

### 1.1. Conceptos básicos

En la vida cotidiana nos encontramos constantemente con ejemplos de procesos estocásticos de diversa naturaleza. En economía, para ser específicos, podemos mencionar las exportaciones de una determinada mercancía, la inflación y el nivel de desempleo. En otras áreas se encuentran casos como el de la temperatura del aire en cierta estación meteorológica y la concentración de dióxido de carbono en la sangre de una persona. En suma, se trata de sistemas cuyas variaciones de estado en el tiempo dependen en cierto sentido del azar.

La definición rigurosa de proceso estocástico está fuera del alcance de este libro. Nuestro estudio se va a limitar a funciones aleatorias escalares del tipo  $X(t)$ , en las que  $X$  es la única variable de interés (por ejemplo, el índice de desempleo) y  $t$ , el argumento de la función, representa el tiempo.

Para entender este concepto, recordamos aquí que, en matemáticas, por sucesión se entiende una colección numerada y ordenada de objetos, finita o infinita. Más formalmente, una sucesión  $\{a(n)\}_{n \in \mathbb{N}}$  o simplemente  $a(n)$  es una función de argumento (índice) natural. Aquí  $\mathbb{N}$  es el conjunto de los números naturales, es decir,  $n = 0, 1, 2, \dots$  o  $n = 1, 2, \dots$  según el contexto. Por tradición, las funciones de variable discreta como  $a(n)$  se denotan  $a_n$ , y el subíndice  $n$  representa la posición del elemento: en el caso  $n = 1, 2, \dots$ ,  $a_1$  es el primer elemento de la sucesión;  $a_2$ , el segundo;  $a_n$ , el  $n$ -ésimo, etc. Como ejemplo, sirvan  $a_n = 2n$ ,  $n = 1, 2, \dots$  (los números pares positivos) y  $a_n = n^2$ ,  $n = 1, 2, \dots$  (los cuadrados perfectos). También se puede definir una sucesión por enumeración o por recurrencia. Volviendo a los números pares

positivos, es posible escribirlos como el conjunto  $(2, 4, 6, \dots)$ , aunque esta forma con puntos suspensivos introduce cierta ambigüedad, y por la recurrencia  $a_n = a_{n-1} + 2$ ,  $a_1 = 2$ .

En las aplicaciones, a menudo la variable  $n$  representa el tiempo. Sea este el caso o no, un elemento de la sucesión  $a_n$  en la posición  $n$  siempre produce el mismo valor. Después de todo, si, como en uno de los ejemplos anteriores,  $a_n = n^2$ , entonces  $a_5 = 25$  cada vez que  $n = 5$ .

Un proceso estocástico  $\{X_t\}_{t \in \mathbb{N}}$ , o  $X_t$  para simplificar, es una sucesión de variables aleatorias indexadas por un parámetro  $t$  que generalmente representa el tiempo. Nosotros vamos a considerar indistintamente variables continuas o discretas, pero el tiempo, como la notación lo indica, se mide en forma discreta (días, semanas, meses, trimestres, etc.). En ese caso hablamos de proceso estocástico en tiempo discreto, sucesión aleatoria o serie de tiempo (serie temporal).

Asumiendo  $t = 0, 1, 2, \dots$ , el proceso está compuesto por la sucesión de variables aleatorias  $(X_0, X_1, X_2, \dots, X_t, X_{t+1}, \dots)$ , finita o infinita. Usualmente, se emplea  $X$  (o  $Y$ ) para denotar la variable aleatoria de interés, y el subíndice  $t$  indica la posición de la variable en el tiempo. Notemos que llamamos  $X_t$  tanto al proceso como a la variable  $X$  en la posición temporal  $t$ , pero eso nunca es causa de confusión.

En un análisis del proceso  $X_t$ , los valores observados a veces se registran como  $(x_0, x_1, x_2, \dots, x_n)$ , con minúscula para distinguirlos de las variables aleatorias, y se considera que corresponden a una muestra finita, una «realización» del proceso. Luego, en una segunda observación del proceso se obtiene otra muestra  $(x'_0, x'_1, x'_2, \dots, x'_n)$  en la que es posible que haya coincidencias con la primera puesto que cada variable toma aleatoriamente un valor de entre los posibles según su distribución probabilística, y así sucesivamente. De aquí en adelante no vamos a distinguir entre la variable aleatoria  $X_t$  y su valor  $x_t$  en el tiempo  $t$  (Shumway y Stoffer, 2006).

En el modelo clásico de regresión lineal se estudia la relación estadística entre la variable dependiente  $y$ , digamos el salario mensual de una persona (en pesos colombianos), y las variables explicativas, por ejemplo, la escolaridad  $X_2$  (en años) y la edad  $X_3$  (en años), asumiendo que las variables se miden en el mismo tiempo  $t$  para diferentes personas. El eje del tiempo «se corta» en el momento de la medición, y por tal razón estos modelos se conocen como de corte transversal:

$$E(y | X_2, X_3) = \beta_1 + \beta_2 X_2 + \beta_3 X_3.$$

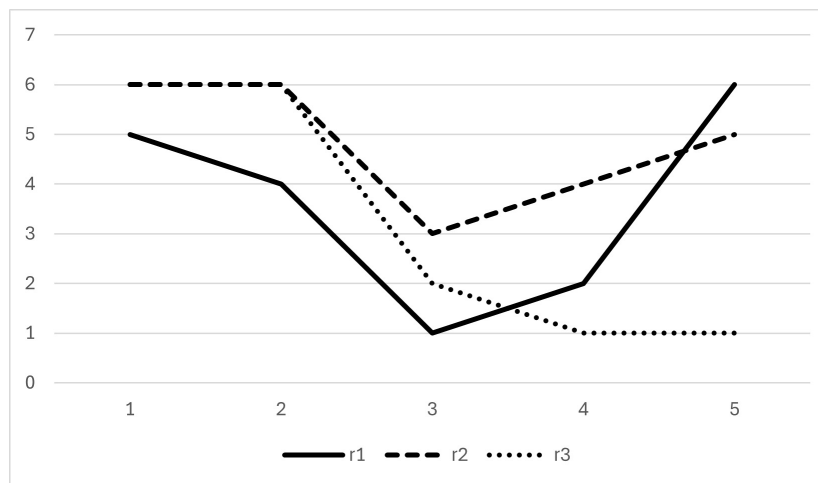
Ahora bien, a diferencia de este modelo, en uno estocástico se analiza la evolución de una única variable aleatoria en el tiempo. De igual forma, es posible estudiar el comportamiento temporal simultáneo de un conjunto (un vector) de variables aleatorias. Por ejemplo, en una ciudad podemos estudiar la evolución temporal conjunta de las ventas diarias de carros nuevos,  $X_t$ , el precio de la gasolina,  $Y_t$ , y el número de usuarios del transporte público,  $Z_t$ , es decir, del vector  $(X_t, Y_t, Z_t)$ . Desafortunadamente, ese tipo de análisis, conocido como VAR, está fuera del alcance de este trabajo.

Un ejemplo sencillo de proceso estocástico, en este caso finito, es el lanzamiento de un dado cinco veces:  $\{X_t\}_{t=1,2,3,4,5}$ . La tabla 1.1 y la figura 1.1 presentan, respectivamente, tres realizaciones (muestras) de este experimento y sus gráficos correspondientes.

Tabla 1.1. Realizaciones (muestras)

$X_t$	$t$	$r_1$	$r_2$	$r_3$
$X_1$	1	5	6	6
$X_2$	2	4	6	6
$X_3$	3	1	3	2
$X_4$	4	2	4	1
$X_5$	5	6	5	1

Figura 1.1. Realizaciones (gráficos)



En la práctica se requiere que, con base en una sola muestra del proceso  $X_t$ , se ajuste el modelo que mejor describe en un sentido técnico el comportamiento de los datos y, si es necesario, hacer predicciones (pronósticos) a partir de dicho modelo. Este proceso consta de varias etapas:

1. Identificación del modelo.
2. Ajuste.
3. Validación.
4. Cálculo de los pronósticos.

La manera de implementar estos pasos depende de la metodología escogida para el ajuste del modelo, tal como se muestra más adelante en ejemplos prácticos. En todo caso, en general, existen tres técnicas principales:

1. Regresión.
2. Metodología Box-Jenkins.
3. Suavizado exponencial (*exponential smoothing*).

El suavizado exponencial, en cuanto a posibilidades, desafortunadamente agrega muy poco a la metodología Box-Jenkins y, por lo tanto, lo excluimos completamente de nuestra consideración.

Un proceso estocástico general  $\{X_t\}_{t \in \mathbb{N}}$  normalmente se considera constituido por una tendencia  $m_t$ , una componente  $S_t$  que, por ser periódica, se conoce como componente estacional (en referencia a las estaciones del año, que se repiten), y una sucesión de choques aleatorios  $u_t$  (los ciclos económicos de largo plazo no forman parte de nuestro análisis), es decir,

$$X_t = m_t + S_t + u_t. \quad (1.1)$$

La tendencia  $m_t$ , como el nombre lo sugiere, indica la propensión del proceso a crecer o decrecer durante el período de observación. Supongamos, por ejemplo, que la población de una ciudad turística muestra una tendencia creciente en los últimos años. A su vez, la componente estacional  $S_t$  representa las fluctuaciones periódicas del proceso. Así, podría asumirse que la población de la ciudad muestra un comportamiento periódico debido a la afluencia del turismo durante las vacaciones. En la práctica,  $m_t$  es un polinomio de grado 1 (una recta) o de grado 2 (una parábola) y  $S_t$  es una función periódica con período  $T$ . Ambos términos son deterministas (no estocásticos).

Por otra parte, el término  $u_t$  es una variable aleatoria, un elemento de un proceso estocástico  $\{u_t\}_{t \in \mathbb{N}}$  conocido como ruido blanco. Este se considera el resultado de la influencia de todos los innumerables factores que afectan a  $X_t$ , tales como el clima (temperatura, precipitación), el estado de la economía (precios, salarios, inflación), la estabilidad política (local y nacional), la calidad de la infraestructura (estado de las carreteras, aeropuertos), la calidad de los servicios (agua, electricidad, internet), la tasa de criminalidad, el nivel de precios de los alimentos y el combustible, la calidad y el precio del transporte, etc. Estas variables son tan numerosas que, con base en el teorema central del límite, es razonable suponer que  $u_t$  tiene una distribución normal.

Las figuras 1.2 a 1.5 ilustran el comportamiento individual de las componentes del proceso y del proceso mismo.

Figura 1.2. Ruido blanco

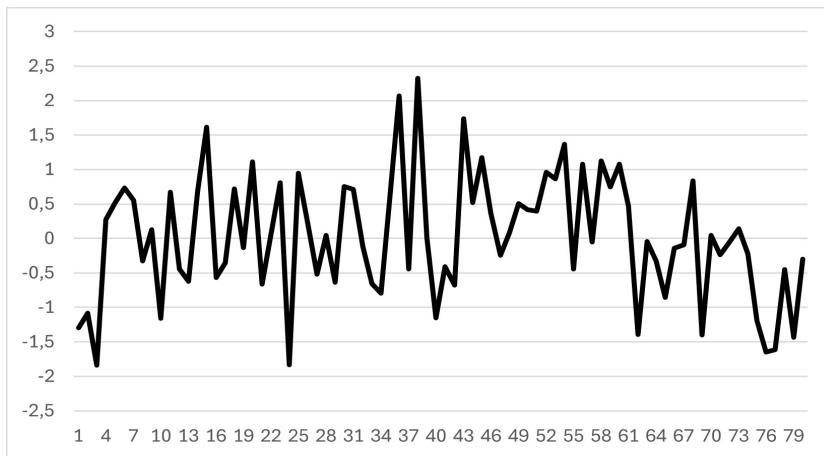


Figura 1.3. Tendencia

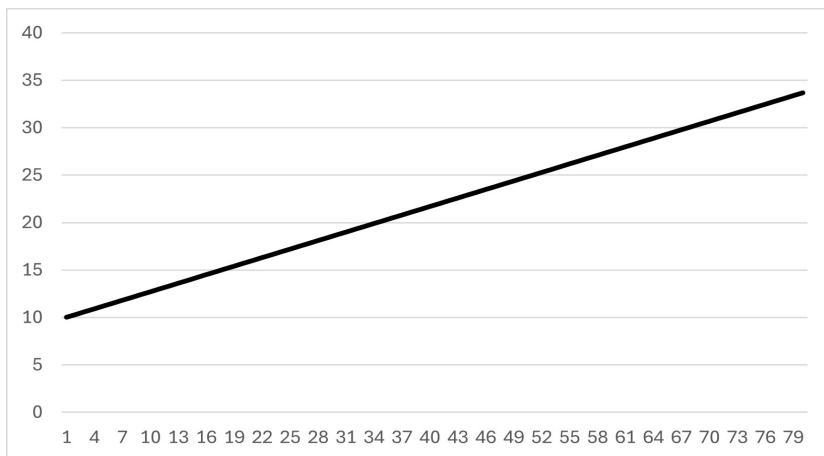


Figura 1.4. Componente estacional

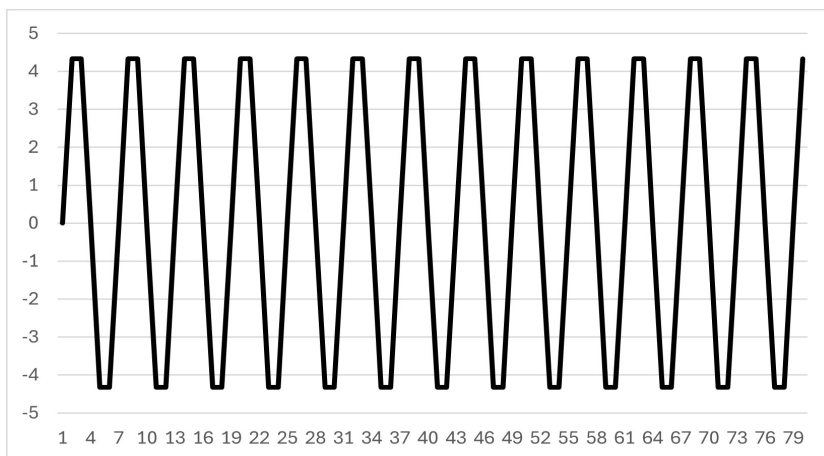
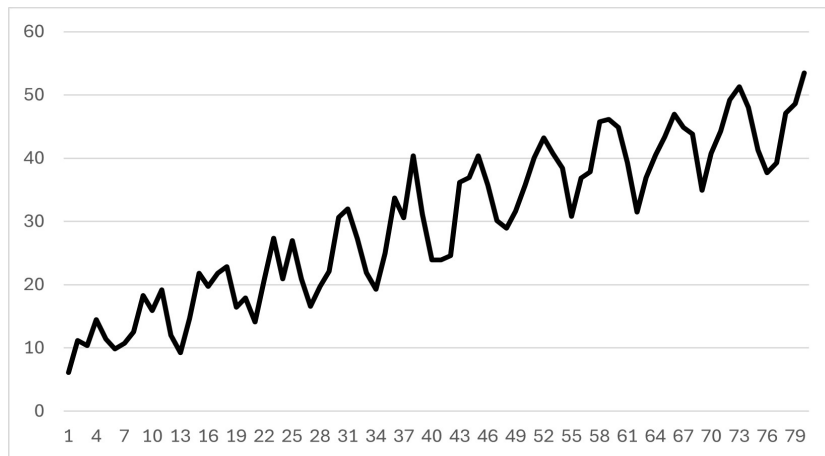


Figura 1.5.  $X_t = m_t + S_t + u_t$



En resumen, nuestro propósito es estudiar, ya sea con la técnicas de regresión o usando la metodología Box-Jenkins, la evolución en el tiempo de una variable aleatoria. El objetivo último es conocer el comportamiento del proceso (el modelo) y hacer pronósticos.

Cada una de las instituciones mencionadas a continuación cuenta con repositorios de información amplios y diversos donde el lector interesado puede encontrar datos susceptibles de tratamiento con las técnicas utilizadas en este libro: Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), Banco Mundial, Banco Interamericano de Desarrollo (BID), Fondo Monetario Internacional (FMI), Organización Internacional del Trabajo (OIT), Organización Mundial de la Salud (OMS), Fondo de las Naciones Unidas para la Infancia (Unicef), Organización de las Naciones Unidas (ONU), entre otras. Invitamos al lector a descargar datos compatibles con cada uno de los modelos estudiados y a realizar el ajuste correspondiente.

### 2.1. Distribuciones de probabilidad discretas

Se dice que una variable aleatoria  $X$  es discreta cuando toma valores en un conjunto contable (posiblemente infinito); por ejemplo, cuando  $X$  describe el resultado del lanzamiento de un dado cuyos lados están numerados del 1 al 6. Por el contrario, si  $X$  puede tomar cualquier valor en un intervalo (también, posiblemente infinito), se dice que es continua, como en el caso de la estatura de una persona.

En el caso discreto, la variable aleatoria se puede fijar por medio de la *función de probabilidad*,  $P(X)$ , la cual establece a través de una tabla o una fórmula la probabilidad que corresponde a cada uno de los valores posibles de  $X$ . Por ejemplo, asumiendo una población de tamaño  $N$  y que  $X = x_i$  con probabilidad  $P(X = x_i) = p_i$ ,  $i = 1, 2, \dots, N$ , la variable aleatoria  $X$  queda completamente especificada por la tabla siguiente:

$$\begin{array}{c|cccccc} X & x_1 & x_2 & x_3 & \dots & x_N \\ \hline P(X = x_i) & p_1 & p_2 & p_3 & \dots & p_N \end{array} \quad (2.1)$$

Se debe cumplir que cada probabilidad es no negativa,  $p_i \geq 0$ , y que la suma de todas las probabilidades es 1, es decir,  $\sum_{i=1}^N p_i = 1$ . En el caso más general escribimos  $\sum_{i=-\infty}^{\infty} p_i = 1$ .

### 2.1.1. Distribución degenerada

Se dice que  $X$  tiene distribución degenerada, concentrada en  $x = a$ , si  $P(X = a) = 1$ . En esencia,  $X$  no es una variable aleatoria en el sentido tradicional, sino una constante. Este sería el caso, por ejemplo, si ambos lados de una moneda fueran «cara» o todos los lados de un dado mostraran el mismo número.

### 2.1.2. Distribución de Bernoulli

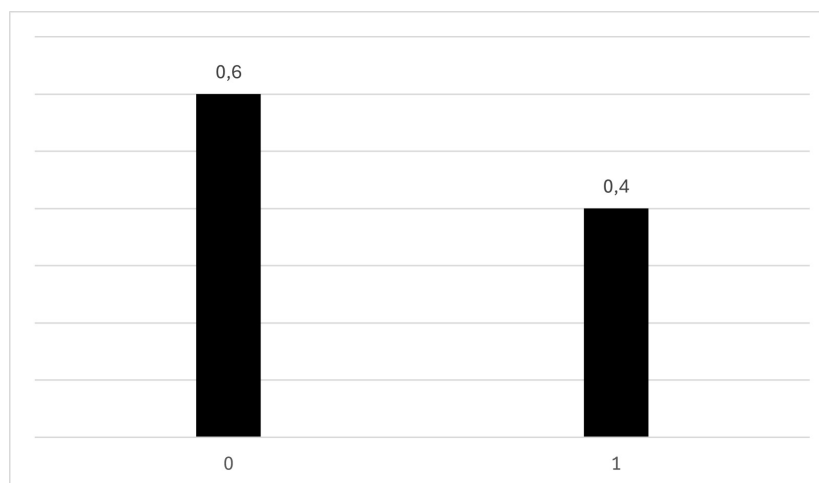
La variable aleatoria  $X$  tiene distribución de Bernoulli con parámetro  $p$  ( $0 < p < 1$ ) si

$$\begin{cases} P(X = 1) = p \\ P(X = 0) = 1 - p. \end{cases} \quad (2.2)$$

Al evento  $X = 1$  se le conoce como «éxito», y a  $X = 0$  como «fracaso». Se denota  $X \sim \text{Bernoulli}(p)$ , y su función de probabilidad se muestra en la figura 2.1 con  $p = 0,4$ .

En una población (no necesariamente de personas) con dos características de interés,  $p$  mide la proporción de éxitos. Por ejemplo, si clasificamos a los habitantes de Santa Marta como samarios y no samarios, y ser samario es la cualidad de interés ( $X = 1$ ), entonces  $p$  es la proporción de samarios en la ciudad.

Figura 2.1. Distribución de Bernoulli



### 2.1.3. Distribución binomial

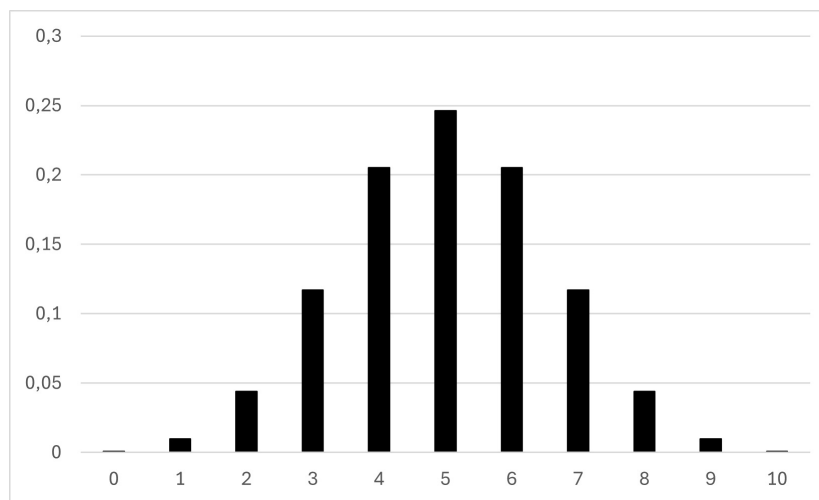
Si  $X_1, X_2, \dots, X_n$  son variables aleatorias independientes e idénticamente distribuidas (iid) y  $X_i \sim \text{Bernoulli}(p)$  para toda  $i$ , entonces

$$Y = \sum_{i=1}^n X_i$$

tiene distribución binomial con parámetros  $n$  y  $p$  y se denota  $Y \sim \text{Binomial}(n, p)$ . Para  $n = 10$  y  $p = 0,5$ , la función de probabilidad correspondiente está dada por la figura 2.2 (Anderson *et al.*, 2011).

La variable  $Y$  mide el número de éxitos en una muestra de tamaño  $n$  tomada de una población de variables con distribución de Bernoulli. En relación con el ejemplo anterior, podemos preguntar la probabilidad de que en una muestra de 50 personas de la ciudad de Santa Marta haya 30 samarios.

Figura 2.2. Distribución binomial:  $n = 10, p = 0,5$



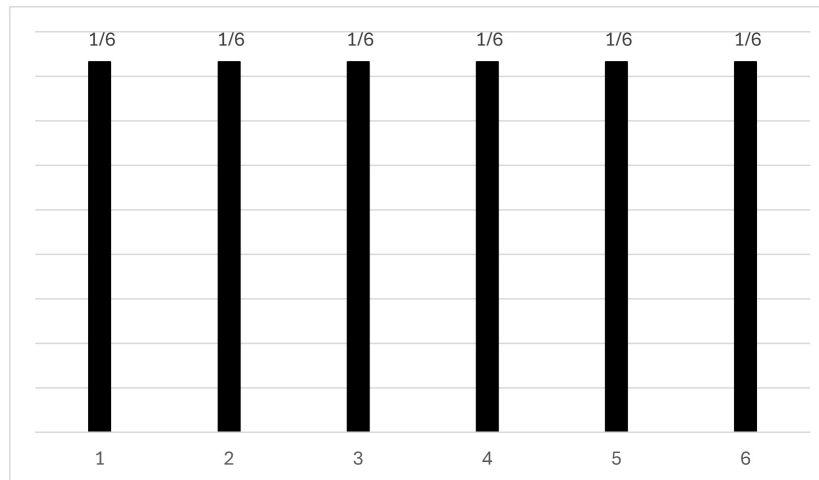
### 2.1.4. Distribución uniforme discreta

$X$  tiene distribución uniforme discreta cuando toma un número finito de valores, todos con la misma probabilidad. En el caso más simple se asume que  $X = i, i = 1, 2, \dots, n$ , y  $P(X = i) = \frac{1}{n}$  (Anderson *et al.*, 2011). Por ejemplo, al lanzar un dado (no cargado) tenemos

$$\begin{array}{l} X \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \\ P(X) \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \end{array} \quad (2.3)$$

o, simplemente,  $P(X = i, i = 1, 2, \dots, 6) = \frac{1}{6}$ . La figura 2.3 muestra la función de probabilidad de  $X$  en este caso.

Figura 2.3. Distribución uniforme discreta



## 2.2. Distribuciones de probabilidad continuas

La contraparte continua de la función de probabilidad es la *función de densidad*  $f(x)$ . Cuando  $X$  es continua, la probabilidad de que  $X$  tome un valor concreto de entre los admisibles es cero. No es el caso de la probabilidad de que  $X$  esté en un intervalo, digamos  $(a, b)$ . Entonces escribimos  $P(X \in (a, b)) = \int_a^b f_X(x)dx$ , que es el área bajo la curva  $f(x)$  en el intervalo especificado. A esto se debe que  $P(X = c) = 0$ , es decir, que entre la curva y el punto dado el área sea cero. También aquí debe cumplirse que  $f_X(x) \geq 0$  y que  $\int_{-\infty}^{\infty} f_X(x)dx = 1$ .

### 2.2.1. Distribución normal

Es la más conocida de las distribuciones continuas. Si la variable  $X$  tiene distribución normal con media  $\mu$  y varianza  $\sigma^2$ , escribimos  $X \sim N(\mu, \sigma^2)$ , y la correspondiente función de densidad de  $X$  resulta ser  $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$ . A menudo es necesario estandarizar  $X$  para obtener una variable aleatoria normal con media  $\mu = 0$  y varianza  $\sigma^2 = 1$ . La variable  $Z = \frac{X-\mu}{\sigma}$  cumple con esos requisitos, y su función de densidad es  $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$  (Anderson *et al.*, 2011). El gráfico de  $\phi(z)$  se muestra en la figura 2.4.