A^2 = 180$ ). Además, calcule el  $P$ -valor.
2. ¿Existe alguna razón para creer, a un nivel de significancia del 0,1, que el peso en gramos luego de 64 días de aplicada la dieta B no es igual a 500 gramos? (suponga que la varianza es desconocida).

*Solución.* Para cada pregunta planteada el procedimiento es el siguiente:

1. Suponiendo que la varianza es conocida:

- a) Se sabe que  $n_A = 10$  y  $\hat{\mu}_A = \bar{x}_A = 274,7$ .
- b) Del enunciado se tiene que  $\mu_0 = 265$ ,  $\sigma_A^2 = 180$  y  $\alpha = 0,05$ .
- c) Las hipótesis nula y alterna son:

$$H_0 : \mu \leq 265 \quad \text{vs.} \quad H_1 : \mu > 265.$$

<sup>46</sup>**Estadístico de contraste.** Es un mecanismo mediante el cual se rechaza la hipótesis nula cuando existen diferencias significativas entre los valores muestrales y los valores teóricos, y se acepta en caso contrario. El estadístico de contraste sigue una distribución determinada conocida y para cada muestra tomará un valor particular.

d) El estadístico para el contraste es:

$$z_c = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{274,7 - 265}{\frac{13,4164}{\sqrt{10}}} = 2,2863.$$

e) El criterio de rechazo según la hipótesis alterna es: se rechaza  $H_0$  cuando

$$z_c \geq z_{1-\alpha}.$$

f) Buscando para  $\alpha = 0,05$  en la tabla C.1:

$$z_{1-\alpha} = z_{0,95} = 1,645.$$

g) Como  $2,2863 > 1,645$ , se toma la decisión de rechazar la hipótesis nula porque los datos apoyan la evidencia para considerar que dicha hipótesis no es cierta. Así, el peso en gramos luego de 64 días de aplicada la dieta A es mayor de 265 gramos con un nivel de significancia del 0,05.

h) **Valor  $P$ .** Como la hipótesis alterna es  $H_1 : \mu > 265$ , entonces

$$P - \text{valor} = 1 - \Phi(z_c) = 1 - \Phi(2,2863) = 1 - 0,9857 = 0,0143.$$

Como el  $P$ -valor es menor que el nivel de significancia propuesto para el test ( $\alpha = 0,05$ ), se rechaza la hipótesis nula y se concluye que el peso en gramos luego de 64 días de aplicada la dieta A es mayor de 265 gramos con un  $P$ -valor de 0,0143, ver figura 7.7.

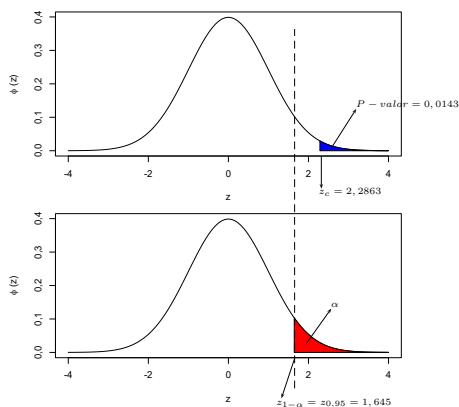


Figura 7.7. Zona de rechazo y  $P$ -valor, ejemplo 7.14, ítem 1

2. Suponiendo que la varianza es desconocida:

a) Se sabe que  $n_B = 8$ ,

$$\hat{\mu}_B = \bar{x}_B = 535,5 \quad \text{y} \quad \hat{\sigma}_B^2 = s_B^2 = 4762,2857.$$

b) Del enunciado se tiene que  $\mu_0 = 500$  y  $\alpha = 0,10$ .

c) Las hipótesis nula y alterna son:

$$H_0 : \mu = 500 \quad \text{vs.} \quad H_1 : \mu \neq 500.$$

d) El estadístico para el contraste es:

$$t_c = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{535,5 - 500}{\frac{69,0093}{\sqrt{8}}} = 1,4550.$$

e) El criterio de rechazo según la hipótesis alterna es: se rechaza  $H_0$  cuando  $|t_c| \geq t_{n-1, 1-\frac{\alpha}{2}}$

$$t_c \leq t_{\frac{\alpha}{2}, n-1} \quad \text{o} \quad t_c \geq t_{1-\frac{\alpha}{2}, n-1}.$$

f) Se busca para  $\alpha = 0,05$  en la tabla C.2:

$$t_{\frac{\alpha}{2}, n-1} = t_{0,05,7} = -1,895 \quad \text{y} \quad t_{1-\frac{\alpha}{2}, n-1} = t_{0,95,7} = 1,895.$$

g) Como  $1,4550 > -1,895$  y  $1,4550 < 1,895$  se toma la decisión de no rechazar la hipótesis nula porque los datos no dan suficiente evidencia para considerar que dicha hipótesis es cierta. Así, el peso en gramos luego de 64 días de aplicada la dieta B no se puede decir que sea diferente de 500 gramos con un nivel de significancia del 0,1.

h) **Valor P.** Como la hipótesis alterna es  $H_1 : \mu \neq 500$ , entonces

$$P - \text{valor} = 2[1 - F(|t_c|, 7)] = 2[1 - F(|1,4550|; 7)].$$

con  $F$  indicando la fda de la distribución  $t$  de Student. Lo usual es no encontrar el valor exacto en la tabla, por lo cual se realiza una aproximación del  $P$ -valor, pero se puede obtener usando un paquete estadístico como R. En este ejemplo el valor más cercano es para  $F(1,415; 7) = 0,900$ , entonces

$$P - \text{valor} \approx 2[1 - F(1,415; 7)] = 2[1 - 0,90] = 0,20.$$

El valor exacto del  $P$ -valor es:

$$P - \text{valor} = 2[1 - F(1,415; 7)] = 2(0,0945) = 0,1890.$$

Como el  $P$ -valor es mayor que el nivel de significancia, se concluye que no se puede rechazar la hipótesis nula, es decir, el peso en gramos luego de 64 días de aplicada la dieta B no se puede decir que sea diferente de 500 gramos con un  $P$ -valor igual a 0,1890, ver figura 7.8.

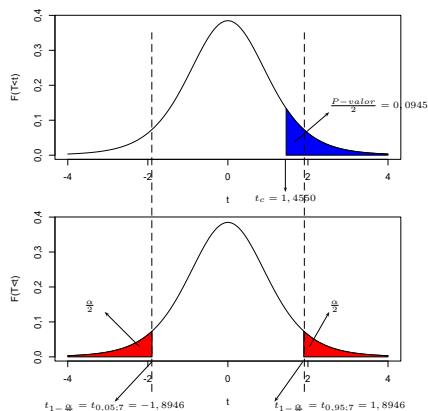


Figura 7.8. Zona de rechazo y  $P$ -valor, ejemplo 7.14, ítem 2



**Ejemplo 7.15.** Se sabe que aproximadamente 1 de 10 fumadores prefieren la marca de cigarrillos  $A$ . Tras una campaña publicitaria, en una región, se entrevistó a 200 fumadores para determinar la eficiencias de la campaña. El resultado de esta muestra es que un total de 32 personas indicaron preferencia por la marca  $A$ . ¿Puede considerarse que estos datos presentan evidencia suficiente para indicar un aumento en la aceptación de la marca  $A$ ? Utilizar un nivel de significancia del 0,01.

**Solución.** Del enunciado se tiene que:

1.  $p_0 = \frac{1}{10} = 0,10$ ,  $\hat{\theta}_A = \frac{x}{n} = \frac{32}{200} = 0,16$  y  $\alpha = 0,01$ .
2. Las hipótesis nula y alterna son:

$$H_0 : p = 0,10 \quad \text{vs.} \quad H_1 : p > 0,10.$$

3. El estadístico para el contraste es:

$$z_c = \frac{\hat{\theta} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0,16 - 0,10}{\sqrt{\frac{0,10 * 0,90}{200}}} = 2,8284.$$

4. El criterio de rechazo según la hipótesis alterna es: se rechaza  $H_0$  cuando

$$z_c \geq z_{1-\alpha}.$$

5. Buscando para  $\alpha = 0,01$  en la tabla C.1:

$$z_{1-\alpha} = z_{0,99} = 2,325.$$

6. Como  $2,8284 > 2,325$  se concluye que la campaña publicitaria sí tuvo un efecto y aumentó la aceptación de la marca A de cigarrillos con nivel de significancia de 0,01.

7. **Valor P.** Como la hipótesis alterna es  $H_1 : p > 0,10$ , entonces

$$P - \text{valor} = 1 - \Phi(z_c) = 1 - \Phi(2,8284) = 1 - 0,9977 = 0,0023.$$

Como el  $P$ -valor es menor que el nivel de significancia se rechaza la hipótesis nula y se concluye que la campaña publicitaria sí tuvo un efecto y aumentó la aceptación de la marca A de cigarrillos con un  $P$ -valor de 0,0023.

El programa en R para el ejemplo 7.15 es<sup>47</sup>:

```
remove(list=ls())
prop.test(32,200,p=0.10,alternative="greater",conf.level=0.99,
correct=FALSE)
prop.test(32,200,p=0.10,alternative="greater",conf.level=0.99,
correct=TRUE)
```

☑

## 7.4.1. Prueba de hipótesis sobre una varianza cuando $\mu$ es una constante desconocida.

Sea

$$H_0 : \theta^2 \leq \theta_0^2 \quad \text{vs.} \quad H_a : \theta^2 > \theta_0^2$$

<sup>47</sup>El test que arroja la función `prop.test` de R calcula el estadístico  $\chi_c^2$  y no el estadístico  $z_c$ .

Así  $\chi_{c,1}^2 = z_c^2$ . Este caso  $z_c = \sqrt{\chi_c^2} = \sqrt{8} = 2,8284$ .

y  $\mu = 0$ , usando la representación de la familia exponencial dada en (5.27):

$$\begin{aligned} p(\mathbf{x}; \theta^2) &= \frac{1}{(2\pi\theta^2)^{\frac{n}{2}}} \exp\left\{\frac{-1}{2\theta^2} \sum_{i=1}^n x_i^2\right\} \\ &= \exp\left\{\underbrace{-\frac{n}{2} \ln(2\pi\theta^2)}_{d(\theta)} + \underbrace{\left(-\frac{1}{2\theta^2}\right)}_{c(\theta)} \underbrace{\sum_{i=1}^n x_i^2}_{T(\mathbf{x})}\right\}, \end{aligned}$$

como  $c(\theta) = \frac{-1}{\theta^2}$  es una función creciente en  $\theta$ , entonces el test rechaza  $H_0 : \theta^2 \leq \theta_0^2$  si y solo si  $T(\mathbf{x}) = \sum_{i=1}^n X_i^2 > t_0$  es un test **UMP** de nivel  $\alpha$ , donde  $\alpha = P_{\theta_0}(T > t_0)$ , ver teorema 7.3 y nota 7.7.

$$\begin{aligned} \alpha &= P_{\theta_0}(T > t_0) = P_{\theta_0}\left(\sum_{i=1}^n X_i^2 > t_0\right) = P_{\theta_0}\left(\frac{1}{\theta_0^2} \sum_{i=1}^n X_i^2 > \frac{t_0}{\theta_0^2}\right) \\ &= P_{\theta_0}\left(\chi_n^2 > \frac{t_0}{\theta_0^2}\right) \end{aligned}$$

$$\chi_{n,1-\alpha}^2 = \frac{t_0}{\theta_0^2} \quad \Leftrightarrow \quad \theta_0^2 \chi_{n,1-\alpha}^2 = t_0,$$

y, por tanto, el test **UMP** rechaza  $H_0 : \theta^2 \leq \theta_0^2$ , cuando  $\mu = 0$ , si

$$\sum_{i=1}^n X_i^2 > \theta_0^2 \chi_{n,1-\alpha}^2,$$

es decir,

$$\frac{1}{\theta_0^2} \sum_{i=1}^n X_i^2 > \chi_{n,1-\alpha}^2.$$

Similarmente (ver [62, p. 239]), el test **UMP** rechaza  $H_0 : \theta^2 \leq \theta_0^2$ , cuando  $\mu$  es constante desconocida, si

$$\sum_{i=1}^n (X_i - \bar{X})^2 > \theta_0^2 \chi_{n-1,1-\alpha}^2,$$

multiplicando y dividiendo por  $(n-1)$  se obtiene que:

$$\frac{(n-1)S^2}{\theta_0^2} > \chi_{n-1,1-\alpha}^2. \quad (7.13)$$

**Ejemplo 7.16.** Considerando la información del ejemplo 6.17 sobre dos grupos de ratas con diferentes dietas, se tenía que:

$$\hat{\mu}_B = 535,5 \quad \hat{\sigma}_B^2 = s_B^2 = 4762,2857 \quad n_B = 8.$$

¿Existe alguna razón para creer, a un nivel de significancia del 0,05, que la varianza del peso en gramos luego de 64 días de aplicada la dieta B sea mayor de 2000?

*Solución.* De los datos del enunciado:

1. Se sabe

$$\hat{\sigma}_B^2 = s_B^2 = 4762,2857 \quad \text{y} \quad n_B = 8.$$

2. Además, se tiene que  $\theta_0^2 = 2000$  y  $\alpha = 0,05$ .
3. Las hipótesis nula y alterna son

$$H_0 : \theta^2 \leq 2000 \quad \text{vs.} \quad H_a : \theta^2 > 2000.$$

4. El estadístico para el contraste es dado en (7.13):

$$\chi_c^2 = \frac{(n-1)S^2}{\theta_0^2} = \frac{(7)4762,2857}{2000} = 16,668.$$

5. El criterio de rechazo según la hipótesis alterna es: se rechaza  $H_0$  cuando

$$\chi_c^2 \geq \chi_{n-1,1-\alpha}^2.$$

6. Buscando para  $\alpha = 0,05$  en la tabla C.3:

$$\chi_{n-1,1-\alpha}^2 = \chi_{7,0,95}^2 = 14,07.$$

7. Como  $16,668 > 14,07$ , se toma la decisión de rechazar la hipótesis nula porque los datos apoyan la evidencia para considerar que dicha hipótesis no es cierta. En consecuencia, la varianza del peso en gramos luego de 64 días de aplicada la dieta B es mayor de 2000 con un nivel de significancia del 0,05.
8. **Valor P.** Como la hipótesis alterna es  $H_a : \theta^2 > 2000$ , entonces

$$\begin{aligned} P\text{-valor} &= 1 - P(\chi_{n-1}^2 < \chi_c^2) = 1 - P(\chi_7^2 < 16,668) \\ &= 1 - 0,9803 = 0,0197. \end{aligned}$$

Como el  $P$ -valor es menor que el nivel de significancia se rechaza la hipótesis nula y se concluye que la varianza del peso en gramos luego de 64 días de aplicada la dieta B es mayor de 2000 con un  $P$ -valor de 0,0197, ver figura 7.9.

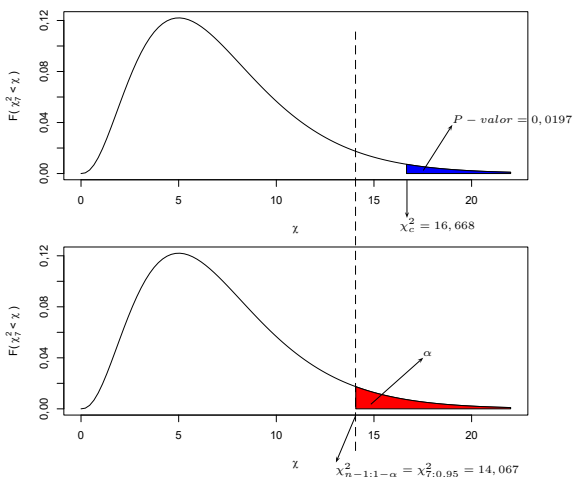


Figura 7.9. Zona de rechazo y  $P$ -valor, ejemplo 7.16



## 7.4.2. Prueba de hipótesis sobre diferencia de medias de dos poblaciones independientes

Supóngase que  $X_1, X_2, X_3, \dots, X_n$  son variables aleatorias iid de modo que  $X_1 \sim N(\mu_X, \sigma_X^2)$  y  $Y_1, Y_2, Y_3, \dots, Y_m$  son variables aleatorias iid tales que  $Y_1 \sim N(\mu_Y, \sigma_Y^2)$ , donde las muestras aleatorias son independientes entre sí. Nos interesa hacer hipótesis sobre la diferencia de sus medias poblacionales. Se consideran tres casos: varianzas conocidas, varianzas desconocidas que se pueden considerar iguales a nivel poblacional y varianzas desconocidas que no se pueden considerar iguales a nivel poblacional.

### 7.4.2.1. Varianzas conocidas

Se sabe que:

$$\bar{X}_n \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right) \quad \text{y} \quad \bar{Y}_m \sim N\left(\mu_Y, \frac{\sigma_Y^2}{m}\right)$$

$$\begin{aligned}\bar{X}_n - \bar{Y}_m &\sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right) \\ \frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} &\sim N(0, 1).\end{aligned}\quad (7.14)$$

Si las hipótesis planteadas son:

$$H_0 : \mu_X - \mu_Y \leq \theta_0 \quad \text{vs.} \quad H_a : \mu_X - \mu_Y > \theta_0,$$

Análogamente a los resultados del ejemplo 7.12, reemplazando  $\mu_X - \mu_Y = \theta_0$  en 7.14, el test rechaza  $H_0$  si

$$\frac{\bar{X} - \bar{Y} - \theta_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} > z_{1-\alpha} \quad (7.15)$$

es un test **UMP** de nivel  $\alpha$ .

Si las hipótesis planteadas fueran  $H_0 : \mu_X - \mu_Y \geq \theta_0$  vs.  $H_a : \mu_X - \mu_Y < \theta_0$  el test rechaza  $H_0$  si

$$\frac{\bar{X} - \bar{Y} - \theta_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} < -z_{1-\alpha}$$

es un test **UMP** de nivel  $\alpha$ .

Si las hipótesis planteadas fueran  $H_0 : \mu_X - \mu_Y = \theta_0$  vs.  $H_a : \mu_X - \mu_Y \neq \theta_0$  el test rechaza  $H_0$  si

$$\left| \frac{\bar{X} - \bar{Y} - \theta_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \right| > z_{1-\frac{\alpha}{2}}$$

es un test insesgado **UMP** de nivel  $\alpha$ .

#### 7.4.2.2. Varianzas desconocidas que se pueden considerar iguales a nivel poblacional

Cuando las varianzas poblacionales no se conocen, pero se puede suponer que  $\sigma_X^2 = \sigma_Y^2 = \sigma^2$  se tiene que:

$$\bar{X}_n \sim N\left(\mu_X, \frac{\sigma^2}{n}\right) \quad \text{y} \quad \bar{Y}_m \sim N\left(\mu_Y, \frac{\sigma^2}{m}\right).$$

Luego, la expresión (7.14) queda:

$$\begin{aligned}\bar{X}_n - \bar{Y}_m &\sim N\left(\mu_X - \mu_Y, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}\right) \\ \frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} &\sim N(0,1).\end{aligned}$$

Del capítulo 6, casos particulares, se sabe que:

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} = T \sim t_{n+m-2}, \quad (7.16)$$

con  $S_p^2$  el estimador de la varianza  $\sigma^2$  dado en (6.38).

- Si las hipótesis planteadas son:

$$H_0 : \mu_X - \mu_Y \leq \theta_0 \quad \text{vs.} \quad H_a : \mu_X - \mu_Y > \theta_0.$$

Reemplazando  $\mu_X - \mu_Y = \theta_0$  en (7.16), el test rechaza  $H_0$  si

$$\frac{\bar{X} - \bar{Y} - \theta_0}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} > t_{n+m-2, 1-\alpha} \quad (7.17)$$

es un test **UMP** de nivel  $\alpha$ .

- Si las hipótesis fueran  $H_0 : \mu_X - \mu_Y \geq \theta_0$  vs.  $H_a : \mu_X - \mu_Y < \theta_0$  y se reemplaza  $\mu_X - \mu_Y = \theta_0$  en (7.16), el test rechaza  $H_0$  si

$$\frac{\bar{X} - \bar{Y} - \theta_0}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} < -t_{n+m-2, 1-\alpha}$$

es un test **UMP** de nivel  $\alpha$ .

- Si las hipótesis fueran  $H_0 : \mu_X - \mu_Y = \theta_0$  vs.  $H_a : \mu_X - \mu_Y \neq \theta_0$  y se reemplaza  $\mu_X - \mu_Y = \theta_0$  en (7.16), el test rechaza  $H_0$  si

$$\left| \frac{\bar{X} - \bar{Y} - \theta_0}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \right| > t_{n+m-2, 1-\frac{\alpha}{2}}$$

es un test insesgado **UMP** de nivel  $\alpha$ .

### 7.4.2.3. Varianzas desconocidas que no se pueden suponer iguales a nivel poblacional

Se sabe que:

$$\begin{aligned}\bar{X}_n &\sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right) \quad \text{y} \quad \bar{Y}_m \sim N\left(\mu_Y, \frac{\sigma_Y^2}{m}\right) \\ \bar{X}_n - \bar{Y}_m &\sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right) \\ \frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} &\sim N(0, 1).\end{aligned}$$

Del capítulo 6, casos particulares, se sabe que se usa como aproximación:

$$\frac{\bar{X}_n - \bar{Y}_m - \theta_0}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \sim t_v,$$

donde  $v$  es dado en (6.40) o (6.42).

- Si las hipótesis planteadas son:

$$H_0 : \mu_X - \mu_Y \leq \theta_0 \quad \text{vs.} \quad H_a : \mu_X - \mu_Y > \theta_0,$$

el test rechaza  $H_0 : \mu_X - \mu_Y \leq \theta_0$  si

$$\frac{\bar{X} - \bar{Y} - \theta_0}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} > t_{v, 1-\alpha} \quad (7.18)$$

es un test **UMP** de nivel  $\alpha$ .

- Si las hipótesis fueran  $H_0 : \mu_X - \mu_Y \geq \theta_0$  vs.  $H_a : \mu_X - \mu_Y < \theta_0$  el test rechaza  $H_0$  si

$$\frac{\bar{X} - \bar{Y} - \theta_0}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} < -t_{v, 1-\alpha}$$

es un test **UMP** de nivel  $\alpha$ .

- Si las hipótesis fueran  $H_0 : \mu_X - \mu_Y = \theta_0$  vs.  $H_a : \mu_X - \mu_Y \neq \theta_0$  el test rechaza  $H_0$  si

$$\left| \frac{\bar{X} - \bar{Y} - \theta_0}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \right| > t_{v, 1 - \frac{\alpha}{2}}$$

es un test insesgado **UMP** de nivel  $\alpha$ .

**Ejemplo 7.17.** Considerando la información del ejemplo 6.17 sobre dos grupos de ratas con diferentes dietas, se tenía que:

Dieta	$n_k$	$\hat{\mu}_k = \bar{x}_k$	$\hat{\sigma}_k^2 = s_k^2$
A	10	274,7	124,9
B	8	535,5	4762,2857

1. ¿Existen evidencias para considerar, a un nivel de significancia del 0,01, que el peso en gramos luego de 64 días de aplicada la dieta B excede a la dieta A en 250 gramos? Suponiendo que las varianzas eran conocidas ( $\sigma_A^2 = 180$  y  $\sigma_B^2 = 3000$ ).
2. Responder la pregunta anterior si las varianzas son desconocidas y se consideran distintas, considerando el resultado obtenido en el ejemplo 6.17.

*Solución.* Para cada pregunta planteada el procedimiento es el siguiente:

1. Suponiendo que las varianzas son conocidas

a) Se sabe que  $n_A = 10$  y  $n_B = 8$ , además que

$$\begin{aligned} \hat{\mu}_A &= 274,7 & \hat{\mu}_B &= 535,5 \\ \sigma_A^2 &= 180 & \sigma_B^2 &= 3000. \end{aligned}$$

b) Del enunciado se tiene que  $\theta_0 = 250$  y  $\alpha = 0,01$ .

c) Las hipótesis nula y alterna son:

$$H_0 : \mu_B - \mu_A \leq 250 \quad \text{vs.} \quad H_a : \mu_B - \mu_A > 250.$$

d) El estadístico para el contraste es dado en (7.15):

$$z_c = \frac{\bar{X} - \bar{Y} - \theta_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} = \frac{535,5 - 274,7 - 250}{\sqrt{\frac{180}{10} + \frac{3000}{8}}} = 0,5448.$$

- e) El criterio de rechazo según la hipótesis alterna es: se rechaza  $H_0$  cuando

$$z_c \geq z_{1-\alpha}.$$

- f) Buscando para  $\alpha = 0,01$  en la tabla C.1:

$$z_{1-\alpha} = z_{0,99} = 2,325.$$

- g) Como  $0,5448 < 2,325$ , se toma la decisión de no rechazar la hipótesis nula. En consecuencia, no se puede rechazar la hipótesis de que el peso en gramos luego de 64 días de aplicada la dieta B sea diferente a la dieta A en 250 gramos con un nivel de significancia del 0,01.
- h) **Valor P.** Como la hipótesis alterna es  $H_a : \mu_B - \mu_A > 250$ , entonces el valor  $P$  es:

$$\begin{aligned} P - \text{valor} &= 1 - P(Z < z_c) = 1 - P(Z < 0,5448) \\ &= 1 - 0,7070545 = 0,293. \end{aligned}$$

Como el  $P$ -valor es mayor que el nivel de significancia no se rechaza la hipótesis de que el peso en gramos luego de 64 días de aplicada la dieta B sea diferente a la dieta A en 250 gramos con un  $P$ -valor de 0,293.

## 2. Suponiendo que las varianzas son desconocidas y distintas

- a) Se sabe que  $n_A = 10$  y  $n_B = 8$ , además que

$$\begin{aligned} \hat{\mu}_A &= 274,7 & \hat{\mu}_B &= 535,5 \\ \hat{\sigma}_A^2 &= s_A^2 = 124,9 & \hat{\sigma}_B^2 &= s_B^2 = 4762,2857. \end{aligned}$$

- b) Del enunciado se tiene que  $\theta_0 = 250$  y  $\alpha = 0,01$ .
- c) Las hipótesis nula y alterna son:

$$H_0 : \mu_B - \mu_A \leq 250 \quad \text{vs.} \quad H_a : \mu_B - \mu_A > 250.$$

- d) El estadístico para el contraste es dado en (7.16):

$$t_c = \frac{\bar{X} - \bar{Y} - \theta_0}{\sqrt{\frac{\hat{\sigma}_X^2}{n} + \frac{\hat{\sigma}_Y^2}{m}}} = \frac{535,5 - 274,7 - 250}{\sqrt{\frac{124,9}{10} + \frac{4762,2857}{8}}} = 0,4381.$$

- e) El criterio de rechazo según la hipótesis alterna es: se rechaza  $H_0$  cuando

$$t_c \geq t_{v,1-\alpha},$$

donde  $v$  es dado por (6.40) o (6.42), es decir,  $v \approx 7,294325$ , ver ejemplo 6.23.

- f) Buscando para  $\alpha = 0,01$  en la tabla C.2:

$$t_{v,1-\alpha} = t_{7;0,99} = 2,998.$$

- g) Como  $0,4381 < 2,998$ , se toma la decisión de no rechazar la hipótesis nula. En consecuencia, no se puede rechazar la hipótesis de que el peso en gramos luego de 64 días de aplicada la dieta B sea diferente a la dieta A en 250 gramos con un nivel de significancia del 0,01.
- h) **Valor  $P$ .** Como la hipótesis alterna es  $H_a : \mu_B - \mu_A > 250$ , entonces el valor  $P$  es:

$$\begin{aligned} P - \text{valor} &= 1 - P(T_v < t_c) = 1 - P(T_7 < 0,4381) \\ &= 1 - 0,6627 = 0,3373. \end{aligned}$$

Como el  $P$ -valor es mayor que el nivel de significancia no se rechaza la hipótesis de que el peso en gramos luego de 64 días de aplicada la dieta B sea diferente a la dieta A en 250 gramos con un  $P$ -valor de 0,3373.

El programa en R para el ejemplo 7.17 ítem 2, es

```
remove(list=ls())
dietaA=c(278, 245, 269, 275, 280, 281, 284, 278, 280, 277)
dietaB=c(478, 496, 472, 628, 495, 490, 625, 600)
var.test(dietaA,dietaB,ratio=1,alternative="two.sided",
conf.level=0.90) ## test F para comparar dos
## varianzas e intervalo de confianza
t.test(dietaB,dietaA,mu=250,alternative="greater",
paired=FALSE,var.equal=FALSE,conf.level=0.98) # test
```

### 7.4.3. Pruebas de hipótesis para la igualdad de varianzas

Para determinar si no se cumple el supuesto de varianzas iguales, frecuentemente se aplica una prueba que compara dos varianzas antes de llevar a cabo una prueba  $t$  sobre dos medias.

El problema de probar la igualdad de las varianzas  $\sigma_X^2$  y  $\sigma_Y^2$  de dos poblaciones consiste en probar la hipótesis nula  $H_0 : \sigma_X^2 = \sigma_Y^2$  contra una de las alternas usuales:

$$\sigma_X^2 < \sigma_Y^2, \quad \sigma_X^2 > \sigma_Y^2, \quad \text{o} \quad \sigma_X^2 \neq \sigma_Y^2.$$

Para muestras aleatorias independientes de tamaños  $n$  y  $m$ , respectivamente, de las dos poblaciones, el valor  $F_c$  para probar  $\sigma_X = \sigma_Y$  es el cociente

$$F_c = \frac{S_{X_n}^2}{S_{Y_m}^2}, \quad (7.19)$$

donde  $S_{X_n}^2$  y  $S_{Y_m}^2$  son las varianzas calculadas de las dos muestras. Si las dos poblaciones se distribuyen de forma aproximadamente normal y la hipótesis nula es verdadera, de manera análoga al intervalo de confianza para el cociente de varianzas de distribuciones normales independientes, la razón  $F_c = S_{X_n}^2 / S_{Y_m}^2$  es un valor de la distribución  $F$  con  $v_1 = n - 1$  y  $v_2 = m - 1$  grados de libertad. Por lo tanto, las regiones críticas de tamaño  $\alpha$  que corresponden a las alternas unilaterales  $\sigma_X^2 < \sigma_Y^2$  y  $\sigma_X^2 > \sigma_Y^2$  son, respectivamente,  $F_c < F_\alpha(v_1, v_2)$  y  $F_c > F_{1-\alpha}(v_1, v_2)$ . Para la alterna bilateral  $\sigma_X^2 \neq \sigma_Y^2$  la región crítica es  $F_c < F_{\alpha/2}(v_1, v_2)$  o  $F_c > F_{1-\alpha/2}(v_1, v_2)$ .

**Ejemplo 7.18.** *Al probar la diferencia en el peso en gramos de las dietas del ejemplo 7.17 se supuso que las dos varianzas desconocidas de las poblaciones eran distintas. ¿Se justifica tal suposición? Utilice un nivel de significancia de 0,10.*

**Solución.** De los datos del enunciado:

1. Se sabe que  $n_A = 10$  y  $n_B = 8$ , además que

$$\begin{aligned} \hat{\mu}_A &= 274,7 & \hat{\mu}_B &= 535,5 \\ \hat{\sigma}_A^2 &= s_A^2 = 124,9 & \hat{\sigma}_B^2 &= s_B^2 = 4762,2857. \end{aligned}$$

2. Las hipótesis nula y alterna son:

$$H_0 : \sigma_A^2 = \sigma_B^2 \quad \text{vs.} \quad H_a : \sigma_A^2 \neq \sigma_B^2$$

y  $\alpha = 0,10$ .

3. El estadístico para el contraste es:

$$F_c = \frac{S_B^2}{S_A^2} = \frac{4762,2857}{124,9} = 38,1287,$$

con  $v_1 = 7$  y  $v_2 = 9$  grados de libertad.

4. El criterio de rechazo según la hipótesis alterna es: se rechaza  $H_0$  cuando

$$F_c < F_{\alpha/2}(v_1, v_2) \quad \text{o} \quad F_c > F_{1-\alpha/2}(v_1, v_2).$$

5. Buscando para  $\alpha = 0,10$  en la tabla de la distribución  $F$ :

$$\begin{aligned} F_{\alpha/2}(v_1, v_2) &= F_{0,05}(7, 9) = 0,2720 & \text{y} \\ F_{1-\alpha/2}(v_1, v_2) &= F_{0,95}(7, 9) = 3,2927. \end{aligned}$$

6. Como  $38,1287 > 3,2927$ , se toma la decisión de rechazar la hipótesis nula. En consecuencia, las varianzas de las dietas son distintas con un nivel de significancia del 0,10.  $\square$

#### 7.4.4. Prueba de hipótesis para diferencia de medias de dos poblaciones no independientes

Sean  $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}, \begin{pmatrix} X_3 \\ Y_3 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$  vectores aleatorios iid tal que

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} \sim N_2 \left[ \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix} \right],$$

luego

$$\begin{aligned} \mathbb{E}(X_1) &= \mu_X & \text{y} & \mathbb{E}(Y_1) = \mu_Y \\ \mathbb{V}(X_1) &= \sigma_X^2 & \text{y} & \mathbb{V}(Y_1) = \sigma_Y^2 & \text{y} & \text{Cov}(X_1, Y_1) = \rho\sigma_X\sigma_Y \\ \mathbb{E}(Y_1 - X_1) &= \mu_Y - \mu_X & \text{y} & \mathbb{V}(Y_1 - X_1) &= \sigma_Y^2 + \sigma_X^2 - 2\rho\sigma_X\sigma_Y. \end{aligned}$$

Si  $D_i = Y_i - X_i$ ,  $\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$  y  $S_{D,n}^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2$ , se sabe que

$$\frac{\sqrt{n} [\bar{D}_n - (\mu_Y - \mu_X)]}{S_{D,n}} \sim t_{n-1}.$$

Si las hipótesis planteadas son:

$$H_0 : \mu_Y - \mu_X \leq \theta_0 \quad \text{vs.} \quad H_a : \mu_Y - \mu_X > \theta_0.$$

El test rechaza  $H_0 : \mu_Y - \mu_X \leq \theta_0$  si

$$\bar{D} > t_{n-1, 1-\alpha} \frac{S_{D,n}}{\sqrt{n}} + \theta_0 \quad \Rightarrow \quad \frac{\bar{D} - \theta_0}{\frac{S_{D,n}}{\sqrt{n}}} > t_{n-1, 1-\alpha}$$

es un test **UMP** de nivel  $\alpha$ .

Si las hipótesis fueran  $H_0 : \mu_Y - \mu_X \geq \theta_0$  vs.  $H_a : \mu_Y - \mu_X < \theta_0$  el test rechaza  $H_0$  si

$$\frac{\bar{D} - \theta_0}{\frac{S_{D,n}}{\sqrt{n}}} < -t_{n-1, 1-\alpha}$$

es un test **UMP** de nivel  $\alpha$ .

Si las hipótesis fueran  $H_0 : \mu_Y - \mu_X = \theta_0$  vs.  $H_a : \mu_Y - \mu_X \neq \theta_0$  el test rechaza  $H_0$  si

$$\left| \frac{\bar{D} - \theta_0}{\frac{S_{D,n}}{\sqrt{n}}} \right| > t_{n-1, 1-\frac{\alpha}{2}}$$

es un test insesgado **UMP** de nivel  $\alpha$ .

**Ejemplo 7.19.** Según [89, p. 358], cinco muestras de una sustancia tipo ferrosa se utilizan para determinar si existe una diferencia entre un análisis químico de laboratorio y un análisis de fluorescencia de rayos X. Cada muestra se divide en dos submuestras y se aplican los dos tipos de análisis. Los datos codificados que muestran los análisis de contenido de hierro se presentan en la tabla 7.3.

Tabla 7.3. Datos codificados, ejemplo 7.19

Muestras \ Análisis	1	2	3	4	5
Rayos X	2,0	2,0	2,3	2,1	2,4
Químico	2,2	1,9	2,5	2,3	2,4
$D_i$	-0,2	0,1	-0,2	-0,2	0,0

Fuente: adaptada de [89, p. 359].

Si se supone que las poblaciones son normales, pruebe, al nivel de significancia de 0,05, si los dos métodos de análisis dan, en promedio, el mismo resultado.

**Solución.** De los datos del enunciado:

1. Se sabe que:

$$\bar{D} = -0,1 \quad n = 5 \quad S_{D,n} = 0,1414.$$

2. Además, se tiene que  $\theta_0 = 0$  y  $\alpha = 0,05$ .

3. Las hipótesis nula y alterna son:

$$H_0 : \mu_B - \mu_A = 0 \quad \text{vs.} \quad H_a : \mu_B - \mu_A \neq 0.$$

4. El estadístico para el contraste es:

$$t_c = \frac{\bar{D} - \theta_0}{\frac{S_{D,n}}{\sqrt{n}}} = \frac{-0,1 - 0}{\frac{0,1414}{\sqrt{5}}} = -1,5814.$$

5. El criterio de rechazo según la hipótesis alterna es: se rechaza  $H_0$  cuando

$$|t_c| \geq t_{n-1, 1-\frac{\alpha}{2}}.$$

6. Buscando para  $\alpha = 0,05$  en la tabla [C.2](#):

$$t_{n-1, 1-\frac{\alpha}{2}} = t_{4, 1-\frac{0,05}{2}} = 2,776.$$

7. Como  $|-1,5814| < 2,776$ , se toma la decisión de no rechazar la hipótesis nula. En consecuencia, no se puede rechazar la hipótesis de que los dos métodos de análisis dan, en promedio, el mismo resultado con un nivel de significancia del 0,05.

8. **Valor  $P$ .** Como la hipótesis alterna es  $H_a : \mu_B - \mu_A \neq 0$ , entonces

$$\begin{aligned} P\text{-valor} &= 2[1 - P(T_{n-1} < |t_c|)] = 2[1 - P(T_4 < |-1,5814|)] \\ &= 2(1 - 0,9055) = 0,189. \end{aligned}$$

Como el  $P$ -valor es mayor que el nivel de significancia no se rechaza la hipótesis de que los dos métodos de análisis dan, en promedio, el mismo resultado con un  $P$ -valor de 0,189.

El programa en  $R$  para el ejemplo [7.19](#) es:

```
remove(list=ls())
rayosX=c(2.0,2.0,2.3,2.1,2.4)
quimico=c(2.2,1.9,2.5,2.3,2.4)
## varianzas e intervalo de confianza
```

<sup>48</sup>El test prop.test realiza una prueba chi-cuadrado. El estadístico calculado que arroja para este ejemplo es  $\chi_c^2 = 8$ .

```
t.test(rayosX,quimico,mu=0,alternative="two.sided",
paired=TRUE,conf.level=0,95) # test
## t para comparar dos medias poblaciones no
## independientes
```

✓

## 7.4.5. Pruebas de hipótesis para diferencia entre dos proporciones poblacionales, muestras independientes

Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid Bernoulli de parámetro  $p_X$  y sean  $Y_1, Y_2, \dots, Y_m$  variables aleatorias iid Bernoulli de parámetro  $p_Y$ , donde las muestras aleatorias son independientes entre sí. Si se denota  $\bar{X}_n = \hat{\theta}_X$  y  $\bar{Y}_m = \hat{\theta}_Y$  y si los tamaños de muestra son suficientemente grandes para cada una de las muestras, se tiene que:

$$\begin{aligned} \sqrt{n} (\hat{\theta}_X - p_X) &\xrightarrow{D} N(0, p_X [1 - p_X]) \\ \frac{\hat{\theta}_X - p_X}{\sqrt{\frac{p_X(1-p_X)}{n}}} &\xrightarrow{D} N(0, 1) \\ \sqrt{m} (\hat{\theta}_Y - p_Y) &\xrightarrow{D} N(0, p_Y [1 - p_Y]) \\ & \text{y} \\ \frac{\hat{\theta}_Y - p_Y}{\sqrt{\frac{p_Y(1-p_Y)}{m}}} &\xrightarrow{D} N(0, 1), \end{aligned}$$

ver ejemplo 3.33. Luego para  $n$  y  $m$  suficientemente grandes:

$$\frac{\hat{\theta}_X - \hat{\theta}_Y - (p_X - p_Y)}{\sqrt{\frac{p_X(1-p_X)}{n} + \frac{p_Y(1-p_Y)}{m}}} \xrightarrow{D} N(0, 1).$$

Si las hipótesis planteadas son:

$$H_0 : p_X - p_Y = 0 \quad \text{vs.} \quad H_a : p_X - p_Y \neq 0,$$

usando pruebas de hipótesis sobre diferencia de medias de dos poblaciones independientes cuando las varianzas son desconocidas, pero se consideran iguales (suponiendo que la varianza depende de  $p$  y bajo la hipótesis nula  $p_X = p_Y$ ), el test rechaza  $H_0 : p_X - p_Y = 0$  si

$$\left| \frac{\hat{\theta}_X - \hat{\theta}_Y}{\sqrt{\bar{\theta} (1 - \bar{\theta}) \left( \frac{1}{n} + \frac{1}{m} \right)}} \right| > z_{1-\frac{\alpha}{2}}$$

es un test insesgado UMP de nivel  $\alpha$ , donde la proporción ponderada es:

$$\bar{\theta} = \frac{1}{m+n} \left[ \sum_{i=1}^n X_i + \sum_{j=1}^m Y_j \right].$$

**Ejemplo 7.20.** Una firma manufacturera de cigarrillos distribuye dos marcas. Si se encuentra que 56 de 200 fumadores prefieren la marca A y que 29 de 150 fumadores prefieren la marca B, ¿puede concluirse en el nivel de significancia de 0,05 que la preferencia por marca A difiere de la preferencia por la marca B?

*Solución.* De los datos del enunciado:

1. Se sabe que:

$$\begin{aligned} \hat{\theta}_A &= \frac{56}{200} = 0,28 & \text{y} & & n_A &= 200, \\ \hat{\theta}_B &= \frac{29}{150} = 0,1933 & \text{y} & & n_B &= 150, \\ \bar{\theta} &= \frac{\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j}{m+n} = \frac{56+29}{200+150} = \frac{85}{350} = 0,2429. \end{aligned}$$

2. Además, se tiene que  $p_0 = 0$  y  $\alpha = 0,05$ .

3. Las hipótesis nula y alterna son:

$$H_0 : p_A - p_B = 0 \quad \text{vs.} \quad H_a : p_A - p_B \neq 0.$$

4. El estadístico para el contraste es:

$$\begin{aligned} z_c &= \frac{\hat{\theta}_X - \hat{\theta}_Y}{\sqrt{\bar{\theta} (1 - \bar{\theta}) \left( \frac{1}{n} + \frac{1}{m} \right)}} = \frac{0,28 - 0,1933}{\sqrt{0,2429 (1 - 0,2429) \left( \frac{1}{200} + \frac{1}{150} \right)}} \\ &= 1,8718. \end{aligned}$$

5. El criterio de rechazo según la hipótesis alterna es: se rechaza  $H_0$  cuando

$$|z_c| \geq z_{1-\frac{\alpha}{2}}.$$

6. Buscando para  $\alpha = 0,05$  en la tabla C.1:

$$z_{1-\frac{\alpha}{2}} = z_{1-\frac{0,05}{2}} = 1,96.$$

7. Como  $1,8718 < 1,96$ , se toma la decisión de no rechazar la hipótesis nula, luego no hay suficiente evidencia para considerar que la preferencia por marca A difiere de la preferencia por la marca B.
8. **Valor  $P$ .** Como la hipótesis alterna es  $H_a : p_A - p_B \neq 0$ , entonces

$$\begin{aligned} P\text{-valor} &= 2 [1 - \Phi(z_c)] = 2(1 - \Phi(1,8718)) = 2(1 - 0,96934) \\ &= 0,06132. \end{aligned}$$

Como el  $P$ -valor es mayor que el nivel de significancia propuesto ( $\alpha = 0,05$ ), no se rechaza la hipótesis nula y se concluye que no hay suficiente evidencia para considerar que la preferencia por marca A difiere de la preferencia por la marca B, con un  $P$ -valor de 0,06132.

La instrucción en R para comparación de dos proporciones, muestras independientes, es:

```
prop.test(x=c(56,29),n=c(200,150),alternative=c("two.sided"),
          conf.level=0.95,correct=FALSE)
prop.test(x=c(56,29),n=c(200,150),alternative=c("two.sided"),
          conf.level=0.95,correct=TRUE)
```

Note que en este caso  $z_c = \underbrace{\sqrt{\chi_c^2}}_{49} = \sqrt{3,5013} = 1,871176$ . ☑

## 7.5. Bondad de ajuste

En este tipo de pruebas se comparan los datos de una muestra aleatoria, clasificados en un cierto tipo de categorías, con las frecuencias esperadas para las categorías bajo la hipótesis nula. Se rechazará la hipótesis nula si las frecuencias esperadas y las frecuencias observadas difieren suficientemente.

### 7.5.1. Test de Pearson para bondad de ajuste

Sea  $n$  el tamaño de muestra (cantidad fija) y considerando  $n$  ensayos independientes, donde cada ensayo tiene su resultado en uno de  $J$  posibles

<sup>49</sup>El test `prop.test` realiza un prueba chi-cuadrado. El estadístico calculado que arroja para este ejemplo es  $\chi_c^2 = 3,5013$ .

resultados, se dice que la distribución asociada respectiva es la distribución multinomial (ver definición 2.9). Cabe notar que la distribución binomial es un caso particular de la distribución multinomial donde  $J = 2$ , ver tabla 7.4.

Tabla 7.4. Tablas de frecuencias esperadas y observadas

Frecuencias esperadas		Frecuencias observadas	
Categoría	Proporción	Categoría	Proporción
1	$\pi_{10}$	1	$\frac{n_1}{n}$
2	$\pi_{20}$	2	$\frac{n_2}{n}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i$	$\pi_{i0}$	$i$	$\frac{n_i}{n}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$J$	$\pi_{J0}$	$J$	$\frac{n_J}{n}$
Total	$\sum_{i=1}^J \pi_{i0} = 1$	Total	$\sum_{i=1}^J \frac{n_i}{n} = 1$

Dado que las  $J$  probabilidades  $\theta_1, \theta_2, \dots, \theta_J$  suman 1, solo hay  $J - 1$  parámetros independientes en el modelo multinomial básico, por lo tanto,

$$\theta_J = 1 - \sum_{i=1}^{J-1} \theta_i \quad \text{y} \quad x_J = n - \sum_{i=1}^{J-1} x_i,$$

luego, la función de log-verosimilitud es:

$$\begin{aligned} \ell(\theta_1, \theta_2, \theta_3, \dots, \theta_{J-1}; x_1, x_2, x_3, \dots, x_{J-1}) &= \ln \left[ n! \left( \prod_{i=1}^J \frac{\theta_i^{x_i}}{x_i!} \right) \right] \\ &= \ln(n!) + \sum_{i=1}^J x_i \ln(\theta_i) - \sum_{i=1}^J \ln(x_i!) \\ &= \ln(n!) + \sum_{i=1}^{J-1} x_i \ln(\theta_i) + \left( n - \sum_{i=1}^{J-1} x_i \right) \ln \left( 1 - \sum_{i=1}^{J-1} \theta_i \right) - \sum_{i=1}^J \ln(x_i!), \end{aligned}$$

derivando con respecto a  $\theta_i, i = 1, 2, 3, \dots, J - 1$ ,

$$\frac{\partial \ell(\theta_1, \theta_2, \theta_3, \dots, \theta_{J-1}; x_1, x_2, x_3, \dots, x_{J-1})}{\partial \theta_i} = \frac{\partial \left[ \ln(n!) + \sum_{i=1}^{J-1} x_i \ln(\theta_i) + \left( n - \sum_{i=1}^{J-1} x_i \right) \ln \left( 1 - \sum_{i=1}^{J-1} \theta_i \right) - \sum_{i=1}^J \ln(x_i!) \right]}{\partial \theta_i}$$

$$= \frac{x_i}{\theta_i} + (-1) \underbrace{\frac{n - \sum_{i=1}^{J-1} x_i}{1 - \sum_{i=1}^{J-1} \theta_i}}_{\theta_J} = \frac{x_i}{\theta_i} - \frac{x_J}{\theta_J},$$

igualando a cero [76, p. 304],

$$\frac{x_i}{\hat{\theta}_i} - \frac{\underbrace{X_J}_{n - \sum_{i=1}^{J-1} x_i}}{\underbrace{1 - \theta_1 - \theta_2 - \dots - \hat{\theta}_i - \theta_{i+1} - \dots - \theta_{J-1}}_{\hat{\theta}_J}} = 0, \quad i = 1, 2, \dots, J - 1$$

$$\frac{x_i}{\hat{\theta}_i} = \frac{X_J}{\hat{\theta}_J} \quad i = 1, 2, 3, \dots, J - 1$$

$$\frac{x_i \hat{\theta}_J}{X_J} = \hat{\theta}_i \quad i = 1, 2, 3, \dots, J - 1,$$

como  $\sum_{i=1}^J \theta_i = 1$ , entonces:

$$\sum_{i=1}^J \hat{\theta}_i = 1 \quad \Leftrightarrow \quad \sum_{i=1}^{J-1} \frac{x_i \hat{\theta}_J}{X_J} + \hat{\theta}_J = 1 \quad \Leftrightarrow$$

$$\hat{\theta}_J \left( \underbrace{\sum_{i=1}^{J-1} x_i + X_J}_n \right) = X_J \quad \Leftrightarrow \quad \hat{\theta}_J = \frac{X_J}{n},$$

luego,

$$\frac{x_i \frac{X_J}{n}}{X_J} = \hat{\theta}_i \quad i = 1, 2, 3, \dots, J - 1$$

$$\frac{x_i}{n} = \hat{\theta}_i \quad i = 1, 2, 3, \dots, J - 1,$$

así el EMV para  $\theta_i$  es  $\hat{\theta}_i = \frac{x_i}{n}$ ,  $i = 1, 2, \dots, J$ .

Ahora, considérese las hipótesis

$$H_0 : \theta_i = \theta_{i0} \quad \text{vs.} \quad H_a : \theta_i \neq \theta_{i0} \quad i = 1, 2, 3, \dots, J, \quad (7.20)$$

entonces la **RVM** dada en (7.2) queda:

$$\begin{aligned} \lambda(\mathbf{x}) &= \frac{\sup_{\Theta_0} L(\theta; \mathbf{x})}{\sup_{\Theta} L(\theta; \mathbf{x})} = \frac{\sup_{\Theta_0} L(\theta_1, \theta_2, \theta_3, \dots, \theta_{J-1}; x_1, x_2, x_3, \dots, x_{J-1})}{\sup_{\Theta} L(\theta_1, \theta_2, \theta_3, \dots, \theta_{J-1}; x_1, x_2, x_3, \dots, x_{J-1})} \\ &= \frac{\left( \frac{n!}{\prod_{i=1}^J x_i!} \right) \left( \prod_{i=1}^J \theta_{i0}^{x_i} \right)}{\left( \frac{n!}{\prod_{i=1}^J x_i!} \right) \left( \prod_{i=1}^J \binom{x_i}{n} \right)} = \prod_{i=1}^J \frac{\theta_{i0}^{x_i}}{\left( \frac{x_i}{n} \right)^{x_i}} = \prod_{i=1}^J \frac{n^{x_i} \theta_{i0}^{x_i}}{x_i^{x_i}} = \prod_{i=1}^J \left( \frac{n\theta_{i0}}{x_i} \right)^{x_i}. \end{aligned}$$

Luego, para  $0 < c < 1$ , la región de rechazo es de la forma:

$$\{ \mathbf{x} : \lambda(\mathbf{x}) \leq c \} = \left\{ \mathbf{x} : \prod_{i=1}^J \left( \frac{n\theta_{i0}}{x_i} \right)^{x_i} \leq c \right\},$$

que corresponde a un test **LRT** para probar la hipótesis (7.20).

Usando el teorema 7.4 se tiene que:

$$\begin{aligned} -2 \ln [\lambda(\mathbf{x})] &= -2 \ln \left[ \prod_{i=1}^J \left( \frac{n\theta_{i0}}{x_i} \right)^{x_i} \right] \\ &= -2 \sum_{i=1}^J x_i \ln \left( \frac{n\theta_{i0}}{x_i} \right) \xrightarrow{D} \chi_{J-1}^2, \end{aligned} \quad (7.21)$$

entonces,

$$\begin{aligned} \alpha &= \lim_{n \rightarrow \infty} P_{\theta} \{ \mathbf{x} : \lambda \{ \mathbf{x} \} \leq c \} \\ &= \lim_{n \rightarrow \infty} P_{\theta} \left\{ \mathbf{x} : \prod_{i=1}^J \left( \frac{n\theta_{i0}}{x_i} \right)^{x_i} \leq c \right\} \\ &= \lim_{n \rightarrow \infty} P_{\theta} \left\{ \mathbf{x} : -2 \ln \left[ \prod_{i=1}^J \left( \frac{n\theta_{i0}}{x_i} \right)^{x_i} \right] \geq -2 \ln(c) \right\} \\ &= \lim_{n \rightarrow \infty} P_{\theta} \left\{ \mathbf{x} : -2 \sum_{i=1}^J x_i \ln \left[ \frac{n\theta_{i0}}{x_i} \right] \geq -2 \ln(c) \right\}. \end{aligned}$$

Por la regla del complemento:

$$\lim_{n \rightarrow \infty} P_{\theta} \left\{ \mathbf{x} : -2 \sum_{i=1}^J x_i \ln \left[ \frac{n\theta_{i0}}{x_i} \right] < -2 \ln(c) \right\} = 1 - \alpha.$$

Despejando, se obtiene el cuantil

$$\chi_{J-1,1-\alpha}^2 = -2 \ln(c) \quad \Rightarrow \quad \exp \left\{ \frac{-\chi_{J-1,1-\alpha}^2}{2} \right\} = c,$$

por tanto, la región de rechazo es de la forma

$$\begin{aligned} \{x : \lambda(x) \leq c\} &= \left\{ x : \prod_{i=1}^J \left( \frac{n\theta_{i0}}{x_i} \right)^{x_i} \leq \exp \left\{ \frac{-\chi_{J-1,1-\alpha}^2}{2} \right\} \right\} \\ &= \left\{ x : \ln \left[ \prod_{i=1}^J \left( \frac{n\theta_{i0}}{x_i} \right)^{x_i} \right] \leq \ln \left[ \exp \left\{ \frac{-\chi_{J-1,1-\alpha}^2}{2} \right\} \right] \right\} \\ &= \left\{ x : \sum_{i=1}^J x_i \ln \left( \frac{n\theta_{i0}}{x_i} \right) \leq \frac{-\chi_{J-1,1-\alpha}^2}{2} \right\} \\ &= \left\{ x : -2 \sum_{i=1}^J x_i \ln \left( \frac{n\theta_{i0}}{x_i} \right) \geq \chi_{J-1,1-\alpha}^2 \right\}, \end{aligned}$$

que corresponde a un test **LRT** para probar la hipótesis  $H_0 : \theta_i = \theta_{i0}$ ,  $i = 1, 2, 3, \dots, J$ .

**Ejemplo 7.21.** Según [19, p. 374], a partir de los registros de un almacén, el 50% de los vestidos adquiridos en una temporada se venderá a precio de menudeo, el 25% a un 20% menos del precio de menudeo, el 15% se venderá tras una reducción de su precio al 40% y los restantes con una disminución en su precio del 60%. Para la temporada se compraron 300 vestidos y su venta se presenta en la tabla 7.5.

Tabla 7.5. Distribución del número de ventas de vestidos por porcentaje de disminución en su precio de venta

Descuento	0 %	20 %	40 %	60 %	Total
Ventas					
Número de vestidos	140	90	30	40	300

Fuente: adaptada de [19, p. 374].

¿Existe alguna razón para creer que el porcentaje de ventas según porcentaje de descuento fue diferente en esta temporada con respecto a las temporadas anteriores? Usar un nivel de significancia de 0,05 y calcular el P-valor.

**Solución.** De los datos del enunciado:

1. Se sabe que  $n = 300$ ,  $J = 4$  y  $\alpha = 0,05$ .
2. Sean  $p_1$  la proporción de ventas de vestidos al precio de menudeo,  $p_2$  la proporción de ventas de vestidos a un 20% menos del precio de

menudeo,  $p_3$  la proporción de ventas de vestidos a un 40 % menos del precio de menudeo y  $p_4$  la proporción de ventas de vestidos a un 60 % menos del precio de menudeo.

3. Las hipótesis nula y alterna son:

$$H_0 : \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} 0,50 \\ 0,25 \\ 0,15 \\ 0,10 \end{pmatrix} \quad \text{vs.} \quad H_1 : \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} \neq \begin{pmatrix} 0,50 \\ 0,25 \\ 0,15 \\ 0,10 \end{pmatrix}.$$

4. El estadístico para el contraste es dado en (7.21):

$$\begin{aligned} \chi_c^2 &= -2 \sum_{i=1}^J x_i \ln \left( \frac{np_{i0}}{x_i} \right) \\ &= -2 \left[ 140 \ln \left( \frac{300 \times 0,50}{140} \right) + 90 \ln \left( \frac{300 \times 0,25}{90} \right) + \right. \\ &\quad \left. 30 \ln \left( \frac{300 \times 0,15}{30} \right) + 40 \ln \left( \frac{300 \times 0,10}{40} \right) \right] = 12,1865. \end{aligned}$$

5. El criterio de rechazo es: se rechaza  $H_0$  cuando

$$\chi_c^2 \geq \chi_{J-1,1-\alpha}^2.$$

6. Buscando para  $\alpha = 0,05$  en la tabla C.3:

$$\chi_{J-1,1-\alpha}^2 = \chi_{3;0,95}^2 = 7,8147.$$

7. Como  $12,1865 > 7,8147$  se concluye que el porcentaje de ventas según porcentaje de descuento fue diferente en esta temporada con respecto a las temporadas anteriores con un nivel de significancia de 0,05.

8. **Valor  $P$ .** En este caso se evalúa como área a la derecha del estadístico calculado.

$$\begin{aligned} P - \text{valor} &= P(\chi_{J-1}^2 > \chi_c^2) = 1 - P(\chi_3^2 \leq \chi_c^2) \\ &= 1 - P(\chi_3^2 \leq 12,1865) = 1 - 0,99323 = 0,00677 \end{aligned}$$

Como el  $P$ -valor es menor que el nivel de significancia propuesto se rechaza la hipótesis nula y se concluye que el porcentaje de ventas según porcentaje de descuento fue diferente en esta temporada con respecto a las temporadas anteriores, con un  $P$ -valor de 0,00677.  $\checkmark$

**Nota 7.8.** Otro test muy utilizado en la práctica para probar las hipótesis presentadas en (7.20) fue propuesto por Pearson y es dado por:

$$\chi_c^2 = \sum_{i=1}^J \frac{(x_i - np_{i0})^2}{np_{i0}}, \quad (7.22)$$

donde se rechaza  $H_0 : p_i = p_{i0}, i = 1, 2, 3, \dots, J$ , si y solo si

$$\chi_c^2 > \chi_{J-1, 1-\alpha}^2.$$

**Ejemplo 7.22.** Considerando la información del ejemplo 7.21, resolver la pregunta usando el test de Pearson.

*Solución.* De los datos del enunciado:

1. Se sabe que  $n = 300$ ,  $J = 4$  y  $\alpha = 0,05$ .
2. Sean  $p_1$  la proporción de ventas de vestidos al precio de menudeo,  $p_2$  la proporción de ventas de vestidos a un 20 % menos del precio de menudeo,  $p_3$  la proporción de ventas de vestidos a un 40 % menos del precio de menudeo y  $p_4$  la proporción de ventas de vestidos a un 60 % menos del precio de menudeo.
3. Las hipótesis nula y alterna son:

$$H_0 : \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} 0,50 \\ 0,25 \\ 0,15 \\ 0,10 \end{pmatrix} \quad \text{vs.} \quad H_1 : \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} \neq \begin{pmatrix} 0,50 \\ 0,25 \\ 0,15 \\ 0,10 \end{pmatrix}.$$

4. El estadístico para el contraste es dado en (7.22):

$$\begin{aligned} \chi_c^2 &= \sum_{i=1}^J \frac{(x_i - np_{i0})^2}{np_{i0}} \\ &= \frac{(140 - 300 \times 0,50)^2}{300 \times 0,50} + \frac{(90 - 300 \times 0,25)^2}{300 \times 0,25} \\ &\quad + \frac{(30 - 300 \times 0,15)^2}{300 \times 0,15} + \frac{(40 - 300 \times 0,10)^2}{300 \times 0,10} = 12. \end{aligned}$$

5. El criterio de rechazo es: se rechaza  $H_0$  cuando

$$\chi_c^2 \geq \chi_{J-1, 1-\alpha}^2.$$

6. Buscando para  $\alpha = 0,05$  en la tabla C.3:

$$\chi_{J-1, 1-\alpha}^2 = \chi_{3; 0,95}^2 = 7,8147.$$

7. Como  $12 > 7,8147$  se concluye que el porcentaje de ventas según porcentaje de descuento fue diferente en esta temporada con respecto a las temporadas anteriores con nivel de significancia de 0,05.
8. **Valor  $P$ .** En este caso se evalúa como área a la derecha del estadístico calculado.

$$\begin{aligned} P\text{-valor} &= P(\chi_{J-1}^2 > \chi_c^2) = 1 - P(\chi_3^2 \leq \chi_c^2) = 1 - P(\chi_3^2 \leq 12) \\ &= 1 - 0,9926 = 0,0074. \end{aligned}$$

Como el  $P$ -valor es menor que el nivel de significancia propuesto se rechaza la hipótesis nula y se concluye que el porcentaje de ventas según porcentaje de descuento fue diferente en esta temporada con respecto a las temporadas anteriores, con un  $P$ -valor de 0,0074.

Las instrucciones en  $R$  para este test son:

```
remove(list=ls())
ventas=matrix(c(140, 90, 30, 40))
dimnames(ventas) = list(c("Precio de venta",
    "20% de descuento", "40% de descuento",
    "60% de descuento"))
ventas
chisq.test(ventas, p=c(0.50, 0.25, 0.15, 0.10))
## test chi-cuadrado 
```

**DEFINICIÓN 7.14.** De acuerdo con [22, p. 79], dadas  $X_1, X_2, X_3, \dots, X_n$  variables aleatorias iid. La función de distribución empírica  $S_n(x)$  es una función de  $x$  que corresponde a la fracción de las  $X_i$  que son menores o iguales a  $x$  para cada  $x$ ,  $-\infty < x < \infty$ , es decir,

$$S_n(x) = \frac{\text{número de valores en la muestra } \leq x}{\text{tamaño de la muestra } (n)}.$$

Note que  $S_n(x)$  es una función aleatoria, ya que depende de las realizaciones de las variables aleatorias  $X_i$ .

**Ejemplo 7.23.** Considerando la información del ejemplo 6.17 determine la función de distribución empírica para los datos asociados para la dieta A.

Tabla 7.6. Distribución empírica, dieta A, ejemplo 7.23

$x$	245	269	275	277	278	280	281	284
$S_{10}(x)$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{4}{10}$	$\frac{6}{10}$	$\frac{8}{10}$	$\frac{9}{10}$	1

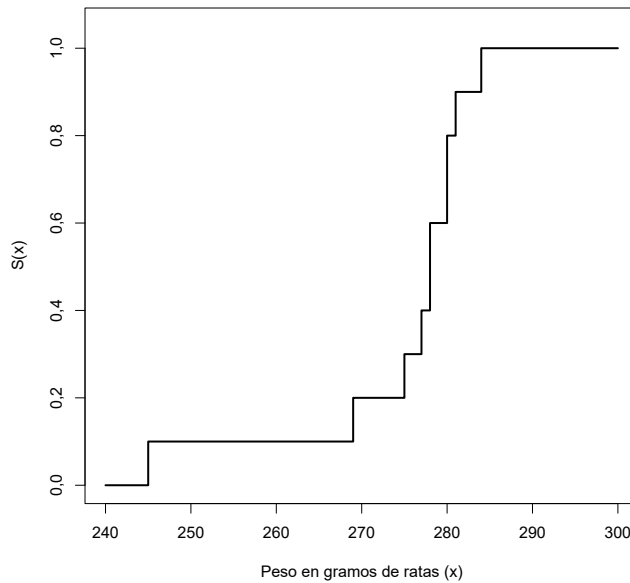


Figura 7.10. Distribución empírica, dieta A, ejemplo 7.23

*Solución.* La  $S_n(x)$  se presenta en la tabla 7.6 y su gráfica en la figura 7.10.

Las instrucciones en *R* para el gráfico son:

```
remove(list=ls())
x=c(240,245, 269, 275, 277, 278, 280, 281, 284,300)
empirica=c(0,1/10,2/10,3/10,4/10,6/10,8/10,9/10,1,1)
plot(empirica~x,ylim=c(0,1.05),type="s",
      xlab="Peso en gramos de ratas (x)",ylab="S(x)",pch=19,
      lwd=2,cex=0.4)
```

☑

## 7.5.2. Test de Kolmogorov para bondad de ajuste

Es un test para bondad de ajuste para datos ordinales [22, cap. 6]. La hipótesis nula plantea que los datos provienen de una población con fda  $F^*(x)$ . Luego una muestra aleatoria  $X_1, X_2, X_3, \dots, X_n$  es seleccionada y se compara  $S_n(x)$  con  $F^*(x)$ . El estadístico,  $T$ , consiste en la distancia más grande entre  $S_n(x)$

y  $F^*(x)$ , es decir,

$$T = \sup_x |F^*(x) - S_n(x)|. \quad (7.23)$$

Las hipótesis a considerar son:

$$H_0 : F(x) = F^*(x) \quad \text{vs.} \quad H_a : F(x) \neq F^*(x).$$

Se rechaza  $H_0$  al nivel de significancia  $\alpha$  si

$$T > w_{n,1-\alpha},$$

con  $w_{n,1-\alpha}$  dado por los cuantiles del test estadístico de Kolmogorov (ver [22, pp. 347 y 462]).

**Ejemplo 7.24.** Continuando con el ejemplo 7.23, encuentre el estadístico  $T$  dado en (7.23).

*Solución.* En la tabla 7.7 se presentan las diferencias entre  $S_n(x)$  y  $F^*(x)$ , donde  $F^*(x)$  corresponde a la fda normal con media 274,7 y varianza 124,9.

Tabla 7.7. fda empírica,  $S_{10}(x)$ , y fda teórica,  $\Phi(x)$ , dieta A, ejemplo 7.23

$x$	245	269	275	277
$z = \frac{x-274,7}{\sqrt{124,9}}$	-2,66	-0,51	0,03	0,21
$S_{10}(x)$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{4}{10}$
$\Phi(z)$	0,004	0,305	0,511	0,582
$ \Phi(x) - S_{10}(x) $	-0,096	0,105	<b> 0,211 </b>	0,182
$x$	278	280	281	284
$z = \frac{x-274,7}{\sqrt{124,9}}$	0,30	0,47	0,56	0,83
$S_{10}(x)$	$\frac{6}{10}$	$\frac{8}{10}$	$\frac{9}{10}$	1
$\Phi(z)$	0,616	0,682	0,714	0,797
$ \Phi(x) - S_{10}(x) $	0,016	-0,118	-0,186	-0,203

En la tabla 7.7 se observa que  $T = 0,211$ . Si  $\alpha = 0,05$ , de la tabla de Kolmogorov se tiene  $w_{n,1-\alpha} = w_{10;0,95} = 0,409$ , por tanto  $T < w$ , y se concluye que no hay suficientes evidencias para considerar que la distribución muestral de los datos asociados a la dieta A no provengan de una distribución normal, ver figura 7.11. Cabe notar que el test supone una distribución continua y por lo mismo no deberían haber empates en los datos; de igual forma, con pequeñas muestras el cálculo de  $T$  no es bien aproximado. Este ejemplo muestra el problema.

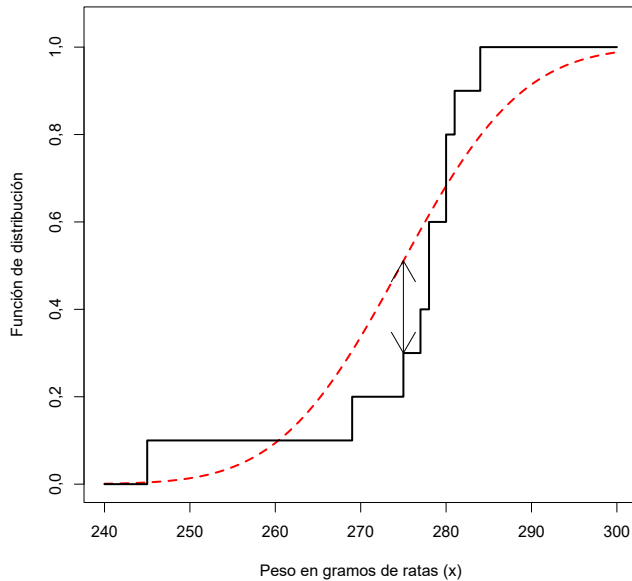


Figura 7.11. Distribución empírica (línea continua) y teórica (línea intermitente), dieta A, ejemplo 7.23

Las instrucciones en *R* para el test son:

```
dietaA=c(278, 245, 269, 275, 280, 281, 284, 278, 280, 277)
ks.test(dietaA,"pnorm",alternative="two.sided",mean=274.7,
sd=sqrt(124,9))
library(kolmim)
1-pkolm(0.211,10) #T=0.211 y n=10, da el $$$-valor
```

y las instrucciones para el gráfico son:

```
curve(pnorm(x,mean=mean(dietaA),sd=sqrt(var(dietaA))),from=240,
to=300,lwd=2,lty=2,col="red",ylim=c(0,1.05),
xlab="Peso en gramos de ratas (x)",
ylab="Función de distribución")
lines(empirica~x,type="s",ylab="S(x)", pch=19,lwd=2,cex=0.4)
arrows(275, 0.3, 275, 0.511,cod=3)
```

Para los datos repetidos  $x_i$ , se realiza un pequeño cambio  $x_i \pm 0,01$  en estos datos para evitar los empates, en este caso se tiene que:

```
dietaAmod=c(277.99,245,269,275,279.99,281,284,278.01,280.01,277)
ks.test(dietaAmod,"pnorm",alternative="two.sided",mean=274.7,
sd=sqrt(124.9))
1-pkolm(0.31071,10) #T=0.31071 y n=10, da el $$$-valor
```



**Nota 7.9.** (Gráfico Q-Q para los datos  $x_1, x_2, \dots, x_n$ ). El gráfico cuantil-cuantil se construye a través de los siguientes pasos:

1. Se ordenan los datos de menor a mayor, generando  $x_{(1)}, x_{(2)}, \dots, x_{(k)}, \dots, x_{(n-1)}, x_{(n)}$ , donde  $x_{(1)}$  es el menor dato y  $x_{(n)}$  es el mayor dato.
2. Para cada  $k, k = 1, 2, \dots, n$ , se calcula  $f_k = \frac{k - 0,5}{n}$  y  $z_k = \Phi^{-1}(f_k)$ , que es el percentil de la distribución normal estándar asociado a  $f_k$ .
3. Se grafica los  $z_k$  vs.  $x_{(k)}$ .

**Ejemplo 7.25.** Retomando los datos de la dieta A del ejemplo 6.17 en la tabla 7.8 se presentan los valores ordenados  $x_{(k)}$ ,  $f_k$  y los percentiles teóricos para una distribución normal estándar. En la figura 7.12 se presenta el gráfico Q-Q.

Tabla 7.8. Datos dieta A ordenados, ejemplo 6.17

$k$	1	2	3	4	5	6	7	8	9	10
$x_{(k)}$	245	269	275	277	278	278	280	280	281	284
$f_k$	0,05	0,15	0,25	0,35	0,45	0,55	0,65	0,75	0,85	0,95
$\Phi^{-1}(f_k)$	-1,645	-1,036	-0,674	-0,385	-0,126	0,126	0,385	0,674	1,036	1,645

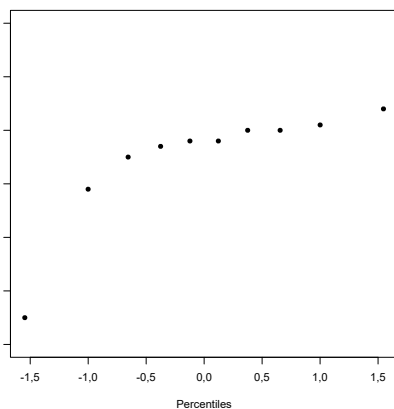


Figura 7.12. Gráfico Q-Q para los datos de la dieta A considerando una **fdp** normal

La instrucción en R para el gráfico Q-Q es:

```
qqnorm(dietaA,xlab="Percentiles",ylab="Peso en gramos",
        ylim=c(240,300),pch=16,main="")
```

**Nota 7.10.** Para una simulación de muestras generadas a partir de una distribución normal y sus gráficos Q-Q, se puede usar el paquete *animation*.

```
library(animation)
oopt = ani.options(interval = 0.8, nmax = 30)
sim.qnorm(n = 20, last.plot = expression(abline(0, 1)))
```

**Nota 7.11.** Para un ajuste de la distribución por **simulación**, se usa el **principio del envelope**, propuesto inicialmente por [9] y se refiere a una banda de confianza empírica. El procedimiento para construir estas bandas se describe a continuación:

1. Definir  $m$ , el número de muestras a generar (simulaciones) para la construcción de las bandas del envelope, por ejemplo  $m = 100$ .
2. Generar  $m$  vectores aleatorios  $n$ -dimensionales correspondientes a la distribución deseada,  $\mathbf{r}_j = (r_{1j}, r_{2j}, \dots, r_{ij}, \dots, r_{nj})^T$  con  $i = 1, 2, \dots, n$  y  $j = 1, 2, \dots, m$ , por ejemplo  $N_n(274, 71\mathbf{1}_n, \sqrt{124, 9}\mathbf{I}_n)$ , que corresponden a la distribución teórica.
3. Para cada muestra generada, ordenar de menor a mayor, es decir,  $r_{(1)j} \leq r_{(2)j} \leq \dots \leq r_{(n)j}$ , ver tabla 7.9.

Tabla 7.9. Organización de los residuales estandarizados de las  $m$  simulaciones

Simulación							
1	$r_{(1)1}$	$\leq$	$r_{(2)1}$	$\leq$	$\dots$	$\leq$	$r_{(n)1}$
2	$r_{(1)2}$	$\leq$	$r_{(2)2}$	$\leq$	$\dots$	$\leq$	$r_{(n)2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$m$	$r_{(1)m}$	$\leq$	$r_{(2)m}$	$\leq$	$\dots$	$\leq$	$r_{(n)m}$
Percentil 5:	$P_{5(1)}$		$P_{5(2)}$		$\dots$		$P_{5(n)}$
Percentil 50:	$P_{50(1)}$		$P_{50(2)}$		$\dots$		$P_{50(n)}$
Percentil 95:	$P_{95(1)}$		$P_{95(2)}$		$\dots$		$P_{95(n)}$

4. En cada columna de la tabla 7.9, calcular los percentiles 5, 50 y 95. Así se tendrán  $n$  percentiles 5,  $n$  percentiles 50 y  $n$  percentiles 95, y construir sobre una misma figura un gráfico Q-Q para los  $n$  percentiles 5, un gráfico Q-Q para los  $n$  percentiles 50 y un gráfico Q-Q para los  $n$  percentiles 95.
5. Sobre la figura anterior, realizar un gráfico Q-Q para los datos originales.

**Ejemplo 7.26.** Retomando los datos de la dieta A del ejemplo 6.17, construya el gráfico de envelope.

*Solución.* En la figura 7.13 se presenta el gráfico de envelope

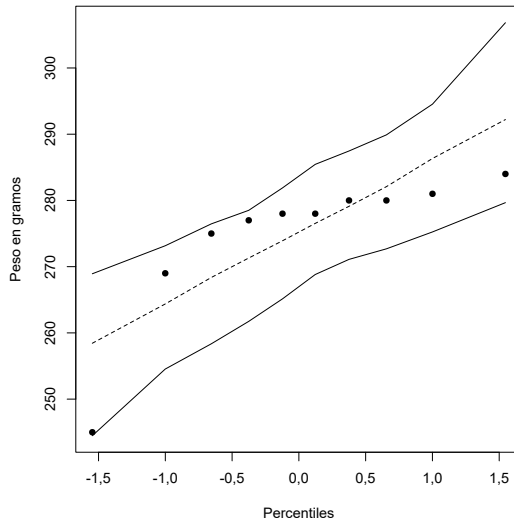


Figura 7.13. Envelope para los datos de la dieta A considerando una [fdp](#) normal con media 274,7 y varianza 124,9

Las instrucciones en *R* para la construcción del envelope son:

```
remove(list=ls())
dietaA=c(278, 245, 269, 275, 280, 281, 284, 278, 280, 277)
m=100
n=length(dietaA)
generado = matrix(0,n,100)
e1 = numeric(n)
e2 = numeric(n)
for(i in 1:m){
  generado[,i] = rnorm(n,mean=274.7,sd=sqrt(124.9))
  generado[,i] = sort(generado[,i])
}
for(i in 1:n){
  gen.ord = sort(generado[i,])
  e1[i] = (gen.ord[2]+gen.ord[3])/2
  e2[i] = (gen.ord[97]+gen.ord[98])/2
}
med = apply(generado,1,mean)
franja = range(dietaA,e1,e2)
#postscript(file="envelopedietaA.eps",width=7, height=7)
par(pty="s")
qqnorm(dietaA,xlab="Percentiles",
ylab="Peso en gramos",ylim=franja,pch=16,main="")
```

```

par(new=T)
qqnorm(e1, axes=F, xlab="", ylab="", type="l", ylim=franja, lty=1,
      main="")
par(new=T)
qqnorm(e2, axes=F, xlab="", ylab="", type="l", ylim=franja, lty=1,
      main="")
par(new=T)
qqnorm(med, axes=F, xlab="", ylab="", type="l", ylim=franja, lty=2,
      main="")
#dev.off()

```



## 7.6. Ejercicios

- Sean  $X_1, X_2, X_3, \dots, X_{16}$  variables aleatorias **iid**. Construya un test más potente para probar la hipótesis nula sobre que la distribución de las  $X$  es normal con  $\mu = 0$  y  $\sigma^2 = 9$  vs. la hipótesis alterna sobre que la distribución de las  $X$  es normal con  $\mu = 1$  y  $\sigma^2 = 9$  al nivel de significancia  $\alpha = 0,05$  [76, p. 336]. Encuentre la potencia del test.
- Sean  $X_1, X_2, X_3, \dots, X_n$  variables aleatorias **iid** con **fdp** [76, p. 349]:

$$f(x; \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}} I_{(0, \infty)}(x), \quad \theta \in \Theta = (0, \infty).$$

- Derive el test **UMP** para probar la hipótesis nula  $\theta \geq \theta_0$  contra la hipótesis alterna  $\theta < \theta_0$  al nivel de significancia  $\alpha$ .
  - Determine el tamaño de muestra mínimo  $n$  requerido para obtener una potencia de al menos 0,95 contra la alternativa  $\theta_1 = 500$  cuando  $\theta_0 = 1000$  y  $\alpha = 0,05$ .
- Para cada situación calcule el  $P$ -valor de acuerdo a los datos observados [20, p. 411]:
    - Para probar  $H_0 : \theta \leq \frac{1}{2}$  vs.  $H_1 : \theta > \frac{1}{2}$ , se observan 7 éxitos en un total de 9 ensayos Bernoulli.
    - Para probar  $H_0 : \lambda \leq 1$  vs.  $H_1 : \lambda > 1$ , se observa  $X = 3$ , donde  $X \sim Poisson(\lambda)$ .
    - Para probar  $H_0 : \lambda \leq 1$  vs.  $H_1 : \lambda > 1$ , se observa  $X_1 = 3$ ,  $X_2 = 5$  y  $X_3 = 1$ , donde  $X_i \sim Poisson(\lambda)$  y  $X_1, X_2$  y  $X_3$  son independientes.
  - En una universidad se estima que el 20% de los estudiantes van en bicicleta. Muestre si la afirmación está sustentada por una muestra de

120 estudiantes, en la cual 35 usan la bicicleta para ir a la universidad. Usar un nivel de significancia del 5 %. Adicionalmente dé el  $P$ -valor.

5. En el hospital A, el número de nacimientos observados para cada mes de un año concreto se presenta en la tabla 7.10.

Tabla 7.10. N.º de nacimientos por mes en un año concreto en el hospital A

Enero	Febrero	Marzo	Abril	Mayo	Junio
101	91	101	97	99	96
Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
101	103	106	103	100	103

¿Existe alguna razón para creer que, en ese año concreto, el número de nacimientos se encuentra distribuido en forma uniforme durante los doce meses? Usar un nivel de significancia de 0,05 y calcular el  $P$ -valor. Usar el LRT generalizada y el test de Pearson.

6. En la tabla 7.11 se presentan los valores de extracto de malta provenientes de maltas de cebada Kindred cultivada en 20 localidades en los viveros de cebada de un valle en un año particular, expresados en porcentaje de base seca.

Tabla 7.11. Porcentaje de base seca, ejercicio 6

77,7	76,0	76,9	74,6	74,7	76,5	74,2	77,5	74,2	76,1
75,4	76,0	73,9	77,4	76,6	77,3	74,3	76,5	73,8	76,4

Utilice el test de Kolmogorov-Smirnov para verificar si estos datos (en porcentaje) se pueden suponer que son realizaciones de:

- a) Una variable aleatoria  $X$  con fdp normal. Presente la distribución empírica y la distribución teórica en una tabla. Use  $\alpha = 0,10$ .
- b) Una variable aleatoria  $X$  con fdp beta de parámetros  $a$  y  $b$ ,

$$f(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}$$

$$\text{con } \mathbb{E}(X) = \frac{a}{a+b} \text{ y } \mathbb{V}(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$





Las estadísticas de prueba que se van a presentar en este capítulo son las estadísticas de Wald, Wilks y Rao. La estadística de Wald está relacionada con la convergencia asintótica a la distribución normal de un estimador  $\hat{\theta}$ , la estadística de Wilks utiliza la razón de verosimilitudes, y la estadística de Rao está relacionada con la función score y su distribución asintótica. Por lo anterior, inicialmente en este capítulo se define la función score y se presentan algunas de sus propiedades asintóticas. También se presentan dos teoremas que dan las condiciones necesarias y suficientes para que los EMV sean consistentes y tengan distribución asintótica normal.

## 8.1. Función de estimación. Caso uniparamétrico

Para encontrar el EMV de  $\theta$ ,  $\hat{\theta}_{MV}$ , en la mayoría de los casos el problema se reduce a resolver las ecuaciones de estimación dadas por (ver [82, p. 202]),

$$\left. \frac{\partial \ell_n(\theta; \mathbf{y})}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0,$$

en que  $\ell_n(\theta; \mathbf{y})$  corresponde a la función de log-verosimilitud.

Sea

$$U_n(\theta) := \frac{\partial \ell_n(\theta; \mathbf{Y})}{\partial \theta},$$

a  $U_n(\theta) = U_n(\theta, Y_1, Y_2, \dots, Y_n)$  se le llama *función score o de puntajes* cuando es considerada como una función de  $Y_1, Y_2, \dots, Y_n$ . Por otro lado,  $U_n(\theta)$  se conoce como *función de estimación* cuando es considerada como función de  $\theta$ .

**Nota 8.1.** En general, bajo el punto de vista de ecuaciones de estimación, se consideran los estimadores que pueden ser expresados como solución de una ecuación

$$\psi(\hat{\theta}; \mathbf{y}) = 0,$$

en que  $\psi$  es una función de los datos,  $\mathbf{y}$ , y el parámetro,  $\theta$ , de un modelo estadístico. Esta  $\psi$  también se le llama *función de inferencia o función de estimación* [47, p. 135].

**Ejemplo 8.1.** Sean  $Y_1, Y_2, \dots, Y_n$  variables aleatorias iid, tal que  $\mathbb{E}(Y_1) = 0$ ,  $\mathbb{V}(Y_1) = \theta$  y  $0 < \mathbb{E}(Y_1^4) < \infty$ . Presente una función de estimación para el parámetro  $\theta$  [7, p. 15].

*Solución.* Una función de estimación para  $\theta$  es:

$$\psi(\mathbf{Y}, \theta) = \sum_{i=1}^n (Y_i^2 - \theta). \quad \square$$

### 8.1.1. Propiedades de una función de estimación

Suponiendo que

$$\begin{aligned} |U_n(\theta)| &= \left| \frac{\partial \ln [f(\mathbf{y}; \theta)]}{\partial \theta} \right| = \left| \frac{\partial \ell(\theta; \mathbf{y})}{\partial \theta} \right| < \infty \quad \text{y} \\ \left| \frac{\partial U_n(\theta)}{\partial \theta} \right| &= \left| \frac{\partial^2 \ln [f(\mathbf{y}; \theta)]}{\partial \theta^2} \right| = \left| \frac{\partial^2 \ell(\theta; \mathbf{y})}{\partial \theta^2} \right| < \infty \end{aligned}$$

y que se puede derivar bajo el signo de integración, la esperanza y varianza de  $U_n(\theta)$  son dadas por:

$$\mathbb{E}[U_n(\theta)] = \mathbb{E} \left( \frac{\partial \ell_n(\theta; \mathbf{y})}{\partial \theta} \right) \underbrace{=} 0 \quad (8.1)$$

Ver (5.13)

$$\mathbb{V}[U_n(\theta)] = \mathbb{E}[U_n^2(\theta)] - \{\mathbb{E}[U_n(\theta)]\}^2 = \mathbb{E}[U_n^2(\theta)], \quad (8.2)$$

en que el último término se puede obtener en función de

$$\begin{aligned} \mathbb{E} \left( \frac{\partial U_n(\theta)}{\partial \theta} \right) &= \mathbb{E} \left( \frac{\partial^2 \ell_n(\theta; \mathbf{y})}{\partial \theta^2} \right) = \mathbb{E} \left( \frac{\partial}{\partial \theta} \left[ \frac{f'(\mathbf{y}; \theta)}{f(\mathbf{y}; \theta)} \right] \right) \\ &= \mathbb{E} \left( \frac{f''(\mathbf{y}; \theta)f(\mathbf{y}; \theta) - f'(\mathbf{y}; \theta)f'(\mathbf{y}; \theta)}{[f(\mathbf{y}; \theta)]^2} \right) \\ &= \mathbb{E} \left( \frac{f''(\mathbf{y}; \theta)}{f(\mathbf{y}; \theta)} \right) - \mathbb{E} \left( \left[ \frac{f'(\mathbf{y}; \theta)}{f(\mathbf{y}; \theta)} \right]^2 \right) \\ &= \int \dots \int_{\mathbb{R}^n} \frac{f''(\mathbf{y}; \theta)}{f(\mathbf{y}; \theta)} \cancel{f(\mathbf{y}; \theta)} dy_1 \dots dy_n - \mathbb{E} \left( \left[ \frac{f'(\mathbf{y}; \theta)}{f(\mathbf{y}; \theta)} \right]^2 \right) \\ &= \int \dots \int_{\mathbb{R}^n} \frac{\partial^2 f(\mathbf{y}; \theta)}{\partial \theta^2} dy_1 \dots dy_n - \mathbb{E} \left[ \left( \frac{\partial \ell_n(\theta; \mathbf{y})}{\partial \theta} \right)^2 \right] \\ &= \frac{\partial^2}{\partial \theta^2} \underbrace{\int \dots \int_{\mathbb{R}^n} f(\mathbf{y}; \theta) dy_1 \dots dy_n}_1 - \mathbb{E} \left[ \left( \frac{\partial \ell_n(\theta; \mathbf{y})}{\partial \theta} \right)^2 \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{\partial^2}{\partial \theta^2} 1 - \mathbb{E} \left[ \left( \frac{\partial \ell_n(\theta; \mathbf{y})}{\partial \theta} \right)^2 \right] = -\mathbb{E} \left[ \left( \frac{\partial \ell_n(\theta; \mathbf{y})}{\partial \theta} \right)^2 \right] \\
\mathbb{E} \left( \frac{\partial U_n(\theta)}{\partial \theta} \right) &= -\mathbb{E}[U_n^2(\theta)].
\end{aligned} \tag{8.3}$$

Luego,  $\mathbb{V}[U_n(\theta)]$  se puede expresar como:

$$\mathbb{V}[U_n(\theta)] = -\mathbb{E} \left( \frac{\partial U_n(\theta)}{\partial \theta} \right). \tag{8.4}$$

**Ejemplo 8.2.** Si  $Y \sim \text{Exp}(\theta)$ , se sabe que  $\mathbb{E}(Y) = \frac{1}{\theta}$  y  $\mathbb{V}(Y) = \frac{1}{\theta^2}$ . Presente la función score y su respectiva esperanza y varianza.

*Solución.* La función de verosimilitud para una muestra aleatoria de  $Y$  es:

$$L_n(\theta; \mathbf{y}) = \prod_{i=1}^n [f(y_i; \theta)] = \prod_{i=1}^n [\theta e^{-\theta y_i}] = \theta^n e^{-\theta \sum_{i=1}^n y_i},$$

y la respectiva función de log-verosimilitud es:

$$\ell_n(\theta; \mathbf{y}) = \ln \left[ \theta^n e^{-\theta \sum_{i=1}^n y_i} \right] = n \ln(\theta) - \theta \sum_{i=1}^n y_i,$$

luego, la correspondiente función score, que se puede usar como función de estimación, es:

$$U_n(\theta) = \frac{\partial \ell_n(\theta; \mathbf{Y})}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^n Y_i,$$

y la ecuación de estimación para  $\theta$  sería:

$$U_n(\hat{\theta}) = 0 \quad \iff \quad n - \hat{\theta}_{MV} \sum_{i=1}^n Y_i = 0.$$

Por tanto,  $\hat{\theta}_{MV} = \frac{n}{\sum_{i=1}^n Y_i} = (\bar{Y})^{-1}$ .

La esperanza y varianza de  $U_n(\theta)$  son:

$$\mathbb{E}[U_n(\theta)] = \mathbb{E} \left[ \frac{n}{\theta} - \sum_{i=1}^n Y_i \right] = \frac{n}{\theta} - \sum_{i=1}^n \mathbb{E}(Y_i) = \frac{n}{\theta} - \frac{n}{\theta} = 0,$$

que ya se tenía por (8.1), y

$$\begin{aligned} \mathbb{V}[U_n(\theta)] &\stackrel{\text{por (8.4)}}{=} -\mathbb{E}\left[\frac{\partial U_n(\theta)}{\partial \theta}\right] = -\mathbb{E}\left[\frac{\partial}{\partial \theta}\left(\frac{n}{\theta} - \sum_{i=1}^n Y_i\right)\right] \\ &= -\mathbb{E}\left[-n\theta^{-2}\right] = \frac{n}{\theta^2} \end{aligned}$$

o por (8.2):

$$\begin{aligned} \mathbb{E}[U_n^2(\theta)] &= \mathbb{E}\left[\left(\frac{n}{\theta} - \sum_{i=1}^n Y_i\right)^2\right] = \mathbb{E}\left[\frac{n^2}{\theta^2} - \frac{2n^2}{\theta}\bar{Y} + n^2\bar{Y}^2\right] \\ \mathbb{V}[U_n(\theta)] &= \frac{n^2}{\theta^2} - \frac{2n^2}{\theta} \times \frac{1}{\theta} + n^2 \left\{ \frac{1}{n\theta^2} + \left(\frac{1}{\theta}\right)^2 \right\} = \frac{n}{\theta^2}. \quad \square \end{aligned}$$

**Nota 8.2.** (Algunas propiedades asintóticas de la función score). *Notando que  $U_n(\theta)$  se puede expresar como una sumatoria, esto es,*

$$\begin{aligned} U_n(\theta) &= \frac{\partial \ell_n(\theta; \mathbf{y})}{\partial \theta} = \frac{\partial \{\ln [f(\mathbf{y}; \theta)]\}}{\partial \theta} = \frac{\partial}{\partial \theta} \left\{ \ln \left( \prod_{i=1}^n f(y_i; \theta) \right) \right\} \\ &= \sum_{i=1}^n \frac{\partial}{\partial \theta} \ln [f(y_i; \theta)], \end{aligned}$$

luego, la estandarización para la variable  $U_n(\theta)$  está dada por:

$$\frac{U_n(\theta) - \mathbb{E}[U_n(\theta)]}{\sqrt{\mathbb{V}[U_n(\theta)]}} = \frac{\sum_{i=1}^n \frac{\partial}{\partial \theta} \ln [f(y_i; \theta)] - n\mathbb{E}\left(\frac{\partial}{\partial \theta} \ln [f(y_1; \theta)]\right)}{\sqrt{n\mathbb{V}\left(\frac{\partial}{\partial \theta} \ln [f(y_1; \theta)]\right)}}, \quad (8.5)$$

nótese que el término del denominador del lado derecho se puede reescribir como

$$n\mathbb{V}\left(\frac{\partial}{\partial \theta} \ln [f(y_1; \theta)]\right) = n\mathbb{V}[U_1(\theta)] = n\mathbb{E}[U_1^2(\theta)] = nI_1(\theta),$$

donde  $I_1(\theta)$  es la información de Fisher dada en (5.9). Al reemplazar en (8.5) y usando el teorema del límite central clásico, teorema 3.22, se tiene que:

$$\frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \theta} \ln [f(y_i; \theta)]}{\sqrt{I_1(\theta)}} \xrightarrow{D} N(0, 1)$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \theta} \ln [f(y_i; \theta)] \xrightarrow{D} N(0, I_1(\theta))$$

$$\frac{1}{\sqrt{n}} U_n(\theta) \xrightarrow{D} N(0, I_1(\theta)).$$

De otro lado, por la ley débil de los grandes números de Khintchine, teorema 3.19,

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \ln [f(y_i; \theta)] \xrightarrow{P} \mathbb{E} \left( \frac{\partial^2}{\partial \theta^2} \ln [f(y_1; \theta)] \right) = \mathbb{E} \left( \frac{\partial U_1(\theta)}{\partial \theta} \right)$$

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \ln [f(y_i; \theta)] \xrightarrow{P} -\mathbb{E} \left( U_1^2(\theta) \right) \quad \text{por (8.3)}$$

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \ln [f(y_i; \theta)] \xrightarrow{P} -I_1(\theta)$$

$$\frac{1}{n} \frac{\partial U_n(\theta)}{\partial \theta} \xrightarrow{P} -I_1(\theta).$$

**Ejemplo 8.3.** Continuando con el ejemplo 8.2, sobre la función de estimación para el parámetro de la distribución exponencial,  $\theta$ , encuentre la distribución asintótica de la función de estimación cuando es convenientemente normalizada.

*Solución.* Sea

$$U_1(\theta) = \frac{\partial \ell(\theta; y_1)}{\partial \theta} = \frac{1}{\theta} - y_1, \quad \Rightarrow \quad \frac{\partial U_1(\theta)}{\partial \theta} = -\theta^{-2}$$

$$I_1(\theta) = \mathbb{E}[U_1^2(\theta)] \underset{\text{Ver (8.2)}}{=} \mathbb{V}[U_1(\theta)] \underset{\text{Ver (8.4)}}{=} -\mathbb{E} \left( \frac{\partial U_1(\theta)}{\partial \theta} \right) = \frac{1}{\theta^2}, \quad (8.6)$$

luego,

$$\frac{1}{\sqrt{n}} U_n(\theta) \xrightarrow{D} N(0, I_1(\theta))$$

$$\frac{1}{\sqrt{n}} \left( \frac{n}{\theta} - \sum_{i=1}^n Y_i \right) \xrightarrow{D} N \left( 0, \frac{1}{\theta^2} \right)$$

$$\sqrt{n} \left( \bar{Y} - \frac{1}{\theta} \right) \xrightarrow{D} N \left( 0, \frac{1}{\theta^2} \right).$$

Cabe resaltar que

$$\frac{1}{n} \frac{\partial U_n(\theta)}{\partial \theta} = \frac{1}{n} \frac{\partial}{\partial \theta} \left( \frac{n}{\theta} - \sum_{i=1}^n y_i \right) = \frac{1}{n} \times (-n\theta^{-2}) = -\theta^{-2}$$

no depende ni de  $n$ , ni de los  $Y_i$ . ☑

**Teorema 8.1.** Según [82, p. 205], si  $Y_1, Y_2, \dots, Y_n$  son variables aleatorias iid con *pdf*  $f(y_1; \theta)$ ,  $y \in \mathbb{R}$ ,  $\theta \in \Theta \subset \mathbb{R}$ , que satisfacen las siguientes condiciones:

1. Las derivadas parciales de primer y segundo orden

$$\frac{\partial}{\partial \theta} f(\mathbf{y}; \theta) \quad y \quad \frac{\partial^2}{\partial \theta^2} f(\mathbf{y}; \theta)$$

existen casi en toda parte y son tales que:

$$\left| \frac{\partial}{\partial \theta} f(\mathbf{y}; \theta) \right| \leq H_1(\mathbf{y}) \quad y \quad \left| \frac{\partial^2}{\partial \theta^2} f(\mathbf{y}; \theta) \right| \leq H_2(\mathbf{y})$$

en que

$$\int \cdots \int_{\mathbb{R}^n} H_j(\mathbf{y}) dy_1 \cdots dy_n < \infty, \quad j = 1, 2.$$

2.

$$U_n(\theta) = \frac{\partial}{\partial \theta} \ln [f(\mathbf{y}; \theta)] \quad y \quad \frac{\partial U_n(\theta)}{\partial \theta} = \frac{\partial^2}{\partial \theta^2} \ln [f(\mathbf{y}; \theta)]$$

existen casi en toda parte y son tales que:

a)  $Y_1$  tiene **información de Fisher** finita, es decir,

$$0 < I_1(\theta) = \mathbb{E}[U_1^2(\theta)] = \mathbb{E}\left\{\left[\frac{\partial}{\partial \theta} \ln [f(Y_1; \theta)]\right]^2\right\} < \infty.$$

b) Cuando  $\delta \rightarrow 0$ , se tiene que:

$$\mathbb{E}\left\{\sup_{\{h: |h| \leq \delta\}} \left| \frac{\partial U_1(\theta + h)}{\partial \theta} - \frac{\partial U_1(\theta)}{\partial \theta} \right|\right\} = \psi_\delta \rightarrow 0.$$

Entonces, el **EMV**,  $\hat{\theta}_n$ , de  $\theta$  es tal que

$$\hat{\theta}_n \xrightarrow{P} \theta \quad y \quad \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N(0, I_1^{-1}(\theta)).$$

*Demostración.* Ver detalles de la prueba en [82, p. 205]. ☑

**Ejemplo 8.4.** Continuando con el ejemplo 8.2, sobre la función de estimación para el parámetro de la distribución exponencial,  $\theta$ , encuentre la convergencia en probabilidad del **EMV** de  $\theta$ ,  $\hat{\theta}_{MV}$ , y su distribución asintótica cuando es convenientemente normalizado.

*Solución.* Suponiendo que se cumplen las condiciones del teorema 8.1, se tiene que:

$$\frac{n}{\sum_{i=1}^n Y_i} \xrightarrow{P} \theta$$

$$\sqrt{n} \left( \frac{n}{\sum_{i=1}^n Y_i} - \theta \right) \xrightarrow{D} N(0, I_1^{-1}(\theta)).$$

Usando (8.6),

$$I_1^{-1}(\theta) = \left( \frac{1}{\theta^2} \right)^{-1} = \theta^2,$$

luego,

$$\sqrt{n} \left( \frac{1}{\bar{Y}_n} - \theta \right) \xrightarrow{D} N(0, \theta^2).$$

Esto coincide con los resultados asintóticos obtenidos en el ejemplo 5.23 con respecto a la convergencia en distribución del EMV de  $\theta$ .  $\checkmark$

## 8.2. Función de estimación. Caso multiparamétrico y vectorial

Sea una función  $\psi_n : \Omega^n \times \Theta \rightarrow \mathbb{R}^r$ ,

$$\psi_n(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^n \psi(\boldsymbol{\theta}; y_i),$$

donde  $\boldsymbol{\theta}$  es un vector  $r$ -dimensional y  $\mathbf{y}$  un vector  $n$ -dimensional.

Sea el estimador  $\hat{\boldsymbol{\theta}}_n$  la solución a la ecuación

$$\sum_{i=1}^n \psi(\hat{\boldsymbol{\theta}}_n; y_i) = \mathbf{0}. \quad (8.7)$$

Note que (8.7) es una ecuación de estimación.

**Ejemplo 8.5.** (Distribución normal univariada). Considere  $Y_1, Y_2, \dots, Y_n$  variables aleatorias *iid*, tal que  $Y_1 \sim N(\mu, \sigma^2)$ . Si  $\mu$  y  $\sigma^2$  son parámetros de interés y son desconocidos, presente una función de estimación.

*Solución.* El vector de parámetros a estimar es

$$\boldsymbol{\theta} = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}.$$

La respectiva función de log-verosimilitud es

$$\ell(\boldsymbol{\theta}; \mathbf{y}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2,$$

luego una función de estimación para  $\mu$  y  $\sigma^2$  sería dada por:

$$\begin{aligned} \boldsymbol{\psi}_n(\boldsymbol{\theta}; \mathbf{y}) &= U_n(\boldsymbol{\theta}) \\ &= \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{y})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{y})}{\partial \mu} \\ \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{y})}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 \end{pmatrix}. \quad \square \end{aligned}$$

**Ejemplo 8.6.** Sean  $Y_1, Y_2, \dots, Y_n$  variables aleatorias independientes con media  $\mathbb{E}(Y_i) = \mu_i(\boldsymbol{\theta})$ ,  $\mu_i$  duplamente diferenciable, y  $\mathbb{V}(Y_i) = \sigma^2$ . Presente una función de estimación.

*Solución.* Una función definida como

$$\boldsymbol{\psi}_n(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^n \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \{y_i - \mu_i(\boldsymbol{\theta})\}$$

es una función de estimación. Las correspondientes ecuaciones de estimación son dadas por:

$$\boldsymbol{\psi}_k(\hat{\boldsymbol{\theta}}_k; \mathbf{y}) = \mathbf{0} \quad \text{para } k = 1, 2, \dots, n. \quad (8.8)$$

Cabe resaltar que las ecuaciones (8.8) son conocidas como *ecuaciones normales* y sus raíces son los estimadores de mínimos cuadrados de  $\boldsymbol{\theta}$ .  $\square$

**DEFINICIÓN 8.1.** Según [47, p. 138], una función de estimación  $\boldsymbol{\psi}_n(\boldsymbol{\theta}; \mathbf{y})$  es *insesgada*<sup>50</sup> si

$$\mathbb{E}_{\boldsymbol{\theta}} \{\boldsymbol{\psi}(\boldsymbol{\theta}; \mathbf{y})\} = \mathbf{0} \quad \forall \boldsymbol{\theta} \in \Theta.$$

**Ejemplo 8.7.** Considerando la información del ejemplo 8.5 sobre la distribución normal univariada, muestre que  $\boldsymbol{\psi}_n(\boldsymbol{\theta}; \mathbf{y})$  es una función de estimación insesgada.

<sup>50</sup>Función de estimación insesgada y no estimador insesgado.

*Solución.* Se tiene que:

$$\psi_n(\boldsymbol{\theta}; \mathbf{y}) = \begin{pmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 \end{pmatrix},$$

aplicando el operador esperanza:

$$\begin{aligned} \mathbb{E}(\psi_n(\boldsymbol{\theta}; \mathbf{y})) &= \mathbb{E} \left[ \begin{pmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \mu) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (Y_i - \mu)^2 \end{pmatrix} \right] \\ &= \begin{bmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n (\mathbb{E}(Y_i) - \mu) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n \mathbb{E}[(Y_i - \mu)^2] \end{bmatrix} = \begin{pmatrix} 0 \\ -\frac{n}{2\sigma^2} + \frac{n\sigma^2}{2\sigma^4} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

Luego  $\psi_n(\boldsymbol{\theta}; \mathbf{y})$  es una función de estimación insesgada. ✓

**Ejemplo 8.8.** Considerando la información del ejemplo 8.1, muestre que  $\psi_n(\boldsymbol{\theta}; \mathbf{y})$  es una función de estimación insesgada.

*Solución.* Se tiene la función de estimación

$$\psi(\mathbf{Y}; \theta) = \sum_{i=1}^n (Y_i^2 - \theta),$$

aplicando el operador esperanza:

$$\begin{aligned} \mathbb{E}(\psi(\mathbf{Y}; \theta)) &= \mathbb{E} \left[ \sum_{i=1}^n (Y_i^2 - \theta) \right] = \sum_{i=1}^n \left[ \mathbb{E}(Y_i^2) - \theta \right] = \sum_{i=1}^n [\mathbb{V}(Y_i) - \theta] \\ &= n(\theta - \theta) = 0. \end{aligned}$$

Luego  $\psi_n(\mathbf{Y}, \theta)$  es una función de estimación insesgada. ✓

## 8.2.1. Propiedades de una función de estimación multivariada

Notando que  $U_n(\boldsymbol{\theta})$  se puede expresar como una sumatoria, esto es,

$$U_n(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \ln[f(\boldsymbol{\theta}; \mathbf{y})] = \frac{\partial}{\partial \boldsymbol{\theta}} \ln \left[ \prod_{i=1}^n f(\boldsymbol{\theta}; y_i) \right]$$

$$= \frac{\partial}{\partial \boldsymbol{\theta}} \sum_{i=1}^n \ln [f(\boldsymbol{\theta}; y_i)] = \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\theta}} \ln [f(\boldsymbol{\theta}; y_i)], \quad (8.9)$$

las respectivas esperanzas de  $U_n(\boldsymbol{\theta})$  y de  $\frac{\partial U_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ , y su convergencia en probabilidad son dadas por:

- Para cada  $\theta_i$ ,

$$\begin{aligned} \mathbb{E} \left( \frac{\partial \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \theta_i} \right) &= \mathbb{E} \left( \frac{\partial}{\partial \theta_i} \log f(\mathbf{y}; \boldsymbol{\theta}) \right) = \mathbb{E} \left( \frac{1}{f(\mathbf{y}; \boldsymbol{\theta})} \frac{\partial f(\mathbf{y}; \boldsymbol{\theta})}{\partial \theta_i} \right) \\ &= \int \cdots \int_{\mathbb{R}^n} \left[ \frac{1}{f(\mathbf{y}; \boldsymbol{\theta})} \frac{\partial}{\partial \theta_i} f(\mathbf{y}; \boldsymbol{\theta}) \right] f(\mathbf{y}; \boldsymbol{\theta}) d\mathbf{y} \\ &= \frac{\partial}{\partial \theta_i} \underbrace{\int \cdots \int_{\mathbb{R}^n} f(\mathbf{y}; \boldsymbol{\theta}) dy_1 \dots dy_n}_{1} = \frac{\partial}{\partial \theta_i} 1 = 0 \\ \mathbb{E} [U_n(\boldsymbol{\theta})] &= \mathbf{0}. \end{aligned} \quad (8.10)$$

- Denotando a

$$f'_i(\mathbf{y}; \boldsymbol{\theta}) = \frac{\partial f(\mathbf{y}; \boldsymbol{\theta})}{\partial \theta_i}, \quad f'_j(\mathbf{y}; \boldsymbol{\theta}) = \frac{\partial f(\mathbf{y}; \boldsymbol{\theta})}{\partial \theta_j} \quad \text{y} \quad f''(\mathbf{y}; \boldsymbol{\theta}) = \frac{\partial^2 f(\mathbf{y}; \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j},$$

se tiene que:

$$\begin{aligned} \mathbb{E} \left( \frac{\partial^2 \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \theta_i \partial \theta_j} \right) &= \mathbb{E} \left( \frac{\partial}{\partial \theta_j} \left[ \frac{f'_i(\mathbf{y}; \boldsymbol{\theta})}{f(\mathbf{y}; \boldsymbol{\theta})} \right] \right) \\ &= \mathbb{E} \left( \frac{f''(\mathbf{y}; \boldsymbol{\theta}) f(\mathbf{y}; \boldsymbol{\theta}) - f'_i(\mathbf{y}; \boldsymbol{\theta}) f'_j(\mathbf{y}; \boldsymbol{\theta})}{[f(\mathbf{y}; \boldsymbol{\theta})]^2} \right) \\ &= \mathbb{E} \left( \frac{f''(\mathbf{y}; \boldsymbol{\theta})}{f(\mathbf{y}; \boldsymbol{\theta})} \right) - \mathbb{E} \left( \frac{f'_i(\mathbf{y}; \boldsymbol{\theta}) f'_j(\mathbf{y}; \boldsymbol{\theta})}{[f(\mathbf{y}; \boldsymbol{\theta})]^2} \right) \\ &= \int \cdots \int_{\mathbb{R}^n} \frac{f''(\mathbf{y}; \boldsymbol{\theta})}{f(\mathbf{y}; \boldsymbol{\theta})} f(\mathbf{y}; \boldsymbol{\theta}) dy_1 \dots dy_n - \mathbb{E} \left( \frac{f'_i(\mathbf{y}; \boldsymbol{\theta}) f'_j(\mathbf{y}; \boldsymbol{\theta})}{[f(\mathbf{y}; \boldsymbol{\theta})]^2} \right) \\ &= \int \cdots \int_{\mathbb{R}^n} \frac{\partial^2 f(\mathbf{y}; \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} dy_1 \dots dy_n - \mathbb{E} \left[ \left( \frac{\partial \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \theta_i} \right) \left( \frac{\partial \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \theta_j} \right) \right] \\ &= \frac{\partial^2}{\partial \theta_i \partial \theta_j} \int \cdots \int_{\mathbb{R}^n} f(\mathbf{y}; \boldsymbol{\theta}) dy_1 \dots dy_n - \mathbb{E} \left[ \left( \frac{\partial \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \theta_i} \right) \left( \frac{\partial \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \theta_j} \right) \right] \\ &= \frac{\partial^2}{\partial \theta_i \partial \theta_j} 1 - \mathbb{E} \left[ \left( \frac{\partial \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \theta_i} \right) \left( \frac{\partial \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \theta_j} \right) \right], \end{aligned}$$

luego,

$$\mathbb{E} \left( \frac{\partial^2 \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \theta_i \partial \theta_j} \right) = -\mathbb{E} \left[ \left( \frac{\partial \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \theta_i} \right) \left( \frac{\partial \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \theta_j} \right) \right]. \quad (8.11)$$

Cabe notar que

$$\mathbb{E} \left[ \left( \frac{\partial \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \theta_i} \right) \left( \frac{\partial \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \theta_j} \right) \right]$$

es similar al elemento  $(i, j)$  de la matriz de *información de Fisher* de la muestra dada en (5.30).

- Por la ley débil de los grandes números de Khintchine, teorema 3.19, y por el teorema 3.8,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta_i} \ln [f(y_i; \boldsymbol{\theta})] &\xrightarrow{P} \mathbb{E} \left( \frac{\partial}{\partial \theta_i} \ln [f(y_1; \boldsymbol{\theta})] \right) = 0 \\ \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\theta}} \ln [f(y_i; \boldsymbol{\theta})] &\xrightarrow{P} \mathbf{0} \quad \Rightarrow \quad \frac{1}{n} \mathbf{U}_n(\boldsymbol{\theta}) \xrightarrow{P} \mathbf{0} \end{aligned}$$

y

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln [f(y_i; \boldsymbol{\theta})] &\xrightarrow{P} \mathbb{E} \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln [f(y_1; \boldsymbol{\theta})] \right) \\ \frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln [f(y_i; \boldsymbol{\theta})] &\xrightarrow{P} -\mathbb{E} \left[ \left( \frac{\partial \ell_n(\boldsymbol{\theta}; y_1)}{\partial \theta_i} \right) \left( \frac{\partial \ell_n(\boldsymbol{\theta}; y_1)}{\partial \theta_j} \right) \right] \\ \frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \ln [f(y_i; \boldsymbol{\theta})] &\xrightarrow{P} -\mathbf{I}_1(\boldsymbol{\theta}) \\ \frac{1}{n} \frac{\partial \{[\mathbf{U}_n(\boldsymbol{\theta})]^\top\}}{\partial \boldsymbol{\theta}} &\xrightarrow{P} -\mathbf{I}_1(\boldsymbol{\theta}). \end{aligned} \quad (8.12)$$

**Ejemplo 8.9.** Considerando la información del ejemplo 8.5 sobre la distribución normal univariada, presente la convergencia asintótica de la función score y de la derivada de la función score.

*Solución.* Se tiene

$$\mathbf{U}_n(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) \\ \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 - \frac{n}{2\sigma^2} \end{pmatrix} \xrightarrow{n=1} \mathbf{U}_1(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{\sigma^2} (y_1 - \mu) \\ \frac{1}{2\sigma^4} (y_1 - \mu)^2 - \frac{1}{2\sigma^2} \end{pmatrix}$$

. Luego,

$$\begin{aligned} \frac{\partial \{[U_n(\boldsymbol{\theta})]^\top\}}{\partial \boldsymbol{\theta}} &= \begin{pmatrix} \frac{\partial \left[ \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) \right]}{\frac{\partial \mu}{\partial \sigma^2}} & \frac{\partial \left[ \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 - \frac{n}{2\sigma^2} \right]}{\frac{\partial \mu}{\partial \sigma^2}} \\ \frac{\partial \left[ \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) \right]}{\partial \sigma^2} & \frac{\partial \left[ \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 - \frac{n}{2\sigma^2} \right]}{\partial \sigma^2} \end{pmatrix} \\ &= \begin{pmatrix} \frac{-n}{\sigma^2} & \frac{-1}{\sigma^4} \sum_{i=1}^n (y_i - \mu) \\ \frac{-1}{\sigma^4} \sum_{i=1}^n (y_i - \mu) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (y_i - \mu)^2 \end{pmatrix} \\ &= \frac{-1}{\sigma^4} \begin{pmatrix} n\sigma^2 & \sum_{i=1}^n (y_i - \mu) \\ \sum_{i=1}^n (y_i - \mu) & \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 - \frac{n}{2} \end{pmatrix}, \end{aligned}$$

para  $n = 1$  se obtiene que:

$$\frac{\partial \{[U_1(\boldsymbol{\theta})]^\top\}}{\partial \boldsymbol{\theta}} = \frac{-1}{\sigma^4} \begin{pmatrix} \sigma^2 & (y_1 - \mu) \\ (y_1 - \mu) & \frac{1}{\sigma^2} (y_1 - \mu)^2 - \frac{1}{2} \end{pmatrix},$$

en que

$$\begin{aligned} \mathbb{E} \left[ \frac{\partial \{[U_1(\boldsymbol{\theta})]^\top\}}{\partial \boldsymbol{\theta}} \right] &= \mathbb{E} \left[ \frac{-1}{\sigma^4} \begin{pmatrix} \sigma^2 & (y_1 - \mu) \\ (y_1 - \mu) & \frac{1}{\sigma^2} (y_1 - \mu)^2 - \frac{1}{2} \end{pmatrix} \right] \\ &= \frac{-1}{\sigma^4} \begin{pmatrix} \sigma^2 & \mathbb{E}(y_1 - \mu) \\ \mathbb{E}(y_1 - \mu) & \frac{1}{\sigma^2} \mathbb{E}[(y_1 - \mu)^2] - \frac{1}{2} \end{pmatrix} \\ &= \frac{-1}{\sigma^4} \begin{pmatrix} \sigma^2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}. \end{aligned}$$

Entonces,

$$\frac{1}{n} U_n(\boldsymbol{\theta}) = \frac{1}{n} \begin{pmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 \end{pmatrix} \xrightarrow{P} \mathbf{0}$$

y

$$\frac{1}{n} \frac{\partial \{[U_n(\boldsymbol{\theta})]^\top\}}{\partial \boldsymbol{\theta}} \xrightarrow{P} -I_1(\boldsymbol{\theta})$$

$$\begin{aligned} \frac{1}{n} \times \frac{-1}{\sigma^4} & \begin{pmatrix} n\sigma^2 & \sum_{i=1}^n (y_i - \mu) \\ \sum_{i=1}^n (y_i - \mu) & \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 - \frac{n}{2} \end{pmatrix} \xrightarrow{P} \frac{-1}{\sigma^4} \begin{pmatrix} \sigma^2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \\ & \frac{-1}{\sigma^4} \begin{pmatrix} \sigma^2 & \bar{Y} - \mu \\ \bar{Y} - \mu & \frac{1}{n\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 - \frac{1}{2} \end{pmatrix} \xrightarrow{P} \frac{-1}{\sigma^4} \begin{pmatrix} \sigma^2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \\ & \begin{pmatrix} \sigma^2 & \bar{Y} - \mu \\ \bar{Y} - \mu & \frac{1}{n\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 - \frac{1}{2} \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \sigma^2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}. \quad \square \end{aligned}$$

**Teorema 8.2.** De acuerdo con [82, p. 209], dadas  $Y_1, Y_2, \dots, Y_n$  variables aleatorias iid con fdp  $f(y_1; \theta)$ ,  $y \in \mathbb{R}$ ,  $\theta \in \Theta \subset \mathbb{R}^q$ , que satisfacen las condiciones:

1. Para  $i, j = 1, 2, 3, \dots, q$ ,

$$\frac{\partial}{\partial \theta_i} f(y_1; \theta) \quad y \quad \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(y_1; \theta)$$

existen casi en toda parte y son tales que

$$\left| \frac{\partial}{\partial \theta_i} f(y_1; \theta) \right| \leq H_i(y) \quad y \quad \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(y_1; \theta) \right| \leq G_{ij}(y),$$

en que

$$\int_{\mathbb{R}} H_i(y) dy < \infty \quad y \quad \int_{\mathbb{R}} G_{ij}(y) dy < \infty.$$

2. Para  $i, j = 1, 2, 3, \dots, q$ ,

$$U_1(\theta) = \left\{ \frac{\partial}{\partial \theta_i} \ln [f(y_1; \theta)] \right\}_{q \times 1} \quad y \quad \frac{\partial U_1(\theta)}{\partial \theta_j} = \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln [f(y_1; \theta)]$$

existen casi en toda parte y son tales que:

a) La matriz de información de Fisher dada en (5.29) queda:

$$\begin{aligned} I_1(\theta) &= \mathbb{E} \{ U_1(\theta) [U_1(\theta)]^\top \} \\ &= \mathbb{E} \left\{ \left[ \frac{\partial}{\partial \theta} \ln [f(y_1; \theta)] \right] \left[ \frac{\partial}{\partial \theta} \ln [f(y_1; \theta)] \right]^\top \right\} \end{aligned}$$

es finita y definida positiva, y

b) Cuando  $\delta \rightarrow 0$ , se tiene que:

$$\mathbb{E} \left\{ \sup_{\{h: |h| \leq \delta\}} \left\| \frac{\partial \{[U_1(\boldsymbol{\theta} + h)]^\top\}}{\partial \boldsymbol{\theta}} - \frac{\partial \{[U_1(\boldsymbol{\theta})]^\top\}}{\partial \boldsymbol{\theta}} \right\| \right\} = \psi_\delta \rightarrow 0.$$

Entonces, el **EMV**  $\hat{\boldsymbol{\theta}}_n$  de  $\boldsymbol{\theta}$  es tal que:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_n &\xrightarrow{P} \boldsymbol{\theta} \\ \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) &\xrightarrow{D} N(\mathbf{0}, [I_1(\boldsymbol{\theta})]^{-1}). \end{aligned}$$

*Demostración.* Ver detalles de la prueba en [82, p. 209]. ☑

**Ejemplo 8.10.** Sean  $X_1, X_2, \dots, X_n$  y  $Y_1, Y_2, \dots, Y_m$  variables aleatorias *iid*, tales que  $X_1 \sim \text{Exp}(\theta_1)$  y  $Y_1 \sim \text{Exp}(\theta_2)$ , respectivamente. Sea el vector de parámetros de interés dado por:

$$\boldsymbol{\theta}_{2 \times 1} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix},$$

encuentre el **EMV**  $\hat{\boldsymbol{\theta}}_n$  de  $\boldsymbol{\theta}$ , su convergencia en probabilidad y su distribución asintótica.

*Solución.* Usando los resultados del ejemplo 8.4 se tiene que:

$$\frac{1}{\bar{X}} \xrightarrow{P} \theta_1 \quad \text{y} \quad \frac{1}{\bar{Y}} \xrightarrow{P} \theta_2,$$

por tanto, usando el teorema 3.8,

$$\begin{pmatrix} \frac{1}{\bar{X}_n} \\ \frac{1}{\bar{Y}_m} \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}.$$

Por otro lado,

$$\begin{aligned} \sqrt{n} \left( \frac{1}{\bar{X}} - \theta_1 \right) &\xrightarrow{D} N(0, \theta_1^2) \\ \sqrt{m} \left( \frac{1}{\bar{Y}} - \theta_2 \right) &\xrightarrow{D} N(0, \theta_2^2). \end{aligned}$$

Como los  $X$  y los  $Y$  son dos muestras aleatorias independientes, y usando Cramér-Wold, se tiene que:

$$\begin{pmatrix} \sqrt{n} \left( \frac{1}{\bar{X}_n} - \theta_1 \right) \\ \sqrt{m} \left( \frac{1}{\bar{Y}_m} - \theta_2 \right) \end{pmatrix} \xrightarrow{D} N_2 \left( \mathbf{0}, \begin{pmatrix} \theta_1^2 & 0 \\ 0 & \theta_2^2 \end{pmatrix} \right). \quad \text{☑}$$

**Nota 8.3.** Según [28]:

- $A \mathbb{E} \left( -\frac{\partial^2 \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right)$  se le llama **información de Fisher**.
- $A -\frac{\partial^2 \ell_n(\boldsymbol{\theta}; \mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$  se le conoce como **información de Fisher observada**.

## 8.3. Estadísticas de prueba

Sea  $U_n(\boldsymbol{\theta})$  la función de estimación dada en 8.9, esto es,

$$U_n(\boldsymbol{\theta}) = \left\{ \sum_{i=1}^n \frac{\partial}{\partial \theta_i} \ln [f(y_i; \boldsymbol{\theta})] \right\}_{q \times 1}.$$

Considere las hipótesis de interés

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{vs.} \quad H_a : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \quad (8.13)$$

en que  $\boldsymbol{\theta}_0$  es un vector de valores fijos. Las tres estadísticas más usadas (ver [82, p. 235]) para probar  $H_0$  son:

1. La *estadística de wald*, que se define como

$$Q_W = n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \mathbf{I}(\hat{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

en que  $\hat{\boldsymbol{\theta}}_n$  corresponde al **EMV** u otro estimador BAN (*Best Asymptotically Normal*) de  $\boldsymbol{\theta}_0$  y

$$\mathbf{I}(\hat{\boldsymbol{\theta}}_n) = -\frac{1}{n} \left. \frac{\partial \{ [U_n(\boldsymbol{\theta})]^\top \}}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n}.$$

2. La *estadística de Wilks* (o de razón de verosimilitud), que se define como

$$Q_L = -2 \ln(\lambda(\mathbf{y})) = 2 \{ \ln [L_n(\hat{\boldsymbol{\theta}}_n; \mathbf{y})] - \ln [L_n(\boldsymbol{\theta}_0; \mathbf{y})] \},$$

con

$$\lambda(\mathbf{y}) = \frac{\sup_{\boldsymbol{\theta}_0} L(\boldsymbol{\theta}; \mathbf{y})}{\sup_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \mathbf{y})} = \frac{L_n(\boldsymbol{\theta}_0; \mathbf{y})}{\sup_{\boldsymbol{\theta} \in \Theta} L_n(\boldsymbol{\theta}; \mathbf{y})}.$$

3. La *estadística de Rao* (estadística score):

$$Q_R = \frac{1}{n} [U_n(\boldsymbol{\theta}_0)]^\top [\mathbf{I}(\boldsymbol{\theta}_0)]^{-1} [U_n(\boldsymbol{\theta}_0)],$$

en que  $U_n(\theta_0) = U_n(\theta)|_{\theta=\theta_0}$ . Cabe resaltar que la estadística de Rao no necesita el cálculo del EMV de  $\theta$ .

**Teorema 8.3.** Según [82, p. 137], si  $\{T_n\}$  es una sucesión de vectores alatorios  $p$ -dimensionales tal que  $\sqrt{n}(T_n - \theta) \xrightarrow{D} N_p(\mathbf{0}, \Sigma)$ ,  $\text{rango}(\Sigma) = q \leq p$ , y  $\{A_n\}$  una sucesión de matrices no estocásticas semidefinidas positivas. Entonces

$$Q_n = n(T_n - \theta)^\top A_n (T_n - \theta) \xrightarrow{D} \chi_q^2,$$

si y solo si  $A_n$  converge a alguna inversa generalizada de  $\Sigma$ .

*Demostración.* Ver detalles de la prueba en apéndice B.3. □

**Teorema 8.4.** De acuerdo con [82, p. 236], dadas  $X_1, X_2, \dots, X_n$  variables aleatorias iid con fdp  $f(x; \theta)$ ,  $\theta \in \Theta \subset \mathbb{R}^q$  las cuales satisfacen las condiciones del teorema 8.2. Considere el problema de probar las hipótesis  $H_0 : \theta = \theta_0$  vs.  $H_a : \theta \neq \theta_0$ . Entonces, bajo  $H_0$  cierta las estadísticas de Wald, Wilks y Rao tienen distribución asintótica chi-cuadrado con  $q$  grados de libertad, es decir,

$$Q_W \xrightarrow{D} \chi_q^2, \quad Q_L \xrightarrow{D} \chi_q^2, \quad Q_R \xrightarrow{D} \chi_q^2.$$

*Demostración.* Solo se mostrará la distribución de  $Q_W$ , las pruebas para las otras dos estadísticas, al igual que la demostración para  $Q_W$  se puede consultar en [82, p. 236].

Por el teorema 8.2 se sabe que si  $\hat{\theta}_n$  es el EMV de  $\theta$ , entonces

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N(\mathbf{0}, [I(\theta)]^{-1}).$$

Cuando  $\theta = \theta_0$ , es decir, bajo la hipótesis nula,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(\mathbf{0}, [I(\theta_0)]^{-1}).$$

Aplicando el teorema 8.3, se tiene que:

$$Q_W = n(\hat{\theta}_n - \theta_0)^\top I(\theta_0)(\hat{\theta}_n - \theta_0) \xrightarrow{D} \chi_q^2.$$

Usando la nota B.4 con  $x = \sqrt{n}(\hat{\theta}_n - \theta_0)$ ,  $A = I(\hat{\theta}_n)$  y  $B = I(\theta_0)$ , y suponiendo que se cumplen los supuestos de la nota B.4, se tiene que:

$$\lambda_{\min} \leq \frac{x^\top I(\hat{\theta}_n)x}{x^\top I(\theta_0)x} \leq \lambda_{\max},$$

en que  $\lambda_{\min}$  y  $\lambda_{\max}$  son los valores propios menor y mayor, respectivamente, de  $\mathbf{I}(\hat{\boldsymbol{\theta}}_n)\mathbf{I}(\boldsymbol{\theta}_0)^{-1}$ . Se puede mostrar que

$$\mathbf{I}(\hat{\boldsymbol{\theta}}_n)\mathbf{I}(\boldsymbol{\theta}_0)^{-1} \xrightarrow{P} \mathbf{I}_q \quad \text{y} \quad \frac{\mathbf{x}^\top \mathbf{I}(\hat{\boldsymbol{\theta}}_n)\mathbf{x}}{\mathbf{x}^\top \mathbf{I}(\boldsymbol{\theta}_0)\mathbf{x}} \xrightarrow{P} 1.$$

Como

$$\begin{aligned} Q_W &= n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \mathbf{I}(\hat{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ &= n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \mathbf{I}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \times \frac{\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \mathbf{I}(\hat{\boldsymbol{\theta}}_n)\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)}{\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \mathbf{I}(\boldsymbol{\theta}_0)\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)} \\ &= \underbrace{\mathbf{x}^\top \mathbf{I}(\boldsymbol{\theta}_0)\mathbf{x}}_{\downarrow D} \times \underbrace{\frac{\mathbf{x}^\top \mathbf{I}(\hat{\boldsymbol{\theta}}_n)\mathbf{x}}{\mathbf{x}^\top \mathbf{I}(\boldsymbol{\theta}_0)\mathbf{x}}}_{\downarrow P}, \\ &\quad \chi_q^2 \qquad \qquad \qquad 1 \end{aligned}$$

usando Slutsky se tiene que:

$$Q_W \xrightarrow{D} \chi_q^2. \quad \checkmark$$

Considere la hipótesis compuesta

$$H_0 : \mathbf{h}(\boldsymbol{\theta}) = \mathbf{0} \quad \text{vs.} \quad H_a : \mathbf{h}(\boldsymbol{\theta}) \neq \mathbf{0} \quad (8.14)$$

en que  $\mathbf{h} : \mathbb{R}^q \rightarrow \mathbb{R}^r$  es una función vectorial de valor real tal que la matriz  $\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial \mathbf{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$  existe y es continua en  $\boldsymbol{\theta}$  con  $\text{rango} [\mathbf{H}(\boldsymbol{\theta})] = r$ .

Sea  $\hat{\boldsymbol{\theta}}_n$  el **EMV** irrestricto (considerando el espacio de parámetros sin restricción) y  $\tilde{\boldsymbol{\theta}}_n$  el **EMV** restringido (considerando el espacio de parámetros con restricción) de  $\boldsymbol{\theta}$ , es decir, sujeto a  $\mathbf{h}(\tilde{\boldsymbol{\theta}}_n) = \mathbf{0}$ . Entonces las estadísticas para probar  $H_0$  más usadas son:

1. La *estadística de Wald*:

$$Q_W = n [\mathbf{h}(\hat{\boldsymbol{\theta}}_n)]^\top \{ [\mathbf{H}(\hat{\boldsymbol{\theta}}_n)]^\top [\mathbf{I}(\hat{\boldsymbol{\theta}}_n)]^{-1} \mathbf{H}(\hat{\boldsymbol{\theta}}_n) \}^{-1} \mathbf{h}(\hat{\boldsymbol{\theta}}_n),$$

también se puede reemplazar  $\hat{\boldsymbol{\theta}}$  por un estimador BAN (Best Asymptotically Normal) de  $\boldsymbol{\theta}$ , y

$$\mathbf{I}(\hat{\boldsymbol{\theta}}_n) = -\frac{1}{n} \left. \frac{\partial \{ [\mathbf{U}_n(\boldsymbol{\theta})]^\top \}}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n}.$$

2. La *estadística de Wilks* (o de razón de verosimilitud):

$$Q_L = -2 \ln(\lambda_n) = 2 \{ \ln [L_n(\hat{\theta}_n)] - \ln [L_n(\bar{\theta}_n)] \},$$

en que

$$\lambda_n = \frac{\sup_{\{\theta \in \Theta: h(\theta)=0\}} L_n(\theta)}{\sup_{\theta \in \Theta} L_n(\theta)}.$$

3. El *test de Rao* (estadística score):

$$Q_R = n^{-1} [U_n(\bar{\theta})]^\top [I(\bar{\theta}_n)]^{-1} U_n(\bar{\theta}),$$

en que  $U_n(\bar{\theta}) = U_n(\theta)|_{\theta=\bar{\theta}}$ .

Según [82, p. 240], bajo  $H_0$  ciertas las estadísticas de Wald, Wilks y Rao, presentadas arriba tienen distribución asintótica chi-cuadrado con  $r$  grados de libertad, es decir,

$$Q_W \xrightarrow{D} \chi_r^2, \quad Q_L \xrightarrow{D} \sim \chi_r^2, \quad Q_R \xrightarrow{D} \sim \chi_r^2.$$

**Ejemplo 8.11.** Considerando la información del ejemplo 8.10 sobre las dos distribuciones exponenciales, con parámetros  $\theta_1$  y  $\theta_2$ ,

1. Presente el test de razón de verosimilitud para probar  $H_0 : \theta_1 = \theta_2$  vs.  $H_a : \theta_1 \neq \theta_2$  y muestre como queda la región de rechazo.
2. Presente el test de Wald para probar las hipótesis del punto 1 y muestre como queda la región de rechazo.
3. Presente el test de score para probar la hipótesis del punto 1 y muestre como queda la región de rechazo.
4. Una muestra con  $n = 10$  de  $X$  presentó los valores: 0,33; 0,24; 1,04; 0,13; 0,60; 0,19; 0,02; 0,25; 0,63 y 0,74. Una muestra con  $m = 12$  de  $Y$  presentó los valores: 2,09; 0,66; 0,02; 0,14; 0,06; 0,51; 0,02; 0,05; 0,66; 0,02; 0,22 y 2,30. ¿Cómo queda la región de (rechazo, aceptación) de los tests obtenidos en 1, 2 y 3 para  $\alpha = 0,05$ ?

*Solución.* Sea el vector de parámetros de interés dado por:

$$\theta_{2 \times 1} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix},$$

para probar las hipótesis  $H_0 : \theta_1 = \theta_2$  vs.  $H_a : \theta_1 \neq \theta_2$  se tiene de (8.14):

$$h(\theta)_{1 \times 1} = \theta_1 - \theta_2 \quad \text{y} \quad H(\theta) = \frac{\partial h(\theta)}{\partial \theta} = \frac{\partial (\theta_1 - \theta_2)}{\partial \theta} = \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

luego  $q = 2$  y  $r = 1$ . El **EMV** de  $\theta$  irrestricto se encontró en el ejemplo 8.10:

$$\hat{\theta}_1 = \frac{1}{\bar{X}} \quad \text{y} \quad \hat{\theta}_2 = \frac{1}{\bar{Y}}. \quad (8.15)$$

Por otro lado, el **EMV** de  $\theta$  restringido, esto es, bajo  $H_0 : \theta_1 = \theta_2 = \theta$  cierta, es

$$L_{n+m}(\theta; \mathbf{x}, \mathbf{y}) = \theta^{n+m} \exp \left\{ -\theta \sum_{i=1}^n x_i \right\} \exp \left\{ -\theta \sum_{j=1}^m y_j \right\}$$

$$l_{n+m}(\theta; \mathbf{x}, \mathbf{y}) = (n+m) \ln(\theta) - \theta(n\bar{X} + m\bar{Y}),$$

derivando con respecto a  $\theta$  se tiene que:

$$\frac{\partial l_{n+m}(\theta; \mathbf{x}, \mathbf{y})}{\partial \theta} = \frac{n+m}{\theta} - n\bar{X} - m\bar{Y},$$

igualando a cero se encuentra  $\bar{\theta}$ ,

$$\frac{n+m}{\bar{\theta}} - n\bar{X} - m\bar{Y} = 0 \quad \Rightarrow \quad \bar{\theta} = \frac{n+m}{n\bar{X} + m\bar{Y}}.$$

1. El test de razón de verosimilitud para probar las hipótesis  $H_0 : \theta_1 = \theta_2$  vs.  $H_a : \theta_1 \neq \theta_2$  es dado por:

$$Q_L = -2 \ln(\lambda_n) = 2 \left\{ \ln [L_n(\hat{\theta}_n)] - \ln [L_n(\bar{\theta}_n)] \right\},$$

usando (8.15):

$$\begin{aligned} \ln [L_n(\hat{\theta}_n)] &= n \ln(\theta_1) + m \ln(\theta_2) - n\theta_1\bar{X} - m\theta_2\bar{Y} \Big|_{\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}} \\ &= n \ln \left( \frac{1}{\bar{X}} \right) + m \ln \left( \frac{1}{\bar{Y}} \right) - n \left( \frac{\bar{X}}{\bar{X}} \right) - m \left( \frac{\bar{Y}}{\bar{Y}} \right) \\ &= -n \ln(\bar{X}) - m \ln(\bar{Y}) - n - m \\ &= (n+m) \ln(\theta) - \theta(n\bar{X} + m\bar{Y}) \Big|_{\theta = \frac{n+m}{n\bar{X} + m\bar{Y}}} \\ &= (n+m) \ln \left( \frac{n+m}{n\bar{X} + m\bar{Y}} \right) - \frac{n+m}{n\bar{X} + m\bar{Y}} \times (n\bar{X} + m\bar{Y}) \\ &= (n+m) \ln \left( \frac{n+m}{n\bar{X} + m\bar{Y}} \right) - (n+m). \end{aligned}$$

Entonces el estadístico de razón de verosimilitud es

$$\begin{aligned}
 Q_L &= 2 \{ \ln [L_n(\hat{\theta}_n)] - \ln [L_n(\bar{\theta}_n)] \} \\
 &= 2 \left\{ -n \ln(\bar{X}) - m \ln(\bar{Y}) - (n+m) \ln \left( \frac{n+m}{n\bar{X} + m\bar{Y}} \right) + \right. \\
 &\quad \left. n + m \right\} = -2 \left\{ n \ln(\bar{X}) + m \ln(\bar{Y}) + (n+m) \ln \left( \frac{n+m}{n\bar{X} + m\bar{Y}} \right) \right\}.
 \end{aligned}$$

Bajo  $H_0 : \theta_1 = \theta_2$  verdadera  $Q_L \xrightarrow{D} \chi_1^2$  y se rechaza  $H_0$  si

$$Q_{Lc} > \chi_{1,1-\alpha}^2,$$

en que

$$Q_{Lc} = -2 \left\{ n \ln(\bar{X}) + m \ln(\bar{Y}) + (n+m) \ln \left( \frac{n+m}{n\bar{x} + m\bar{y}} \right) \right\}$$

y  $P(\chi \leq \chi_{1,1-\alpha}^2) = 1 - \alpha$ ,  $\chi \sim \chi_1^2$ .

2. El test de Wald para probar  $H_0 : \theta_1 = \theta_2$  vs.  $H_a : \theta_1 \neq \theta_2$  es dado por:

$$Q_W = [\mathbf{h}(\hat{\theta}_{n+m})]^\top \{ [\mathbf{H}(\hat{\theta}_{n+m})]^\top [\mathbf{I}_{F_N}(\hat{\theta})]^{-1} \mathbf{H}(\hat{\theta}_{n+m}) \}^{-1} \mathbf{h}(\hat{\theta}_{n+m}),$$

en que

$$\begin{aligned}
 \mathbf{h}(\hat{\theta}_n) &= \mathbf{h}(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\hat{\theta}_n} = \theta_1 - \theta_2 \Big|_{\boldsymbol{\theta}=\hat{\theta}_n} = \frac{1}{\bar{X}} - \frac{1}{\bar{Y}} \\
 \mathbf{H}(\hat{\theta}_n) &= \begin{pmatrix} 1 \\ -1 \end{pmatrix}.
 \end{aligned}$$

Para la matriz de información de Fisher,  $\mathbf{I}(\hat{\theta}_N)$ , con  $N = n + m$ , sean dos conjuntos de variables aleatorias independientes  $X_1, X_2, \dots, X_n$  y  $Y_1, Y_2, \dots, Y_m$ , que no son idénticamente distribuidas en conjunto, entonces se evaluará esta matriz para  $N$  como sigue:

$$\begin{aligned}
 \mathbf{I}_{F_N}(\hat{\theta}) &= - \left( \begin{array}{cc} \sum_{i=1}^n \frac{\partial^2 [\ln f(x_i; \theta_1)]}{\partial \theta_1^2} & 0 \\ 0 & \sum_{i=1}^m \frac{\partial^2 [\ln f(y_i; \theta_2)]}{\partial \theta_2^2} \end{array} \right) \Bigg|_{\boldsymbol{\theta}=\hat{\theta}_{n+m}} \\
 &= - \left( \begin{array}{cc} \frac{\partial \left[ \frac{n}{\theta_1} - n\bar{X} \right]}{\partial \theta_1} & 0 \\ 0 & \frac{\partial \left[ \frac{m}{\theta_2} - m\bar{Y} \right]}{\partial \theta_2} \end{array} \right) \Bigg|_{\boldsymbol{\theta}=\hat{\theta}_{n+m}} = \left( \begin{array}{cc} \frac{n}{\theta_1^2} & 0 \\ 0 & \frac{m}{\theta_2^2} \end{array} \right) \Bigg|_{\boldsymbol{\theta}=\hat{\theta}_{n+m}} \quad (8.16)
 \end{aligned}$$

$$= \begin{pmatrix} n\bar{X}^2 & 0 \\ 0 & m\bar{Y}^2 \end{pmatrix},$$

entonces  $Q_W$  sería dado por:

$$\begin{aligned} Q_W &= [\mathbf{h}(\hat{\theta}_n)]^\top \{[\mathbf{H}(\hat{\theta}_n)]^\top [\mathbf{I}_{F_N}(\hat{\theta})]^{-1} \mathbf{H}(\hat{\theta}_n)\}^{-1} \mathbf{h}(\hat{\theta}_n) \\ &= \left[ \frac{1}{\bar{X}} - \frac{1}{\bar{Y}} \right]^\top \left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix}^\top \begin{pmatrix} n\bar{X}^2 & 0 \\ 0 & m\bar{Y}^2 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}^{-1} \left[ \frac{1}{\bar{X}} - \frac{1}{\bar{Y}} \right] \\ &= \left[ \frac{1}{\bar{X}} - \frac{1}{\bar{Y}} \right]^2 \left\{ \left( \frac{1}{n\bar{X}^2} - \frac{1}{m\bar{Y}^2} \right) \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}^{-1} \\ &= \left[ \frac{\bar{Y} - \bar{X}}{\bar{X}\bar{Y}} \right]^2 \left\{ \frac{1}{n\bar{X}^2} + \frac{1}{m\bar{Y}^2} \right\}^{-1} = \frac{[\bar{Y} - \bar{X}]^2}{\bar{X}^2\bar{Y}^2} \left\{ \frac{n\bar{X}^2 + m\bar{Y}^2}{n\bar{X}^2m\bar{Y}^2} \right\}^{-1} \\ &= \frac{[\bar{Y} - \bar{X}]^2}{\bar{X}^2\bar{Y}^2} \frac{nm\bar{X}^2\bar{Y}^2}{n\bar{X}^2 + m\bar{Y}^2} = \frac{[\bar{Y} - \bar{X}]^2}{\frac{\bar{X}^2}{m} + \frac{\bar{Y}^2}{n}}. \end{aligned}$$

Bajo  $H_0 : \theta_1 = \theta_2$  verdadera  $Q_W \xrightarrow{D} \chi_1^2$  y se rechaza  $H_0$  si

$$Q_{Wc} > \chi_{1,1-\alpha}^2$$

en que  $Q_{Wc} = \frac{[\bar{y}-\bar{x}]^2}{\frac{\bar{x}^2}{m} + \frac{\bar{y}^2}{n}}$  y  $P(\chi \leq \chi_{1,1-\alpha}^2) = 1 - \alpha$ ,  $\chi \sim \chi_1^2$ .

3. El test de score para probar  $H_0 : \theta_1 = \theta_2$  vs.  $H_a : \theta_1 \neq \theta_2$  es dado por:

$$Q_R = [\mathbf{U}_{n+m}(\bar{\theta})]^\top [\mathbf{I}_{F_N}(\bar{\theta})]^{-1} \mathbf{U}_{n+m}(\bar{\theta}),$$

en que

$$\begin{aligned} \mathbf{U}_{n+m}(\bar{\theta}) &= \mathbf{U}_{n+m}(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} = \begin{pmatrix} \frac{n}{\theta_1} - n\bar{X} \\ \frac{m}{\theta_2} - m\bar{Y} \end{pmatrix} \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} = \begin{pmatrix} \frac{\frac{n}{n\bar{X}+m\bar{Y}} - n\bar{X}}{\frac{n}{n\bar{X}+m\bar{Y}}} \\ \frac{\frac{m}{n\bar{X}+m\bar{Y}} - m\bar{Y}}{\frac{n}{n\bar{X}+m\bar{Y}}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\frac{n^2\bar{X}+mn\bar{Y}}{n+m} - n\bar{X}}{\frac{n}{n+m}} \\ \frac{\frac{mn\bar{X}+m^2\bar{Y}}{n+m} - m\bar{Y}}{\frac{m}{n+m}} \end{pmatrix} = \frac{1}{n+m} \begin{pmatrix} n^2\bar{X} + mn\bar{Y} - n(n+m)\bar{X} \\ nm\bar{X} + m^2\bar{Y} - m(n+m)\bar{Y} \end{pmatrix} \\ &= \frac{mn}{n+m} \begin{pmatrix} \bar{Y} - \bar{X} \\ \bar{X} - \bar{Y} \end{pmatrix} \end{aligned}$$

y a partir de (8.16) la matriz de información de Fisher queda:

$$\mathbf{I}_{F_N}(\bar{\theta}) = \begin{pmatrix} \frac{n}{\theta_1^2} & 0 \\ 0 & \frac{m}{\theta_2^2} \end{pmatrix} \Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}_{n+m}} = \begin{pmatrix} \frac{n}{(n+m)^2} & 0 \\ \frac{m}{(n\bar{X}+m\bar{Y})^2} & 0 \\ 0 & \frac{m}{(n+m)^2} \\ 0 & \frac{m}{(n\bar{X}+m\bar{Y})^2} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{n(n\bar{X}+m\bar{Y})^2}{(n+m)^2} & 0 \\ 0 & \frac{m(n\bar{X}+m\bar{Y})^2}{(n+m)^2} \end{pmatrix} = \frac{(n\bar{X} + m\bar{Y})^2}{(n + m)^2} \begin{pmatrix} n & 0 \\ 0 & m \end{pmatrix},$$

entonces  $Q_R$  sería dado por:

$$\begin{aligned} Q_R &= [U_{n+m}(\bar{\theta})]^\top [I_{F_N}(\bar{\theta})]^{-1} U_{n+m}(\bar{\theta}) \\ &= \left[ \frac{mn}{n+m} \begin{pmatrix} \bar{Y} - \bar{X} \\ \bar{X} - \bar{Y} \end{pmatrix} \right]^\top \left[ \frac{(n\bar{X} + m\bar{Y})^2}{(n+m)^2} \begin{pmatrix} n & 0 \\ 0 & m \end{pmatrix} \right]^{-1} \left[ \frac{mn}{n+m} \begin{pmatrix} \bar{Y} - \bar{X} \\ \bar{X} - \bar{Y} \end{pmatrix} \right] \\ &= \frac{m^2 n^2}{(n\bar{X} + m\bar{Y})^2} \begin{pmatrix} \bar{Y} - \bar{X} \\ -(\bar{Y} - \bar{X}) \end{pmatrix}^\top \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{m} \end{pmatrix} \begin{pmatrix} \bar{Y} - \bar{X} \\ -(\bar{Y} - \bar{X}) \end{pmatrix} \\ &= \frac{m^2 n^2}{(n\bar{X} + m\bar{Y})^2} \begin{pmatrix} \frac{1}{n}(\bar{Y} - \bar{X}) & -\frac{1}{m}(\bar{Y} - \bar{X}) \end{pmatrix} \begin{pmatrix} \bar{Y} - \bar{X} \\ -(\bar{Y} - \bar{X}) \end{pmatrix} \\ &= \frac{m^2 n^2}{(n\bar{X} + m\bar{Y})^2} \left( \frac{1}{n}(\bar{Y} - \bar{X})^2 + \frac{1}{m}(\bar{Y} - \bar{X})^2 \right) \\ &= \frac{m^2 n^2 (\bar{Y} - \bar{X})^2}{(n\bar{X} + m\bar{Y})^2} \left( \frac{m+n}{nm} \right). \end{aligned}$$

Luego,

$$Q_R = \frac{mn(m+n)(\bar{Y} - \bar{X})^2}{(n\bar{X} + m\bar{Y})^2}.$$

Bajo  $H_0 : \theta_1 = \theta_2$  verdadera  $Q_R \xrightarrow{D} \chi_1^2$  y se rechaza  $H_0$  si

$$Q_{Rc} > \chi_{1,1-\alpha}^2,$$

en que  $Q_{Rc} = \frac{mn(m+n)(\bar{Y} - \bar{X})^2}{(n\bar{X} + m\bar{Y})^2}$  y  $P(\chi \leq \chi_{1,1-\alpha}^2) = 1 - \alpha$ ,  $\chi \sim \chi_1^2$ .

4. De los datos se tiene que:

$$n = 10, \quad \bar{x} = 0,417, \quad m = 12, \quad \bar{y} = 0,5625, \quad \chi_{1,0,95}^2 = 3,841$$

$$\begin{aligned} Q_{Lc} &= -2 \left\{ n \ln(\bar{X}) + m \ln(\bar{Y}) + (m+n) \ln \left( \frac{n+m}{n\bar{x} + m\bar{y}} \right) \right\} \\ &= -2 \{ 10 \ln(0,417) + 12 \ln(0,5625) + \\ &\quad (22) \ln \left( \frac{10+12}{10(0,417) + 12(0,5625)} \right) \} \end{aligned}$$

$$Q_{Lc} = 0,4825$$

$$Q_{Wc} = \frac{[\bar{y} - \bar{x}]^2}{\frac{\bar{x}^2}{m} + \frac{\bar{y}^2}{n}} = \frac{[0,5625 - 0,417]^2}{\frac{0,417^2}{12} + \frac{0,5625^2}{10}} = 0,4589$$

$$Q_{Rc} = \frac{mn(m+n)(\bar{y} - \bar{x})^2}{(n\bar{x} + m\bar{y})^2} = \frac{12 \times 10(12+10)(0,5625 - 0,417)^2}{(10(0,417) + 12(0,5625))^2} \\ = 0,4687.$$

Por otra parte, para los tres test la región de rechazo es dada por  $Q_{Lc}, Q_{Wc}, Q_{Rc} > 3,841$  y, por tanto, no hay suficiente evidencia para rechazar la hipótesis nula  $H_0 : \theta_1 = \theta_2$  usando los tres test a un nivel de significancia de 0,05.  $\checkmark$

## 8.4. Ejercicios

1. Considere la distribución del ejemplo 8.11 y los datos dados en el ejemplo. Evalúe los tres test para probar  $H_0 : \theta_1 = 2$  y  $\theta_2 = 1$ .
2. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias independientes, tal que cada  $X_i \sim \text{Poisson}(\alpha + \beta Z_i)$ .
  - a) Obtenga el EMV de  $\alpha$  y  $\beta$ .
  - b) Obtenga las estadísticas de razón de verosimilitud, de Wald y de Rao para probar la hipótesis  $\beta = 0$ .
3. Sean  $X_1, X_2, \dots, X_n$  variables aleatorias iid, tal que  $X_1$  tiene fdp de Pareto dada por [52, p. 565]:

$$f(x; \theta) = \frac{\theta c^\theta}{x^{\theta+1}}, \quad 0 < c < x, \quad 0 < \theta.$$

- a) Determine la solución  $\hat{\theta}_n$  de la ecuación de verosimilitud y encuentre su distribución cuando es convenientemente normalizada. Nota: resuelva primero el problema para  $\eta = \frac{1}{\theta}$ .
  - b) Determine los test de Rao, Wald y Wilks para probar  $H_0 : \theta = \theta_0$  vs.  $H_0 : \theta \neq \theta_0$ .
4. [12, p. 351]. Sección 5.3, ejercicio 13.







Puesto que el concepto de serie de Taylor se usa en algunas partes del libro y dado que algunos lectores quizás tengan solo una breve noción de esta temática, este apéndice contiene un breve repaso de las propiedades más importantes sobre este tipo de expansiones. El lector que desee profundizar puede consultar cualquier texto de cálculo, por ejemplo, [5], [23] o [56].

## A.1. Expansiones de Taylor

Para [5, p. 532] y [23, p. 440], una función  $f(x)$  que tenga derivadas de cualquier orden en un intervalo abierto en torno al punto  $x_0$  es infinitamente derivable en este intervalo y se puede expresar como la serie de potencias

$$f(x) = \sum_{k=0}^{\infty} a_k (x - x_0)^k = f(x_0) + \sum_{k=1}^{\infty} a_k (x - x_0)^k,$$

donde

$$a_k = \frac{f^{(k)}(x_0)}{k!}$$

se llama *la serie de Taylor* generada por  $f$  alrededor de  $x_0$ .

**Teorema A.1.** *Suponga que  $f$  es una función de valor real de una variable real que tiene derivadas de todos los órdenes en un intervalo  $(x_0 - r, x_0 + r)$ . Si existe un número real positivo  $M$  (que puede depender de  $x_0$ ) y un número entero positivo  $n_0$  tal que, para todo  $n \geq n_0$ ,*

$$\left| f^{(n)}(x) \right| \leq M, \quad \forall x \in (x_0 - r, x_0 + r),$$

entonces la serie de Taylor de  $f$  definida en torno de  $x_0$  converge para  $f(x)$ , es decir,

$$f(x) = \sum_{j=0}^{\infty} \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j. \quad (\text{A.1})$$

Un ejemplo sencillo se puede consultar en

[www.mat.uson.mx/eduardo/calculo2/soltaylor/soltaylorHTML/taylor.htm](http://www.mat.uson.mx/eduardo/calculo2/soltaylor/soltaylorHTML/taylor.htm)

**Ejemplo A.1.** *Según [5, p. 534], si  $f(x) = e^x$  y  $x_0 = 0$ , encuentre la serie de Taylor generada por  $f(x)$  alrededor de  $x_0 = 0$ .*

*Solución.* Nótese que las derivadas de orden superior de  $f(x)$  satisfacen que

$$f^{(n)}(x_0) = e^{x_0} \quad \text{para} \quad n = 0, 1, 2, \dots$$

luego, la serie de Taylor dada en (A.1) en torno a  $x_0 = 0$  queda:

$$e^x = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n = \sum_{n=0}^{\infty} \frac{x^n}{n!}. \quad (\text{A.2})$$

Esta serie de potencias converge para todo  $x \in \mathbb{R}$ . ☑

**Ejemplo A.2.** Según [5, p. 462], si  $f(x) = (1+x)^n$  para  $n \in \mathbb{N}$  y  $x_0 = 0$ , encuentre la serie de Taylor generada por  $f(x)$  alrededor de  $x_0 = 0$ .

*Solución.* Las primeras derivadas de  $f(x)$  son:

$$\begin{aligned} f^{(0)}(x) &= f(x) = (1+x)^n; & f^{(1)}(x) &= n(1+x)^{n-1} \\ f^{(2)}(x) &= n(n-1)(1+x)^{n-2}; & f^{(3)}(x) &= \frac{n!}{(n-3)!} (1+x)^{n-3}, \end{aligned}$$

nótese que las derivadas de orden superior de  $f(x)$  satisfacen que:

$$f^{(k)}(x_0) = \begin{cases} \frac{n!}{(n-k)!} (1+x_0)^{n-k} & \text{si } k = 0, 1, 2, \dots, n \\ 0 & \text{si } k > n, \end{cases}$$

luego, la serie de Taylor dada en (A.1) en torno a  $x_0 = 0$  queda:

$$(1+x)^n = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n = \sum_{k=0}^n \frac{n!}{k!(n-k)!} x^k = \sum_{k=0}^n \binom{n}{k} x^k. \quad (\text{A.3})$$

Esta serie converge para cualquier entero  $n \geq 0$ . ☑

A partir de la expresión (A.3) se obtiene la fórmula del binomio de Newton [20, p. 90]:

$$(a+b)^n = b^n \underbrace{\left(1 + \frac{a}{b}\right)^n}_x = b^n \sum_{k=0}^n \binom{n}{k} \left(\frac{a}{b}\right)^k = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}. \quad (\text{A.4})$$

**Ejemplo A.3.** Según [5, p. 478], dada  $f(x) = (1-x)^{-n}$  con  $n \in \mathbb{N}$ ,  $x \neq 1$ , y  $x_0 = 0$ , encuentre la serie de Taylor generada por  $f(x)$  alrededor de  $x_0 = 0$ .

*Solución.* Las primeras derivadas de  $f(x)$  son:

$$\begin{aligned} f^{(0)}(x) &= f(x) = (1-x)^{-n}; & f^{(1)}(x) &= n(1-x)^{-n-1} \\ f^{(2)}(x) &= n(n+1)(1-x)^{-n-2}; & f^{(3)}(x) &= \frac{(n+2)!}{(n-1)!} (1-x)^{-n-3}, \end{aligned}$$

nótese que las derivadas de orden superior de  $f(x)$  satisfacen que:

$$f^{(k)}(x_0) = \frac{(n+k-1)!}{(n-1)!} (1-x_0)^{-n-k} \quad \text{si } k = 0, 1, 2, \dots,$$

luego, la serie de Taylor dada en (A.1) en torno a  $x_0 = 0$  queda:

$$(1-x)^{-n} = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n = \sum_{k=0}^{\infty} \frac{(n+k-1)!}{k!(n-1)!} x^k = \sum_{k=0}^{\infty} \binom{n+k-1}{k} x^k. \quad (\text{A.5})$$

Esta expansión se llama serie binomial negativa y converge para  $|x| < 1$ . Para  $n = 1$  se tiene que:

$$(1-x)^{-1} = \sum_{k=0}^{\infty} \frac{(1+k-1)!}{k!(1-1)!} x^k = \sum_{k=0}^{\infty} x^k. \quad (\text{A.6})$$

Esta serie se denomina *serie geométrica*, la cual converge para  $|x| < 1$  y diverge para  $|x| > 1$  (ver [5, p. 475]).  $\checkmark$

**Ejemplo A.4.** Sea  $f(x) = \text{sen}(x)$  y sea  $x_0 = 0$ , encuentre la serie de Taylor generada por  $f(x)$  en  $x_0 = 0$ .

*Solución.* Nótese que las derivadas de orden superior de  $f(x)$  satisfacen que:

$$f^{(j)}(x_0) = \begin{cases} 0 & \text{si } j = 2n \\ (-1)^n & \text{si } j = 2n + 1. \end{cases}$$

En la tabla A.1 se dan los primeros coeficientes de la serie de Taylor:

Tabla A.1. Coeficientes de la serie de Taylor

$j$	0	1	2	3	4	5	6	7	8	9
$a_j x^j$	0	$x$	0	$-\frac{x^3}{6}$	0	$\frac{x^5}{120}$	0	$-\frac{x^7}{7!}$	0	$\frac{x^9}{9!}$

luego, la serie de Taylor dada en (A.1) en torno a  $x_0 = 0$  queda:

$$\text{sen}(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = x - \frac{x^3}{6} + \frac{x^5}{120} + \dots$$

Una ilustración se puede ver en:

<https://youtu.be/g9yu8mLPQ0A?si=63leU2gl0h7s7Krn>

$\checkmark$

**Nota A.1.** En general:

1. Si  $x = x_0$ , la serie anterior converge para  $f(x_0)$ , pero
2. la serie de Taylor NO siempre converge para cualquier punto  $x \neq x_0$ , ver ejemplo A.3.
3. Si la serie converge para algún  $x \neq x_0$ , ella NO necesariamente converge para  $f(x)$ .

**Ejemplo A.5.** Según [56, p. 982] o [23, p. 462], dada la función

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}}, & \text{si } x \neq 0 \\ 0, & \text{si } x = 0 \end{cases}$$

y  $x_0 = 0$ , encontrar la serie de Taylor generada por  $f(x)$  en  $x_0 = 0$ .

*Solución.* Las primeras derivadas de  $f(x)$  son:

$$f^{(0)}(0) = 0$$

$$f^{(1)}(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

$$f^{(1)}(0) = \lim_{h \rightarrow 0} \frac{f(0+h) - f(0)}{h} = \lim_{h \rightarrow 0} \frac{e^{-h^{-2}} - 0}{h} = \lim_{h \rightarrow 0} \frac{1}{h e^{h^{-2}}},$$

como

$$\lim_{h \rightarrow 0} \frac{1}{h} = +\infty \quad \text{y} \quad \lim_{h \rightarrow 0} e^{h^{-2}} = +\infty,$$

se usa la regla de L'Hôpital (ver nota B.9) y

$$f^{(1)}(0) = \lim_{h \rightarrow 0} \frac{-h^{-2}}{-2h^{-3}e^{h^{-2}}} = \lim_{h \rightarrow 0} \frac{h}{2e^{h^{-2}}} = 0,$$

de forma similar se muestra que  $f^{(j)}(0) = 0$  para toda  $j$ . Así la serie de Taylor generada por  $f$  en torno al punto 0 converge en todo el eje real, pero representa a  $f$  solamente en el origen, es decir, cuando  $x = 0$ .  $\square$

## A.1.1. Fórmula de Taylor con resto

La *fórmula de Taylor con resto* es dada por:

$$f(x) = \underbrace{\sum_{j=0}^k \frac{f^{(j)}(x_0)}{j!} (x-x_0)^j}_{P_k(x)} + R_k(x, x_0), \quad (\text{A.7})$$

en que la suma finita es el *polinomio de Taylor* ( $P_k(x)$ ) de grado  $k$  generado por  $f$  en torno del punto  $x_0$  y  $R_k(x, x_0)$  es el error de truncación cometido al aproximar  $f$  con su polinomio de Taylor.

**Teorema A.2.** Según [5, p. 342], dada una función  $f$  con derivada continua de orden  $n + 1$  en un intervalo  $I$  que contiene a  $x_0$ . Entonces, se tiene que:

$$f(x) = \sum_{j=0}^n \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j + R_n(x, x_0) \quad \forall x \in I,$$

donde

$$R_n(x, x_0) = \frac{1}{n!} \int_{x_0}^x (x - t)^n f^{(n+1)}(t) dt. \quad (\text{A.8})$$

*Demostración.* El teorema se demuestra por inducción matemática respecto a  $n$ . Para  $n = 1$  supóngase que  $f$  tiene derivada segunda  $f''$  continua alrededor de  $x_0$ . Entonces, para todo  $x$  en ese entorno, se tiene

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + R_1(x, x_0),$$

que corresponde a la aproximación lineal (o linealización) de la función  $f$  alrededor de  $x_0$ ; luego

$$\begin{aligned} R_1(x, x_0) &= [f(x) - f(x_0)] - f'(x_0)(x - x_0) \\ &= \int_{x_0}^x f'(t) dt - f'(x_0) \int_{x_0}^x dt = \int_{x_0}^x [f'(t) - f'(x_0)] dt; \end{aligned}$$

integrando esta expresión por partes de la siguiente manera:

$$\begin{array}{lll} u = f'(t) - f'(x_0) & \text{(derivamos)} & du = f''(t) dt \\ dv = dt & \text{(integramos)} & v = t - x_0. \end{array}$$

Sustituyendo en la expresión de  $R_1(x, x_0)$  se obtiene

$$\begin{aligned} R_1(x, x_0) &= (t - x_0) [f'(t) - f'(x_0)] \Big|_{x_0}^x - \int_{x_0}^x (t - x_0) f''(t) dt \\ &= \int_{x_0}^x (x - t) f''(t) dt. \end{aligned}$$

Ahora, supóngase que es cierto para un cierto  $m$  y demostremos para  $m + 1$ . Por lo tanto,

$$f(x) = \sum_{j=0}^m \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j + \frac{1}{m!} \int_{x_0}^x (x - t)^m f^{(m+1)}(t) dt.$$

Puesto que:

$$\begin{aligned} R_m(x, x_0) &= \sum_{j=m+1}^{\infty} \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j \\ &= \frac{f^{(m+1)}(x_0)}{(m+1)!} (x - x_0)^{m+1} + \underbrace{\sum_{j=m+2}^{\infty} \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j}_{R_{m+1}(x, x_0)}, \end{aligned}$$

luego,

$$R_{m+1}(x, x_0) = R_m(x, x_0) - \frac{f^{(m+1)}(x_0)}{(m+1)!} (x - x_0)^{m+1}.$$

Usando la expresión integral (A.8) de  $R_m(x, x_0)$  y dado que el factor

$$\frac{(x - x_0)^{m+1}}{m+1} = \int_{x_0}^x (x - t)^m dt$$

se obtiene que

$$\begin{aligned} R_{m+1}(x, x_0) &= \frac{1}{m!} \int_{x_0}^x (x - t)^m f^{(m+1)}(t) dt - \frac{f^{(m+1)}(x_0)}{m!} \int_{x_0}^x (x - t)^m dt \\ &= \frac{1}{m!} \int_{x_0}^x (x - t)^m [f^{(m+1)}(t) - f^{(m+1)}(x_0)] dt, \end{aligned}$$

integrando esta expresión por partes de la siguiente manera:

$$\begin{aligned} u &= f^{(m+1)}(t) - f^{(m+1)}(x_0) && \text{(derivamos)} && du &= f^{(m+2)}(t) dt \\ dv &= (x - t)^m dt && \text{(integramos)} && v &= -\frac{(x - t)^{m+1}}{m+1}. \end{aligned}$$

Sustituyendo en la expresión de  $R_{m+1}(x, x_0)$  se obtiene que:

$$\begin{aligned} R_{m+1}(x, x_0) &= -\frac{(x - t)^{m+1}}{(m+1)!} [f^{(m+1)}(t) - f^{(m+1)}(x_0)] \Big|_{x_0}^x + \\ &\quad \int_{x_0}^x \frac{(x - t)^{m+1}}{(m+1)!} f^{(m+2)}(t) dt \\ &= \frac{1}{(m+1)!} \int_{x_0}^x (x - t)^{m+1} f^{(m+2)}(t) dt. \end{aligned}$$

Por tanto, hemos demostrado que si la relación (A.8) es verdadera para  $m$ , entonces es verdadera para  $m + 1$ .  $\square$

**Corolario A.1.** Usando la notación  $o(\cdot)$  (ver apartado 3.3.1), la fórmula de Taylor dada en (A.7) se puede expresar como [5, p. 351],

$$f(x) = f(x_0) + \sum_{j=1}^k \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j + \underbrace{o((x - x_0)^k)}_{R_k(x, x_0)} \quad \text{cuando } x \rightarrow x_0.$$

*Demostración.* Por la definición 3.5 se sabe que  $R_k(x, x_0) = o((x - x_0)^k)$ , si  $\forall \varepsilon > 0, \exists n_0 = n_0(\varepsilon)$  tal que

$$\left| \frac{R_k(x, x_0)}{(x - x_0)^k} \right| < \varepsilon, \quad \forall n \geq n_0.$$

De la expresión (A.7), se tiene que para todo  $x \neq x_0$ :

$$\frac{R_k(x, x_0)}{(x - x_0)^k} = \frac{f(x) - f(x_0)}{(x - x_0)^k} - \sum_{j=1}^k \frac{f^{(j)}(x_0)}{j!} \frac{(x - x_0)^j}{(x - x_0)^k},$$

tomando límite cuando  $x \rightarrow x_0$ :

$$\lim_{x \rightarrow x_0} \frac{R_k(x, x_0)}{(x - x_0)^k} = \lim_{x \rightarrow x_0} \left[ \frac{f(x) - f(x_0)}{(x - x_0)^k} - \sum_{j=1}^k \frac{f^{(j)}(x_0)}{j!} \frac{(x - x_0)^j}{(x - x_0)^k} \right],$$

esto resulta indeterminado, entonces aplicando regla de L'Hôpital:

$$\lim_{x \rightarrow x_0} \frac{R_k(x, x_0)}{(x - x_0)^k} = \lim_{x \rightarrow x_0} \left[ \frac{f'(x) - f'(x_0)}{k(x - x_0)^{k-1}} - \sum_{j=2}^k \frac{f^{(j)}(x_0)}{j!} \frac{j(x - x_0)^{j-1}}{k(x - x_0)^{k-1}} \right],$$

aquí se reorganizaron los términos de la serie. Aplicando sucesivamente la regla de L'Hôpital hasta la derivada de orden  $m$  y teniendo en cuenta que

$$\frac{d^m}{du^m} u^k = \frac{k!}{(k - m)!} u^{k-m} \quad m < k,$$

se llega a

$$\begin{aligned} \lim_{x \rightarrow x_0} \frac{R_k(x, x_0)}{(x - x_0)^k} &= \lim_{x \rightarrow x_0} \left[ \frac{f^{(m)}(x) - f^{(m)}(x_0)}{\frac{k!}{(k-m)!} (x - x_0)^{k-m}} \right. \\ &\quad \left. - \sum_{j=m+1}^k \frac{f^{(j)}(x_0)}{k!} \frac{(k-m)! (x - x_0)^{j-m}}{(j-m)! (x - x_0)^{k-m}} \right], \end{aligned}$$

nótese que si se aplica regla de L'Hôpital hasta  $m = k - 1$ , se tiene que:

$$\begin{aligned} \lim_{x \rightarrow x_0} \frac{R_k(x, x_0)}{(x - x_0)^k} &= \lim_{x \rightarrow x_0} \left[ \frac{f^{(k-1)}(x) - f^{(k-1)}(x_0)}{k!(x - x_0)} - \frac{f^{(k)}(x_0)}{k!} \right] \\ &= \lim_{x \rightarrow x_0} \frac{1}{k!} \left[ \frac{f^{(k-1)}(x) - f^{(k-1)}(x_0)}{x - x_0} \right] - \frac{f^{(k)}(x_0)}{k!} \\ &= \frac{1}{k!} [f^{(k)}(x_0) - f^{(k)}(x_0)] = 0, \end{aligned}$$

luego  $R_k(x, x_0) \xrightarrow{x \rightarrow x_0} 0$  más rápido que  $(x - x_0)^k \xrightarrow{x \rightarrow x_0} 0$ . □

Puesto que en la expresión (A.8) que expresa el error como una integral, el factor  $(x - t)^n$  del integrando nunca cambia de signo en el intervalo de integración, y  $f^{(n+1)}(t)$  es continua en este intervalo. Por el *teorema del valor medio ponderado para integrales* dado por [5, p. 189], se tiene que:

$$\begin{aligned} \int_{x_0}^x (x - t)^n f^{(n+1)}(t) dt &= f^{(n+1)}(c) \int_{x_0}^x (x - t)^n dt \\ &= f^{(n+1)}(c) \frac{(x - x_0)^{n+1}}{n + 1}, \end{aligned}$$

para algún  $x_0 < c < x$ . Por consiguiente, la expresión (A.8) se puede expresar en la forma del residuo de Lagrange (ver [5, p. 347]):

$$R_n(x, x_0) = \frac{f^{(n+1)}(c)}{(n + 1)!} (x - x_0)^{n+1}.$$

Sea  $f(x) : \mathbb{R} \rightarrow \mathbb{R}$ ,  $a \leq x \leq b$ , una función continua y  $(n + 1)$  diferenciable para algún  $n \geq 0$ . Entonces, para cada  $x \in [a, b]$ ,  $x_0 \in (a, b)$ ,

$$f(x) = f(x_0) + \sum_{j=1}^n \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j + \frac{f^{(n+1)}(c)}{(n + 1)!} (x - x_0)^{n+1} \quad (\text{A.9})$$

para algún  $x_0 < c < x$ . El punto  $c$  depende de  $x$ , de  $n$  y de  $f$  (ver [82, p. 17]).

El error  $R_n(x, x_0)$  se puede expresar de otras formas, por ejemplo, la forma de Cauchy del resto es dada por:

$$R_n(x, x_0) = \frac{f^{(n+1)}(c)}{n!} (x - c)^n (x - x_0)$$

para algún  $x_0 < c < x$ .

Tabla A.2. Fórmulas del resto

Forma	$R_n(x, x_0)$	Región
Integral	$\frac{1}{n!} \int_{x_0}^x (x-t)^n f^{(n+1)}(t) dt$	$x > x_0$
Lagrange	$\frac{f^{(n+1)}(c)}{(n+1)!} (x-x_0)^{n+1}$	$x_0 < c < x$
Cauchy	$\frac{f^{(n+1)}(c)}{n!} (x-c)^n (x-x_0)$	$x_0 < c < x$

## A.2. Ejercicios

Obtenga la serie de Taylor para las siguientes funciones en torno al número dado [56, p. 994].

1.  $f(x) = \text{sen}(3x)$ ; en  $x_0 = \frac{-\pi}{3}$ .
2.  $f(x) = \frac{1}{x}$ ; en  $x_0 = 2$ .
3.  $f(x) = \ln|x|$ ; en  $x_0 = -1$ .
4.  $f(x) = e^{x-2}$ ; en  $x_0 = 2$ .



Apéndice

B

Tópicos especiales  
de matemáticas



En este apéndice aparece un compendio de distintas temáticas, que son utilizadas a lo largo del texto, las cuales se clasificaron para una mayor comprensión.

## B.1. Tópicos de álgebra matricial

**Nota B.1.** El espacio fila de  $\mathbf{A}_{m \times n}$ , denotado por  $\mathcal{R}(\mathbf{A})$ , es el conjunto de todos los vectores fila  $n$ -dimensionales que se pueden expresar como combinaciones lineales de las  $m$  filas de  $\mathbf{A}$ .

$$x_1 \mathbf{a}_1^\top + x_2 \mathbf{a}_2^\top + \cdots + x_m \mathbf{a}_m^\top = \mathbf{x}^\top \mathbf{A},$$

con  $x_i$  escalar y  $\mathbf{a}_i^\top$  la  $i$ -ésima fila de  $\mathbf{A}$ . Similarmente, el espacio columna de  $\mathbf{A}_{m \times n}$ , denotado por  $\mathcal{C}(\mathbf{A})$ , es el conjunto de todos los vectores columna  $m$ -dimensionales tales que se pueden expresar como combinaciones lineales de las  $n$  columnas de  $\mathbf{A}$ .

$$x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \cdots + x_n \mathbf{a}_n = \mathbf{A} \mathbf{x},$$

con  $x_i$  escalar y  $\mathbf{a}_i$  la  $i$ -ésima columna de  $\mathbf{A}$ .

**DEFINICIÓN B.1.** Si  $\mathbf{a} = (a_1, a_2, \dots, a_n)^\top$  es un vector en  $\mathbb{R}^n$ , su longitud o norma se designa con  $\|\mathbf{a}\|$  y se define como (ver [5, p. 554]):

$$\|\mathbf{a}\| = \sqrt{\mathbf{a}^\top \mathbf{a}} = \left( \sum_{j=1}^n a_j^2 \right)^{\frac{1}{2}}. \quad (\text{B.1})$$

**Ejemplo B.1.** Hallar la norma de los siguientes vectores de  $\mathbb{R}^3$ :

$$\mathbf{a} = \begin{pmatrix} 0,5 \\ 0,25 \\ 0,5 \end{pmatrix}; \quad \mathbf{b} = \begin{pmatrix} 0,1 \\ 0,01 \\ 0,9 \end{pmatrix}; \quad \mathbf{c} = \begin{pmatrix} 0,01 \\ 0,0001 \\ 0,99 \end{pmatrix}$$

*Solución.* Usando la expresión (B.1) se tiene que:

$$\|\mathbf{a}\| = 0,750; \quad \|\mathbf{b}\| = 0,906; \quad \|\mathbf{c}\| = 0,990. \quad \checkmark$$

**DEFINICIÓN B.2.** De acuerdo con [39, p. 40], dado un vector  $\mathbf{X}$  y las matrices  $\mathbf{A} = \{a_{ij}\}$  y  $\mathbf{B} = \{b_{kl}\}$  de dimensión  $n \times n$  y  $t \times n$ , respectivamente, donde los elementos de  $\mathbf{A}$  y  $\mathbf{B}$  son constantes conocidos y  $\mathbf{A}$  es simétrica, entonces:

1.  $\mathbf{Y} = \mathbf{B}\mathbf{X}$  es llamada una forma lineal en  $\mathbf{X}$  y
2.  $\mathbf{Q}(\mathbf{X}) = \mathbf{X}^\top \mathbf{A}\mathbf{X}$  es llamada una forma cuadrática en  $\mathbf{X}$ .

**Ejemplo B.2.** Expresar  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  como una forma lineal.

*Solución.* Sea  $\mathbf{1}_n$  un vector columna de unos, entonces:

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n} \implies \bar{X} = \frac{1}{n} \mathbf{1}_n^\top \mathbf{X}$$

donde

$$\mathbf{B} = \frac{1}{n} \mathbf{1}_n^\top. \quad \square$$

**Ejemplo B.3.** Represente  $S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  como una forma cuadrática.

*Solución.* Teniendo en cuenta la representación de  $\bar{X}$ , se tiene que:

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \implies \mathbf{A} = \frac{1}{n-1} \left( \mathbf{I} - \frac{1}{n} \mathbf{J}_n \right)$$

$$S_X^2 = \mathbf{X}^\top \left[ \frac{1}{n-1} \left( \mathbf{I} - \frac{1}{n} \mathbf{J}_n \right) \right] \mathbf{X} = \frac{1}{n-1} \left( \mathbf{X}^\top \mathbf{X} - \frac{1}{n} \mathbf{X}^\top \mathbf{J}_n \mathbf{X} \right),$$

donde  $\mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n^\top$ . Luego,

$$S_X^2 = \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - \frac{1}{n} \left[ \sum_{i=1}^n X_i \right]^2 \right) = \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n \bar{X}^2 \right). \quad \square$$

**Nota B.2.** Si  $\mathbf{A}$  no es simétrica, es decir,  $a_{ij} \neq a_{ji}$ , siempre se puede expresar  $Q(\mathbf{X}) = \mathbf{X}^\top \mathbf{A} \mathbf{X}$  usando la matriz simétrica:

$$a_{ij}^* = a_{ji}^* = \frac{1}{2} (a_{ij} + a_{ji}).$$

### B.1.1. Transformación ortogonal

La transformación lineal  $\mathbf{Y} = \mathbf{Q}^\top \mathbf{X}$ , donde  $\mathbf{Q}$  es una matriz ortogonal, se llama *transformación ortogonal*. En este apartado, se interpreta geoméricamente este tipo de transformaciones en el plano cartesiano.

**Teorema B.1. Rotación de ejes en dos dimensiones** Sean  $\mathbf{X}$  y  $\mathbf{Y}$  dos sistemas coordenados y  $\mathfrak{B} = \{\mathbf{e}_1, \mathbf{e}_2\}$  y  $\mathfrak{B}' = \{\mathbf{e}'_1, \mathbf{e}'_2\}$  las bases correspondientes de estos sistemas. Si

$$\mathbf{e}'_j = a_{1j} \mathbf{e}_1 + a_{2j} \mathbf{e}_2, \quad \text{para } j = 1, 2,$$

sea un punto cualquiera  $P$  con coordenadas  $(x_1, x_2)$  en el sistema  $\mathbf{X}$  y con coordenadas  $(y_1, y_2)$  en el sistema  $\mathbf{Y}$ , entonces este par de coordenadas están relacionadas como sigue:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{A} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix},$$

donde  $\mathbf{A} = [a_{ij}]$  es la matriz de cambio de base (o de transición) de  $\mathfrak{B}$  a  $\mathfrak{B}'$ .

*Demostración.* Ver detalles de la prueba en [44, p. 290]. □

### B.1.1.1. Rotación de ejes en el plano manteniendo el mismo origen

Considérese un sistema de coordenadas rectangulares  $\tilde{\mathbf{X}} = \mathbf{X} - \boldsymbol{\mu}$  (en  $\mathbb{R}^2$ ) que tiene unidades iguales sobre ambos ejes, es decir, si  $\mathbf{e}_1$  y  $\mathbf{e}_2$  son los vectores base (unitarios), entonces estos son perpendiculares entre sí.

Supóngase que los ejes del sistema de coordenadas  $\mathbf{Y}$  se obtienen, conservando la ortogonalidad, rotando mediante un ángulo de giro  $\theta$  el sistema  $\tilde{\mathbf{X}}$ , en sentido contrario al movimiento de las manecillas del reloj.

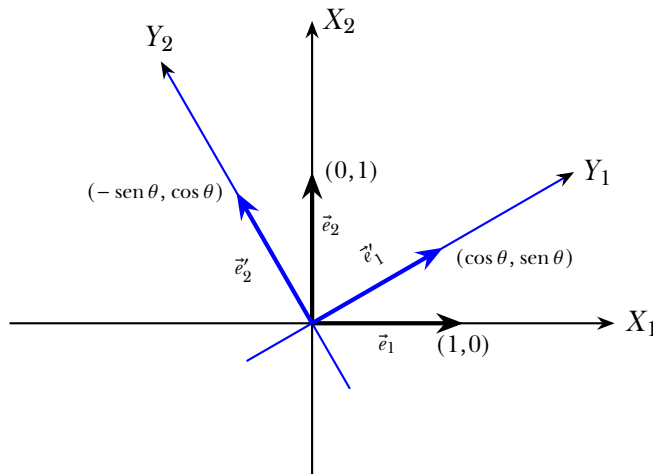


Figura B.1. Rotación de ejes en  $\mathbb{R}^2$

Fuente: tomada de [44, p. 291]

Nótese que los vectores base  $\mathbf{e}'_1$  y  $\mathbf{e}'_2$  del sistema  $\mathbf{Y}$ , forman también una base ortonormal y están dados por

$$\mathbf{e}'_1 = [\cos \theta, \sen \theta] = \cos \theta \mathbf{e}_1 + \sen \theta \mathbf{e}_2.$$

$$e_2' = [-\operatorname{sen} \theta, \operatorname{cos} \theta] = -\operatorname{sen} \theta e_1 + \operatorname{cos} \theta e_2.$$

Luego, por el teorema B.1, se deduce que las coordenadas de un punto en ambos sistemas están relacionadas por:

$$\tilde{x}_1 = y_1 \operatorname{cos} \theta - y_2 \operatorname{sen} \theta \quad \text{y} \quad \tilde{x}_2 = y_1 \operatorname{sen} \theta + y_2 \operatorname{cos} \theta, \quad (\text{B.2})$$

que corresponde a las ecuaciones de una rotación levógira de ejes, cuando el ángulo de giro es  $\theta$ . Si se denota por  $\mathbf{R}_\theta$  la matriz ortogonal,

$$\mathbf{R}_\theta = \begin{bmatrix} \operatorname{cos} \theta & -\operatorname{sen} \theta \\ \operatorname{sen} \theta & \operatorname{cos} \theta \end{bmatrix}, \quad (\text{B.3})$$

entonces la expresión (B.2) se puede expresar matricialmente como

$$\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} = \mathbf{R}_\theta \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad (\text{B.4})$$

esta es una transformación ortogonal propia, puesto que el  $\det(\mathbf{R}_\theta) = 1$  y representa una rotación de ejes en  $\mathbb{R}^2$ .

El siguiente teorema permite obtener el ángulo de rotación  $\theta$  de una ecuación cuadrática de segundo grado.

**Teorema B.2.** Sea  $Q : \mathbb{R}^2 \rightarrow \mathbb{R}$  cuya expresión matricial es:

$$x^\top \mathbf{A} x + 2\mathbf{k}^\top x + f = 0, \quad (\text{B.5})$$

donde

$$\mathbf{A} = \begin{bmatrix} a & \frac{c}{2} \\ \frac{c}{2} & b \end{bmatrix} \quad \text{y} \quad \mathbf{k}^\top = \left[ \frac{d}{2} \quad \frac{e}{2} \right].$$

Entonces existe un número único  $\theta$ , llamado ángulo de rotación, tal que

$$\tan(2\theta) = \frac{c}{a-b}, \quad \theta \in \left[ 0, \frac{\pi}{2} \right). \quad (\text{B.6})$$

*Demostración.* Ver detalles de la prueba en [44, p. 292].  $\square$

**Nota B.3.** De acuerdo con [44, p. 294], el ángulo de rotación se establece como

$$\tan \theta = \frac{1}{c} \left[ b - a \pm \sqrt{\operatorname{tr}^2(\mathbf{A}) - 4 \det(\mathbf{A})} \right], \quad (\text{B.7})$$

donde  $\operatorname{tr}(\cdot)$  denota la traza y  $\det(\cdot)$  el determinante.

## B.1.2. Forma cuadrática

De acuerdo con [38, p. 210], si  $Q(x) = x^T A x$  representa una forma cuadrática arbitraria [en un vector  $n$ -dimensional  $x = (x_1, x_2, \dots, x_n)^T$ ], entonces,

- La forma cuadrática  $Q(x)$  se dice *definida no negativa* si  $Q(x) \geq 0$  para todo  $x \in \mathbb{R}^n$ . En este caso,  $A_{n \times n}$  se dice que es *definida no negativa*.
- Si  $Q(x) > 0$  para todo  $x$  excepto cuando  $x = \mathbf{0}$ , se dice que  $Q(x)$  es *definida positiva*. En este caso,  $A_{n \times n}$  se dice que es *definida positiva*.
- $Q(x)$  es *semidefinida positiva* si  $Q(x) \geq 0$  para todo  $x \in \mathbb{R}^n$  y  $Q(x) = 0$  para algún  $x \neq \mathbf{0}$ . En este caso,  $A_{n \times n}$  se dice que es *semidefinida positiva*.
- $Q(x)$  se dice *definida no positiva* si  $-Q(x)$  es definida no negativa. En este caso,  $A_{n \times n}$  se dice que es *definida no positiva*.
- $Q(x)$  se dice *definida negativa* si  $-Q(x)$  es definida positiva. En este caso,  $A_{n \times n}$  se dice que es *definida negativa*.
- $Q(x)$  se dice *semidefinida negativa* si  $-Q(x)$  es semidefinida positiva. En este caso,  $A_{n \times n}$  se dice que es *semidefinida negativa*.
- $Q(x)$  se dice *indefinida* si  $Q(x) < 0$  para algún  $x$  y  $Q(x) > 0$  para algún (otro)  $x$ . En este caso,  $A_{n \times n}$  se dice que es *indefinida*.

**Nota B.4.** De acuerdo con [38, p. 581], dada  $A$  una matriz simétrica de orden  $n \times n$ , y  $B$  una matriz simétrica definida positiva de orden  $n \times n$ . Si  $\lambda_{\min}$  y  $\lambda_{\max}$  son los valores propios mínimo y máximo respectivamente de  $AB^{-1}$ , entonces:

$$\lambda_{\min} \leq \frac{x^T A x}{x^T B x} \leq \lambda_{\max}$$

para cada vector  $x \in \mathbb{R}^n$ .

## B.1.3. Inversa generalizada

Una inversa generalizada de una matriz  $A_{p \times q}$ , digamos  $G$ , satisface que (ver [38, p. 107]):

$$AGA = A.$$

Un método para obtener  $G$  dada una matriz  $A_{p \times q}$  de rango  $r$  es:

1. Obtener una submatriz  $A_{11}$  de rango  $r$

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

2. Obtener  $A_{11}^{-1}$ .
3. Substituir en  $A^T$  los elementos de  $A_{11}^T$  por sus correspondientes  $A_{11}^{-1}$ .

4. Hacer todos los otros elementos iguales a cero.

**Ejemplo B.4.** Encontrar la inversa generalizada de la matriz

$$\mathbf{A}_{3 \times 4} = \begin{pmatrix} 1 & 2 & 4 & 3 \\ 3 & -1 & 2 & -2 \\ 5 & -4 & 0 & -7 \end{pmatrix}$$

*Solución.* Se puede verificar que el rango de la matriz  $\mathbf{A}$  es 2.

1. Obtener una submatriz  $\mathbf{A}_{11}$  de rango 2:  $\mathbf{A}_{11} = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}$ .
2. Sustituir en  $\mathbf{A}^\top$  los elementos de  $\mathbf{A}_{11}^\top$  por sus correspondientes  $\mathbf{A}_{11}^{-1}$  y hacer todos los otros elementos iguales a cero:

$$\mathbf{A}_{11}^{-1} = \frac{1}{-7} \begin{pmatrix} -1 & -2 \\ -3 & 1 \end{pmatrix} \quad \Rightarrow \quad \mathbf{G} = \begin{pmatrix} 1/7 & 2/7 & 0 \\ 3/7 & -1/7 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad \checkmark$$

**Nota B.5.** Para cualquier matriz idempotente  $\mathbf{A}$ ,  $\text{rango}(\mathbf{A}) = \text{tr}(\mathbf{A})$ . Sea  $\mathbf{A}$  una matriz de orden  $m \times n$  y  $\mathbf{B}$  una matriz  $n \times m$ . Entonces,  $\mathbf{B}$  es una inversa generalizada de  $\mathbf{A}$  si y solo si  $\mathbf{BA}$  es idempotente y  $\text{rango}(\mathbf{BA}) = \text{rango}(\mathbf{A})$  [38, p. 135].

**Teorema B.3.** La matriz  $\mathbf{X}^\top \mathbf{X}$ : si  $\mathbf{G}$  es una inversa generalizada de  $\mathbf{X}^\top \mathbf{X}$ , se satisfacen las siguientes propiedades:

1.  $\mathbf{G}^\top$  también es una inversa generalizada de  $\mathbf{X}^\top \mathbf{X}$ .
2.  $\mathbf{XGX}^\top \mathbf{X} = \mathbf{X}$ , es decir,  $\mathbf{GX}^\top$  es una inversa generalizada de  $\mathbf{X}$ .
3.  $\mathbf{XGX}^\top$  es invariante a  $\mathbf{G}$ .
4.  $\mathbf{XGX}^\top$  es simétrica.

*Demostración.*

1. Se sabe que  $\mathbf{X}^\top \mathbf{XGX}^\top \mathbf{X} = \mathbf{X}^\top \mathbf{X}$ , luego

$$\begin{aligned} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} &= \mathbf{X}^\top \mathbf{X} \\ [\mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}]^\top &= (\mathbf{X}^\top \mathbf{X})^\top \\ \mathbf{X}^\top \mathbf{X} [(\mathbf{X}^\top \mathbf{X})^{-1}]^\top \mathbf{X}^\top \mathbf{X} &= \mathbf{X}^\top \mathbf{X}. \end{aligned}$$

Entonces  $[(\mathbf{X}^\top \mathbf{X})^{-1}]^\top = \mathbf{G}^\top$  es una inversa generalizada de  $\mathbf{X}^\top \mathbf{X}$ .

2. Teniendo en cuenta que (ver [38, p. 52]),

- Para cualquier matriz  $A_{m \times n}$ , se puede probar que  $A = \mathbf{0}$  si y solo si  $A^T A = \mathbf{0}$ , y que
- Para cualquier matrices  $A_{m \times n}$ ,  $B_{n \times p}$  y  $C_{n \times p}$ , se puede probar que  $AB = AC$  si y solo si  $A^T AB = A^T AC$ ,

entonces

$$\underbrace{X^T}_{A^T} \underbrace{X}_A \underbrace{(X^T X)^{-1} X^T X}_B = \underbrace{X^T}_{A^T} \underbrace{X}_A \underbrace{I}_C$$

$$X \underbrace{(X^T X)^{-1} X^T X}_G = X,$$

es decir,  $(X^T X)^{-1} X^T$  es una inversa generalizada de  $X$ .

3. Usando el resultado de [38, p. 119]: sean las matrices  $A_{m \times n}$ ,  $B_{p \times n}$  y  $C_{m \times q}$ . Si  $\mathcal{R}(B) \subset \mathcal{R}(A)$  y  $\mathcal{C}(C) \subset \mathcal{C}(A)$  (ver nota B.1), se tiene que  $BA^{-1}C$  es invariante a la escogencia de la inversa generalizada  $A^{-1}$ , sea

$$BA^{-1}C = \underbrace{B}_B \underbrace{(X^T X)^{-1} X^T}_{A^{-1}} \underbrace{C}_C,$$

luego, se necesita encontrar  $L$  y  $R$  tal que

$$B = LA \quad \text{y} \quad C = AR,$$

es decir,

$$X = LX^T X \quad \text{y} \quad X^T = X^T X R,$$

pero del punto 2 se tiene que:

$$X(X^T X)^{-1} X^T X = X \quad \Rightarrow \quad L = X(X^T X)^{-1}$$

$$X^T X [(X^T X)^{-1}]^T X^T = X^T \quad \Rightarrow \quad R = [X(X^T X)^{-1}]^T.$$

Así,  $X(X^T X)^{-1} X^T$  es invariante a  $(X^T X)^{-1}$ .

4. Usando el hecho de que para cualquier matriz  $A$ ,  $(A^{-1})^T$  es una inversa generalizada de  $A^T$  [38, p. 117]. Luego, si  $A^{-1} = (X^T X)^{-1}$  es la inversa generalizada de  $A = X^T X$ , entonces  $(A^{-1})^T = [(X^T X)^{-1}]^T$  es una inversa generalizada de  $(X^T X)^T = X^T X$ , y

$$[X(X^T X)^{-1} X^T]^T = X [(X^T X)^{-1}]^T X^T$$

la cual, por el punto 3, es invariante a  $G$ , entonces la matriz

$$\mathbf{X} [(\mathbf{X}^\top \mathbf{X})^{-}]^\top \mathbf{X}^\top = \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-} \mathbf{X}^\top. \quad \checkmark$$

**Teorema B.4.** Sea  $\mathbf{A}^-$  una inversa generalizada de  $\mathbf{A}$ . Entonces,

- i) La ecuación  $\mathbf{A}\mathbf{x} = \mathbf{d}$  es consistente si  $\mathbf{A}\mathbf{A}^-\mathbf{d} = \mathbf{d}$ .
- ii) Cualquier solución es de la forma

$$\mathbf{x} = \mathbf{A}^-\mathbf{d} + (\mathbf{I} - \mathbf{A}\mathbf{A}^-) \mathbf{z}$$

para algún  $\mathbf{z}$ .

**Nota B.6.** Retomando la ecuación (4.31), si  $\mathbf{X}^\top \mathbf{X}$  es singular,

$$\underbrace{\mathbf{X}^\top \mathbf{X}}_{\mathbf{A}} \underbrace{\hat{\boldsymbol{\beta}}}_x = \underbrace{\mathbf{X}^\top \mathbf{y}}_d,$$

veamos si satisface las condiciones del teorema B.4:

- i) Multiplicando a ambos por  $\mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-}$ , se tiene que:

$$\mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-} \mathbf{X}^\top \mathbf{y} = \mathbf{X}^\top \mathbf{y} \mathbf{X}^\top \mathbf{y} = \mathbf{d},$$

aquí se usó la inversa generalizada de  $\mathbf{X}^\top$ , luego  $\mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{y}$  es una ecuación consistente.

- ii) Se tiene una solución para  $\boldsymbol{\beta}$  y es dada por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-} \mathbf{X}^\top \mathbf{y} + [\mathbf{I} - \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-}] \mathbf{z}$$

para algún  $\mathbf{z}$ .

## B.2. Tópicos de cálculo

**Nota B.7.** (Fórmula cuadrática). Una ecuación  $ax^2 + bx + c = 0$ ,  $a \neq 0$ , con coeficientes reales o complejos tiene dos soluciones dadas por:  $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ .

**Nota B.8.** (Cotas superior e inferior). Según [5, p. 29], si  $S$  es un conjunto no vacío de números reales y existe un número  $B$  tal que:

1. Para todo  $x$  de  $S$ ,  $x \leq B$ , entonces se dice que  $S$  está acotado superiormente por  $B$ . Al número  $B$  se le denomina **cota superior** para  $S$ .

2. Para todo  $x$  de  $S$ ,  $x \geq B$ . entonces se dice que  $S$  está acotado inferiormente por  $B$ . Al número  $B$  se le denomina **cota inferior** para  $S$ .

De acuerdo con [5, p. 31], un número  $B$  se llama:

1. Extremo superior (o **supremo**) de  $S$  si:

- a)  $B$  es una cota superior para  $S$ , y  
b) ningún número menor que  $B$  es cota superior para  $S$ .

El extremo superior de  $S$ , cuando existe, es único y se denota por  $\sup S$ .

2. Extremo inferior (o **ínfimo**) de  $S$  si

- a)  $B$  es una cota inferior para  $S$ .  
b) Ningún número mayor que  $B$  es cota inferior para  $S$ .

El extremo inferior de  $S$ , cuando existe, es único y se denota por  $\inf S$ .

**Nota B.9.** (Regla de L'Hôpital). Sean  $f$  y  $g$  funciones diferenciables en un intervalo abierto  $I$ , excepto posiblemente en el número  $a$  en  $I$ , y suponiendo que para toda  $x \neq a$  en  $I$ ,  $g(x) \neq 0$ . Entonces, si  $\lim_{x \rightarrow a} f(x) = +\infty$  o  $-\infty$  y  $\lim_{x \rightarrow a} g(x) = +\infty$  o  $-\infty$  y si (ver [56, p. 853]):

$$\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)} = L \quad \text{entonces} \quad \lim_{x \rightarrow a} \frac{f(x)}{g(x)} = L.$$

Esto es válido si todos los límites son por la derecha o por la izquierda del punto  $a$ .

**Nota B.10.** (Gradiente). Si  $f(\mathbf{t})$  es una función real de variable múltiple, es decir,  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , y las primeras derivadas parciales de  $f$  con respecto a las variables  $t_i$  existen, entonces el gradiente de  $f$  se define como [56, p. 982]:

$$\nabla f(\mathbf{t}) = \left( \frac{\partial f(\mathbf{t})}{\partial t_1}, \frac{\partial f(\mathbf{t})}{\partial t_2}, \dots, \frac{\partial f(\mathbf{t})}{\partial t_n} \right)^\top.$$

**Nota B.11.** (Primer teorema fundamental del cálculo). Sea la función  $f$  continua en el intervalo cerrado  $[a, b]$  y sea  $x$  cualquier número en  $[a, b]$ . Si  $F$  es la función definida por:

$$F(x) = \int_a^x f(t) dt \quad \Rightarrow \quad \frac{dF(x)}{dx} = f(x),$$

si  $x = a$ , la derivada puede ser una derivada por derecha, y si  $x = b$ , la derivada puede ser una derivada por izquierda (ver [56, p. 450]).

**Nota B.12.** (Regla de Leibniz). Si  $f(x, \theta)$ ,  $a(\theta)$  y  $b(\theta)$  son diferenciables con respecto a  $\theta$ , entonces (ver [20, p. 69]):

$$\frac{\partial}{\partial \theta} \int_{a(\theta)}^{b(\theta)} f(x, \theta) dx = f(b(\theta), \theta) \frac{db(\theta)}{d\theta} - f(a(\theta), \theta) \frac{da(\theta)}{d\theta} + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial \theta} f(x, \theta) dx.$$

## B.2.1. Máximo o mínimo local

I. **Caso univariado:** una función  $f$  tiene un *valor máximo* en  $c$  si existe un intervalo abierto que contenga a  $c$ , en el cual  $f$  esté definida, tal que  $f(c) \geq f(x)$  para toda  $x$  en este intervalo.

Si  $f(x)$  existe para todos los valores de  $x$  en el intervalo abierto  $(a, b)$  y si  $f$  tiene un extremo relativo en  $c$ , donde  $a < c < b$ , entonces  $\frac{\partial f(x)}{\partial x} = f'(x)$  existe y  $f'(c) = 0$ .

Se dice que  $f$  tiene un *valor máximo absoluto* en un intervalo si existe algún número  $c$  en el intervalo, tal que  $f(c) \geq f(x)$  para toda  $x$  en el mismo. En tal caso,  $f(c)$  es el valor máximo absoluto de  $f$  en el intervalo.

Se dice que  $f(c)$  es el *valor máximo absoluto* de la función  $f$  si  $c$  está en el dominio de  $f$  y si  $f(c) \geq f(x)$  para todos los valores de  $x$  en el dominio de  $f$  (ver [56, p. 285]).

II. **Caso bivariado:** una función  $f(x, y)$  tiene un máximo o mínimo local en  $\mathbf{c} = (x_0, y_0)$  si se verifican las siguientes condiciones:

- 1) Las derivadas parciales de primer orden evaluadas en  $(x_0, y_0)$  son 0, es decir,  $\nabla f(x_0, y_0) = (0, 0)$ .
- 2) La matriz de segundas derivadas (hessiana), introducida por Ludwig Otto Hesse,

$$\mathbf{H}_f(x, y) = \begin{pmatrix} \frac{\partial^2}{\partial x^2} f(x, y) & \frac{\partial^2}{\partial x \partial y} f(x, y) \\ \frac{\partial^2}{\partial y \partial x} f(x, y) & \frac{\partial^2}{\partial y^2} f(x, y) \end{pmatrix} = \begin{pmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{xy}(x, y) & f_{yy}(x, y) \end{pmatrix},$$

se evalúa en el punto  $(x_0, y_0)$  y se establecen los valores propios,  $\lambda$ , de esta matriz [6, p. 379]:

$$\begin{aligned} 2\lambda &= \text{tr}(\mathbf{H}_f(x_0, y_0)) \pm \sqrt{[\text{tr}(\mathbf{H}_f(x_0, y_0))]^2 - 4 \det(\mathbf{H}_f(x_0, y_0))} \\ &= f_{xx}(x_0, y_0) + f_{yy}(x_0, y_0) \pm \\ &\quad \sqrt{(f_{xx}(x_0, y_0) - f_{yy}(x_0, y_0))^2 + 4f_{xy}^2(x_0, y_0)} \end{aligned}$$

- a) Si  $\lambda_1$  y  $\lambda_2$  son positivos, entonces la función  $f(x, y)$  tiene un mínimo relativo en  $(x_0, y_0)$ .
- b) Si  $\lambda_1$  y  $\lambda_2$  son negativos, entonces la función  $f(x, y)$  tiene un máximo relativo en  $(x_0, y_0)$ .
- c) Si  $\lambda_1$  y  $\lambda_2$  tienen signos contrarios, entonces la función  $f(x, y)$  tiene un punto de silla en  $(x_0, y_0)$ .

**Ejemplo B.5.** Determine y clasifique los puntos críticos de la función

$$f(x, y) = x^4 + y^4 - 2(x - y)^2.$$

*Solución.* Primero se determina el gradiente y se iguala a cero

$$\nabla f(x, y) = (4x^3 - 4(x - y), 4y^3 + 4(x - y)) = (0, 0).$$

A partir de estas ecuaciones se tiene que

$$\begin{aligned} 4x^3 - 4(x - y) &= 0 \\ 4y^3 + 4(x - y) &= 0 \end{aligned} \quad \text{sumando } x^3 + y^3 = (x + y)(x^2 - xy + y^2) = 0$$

y la solución es  $y = -x$ , reemplazando en la primera ecuación se tiene

$$4(x^3 - 2x) = 4x(x^2 - 2) = 4x(x - \sqrt{2})(x + \sqrt{2}) = 0.$$

Luego, los puntos críticos son:

$$a_1 = (0, 0), \quad a_2 = (\sqrt{2}, -\sqrt{2}) \quad \text{y} \quad a_3 = (-\sqrt{2}, \sqrt{2}).$$

Se encuentran las segundas derivadas y se construye la matriz hessiana

$$\mathbf{H}_f(x, y) = \begin{pmatrix} \frac{\partial^2}{\partial x^2} f(x, y) & \frac{\partial^2}{\partial x \partial y} f(x, y) \\ \frac{\partial^2}{\partial y \partial x} f(x, y) & \frac{\partial^2}{\partial y^2} f(x, y) \end{pmatrix} = \begin{pmatrix} 12x^2 - 4 & 4 \\ 4 & 12y^2 - 4 \end{pmatrix}.$$

- i) Se sustituye en  $\mathbf{H}_f(x, y)$  el punto crítico  $a_1 = (0, 0)$

$$\mathbf{H}_f(0, 0) = \begin{pmatrix} -4 & 4 \\ 4 & -4 \end{pmatrix}$$

cuyos valores propios son: 0 y  $-8$ , es decir, es un punto de silla

- ii) Se sustituye en  $\mathbf{H}_f(x, y)$  el punto crítico  $a_2 = (\sqrt{2}, -\sqrt{2})$

$$\mathbf{H}_f(\sqrt{2}, -\sqrt{2}) = \begin{pmatrix} 20 & 4 \\ 4 & 20 \end{pmatrix}$$

cuyos valores propios son: 24 y 16, es decir, es un mínimo relativo.

iii) Para el punto crítico  $a_3 = (-\sqrt{2}, \sqrt{2})$  es igual al ítem ii).  $\square$

**Nota B.13.** (Multiplicadores de Lagrange). *Este método permite hallar los extremos relativos de una función  $h(x)$  sujeta a una restricción  $g(x) = 0$ . Para esto se forma la función auxiliar  $H(x, \lambda) = h(x) - \lambda g(x)$ , donde  $\lambda (\neq 0)$ , es llamado el multiplicador de Lagrange; luego, se deriva  $H(x, \lambda)$  con respecto a  $x$  y a  $\lambda$ , y el sistema de ecuaciones se iguala a cero [56, p. 1283].*

## B.2.2. Desigualdad de Jensen

**DEFINICIÓN B.3.** De acuerdo con [5, p. 151], una función  $g$  se llama **convexa** en un intervalo  $[a, b]$  si, cualquiera que sean  $x$  y  $y$  de  $[a, b]$  y para todo  $\alpha$  tal que  $0 < \alpha < 1$ , se tiene

$$g(z) \leq \alpha g(y) + (1 - \alpha)g(x), \quad \text{siendo } z = \alpha y + (1 - \alpha)x.$$

Se dice que  $g$  es **cóncava** en  $[a, b]$  si es válida la desigualdad invertida,

$$g(z) \geq \alpha g(y) + (1 - \alpha)g(x), \quad \text{siendo } z = \alpha y + (1 - \alpha)x.$$

En la figura B.2 se ilustran una función cóncava y una convexa [5, p. 150].

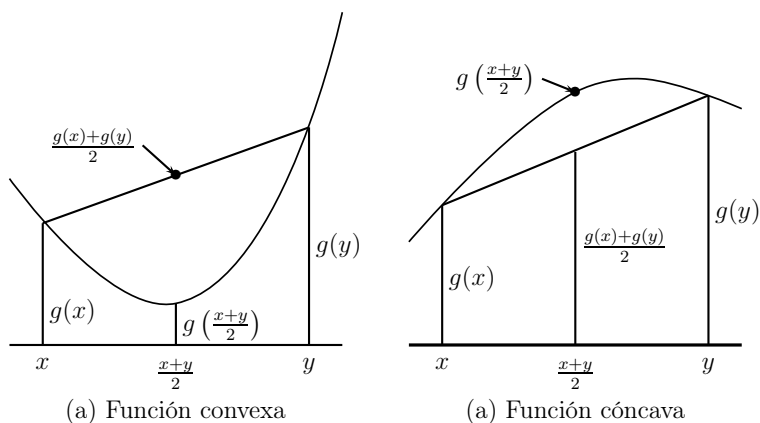


Figura B.2. Ilustración de función cóncava y convexa

**Teorema B.5.** Para cualquier variable aleatoria  $X$ , si  $g(x)$  es una función convexa, entonces [20, p. 190]:

$$\mathbb{E}[g(X)] \geq g[\mathbb{E}(X)].$$

La igualdad se tiene si y solo si para cada línea  $a + bx$  que es tangente a  $g(x)$  en  $x = \mathbb{E}(X)$ ,  $P(g[X] = a + bX) = 1$ .

*Demostración.* Para establecer la desigualdad, sea  $l(x)$  la recta tangente a  $g(x)$  en el punto  $g(\mathbb{E}[X])$ , digamos  $l(x) = a + bx$ . Por la convexidad de  $g(x)$  se tiene que  $g(x) \geq a + bx$ . Aplicando el operador esperanza se tiene que:

$$\mathbb{E}[g(X)] \geq \mathbb{E}[a + bX] = a + b\mathbb{E}[X] = l(\mathbb{E}[X]) = g(\mathbb{E}[X]),$$

aquí se uso el hecho de que  $l(x)$  es tangente en  $\mathbb{E}[X]$ . ☑

**Lema B.1.** Sean  $X_1, X_2, X_3, \dots, X_n$  variables aleatorias. Para todo  $\varepsilon > 0$ ,

$$P\left(\left|\sum_{j=1}^n X_j\right| \geq \varepsilon\right) \leq \sum_{j=1}^n P\left(|X_j| \geq \frac{\varepsilon}{n}\right).$$

*Demostración.* Veamos:

$$\bigcap_{j=1}^n \left\{|X_j| < \frac{\varepsilon}{n}\right\} \subseteq \left\{\sum_{j=1}^n |X_j| < \varepsilon\right\} \subseteq \left\{\left|\sum_{j=1}^n X_j\right| < \varepsilon\right\},$$

luego,

$$\begin{aligned} \left(\bigcap_{j=1}^n \left\{|X_j| < \frac{\varepsilon}{n}\right\}\right)^C &\supseteq \left\{\left|\sum_{j=1}^n X_j\right| \geq \varepsilon\right\}^C \\ \bigcup_{j=1}^n \left\{|X_j| \geq \frac{\varepsilon}{n}\right\} &\supseteq \left\{\sum_{j=1}^n X_j \geq \varepsilon\right\}, \end{aligned}$$

Por la propiedad 5. del teorema 1.1, se tiene que:

$$P\left(\left|\sum_{j=1}^n X_j\right| \geq \varepsilon\right) \leq P\left(\bigcup_{j=1}^n \left\{|X_j| \geq \frac{\varepsilon}{n}\right\}\right) \stackrel{(1.1)}{=} \sum_{j=1}^n P\left(|X_j| \geq \frac{\varepsilon}{n}\right). \quad \square$$

### B.3. Distribuciones de algunas formas cuadráticas

Supóngase que en la definición B.2,  $\mathbf{X}$  es un vector de variables aleatorias. En este apartado se estudia cómo se distribuyen algunas formas lineales o cuadráticas asociadas al vector aleatorio  $\mathbf{X}$ . Por ejemplo, si  $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , entonces  $\mathbf{c}^\top \mathbf{X} \sim N(\mathbf{c}^\top \boldsymbol{\mu}, \mathbf{c}^\top \boldsymbol{\Sigma} \mathbf{c})$  (ver ejemplo 2.23).

Por otra parte, si  $\mathbf{X} \sim N_n(\theta \mathbf{1}_n, \sigma^2 \mathbf{I})$ , entonces la distribución de  $\bar{X}$  es:

$$\begin{aligned} \bar{X} &= \frac{1}{n} \mathbf{1}_n^\top \mathbf{X} \sim N_n \left( \frac{1}{n} \mathbf{1}_n^\top [\theta \mathbf{1}_n], \frac{1}{n} \mathbf{1}_n^\top [\sigma^2 \mathbf{I}] \frac{1}{n} \mathbf{1}_n \right) \\ &\sim N_n \left( \frac{\theta}{n} \mathbf{1}_n^\top \mathbf{1}_n, \frac{\sigma^2}{n^2} \mathbf{1}_n^\top \mathbf{1}_n \right) \sim N \left( \theta, \frac{\sigma^2}{n} \right), \end{aligned}$$

el estadístico de contraste

$$\frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{X} - \theta)}{\sigma} \sim N(0, 1). \quad (\text{B.8})$$

**Ejemplo B.6.** Si  $\mathbf{X} \sim N_n(\theta \mathbf{1}_n, \sigma^2 \mathbf{I})$ , determine la distribución de  $\frac{n-1}{\sigma^2} S^2$ .

*Solución.* Expresando  $S_n^2$  como en el ejemplo B.3, se tiene que:

$$\frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \mathbf{X}^\top \left( \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \right) \mathbf{X} = \mathbf{X}^\top \left( \frac{1}{\sigma^2} \left[ \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \right] \right) \mathbf{X}.$$

Sea  $\mathbf{A} = \frac{1}{\sigma^2} \left[ \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \right]$ . Para usar el resultado del pie de página 33, se verifica si  $\mathbf{A}\mathbf{V}$  es idempotente:

$$\begin{aligned} \mathbf{A}\mathbf{V} &= \frac{1}{\sigma^2} \left[ \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \right] \sigma^2 \mathbf{I}_n = \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \\ \mathbf{A}\mathbf{V}\mathbf{A}\mathbf{V} &= \left[ \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \right] \left[ \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \right] \\ &= \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n - \frac{1}{n} \mathbf{J}_n + \frac{1}{n^2} \mathbf{J}_n \mathbf{J}_n \\ &= \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n = \mathbf{A}\mathbf{V}, \end{aligned}$$

entonces  $\mathbf{A}\mathbf{V}$  es idempotente. Por tanto,

$$\mathbf{X}^\top \mathbf{A}\mathbf{X} = \frac{n-1}{\sigma^2} S^2 \sim \chi^2(r, \lambda),$$

donde

$$\begin{aligned} r = \text{rango}(\mathbf{A}) &= \text{rango}\left(\frac{1}{\sigma^2} \left[ \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \right]\right) = \text{rango}\left(\mathbf{I}_n - \frac{1}{n} \mathbf{J}_n\right) \\ &= \underbrace{\text{rango}(k\mathbf{B})}_{=\text{rango}(\mathbf{B})} \\ &= \text{tr}\left(\mathbf{I}_n - \frac{1}{n} \mathbf{J}_n\right) = n \left(1 - \frac{1}{n}\right) = n - 1, \end{aligned}$$

aquí se consideró que  $\text{rango}(\mathbf{A}) = \text{tr}(\mathbf{A})$ , por ser  $\mathbf{A}$  idempotente. Luego,

$$\begin{aligned} \lambda &= \frac{1}{2} \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} = \frac{1}{2} \theta^2 \mathbf{1}_n^\top \frac{1}{\sigma^2} \left[ \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \right] \mathbf{1}_n \\ &= \frac{\theta^2}{2\sigma^2} \left[ \mathbf{1}_n^\top \mathbf{I}_n \mathbf{1}_n - \frac{1}{n} \underbrace{\mathbf{1}_n^\top \mathbf{1}_n}_{n} \underbrace{\mathbf{1}_n^\top \mathbf{1}_n}_{n} \right] = 0. \end{aligned}$$

Por lo tanto,

$$\frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2, \quad (\text{B.9})$$

es un estadístico de contraste.  $\square$

Para determinar la independencia entre  $\frac{\sqrt{n}(\bar{X}-\theta)}{\sigma}$  y  $\frac{n-1}{\sigma^2} S^2$ , se usará el teorema B.6.

**Teorema B.6.** Según [39, p. 52], si  $\mathbf{X} \sim N(\boldsymbol{\mu}, \mathbf{V})$ ,  $\mathbf{Q} = \mathbf{X}^\top \mathbf{A} \mathbf{X}$  y  $\mathbf{L} = \mathbf{B} \mathbf{X}$ , entonces  $\mathbf{Q}$  y  $\mathbf{L}$  son independientes si y solo si  $\mathbf{B} \mathbf{V} \mathbf{A} = \mathbf{0}$ .

**Ejemplo B.7.** Si  $\mathbf{X} \sim N_n(\theta \mathbf{1}_n, \sigma^2 \mathbf{I})$ , verifique que  $\frac{\sqrt{n}(\bar{X}-\theta)}{\sigma}$  y  $\frac{n-1}{\sigma^2} S^2$  son independientes.

*Solución.* Para probar la independencia es suficiente mostrar que  $\mathbf{B} \mathbf{V} \mathbf{A} = \mathbf{0}$ , donde  $\mathbf{B} = \frac{1}{n} \mathbf{1}_n^\top$ ,  $\mathbf{V} = \sigma^2 \mathbf{I}$  y  $\mathbf{A} = \frac{1}{n-1} \left( \mathbf{I} - \frac{1}{n} \mathbf{J}_n \right)$ , luego

$$\begin{aligned} \mathbf{B} \mathbf{V} \mathbf{A} &= \left( \frac{1}{n} \mathbf{1}_n^\top \right) (\sigma^2 \mathbf{I}) \left( \frac{1}{n-1} \left( \mathbf{I} - \frac{1}{n} \mathbf{J}_n \right) \right) = \frac{\sigma^2}{n(n-1)} \mathbf{1}_n^\top \left( \mathbf{I} - \frac{1}{n} \mathbf{J}_n \right) \\ &= \frac{\sigma^2}{n(n-1)} \left( \mathbf{1}_n^\top - \frac{1}{n} \underbrace{\mathbf{1}_n^\top \mathbf{1}_n}_{n} \mathbf{1}_n^\top \right) = \mathbf{0}. \quad \square \end{aligned}$$

Finalmente, para establecer la distribución de  $\frac{\sqrt{n}(\bar{X}-\theta)}{S_n}$  se usa la definición B.4.

**DEFINICIÓN B.4.** Según [39, p. 57], si  $Y \sim N(\delta, 1)$  y  $X \sim \chi_d^2$  son variables aleatorias independientes, la razón

$$t = \frac{y}{\sqrt{\frac{X}{d}}},$$

se llama *fdp t-Student no central con d grados de libertad y parámetro de no centralidad  $\delta$* . Se denota por  $t(d, \delta)$ . (El caso  $\delta = 0$  fue dado en el ejemplo 2.21).

De las expresiones (B.8) y (B.9) se tiene que:

$$\frac{\sqrt{n}(\bar{Y} - \theta)}{\sigma} \sim N(0, 1) \quad \text{y} \quad \frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2.$$

Además, en el ejemplo B.7 se mostró que  $\bar{Y}$  y  $S_n^2$  son independientes, luego  $\frac{\sqrt{n}(\bar{Y} - \theta)}{\sigma}$  y  $\frac{n-1}{\sigma^2} S^2$  son independientes. Por último, usando la definición B.4 se tiene que:

$$t = \frac{y}{\sqrt{\frac{X}{d}}} = \frac{\frac{\sqrt{n}(\bar{X}_n - \theta)}{\sigma}}{\sqrt{\frac{\frac{n-1}{\sigma^2} S_n^2}{n-1}}} \sim t_{(n-1, \delta)}$$

$$\frac{\sqrt{n}(\bar{X}_n - \theta)}{S_n} \sim t_{n-1}.$$

### Demostración teorema 8.3.

De acuerdo con [82, p. 137], como  $\Sigma$  es una matriz simétrica semidefinida positiva, existe una matriz  $B$  de orden  $p \times q$  (ver [60, p. 21]), tal que:

$$\Sigma = BB^T, \quad B^T B = \Lambda,$$

con  $\Lambda$  una matrix diagonal de orden  $q$ , en que  $\text{rango}(\Sigma) = q$ , que contiene los valores propios positivos de la matrix  $\Sigma$ . Sea

$$Z_n = (B^T B)^{-1} B^T \sqrt{n}(T_n - \theta),$$

entonces

$$\begin{aligned} \mathbb{E}(Z_n) &= \mathbb{E} \left[ (B^T B)^{-1} B^T \sqrt{n}(T_n - \theta) \right] \\ &= (B^T B)^{-1} B^T \sqrt{n} \underbrace{\mathbb{E}[(T_n - \theta)]}_{\mathbf{0}} = \mathbf{0} \end{aligned}$$

y

$$\begin{aligned} \mathbb{V}(Z_n) &= \mathbb{V} \left[ (B^\top B)^{-1} B^\top \sqrt{n}(T_n - \theta) \right] \\ &= (B^\top B)^{-1} B^\top \mathbb{V} \left[ \sqrt{n}(T_n - \theta) \right] B (B^\top B)^{-1}. \end{aligned}$$

Note que

$$\mathbb{V}(Z_n) \rightarrow (B^\top B)^{-1} B^\top \underbrace{\Sigma}_{BB^\top} B (B^\top B)^{-1} = I_q.$$

Por el teorema del límite central clásico, teorema 3.22,

$$\{Z_n\}_j \xrightarrow{D} Z,$$

en que  $Z \sim N(0, 1)$ , y por el Cramér-Wold, teorema 3.13, se tiene que:

$$Z_n \xrightarrow{D} Z,$$

en que  $Z \sim N_q(\mathbf{0}, I_q)$ . Sea  $Q_n = Z_n^\top (B^\top A_n B) Z_n$  una forma cuadrática en  $Z_n$ ,

$$\begin{aligned} Q_n &= Z_n^\top (B^\top A_n B) Z_n \\ &= \left[ (B^\top B)^{-1} B^\top \sqrt{n}(T_n - \theta) \right]^\top B^\top A_n B \left[ (B^\top B)^{-1} B^\top \sqrt{n}(T_n - \theta) \right] \\ &= \sqrt{n}(T_n - \theta)^\top \underbrace{B (B^\top B)^{-1} B^\top A_n B (B^\top B)^{-1} B^\top}_{A_n \text{ (ejercicio)}} \sqrt{n}(T_n - \theta), \end{aligned}$$

por lo tanto,

$$Q_n = n(T_n - \theta)^\top A_n (T_n - \theta).$$

Tomando límite cuando  $n \rightarrow \infty$ ,

$$\begin{array}{ccc} Q_n = \left[ \begin{array}{c} B \\ \underbrace{Z_n} \end{array} \right]^\top & \underbrace{A_n} & B \underbrace{Z_n} \xrightarrow{D} Q, \\ \downarrow D & \downarrow & \downarrow D \\ Z & A & Z \end{array}$$

cuando  $n \rightarrow \infty$

con  $Q = Z^\top B^\top A B Z$ . Aplicando la nota de pie de página 33,

$$Q \sim \chi^2(r, \lambda),$$

con  $r = \text{rango}(\mathbf{B}^\top \mathbf{A} \mathbf{B})$  y  $\lambda = \frac{1}{2} \mathbf{0}^\top \mathbf{B}^\top \mathbf{A} \mathbf{B} \mathbf{0} = 0$  si y solo si  $\mathbf{B}^\top \mathbf{A} \mathbf{B} \mathbf{I}_q$  es idempotente. Veamos:

$$\begin{aligned} \mathbf{B}^\top \mathbf{A} \mathbf{B} \mathbf{I}_q \mathbf{B}^\top \mathbf{A} \mathbf{B} \mathbf{I}_q &= \mathbf{B}^\top \mathbf{A} \underbrace{\mathbf{B} \mathbf{B}^\top}_{\Sigma} \mathbf{A} \mathbf{B} = \mathbf{B}^\top \mathbf{A} \Sigma \mathbf{A} \mathbf{B} \\ &= \underbrace{\mathbf{B}^\top \mathbf{A} \Sigma \mathbf{A}}_{\text{si y solo si } \mathbf{A} \Sigma \mathbf{A} = \mathbf{A}} \mathbf{B} = \mathbf{B}^\top \mathbf{A} \mathbf{B}. \end{aligned}$$

Si  $\Sigma$  es una inversa generalizada de  $\mathbf{A}$ ,  $\mathbf{B}^\top \mathbf{A} \mathbf{B}$  es idempotente,

$$\begin{aligned} r = \text{rango}(\mathbf{B}^\top \mathbf{A} \mathbf{B}) &= \text{tr}(\mathbf{B}^\top \mathbf{A} \mathbf{B}) = \text{tr}(\mathbf{A} \mathbf{B} \mathbf{B}^\top) \\ &= \text{tr}(\mathbf{A} \Sigma) = \text{rango}(\mathbf{A} \Sigma) = \text{rango}(\Sigma) = q. \end{aligned}$$

Aquí se usó el hecho de que la matriz es idempotente (ver B.5). Por tanto  $Q \sim \chi_q^2$ , concluyendo que:

$$n(\mathbf{T}_n - \boldsymbol{\theta})^\top \mathbf{A}_n (\mathbf{T}_n - \boldsymbol{\theta}) \xrightarrow{D} \chi_q^2.$$

## B.4. Otros tópicos

### B.4.1. Condiciones de regularidad

Sea  $\mathbf{X}$  un vector aleatorio con fdp conjunta  $p(\mathbf{x}; \boldsymbol{\theta})$ ,  $\boldsymbol{\theta} \in \Theta$  (ver [37]), entonces se considera que:

- C1.  $\Theta$  es un intervalo abierto en  $\mathbb{R}$  (finito o no).
- C2.  $p(\mathbf{x}; \boldsymbol{\theta})$  es positiva sobre un conjunto  $S$  independiente de  $\boldsymbol{\theta} \in \Theta$ .
- C3.  $\frac{\partial}{\partial \boldsymbol{\theta}} p(\mathbf{x}; \boldsymbol{\theta})$  existe para todo  $\boldsymbol{\theta} \in \Theta$  y todo  $\mathbf{x} \in S$ , excepto posiblemente en un conjunto  $N \subset S$ , el cual es independiente de  $\boldsymbol{\theta}$  y tal que para todo  $\boldsymbol{\theta} \in \Theta$ ,  $p_{\boldsymbol{\theta}}(\mathbf{x} \in N) = 0$ .
- C4. Para todo  $\boldsymbol{\theta} \in \Theta$ :

$$\int \cdots \int_{S^n} p(x_1; \boldsymbol{\theta}) \cdots p(x_n; \boldsymbol{\theta}) dx_1 \cdots dx_n$$

o

$$\sum_S \cdots \sum_S p(x_1; \boldsymbol{\theta}) \cdots p(x_n; \boldsymbol{\theta})$$

puede ser derivable bajo el signo de la integral o bajo el signo de la suma, respectivamente [86].

C5. Si

$$\mathbb{E}_\theta \left[ \left( \frac{\partial}{\partial \theta} \ln [p(\mathbf{x}; \theta)] \right)^2 \right] = \mathbb{E}_\theta \left[ \left[ \frac{\partial}{\partial \theta} \ell(\theta; \mathbf{x}) \right]^2 \right]$$

denotado por  $I_n(\theta)$  es positivo para todo  $\theta \in \Theta$ , entonces

$$I_n(\theta) = \mathbb{E}_\theta \left[ \left[ \frac{\partial}{\partial \theta} \ell(\theta; \mathbf{x}) \right]^2 \right]. \tag{B.10}$$

C6. Si  $W(\mathbf{x})$  es un estimador insesgado cualquiera de  $q(\theta)$ , entonces

$$\int \cdots \int_{S^n} W(\mathbf{x}) p(x_1; \theta) \dots p(x_n; \theta) dx_1 \dots dx_n$$

o

$$\sum_S \cdots \sum_S W(\mathbf{x}) p(x_1; \theta) \dots p(x_n; \theta),$$

puede ser derivable bajo el signo de la integral o bajo el signo de la suma, respectivamente.

### B.4.2. Ejemplo numérico de máxima verosimilitud

El parámetro  $r$  se determina numéricamente mediante la **EMV** usando la expresión (4.12). Si los datos están agrupados, se establece la frecuencia relativa  $h(k) = \frac{n_k}{n}$  y, luego,

$$f(r) = \ln \left( 1 + \frac{\bar{n}}{r} \right) - \sum_{k=1}^n h(k) \left[ \sum_{m=0}^{k-1} \frac{1}{m+r} \right] \quad y$$

$$f'(r) = - \frac{\bar{n}}{r^2 + \bar{n}r} + \sum_{k=1}^n h(k) \left[ \sum_{m=0}^{k-1} \frac{1}{(m+r)^2} \right].$$

**Ejemplo B.8.** En la tabla B.1 se presentan los datos dados en [33, p. 167] correspondientes a una investigación realizada durante tres meses consecutivos sobre el número de accidentes de trabajo ocurridos a los 414 operarios de máquinas.

Tabla B.1. Número de accidentes de los operarios de maquinaria

$k$	0	1	2	3	4	5	6	7	8	Total
$n_k$	296	74	26	8	4	4	1	0	1	414

Fuente: adaptada de [33, p. 167].

donde  $n_k$  es el número de operarios que han tenido  $k$  accidentes reales. Determine la fmp que más se ajuste a los datos, presentando las respectivas aproximaciones de las fmp con la estimación de sus parámetros.

*Solución.* Primero, se calcula la frecuencia relativa  $h(k) = \frac{n_k}{n}$  y, luego, se halla

$$\bar{n} = \frac{100}{207} \quad \text{y} \quad S_n^2 = \frac{43\,199}{42\,849}.$$

Entre las distribuciones dadas se buscará la distribución de probabilidad que más se ajuste a los datos.

1. Distribución de Poisson: aunque se obtuvo que  $S_n^2 > \bar{n}$ , se presenta también la respectiva aproximación con la distribución Poisson, para realizar comparaciones. Por los métodos de momentos y máxima verosimilitud, se obtiene que  $\hat{\lambda} = \bar{n}$ .
2. Distribución binomial negativa: por el método de frecuencias, el valor de los parámetros se establece como sigue:

$$\hat{p} = 2 \frac{n_2}{n_1} - \frac{n_1}{n_0} = 0,452702703 \quad \text{y} \quad \hat{r} = \frac{n_1^2}{2n_0n_2 - n_1^2} = 0,55223881.$$

Por el método de momentos, el valor de los parámetros se establece mediante la expresión (4.10), en este caso, se obtiene que:

$$\hat{q} = \frac{22499}{43199} \approx 0,520822241 \quad \text{y} \quad \hat{r} = \frac{10000}{22499} \approx 0,444464198.$$

Por el método de máxima verosimilitud se determina el parámetro  $r$  de la fmp binomial negativa usando la expresión (4.12) y se llega a:

Tabla B.2. Iteraciones de Newton Raphson para encontrar  $r$

$i$	$r_i$
0	0,444464198
1	0,47099858
2	0,47423699
3	0,47427900
4	0,47427901
5	0,47427901

el valor de  $\hat{q}$  se despeja de la expresión (4.13),

$$\hat{q} = 0,504602594.$$

En la siguiente tabla se resumen las estimaciones obtenidas

Método	$\hat{q}$	$\hat{r}$
Frecuencia	0,547297297	0,55223881
EMM	0,520822241	0,444464198
EMV	0,504602594	0,47427901

En la tabla B.3 se muestra que la aproximación BN resulta más apropiada para ajustar los datos que la obtenida con la distribución Poisson.

Tabla B.3. Comparación de distribuciones ajustadas

No. de acc. $k$	Accidentes observados	Distribución teórica ajustada			
		Poisson	bin. negat. <sup>1</sup>	bin. negat. <sup>2</sup>	bin. negat. <sup>3</sup>
0	296	255,4	267,3	298,5	296,7
1	74	123,4	80,8	69,1	71,0
2	26	29,8	34,3	26,0	26,4
3	8	4,8	16,0	11,0	11,0
4	4	0,6	7,8	4,9	4,8
5	4	0,1	3,9	2,3	2,2
6	1	0,0	2,0	1,1	1,0
7	0	0,0	1,0	0,5	0,5
8	1	0,0	0,5	0,3	0,2
Total	414	414,0	413,4	413,8	413,8
$\chi_c^2$		55,7	12,1	2,7	2,7
$\nu$		2	4	3	3
1. Frecuencia		2. EMM		3. EMV	

En este caso, la prueba chi-cuadrado (expresión (7.22)) de bondad de ajuste se modifica y se calcula como:

$$\chi_c^2 = \sum_{k=1}^m \frac{(n_k - n\hat{p}_k)^2}{n\hat{p}_k} = \sum_{k=1}^m \frac{(O_k - E_k)^2}{E_k} \quad \text{siempre que } E_k \geq 2 \forall k,$$

donde  $m$  es el número de clases (categorías). Si  $E_k < 2$  se agrupan las clases con frecuencia pequeña con las clases adyacentes. Con esta prueba se verifica si los datos observados en una muestra aleatoria se ajustan a una fmp teórica ( $\hat{p}_k$ ) con  $q$  parámetros desconocidos. Según [25, p. 425], el valor tabulado de la  $\chi^2$  tiene un grado de libertad menos por cada parámetro estimado; es decir,  $\nu = m - q - 1$ . Para 2 y 3 grados de libertad y  $\alpha = 0,05$ , de la tabla C.3 se obtiene que los límites de significancia de la  $\chi^2$  son

$$\chi_{2,\alpha}^2 = 5,99146 \quad \text{y} \quad \chi_{3,\alpha}^2 = 7,81473.$$

Luego, se rechaza  $H_0$  si  $\chi_c^2 > 7,81473$ , para los respectivos  $\chi_c^2$  menores a  $7,81473$ , no hay evidencia suficiente para rechazar los modelos propuestos, de modo que al nivel de 5%, la distribución **BN** da un ajuste razonable a los datos.  $\checkmark$

## B.5. Ejercicios

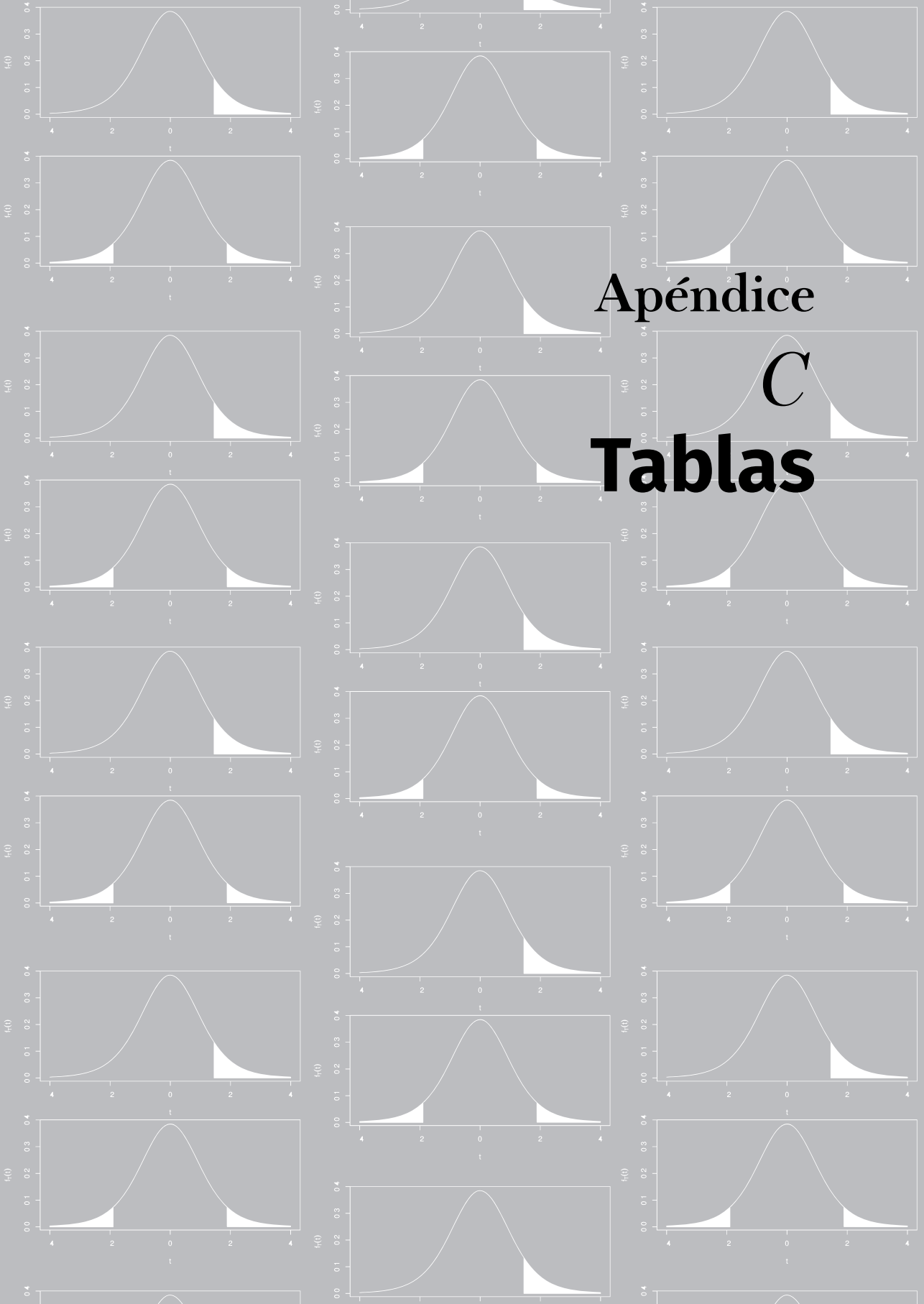
1. Completar la demostración de la desigualdad de Jensen.
2. Muestre que si  $g(x)$  es una función cóncava, entonces:

$$\mathbb{E}[g(X)] \leq g[\mathbb{E}(X)].$$

3. Verificar que  $\mathbf{G}$  en el ejemplo **B.4** es una inversa generalizada de  $\mathbf{A}$ .
4. Para  $\mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{y}$ , encontrar una solución de  $\hat{\boldsymbol{\beta}}$ , donde

$$\mathbf{Y} = \begin{pmatrix} 10 \\ 5 \\ 12 \\ 3 \\ 4 \\ 3 \\ 8 \\ 7 \\ 6 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\tau}_3 \end{pmatrix}.$$

# Apéndice C Tablas



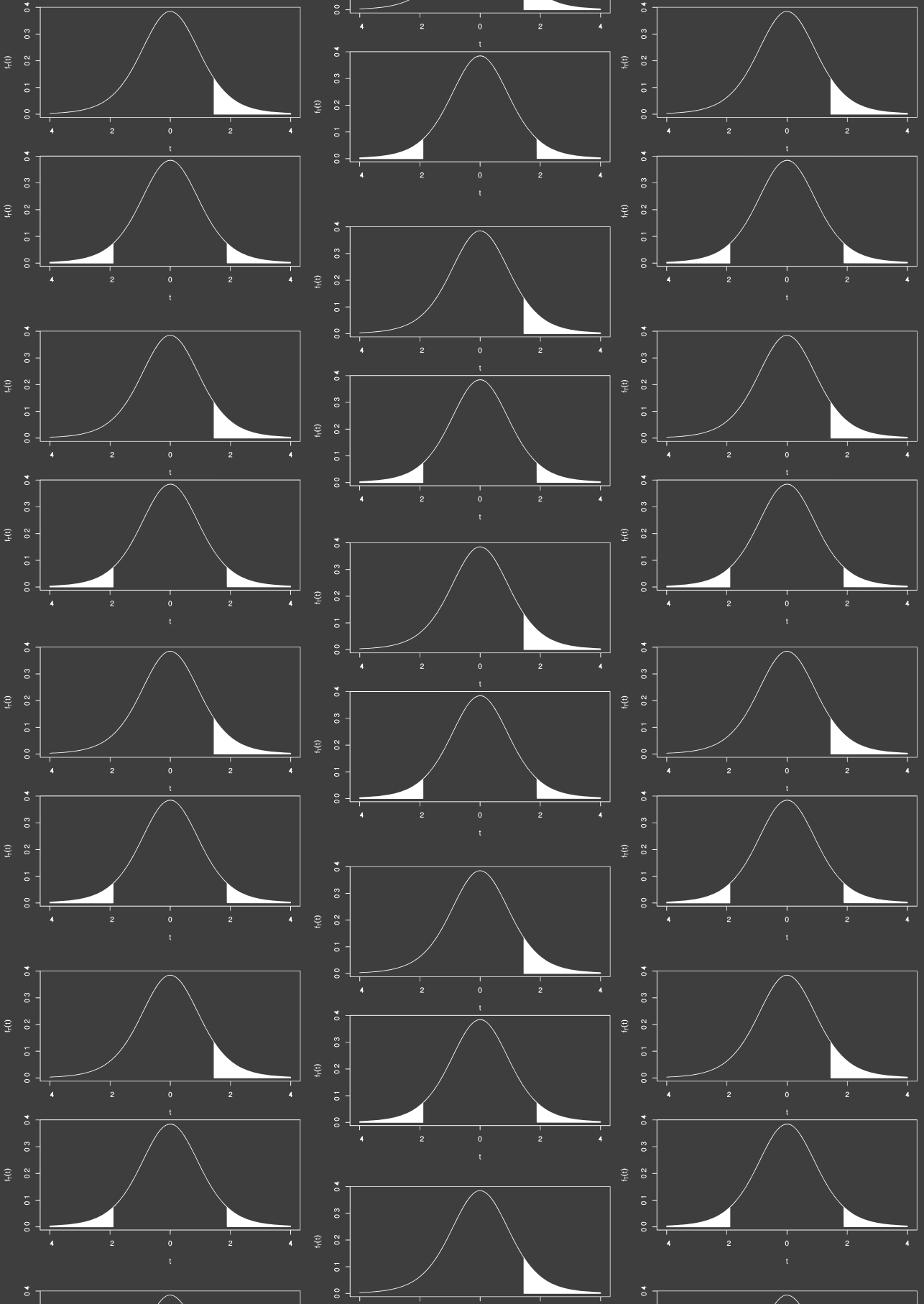


Tabla C.1. Valores de la *fda* normal estándar

$z$	0,0000	0,0100	0,0200	0,0300	0,0400	0,0500	0,0600	0,0700	0,0800	0,0900
-3,9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
-3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,6	0,0002	0,0002	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641

Continúa en la siguiente página



Tabla C.2. Valores de cuantiles de la *fda t*-Student

$\nu$	$t_{0,001}$	$t_{0,005}$	$t_{0,01}$	$t_{0,025}$	$t_{0,05}$	$t_{0,1}$	$t_{0,2}$
1	-318,3088	-63,6567	-31,8205	-12,7062	-6,3138	-3,0777	-1,3764
2	-22,3271	-8,9248	-6,9646	-4,3027	-2,9200	-1,8856	-1,0607
3	-10,2145	-5,8409	-4,5407	-3,1824	-2,3534	-1,6377	-0,9785
4	-7,1732	-4,6041	-3,7469	-2,7764	-2,1318	-1,5332	-0,9410
5	-5,8934	-4,0321	-3,3649	-2,5706	-2,0150	-1,4759	-0,9195
6	-5,2076	-3,7074	-3,1427	-2,4469	-1,9432	-1,4398	-0,9057
7	-4,7853	-3,4995	-2,9980	-2,3646	-1,8946	-1,4149	-0,8960
8	-4,5008	-3,3554	-2,8965	-2,3060	-1,8595	-1,3968	-0,8889
9	-4,2968	-3,2498	-2,8214	-2,2622	-1,8331	-1,3830	-0,8834
10	-4,1437	-3,1693	-2,7638	-2,2281	-1,8125	-1,3722	-0,8791
11	-4,0247	-3,1058	-2,7181	-2,2010	-1,7959	-1,3634	-0,8755
12	-3,9296	-3,0545	-2,6810	-2,1788	-1,7823	-1,3562	-0,8726
13	-3,8520	-3,0123	-2,6503	-2,1604	-1,7709	-1,3502	-0,8702
14	-3,7874	-2,9768	-2,6245	-2,1448	-1,7613	-1,3450	-0,8681
15	-3,7328	-2,9467	-2,6025	-2,1314	-1,7531	-1,3406	-0,8662
16	-3,6862	-2,9208	-2,5835	-2,1199	-1,7459	-1,3368	-0,8647
17	-3,6458	-2,8982	-2,5669	-2,1098	-1,7396	-1,3334	-0,8633
18	-3,6105	-2,8784	-2,5524	-2,1009	-1,7341	-1,3304	-0,8620
19	-3,5794	-2,8609	-2,5395	-2,0930	-1,7291	-1,3277	-0,8610
20	-3,5518	-2,8453	-2,5280	-2,0860	-1,7247	-1,3253	-0,8600
21	-3,5272	-2,8314	-2,5176	-2,0796	-1,7207	-1,3232	-0,8591
22	-3,5050	-2,8188	-2,5083	-2,0739	-1,7171	-1,3212	-0,8583
23	-3,4850	-2,8073	-2,4999	-2,0687	-1,7139	-1,3195	-0,8575
24	-3,4668	-2,7969	-2,4922	-2,0639	-1,7109	-1,3178	-0,8569
25	-3,4502	-2,7874	-2,4851	-2,0595	-1,7081	-1,3163	-0,8562
26	-3,4350	-2,7787	-2,4786	-2,0555	-1,7056	-1,3150	-0,8557
27	-3,4210	-2,7707	-2,4727	-2,0518	-1,7033	-1,3137	-0,8551
28	-3,4082	-2,7633	-2,4671	-2,0484	-1,7011	-1,3125	-0,8546
29	-3,3962	-2,7564	-2,4620	-2,0452	-1,6991	-1,3114	-0,8542
30	-3,3852	-2,7500	-2,4573	-2,0423	-1,6973	-1,3104	-0,8538
35	-3,3400	-2,7238	-2,4377	-2,0301	-1,6896	-1,3062	-0,8520
40	-3,3069	-2,7045	-2,4233	-2,0211	-1,6839	-1,3031	-0,8507
45	-3,2815	-2,6896	-2,4121	-2,0141	-1,6794	-1,3006	-0,8497
50	-3,2614	-2,6778	-2,4033	-2,0086	-1,6759	-1,2987	-0,8489
60	-3,2317	-2,6603	-2,3901	-2,0003	-1,6706	-1,2958	-0,8477
70	-3,2108	-2,6479	-2,3808	-1,9944	-1,6669	-1,2938	-0,8468
80	-3,1953	-2,6387	-2,3739	-1,9901	-1,6641	-1,2922	-0,8461
90	-3,1833	-2,6316	-2,3685	-1,9867	-1,6620	-1,2910	-0,8456
100	-3,1737	-2,6259	-2,3642	-1,9840	-1,6602	-1,2901	-0,8452
200	-3,1315	-2,6006	-2,3451	-1,9719	-1,6525	-1,2858	-0,8434
500	-3,1066	-2,5857	-2,3338	-1,9647	-1,6479	-1,2832	-0,8423

Continúa en la siguiente página

Tabla C.2: Continuación tabla

$\nu$	$t_{0,8}$	$t_{0,9}$	$t_{0,95}$	$t_{0,975}$	$t_{0,99}$	$t_{0,995}$	$t_{0,999}$
1	1,3764	3,0777	6,3138	12,7062	31,8205	63,6567	318,3088
2	1,0607	1,8856	2,9200	4,3027	6,9646	8,9248	22,3271
3	0,9785	1,6377	2,3534	3,1824	4,5407	5,8409	10,2145
4	0,9410	1,5332	2,1318	2,7764	3,7469	4,6041	7,1732
5	0,9195	1,4759	2,0150	2,5706	3,3649	4,0321	5,8934
6	0,9057	1,4398	1,9432	2,4469	3,1427	3,7074	5,2076
7	0,8960	1,4149	1,8946	2,3646	2,9980	3,4995	4,7853
8	0,8889	1,3968	1,8595	2,3060	2,8965	3,3554	4,5008
9	0,8834	1,3830	1,8331	2,2622	2,8214	3,2498	4,2968
10	0,8791	1,3722	1,8125	2,2281	2,7638	3,1693	4,1437
11	0,8755	1,3634	1,7959	2,2010	2,7181	3,1058	4,0247
12	0,8726	1,3562	1,7823	2,1788	2,6810	3,0545	3,9296
13	0,8702	1,3502	1,7709	2,1604	2,6503	3,0123	3,8520
14	0,8681	1,3450	1,7613	2,1448	2,6245	2,9768	3,7874
15	0,8662	1,3406	1,7531	2,1314	2,6025	2,9467	3,7328
16	0,8647	1,3368	1,7459	2,1199	2,5835	2,9208	3,6862
17	0,8633	1,3334	1,7396	2,1098	2,5669	2,8982	3,6458
18	0,8620	1,3304	1,7341	2,1009	2,5524	2,8784	3,6105
19	0,8610	1,3277	1,7291	2,0930	2,5395	2,8609	3,5794
20	0,8600	1,3253	1,7247	2,0860	2,5280	2,8453	3,5518
21	0,8591	1,3232	1,7207	2,0796	2,5176	2,8314	3,5272
22	0,8583	1,3212	1,7171	2,0739	2,5083	2,8188	3,5050
23	0,8575	1,3195	1,7139	2,0687	2,4999	2,8073	3,4850
24	0,8569	1,3178	1,7109	2,0639	2,4922	2,7969	3,4668
25	0,8562	1,3163	1,7081	2,0595	2,4851	2,7874	3,4502
26	0,8557	1,3150	1,7056	2,0555	2,4786	2,7787	3,4350
27	0,8551	1,3137	1,7033	2,0518	2,4727	2,7707	3,4210
28	0,8546	1,3125	1,7011	2,0484	2,4671	2,7633	3,4082
29	0,8542	1,3114	1,6991	2,0452	2,4620	2,7564	3,3962
30	0,8538	1,3104	1,6973	2,0423	2,4573	2,7500	3,3852
35	0,8520	1,3062	1,6896	2,0301	2,4377	2,7238	3,3400
40	0,8507	1,3031	1,6839	2,0211	2,4233	2,7045	3,3069
45	0,8497	1,3006	1,6794	2,0141	2,4121	2,6896	3,2815
50	0,8489	1,2987	1,6759	2,0086	2,4033	2,6778	3,2614
60	0,8477	1,2958	1,6706	2,0003	2,3901	2,6603	3,2317
70	0,8468	1,2938	1,6669	1,9944	2,3808	2,6479	3,2108
80	0,8461	1,2922	1,6641	1,9901	2,3739	2,6387	3,1953
90	0,8456	1,2910	1,6620	1,9867	2,3685	2,6316	3,1833
100	0,8452	1,2901	1,6602	1,9840	2,3642	2,6259	3,1737
200	0,8434	1,2858	1,6525	1,9719	2,3451	2,6006	3,1315
500	0,8423	1,2832	1,6479	1,9647	2,3338	2,5857	3,1066

Tabla C.3. Valores de cuantiles de la  $\chi^2$  chi-cuadrada

$\nu$	$\chi^2_{0,005}$	$\chi^2_{0,010}$	$\chi^2_{0,025}$	$\chi^2_{0,050}$	$\chi^2_{0,10}$	$\chi^2_{0,9}$	$\chi^2_{0,95}$	$\chi^2_{0,975}$	$\chi^2_{0,99}$	$\chi^2_{0,995}$
1	0,0000	0,0002	0,0010	0,0039	0,0158	2,7055	3,8415	5,0239	6,6349	7,8794
2	0,0100	0,0201	0,0506	0,1026	0,2107	4,6052	5,9915	7,3778	8,2103	10,5966
3	0,0717	0,1148	0,2158	0,3518	0,5844	6,2514	7,8147	8,3484	11,3449	12,8382
4	0,2070	0,2971	0,4844	0,7107	1,0636	7,7794	8,4877	11,1433	13,2767	14,8603
5	0,4117	0,5543	0,8312	1,1455	1,6103	8,2364	11,0705	12,8325	15,0863	16,7496
6	0,6757	0,8721	1,2373	1,6354	2,2041	10,6446	12,5916	14,4494	16,8119	18,5476
7	0,9893	1,2390	1,6899	2,1673	2,8331	12,0170	14,0671	16,0128	18,4753	20,2777
8	1,3444	1,6465	2,1797	2,7326	3,4895	13,3616	15,5073	17,5345	20,0902	21,9550
9	1,7349	2,0879	2,7004	3,3251	4,1682	14,6837	16,9190	18,0228	21,6660	23,5894
10	2,1559	2,5582	3,2470	3,9403	4,8652	15,9872	18,3070	20,4882	23,2093	25,1882
11	2,6032	3,0535	3,8157	4,5748	5,5778	17,2750	18,6751	21,9200	24,7250	26,7568
12	3,0738	3,5706	4,4038	5,2260	6,3038	18,5493	21,0261	23,3367	26,2170	28,2995
13	3,5650	4,1069	5,0088	5,8919	7,0415	18,8119	22,3620	24,7356	27,6882	28,8195
14	4,0747	4,6604	5,6287	6,5706	7,7895	21,0641	23,6848	26,1189	28,1412	31,3193
15	4,6009	5,2293	6,2621	7,2609	8,5468	22,3071	24,9958	27,4884	30,5779	32,8013
16	5,1422	5,8122	6,9077	7,9616	9,3122	23,5418	26,2962	28,8454	31,9999	34,2672
17	5,6972	6,4078	7,5642	8,6718	10,0852	24,7690	27,5871	30,1910	33,4087	35,7185
18	6,2648	7,0149	8,2307	9,3905	10,8649	25,9894	28,8693	31,5264	34,8053	37,1565
19	6,8440	7,6327	8,9065	10,1170	11,6509	27,2036	30,1435	32,8523	36,1909	38,5823
20	7,4338	8,2604	9,5908	10,8508	12,4426	28,4120	31,4104	34,1696	37,5662	38,9968
21	8,0337	8,8972	10,2829	11,5913	13,2396	29,6151	32,6706	35,4789	38,9322	41,4011
22	8,6427	9,5425	10,9823	12,3380	14,0415	30,8133	33,9244	36,7807	40,2894	42,7957
23	8,2604	10,1957	11,6886	13,0905	14,8480	32,0069	35,1725	38,0756	41,6384	44,1813
24	8,8862	10,8564	12,4012	13,8484	15,6587	33,1962	36,4150	38,3641	42,9798	45,5585
25	10,5197	11,5240	13,1197	14,6114	16,4734	34,3816	37,6325	40,6465	44,3141	46,9279

Continúa en la siguiente página

Tabla C.3: Continuación tabla

$\nu$	$\chi^2_{0,005}$	$\chi^2_{0,010}$	$\chi^2_{0,025}$	$\chi^2_{0,050}$	$\chi^2_{0,10}$	$\chi^2_{0,9}$	$\chi^2_{0,95}$	$\chi^2_{0,975}$	$\chi^2_{0,99}$	$\chi^2_{0,995}$
26	11,1602	12,1981	13,8439	15,3792	17,2919	35,5632	38,8851	41,9232	45,6417	48,2899
27	11,8076	12,8785	14,5734	16,1514	18,1139	36,7412	40,1133	43,1945	46,9629	48,6449
28	12,4613	13,5647	15,3079	16,9279	18,9392	37,9159	41,3371	44,4608	48,2782	50,9934
29	13,1211	14,2565	16,0471	17,7084	18,7677	38,0875	42,5570	45,7223	48,5879	52,3356
30	13,7867	14,9535	16,7908	18,4927	20,5992	40,2560	43,7730	46,9792	50,8922	53,6720
35	17,1918	18,5089	20,5694	22,4650	24,7967	46,0588	48,8018	53,2033	57,3421	60,2748
40	20,7065	22,1643	24,4330	26,5093	28,0505	51,8051	55,7585	58,3417	63,6907	66,7660
45	24,3110	25,9013	28,3662	30,6123	33,3504	57,5053	61,6562	65,4102	68,9568	73,1661
50	27,9907	28,7067	32,3574	34,7643	37,6886	63,1671	67,5048	71,4202	76,1539	78,4900
55	31,7348	33,5705	36,3981	38,9580	42,0596	68,7962	73,3115	77,3805	82,2921	85,7490
60	35,5345	37,4849	40,4817	43,1880	46,4589	74,3970	78,0819	83,2977	88,3794	91,9517
65	38,3831	41,4436	44,6030	47,4496	50,8829	78,9730	84,8206	88,1771	94,4221	98,1051
70	43,2752	45,4417	48,7576	51,7393	55,3289	85,5270	90,5312	95,0232	100,4252	104,2149
75	47,2060	48,4750	52,9419	56,0541	58,7946	91,0615	96,2167	100,8393	106,3929	110,2856
80	51,1719	53,5401	57,1532	60,3915	64,2778	96,5782	101,8795	106,6286	112,3288	116,3211
85	55,1696	57,6339	61,3888	64,7494	68,7772	102,0789	107,5217	112,3984	118,2357	122,3246
90	58,1963	61,7541	65,6466	68,1260	73,2911	107,5650	113,1453	118,1359	124,1163	128,2989
95	63,2496	65,8984	68,9249	73,5198	77,8184	113,0377	118,7516	123,8580	128,9727	134,2465
100	67,3276	70,0649	74,2219	77,9295	82,3581	118,4980	124,3421	128,5612	135,8067	140,1695

## Referencias

- [1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions, with Formulas, Graphs and Mathematical Tables*. Dover Publications, Nueva York, 1965.
- [2] A. Agresti and B. Caffo. “Simple and Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures.” *The American Statistician*, volume 54, no. 4, pp. 280–288, 2000. <http://doi.org/10.1080/00031305.2000.10474560>.
- [3] J. Albert. *Bayesian Computation With R*. Springer-Verlag, Nueva York, 2 edition, 2009.
- [4] T. M. Apostol. *Mathematical Analysis*. Addison Wesley, Reading, MA, 2 edition, 1974.
- [5] T. M. Apostol. *Calculus I. Cálculo con funciones de una variable, con una introducción al álgebra Lineal*. Editorial Reverté, Barcelona, 2009.
- [6] T. M. Apostol. *Calculus II. Cálculo con funciones de una variable, con una introducción al álgebra Lineal*. Editorial Reverté, Barcelona, 2009.
- [7] R. Artes and D. A. Botter. *Funções de Estimacão em Modelos de Regressão*. ABE, São Paulo, 2005. 9na Escola de Modelos de Regressão.
- [8] E. Artin. *The gamma function*. Athena series; selected topics in mathematics. Holt, Rinehart and Winston, Nueva York, 1st edition, 1964.
- [9] A. C. Atkinson. “Two Graphical Displays for Outlying and Influential Observations in Regression.” *Biometrika*, volume 68, no. 1, pp. 13–20, 1981. <http://doi.org/10.2307/2335801>.
- [10] J. O. Berger. *Statistical decision theory and Bayesian analysis*. Springer series in statistics. Springer-Verlag, Nueva York, 2nd edition, 1985.

- [11] B. R. Bhat. *Modern Probability Theory*. John Wiley & Sons, Nueva York, 2nd edition, 1985.
- [12] P. J. Bickel and K. A. Doksum. *Mathematical statistics: basic ideas and selected topics*, volume I. CRC Press - Taylor & Francis Group, San Francisco, 2nd. edition, 2015.
- [13] P. Billingsley. *Convergence of probability measures*. Wiley series in probability and statistics. Probability and statistics section. Wiley, Hoboken, NJ, 2nd edition, 1999.
- [14] Y. M. Bishop, S. E. Fienberg, P. W. Holland, *et al.* *Discrete Multivariate Analysis, Theory and Practice*. Springer Monographs in Mathematics. Springer, Nueva York, 1 edition, 2007.
- [15] D. Blackwell and M. A. Girshick. *Theory of games and statistical decisions*. Wiley publication in statistics. Wiley, Nueva York, 1954.
- [16] L. Blanco. *Probabilidad*. Universidad Nacional de Colombia, Bogotá, 2a. edition, 2010.
- [17] J. I. Bula. “Estadísticas e indicadores de la Universidad Nacional de Colombia.” *Revista de la Oficina Nacional de Planeación*, volume 13, pp. 1–143, 2007. [https://estadisticas.unal.edu.co/fileadmin/user\\_upload/Revista\\_2007.pdf](https://estadisticas.unal.edu.co/fileadmin/user_upload/Revista_2007.pdf).
- [18] C. W. Burrill. *Measure, Integration, and Probability*. McGraw-Hill, Nueva York, 1972.
- [19] G. C. Canavos. *Probabilidad y estadística. Aplicaciones y Métodos*. McGraw-Hill, México, 1988.
- [20] G. Casella and R. L. Berger. *Statistical Inference*. Thompson Learning, Pacific Grove, 2nd. edition, 2002.
- [21] E. Cepeda. *Estadística matemática*. Universidad Nacional de Colombia, Bogotá, 1a. edition, 2015.
- [22] W. J. Conover. *Practical nonparametric statistics*. Wiley series in probability and statistics. John Wiley & Sons, Nueva York, 3rd edition, 1999.
- [23] R. Courant and F. John. *Introduction to calculus and analysis*, volume 1 of *Classics in Mathematics*. Springer-Verlag Berlin Heidelberg, Nueva York, 1999.

- [24] D. R. Cox and D. V. Hinkley. *Theoretical statistics*. CRC Press - Taylor & Francis Group, Nueva York, 2nd edition, 2017.
- [25] H. Cramér. *Mathematical methods of statistics*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, New York, 9th edition, 1962.
- [26] A. J. Dobson and A. G. Barnett. *An introduction to generalized linear models*. Chapman & Hall Statistics Texts. CRC Press: Taylor & Francis Group, Boca Ratón, 4th edition, 2018.
- [27] E. J. Dudewicza, Y. Ma, E. Mai, *et al.* “Exact solutions to the Behrens–Fisher Problem: Asymptotically optimal and finite sample efficient choice among.” *Journal of Statistical Planning and Inference*, volume 137, no. 5, pp. 1584–1605, 2007. <http://doi.org/10.1016/j.jspi.2006.09.007>.
- [28] B. Efron and D. V. Hinkley. “Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information.” *Biometrika*, volume 65, no. 3, pp. 457–483, 1978. <https://doi.org/10.1093/biomet/65.3.457>.
- [29] B. S. Everitt. *The Cambridge Dictionary of Statistics*. Cambridge University Press, Cambridge, 2006.
- [30] T. S. Ferguson. *A course in large sample theory*. Chapman & Hall texts in statistical science. Chapman and Hall/CRC, Londres, 1st edition, 1996.
- [31] R. A. Fisher. “Two New Properties of Mathematical Likelihood.” *Proceedings of the Royal Society of London. Series A*, volume 144, no. 852, pp. 285–307, 1934. <http://doi.org/10.1098/rspa.1934.0050>.
- [32] R. A. Fisher. “The Negative Binomial Distribution.” *Annals of Eugenics*, volume 11, no. 1, pp. 182–187, 1941. <https://doi.org/10.1111/j.1469-1809.1941.tb02284.x>.
- [33] M. Fisz. *Probability theory and mathematical statistics*. John Wiley & Sons, Nueva York, 3ed edition, 1980.
- [34] S. French. *Decision theory: An introduction to the mathematics of rationality*. John Wiley & Sons, Nueva York, first edition, 1988.

- [35] M. Ghosh, D. Reid, and D. A. S. Frase. “Ancillary statistics: a review.” *Statista Sinica*, volume 20, pp. 1309–1332, 2010. <https://utstat.utoronto.ca/reid/research/A20n41.pdf>.
- [36] B. V. Gnedenko. *Theory of Probability*. Taylor & Francis, India, sixth edition, 2018.
- [37] V. P. Godambe. “An optimum property of regular maximum likelihood estimation.” *The Annals of Mathematical Statistics*, volume 31, no. 4, pp. 1208–1211, 1960. <http://doi.org/10.1214/aoms/1177705693>.
- [38] D. A. Harville. *Matrix Algebra from a Statistician’s Perspective*. Springer, Nueva York, 1997.
- [39] R. R. Hocking. *Methods and applications of linear models*. John Wiley & Sons, Hoboken, N.J., 1996.
- [40] P. J. Huber. “Robust Estimation of a Location Parameter.” *The Annals of Mathematical Statistics*, volume 35, no. 1, pp. 73–101, 1964. <http://doi.org/10.1214/aoms/1177703732>.
- [41] V. S. Huzurbazar. “The likelihood equation, consistency and the maxima of the likelihood function.” *Annals of Human Genetics*, volume 14, no. 1, pp. 185–200, 1947. <http://doi.org/10.1111/j.1469-1809.1947.tb02394.x>.
- [42] V. S. Huzurbazar. “Confidence Intervals for the Parameter of a Distribution Admitting A Sufficient Statistic when The Range Depends on the Parameter.” *Journal of the Royal Statistical Society Series B (Methodological)*, volume 17, no. 1, pp. 86–90, 1955. <http://doi.org/10.1111/j.2517-6161.1955.tb00182.x>.
- [43] B. R. James. *Probabilidade: um curso em nível intermediário*. Associação Instituto Nacional de Matemática Pura e Aplicada, Rio de Janeiro, 2da edition, 2002.
- [44] J. A. Jiménez. *Álgebra matricial con aplicaciones en estadística*. Universidad Nacional de Colombia, Bogotá, 3a edition, 2017.
- [45] N. L. Johnson, A. W. Kemp, and S. Kotz. *Univariate Discrete Distributions*. Wiley Series in Probability and Statistics. Wiley-Interscience, Nueva York, 3 edition, 2005.

- [46] R. A. Johnson, J. Ladalla, and S. T. Liu. “Differential Relations, in the Original Parameters, which Determine the First Two Moments of the Multiparameter Exponential Family.” *The Annals of Statistics*, volume 7, no. 1, pp. 232–235, 1979. <http://doi.org/10.1214/aos/1176344569>.
- [47] B. Jørgensen and R. Labouriau. *Exponential families and theoretical inference*, volume 52 of *Monografías de Matemática*. Springer, Rio de Janeiro, 2 edition, 2012.
- [48] J. G. Kalbfleisch. *Probability and Statistical Inference*, volume 2 of *Springer Texts in Statistics*. Springer-Verlag, Nueva York, second edition, 1985.
- [49] S. H. Kim and A. S. Cohen. “On the Behrens–Fisher Problem: A Review.” *Journal of Educational and Behavioral Statistics*, volume 23, no. 4, pp. 356–377, 1998. <http://doi.org/10.3102/10769986023004356>.
- [50] M. Knott. “*Statistics 2.*”, 2002. <http://stats.lse.ac.uk/knott/tiger.links/stats2/st2.pdf>.
- [51] E. L. Lehmann. “An Interpretation of Completeness and Basu’s Theorem.” *Journal of the American Statistical Association*, volume 76, pp. 335–340, 1981. <http://doi.org/10.2307/2287834>.
- [52] E. L. Lehmann. *Elements of Large-Sample Theory*. Springer Texts in Statistics. Springer-Verlag, Nueva York, 1999.
- [53] E. L. Lehmann and G. Casella. *Theory of point estimation*. Springer texts in statistics. Springer-Verlag, Nueva York, 2nd edition, 2003.
- [54] E. L. Lehmann and J. P. Romano. *Testing statistical hypotheses*. Springer-Verlag, Nueva York, third edition, 2008.
- [55] J. G. Leite and J. M. Singer. *Métodos assintóticos em estatística. Fundamentos e Aplicações*. Associação Brasileira de Estatística, São Paulo, 1990.
- [56] L. Leithold. *El cálculo con geometría analítica*. Harla, México, 1992.
- [57] D. V. Lindley. *Introduction to Probability and Statistics from a Bayesian Viewpoint, Inference*, volume 2. Cambridge University Press, Nueva York, 1970.
- [58] E. C. MacRae. “Matrix Derivatives with an Application to an Adaptive Linear Decision Problem.” *The Annals of Statistics*, volume 2, no. 2, pp. 337 – 346, 1974. <https://doi.org/10.1214/aos/1176342667>.

- [59] M. Magalhães. *Probabilidade e Variáveis Aleatórias*. Editora da Universidade de São Paulo, São Paulo, 2006.
- [60] J. R. Magnus and H. Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons, Chichester, 1995.
- [61] M. J. Marqués De Cantú. *Probabilidad y estadística para ciencias químico-biológicas*. McGraw-Hill Interamericana, México, 1991.
- [62] H. Mayorga. *Inferencia estadística*. Universidad Nacional de Colombia, Bogotá, 2004.
- [63] J. S. Milton. *Estadística para biología y ciencias de la salud*. McGraw-Hill, Interamericana, México, 3a edition, 2001.
- [64] D. C. Montgomery and G. C. Runger. *Probabilidad y Estadística aplicadas a la ingeniería*. McGraw-Hill, México, 1a edition, 1996.
- [65] A. M. Mood, F. A. Graybill, and D. C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill, Singapur, third edition, 1974.
- [66] M. Muñoz and L. Blanco. *Introducción a la teoría avanzada de la probabilidad*. Universidad Nacional de Colombia, Bogotá, 2002.
- [67] N. Mukhopadhyay. *Probability and statistical inference*. Statistics: A Series of Textbooks and Monographs. Marcel Dekker, Nueva York, 1 edition, 2000.
- [68] V. N. Murty. “Integration by Parts.” *The Two-Year College Mathematics Journal*, volume 11, no. 2, pp. 90–94, 1980. <http://doi.org/10.2307/3026660>.
- [69] F. Oberhettinger. *Fourier Transforms of Distributions and Their Inverses: A Collection of Tables*. Probability & Mathematical Statistics Monograph. Academic Press, Nueva York, 1973.
- [70] J. Pfanzagl. “On the Behrens–Fisher problem.” *Biometrika*, volume 61, no. 1, pp. 39–47, 1974. <http://doi.org/10.1093/biomet/61.1.39>.
- [71] E. J. G. Pitman and J. Wishart. “Sufficient statistics and intrinsic accuracy.” *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 32, no. 4, pp. 567–579, 1936. <http://doi.org/10.1017/S0305004100019307>.

- [72] A. Prada and D. A. Ruiz. *Respuesta de la pitahaya (Cereus triangularis Haw.) a la fertilización edáfica y foliar en el municipio de Sasaima*. Trabajo de grado, Universidad Nacional de Colombia, Facultad de Agronomía, Bogotá, 1990.
- [73] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, 2011. [www.R-project.org](http://www.R-project.org).
- [74] A. C. Rencher. *Methods of multivariate analysis*. Wiley series in probability and mathematical statistics. John Wiley & Sons, Nueva York, 2nd edition, 2002.
- [75] S. M. Ross. *Introduction to Probability Models*. Academic Press, Nueva York, 12th edition, 2019.
- [76] G. G. Roussas. *A Course in Mathematical Statistics*. Academic Press, San Diego, 1997.
- [77] G. G. Roussas. *An introduction to probability and statistical inference*. Elsevier Inc., Londres, 2nd edition, 2015.
- [78] F. J. Samaniego. *A comparison of the Bayesian and frequentist approaches to estimation*. Springer series in statistics. Springer-Verlag, Nueva York, 1 edition, 2010.
- [79] L. J. Savage. “The Theory of Statistical Decision.” *Journal of the American Statistical Association*, volume 46, no. 253, pp. 55–67, 1951. <http://doi.org/10.2307/2280094>.
- [80] H. Scheffé. “Practical Solutions of the Behrens–Fisher Problem.” *Journal of the American Statistical Association*, volume 65, no. 332, pp. 1501–1508, 1970. <http://doi.org/10.1080/01621459.1970.10481179>.
- [81] R. L. Schilling. *Measures, Integrals and Martingales*. Cambridge: Cambridge University Press, 2 edition, 2017.
- [82] P. K. Sen and J. M. Singer. *Large Sample Methods in Statistics: An Introduction with Applications*. CRC Press - Taylor & Francis Group, Nueva York, 2017.
- [83] T. A. Severini. *Elements of Distribution Theory*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, Nueva York, 2005.

- [84] J. M. Singer and L. M. Magalhães. “Efeito do sódio na pressão arterial em hipertensos essenciais.” Technical Report RAE-CEA-97P37, Centro de Estatística Aplicada, IME, Universidade de São Paulo, 1998. <https://repositorio.usp.br/directbitstream/3bdd911c-4f66-4532-860a-d67167f2b4f0/979489.pdf>.
- [85] A. Stuart and J. K. Ord. *Kendall's Advanced Theory of Statistics*, volume I of *Kendall's Advanced Theory of Statistics*. Griffin, Londres, 6th edition, 1994.
- [86] E. Talvila. “Necessary and sufficient conditions for differentiating under the integral sign.” *The American Mathematical Monthly*, volume 108, no. 6, pp. 544–548, 2001. <http://doi.org/10.1080/00029890.2001.11919782>.
- [87] R. C. Tiwari, Y. Yang, and J. N. Zalkikar. “Bayes estimation for the Pareto failure-model using Gibbs sampling.” *IEEE Transactions on Reliability*, volume 45, no. 3, pp. 471–476, 1996. <http://doi.org/10.1109/24.537018>.
- [88] D. D. Wackerly, W. Mendenhall III, and R. L. Scheaffer. *Estadística matemática con aplicaciones*. Cengage Learning Editores, México, 7 edition, 2009.
- [89] R. E. Walpole, R. H. Myers, S. L. Myers, *et al.* *Probability & statistics for engineers & scientists*. Pearson Education, Nueva York, 9th edition, 2012.
- [90] B. L. Welch. “The Significance of the Difference Between Two Means when the Population Variances are Unequal.” *Biometrika*, volume 29, no. 3-4, pp. 350–362, 1938. <http://doi.org/10.1093/biomet/29.3-4.350>.
- [91] B. L. Welch. “The Generalization of ‘Student’s’ Problem when Several Different Population Variances are Involved.” *Biometrika*, volume 34, no. 1-2, pp. 28–35, 1947. <http://doi.org/10.1093/biomet/34.1-2.28>.
- [92] J. E. Wilkins. “A Note on Skewness and Kurtosis.” *The Annals of Mathematical Statistics*, volume 15, no. 3, pp. 333–335, 1944. <http://doi.org/10.1214/aoms/1177731243>.

# Índice analítico

- Ángulo(s)
  - de rotación, 486
- Behrens-Fisher
  - problema de, 355
- Binomio de Newton, 474
- Bondad
  - de ajuste, 429
- Coefficiente
  - de correlación, 111
- Completez, 272
- Condiciones de regularidad, 502
- Consistencia, 293
- Convergencia
  - casí siempre, 181
  - de vectores aleatorios, 182
  - débil, 182
  - en distribución, 182
    - de vectores aleatorios, 185
  - en media, 182
  - en probabilidad, 176
    - de vectores aleatorios, 177
- Convolución de
  - variables, 115
- Correlación
  - coeficiente, 111
  - entre variables, 111
- Cota
  - inferior
    - de Cramér-Rao, 282
- Covarianza, 63
- Criterio
  - Factorización
    - de Fisher-Neyman, 265
    - Suficiencia minimal, 269
- Curtosis, 66
  - estandarizada, 312
- Descomposición
  - de la covarianza, 110
  - de la varianza, 110
- Desigualdad
  - de Cauchy-Schwarz, 111
  - de Chebyshev, 69
  - de Cramér-Rao, 282
  - de Jensen, 496
  - triangular
    - para integrales, 85
- Distribución
  - a priori, 227
  - beta, 46
  - binomial, 37
    - negativa, 43
  - de los estadísticos de orden,
    - 159
  - degenerada, 176
  - del máximo, 153
  - del mínimo, 154
  - del rango, 162
  - $F$  (de Fisher), 143
  - gamma, 51

- gamma, 51
  - Gumbel, 98, 257
  - hipergeométrica, 44
  - log-normal, 54
  - multinomial, 133
  - normal, 47
    - bivariada, 120
    - multivariada, 145
  - Poisson, 39
  - Rayleigh, 306
  - t*-Student, 141
    - no central, 499
  - uniforme
    - discreta, 35
    - generalizada, 161
  - Weibull, 82
- Eficiencia
- relativa asintótica, 297
- Error
- cuadrático
    - medio, 274
  - de truncación, 477
  - tipo I, 382
  - tipo II, 382
- Espacio
- de probabilidad, 11
  - medible, 8
  - muestral, 3
- Esperanza
- matemática, 55
- Estadística
- completa, 272
  - suficiente, 263
    - minimal, 269
- Estadísticos
- de orden, 153
- Estimación
- bayesiana
    - por intervalo, 365
    - puntual, 234
  - máxima verosimilitud, 213
  - método de momentos, 209
    - por sustitución de frecuencias, 207
- Estimador, 207
- bayesiano, 235
  - por intervalo, 318
- Estimadores
- propiedades, 263
- Experimento
- aleatorio, 3
  - Bernoulli, 36
  - Poisson, 39
- Familia
- conjugada, 234
  - exponencial, 300
    - en forma natural, 304
  - lineal, 125
  - multivariada, 308
  - independiente, 23
    - dos a dos, 23
- Fisher
- información de, 282
- Forma(s)
- cuadrática(s), 489
    - clasificación, 489
    - distribuciones, 497
- Fórmula
- del binomio de Newton, 474
  - cuadrática, 492
  - de inversión, 85, 135
  - de Stirling, 67
  - de Taylor, 473
    - con resto, 476
  - de Welch, 358
  - del resto, 476
- Función
- beta

- incompleta regularizada, 97
  - cóncava, 496
  - característica, 83
    - multidimensional, 134
  - convexa, 496
  - de decisión, 253
  - de densidad de probabilidad, 35
  - de masa de probabilidad, 34
  - de pérdida, 253
  - de potencia, 382
  - de riesgo, 254
  - de verosimilitud
    - inducida, 299
  - generadora
    - de momentos, 75
    - de probabilidad, 72
  - indicadora, 30
- Gradiente, 493
- Índice
  - de curtosis, 66
  - de dispersión, 66
- Información
  - de Fisher, 282
- Intervalo(s)
  - aleatorio, 318
  - de confianza
    - bayesiano, 371
    - única estadística, 322
  - unilateral, 318
- Lema
  - de Neyman-Pearson, 393
  - Helly-Bray, 183
- Ley
  - débil de los grandes números, 191
  - números de Khintchine, 192
  - del estadístico inconsciente, 59, 106
  - esperanza total, 107
  - fuerte de los grandes números
    - de Khintchine, 194
    - de Kolmogorov, 194
- Matriz
  - de covarianzas, 130
  - de información de Fisher, 310
  - de rotación, 488
  - hat, 260
  - hessiana, 494
  - inversa generalizada, 489
  - acobiana, 138
- Mediana
  - muestral, 153
- Medida
  - de probabilidad, 11
- Método(s)
  - de estimación, 207–257
    - por intervalo, 317
    - puntual, 207
  - de integración tabular, 62
  - de mínimos cuadrados, 237
  - delta, 190
  - minimax, 250
  - pivotal, 339
- Momentos
  - centrales, 59
  - factoriales, 59
  - ordinarios, 59
- Muestra
  - aleatoria, 151
- Multiplicadores
  - de Lagrange, 496

- Norma
  - de un vector, 485
- Operador
  - Vec, 300
- Órdenes
  - de magnitud
    - de funciones, 169
    - de sucesiones, 163
    - en probabilidad, 171
    - propiedades, 166
- Polinomio
  - de Taylor, 476
- Principio
  - de inclusión-exclusión, 15
  - de invarianza del EMV, 298
  - de sustitución, 207
  - de verosimilitud, 214
- Probabilidad
  - clásica, 9
  - condicional, 17
  - de cobertura, 319
  - frecuencial, 10
- Propiedades
  - de estimadores, 263
  - de la fda, 33
  - espacio de probabilidad, 13
- Prueba
  - de hipótesis, 381-443
  - de Kolmogorov, 438
  - de Pearson, 430
- Razón
  - de chances, 209
  - de Katz, 66
  - de verosimilitud, 385
    - monótona, 402
- Regla
  - de Bayes, 25
  - de decisión, 253
- de L'Hôpital, 493
- de Leibniz, 494
- de multiplicación, 18
- Regresión
  - lineal
    - múltiple, 241
    - simple, 237
- Residuo
  - de Cauchy, 480
  - de Lagrange, 480
- Suficiencia, 263
- Teorema
  - central del límite, 196
  - de Liapunov, 199
  - de Lindeberg, 199
  - Cramér-Wold, 135, 186
  - de transformación
    - multivariado, 138
    - univariado, 119
  - Lehmann-Scheffé, 282
  - Rao-Blackwell, 277
  - Slutsky, 187
- Valor
  - esperado, 54
  - condicional, 107
- Variable(s)
  - aleatoria(s), 27
    - independientes, 102
  - continua, 34
  - discreta, 34
  - pivote, 339
- Varianza
  - condicional, 109
  - ponderada, 353
- Vector
  - aleatorio, 127
  - bidimensional, 101

*Optimización estática y dinámica en economía,*  
tercera edición, fue editado por el Centro  
Editorial de la Facultad de Ciencias,  
Universidad Nacional de Colombia, sede Bogotá.  
Fuente principal Baskerville y Fira Sans. Bogotá D. C.

