Apuntes de

ESTADÍSTICA

Variable aleatoria

UNIVERSIDAD
DE MÁLAGA

Sixto Sánchez Merino

Dpto. de Matemática Aplicada
Universidad de Málaga

Matemática Aplicada

Mi agradecimiento a los profesores Carlos Cerezo Casermeiro y Carlos Guerrero García, por sus correcciones y sugerencias en la elaboración de estos apuntes.

Apuntes de Estadística

Capítulo 5

Variable aleatoria

Los posibles resultados de un experimento son todos los sucesos que constituyen el espacio muestral. A menudo, nos interesa que estos resultados sean numéricos. En este caso, utilizamos una función que permita clasificar a los sucesos, asignado valores numéricos a cada uno de ellos.

Por ejemplo, si el experimento aleatorio consiste en lanzar tres veces una moneda, entonces el espacio muestral se puede representa por $\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$ donde H al suceso "salir cara" y T es el suceso "salir cruz". Pero si estamos interesados en determinar el número de caras obtenidas en los tres lanzamientos de la moneda, entonces podemos definir una función X que asigna un valor numérico (número de caras) a cada resultado del experimento. De esta manera, tenemos, por ejemplo, que $X(HTH) = 2$ o que $X(TTT) = 0$. Este tipo de funciones, cuyos valores dependen de los resultados de un experimento aleatorio, se llaman variables aleatorias.

Las variables aleatorias y sus distribuciones de probabilidad, pueden considerarse una generalización del concepto frecuentista de probabilidad. Se introducen como el modelo matemático ideal al que se aproximan las distribuciones de frecuencias que se obtendrían en una repetición indefinida de pruebas de este experimento. Por ello, nos recuerdan a las variables estadísticas y a sus distribuciones de frecuencia que ya hemos estudiado en estadística descriptiva.

Las variables aleatorias se clasifican conforme al rango de valores que pueden asumir, y llamaremos *soporte* a ese conjunto de posibles valores (números reales) que puede tomar una variable aleatoria.

En este capítulo estudiaremos principalmente las variables aleatorias discretas, cuyo soportes está formado por un número finito, o infinito numerable de valores (p.e. número de defectos en una inspección de productos, número de elementos en espera en una cola, etc.) y las variables aleatorias continuas cuyo soporte es un intervalo o conjunto de intervalos de números reales (p.e. durabilidad de un dispositivo, velocidad de un automóvil, resistencia a la tensión de una nueva aleación, etc.).

Además, al final del tema, estudiaremos las variables aleatorias bidimensionales y algunos aspectos asociados relativos a las distribuciones, medidas y regresión. Su analogía con ellas, nos recordará lo estudiado en el tema de regresión y correlación, sin más que cambiar la frecuencia por la probabilidad, en la mayoría de las fórmulas.

5.1. Variable aleatoria unidimensional

Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico asociado a un experimento aleatorio. Una variable aleatoria X es una función definida sobre el espacio muestral Ω (conjunto de resultados de un experimento aleatorio) que toma valores en un conjunto de número reales, llamado *soporte*, y que denotaremos por S_x. Se suelen utilizar las abreviaturas "v.a.u." o simplemente "v.a." para referirse a las variables aleatorias unidimensionales.

En términos matemáticos precisos, una variable aleatoria unidimensional es una aplicación $X : \Omega \to \mathbb{R}$ que verifica la siguiente propiedad:

$$\text{para todo } x \in \mathbb{R} \text{ el conjunto } \{\omega \in \Omega \,/\, X(\omega) \leq x\} \in \mathcal{A}.$$

Ejemplo 5.1 *Utilice una variable aleatoria para modelizar el experimento que consiste en lanzar dos veces un dado y anotar la suma de las puntuaciones obtenidas.*

El espacio muestral del experimento que consiste en lanzar dos veces un dado se puede representar así:

$$\Omega = \{(11), (12), (13), (14), (15), (16), (21), (22), \ldots, (66)\}$$

y nos permite considerar la variable aleatoria X que suma el valor de las puntuaciones obtenidas en los dos dados:

$$X : \Omega \to \mathbb{R} \quad \text{tal que} \quad X(ij) = i + j \quad \text{siendo} \quad (ij) \in \Omega$$

Así, por ejemplo, $X(11) = 2$, $X(36) = 9$ ó $X(66) = 12$, de manera que el soporte de esta variable es el conjunto

$$S_x = \{2, 3, 4, 5, ..., 12\}$$

y, por lo tanto, la variable aleatoria X es discreta. □

5.2. Función de distribución

Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico asociado a un experimento aleatorio y sea X una variable aleatoria. Definimos la función de distribución $F : \mathbb{R} \to [0, 1]$, asociada a la variable aleatoria X, de la siguiente manera:

$$F(x) = P(\{\omega \in \Omega \,/\, X(\omega) \leq x\}) = P(X \leq x) \quad \text{para todo} \quad x \in \mathbb{R}$$

La función de distribución es única para cada variable aleatoria a la que caracteriza, resulta especialmente útil para calcular probabilidades ya que:

- $P(X \leq x) = F(x)$
- $P(X > x) = 1 - P(X \leq x) = 1 - F(x)$
- $P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1)$

y sus principales propiedades son:

1. $0 \leq F(x) \leq 1$ para todo $x \in \mathbb{R}$
2. $F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0$
3. $F(\infty) = \lim\limits_{x \to \infty} F(x) = F(\infty) = 1$
4. F es monótona no decreciente, es decir, si $x_1 < x_2$ entonces $F(x_1) \leq F(x_2)$.
5. F es continua por la derecha, es decir, $\lim\limits_{h \to 0^+} F(x+h) = F(x)$

Veamos ahora que forma tiene esta función de distribución en cada uno de los tipos de variables (discretas y continuas) que vamos a estudiar.

5.3. Variable aleatoria discreta

Una variable aleatoria X se dice que es *discreta* si el soporte S_x es un conjunto discreto, es decir, cuando la variable X toma un número finito o infinito numerable de valores reales. Por ejemplo, el número de defectos observados en un control de calidad o el número de elementos que esperan en una cola son variables aleatorias discretas. Se suelen utilizar la abreviatura "v.a.d." para referirse a las variables aleatorias discretas.

A continuación, haremos corresponder una probabilidad a cada valor de la variable aleatoria, lo cual constituye la distribución de probabilidad de la variable aleatoria, que nos recuerda a las distribuciones de frecuencias asociadas a las variables estadísticas.

5.3.1. Distribución de probabilidad

Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico asociado a un experimento aleatorio y sea X una variable aleatoria discreta que toma los valores en el conjunto $S_x = \{x_1, x_2, x_3, ...\}$. Definimos la probabilidad $p(x_i)$ para cada uno de los elementos del soporte, de la siguiente manera:

$$p(x_i) = P(\{\omega \in \Omega \ / \ X(\omega) = x_i\}) = P(X = x_i)$$

La distribución de probabilidad de la variable X está constituida por los elementos del soporte S_x junto a sus correspondientes valores de probabilidad. Normalmente, se representa en forma de tabla, de la siguiente manera:

x	$p(x)$
x_1	$p(x_1)$
x_2	$p(x_2)$
$\vdots$	$\vdots$
x_n	$p(x_n)$
$\vdots$	$\vdots$

La representación gráfica de la distribución de probabilidad se realiza en un diagrama de barras. En el eje OX se representan los distintos elementos del soporte, y en el eje OY se representa la probabilidad correspondiente a cada uno de ellos.

Ejemplo 5.2 *Consideramos el experimento consistente en lanzar una moneda tres veces al aire. Definimos la variable aleatoria X que determina el número de caras (H) que aparecen en cada serie de tres lanzamientos. Obtener y representar su distribución de probabilidad.*

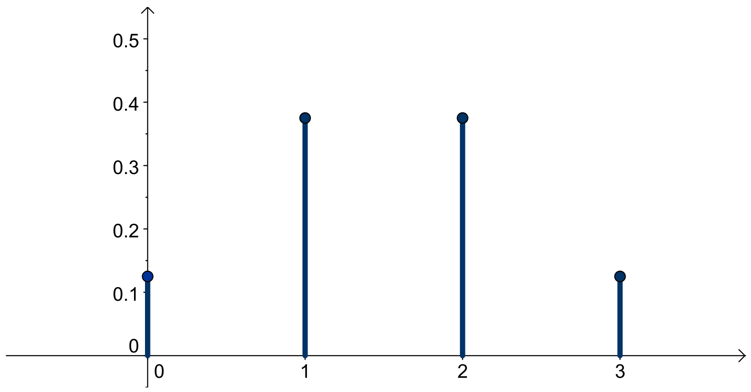
La variable X toma los valores 0, 1, 2 y 3, que constituyen el soporte. Para calcular la probabilidad de cada uno de ellos recurrimos a los sucesos correspondientes:

$$\begin{aligned} p(0) = P(X=0) &= P(\{TTT\}) = 1/8 \\ p(1) = P(X=1) &= P(\{TTH, THT, HTT\}) = 3/8 \\ p(2) = P(X=2) &= P(\{THH, HTH, HHT\}) = 3/8 \\ p(3) = P(X=3) &= P(\{HHH\}) = 1/8 \end{aligned}$$

Por tanto la distribución de probabilidad de la variable X es:

x	$p(x)$
0	1/8
1	3/8
2	3/8
3	1/8

y su representación gráfica mediante diagrama de barras es:



Obsérvese la analogía de la distribución de probabilidad de esta variable aleatoria discreta X, con las distribuciones de frecuencia estudiadas en el tema de estadística descriptiva. □

A continuación vamos a definir los conceptos de función de distribución, media, varianza y momentos de una variable aleatoria discreta a partir de su distribución de probabilidad. Por analogía, usando el concepto frecuentista de la probabilidad, podríamos definir el resto de las medidas de centralización, dispersión, simetría y apuntamiento tal y como se hizo en el tema de estadística descriptiva.

Para las definiciones que siguen a continuación, consideraremos una variable aleatoria discreta X que toma los valores en el conjunto $S_x = \{x_1, x_2, \dots\}$ con probabilidades $p(x_1), p(x_2), \dots$ Si el número de valores que toma la variable es infinito numerable es necesario asegurarse de que las series correspondientes, que aparecen en las fórmulas, son absolutamente convergentes.

5.3.2. Función de distribución

Dada una variable aleatoria discreta X, para todo número real x se define su *función de distribución* asociada $F(x)$, de la siguiente manera:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$$

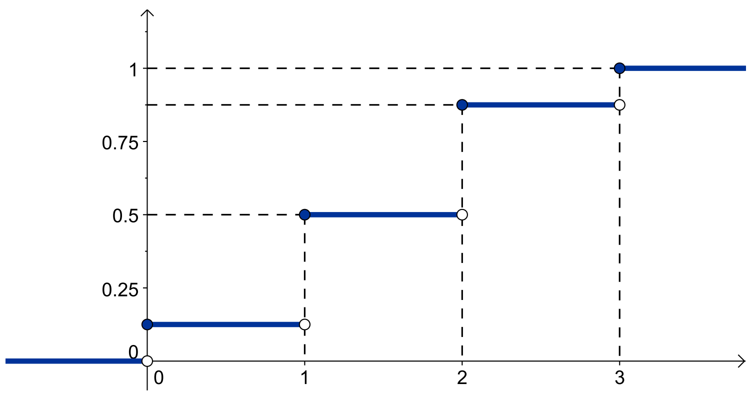
Gráficamente, esta función "acumulativa" adopta una forma de escalera, donde los saltos se producen en los puntos del soporte, siendo $F(x)$ continua por la derecha, en cada uno de ellos. Además, la altura del salto en cada punto corresponde con la probabilidad de que la variable tome ese valor.

Ejemplo 5.3 *Obtener y representar la función de distribución de la variable aleatoria definida en el ejemplo 5.2 de la página 176.*

La función de distribución de la variable aleatoria que determina el número de caras que aparecen en cada serie de tres lanzamientos de una moneda perfecta es:

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ 1/8 & \text{si} \quad 0 \leq x < 1 \\ 4/8 & \text{si} \quad 1 \leq x < 2 \\ 7/8 & \text{si} \quad 2 \leq x < 3 \\ 1 & \text{si} \quad x \geq 3 \end{cases}$$

y su representación gráfica es



que tiene forma de escalera donde los saltos se producen en los puntos del soporte, y la altura del salto corresponde con la probabilidad en el punto. □

5.3.3. Función generatriz de probabilidad

Cuando el soporte de una variable aleatoria discreta X es el conjunto $\mathbb{N} = \{0, 1, 2, \dots\}$, podemos definir la *función generatriz de probabilidad* de la variable aleatoria X como la serie de potencias

$$G(s) = \sum_{n=0}^{\infty} s^n \cdot p_n \qquad \text{con} \qquad s \in (-1, 1)$$

donde $p_n = p(n) = P(X = n)$. Además, se suelen utilizar la abreviatura "f.g.p." para referirse a la función generatriz de probabilidad.

La función generatriz es infinitamente derivable y nos permite obtener una de las propiedades más importantes conocida como *teorema de inversión* que establece la distribución de probabilidad de la variable aleatoria X en términos de su función generatriz:

$$p_n = \frac{G^{(n}(0)}{n!} \qquad \text{para todo } n = 0, 1, 2, \dots$$

Ejemplo 5.4 *Sea X la variable aleatoria discreta que determina el número de caras (H) antes de obtener la primera cruz (T) en lanzamientos consecutivos de una misma moneda equilibrada. Determine la función generatriz de probabilidad, compruebe que se verifica el teorema de inversión y responda a la siguiente pregunta: ¿cómo cambiaría esta función si la moneda estuviese trucada con probabilidad p de salir cara?*

El soporte de la variable X es el conjunto de los números naturales $S_x = \{0, 1, 2, \dots\}$ y su distribución de probabilidad se determina así:

$$p_n = P(HH \overset{(n}{\dots} HT) = P(H) \cdot P(H) \cdot \overset{(n}{\dots} \cdot P(H) \cdot P(T) = \left(\frac{1}{2}\right)^n \frac{1}{2} = \left(\frac{1}{2}\right)^{n+1}$$

siendo $p_n = p(n) = P(X = n)$. Por lo tanto, la función generatriz de probabilidad de X es:

$$G(s) = \sum_{n=0}^{\infty} s^n \cdot p_n = \sum_{n=0}^{\infty} s^n \frac{1}{2}\left(\frac{1}{2}\right)^n = \sum_{n=0}^{\infty} \frac{1}{2}\left(\frac{s}{2}\right)^n \overset{*}{=} \frac{1}{2-s}$$

Obsérvese (*) que para obtener la expresión explícita de la función en términos de funciones elementales, hemos utilizado que la serie de potencias correspondiente era una serie geométrica convergente para $s \in (-1, 1)$.

Calculando las derivadas sucesivas de la función generatriz de probabilidad $G(s)$, se puede comprobar que

$$G^{(n}(s) = \frac{n!}{(2-s)^{n+1}} \qquad \longrightarrow \qquad G^{(n}(0) = \frac{n!}{2^{n+1}}$$

y, por lo tanto,

$$p_n = \frac{G^{(n}(0)}{n!} = \frac{n!/2^{n+1}}{n!} = \left(\frac{1}{2}\right)^{n+1}$$

que pone de manifiesto el teorema de inversión.

Si la moneda estuviese trucada y fuese p la probabilidad de salir cara, entonces

$$p_n = P(HH \overset{(n}{\dots} HT) = P(H) \cdot P(H) \cdot \overset{(n}{\dots} \cdot P(H) \cdot P(T) = p^n(1-p)$$

y, por lo tanto, la función generatriz de probabilidad sería:

$$G(s) = \sum_{n=0}^{\infty} s^n \cdot p_n = \sum_{n=0}^{\infty} s^n p^n (1-p) = \frac{1-p}{1-ps}$$

para todo $s \in (-1, 1)$. □

5.4. Variable aleatoria continua

Muchas variables aleatorias que se observan en la vida real no son discretas porque pueden tomar cualquier valor en un intervalo de números, o en uniones de ellos. Por ejemplo, el tiempo de espera en una cola, la durabilidad de un componente electrónico, la velocidad de un automóvil o la resistencia a la tensión de una nueva aleación. A las variables de este tipo las definiremos como *variables aleatorias continuas.*

Matemáticamente, una variable aleatoria X se dice que es *continua* si su función de distribución $F(x)$ correspondiente es continua. Se suelen utilizar la abreviatura "v.a.c." para referirse a las variables aleatorias continuas.

Asociada a cada variable aleatoria continua, existe una función, llamada *función de densidad* que determina la distribución de probabilidad de la variable aleatoria. Veamos, en primer lugar, esta función de densidad y, después, estudiaremos la función de distribución y la relación entre ambas funciones.

5.4.1. Función de densidad

Dada una variable aleatoria continua X, decimos que una función real $f(x)$, integrable y no negativa, es la *función de densidad de probabilidad* (o simplemente *función de densidad*) *de la variable aleatoria* X si el área encerrada entre la curva y el eje OX es igual a la unidad y, además, la probabilidad de que X se encuentre entre dos valores x_1 y x_2 con $x_1 \leq x_2$ es igual al área comprendida entre estos dos valores, es decir,

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \qquad \text{y} \qquad P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)\, dx$$

Y además, podemos calcular la probabilidad de que la variable tome valores en cualquier otro intervalo. Por ejemplo,

$$P(X \leq x_1) = \int_{-\infty}^{x_1} f(x)\, dx \qquad \text{y} \qquad P(X > x_2) = \int_{x_2}^{\infty} f(x)\, dx$$

El soporte de una variable aleatoria continua es el conjunto de números reales donde la función de densidad $f(x)$ sea estrictamente positiva. Si este soporte es un intervalo, por ejemplo $S_x = (a, b)$, entonces las integrales impropias se reducen a integrales definidas. De esta manera

$$\int_a^b f(x)\mathrm{d}x = 1$$

y si c es un número comprendido entre a y b ($a < c < b$) entonces

$$P(X \leq c) = \int_a^c f(x)\,dx \qquad \text{o bien} \qquad P(X \geq c) = \int_c^b f(x)\,dx$$

Obsérvese que la probabilidad de que una variable aleatoria continua tome un valor particular es cero, aunque sea posible. Es decir, la probabilidad medirá intervalos de ocurrencia de la variable, no instancias puntuales. Por lo tanto, no será relevante que una desigualdad sea o no estricta. Por ejemplo $P(X \leq x) = P(X < x)$ y $P(X \geq x) = P(X > x)$ o bien

$$P(x_1 < X < x_2) = P(x_1 \leq X < x_2) = P(x_1 < X \leq x_2) = P(x_1 \leq X \leq x_2)$$

5.4.2. Función de distribución

Dada una variable aleatoria continua X, a la función acumulativa

$$F(x) = P(X \leq x)$$

se la denomina *función de distribución de* X, y su representación gráfica corresponde a una función continua, creciente, definida en el intervalo $(-\infty, \infty)$ y con asíntotas horizontales para los valores de $y = 0$ e $y = 1$.

La función de distribución se define en términos de la función de densidad, de la siguiente manera:

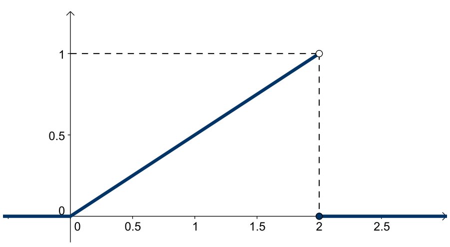
$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

y por tanto, en los valores de x donde exista la derivada de $F(x)$, se verifica la igualdad

$$f(x) = F'(x)$$

que relaciona las funciones de distribución y densidad.

Ejemplo 5.5 *Consideremos la variable aleatoria* X *que determina la duración en unidades de tiempo (u.t.) de un componente electrónico y cuya función de densidad viene representada en el siguiente gráfico:*



Determinar y representar su función de distribución, y calcular las probabilidades de que el componente dure más de 1 u.t., exactamente 1 u.t. y más de una unidad de tiempo sabiendo que dura menos de 1'5 u.t.

A la vista de la representación gráfica, deducimos que la función de densidad es

$$f(x) = \begin{cases} x/2 & \text{si } 0 < x < 2 \\ 0 & \text{en el resto} \end{cases}$$

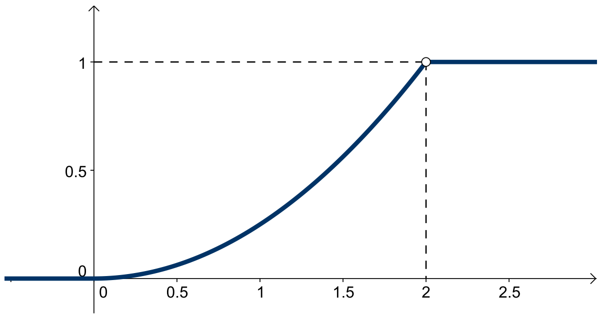
y, a partir de ella, podemos calcular la función de distribución $F(x) = P(X \leq x)$ de la siguiente manera

$$\begin{array}{llll} \text{Si} & x < 0 & \text{entonces} & F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = 0 \\ \text{Si} & 0 \leq x < 2 & \text{entonces} & F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x} \frac{t}{2}\,dt = \frac{x^2}{4} \\ \text{Si} & x \geq 2 & \text{entonces} & F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{2} \frac{t}{2}\,dt = 1 \end{array}$$

y, por lo tanto,

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ x^2/4 & \text{si} \quad 0 \leq x < 2 \\ 1 & \text{si} \quad x \geq 2 \end{cases}$$

cuya gráfica es



que representa una función continua pues corresponde a una variable aleatoria continua.

Ahora utilizamos estas dos funciones (densidad y distribución) para calcular las probabilidades. En primer lugar, calculamos la probabilidad de que el componente dure más de 1 u.t.

$$P(X > 1) = \int_{1}^{\infty} f(x)\,dx = \int_{1}^{2} \frac{x}{2}\,dx = \frac{3}{4}$$

o bien

$$P(X > 1) = 1 - P(X \leq 1) = 1 - F(1) = 1 - \frac{1}{4} = \frac{3}{4}$$

En segundo lugar, la probabilidad de que el componente dure exactamente 1 u.t. es cero pues la variable aleatoria es continua. Y, por último, la probabilidad de que el componente dure más de 1 u.t. sabiendo que dura menos de 1'5 u.t., es una probabilidad condicionada que se calcula así:

$$P(X > 1 \,|\, X < 1'5) = \frac{P(1 < x < 1'5)}{P(X < 1'5)} = \frac{F(1'5) - F(1)}{F(1'5)} = \frac{9/16 - 1/4}{9/16} = \frac{5}{9}$$

Obsérvese que las probabilidades se han calculado utilizando la función de distribución. □

5.5. Esperanza matemática y otras medidas

En esta sección vamos a introducir el concepto de esperanza matemática que permite definir los momentos de una variable aleatoria. La analogía con las variables estadísticas nos permitirá deducir las principales medidas de centralización, dispersión, simetría y apuntamiento.

Para las definiciones que siguen consideraremos la variable aleatoria X con soporte S_x.

5.5.1. Esperanza matemática

Para definir la *esperanza matemática* (o simplemente, *esperanza*) distinguiremos entre variables aleatorias discretas y continuas.

- Si X es una variable aleatoria discreta entonces su esperanza matemática es:

$$E\left[X\right] = \sum_{x_i \in S_x} x_i \cdot p(x_i)$$

- Si X es una variable aleatoria continua entonces su esperanza matemática es:

$$E\left[X\right] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

La esperanza matemática está definida a partir de una serie (en el caso discreto) o de una integral impropia (en el caso continuo), de manera que la esperanza matemática no existe, o no se puede definir, si la serie o integral correspondiente no es convergente. Por lo tanto, algunas de las definiciones que se presentan a continuación, donde interviene la esperanza matemática, están condicionadas a la existencia de esta esperanza matemática.

5.5.2. Momentos

Se llama *momento de orden k respecto del parámetro c*, y se denota por $M_k(c)$, a la esperanza matemática de la variable $(X-c)^k$, es decir

$$M_k(c) = E\left[(X-c)^k\right]$$

Y en función de que la variable aleatoria sea discreta o continua, se define respectivamente así:

$$M_k(c) = \sum_{x_i \in S_x} (x_i - c)^k \cdot p(x_i) \qquad , \qquad M_k(c) = \int_{-\infty}^{\infty} (x-c)^k \cdot f(x)\, dx$$

Como casos particulares, y por su importancia, se definen los dos siguientes tipos de momentos:

- Si $c = 0$ tenemos los *momentos ordinarios* que representamos por m_k y definidos como $m_k = E\left[X^k\right]$ para cada uno de los tipos de variables, discretas y continuas, respectivamente:

$$m_k = \sum_{x_i \in S_x} x_i^k \cdot p(x_i) \qquad , \qquad m_k = \int_{-\infty}^{\infty} x^k \cdot f(x)\, dx$$

- Si $c = \mu_x$ tenemos los *momentos centrales* que representamos por μ_k y definidos como $\mu_k = E\left[(X-\mu_x)^k\right]$ para cada uno de los tipos de variables, discretas y continuas, respectivamente:

$$\mu_k = \sum_{x_i \in S_x} (x_i - \mu_x)^k \cdot p(x_i) \qquad , \qquad \mu_k = \int_{-\infty}^{\infty} (x-\mu_x)^k \cdot f(x)\, dx$$

Los momentos son de gran importancia porque forman parte de la definición de muchas medidas, por ejemplo, la media, la varianza o los coeficientes de asimetría o aplastamiento. Veamos, ahora, una función asociada a cada variable aleatoria que la caracteriza porque permite calcular sus momentos ordinarios.

5.5.3. Función generatriz de momentos

La función generatriz de probabilidad sólo se define para variables aleatorias discretas que toman valores en $\mathbb{N}$. Por lo tanto, se hace necesario definir una función más general, asociada a cualquier tipo de variable aleatoria, continua o discreta, y que caracterice la distribución de probabilidad de esa variable.

Sea X una variable aleatoria. Se define la *función generatriz de momentos* asociada a la variable X como la función

$$M(t) = E(\mathrm{e}^{tX})$$

siempre que la esperanza exista en un entorno del cero $(-t_0, t_0)$. Se suelen utilizar la abreviatura "f.g.m." para referirse a la función generatriz de momentos, y según sea discreta o continua la variable aleatoria la expresión, respectivamente será:

$$M(t) = \sum_{x_i \in S_x} \mathrm{e}^{tx_i} \cdot p(x_i) \qquad , \qquad M(t) = \int_{-\infty}^{\infty} \mathrm{e}^{tx} \cdot f(x)\, dx$$

Si X es una variable aleatoria con función generatriz de momentos $M(t)$ que es finita para $|t| < t_0$ con $t_0 > 0$, entonces X posee momentos ordinarios de todos los órdenes y además

$$E(X^n) = M^{(n}(0)$$

Esta propiedad justifica el nombre de esta función generatriz pues determina los momentos ordinarios a partir de las derivadas sucesivas de la función en el cero.

Ejemplo 5.6 *Sea X la variable aleatoria discreta, definida en el ejemplo 5.4 de la página 178, que determina el número de caras antes de obtener la primera cruz en el lanzamiento de una moneda equilibrada. Determine la función generatriz de momentos y, a partir de ella, calcule los momentos ordinarios de pimer y segundo orden.*

En el ejemplo 5.4 determinamos que la distribución de probabilidad de la variable X era:

$$p(n) = \left(\frac{1}{2}\right)^{n+1} \qquad \text{para todo} \quad n \in \mathbb{N}$$

y, por lo tanto, su función generatriz de momentos es:

$$M(t) = \sum_{n=0}^{\infty} \mathrm{e}^{tn} p(n) = \sum_{n=0}^{\infty} \mathrm{e}^{tn} \left(\frac{1}{2}\right)^{n+1} = \sum_{n=0}^{\infty} \frac{1}{2} \left(\frac{\mathrm{e}^t}{2}\right)^n = \frac{1}{2-\mathrm{e}^t}$$

Ahora, calculando el valor en 0 de las derivadas sucesivas de esta función, obtenemos los momentos ordinarios. Por ejemplo, la primera derivada

$$M'(t) = \frac{\mathrm{e}^t}{(2-\mathrm{e}^t)^2} \qquad \longrightarrow \qquad M'(0) = 1$$

determina que $E[X] = 1$ que, como veremos, corresponde a la media de la variable. □

5.5.4. Medidas de posición

A continuación, definimos las principales medidas de posición.

Media

La esperanza matemática de la variable aleatoria X recibe el nombre de *media* de la variable y se denota por $\bar{x}$, o bien, μ_x. La estructura de su fórmula y la interpretación de su valor es similar a la media aritmética definida en estadística descriptiva pero sustituyendo las frecuencias relativas (de los datos que se han observado) por la probabilidad de los valores de la variable (resultados posibles).

El comportamiento de la esperanza respecto de las transformaciones lineales es el siguiente:

$$\text{Si} \qquad Y = a + bX \qquad \text{entonces} \qquad E[Y] = a + bE[X]$$

Moda

La *moda* de una variable aleatoria X es el valor del soporte que tiene mayor probabilidad (variable discreta) o densidad (variable continua).

Cuantiles

El cuantil de orden k de una variable aleatoria X es el punto c_k del soporte que verifica las dos siguientes condiciones:

$$P(X \leq c_k) \geq k \qquad \text{y} \qquad P(X \geq c_k) \geq 1-k \qquad \text{con} \qquad 0 < k < 1$$

que pueden resumirse en la siguiente condición

$$k \leq F(c_k) \leq k + P(X = c_k)$$

y que, en el caso de una variable continua, equivale a $F(c_k) = k$.

En general, el cuantil de orden k no es único y, además, si c_k y c'_k son dos cuantiles de orden k de una misma variable aleatoria, con $c_k < c'_k$, entonces cualquier valor del intervalo (c_k, c'_k) es también, un cuantil de orden k.

A partir de esta definición, y por analogía a la definición de las medidas de estadística descriptivas, podemos considerar los distintos cuantiles (cuartiles, deciles y percentiles).

Como caso particular, definimos la *mediana* de una variable aleatoria X como el punto "Me" que verifica las dos siguientes condiciones: $P(X \leq \text{Me}) \geq 1/2$ y $P(X \geq \text{Me}) \geq 1/2$. Obsérvese que si X es una variable aleatoria continua entonces la mediana verifica que $F(\text{Me}) = 1/2$.

Ejemplo 5.7 *Calcule la media, la mediana y la moda de la variable aleatoria discreta definida en el ejemplo 5.2 de la página 176.*

Ejemplo 5.8 *Calcule la media y la mediana de la variable aleatoria continua definida en el ejemplo 5.5 de la página 180, y compruebe que no existe la moda de esta distribución.*

5.5.5. Medidas de dispersión

A continuación, definimos las principales medidas de dispersión y veremos que la estructura de sus fórmulas y la interpretación de sus valores son similares a las de sus homónimos en estadística descriptiva pero sustituyendo las frecuencias relativas (de los datos que se han observado) por la probabilidad de los valores de la variable (resultados posibles).

Rangos

El *rango* de una variable aleatoria es la diferencia entre los valores extremos del soporte, si son finitos, e infinito, en otro caso.

A partir de los cuantiles, también podemos definir los rangos intercuartílico, interdecílico e intercentílico, análogamente a como se define en estadística descriptiva.

Varianza y desviación típica

La varianza de una variable aleatoria X, que denotaremos por σ_x^2, o bien, por $V[X]$, se define como el momento central de orden 2, es decir

$$\sigma_x^2 = V[X] = E\left[(X - E[X])^2\right]$$

Según sea discreta o continua la variable aleatoria la expresión, respectivamente será:

$$\sigma_x^2 = \sum_{x_i \in S_x} (x_i - \bar{x})^2 \cdot p(x_i) \qquad , \qquad \sigma_x^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 \cdot f(x)\, dx$$

Si desarrollamos el cuadrado y aplicamos las propiedades de la esperanza (serie o integral) obtenemos la siguiente fórmula:

$$\sigma_x^2 = V[X] = E\left[X^2\right] - (E[X])^2$$

que permite calcular la varianza de una manera más sencilla.

También se define la *desviación típica* de una variable aleatoria discreta X como la raíz cuadrada de la varianza. Según sea discreta o continua la variable aleatoria la expresión, respectivamente será:

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{\sum_{x_i \in S_x} (x_i - \bar{x})^2 \cdot p(x_i)} \qquad , \qquad \sigma_x = \sqrt{\sigma_x^2} = \sqrt{\int_{-\infty}^{\infty} (x - \bar{x})^2 \cdot f(x)\, dx}$$

El comportamiento de la varianza respecto de las transformaciones lineales es el siguiente:

$$\text{Si} \qquad Y = a + bX \qquad \text{entonces} \qquad V[Y] = b^2 V[X]$$

Coeficiente de variación

A partir de los conceptos de media ($\bar{x}$) y desviación típica (σ_x) de una variable aleatoria X, se define el coeficiente de variación de la siguiente manera:

$$CV(X) = \frac{\sigma_x}{|\bar{x}|}$$

siempre que la media de la variable sea distinta de cero.

Este coeficiente nos permite comparar la dispersión de dos variables aleatorias.

Ejemplo 5.9 *Calcule el rango intercuartílico y el coeficiente de variación de la variable aleatoria discreta definida en el ejemplo 5.2 de la página 176.*

Ejemplo 5.10 *Calcule el rango intercuartílico y el coeficiente de variación de la variable aleatoria continua definida en el ejemplo 5.5 de la página 180.*

5.5.6. Medidas de forma

La simetría y el apuntamiento de una variable aleatoria se estudia de manera similar al de una variable estadística y para medir ambas características se utilizan los mismos coeficientes adimensionales, con sus mismas interpretaciones en función de su valor.

Coeficiente de asimetría

Para medir la simetría de una variable aleatoria X se define el coeficiente de asimetría de Fisher, que se denota por g_1, de la siguiente manera:

$$g_1 = \frac{\mu_3}{\sigma^3}$$

Coeficiente de aplastamiento

Para medir la curtosis de una variable aleatoria X se define el coeficiente de aplastamiento de Fisher, que se denota por g_2, de la siguiente manera:

$$g_2 = \frac{\mu_4}{\sigma^4} - 3$$

Ejemplo 5.11 *Estudiar la simetría y la curtosis de la variable aleatoria discreta definida en el ejemplo 5.2 de la página 176.*

Ejemplo 5.12 *Estudiar la simetría y la curtosis de la variable aleatoria continua definida en el ejemplo 5.5 de la página 180.*

5.6. Variable aleatoria bidimensional

Vamos a generalizar el concepto de variable aleatoria y de función de distribución para considerar el estudio conjunto de dos variables aleatorias. Los resultados obtenidos reflejan un paralelismo con los contenidos del tema de regresión y correlación.

Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilizable y sean X e Y dos variables aleatorias definidas sobre ese espacio. Una *variable aleatoria bidimensional* es una aplicación $(X,Y) : \Omega \to \mathbb{R}^2$ que verifica la siguiente propiedad:

$$\text{para todo } (x,y) \in \mathbb{R}^2 \text{ el conjunto } \{\omega \in \Omega \,/\, X(\omega) \leq x, Y(\omega) \leq y\} \in \mathcal{A}.$$

Ejemplo 5.13 *Consideremos el experimento consistente en lanzar una moneda al aire tres veces. Sea X la variable aleatoria que determina el número de caras (H) obtenidas, y sea Y la variable aleatoria que toma los valores 0, si la primera vez salió cara (H), y 1, si la primera vez salió cruz (T). Determine la variable aleatoria bidimensional (X,Y).*

El espacio muestral del experimento consistente en lanzar tres veces una moneda se puede representar así:

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

y, para cada uno de estos sucesos, representamos el valor de la variable (X,Y) en la siguiente tabla:

$\omega \in \Omega$	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$(X,Y)(\omega)$	$(3,0)$	$(2,0)$	$(2,0)$	$(2,1)$	$(1,0)$	$(1,1)$	$(1,1)$	$(0,1)$

que corresponde a una variable aleatoria discreta cuyo soporte es $S_{xy} = \{0,1,2,3\} \times \{0,1\}$. □

5.6.1. Función de distribución

Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilizable y sean (X,Y) una variable aleatoria bidimensional definida en ese espacio. Llamaremos *función de distribución conjunta* de la variable (X,Y) a la función $F : \mathbb{R}^2 \to \mathbb{R}$ definida por

$$F(x,y) = P(X \leq x, Y \leq y)$$

Las propiedades de esta función de distribución conjunta son similares a las de la función de distribución de una variable aleatoria unidimensional:

1. $F(-\infty, -\infty) = \lim\limits_{(x,y)\to(-\infty,-\infty)} F(x, y) = 0$, y además,

   a) $F(-\infty, y) = \lim\limits_{x\to-\infty} F(x, y) = 0$ para todo $y \in \mathbb{R}$, y

   b) $F(x, -\infty) = \lim\limits_{y\to-\infty} F(x, y) = 0$ para todo $x \in \mathbb{R}$.

2. $F(\infty, \infty) = \lim\limits_{(x,y)\to(\infty,\infty)} F(x, y) = 1$.

3. $F(x, y)$ es monótona no decreciente respecto a cada una de sus variables, es decir

   a) Si $x_1 < x_2$ entonces $F(x_1, y) \leq F(x_2, y)$ para todo $y \in \mathbb{R}$

   b) Si $y_1 < y_2$ entonces $F(x, y_1) \leq F(x, y_2)$ para todo $x \in \mathbb{R}$

4. $F(x, y)$ es continua a la derecha respecto a cada una de sus variables, es decir,

   a) $\lim\limits_{h\to0^+} F(x + h, y) = F(x, y)$ para todo $y \in \mathbb{R}$

   b) $\lim\limits_{k\to0^+} F(x, y + k) = F(x, y)$ para todo $x \in \mathbb{R}$

La función de distribución permite calcular la probabilidad de cualquier rectángulo de $\mathbb{R}^2$ de la forma $(x_1, x_2] \times (y_1, y_2]$, de la siguiente manera:

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$$

Además, si (X, Y) es una variable aleatoria bidimensional con función de distribución conjunta $F(x, y)$, siendo $F_1(x)$ y $F_2(y)$ las funciones de distribución de las variables aleatorias X e Y, respectivamente, entonces decimos que estas variables son independientes si, y sólo si,

$$F(x, y) = F_1(x) \cdot F_2(y) \qquad \text{para todo} \quad (x, y) \in \mathbb{R}^2$$

5.6.2. Tipos de variables aleatorias bidimensionales

Existen varios tipos de variables aleatorias bidimensionales en función de la naturaleza (discreta, continua o mixta) de las variables que la componen vamos a estudiar dos casos: las variables aleatorias bidimensionales discretas y las continuas.

Variables aleatorias bidimensionales discretas

Una variable aleatoria bidimensional (X, Y) se dice que es discreta si X e Y son variables aleatorias discretas.

Supongamos que X toma los valores $\{x_1, x_2, \ldots, x_k\}$, e Y toma los valores $\{y_1, y_2, \ldots, y_p\}$. Entonces la distribución de probabilidad de la variable (X, Y) viene determinada por la tabla de doble entrada

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_p	
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	p_{1p}	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	p_{2p}	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	p_{ip}	$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_k	p_{k1}	p_{k2}	$\cdots$	p_{kj}	$\cdots$	p_{kp}	$p_{k\cdot}$
	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	$p_{\cdot p}$	1

siendo

$$p_{ij} = p(x_i, y_j) = P(X = x_i, Y = y_j) \qquad \text{con} \qquad \sum_{i=1}^{k} \sum_{j=1}^{p} p_{ij} = 1$$

$$p_{i\cdot} = \sum_{j=1}^{p} p_{ij} \qquad \text{y} \qquad p_{\cdot j} = \sum_{i=1}^{k} p_{ij}$$

La función de distribución de la variable aleatoria bidimensional (X, Y) se define así:

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_j \leq y} P(X = x_i, Y = y_j)$$

El momento de orden (r, s) respecto al punto (a, b), de la variable aleatoria bidimensional (X, Y), se definen así:

$$M_{rs}(a, b) = E[(X - a)^r (Y - b)^s] = \sum_{i=1}^{k} \sum_{j=1}^{p} (x_i - a)^r \cdot (y_j - b)^s \cdot p_{ij}$$

denotando por m_{rs} los momentos ordinarios (cuando $a = 0$ y $b = 0$) y por μ_{rs} los momentos centrales (cuando $a = \bar{x}$ y $b = \bar{y}$). Estos momentos definen, entre otras medidas, las medias y varianzas de las variables X e Y, así como su covarianza:

$$\bar{x} = m_{10} \quad , \quad \bar{y} = m_{01} \quad , \quad \sigma_x^2 = \mu_{20} \quad , \quad \sigma_y^2 = \mu_{02} \quad , \quad \sigma_{xy} = \mu_{11}$$

Los conceptos de distribución marginal, distribución condicionada e independencia de variables son similares a los de las variables estadísticas cambiando frecuencia por probabilidad.

Ejemplo 5.14 *Estudiar la variable aleatoria bidimensional (X, Y) cuya distribución de probabilidad se muestra en la siguiente tabla:*

$Y \backslash X$	0	1	2	3	
-1	$0'06$	$0'02$	$0'04$	$0'08$	$0'2$
0	$0'15$	$0'05$	$0'10$	$0'20$	$0'5$
1	$0'09$	$0'03$	$0'06$	$0'12$	$0'3$
	$0'3$	$0'1$	$0'2$	$0'4$	1

En primer lugar observamos que la variable bidimensional (X, Y) es discreta porque son también discretas sus dos componentes, y el soporte es el producto cartesiano de los soportes de cada una de sus componentes, es decir, el conjunto $S_{xy} = \{0, 1, 2, 3\} \times \{-1, 0, 1\}$.

La probabilidad de cualquier región de $\mathbb{R}^2$ se calcula sumando las probabilidades $p(x, y)$ correspondientes a todos los puntos $(x, y) \in S_{xy}$ que pertenecen a la región. Por ejemplo,

$$\begin{aligned} P((X-1)^2 + Y^2 \leq 1) &= p(0,0) + p(1,0) + p(2,0) + p(1,1) + p(1,-1) = \\ &= 0'15 + 0'05 + 0'10 + 0'03 + 0'02 = \\ &= 0'35 \end{aligned}$$

Las distribuciones marginales, que denotaremos por $p_x(x_i)$ y por $p_y(y_j)$, aparecen representadas en el margen de la tabla y son las siguientes:

x_i	$p_x(x_i)$
0	$0'3$
1	$0'1$
2	$0'2$
3	$0'4$

y_j	$p_y(y_j)$
-1	$0'2$
0	$0'5$
1	$0'3$

y, a partir de ellas, podemos comprobar que las variables son independientes pues

$$p(x_i, y_j) = p_x(x_i) \cdot p_y(y_j) \qquad \text{para todo} \qquad (x_i, y_j) \in S_{xy}$$

Además, podemos calcular cualquier medida de cada una de las variables, por ejemplo, la media y la varianza de la variable X:

$$\begin{aligned} E[X] &= \sum_{x=0}^{3} x \cdot p(x) = 1'7 \\ E[X^2] &= \sum_{x=0}^{3} x^2 \cdot p(x) = 4'5 \\ V[X] &= E[X^2] - (E[X])^2 = 4'5 - (1'7)^2 = 1'61 \end{aligned}$$

También podemos calcular las distribuciones condicionadas. Por ejemplo, la distribución de probabilidad de la variable Y condicionada al valor 2 de la variable X, que denotamos por $p_2(y_j)$, es:

$$p_2(y_j) = p(y_j \,|\, X = 2) = \frac{p(x, y)}{p_x(2)} \qquad \longrightarrow$$

y_j	$p_2(y_j)$
-1	$0'2$
0	$0'5$
1	$0'3$

que coincide con la distribución marginal de la variable Y, pues las variables son independientes. □

Variables aleatorias bidimensionales continuas

Una variable aleatoria bidimensional (X, Y) se dice que es continua si X e Y son variables aleatorias continuas.

La distribución de probabilidad de la variable (X, Y) viene determinada por una función de densidad $f(x, y)$, integrable y no negativa, que verifica:

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, dv\, du \qquad \text{para todo} \quad (x, y) \in \mathbb{R}^2$$

y, además,

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

para todo punto $(x, y) \in \mathbb{R}^2$ donde exista esta derivada de segundo orden.

La función de densidad permite calcular la probabilidad de cualquier rectángulo de $\mathbb{R}^2$ de la forma $(x_1, x_2] \times (y_1, y_2]$, de la siguiente manera:

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x, y)\, dy\, dx$$

y, en general, se puede calcular la probabilidad de cualquier región $D \subset \mathbb{R}^2$ integrando (integrales dobles) la función de densidad sobre la región:

$$P(D) = \int\int_D f(x, y)\, dx\, dy$$

El momento de orden (r, s) respecto al punto (a, b), de la variable aleatoria bidimensional (X, Y), se definen así:

$$M_{rs}(a, b) = E[(X - a)^r (Y - b)^s] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - a)^r \cdot (y - b)^s \cdot f(x, y)\, dy\, dx$$

denotando por m_{rs} los momentos ordinarios (cuando $a = 0$ y $b = 0$) y por μ_{rs} los momentos centrales (cuando $a = \mu_x$ y $b = \mu_y$). Estos momentos definen, entre otras medidas, las medias y varianzas de las variables X e Y, así como su covarianza:

$$\mu_x = m_{10} \quad , \quad \mu_y = m_{01} \quad , \quad \sigma_x^2 = \mu_{20} \quad , \quad \sigma_y^2 = \mu_{02} \quad , \quad \sigma_{xy} = \mu_{11}$$

Las distribuciones marginales de la variable aleatoria bidimensional (X, Y) son

$$F_1(x) = F(x, \infty) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u, v)\, dv\, du \quad , \quad F_2(y) = F(\infty, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{y} f(u, v)\, dv\, du$$

siendo

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \qquad \text{y} \qquad f_2(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

las funciones de densidad de las distribuciones de las variables X e Y, respectivamente.

Las distribuciones condicionadas de la variable aleatoria bidimensional (X, Y) son

$$\begin{aligned} F_y(x) = F(x \mid y) &= P(X \leq x \mid Y = y) = \frac{\int_{-\infty}^{x} f(u, y)\, du}{f_2(y)} \\ F_x(y) = F(y \mid x) &= P(Y \leq y \mid X = x) = \frac{\int_{-\infty}^{y} f(x, v)\, dv}{f_1(x)} \end{aligned}$$

siendo

$$f_y(x) = f(x \mid y) = \frac{f(x,y)}{f_2(y)} \quad , \quad f_x(y) = f(y \mid x) = \frac{f(x,y)}{f_1(x)}$$

donde $f_y(x)$ es la función de densidad de la variable X condicionada al valor y de la variable Y, y $f_x(y)$ es la función de densidad de la variable Y condicionada al valor x de la variable X. Obsérvese que para poder definir estas funciones condicionadas es necesario que sea positivo el correspondiente valor de la función de densidad marginal que aparece en el denominador.

Y, por último, diremos que las variables aleatorias X e Y son independientes si, y sólo si,

$$F(x,y) = F_1(x) \cdot F_2(y) \quad \text{o bien} \quad f(x,y) = f_1(x) \cdot f_2(y) \quad \text{para todo} \quad (x,y) \in \mathbb{R}^2$$

Ejemplo 5.15 *Estudiar la variable aleatoria bidimensional (X,Y) con función de densidad*

$$f(x,y) = \begin{cases} cxy & \textit{si} \quad 0 \le y \le 1 - x \le 1 \\ 0 & \textit{en el resto} \end{cases}$$

En primer lugar determinamos el soporte S_{xy} que resulta ser la región de $\mathbb{R}^2$ con forma de triángulo de vértices $(0,0)$, $(1,0)$ y $(0,1)$. Después, podemos determinar el valor de la constante c aplicando que la integral de la función de densidad sobre el soporte es 1.

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dx\, dy = \int_0^1 \int_0^{1-x} cxy\, dy\, dx = \int_0^1 \frac{cx(x-1)^2}{2}\, dx = \frac{c}{24}$$

y determinamos el valor $c = 24$ resolviendo la ecuación correspondiente.

Ahora, podemos calcular las distribuciones marginales:

$$\begin{aligned} f_1(x) &= \int_{-\infty}^{\infty} f(x,y)\, dy = \int_0^{1-x} 24xy\, dy = 12x(x-1)^2 \quad \text{si } 0 \le x \le 1 \\ f_2(y) &= \int_{-\infty}^{\infty} f(x,y)\, dx = \int_0^{1-y} 24xy\, dx = 12y(y-1)^2 \quad \text{si } 0 \le y \le 1 \end{aligned}$$

y, por lo tanto,

$$f_1(x) = \begin{cases} 12x(x-1)^2 & \text{si } 0 \le x \le 1 \\ 0 & \text{en el resto} \end{cases} \qquad \text{y} \qquad f_2(x) = \begin{cases} 12y(y-1)^2 & \text{si } 0 \le y \le 1 \\ 0 & \text{en el resto} \end{cases}$$

y comprobar que las variables no son independientes pues $f(x,y) \neq f_1(x) f_2(y)$.

Con las distribuciones marginales podemos calcular la media y la varianza de cada una de las variables:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f_1(x)\, dx = \int_0^1 12x^2(x-1)^2\, dx = \frac{2}{5} \\ E[X^2] &= \int_{-\infty}^{\infty} x^2 f_1(x)\, dx = \int_0^1 12x^3(x-1)^2\, dx = \frac{1}{5} \\ V[X] &= E[X^2] - (E[X])^2 = \frac{1}{5} - \left(\frac{2}{5}\right)^2 = \frac{1}{25} \end{aligned}$$

y, por simetría, deducimos que $E[Y] = 2/5$ y que $V[Y] = 1/25$.

Y también podemos calcular las distribuciones condicionadas. Por ejemplo, la distribución de probabilidad de la variable Y condicionada al valor $1/2$ de la variable X es

$$f_{1/2}(y) = f(y\,|\,X = 1/2) = \frac{f(1/2, y)}{f_1(1/2)} = \frac{24 \cdot (1/2) \cdot y}{3/2} = 8y \quad \text{para todo} \quad 0 \leq y \leq \frac{1}{2}$$

y, por lo tanto,

$$f_{1/2}(y) = \begin{cases} 8y & \text{si } 0 \leq y \leq \dfrac{1}{2} \\ 0 & \text{en el resto} \end{cases}$$

Con todas estas funciones de densidad calculadas podríamos obtener la probabilidad de cualquier conjunto y las medidas de cualquiera de las variables aplicando las fórmulas correspondientes. □

Regresión y correlación

Los conceptos de regresión y correlación de variables aleatorias son similares a los de las variables estadísticas cambiando frecuencia por probabilidad. El objetivo es el mismo: "encontrar y medir una relación entre las variables X e Y, que nos permita predecir una de ellas en función de la otra". Para ello, determinaremos la línea de regresión, que en el caso lineal, consiste en encontrar los valores de a y b en el modelo $Y^* = a + bX$ que minimice $E[(Y - a - bX)^2]$. Y el resultado es

$$b = \frac{\sigma_{xy}}{\sigma_x^2} \quad , \quad a = \mu_y - b\mu_x \quad \text{y} \quad r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

siendo r un número real en el intervalo $[-1, 1]$, que se denomina coeficiente de correlación lineal y que determina la bondad del ajuste.

Ejemplo 5.16 *Obtener la recta de regresión de Y/X para las variables X e Y estudiadas en el ejemplo 5.13 de la página 187 y determinar la bondad del ajuste.*

La distribución de probabilidad de la variable (X, Y) se representa en la siguiente tabla:

$Y \backslash X$	0	1	2	3	
0	0	1/8	2/8	1/8	1/2
1	1/8	2/8	1/8	0	1/2
	1/8	3/8	3/8	1/8	1

y, a partir de ella, determinamos la curva general de regresión:

x	y	$p(x, y)$
0	1	1/8
1	2/3	3/8
2	1/3	3/8
3	0	1/8

que nos permite calcular, de forma más sencilla, la recta de regresión, pues los puntos de esta curva están alineados y la única recta que pasa por ellos es la recta buscada:

$$y = 1 - \frac{1}{3}x$$

Si queremos estudiar la bondad del ajuste tenemos que calcular el coeficiente de correlación lineal, utilizando los datos de la distribución de probabilidad de la variable (X, Y), presentados en la primera tabla:

$$\bar{x} = 1'5 \quad , \quad \bar{y} = 0'5 \quad , \quad \sigma_x^2 = 0'75 \quad , \quad \sigma_y^2 = 0'25 \quad , \quad \sigma_{xy} = -0'25$$

y el coeficiente de correlación lineal de Pearson es

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{-1/4}{\sqrt{3/4} \cdot \sqrt{1/4}} = -\frac{1}{\sqrt{3}} = -0'577$$

que se interpreta de forma similar a su homónimo en estadística descriptiva. □

Ejemplo 5.17 *Obtener la recta de regresión de Y/X para las variables X e Y estudiadas en el ejemplo 5.15 de la página 192 y determinar la bondad del ajuste.*

En el ejemplo 5.15 ya habíamos calculado la media y la varianza de cada una de las variables que resultaban ser:

$$\bar{x} = \bar{y} = \frac{2}{5} \qquad \text{y} \qquad \sigma_x^2 = \sigma_y^2 = \frac{1}{25}$$

Si calculamos la covarianza

$$E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y \cdot f(x, y)\, dx\, dy = \int_0^1 \int_0^{1-x} 24x^2y^2\, dy\, dx = \cdots = \frac{2}{15}$$

$$\sigma_{xy} = E[XY] - E[X] \cdot E[Y] = \frac{2}{15} - \frac{2}{5} \cdot \frac{2}{5} = -\frac{2}{75}$$

ya tenemos todas las medidas para determinar al recta de regresión de Y/X

$$y = a + bx \qquad \longrightarrow \qquad \left\{ \begin{array}{l} b = \dfrac{\sigma_{xy}}{\sigma_x^2} = \dfrac{-2/75}{1/25} = -\dfrac{2}{3} \\[2ex] a = \bar{y} - b\bar{x} = \dfrac{2}{5} + \dfrac{2}{3} \cdot \dfrac{2}{5} = \dfrac{2}{3} \end{array} \right\} \qquad \longrightarrow \qquad y = \frac{2}{3} - \frac{2}{3}x$$

y la bondad del ajuste queda determinada por el coeficiente de correlación lineal de Pearson:

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{-2/75}{1/5 \cdot 1/5} = -\frac{2}{3}$$

que se interpreta de forma similar a su homónimo en estadística descriptiva. □

5.7. Relación de problemas

1. Sea X el número de años que deben transcurrir antes de que un tipo particular de máquina necesite reemplazo. Supóngase que la distribución de probabilidad de X es $P(1) = 0'3$, $P(2) = 0'4$, $P(3) = 0'2$ y $P(4) = 0'1$. Calcule y represente la función de distribución.

2. Dado el experimento consistente en lanzar un par de dados, consideramos las siguientes variables aleatorias:

   $X =$ "máximo de la puntuación obtenida entre los dos dados".

   $Y =$ "diferencia (en valor absoluto) de los puntos obtenidos en los dados".

   Para cada una de las variables que hemos definido, se pide:

   a) Calcular y representar la distribución de probabilidad.

   b) Calcular y representar la función de distribución.

   c) Calcular la esperanza, la varianza y desviación típica.

   d) Calcular: $P(X \leq \bar{x})$, $P(X > \bar{x})$, $P(2 < X \leq 4)$.

   e) Calcular: $P(Y \leq 2)$, $P(Y > 2)$, $P(Y = 2)$, $P(Y > 7)$.

   f) Determinar la mediana, la moda y los cuartiles.

3. Se lanza cuatro veces una moneda trucada que tiene $2/3$ de probabilidad de salir cara (H) y $1/3$ de probabilidad de salir cruz (T). Consideramos las siguientes variables aleatorias:

   $X =$ "mayor número de caras consecutivas obtenidas en los cuatro lanzamientos".

   $Y =$ "número total de caras obtenidas en los cuatros lanzamientos".

   a) Para cada una de las variables que hemos definido

      1) Calcule y represente la distribución de probabilidad.

      2) Calcule y represente la función de distribución.

      3) Calcule la esperanza, la varianza y la desviación típica.

   b) Utilice alguna de las variables X o Y que hemos definido, para calcular las siguientes probabilidades:

      1) Probabilidad de que salgan a lo sumo dos caras consecutivas.

      2) Probabilidad de que salgan, al menos, dos caras (no necesariamente consecutivas).

4. Consideremos la variable aleatoria X con función de distribución

$$F(x) = \begin{cases} 0 & \text{si} \quad x < -1 \\ 0'3 & \text{si} \quad -1 \leq x < 0 \\ 0'7 & \text{si} \quad 0 \leq x < 1 \\ 1 & \text{si} \quad x \geq 1 \end{cases}$$

   a) Dibuje la función de distribución.

   b) Calcule la distribución de probabilidad.

   c) Calcule las probabilidades $P(X > 0)$, $P(X \leq 2)$ y $P(X = 1 \,|\, X \geq 0)$.

5. *Distribución degenerada.* Sea X una variable aleatoria que sólo toma el valor x_0. Determine su distribución de probabilidad, su función de distribución, su media, su varianza y su función generatriz de momentos.

6. *Distribución de Bernoulli.* Consideramos la variable aleatoria X que sólo toma los valores 0 y 1, y que la probabilidad asociada al punto $x = 1$ es un valor $p \in [0, 1]$.

    a) Calcule la media, la varianza y las funciones generatrices de probabilidad y de momentos. Particularice los resultados para $p = 1/2$.
    b) Determine experimentos y variables que tengan esta distribución de probabilidad.
    c) Compruebe que si $p = 0$, o bien, si $p = 1$, entonces la distribución de probabilidad de nuestra variable es degenerada.

7. *Distribución uniforme discreta.* Consideremos la variable aleatoria discreta U que toma los valores $\{1, 2, \ldots, n\}$, todos ellos con la misma probabilidad. Calcule la distribución de probabilidad, su media y su varianza.

    Nota: para calcular la varianza se necesita saber que $\displaystyle\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$

8. *Distribución geométrica I de parámetro p.* Sea X la variable aleatoria que determina el número de fallos antes del primer éxito, siendo p la probabilidad de éxito y $q = 1 - p$ la probabilidad de fracaso. Pensemos, por ejemplo, en lanzamientos consecutivos de una moneda, siendo éxito, por ejemplo, el suceso F. Se pide:

    a) Determinar la distribución de probabilidad de la variable aleatoria X.
    b) Demostrar que $\bar{x} = q/p$ y que $\sigma_x^2 = q/p^2$.
    c) Determinar la función generatriz de probabilidad y comprobar que se verifica el teorema de inversión.
    d) Determinar la función generatriz de momentos y utilizarla para comprobar los resultados obtenidos para la media y la varianza de la variable.
    e) Particularizar los resultados para $p = 1/2$.
    f) ¿Qué sucede si $p = 0$, o bien, si $p = 1$?

9. *Distribución geométrica II de parámetro p.* Sea X una variable aleatoria discreta con función generatriz de probabilidad

    $$G(s) = \frac{ps}{1 - sq}$$

    para algún $p \in [0, 1]$ con $q = 1 - p$. Se pide:

    a) Determinar la distribución de probabilidad de la variable aleatoria X.
    b) Demostrar que $\bar{x} = 1/p$ y que $\sigma_x^2 = q/p^2$.
    c) Comprobar que la variable X representa el número de pruebas necesarias para obtener el primer éxito, siendo p la probabilidad de éxito y $q = 1-p$ la probabilidad de fracaso.
    d) Determinar la función generatriz de momentos y utilizarla para comprobar los resultados obtenidos para la media y la varianza de la variable.
    e) Particularizar los resultados para $p = 1/2$.

f) ¿Qué sucede si $p = 0$, o bien, si $p = 1$?

10. *Distribución de Poisson.* Sea X una v.a.d que toma los valores $0, 1, 2, \ldots$ con probabilidad

$$P(X = x) = \mathrm{e}^{-\lambda} \frac{\lambda^x}{x!}$$

para algún valor real $\lambda > 0$.

a) Demuestre que $E[x] = V[X] = \lambda$.

b) Determine la función generatriz de probabilidad y compruebe que se verifica el teorema de inversión.

c) Determine la función generatriz de momentos y utilícela para comprobar los resultados obtenidos para la media y la varianza de la variable.

d) Particularice los resultados para $\lambda = 2$.

11. Sea X una variable aleatoria continua con función de densidad

$$f(x) = \begin{cases} x/8 & \text{si } 0 \leq x \leq 4 \\ 0 & \text{en el resto} \end{cases}$$

Se pide:

a) Representar la función de densidad de X.

b) Calcular y representar la función de distribución de X.

c) Calcular la esperanza, la varianza y desviación típica de X.

d) Calcular la mediana, la moda y el rango intercuartílico.

e) Calcular las probabilidades $P(1 \leq X \leq 3)$, $P(X \leq 1)$, $P(X \geq 3)$, $P(X > 0)$ y $P(X \geq 5)$, y las probabilidades condicionadas $P(X > 1 \,|\, X < 3)$ y $P(X \geq Q_1 \,|\, X \leq Me)$.

12. *Distribución Uniforme Continua.* Obtenga k para que $f(x) = k$, sea una función de densidad en el intervalo $[0, 1]$. Halle su función de distribución, su media y su varianza. Obtenga los mismos resultados para el caso en el que la función esté definida en el intervalo $[a, b]$.

13. La demanda diaria de gasolina sin plomo (en litros) en cierta estación de servicio es una variable aleatoria X. Supóngase que X tiene la densidad

$$f(x) = \begin{cases} k & \text{si } 4000 < x < 9000 \\ 0 & \text{en el resto} \end{cases}$$

Se pide:

a) Calcular el valor de k.

b) Representar la función de densidad de X.

c) Calcular y representar la función de distribución de X.

d) Calcular la esperanza, la varianza y desviación típica de X.

e) Calcular la probabilidad de vender más de 5000 litros.

14. Una variable aleatoria X tiene por función de densidad:

$$f(x) = \begin{cases} k \cdot x^2 & \text{si } 0 \leq x \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

Determine el valor de k y encuentre el número c tal que $F(c) = 72'9\,\%$.

15. Una variable aleatoria X tiene por función de densidad:

$$f(x) = \begin{cases} c \cdot \mathrm{e}^{-x} & \text{si } x > 0 \\ 0 & \text{en el resto} \end{cases}$$

Calcule el valor de c y determine la función de distribución, la media, la mediana, la varianza y la función generatriz de momentos.

16. Una variable aleatoria X tiene por función de distribución:

$$F(x) = \begin{cases} 0 & \text{si} & x < 0 \\ k \cdot x^n & \text{si} & 0 \leq x \leq 1 \\ 1 & \text{si} & x > 1 \end{cases}$$

Determine la función de densidad, la media, la mediana y la varianza, para cualquier valor entero de n que sea mayor que 1.

17. La función de densidad de una variable aleatoria X es $f(x) = kx$ si $x \in (0, 1)$ y $f(x) = 0$ en el resto. Halle:

   a) La función de distribución.
   b) $P(X < 2/3)$.
   c) $P(1/3 < X < 1/2)$.
   d) El valor a tal que $P(X < a) = 0'25$ e interpretar el resultado.
   e) Su media y varianza.

18. Sea X el espesor (en milímetros) de las arandelas que produce una máquina. Supóngase que X tiene una densidad $f(x) = kx$ si $x \in (1'9, 2'1)$ y $f(x) = 0$ en el resto. Halle:

   a) La función de distribución.
   b) La probabilidad de que una arandela tenga espesor $1'95$.
   c) $P(1'95 < X < 2'05)$.
   d) El valor a tal que $P(X < a) = 0'25$ e interpretar el resultado.
   e) Su media y varianza.

19. Una máquina fabrica ejes, cuya medida del radio (X) se distribuyen según la función de densidad $f(x) = k \cdot (x - 9'9) \cdot (x - 10'1)$ si $x \in (9'9, 10'1)$ y cero en caso contrario (x en milímetros).

   a) Determine el valor de k, y calcule la media y la varianza.
   b) Si se desechan todos los ejes cuyos radios se desvían en más de $0'03$ mm. de la media, calcule la proporción de ejes fabricados que se rechazarán.

   c) Determine la nueva función de densidad $f(h)$, si los ejes se midiesen en centímetros, es decir, si $h = x/10$.

20. Sea X una variable aleatoria continua con función de densidad

$$f(x) = \begin{cases} 1/4 & \text{si } 0 \leq x < 1 \\ 1/2 & \text{si } 1 \leq x < 2 \\ a & \text{si } 2 \leq x \leq 4 \\ 0 & \text{en el resto} \end{cases}$$

   a) Determine el valor de a.
   b) Determine y representa la función de distribución.
   c) Calcule la media, la mediana y la moda.
   d) Calcule la varianza
   e) Estudie la simetría y la curtosis.

21. Sea X una variable aleatoria continua con función de densidad

$$f(x) = \begin{cases} 0 & \text{si} & x < -1 \\ a + x & \text{si} & -1 \leq x < 0 \\ a - x & \text{si} & 0 \leq x < 1 \\ 0 & \text{si} & x \geq 1 \end{cases}$$

   a) Determine el valor de a.
   b) Determine y representa la función de distribución.
   c) Calcule la media, la mediana y la moda.
   d) Calcule la varianza
   e) Estudie la simetría y la curtosis.

22. Sea X una variable aleatoria con función de distribución

$$F(x) = \begin{cases} 0 & \text{si} & x \leq 0 \\ x^2 & \text{si} & 0 \leq x < 1 \\ 1 & \text{si} & x \geq 1 \end{cases}$$

   Se pide:

   a) Dibujar la función de distribución.
   b) Calcular y dibujar la función de densidad.
   c) Calcular las probabilidades $P(X < 0'25)$ y $P(X < 0'25 \,|\, X < 0'5)$.

23. El tiempo de reparación (en horas) de un tipo de máquina, tiene la función de distribución:

$$F(x) = \begin{cases} 0 & \text{si} & x \leq 0 \\ x/2 & \text{si} & 0 \leq x < 1 \\ 1/2 & \text{si} & 1 \leq x < 2 \\ x/4 & \text{si} & 2 \leq x < 4 \\ 1 & \text{si} & x \geq 4 \end{cases}$$

   Se pide:

a) Dibujar la función de distribución.

b) Calcular, dibujar e interpretar la función de densidad.

c) Calcular la probabilidad de que si el tiempo de reparación es superior a una hora, lo sea de $3'5$ horas ($P(X \geq 3'5|X \geq 1)$).

24. *Variable aleatoria mixta (v.a.m).* La mayoría de los problemas se modelizan utilizando distribuciones discretas o continuas. Sin embargo, en ocasiones es necesario considerar una mezcla de las dos distribuciones. Una variable aleatoria X se dice que es mixta si su distribución de probabilidad está determinada por la probabilidad en un conjunto de puntos $D = \{x_1, x_2, \dots\}$, a lo sumo numerable, y por una función no negativa $g(x)$ (a modo de función de densidad) que determina la probabilidad de los intervalos de números reales que no contengan puntos de D, de manera que

$$\sum_{x_i \in D} P(X = x_i) + \int_{-\infty}^{\infty} g(x)dx = 1$$

Obsérvese que $g(x)$ no es una función de densidad pues $\int_{-\infty}^{\infty} g(x)dx < 1$. La función de distribución se define de la manera habitual, y la media y la varianza se definen así:

$$E(X) = \mu_x = \sum_{x_i \in D} x_i \cdot P(X = x_i) + \int_{-\infty}^{\infty} xg(x)dx \quad , \quad V(X) = E((X - \mu_x)^2)$$

Sea X una variable aleatoria mixta cuya distribución de probabilidad está definida por

$$P(0) = 0'1 \qquad , \qquad P(1) = 0'2 \qquad , \qquad g(x) = \begin{cases} 0'05x + 0'2 & \text{si } x \in [2,4] \\ 0 & \text{en el resto} \end{cases}$$

Se pide:

a) Determinar las probabilidades: $P(X < 0)$, $P(X \geq 0'5)$ y $P(3 \leq X \leq 7)$.

b) Calcular y representar la función de distribución.

c) Calcular la media, y la varianza.

d) Calcular la mediana y el rango intercuartílico.

e) Estudiar la simetría y la curtosis de la distribución de la variable.

25. Consideremos las tres funciones de distribución:

$$F_1(x) = \begin{cases} 0 & \text{si} & x < 1 \\ (x-1)/2 & \text{si} & 1 \leq x < 2 \\ 1/2 & \text{si} & 2 \leq x < 3 \\ (x-2)/2 & \text{si} & 3 \leq x < 4 \\ 1 & \text{si} & x \geq 4 \end{cases} \quad , \quad F_2(x) = \begin{cases} 0 & \text{si} & x < 1 \\ (x-1)/3 & \text{si} & 1 \leq x < 2 \\ 1/3 & \text{si} & 2 \leq x < 3 \\ (x-1)/3 & \text{si} & 3 \leq x < 4 \\ 1 & \text{si} & x \geq 4 \end{cases}$$

$$\text{y} \qquad F_3(x) = \begin{cases} 0 & \text{si} & x < 1 \\ 1/6 & \text{si} & 1 \leq x < 2 \\ 1/2 & \text{si} & 2 \leq x < 3 \\ 5/6 & \text{si} & 3 \leq x < 4 \\ 1 & \text{si} & x \geq 4 \end{cases}$$

Para cada una de estas distribuciones, se pide:

a) Representar la función y determinar el tipo de variable aleatoria correspondiente (discreta, continua o mixta).

b) Calcular las siguientes probabilidades:
$P(X=1)$, $P(X<3)$, $P(X\leq 3)$, $P(2<X\leq 3)$, $P(X\geq 4)$ $P(X<3\,|\,X\geq 2)$.

c) Calcular la media, la mediana, la varianza y el rango intecuartílico.

d) Determinar la simetría de las distribuciones.

26. Propiedades de las funciones generatrices: Sea X una variable aleatoria discreta tal que $S_x \subset \mathbb{N}$, y sean $G(s)$ y $M(t)$ sus funciones generatrices de probabilidad y momentos, respectivamente. Demuestre que se verifican las siguientes propiedades

    a) $G(1)=1$

    b) $M(0)=1$

    c) $M(t)=G(\mathrm{e}^t)$

27. *Distribución exponencial de parámetro* λ. Sea X una v.a.c con función de densidad

$$f(x)=\begin{cases}\lambda \mathrm{e}^{-\lambda x} & \text{si } x>0\\ 0 & \text{si } x\leq 0\end{cases}$$

para algún valor $\lambda>0$.

a) Represente la función de densidad y verifique sus propiedades.

b) Calcule y representa la función de distribución.

c) Demuestre que $E(x)=1/\lambda$ y que $V(X)=1/\lambda^2$.

d) Determine la función generatriz de momentos y utilícela para comprobar los resultados obtenidos en el apartado anterior.

e) Determine la asimetría de la distribución.

f) Particularice los resultados para $\lambda=2$.

28. Consideremos la variable aleatoria bidimensional (X,Y) con distribución de probabilidad

$X \backslash Y$	1	2	3	4
1	k	0	$0'1$	0
2	$0'3$	0	$0'1$	$0'2$
3	0	$0'2$	0	0

a) Determine el valor de k.

b) Calcule las probabilidades: $P(1<X\leq 3, Y=2)$ y $P(X\geq 2\,|\,Y<2)$.

c) Calcule $F(2,2)$.

d) ¿Son X e Y variables independientes?

e) ¿Qué variable está más dispersa, la X o la Y?

f) Compruebe que $Y/_{X=3}$ es una variable aleatoria degenerada.

29. Consideremos la variable aleatoria bidimensional (X, Y) con distribución de probabilidad

$Y \backslash X$	0	1	4	9
0	0	0	0	k
1	0	0	1/4	0
2	0	1/4	0	0
3	1/4	0	0	0

   a) Determine el valor de k.
   b) ¿Qué variable está más dispersa, la X o la Y?
   c) ¿Son X e Y variables independientes?
   d) Ajuste el modelo de regresión $Y = a + b\sqrt{X}$.

30. Consideremos la variable aleatoria bidimensional (X, Y) con distribución de probabilidad

$$p(x,y) = \begin{cases} c(x+y) & \text{si } (x,y) \in \{0,1,2,3\} \times \{0,1,2\} \\ 0 & \text{en el resto} \end{cases}$$

   a) Determine el valor de c.
   b) Ajuste el modelo lineal de regresión $Y = a + bX$.
   c) Calcule y represente la función de distribución de la variable X.

31. *Distribución uniforme bidimensional discreta.* Consideramos la variable (U, V) que toma todos los valores en el conjunto $\{1, 2, 3, 4, 5\} \times \{1, 2, 3, 4\}$ con la misma probabilidad.

   a) Determine la distribución de probabilidad conjunta.
   b) Compruebe que las distribuciones marginales son también de tipo uniforme.
   c) Calcule las rectas de regresión Y/X y X/Y y estudie su correlación lineal.
   d) Estudie la independencia de las distribuciones marginales.

32. Calcule las rectas de regresión Y/X y determine la bondad de los ajustes para los pares de variables X e Y de los ejercicios 2 y 3 de esta relación de problemas.

33. Consideremos la variable aleatoria continua (X, Y) con función de densidad

$$f(x,y) = \begin{cases} cx^2y & \text{si } x^2 \le y \le 1 \\ 0 & \text{en el resto} \end{cases}$$

   a) Dibuje la región de $\mathbb{R}^2$ que representa el soporte de la variable.
   b) Determine el valor de c.
   c) Calcule las probabilidades $P(X \ge 0)$, $P(Y \le 1/4)$, $P(Y^2 < X)$ y $P(X^2 \le Y < X)$.
   d) Calcule la probabilidad correspondiente al cuadrado de lado 1 que tiene su centro en el origen de coordenadas.
   e) Calcule la probabilidad correspondiente al círculo de radio 1, centrado en el origen.
   f) Determine las distribuciones marginales de las variables X e Y.
   g) ¿Son X e Y variables independientes?
   h) Calcule las rectas de regresión X/Y e Y/X y determine la bondad de los ajustes.

34. Consideremos la variable aleatoria continua (X, Y) con función de densidad

$$f(x,y) = \begin{cases} cy^2 & \text{si } 0 \leq x \leq 2 \text{ y } 0 \leq y \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

  a) Dibuje la región de $\mathbb{R}^2$ que representa el soporte de la variable.

  b) Determine el valor de c.

  c) Determine la función de distribución.

  d) Calcule las probabilidades $P(X \geq 0)$, $P\left(Y \leq \frac{1}{2}\right)$ y $P\left(\frac{1}{2} < X < \frac{3}{2}, \frac{1}{4} < Y < \frac{3}{4}\right)$.

  e) Calcule las probabilidades condicionadas $P(Y \leq X \mid X > 1)$ y $P(Y \leq X \mid Y \leq 1/2)$.

  f) Calcule la probabilidad correspondiente al rectángulo que tiene su centro en el origen de coordenadas y cuya base y altura miden respectivamente dos unidades y una unidad.

  g) Compruebe que la variable aleatoria X se distribuye de manera uniforme.

35. Consideremos la variable aleatoria continua (X, Y) con función de densidad

$$f(x,y) = \begin{cases} c(x^2 + y) & \text{si } 0 \leq y \leq 1 - x^2 \\ 0 & \text{en el resto} \end{cases}$$

  a) Dibuje la región de $\mathbb{R}^2$ que representa el soporte de la variable.

  b) Determine el valor de c.

  c) Calcule las probabilidades $P(X \geq 0)$ y $P(Y >= |3X/2|)$.

  d) Calcule la probabilidad correspondiente al cuadrado de lado 1 que tiene su centro en el origen de coordenadas.

  e) Determine las distribuciones marginales de las variables X e Y.

  f) ¿Son X e Y variables independientes?

  g) Calcule la recta de regresión Y/X y determine la bondad del ajuste.

36. Consideremos la variable aleatoria continua (X, Y) con función de distribución

$$F(x,y) = \begin{cases} 0 & \text{si} & x < 0 & \text{o} & y < 0 \\ kxy(x+y) & \text{si} & 0 \leq x < 2 & \text{y} & 0 \leq y < 2 \\ 2kx(x+2) & \text{si} & 0 \leq x < 2 & \text{y} & y \geq 2 \\ 2ky(2+y) & \text{si} & x \geq 2 & \text{y} & 0 \leq y < 2 \\ 1 & \text{si} & x \geq 2 & \text{y} & y \geq 2 \end{cases}$$

  a) Determine el valor de k.

  b) Determine la función de densidad y el soporte de la variable (X, Y).

  c) Calcule las probabilidades $P(X \leq 1, Y \leq 1)$ y $P(0 \leq X \leq 1, 0 \leq Y \leq 1)$.

  d) Calcule la probabilidad condicionada $P(0 \leq X \leq 1 \mid 0 \leq Y \leq 1)$.

  e) Calcule la probabilidad correspondiente al cuadrado de lado 2 que tiene su centro en el origen de coordenadas.

  f) Calcule la probabilidad correspondiente al círculo de radio 1, centrado en el origen.

g) Determine las distribuciones marginales (función de densidad y de distribución) y calcule la media y la varianza de cada una de las variables.

h) ¿Son X e Y variables independientes?

i) Determine la distribución de Y condicionada al valor $x = 1$ de la variable X.

j) Determine la distribución de X condicionada al valor $y = 1$ de la variable Y.

k) Calcule la recta de regresión Y/X y determine la bondad del ajuste.

37. *Distribución uniforme bidimensional continua.* Sea (X, Y) la variable aleatoria continua con función de densidad constante en todo el soporte $S_{xy} = [0, 1] \times [0, 1]$. Se pide:

a) Determinar el valor constante de la función de densidad.

b) Determinar la función de distribución.

c) Determinar las distribuciones marginales de X e Y y calcular sus medias y varianzas.

d) ¿Son X e Y variables independientes?

e) Calcular las probabilidades $P(X \le 0'5)$, $P(Y \le 0'5)$ y $P(X \le 0'5, Y \le 0'5)$.

f) Calcular las probabilidades $P(Y < X^2)$, $P(X \le Y)$ y $P(X \le Y \le X^2)$.

g) Calcular la probabilidad condicionada $P(X \le 0'5 \,|\, Y \ge 0'5)$.

h) Calcular la probabilidad correspondiente al cuadrado de lado 1 que tiene su centro en el origen de coordenadas.

i) Calcular la probabilidad correspondiente al círculo de radio 1, centrado en el origen.

j) Repetir los cuatro primeros apartado de este ejercicio suponiendo que el soporte original de la distribución uniforme hubiese sido el conjunto $S_{xy} = [a, b] \times [c, d]$

38. Consideremos la variable aleatoria bidimensional (X, Y) cuya función de densidad es

$$f(x, y) = \begin{cases} kx\mathrm{e}^{-xy} & \text{si } 0 \le x \le 1 \text{ e } y > 0 \\ 0 & \text{en el resto} \end{cases}$$

a) Determine el valor de la constante k.

b) Determinar la función de distribución.

c) Determinar las distribuciones marginales de X e Y y calcular sus medias y varianzas.

d) ¿Son X e Y variables independientes?

e) Calcular las probabilidades $P(X \le 0'5)$, $P(Y \le 1)$ y $P(X \le 0'5, Y \le 1)$.