

Capítulo 2

VARIABLES ALEATORIAS

2.1. Introducción

En el capítulo anterior hemos visto ejemplos de fenómenos aleatorios en los que resulta sencillo identificar el espacio muestral y llevar a cabo una asignación de probabilidades. Sin embargo, en muchas aplicaciones ésta no es ni mucho menos una tarea inmediata. Pensemos, por ejemplo, que nuestro objetivo es caracterizar el peso que alcanzan las doradas adultas cultivadas en una piscifactoría. Para conseguir este objetivo necesitaremos un instrumento de medida –en este caso una simple balanza–, que nos dé el peso de cada pez. Es obvio que aún cuando todas las doradas hayan sido cultivadas en las mismas condiciones (misma temperatura, salinidad, alimentación, etc.), habrá diferencias en el peso final alcanzado por cada una. Pesar cada dorada es, pues, un experimento aleatorio en el sentido apuntado en el capítulo anterior: su resultado no se conoce hasta haberlo realizado.

Tras pesar muchas doradas adultas observamos que su peso oscila entre los 300 y los 600 gramos. Podemos entonces asignar como espacio muestral el intervalo $[300, 600]$ (o quizás uno un poco mayor, por ejemplo el $[200, 700]$, si queremos darnos un margen para incluir pesos que quizás puedan darse pero que no se han registrado durante nuestro periodo de observación). ¿Cómo realizamos ahora la asignación de probabilidades? O dicho de otra forma, ¿cómo repartimos (distribuimos) la probabilidad total (que debe ser 1) entre todos los valores de ese intervalo?. Como este intervalo (en realidad, cualquier intervalo) contiene infinitos valores, la regla de Laplace no resulta útil. La asignación mediante frecuencias relativas, todo lo más, nos permitiría asignar probabilidades a subintervalos del espacio muestral; quizás ocurre que un 5% de las doradas observadas pesan entre 300 y 350 gramos, un 15% pesan entre 350 y 400, un 30% entre 400 y 450, etc. Podríamos entonces utilizar estas proporciones como aproximaciones de la probabilidad de que el peso de una dorada se encuentre en cada uno

de estos intervalos. Pero, ¿qué hacemos si queremos saber cuál es la probabilidad de que una dorada pese entre 352 y 353 gramos? Sí, podemos construir subintervalos más finos y volver a evaluar las proporciones; pero para ello necesitaremos muchos más datos experimentales que pueden ser difíciles de conseguir.

Por tanto se hace precisa una herramienta matemática que permita modelar y manejar probabilidades en situaciones como ésta. En este capítulo veremos que los conceptos de variable aleatoria y su distribución de probabilidad son la clave para alcanzar este objetivo. Estos conceptos nos proporcionarán, como veremos, una colección de modelos con la suficiente flexibilidad para adaptarse a un gran número de situaciones. Para conseguir este objetivo deberemos aprender a identificar la estructura probabilista subyacente al problema que nos ocupa; si en lugar de caracterizar el peso de las doradas de piscifactoría, nuestro objetivo fuese caracterizar el peso de las doradas salvajes, o la longitud de las lubinas, o el diámetro del opérculo de las percas, es muy posible que podamos utilizar el mismo modelo, adaptando en cada caso los parámetros de ajuste necesarios.

2.2. Objetivos

Al finalizar este capítulo el alumno deberá:

- Comprender el concepto de variable aleatoria y su función de distribución.
- Saber distinguir variables aleatorias discretas y continuas.
- Entender y saber manejar los conceptos de función de probabilidad (caso discreto) y densidad de probabilidad (caso continuo). Ser capaz de pasar de función de distribución a densidad y viceversa.
- Conocer y saber calcular las principales medidas resumen de una variable aleatoria: momentos, esperanza, varianza y cuantiles. Conocer otras medidas de forma: asimetría y apuntamiento.
- Comprender el concepto de distribución conjunta de variables aleatorias, en particular en el caso de variables independientes.
- Conocer y saber calcular medidas de asociación lineal entre variables continuas: covarianza y correlación.
- Conocer y saber aplicar la desigualdad de Chebyshev.

2.3. Concepto de variable aleatoria

Frecuentemente el resultado de un experimento aleatorio puede medirse de formas distintas, dependiendo de la finalidad con que se haya realizado el experimento. Si se lanza una moneda al aire, el resultado será cara o cruz; pero si hemos apostado 10 euros a que sale cara, desde nuestra perspectiva el resultado del lanzamiento será ganar 10 euros o perder 10 euros. En el curso de una campaña oceanográfica se escogen numerosos puntos de observación; dependiendo del tipo de sensor que se utilice, en un mismo punto se podrán medir velocidad de corriente, temperatura, salinidad, concentración de clorofila,... En un estudio sobre pesca se pueden escoger al azar varias nasas situadas en una misma zona; de cada nasa se puede medir el peso de las capturas, el número de ejemplares capturados, la proporción relativa de sujetos de distintas especies, ...

Así pues, el valor numérico obtenido en un experimento aleatorio resulta de aplicar algún instrumento de medida¹ al objeto observado. La formalización del concepto de instrumento de medida conduce a la definición de *variable aleatoria*.

Formalmente, una *variable aleatoria* es una función que a cada suceso elemental de un espacio muestral le asigna un valor numérico. Más concretamente, dado un experimento aleatorio cuyo *espacio de probabilidad*² asociado es (E, \mathcal{F}, P) , una *variable aleatoria* es una función X definida de E en \mathbb{R} tal que para todo valor $x \in \mathbb{R}$ el conjunto $\{w \in E : X(w) \leq x\}$ pertenece a \mathcal{F} .

Ejemplo 2.1. Consideremos el experimento aleatorio consistente en lanzar dos dados equilibrados. El espacio muestral es el conjunto de parejas de valores:

$$E = \{(i, j), i, j \in \{1, 2, 3, 4, 5, 6\}\}$$

(i es el resultado del primer dado y j el del segundo). Sobre este espacio muestral definimos la variable aleatoria $X = \text{“Suma de las caras superiores de los dados”}$:

$$X(i, j) = i + j$$

Si consideramos el álgebra \mathcal{F} de las partes de E (esto es, el conjunto de todos los conjuntos que pueden formarse con elementos de E), es obvio que para todo $x \in \mathbb{R}$ el conjunto

¹El término *instrumento de medida* se entiende aquí en sentido amplio; puede ser un termómetro que sirve para medir temperatura, o puede ser simplemente nuestro cerebro que traduce la cara de una moneda en una ganancia de 10 euros.

²Recordemos del capítulo anterior que un espacio de probabilidad es una terna (E, \mathcal{F}, P) donde E es el espacio muestral, \mathcal{F} es un álgebra de sucesos asociados a dicho espacio y P es una probabilidad definida sobre \mathcal{F} .

$\{w \in E : X(w) \leq x\}$ pertenece a \mathcal{F} . Así, por ejemplo:

- si $x = 5$, se tiene que:

$$\{w \in E : X(w) \leq 5\} = \{(1, 1), (1, 2), (2, 1), (2, 2), (1, 3), (3, 1), (2, 3), (3, 2)\} \in \mathcal{F};$$
- si $x = 0$, $\{w \in E : X(w) \leq 0\} = \emptyset \in \mathcal{F}$;
- si $x = 17$, $\{w \in E : X(w) \leq 17\} = E \in \mathcal{F}$;
- si $x = 2,83$, $\{w \in E : X(w) \leq 2,83\} = \{(1, 1)\} \in \mathcal{F}$

2.4. Función de distribución de una variable aleatoria.

La condición de que el conjunto $B_x = \{w \in E : X(w) \leq x\}$ sea un suceso perteneciente a \mathcal{F} para todo $x \in \mathbb{R}$, nos asegura que tiene asignada una probabilidad, pues ésta está definida para todos los elementos de \mathcal{F} . La función F_X que a cada valor x le asigna la probabilidad del suceso B_x , esto es,

$$F_X(x) = P(X \leq x) = P(\{w \in E : X(w) \leq x\})$$

recibe el nombre de *función de distribución acumulativa* de la variable X . Esta función toma valores en toda la recta real y tiene por recorrido el intervalo $[0, 1]$.

Ejemplo 2.2. Consideremos de nuevo el experimento aleatorio consistente en lanzar dos dados equilibrados. El resultado de la suma de sus caras superiores es un número entero entre 2 y 12. Si llamamos A_k al suceso consistente en que la suma sea k , tenemos:

$$\begin{aligned} A_2 &= \{(1, 1)\} \\ A_3 &= \{(1, 2), (2, 1)\} \\ A_4 &= \{(1, 3), (3, 1), (2, 2)\} \\ A_5 &= \{(1, 4), (4, 1), (2, 3), (3, 2)\} \\ A_6 &= \{(1, 5), (5, 1), (2, 4), (4, 2), (3, 3)\} \\ A_7 &= \{(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3)\} \\ A_8 &= \{(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)\} \\ A_9 &= \{(3, 6), (6, 3), (4, 5), (5, 4)\} \\ A_{10} &= \{(4, 6), (6, 4), (5, 5)\} \\ A_{11} &= \{(5, 6), (6, 5)\} \\ A_{12} &= \{(6, 6)\} \end{aligned}$$

La probabilidad de cada uno de estos sucesos puede calcularse como $P(A_k) = \frac{N(A_k)}{N(E)} = \frac{N(A_k)}{36}$.

Por tanto las probabilidades de los distintos resultados son:

k	2	3	4	5	6	7	8	9	10	11	12
$P(A_k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Los sucesos B_k , consistentes en que la suma de puntos sea menor o igual que k , pueden obtenerse como:

$$B_k = \{(i, j) : i + j \leq k\} = A_2 \cup A_3 \cup \dots \cup A_k, \quad k = 2, \dots, 12.$$

por lo que la probabilidad de cualquiera de los B_k para $k = 2, 3, \dots, 12$, será:

$$P(B_k) = P(A_2 \cup A_3 \cup \dots \cup A_k) = \sum_{j=1}^k P(A_j) = \sum_{j=1}^k \frac{N(A_j)}{N(E)}$$

Si tenemos en cuenta que, obviamente, $B_x = \emptyset$ si $x < 2$ (no es posible sacar una suma menor que dos al tirar dos dados), $B_x = E$ si $x \geq 12$, y además para cualquier x real tal que $k \leq x < k + 1$ (con $k = 2, 3, \dots, 11$) se tiene que $B_x = B_k$ es inmediato construir la función de distribución de X :

$$F_X(x) = P(X \leq x) = P(B_x) = \begin{cases} 0 & x < 2 \\ 1/36 & 2 \leq x < 3 \\ 3/36 & 3 \leq x < 4 \\ 6/36 & 4 \leq x < 5 \\ 10/36 & 5 \leq x < 6 \\ 15/36 & 6 \leq x < 7 \\ 21/36 & 7 \leq x < 8 \\ 26/36 & 8 \leq x < 9 \\ 30/36 & 9 \leq x < 10 \\ 33/36 & 10 \leq x < 11 \\ 35/36 & 11 \leq x < 12 \\ 1 & x \geq 12 \end{cases}$$

La figura 2.1 muestra gráficamente esta función de distribución.

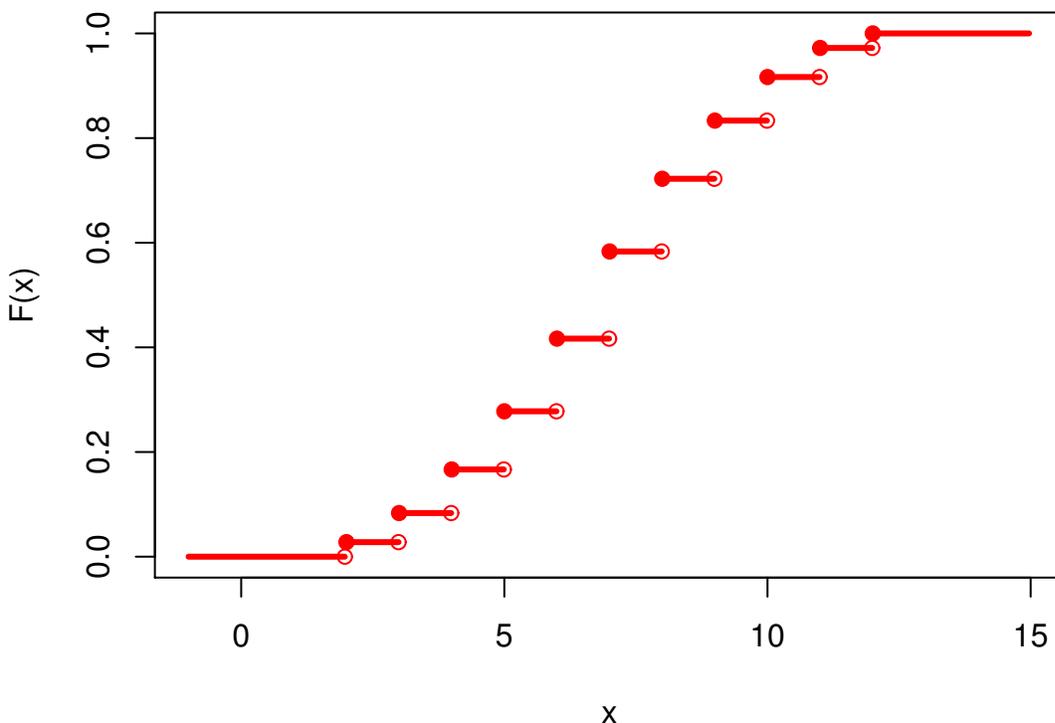


Figura 2.1: Función de distribución de la suma de caras al lanzar dos dados (ejemplo 2.2)

Propiedades de la función de distribución de una variable aleatoria.

1. $0 \leq F(x) \leq 1 \quad \forall x \in \mathbb{R}$
2. $\lim_{x \rightarrow -\infty} F_X(x) = 0$, y $\lim_{x \rightarrow \infty} F_X(x) = 1$
3. $F_X(x)$ es una función monótona no decreciente, esto es, si $a < b$ entonces $F_X(a) \leq F_X(b)$
4. Si $a < b$ entonces $P(a < X \leq b) = F_X(b) - F_X(a)$

2.5. Clasificación de variables aleatorias

Las variables aleatorias pueden clasificarse como *discretas* o *continuas*. Las primeras son aquellas que distribuyen la probabilidad sobre un conjunto finito o numerable de valores. Las variables continuas, por su parte, distribuyen la probabilidad sobre un rango continuo

de valores.

2.5.1. Variables aleatorias discretas,

Una variable aleatoria X es *discreta* cuando el conjunto de valores que puede tomar es finito o numerable. En tal caso, su distribución de probabilidad queda plenamente especificada por la *función de probabilidad* $P(X = k)$, donde k es cualquier valor que pueda tomar la variable. Obviamente se tiene que $\sum_k P(X = k) = 1$.

Ejemplo 2.3. (variable discreta con un número finito de valores). Consideremos el experimento aleatorio consistente en tirar una moneda equilibrada tres veces. Definimos la variable aleatoria $X = \text{“Número de caras”}$. Para este experimento el espacio muestral es

$$E = \{ccc, ccx, cxc, xcc, cxx, xcx, xxc, xxx\}$$

Los únicos valores posibles de X en este experimento son $k = 0, 1, 2, 3$. Para cada k la probabilidad $P(X = k) = P(\{w \in E : X(w) = k\})$ puede obtenerse de manera sencilla utilizando la regla de Laplace y se resume en la tabla siguiente:

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

La función de distribución acumulativa de esta variable aleatoria es:

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

La figura (2.2) muestra gráficamente las funciones de probabilidad y de distribución acumulativa de esta variable aleatoria.

Ejemplo 2.4. (variable discreta con un número infinito numerable de valores) Se realiza el experimento aleatorio consistente en lanzar sucesivas veces una moneda hasta que sale cara por primera vez. El espacio muestral asociado a este experimento es entonces $E = \{c, xc, xxc, xxxc, \dots\}$. Si denotamos por X a la variable aleatoria "Número de lanzamientos

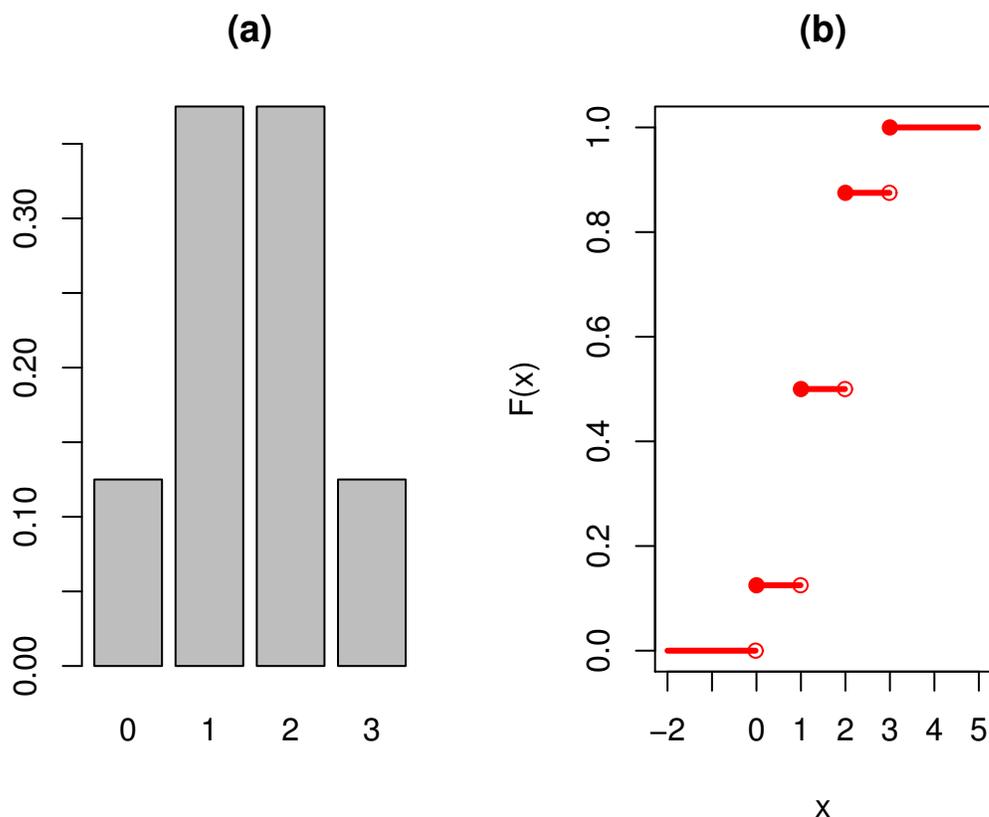


Figura 2.2: (a) Función de probabilidad y (b) Función de distribución acumulativa del número de caras en el lanzamiento de tres monedas (ejemplo 2.3)

hasta que sale cara", teniendo en cuenta que los resultados de los sucesivos lanzamientos constituyen sucesos independientes se tiene:

$$P(X = 1) = P(\{c\}) = \frac{1}{2}$$

$$P(X = 2) = P(\{xc\}) = P(\{x\} \cap \{c\}) = P(\{x\}) P(\{c\}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2^2} = \frac{1}{4}$$

$$\begin{aligned} P(X = 3) &= P(\{x xc\}) = P(\{x\} \cap \{x\} \cap \{c\}) = P(\{x\}) P(\{x\}) P(\{c\}) = \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2^3} = \frac{1}{8} \end{aligned}$$

⋮

$$\begin{aligned} P(X = k) &= P(\{x \dots xc\}) = P(\{x\} \cap \dots \cap \{x\} \cap \{c\}) = \\ &= P(\{x\})^{k-1} P(\{c\}) = \frac{1}{2^{k-1}} \cdot \frac{1}{2} = \frac{1}{2^k} \end{aligned}$$

⋮

(Obsérvese que esta variable aleatoria podría tomar infinitos valores ya que, al menos en teoría, cabe la posibilidad de que en los sucesivos lanzamientos salga siempre cruz, por lo que el experimento no se detiene nunca). Por tanto la función de distribución de esta variable aleatoria, para $n = 1, 2, 3, \dots$, viene dada por³:

$$F(n) = P(X \leq n) = \sum_{k=1}^n P(X = k) = \sum_{k=1}^n \frac{1}{2^k} = \frac{\frac{1}{2} - \frac{1}{2^{n+1}}}{1 - \frac{1}{2}} = 1 - \frac{1}{2^n}$$

La figura 2.3 muestra las gráficas de la función de probabilidad $P(X = k)$ y la función de distribución acumulativa $F(x)$, sólo para los valores $x \in [0, 10]$.

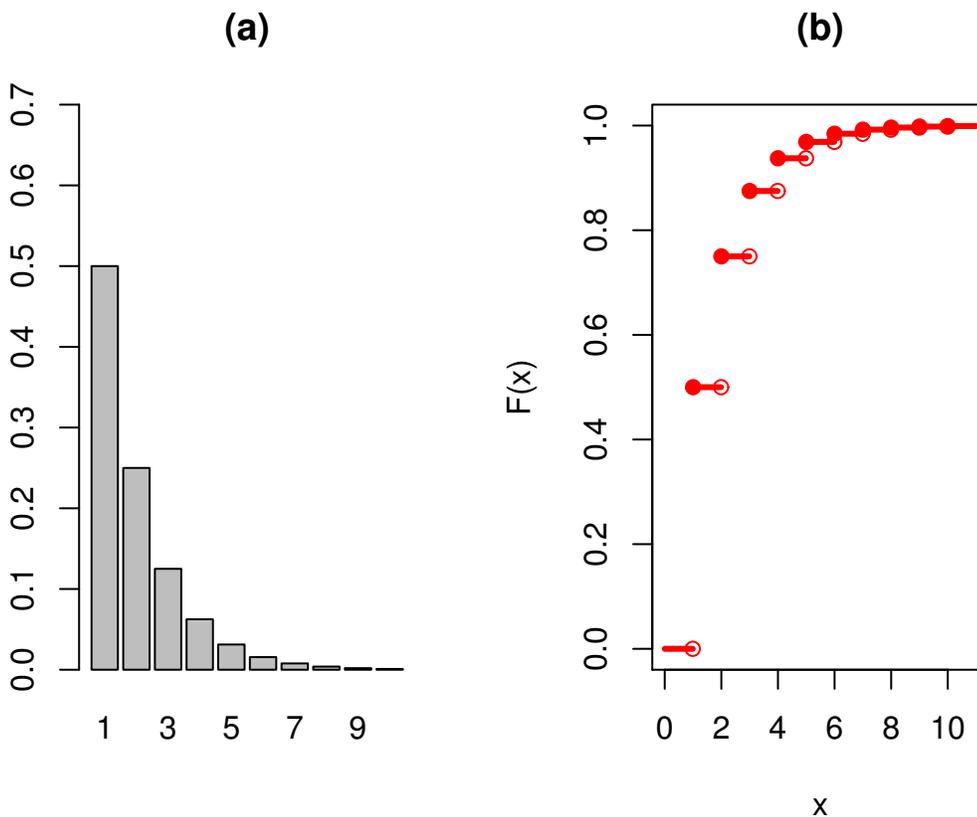


Figura 2.3: (a) Función de probabilidad y (b) Función de distribución acumulativa del número de lanzamientos de una moneda hasta que sale cara por primera vez (ejemplo 2.4).

Como hemos visto en los ejemplos 2.3 y 2.4, las variables aleatorias discretas se caracterizan por tener funciones de distribución acumulativa *escalonadas*, que se van incrementando a

³Es preciso utilizar que la suma de los n primeros términos de una progresión geométrica de razón menor que la unidad es $\sum_{k=1}^n \rho^k = \frac{1-\rho^{n+1}}{1-\rho}$

saltos. Las posiciones de los saltos corresponden a los valores que puede tomar la variable aleatoria. A su vez, la magnitud de cada salto es igual a la probabilidad de observar el valor correspondiente al punto de salto. Entre salto y salto, la función de distribución es constante.

Propiedades de la función de probabilidad de una variable aleatoria discreta

Sea $M = \{n_1, n_2, n_3, \dots\}$ el conjunto (finito o infinito numerable) de todos los posibles valores que puede tomar una variable aleatoria discreta X . Suponemos además que $n_1 < n_2 < n_3 < \dots$, y llamemos $f(n) = P(X = n)$. Las siguientes propiedades se siguen inmediatamente de la definición de $f(n)$:

1. $0 \leq f(x) \leq 1$ para todo $x \in \mathbb{R}$
2. $\sum_{n_j \in M} f(n_j) = 1$
3. $F(n_k) = \sum_{j \leq k} f(n_j)$
4. $f(n_k) = F(n_k) - F(n_{k-1})$

2.5.2. Variables aleatorias continuas.

Las variables aleatorias cuya función de distribución acumulativa es continua reciben el nombre de *variables aleatorias continuas*. Se caracterizan por tomar valores en un rango continuo (intervalo), sin que haya puntos en los que se acumule la probabilidad; dicho de otra forma, si X es una v.a. continua, $P(X = x) = 0$ para cualquier valor $x \in \mathbb{R}$.

Ejemplo 2.5. Realizamos el experimento consistente en tirar de los extremos de una cuerda de 1 metro de longitud hasta que se parte. Suponemos que la cuerda está fabricada con un material completamente homogéneo, de forma que a priori es igualmente probable que se rompa en cualquier punto. Consideremos la variable aleatoria $X = \text{“Posición del punto en que se parte la cuerda”}$.

Dado que existen infinitos puntos entre 0 y 1 en los que la cuerda puede romperse (todos equiprobables, por ser la cuerda homogénea), la regla de Laplace nos indicaría que la probabilidad de que se rompa en un punto x concreto es 0, cualquiera que sea x :

$$P(X = x) = 0 \quad \forall x \in [0, 1]$$

Ahora bien, si consideramos el punto medio ($x = \frac{1}{2}$), por ser la cuerda homogénea la probabilidad de que se parta a la izquierda de ese punto debe ser igual a la probabilidad de que

se parta a la derecha; por tanto $P(X \leq \frac{1}{2}) = \frac{1}{2}$. De igual forma, si consideramos el punto $x = \frac{1}{3}$, como el trozo a la izquierda de este punto mide una tercera parte de la longitud total de la cuerda, nuevamente la homogeneidad de ésta implica que $P(X \leq \frac{1}{3}) = \frac{1}{3}$. En general, el mismo argumento nos permite concluir que para cualquier $x \in [0, 1]$, $P(X \leq x) = x$. Asimismo, como la cuerda no puede partirse antes de $x = 0$, se tiene $P(X < 0) = 0$; y como tampoco puede partirse después de $x = 1$, resulta $P(X \leq x) = 1$ para los $x > 1$.

Observemos, pues, que aunque para esta variable sea $P(X = x) = 0 \quad \forall x \in [0, 1]$, el razonamiento anterior nos ha permitido construir su función de distribución acumulativa $F(x) = P(X \leq x)$ para cualquier valor $x \in \mathbb{R}$:

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

La figura 2.4 muestra gráficamente esta función de distribución.

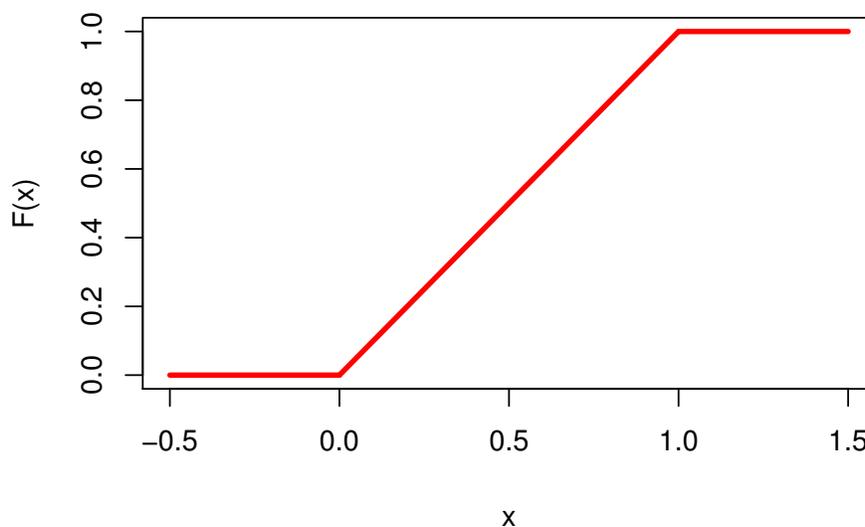


Figura 2.4: Función de distribución acumulativa descrita en el ejemplo 2.5.

Un caso particular de variables aleatorias continuas son las *absolutamente continuas*, que se caracterizan porque su función de distribución es absolutamente continua. Esto significa que existe una función real f , positiva e integrable en el conjunto de números reales, tal que la función de distribución acumulativa F se puede expresar como:

$$F(x) = \int_{-\infty}^x f(u) du \quad (2.1)$$

La función f recibe el nombre de *función de densidad de probabilidad de la variable aleatoria* X . Este nombre no es arbitrario, ya que $f(x)$ admite una interpretación análoga a la del concepto físico de densidad. En efecto de la ecuación (2.1) se sigue que $f(x)$ es la derivada de $F(x)$ y por tanto:

$$\begin{aligned} f(x) &= F'(x) = \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = \\ &= \lim_{\Delta x \rightarrow 0} \frac{P(X \leq x + \Delta x) - P(X \leq x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x} \end{aligned}$$

lo que nos indica que $f(x)$ representa la cantidad de probabilidad en un entorno próximo de x , dividida por la medida Δx de ese entorno. Utilizando un símil físico, $P(x \leq X \leq x + \Delta x)$ puede entenderse como la *masa* total de probabilidad que se concentra en un *volumen* Δx alrededor de x . *Masa* partido por *volumen* es precisamente la definición clásica de densidad, lo que justifica el nombre de la función f .

Asimismo, de la expresión anterior se sigue también que para un valor Δx suficientemente pequeño:

$$P(X \in (x, x + \Delta x]) \cong f(x)\Delta x$$

lo que significa que la probabilidad de que la variable aleatoria X esté dentro de un intervalo muy pequeño que contenga a un valor x es aproximadamente igual a $f(x)$ veces la amplitud de dicho intervalo. Geométricamente, el término $f(x)\Delta x$ representa el área de un rectángulo de base Δx y altura $f(x)$.

Continuación del ejemplo 2.5: Recordemos que en este ejemplo considerábamos la variable aleatoria X = “punto donde se rompe una cuerda homogénea de 1 metro de longitud al tirar de sus extremos”. La función de distribución de esta variable era de la forma:

$$F(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x < 1 \\ 1 & x > 1 \end{cases}$$

Derivando obtenemos la función de densidad :

$$f(x) = \begin{cases} 0 & x < 0 \\ 1 & 0 \leq x \leq 1 \\ 0 & x > 1 \end{cases}$$

Como vemos, esta función es constante en el intervalo $[0, 1]$, lo que se corresponde con la idea intuitiva de que, por ser la cuerda homogénea, es igualmente probable que se rompa en cualquier punto; por tanto la densidad de dicha probabilidad debe ser constante a lo largo de todo el recorrido de la cuerda.

Nota: Si bien es posible definir variables aleatorias continuas que no sean absolutamente continuas, constituyen la excepción antes que la regla. La inmensa mayoría de las variables aleatorias continuas que nos encontramos en las aplicaciones son también absolutamente continuas. Por ello, con el objetivo de simplificar la terminología, cuando en este texto utilicemos la expresión *variable aleatoria continua* nos estaremos refiriendo en realidad a variables aleatorias *absolutamente continuas*, y por tanto con función de densidad bien definida.

Propiedades de la función de densidad de probabilidad de variables aleatorias continuas.

1. $\int_{-\infty}^{\infty} f(x) dx = 1$
2. $f(x) \geq 0$ para todo $x \in \mathbb{R}$
3. $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) = \int_a^b f(x) dx$

La última propiedad nos indica que la probabilidad de que una variable aleatoria continua X tome valores entre dos puntos a y b coincide con el área bajo la función de densidad entre esos dos puntos.

Continuación del ejemplo 2.5: La probabilidad de que la cuerda se parta entre los puntos 0.3 y 0.7 puede calcularse como:

$$P(0,3 < X \leq 0,7) = \int_{0,3}^{0,7} f(x) dx = \int_{0,3}^{0,7} 1 dx = [x]_{0,3}^{0,7} = 0,7 - 0,3 = 0,4$$

donde hemos tenido en cuenta que $f(x) = 1$ para $x \in [0, 1]$. La figura 2.5 muestra el significado geométrico de esta integral. La probabilidad que se ha calculado es el área bajo la función $f(x) = 1$ entre 0,3 y 0,7, que en este caso corresponde simplemente a un rectángulo.

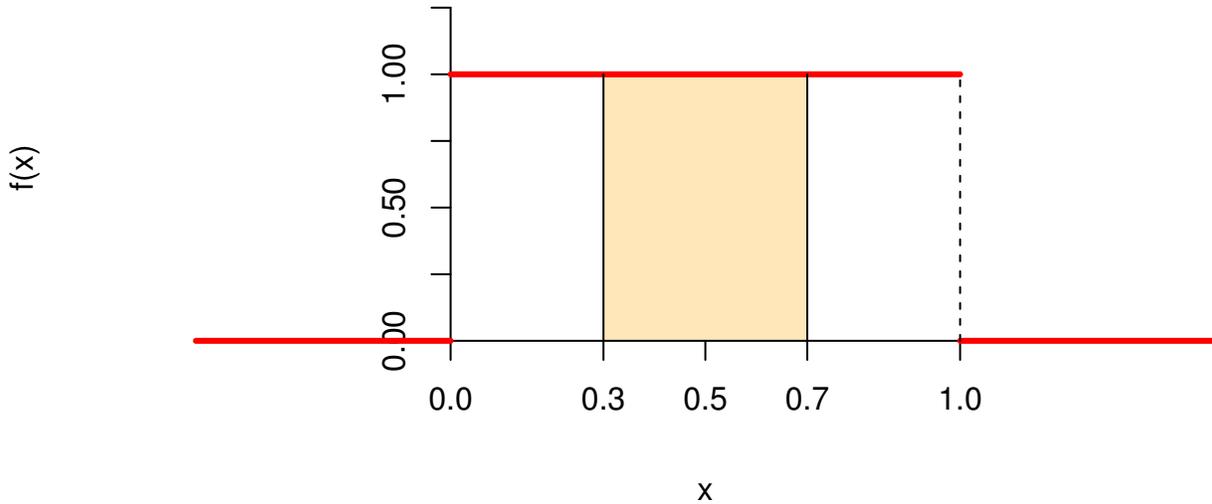


Figura 2.5: La línea de trazo grueso representa la función de densidad $f(x)$ de la variable aleatoria descrita en el ejemplo 2.5 (punto aleatorio en que se rompe una cuerda de un metro). El área coloreada representa la probabilidad de que la cuerda se rompa entre los puntos 0,3 y 0,7.

Ejemplo 2.6. En la desembocadura de muchos ríos es frecuente encontrar radioisótopos (plomo 210, cesio 137 y otros) que pueden ser utilizados como trazadores del arrastre de materiales sedimentarios. Se ha comprobado que la probabilidad de detectar uno de estos radioisótopos disminuye exponencialmente con la profundidad de muestreo en el lecho marino. En particular, en el estuario de cierto río, la variable X = “Profundidad (en cm.) a la que es detectable la presencia de ^{210}Pb ” tiene como función de densidad

$$f(x) = \begin{cases} 0,1e^{-0,1x}, & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Obviamente $f(x)$ está bien definida como función de densidad, ya que $f(x) \geq 0, \forall x$ y además:

$$\int_0^{\infty} 0,1e^{-0,1x} dx = [-e^{-0,1x}]_0^{\infty} = 1$$

Si se desea obtener la probabilidad de detectar ^{210}Pb entre 5 y 15 cm. de profundidad calcu-

lamos simplemente:

$$\begin{aligned} P(5 \leq X \leq 15) &= \int_5^{15} 0,1e^{-0,1x} dx = [-e^{-0,1x}]_5^{15} = \\ &= e^{-0,1 \cdot 5} - e^{-0,1 \cdot 15} = 0,38 \end{aligned}$$

La figura 2.6 muestra la función de densidad de esta variable. La probabilidad que se acaba de calcular corresponde al área bajo esta función entre los valores 5 y 15, que se ha representado también en esta gráfica.

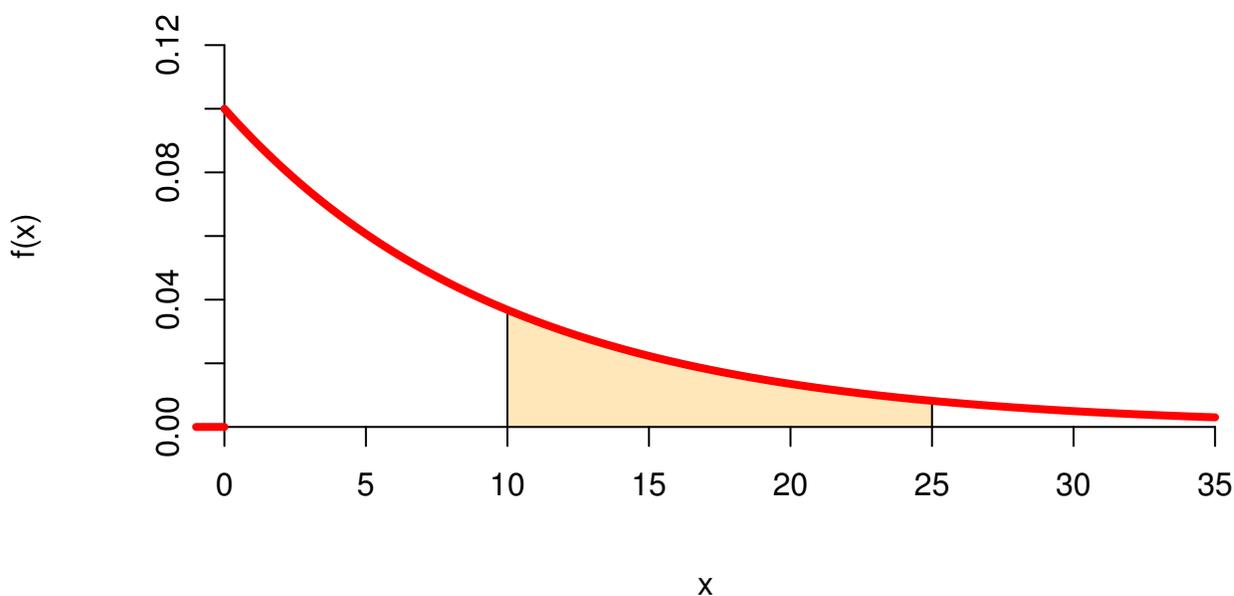


Figura 2.6: Función de densidad de la variable descrita en el ejemplo 2.6

Ejemplo 2.7. En ingeniería de costas resulta de interés modelar la distribución de probabilidad de la altura de ola. En particular es importante conocer la probabilidad de que dicha altura supere ciertos valores, ya que ello determina las características que han de tener las construcciones costeras. La función de densidad:

$$f(x) = \begin{cases} \vartheta x e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

constituye un modelo simple que puede emplearse en algunos casos. Para que esta función de densidad esté bien definida, el área total bajo la misma debe ser 1, esto es:

$$\int_0^{\infty} \vartheta x e^{-\lambda x} dx = 1$$

Resolvemos esta integral (es sencillo integrar por partes):

$$\int_0^{\infty} \vartheta x e^{-\lambda x} dx = \vartheta \left[-\frac{x}{\lambda} e^{-\lambda x} - \frac{1}{\lambda^2} e^{-\lambda x} \right]_0^{\infty} = \frac{\vartheta}{\lambda^2}$$

Por tanto, para que esta integral valga 1 deberá ocurrir que si $\vartheta = \lambda^2$, en cuyo caso $f(x)$ corresponde a una función de densidad correctamente definida cualquiera que sea el valor de λ . Supongamos que $\lambda = 0,9$ y que se desea calcular la probabilidad de que la altura de ola supere los 4 metros. Entonces, si $X = \text{“Altura de ola”}$:

$$\begin{aligned} P(X \geq 4) &= \int_4^{\infty} \vartheta x e^{-\lambda x} dx = \lambda^2 \left[-\frac{x}{\lambda} e^{-\lambda x} - \frac{1}{\lambda^2} e^{-\lambda x} \right]_4^{\infty} = \\ &= 0,9^2 \left(\frac{4}{0,9} e^{-0,9 \cdot 4} + \frac{1}{0,9^2} e^{-0,9 \cdot 4} \right) = e^{-0,9 \cdot 4} (0,9 \cdot 4 + 1) = 0,126 \end{aligned}$$

La figura 2.7 muestra la gráfica de esta función de densidad. La probabilidad que se acaba de calcular corresponde al área bajo esta curva desde el valor 4 en adelante.

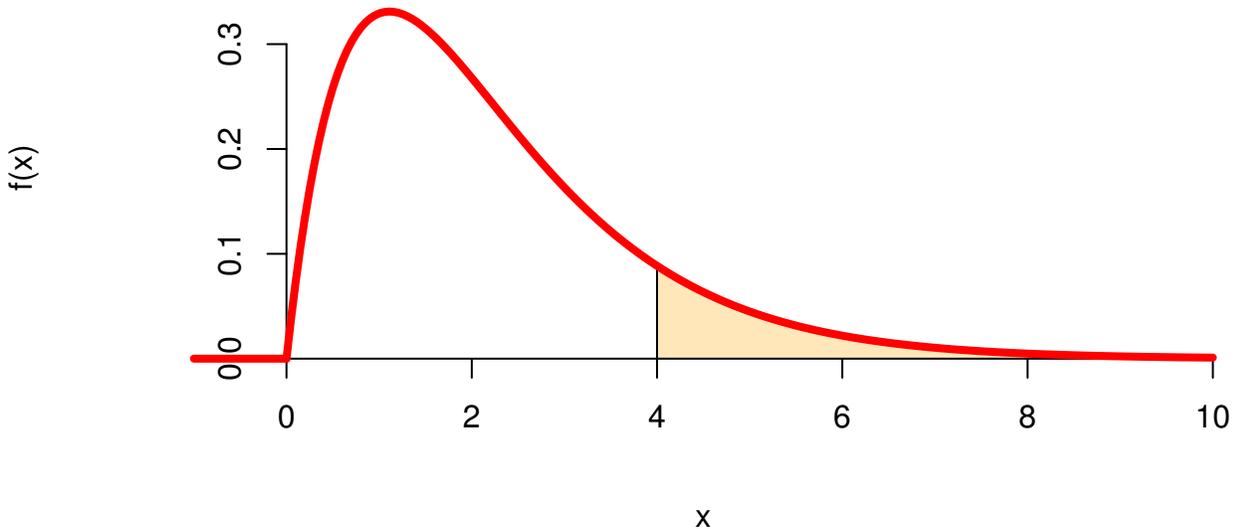


Figura 2.7: Función de densidad de la altura de ola (ejemplo 2.7). Se ha sombreado la probabilidad de que una ola supere los 4 metros.

Así pues, la función de distribución de una variable aleatoria (o sus derivadas, la función de probabilidad en el caso discreto y la función de densidad en el caso continuo) es la herramienta que permite modelar la incertidumbre presente en los procesos de observación o

experimentación. En los ejemplos que acabamos de ver –punto de rotura de una cuerda, profundidad a la que se detecta un isótopo radiactivo, altura de ola– el valor que toma la variable es impredecible *a priori*, pero las funciones de densidad de probabilidad asociadas a estas variables determinan qué rangos de valores tienen más o menos probabilidad de ocurrir. La distribución de probabilidad, pues, modela el efecto del conjunto de causas que dan origen a dichos valores. Permitiéndonos cierto abuso del lenguaje, podemos decir que la distribución de probabilidad es la que *genera* los valores que observamos en las variables aleatorias, produciendo más valores en las regiones con mayor probabilidad y menos en el resto. La figura 2.8 representa esta idea. Se han reproducido de nuevo las funciones de densidad de los últimos ejemplos, pero representando en la base de cada figura puntos correspondientes a 300 observaciones de las respectivas variables (puntos de rotura de 300 cuerdas homogéneas, altura de 300 olas, y profundidad a la que se ha detectado ^{210}Pb en 300 muestras). Como puede apreciarse, en (a) las observaciones se reparten uniformemente en el intervalo $[0, 1]$, en consonancia con una densidad de probabilidad constante; en (b) y en (c) se observa que los valores observados tienden a concentrarse en las regiones con mayor densidad de probabilidad, disminuyendo su número a medida que disminuye la densidad.

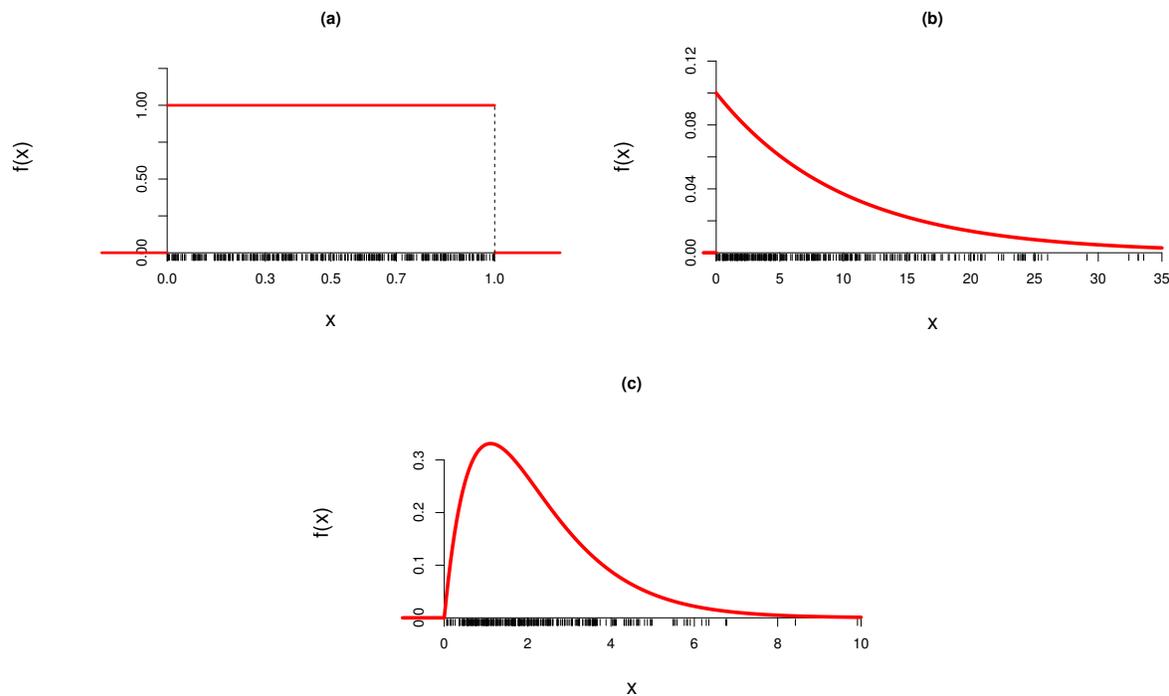


Figura 2.8: Densidades de probabilidad de las variables descritas en los ejemplos 2.5, 2.6 y 2.7. Sobre los ejes de abcisas se han representado las posiciones de 300 valores observados en estas variables.

2.5.3. Variables aleatorias mixtas.

En el caso de que la función de distribución tenga saltos, y además tramos continuos en los que sea estrictamente creciente (no constante), la variable aleatoria es *mixta*. Una variable aleatoria mixta se caracteriza, por tanto, porque toma valores en intervalos continuos, a la vez que existen uno o más valores discretos para los que $P(X = x) > 0$. En este curso no nos ocuparemos de este tipo de variables.

2.6. Variables aleatorias multidimensionales.

En muchas ocasiones se realizan múltiples medidas sobre los objetos de nuestro estudio. Así por ejemplo, en el curso de un trabajo de campo sobre tortugas marinas, en cada ejemplar podemos medir su longitud (X), peso (Y) y perímetro de la concha (Z). De esta forma, cada observación da lugar a un vector (x, y, z) . Este vector es una variable aleatoria dado que a priori, antes de capturar cada ejemplar, no podemos predecir su valor. Por ello este vector recibe el nombre de *variable aleatoria multidimensional* (o *vector aleatorio*).

2.6.1. Distribución conjunta de variables aleatorias.

Los conceptos de función de distribución acumulativa, función de probabilidad y función de densidad de probabilidad se generalizan fácilmente al caso multidimensional. Por simplicidad, a continuación enunciamos estos conceptos sólo para el caso bidimensional. Dado un vector aleatorio (X, Y) :

- La función $F(x, y) = P(X \leq x \cap Y \leq y)$ recibe el nombre de *función de distribución conjunta* del vector (X, Y) .
- Cuando X e Y son discretas, la función $f(x, y) = P(X = x \cap Y = y)$ recibe el nombre de *función de probabilidad conjunta* del vector (X, Y) .
- Cuando X e Y son continuas y existe una función $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, integrable y no negativa, tal que:

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

si dice entonces que el vector (X, Y) tiene distribución absolutamente continua. En tal caso:

$$f(x, y) = \lim_{\Delta x \rightarrow 0, \Delta y \rightarrow 0} \frac{P(\{x < X \leq x + \Delta x\} \cap \{y < Y \leq y + \Delta y\})}{\Delta x \Delta y} = \frac{\partial^2}{\partial x \partial y} F(x, y)$$

recibe el nombre de *función de densidad de probabilidad* del vector (X, Y) .

Ejemplo 2.8. (Vector de variables discretas) Supongamos que en el lanzamiento de dos dados equilibrados consideramos la variable bidimensional (X, Y) , donde $X = \text{“Producto de las caras superiores”}$ e $Y = \text{“Suma de las caras superiores”}$. La tabla 2.1 muestra los posibles valores de la variable (X, Y) , así como los sucesos que los generan y su probabilidad. La figura 2.9 representa la función de probabilidad de esta variable aleatoria.

Suceso	(X, Y)	Probabilidad	Suceso	(X, Y)	Probabilidad
$\{(1, 1)\}$	(1, 2)	1/36	$\{(3, 4), (4, 3)\}$	(12, 7)	2/36
$\{(1, 2), (2, 1)\}$	(2, 3)	2/36	$\{(2, 6), (6, 2)\}$	(12, 8)	2/36
$\{(1, 3), (3, 1)\}$	(3, 4)	2/36	$\{(3, 5), (5, 3)\}$	(15, 8)	2/36
$\{(2, 2)\}$	(4, 4)	1/36	$\{(4, 4)\}$	(16, 8)	1/36
$\{(1, 4), (4, 1)\}$	(4, 5)	2/36	$\{(3, 6), (6, 3)\}$	(18, 9)	2/36
$\{(2, 3), (3, 2)\}$	(6, 5)	2/36	$\{(4, 5), (5, 4)\}$	(20, 9)	2/36
$\{(1, 5), (5, 1)\}$	(5, 6)	2/36	$\{(4, 6), (6, 4)\}$	(24, 10)	2/36
$\{(2, 4), (4, 2)\}$	(8, 6)	2/36	$\{(5, 5)\}$	(25, 10)	1/36
$\{(3, 3)\}$	(9, 6)	1/36	$\{(5, 6), (6, 5)\}$	(30, 11)	2/36
$\{(1, 6), (6, 1)\}$	(6, 7)	2/36	$\{(6, 6)\}$	(36, 12)	1/36
$\{(2, 5), (5, 2)\}$	(10, 7)	2/36			

Tabla 2.1: Función de probabilidad de la variable (X, Y) descrita en el ejemplo 2.8 ($X = \text{“Producto de las caras superiores resultantes de lanzar dos dados”}$ e $Y = \text{“Suma de las caras superiores”}$).

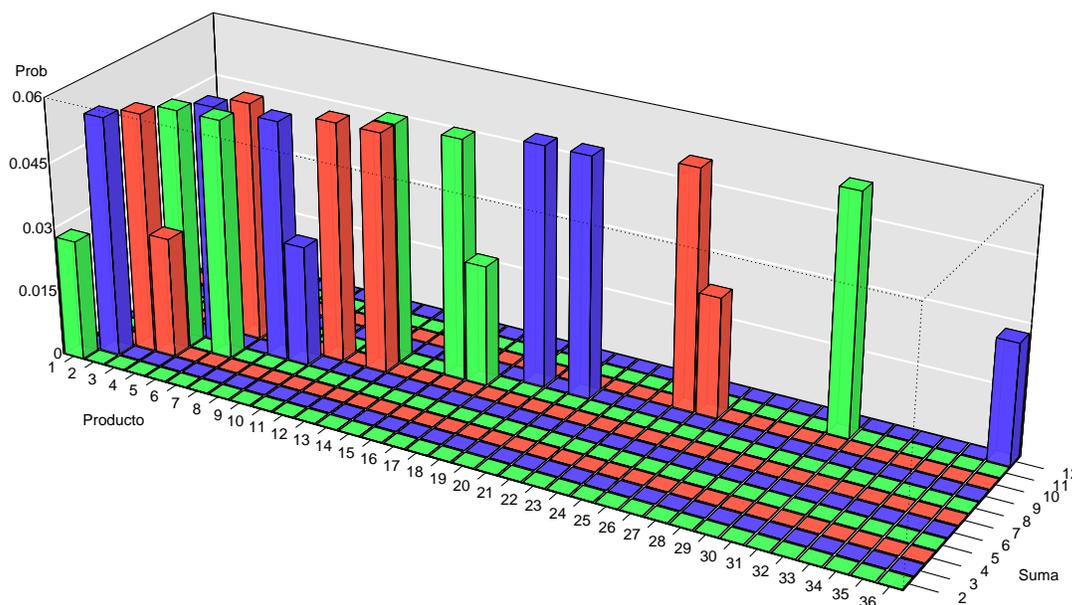


Figura 2.9: Representación gráfica de la función de probabilidad de la variable (X, Y) descrita en el ejemplo 2.8 (Tabla 2.1).

Ejemplo 2.9. (Vector de variables continuas) Un estudio morfométrico de peces de la familia de los *Serránidos*, subfamilia *Epinephelinae* ha permitido obtener una aproximación de la función de densidad conjunta $f(x, y)$ del vector aleatorio (X, Y) siendo $X = \text{“Longitud (cm)”}$ e $Y = \text{“Peso (kg)”}$ de los ejemplares de esta familia⁴. Esta aproximación se muestra en la figura 2.10. Del mismo modo que el área entre dos puntos bajo la función de densidad de una variable aleatoria unidimensional da la probabilidad de que la variable tome valores en ese rango, el *volumen* bajo la función de densidad bivalente sobre un entorno determinado da la probabilidad de que el vector aleatorio (X, Y) tome valores en dicho entorno.

La figura 2.11 muestra un conjunto de 1000 observaciones de $(\text{Longitud}, \text{Peso})$ que obedecen a esta distribución de probabilidad. Como puede apreciarse, donde la densidad de probabilidad encierra un mayor volumen (mayor probabilidad) se produce un mayor número de observaciones, disminuyendo este número a medida que disminuye el volumen; donde la densidad es cero (probabilidad nula), no se producen observaciones.

Obviamente el cálculo de probabilidades con variables aleatorias multidimensionales es más complejo que en el caso unidimensional, y no nos ocuparemos de él en este curso. No obstante

⁴El vector (X, Y) se entiende como aleatorio en el sentido de que, *a priori*, antes de medir cualquier ejemplar de esta familia no se pueden predecir su longitud ni su peso.

existe un caso, que se presenta con frecuencia en las aplicaciones prácticas, en el que las funciones que se acaban de definir adquieren una estructura simple. Es el caso de las *variables aleatorias independientes*.

2.6.2. Independencia de variables aleatorias.

Recordemos que dos sucesos A y B se dicen independientes si $P(A \cap B) = P(A) \cdot P(B)$. Esta definición puede generalizarse al caso de variables aleatorias. Así, dos variables aleatorias X e Y se dicen *estocásticamente independientes* o simplemente, *independientes*, si para cualesquiera $a, b, c, d \in \mathbb{R}$ los sucesos $\{a < X \leq b\}$ y $\{c < Y \leq d\}$ son independientes, esto es:

$$P(\{a < X \leq b\} \cap \{c < Y \leq d\}) = P(a < X \leq b) \cdot P(c < Y \leq d) \quad (2.2)$$

En lo que sigue llamaremos $F_X(x)$ y $F_Y(y)$ a las funciones de distribución respectivas de las variables X e Y . Asimismo, denotaremos por $f_X(x)$ y $f_Y(y)$ las respectivas funciones de probabilidad o densidad de probabilidad (según que X e Y sean discretas o continuas).

En el caso de que dos variables aleatorias X e Y sean independientes se cumplen las siguientes propiedades:

1. $F(x, y) = F_X(x) \cdot F_Y(y)$
2. $f(x, y) = f_X(x) \cdot f_Y(y)$

La demostración de estas propiedades puede encontrarse en el apéndice.

2.7. Parámetros característicos de las distribuciones de probabilidad.

En esta sección presentaremos algunas medidas que tienen como objetivo sintetizar –resumir– la distribución de probabilidad de una variable aleatoria en unos pocos valores característicos:

- *Esperanza*: Valor que describe dónde se encuentra el “centro” de la distribución de probabilidad.
- *Varianza*: Valor que describe el grado de dispersión de los valores que toma la variable aleatoria.

- *Momentos*: Valores que describen la *forma* de la distribución de probabilidad (asimetría, apuntamiento).
- *Cuantiles*: Valores por debajo de los cuales se acumula una determinada probabilidad (normalmente el 1 %, 2.5 %, 5 %, 25 %, 50 %, 75 %, 95 %, 97.5 %, 99 %).
- *Covarianza y Correlación*: Valores que cuantifican el grado de asociación lineal entre dos variables X e Y .

2.7.1. Esperanza matemática

La *esperanza matemática* de una variable aleatoria X se define como:

- Si X es discreta: $E[X] = \sum_k k \cdot P(X = k)$
- Si X es continua y tiene función de densidad $f(x)$: $E[X] = \int_{-\infty}^{\infty} x f(x) dx$

Si en el caso discreto identificamos la probabilidad de un valor con su *masa*, y en el caso continuo la densidad de probabilidad de un valor con la *densidad de masa en un entorno del mismo*, podemos interpretar la esperanza de una variable aleatoria como el *centro de gravedad* de su distribución de probabilidad. Más concretamente, si imaginamos la gráfica de la función de probabilidad (caso discreto) o de la densidad de probabilidad (caso continuo) como un objeto físico, la esperanza coincide con la posición del eje X en que deberíamos apoyar este objeto para que permanezca en equilibrio. La figura 2.12 muestra sendos ejemplos de la posición de la esperanza: en la figura (a) se muestra la función de probabilidad de una variable aleatoria discreta (concretamente la del ejemplo 2.2), y en la figura (b) la función de densidad de probabilidad de la altura de ola vista en el ejemplo 2.7. En ambos casos la posición de la esperanza se ha marcado con un pequeño triángulo. Se puede apreciar a simple vista que la esperanza corresponde al centro de gravedad en ambas figuras.

En ocasiones se requiere calcular la esperanza de alguna función⁵ g de la variable aleatoria X . En tal caso la esperanza de la variable aleatoria $g(X)$ se define de modo análogo a la anterior:

- Si X es discreta: $E[g(X)] = \sum_k g(k) \cdot P(X = k)$

⁵Por ejemplo, si tiramos una moneda y el resultado es una variable X que vale 1 si sale cara y 0 si sale cruz. En este caso $E[X]$ representa el número esperado de caras. Si decidimos apostar y ganamos 10 € cada vez que sale cara, y perdemos 10€ cada vez que sale cruz, podemos representar nuestra apuesta mediante la función $g(X)$, que vale 10 cuando $X = 1$ (cara) y -10 cuando $X = 0$ (cruz). En este caso $E[g(X)]$ representa nuestra ganancia (o pérdida) esperada durante el juego.

- Si X es continua y tiene función de densidad $f(x)$: $E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$

En el caso de variables aleatorias bidimensionales (X, Y) la esperanza de una función $g(X, Y)$ se define como:

- Si (X, Y) es un vector de variables discretas,

$$E[g(X, Y)] = \sum_x \sum_y g(x, y) \cdot P(X = x, Y = y)$$

- Si (X, Y) tiene distribución absolutamente continua con función de densidad $f(x, y)$:

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy$$

Ejemplo 2.10. Para la variable aleatoria X definida en el ejemplo 2.2, correspondiente al resultado de la suma de las caras superiores resultantes al lanzar dos dados, la esperanza se obtiene fácilmente como:

$$\begin{aligned} E[X] &= \sum_{k=2}^{12} kP(X = k) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + \\ &\quad + 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = 7 \end{aligned}$$

Ejemplo 2.11. Para la variable aleatoria X definida en el ejemplo 2.5 (punto en que se parte una cuerda homogénea de un metro), la esperanza es:

$$E[X] = \int_{-\infty}^{\infty} xf(x) dx = \int_0^1 x \cdot 1 \cdot dx = \left[\frac{x^2}{2} \right]_0^1 = \frac{1}{2}$$

Ejercicio 2.1. Calcular la esperanza de las variables aleatorias definidas en los ejemplos 2.6 y 2.7.

Propiedades de la esperanza matemática.

1. Para cualquier constante arbitraria c :

$$E[c] = c$$

2. Dadas una variable aleatoria X , y una constante arbitraria c :

$$E[cX] = cE[X]$$

3. Dadas dos variables aleatorias X e Y :

$$E[X + Y] = E[X] + E[Y]$$

4. Si X e Y son independientes, entonces:

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

La demostración de estas propiedades se encuentra en el apéndice.

2.7.2. Medidas de dispersión de una variable aleatoria.

La *varianza* es una medida de dispersión de los valores de una variable aleatoria X . Si la esperanza es $\mu = E[X]$, la varianza se define como:

$$Var(X) = E[(X - \mu)^2]$$

La varianza es, pues, el valor esperado de la distancia al cuadrado entre los valores que toma la variable aleatoria y su esperanza⁶; si los valores están muy agrupados, estarán muy cerca de su centro (la esperanza) y la varianza será pequeña; por contra, si los valores de X está muy alejados entre sí, lo estarán también de su centro, y la varianza será grande. Por tanto la varianza es, efectivamente, una medida de dispersión.

Dada su definición, es obvio que las unidades en que se mide la varianza corresponden al cuadrado de las unidades en que se mide la variable X . Esto resulta poco práctico en muchas ocasiones, por lo que se suele emplear como medida de dispersión la *desviación típica* definida

⁶Esta distancia se toma al cuadrado para evitar la presencia de valores negativos, que pueden falsear su significado.

como⁷:

$$\text{sd}(X) = \sqrt{\text{var}(X)}$$

Es habitual denotar la desviación típica de una variable aleatoria mediante la letra griega σ . De la misma forma, la varianza suele denotarse como σ^2 .

La figura 2.13 muestra tres funciones de densidad correspondientes a variables aleatorias con la misma esperanza $E[X] = 0$, y con distintas desviaciones típicas. Como puede apreciarse, a medida que aumenta la desviación típica, la densidad se distribuye en un rango más amplio (la variable toma valores más dispersos). Nótese también que como el área total bajo la función de densidad debe ser siempre 1, cuando se incrementa el rango que abarca dicha función, su altura disminuye.

Propiedades de la varianza.

1. Dadas una variable aleatoria X , y una constante arbitraria c :

$$\text{var}(cX) = c^2 \text{var}(X)$$

$$\text{var}(c + X) = \text{var}(X)$$

2. $\text{var}(X) = E[X^2] - (E[X])^2$

3. Si X e Y son variables aleatorias independientes, $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$

La demostración de estas propiedades se encuentra en el apéndice.

Desigualdad de Chebyshev.

La desigualdad de Chebyshev permite utilizar la varianza de una variable aleatoria para acotar el valor de ciertas probabilidades que resultan de interés práctico. Concretamente, si X es una variable aleatoria tal que $E[X] = \mu$ y $\text{var}(X) = \sigma^2$ esta desigualdad establece que para todo $k \geq 1$:

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

En otras palabras, la probabilidad de que X tome valores que disten de su esperanza menos de k veces su desviación típica es al menos $1 - \frac{1}{k^2}$. Así, por ejemplo:

- Eligiendo $k = 2$: $P(|X - \mu| \leq 2\sigma) \geq \frac{3}{4} = 0,75$

⁷Utilizamos aquí la notación *sd* para la desviación típica, que deriva de su denominación inglesa *standard deviation*.

- Eligiendo $k = 3$: $P(|X - \mu| \leq 3\sigma) \geq 1 - \frac{1}{9} = 0,89$
- Eligiendo $k = 4$: $P(|X - \mu| \leq 4\sigma) \geq 1 - \frac{1}{16} = 0,9375$

En cualquier caso, es importante darse cuenta de que la desigualdad de Chebyshev establece una cota inferior para estas probabilidades y puede alejarse mucho de la probabilidad exacta. Así por ejemplo (con $k = 2$) la desigualdad nos indica que la probabilidad de que los valores de X se diferencien de μ en menos de 2 desviaciones típicas es *al menos* 0.75, pero según como sea la distribución de X , esa probabilidad podría en realidad ser 0.8, 0.9, ó 0.95, por ejemplo.

Relación entre esperanza y media aritmética.

Supongamos que la variable aleatoria X mide alguna característica de los sujetos de una población (peso, talla, temperatura, ...), y sean $\mu = E[X]$ y $\sigma^2 = \text{var}(X)$. Se eligen *al azar y de manera independiente* n sujetos de esa población. Llamaremos *muestra aleatoria simple* a los valores X_1, X_2, \dots, X_n que toma la variable X cuando se evalúa sobre cada uno de esos sujetos. X_1, X_2, \dots, X_n son a su vez variables aleatorias, toda vez que sus valores no se conocen antes de haber sido medidos. Asimismo, como todos los sujetos proceden de la misma población, las X_i tendrán la misma distribución de probabilidad de X , por lo que $E[X_i] = \mu$ y $\text{var}(X_i) = \sigma^2$ para $i = 1, \dots, n$.

La media aritmética de las observaciones, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, es también una variable aleatoria, ya que no es posible conocer su valor antes de haber obtenido la muestra. Cada posible muestra producirá unos valores distintos de X_1, X_2, \dots, X_n , y por tanto un valor distinto de \bar{X} . Tiene sentido, por tanto, que nos preguntemos por cuál es el valor esperado de \bar{X} (el centro de masas de todos los posibles valores que puede tomar) y cuál es su varianza. Ambos valores son fáciles de obtener. Aplicando las propiedades de la esperanza, tenemos:

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} E\left[\sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} n\mu = \mu$$

Asimismo, aplicando las propiedades de la varianza:

$$\text{var}(\bar{X}) = \text{var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \text{var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Por tanto, a medida que aumenta el valor de n , la varianza de \bar{X} se va reduciendo, de tal forma que cuando n es grande $\text{var}(\bar{X}) \cong 0$. Ello significa que para valores grandes de n

el valor de \bar{X} apenas se aparta de su valor esperado μ . De esta forma, cuando n es grande $\bar{X} \cong \mu$. Ello nos permite interpretar la esperanza de una variable aleatoria como la media aritmética de los valores observados de la misma en muestras aleatorias muy grandes.

2.7.3. Momentos.

Dada una variable aleatoria X , el *momento de orden k respecto al origen* (o simplemente *momento de orden k*), con $k \in \mathbb{N}$, se define como:

$$\mu_k = E[X^k]$$

Asimismo, si la esperanza de X es $E[X] = \mu$, se define el *momento de orden k respecto a la esperanza* (o *momento central de orden k*) como:

$$M_k = E[(X - \mu)^k]$$

Obviamente $\mu_1 = E[X]$ y $M_2 = \text{var}(X) = \mu_2 - \mu_1^2$. Los momentos centrales está relacionados con la forma de la distribución de probabilidad. Ya hemos visto que la varianza (que coincide con el momento central de orden 2) es una medida de dispersión. A partir del momento central de orden 3 se define el *coeficiente de asimetría*:

$$A = \frac{1}{\sigma^3} E[(X - \mu)^3]$$

y a partir del momento central de orden 4, el *coeficiente de apuntamiento o curtosis*:

$$\kappa = \frac{1}{\sigma^4} E[(X - \mu)^4] - 3$$

La figura 2.14 muestra funciones de densidad con diversos grados de asimetría:

- *Asimetría negativa*: la *masa* de probabilidad tiende a concentrarse a la derecha; en este caso el coeficiente de asimetría es negativo.
- *Asimetría positiva*: la *masa* de probabilidad tiende a concentrarse a la izquierda; en este caso el coeficiente de asimetría es positivo.
- *Simetría*: La masa de probabilidad se reparte simétricamente respecto a su centro (la esperanza). En este caso el coeficiente de asimetría es nulo.

La figura 2.15 muestra las funciones de densidad de tres variables aleatorias con las mismas esperanza y varianza, pero con distintos grados de apuntamiento:

- *Curtois negativa* ($\kappa < 0$): corresponde a funciones de densidad más bien aplanadas y con “colas” cortas. Las curvas con esta forma reciben el nombre de *platicúrticas*.
- *Curtois positiva* ($\kappa > 0$): corresponde a funciones de densidad más bien “puntiagudas” y con colas largas. Las curvas con esta forma se llaman *leptocúrticas*.
- *Curtois nula* ($\kappa = 0$): corresponde al caso intermedio, con un pico redondeado y colas de tamaño intermedio, como ocurre con la curva en forma de campana. Las curvas de este tipo reciben el nombre de *mesocúrticas*.

2.7.4. Cuantiles

Dada una variable aleatoria X cuya función de distribución acumulativa es $F(x)$, se define el α -ésimo cuantil ($0 < \alpha < 1$) como el valor q_α , tal que $F(q_\alpha) = P(X \leq q_\alpha) = \alpha$.

Cuando $F(x)$ es estrictamente creciente la ecuación anterior tiene solución única. En el caso de que la variable aleatoria sea discreta, ya hemos visto que $F(x)$ es escalonada; y aún cuando X sea continua, podría ocurrir que su función de distribución acumulativa presente intervalos en los que su valor sea constante. En estos casos se define el α -ésimo cuantil como $q_\alpha = \min \{x : F(x) \geq \alpha\}$.

Hay algunos cuantiles de uso muy frecuente, que reciben su propio nombre:

- La *mediana* (Me) es el cuantil 0,5. Por tanto, la probabilidad de que la variable tome valores menores o iguales que la mediana es el 50 %, y que tome valores mayores que ella es otro 50 %. Por esta razón, la mediana se usa habitualmente como medida de posición central.
- Los *cuartiles* (Q_1, Q_2 y Q_3): corresponden a los cuantiles 0.25, 0.5 (mediana) y 0.75.
- Los *centiles* o *percentiles* (P_k): corresponden a los cuantiles de la forma $\frac{k}{100}$, $k = 1, \dots, 100$

Ejemplo 2.12. En el ejemplo 2.6 vimos que la profundidad a que se detecta el isótopo ^{210}Pb es una variable aleatoria cuya densidad de probabilidad puede modelarse por $f(x) = 0,1e^{-0,1x}$. La función de distribución es entonces:

$$P(X \leq x) = F(x) = \int_0^x f(s) ds = \int_0^x f(x) = \int_0^x 0,1e^{-0,1s} ds = 1 - e^{-0,1x}$$

2.7. PARÁMETROS CARACTERÍSTICOS DE LAS DISTRIBUCIONES DE PROBABILIDAD.29

Para calcular cualquier cuantil α bastará con resolver la ecuación $F(q_\alpha) = \alpha$, que en este caso queda de la forma:

$$1 - e^{-0,1q_\alpha} = \alpha \Rightarrow e^{-0,1q_\alpha} = 1 - \alpha \Rightarrow q_\alpha = -\frac{1}{0,1} \log(1 - \alpha) = -10 \log(1 - \alpha)$$

Así, por ejemplo, la mediana sería $Me = -10 \log 0,5 = 6,93$, y el percentil 95 sería $P_{95} = -10 \log 0,05 = 29,96$.

Ejemplo 2.13. En el ejemplo 2.7 hemos visto que la altura de ola (en metros) en cierta zona puede modelarse mediante una variable aleatoria con función de densidad $f(x) = \lambda^2 x e^{-\lambda x}$, $x \geq 0$, $\lambda = 0,9$. Se desean calcular los cuantiles 0.025 y 0.975.

Para ello obtenemos primero la función de distribución acumulativa:

$$\begin{aligned} F(x) &= \int_0^x f(s) ds = \int_0^x \lambda^2 s \cdot e^{-\lambda s} ds = \lambda \left[-s e^{-\lambda s} - \frac{1}{\lambda} e^{-\lambda s} \right]_0^x = \\ &= \lambda \left(\frac{1}{\lambda} - x e^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x} \right) = 1 - e^{-\lambda x} (1 + \lambda x) \end{aligned}$$

Para encontrar el cuantil α hemos de resolver $F(q_{0,025}) = 0,025$. Por tanto (teniendo en cuenta que $\lambda = 0,9$):

$$\begin{aligned} 1 - e^{-0,9 \cdot q_{0,025}} (1 + 0,9 q_{0,025}) &= 0,025 \\ 0,975 - e^{-0,9 \cdot q_{0,025}} (1 + 0,9 q_{0,025}) &= 0 \end{aligned}$$

Esta ecuación obviamente no puede resolverse de manera explícita, así que utilizamos la función `uniroot` de R. La figura 2.7 nos indica que el cuantil buscado debe estar en el intervalo (0, 1):

```
Q = function(qa) {
  0.975 - exp(-0.9 * qa) * (1 + 0.9 * qa)
}
uniroot(Q, interval = c(0, 1))$root

## [1] 0.2691
```

El cuantil 0.975 se obtiene de modo análogo, salvo que buscamos en el intervalo (5, 10):

```

Q = function(qa) {
  0.025 - exp(-0.9 * qa) * (1 + 0.9 * qa)
}
uniroot(Q, interval = c(5, 10))$root

## [1] 6.191

```

De esta forma, con una probabilidad 0.95, la altura de ola en esta zona se encuentra entre los 0.269 y los 6.191 metros, esto es, $P(0,269 < X \leq 6,191) = 0,95$

2.7.5. Asociación lineal entre variables aleatorias.

Covarianza.

Dadas dos variables aleatorias X e Y , con esperanzas respectivas $E[X]$ y $E[Y]$, se define la *covarianza* entre ambas variables como:

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

La covarianza es, pues, el valor esperado del producto $(X - E[X])(Y - E[Y])$, lo que significa que:

- Si este valor es positivo X e Y varían conjuntamente en el mismo sentido: en efecto, el producto $(X - E[X])(Y - E[Y])$ es positivo solo si valores positivos de $(X - E[X])$ tienden a ir acompañados de valores positivos de $(Y - E[Y])$, y valores negativos de $(X - E[X])$ tienden a ir acompañados de valores negativos de $(Y - E[Y])$. O, dicho de otra forma, si valores de X superiores a $E[X]$ tienden a ir acompañados de valores de Y mayores que $E[Y]$, y valores de X menores que $E[X]$ tienden a ir acompañados de valores de Y menores que $E[Y]$. Cuanto más fuerte sea esta tendencia, mayor será el valor de la covarianza.
- Si este valor es negativo X e Y varían conjuntamente sentidos opuestos: el producto $(X - E[X])(Y - E[Y])$ es negativo solo si valores positivos de $(X - E[X])$ tienden a ir acompañados de valores negativos de $(Y - E[Y])$, y valores negativos de $(X - E[X])$ tienden a ir acompañados de valores positivos de $(Y - E[Y])$. Dicho de otra forma, valores de X mayores que $E[X]$ tienden a ir acompañados de valores de Y menores que $E[Y]$, y valores de X menores que $E[X]$ tienden a ir acompañados de valores de

Y mayores que $E[Y]$. Cuánto más fuerte sea esta tendencia mayor (en valor absoluto) será la covarianza.

- Si este valor es nulo, entonces valores positivos y negativos de $(X - E[X])$ van acompañados indistintamente por valores positivos o negativos de $(Y - E[Y])$, de tal forma que los productos $(X - E[X])(Y - E[Y])$ positivos se cancelan con los negativos.

La figura 2.16(a) muestra la función de densidad de un vector aleatorio (X, Y) para el que $\text{cov}(X, Y) > 0$. Puede apreciarse que esta función de densidad concentra la mayor parte de la probabilidad a lo largo de una línea en el plano XY . La figura 2.16(b) muestra una nube de puntos generada por la densidad anterior (hay mayor densidad de puntos donde la densidad encierra mayor volumen). Se aprecia aún más claramente el alineamiento de los puntos a lo largo de una recta, que tiene pendiente positiva. En trazos punteados se han marcado las posiciones de las esperanzas de X e Y respectivamente, dividiendo el plano XY en cuatro cuadrantes. Como puede verse, precisamente debido a la presencia de esta relación lineal positiva entre la X y la Y , hay más puntos en los cuadrantes (2) y (4), justamente aquellos en los que $(X - E[X])(Y - E[Y]) > 0$; además, estos puntos se alejan más del centro, esto es de la posición de $(E[X], E[Y])$, por lo que la magnitud absoluta de los valores $(X - E[X])(Y - E[Y])$ asociados será también mayor. Todo ello indica que la existencia de una asociación lineal con pendiente positiva entre la X y la Y implica un valor positivo de la covarianza, tanto más grande cuanto mayor sea el grado de asociación lineal entre las variables (mejor el ajuste de los puntos a una recta).

Un razonamiento análogo sobre la figura 2.17 nos muestra que la existencia de una relación lineal de pendiente negativa entre X e Y se asocia con una covarianza negativa, tanto mayor en valor absoluto cuanto mejor sea el ajuste a una recta. Por último, la figura 2.18 nos muestra que cuando no hay asociación lineal entre las variables X e Y , se tiene que $\text{cov}(X, Y) = 0$, ya que los puntos se reparten por igual en los cuatro cuadrantes, cancelándose los términos $(X - E[X])(Y - E[Y])$ positivos con los negativos.

La figura 2.19 nos muestra otra situación. Nuevamente tenemos la densidad a la izquierda y una nube de puntos generada por esta densidad a la derecha. Claramente las variables X e Y no son independientes (*conocer el valor de X nos informa aproximadamente de cuál puede ser el valor de Y*). Sin embargo, en los cuadrantes (1) y (2) los valores de $(X - E[X])(Y - E[Y])$ son iguales y de signo contrario; lo mismo sucede con los cuadrantes (3) y (4), por lo que $\text{cov}(X, Y) = 0$. Por tanto *una covarianza nula no significa que no haya asociación entre las variables*, ya que de hecho podría existir una asociación no lineal como en este caso.

Propiedades de la covarianza.

1. $\text{cov}(aX, bY) = ab \text{cov}(X, Y)$
2. $\text{cov}(X, X) = \text{var}(X)$
3. $\text{cov}(X, Y) = E[XY] - E[X]E[Y]$
4. Si X e Y son independientes, $\text{cov}(X, Y) = 0$

La demostración de estas propiedades se encuentra en el apéndice.

Ejercicio 2.2. Calcular la covarianza de las variables $U = X \cdot Y$ y $V = X + Y$ siendo X e Y los resultados de las caras superiores del lanzamiento de dos dados (ver ejemplo 2.8).

Correlación.

Hemos visto que el signo de la covarianza nos indica si entre las variables existe una relación lineal con pendiente positiva o negativa. Sin embargo no nos informa de la *intensidad* de esa relación, ya que el valor de la covarianza depende de las unidades en que se midan las variables X e Y . Para evitar este problema se define el *coeficiente de correlación lineal de Pearson* como:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

siendo σ_X^2 y σ_Y^2 las varianzas de X e Y respectivamente. De esta definición se sigue inmediatamente que $\rho_{X,Y}$ es adimensional.

Propiedades del coeficiente de correlación.

1. Si X e Y son independientes, entonces $\rho_{X,Y} = 0$
2. $-1 \leq \rho \leq 1$
3. Si $|\rho| = 1$ entonces $Y = aX + b$ (los valores (X, Y) se disponen exactamente a lo largo de una recta)

La demostración de estas propiedades se encuentra en el apéndice.

Cuando $\rho_{X,Y} = 0$, las variables X e Y se dicen *in correladas*.

La primera de las propiedades anteriores nos indica que la independencia entre dos variables implica la incorrelación. Lo contrario en general no es cierto como se ha visto con las variables

representadas en la figura 2.19; estas variables están asociadas, pero como su covarianza es cero, también su correlación es cero.

Ejercicio 2.3. Calcular el coeficiente de correlación entre las variables del ejercicio 2.2.

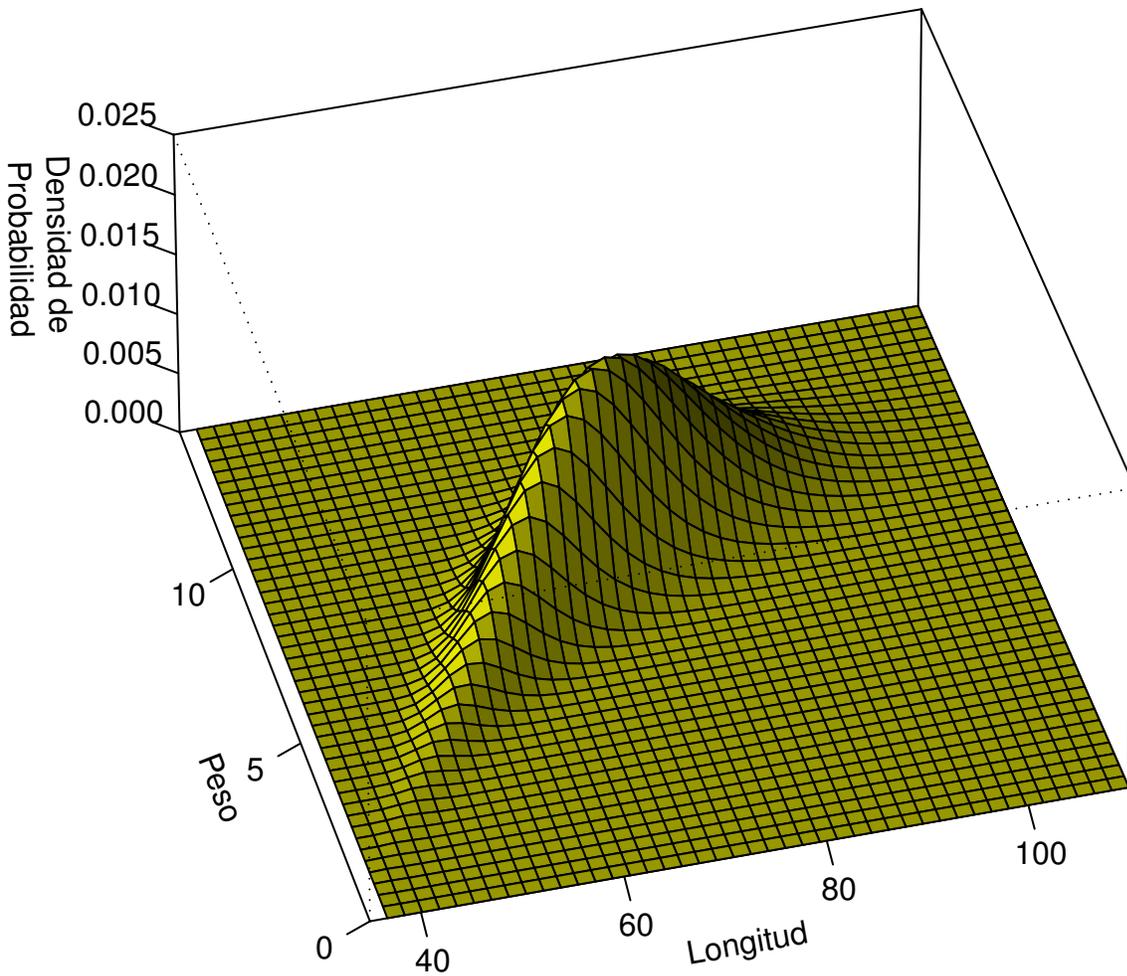


Figura 2.10: Función de densidad del vector aleatorio $(X, Y) = (\text{Longitud}, \text{Peso})$ para una población de peces de la familia *Serránidos*, subfamilia *Epinephelina* (ejemplo 2.9)

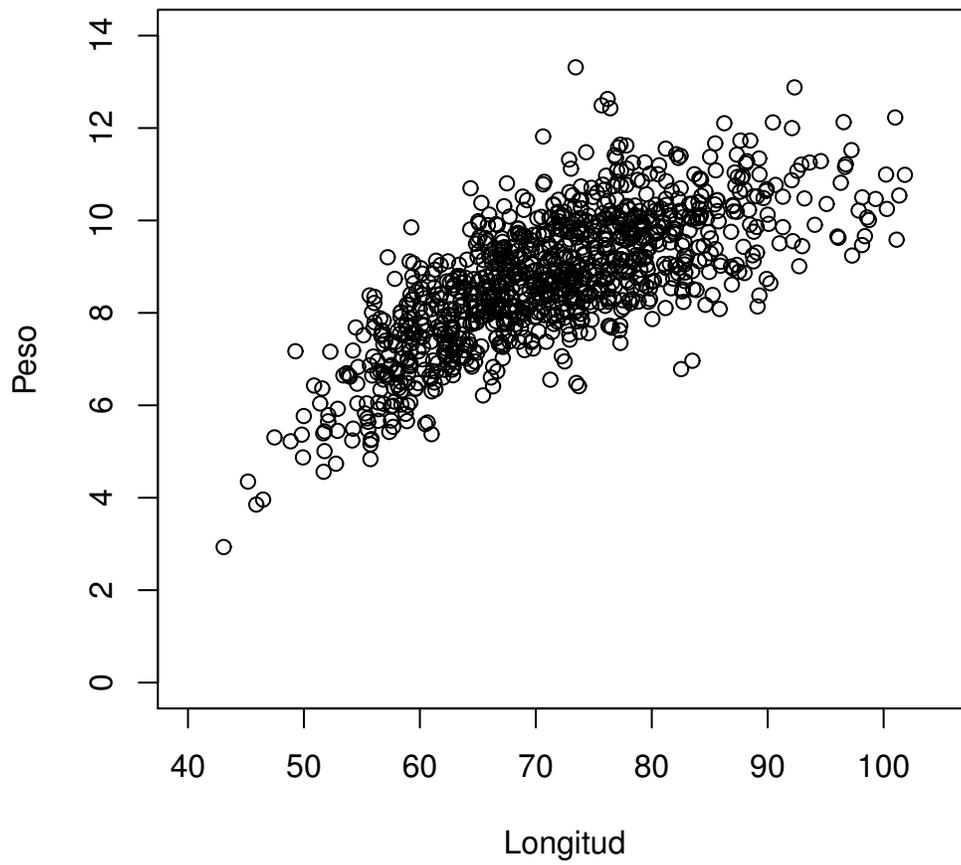


Figura 2.11: Nube de puntos correspondiente a la observación de la longitud y peso de 1000 peces del estudio descrito en el ejemplo 2.9.

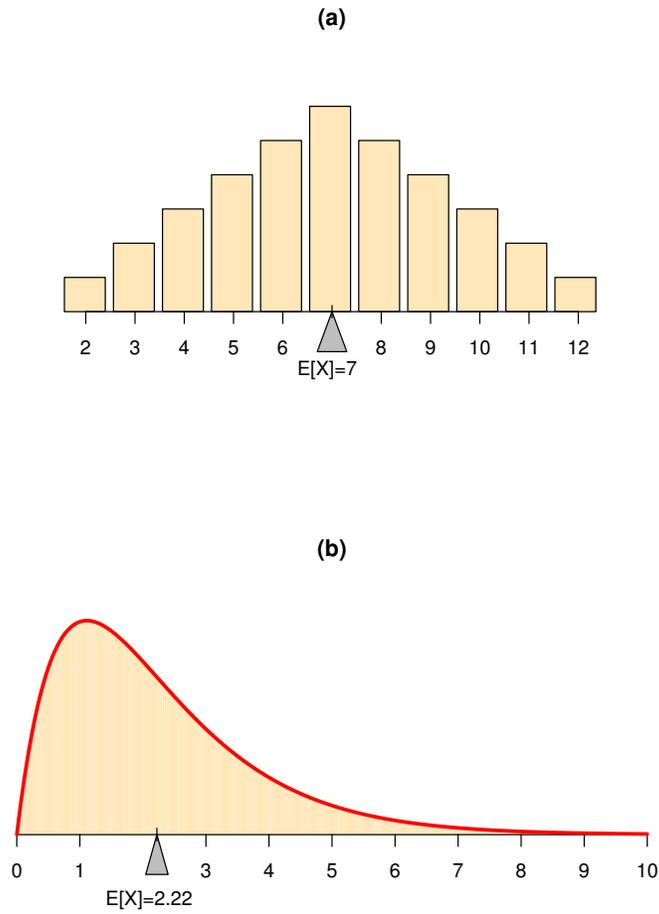


Figura 2.12: (a) Representación de la función de probabilidad de una variable aleatoria discreta (b) Representación de la densidad de probabilidad de una variable aleatoria continua. En ambos casos la posición de su esperanza (centro de gravedad de la figura) se representa mediante un triángulo.

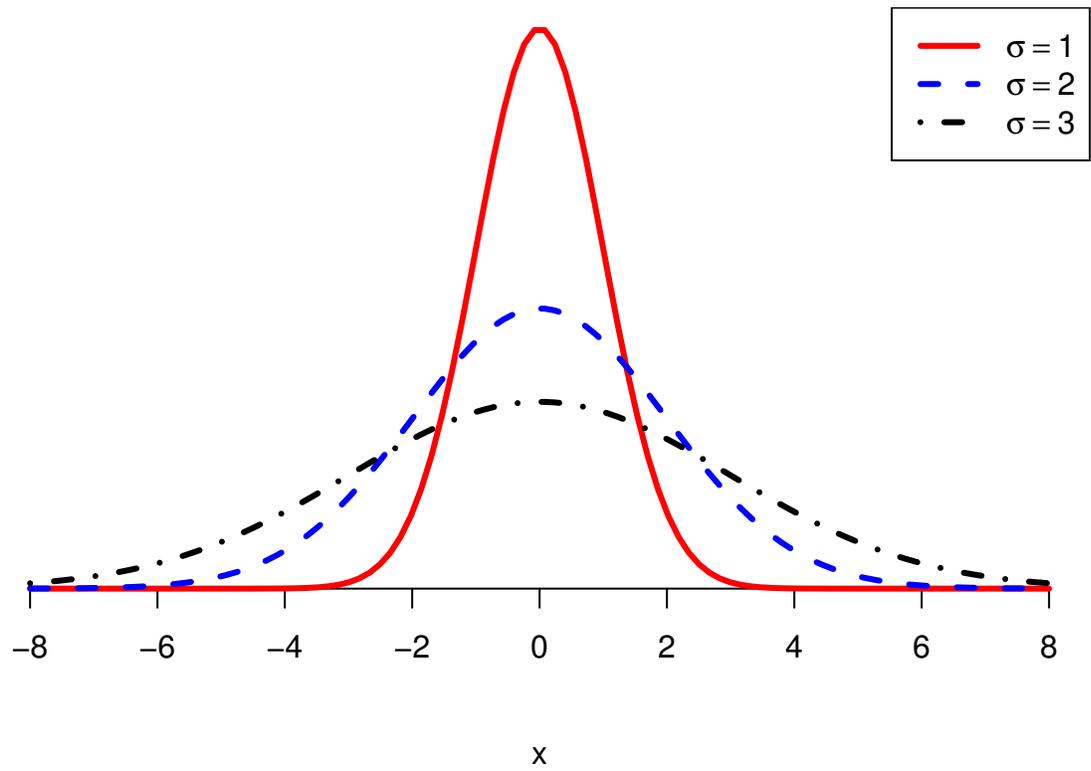


Figura 2.13: Funciones de densidad de tres variables aleatorias con distintas desviaciones típicas.

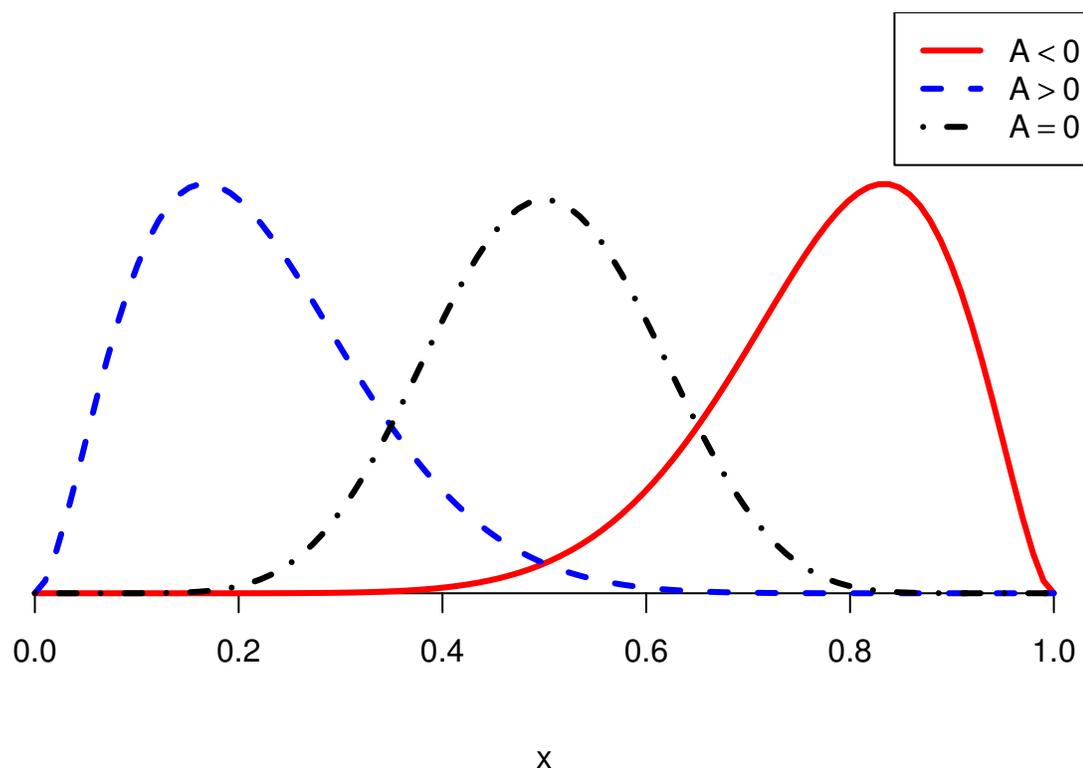


Figura 2.14: Funciones de densidad con diversos grados de asimetría.

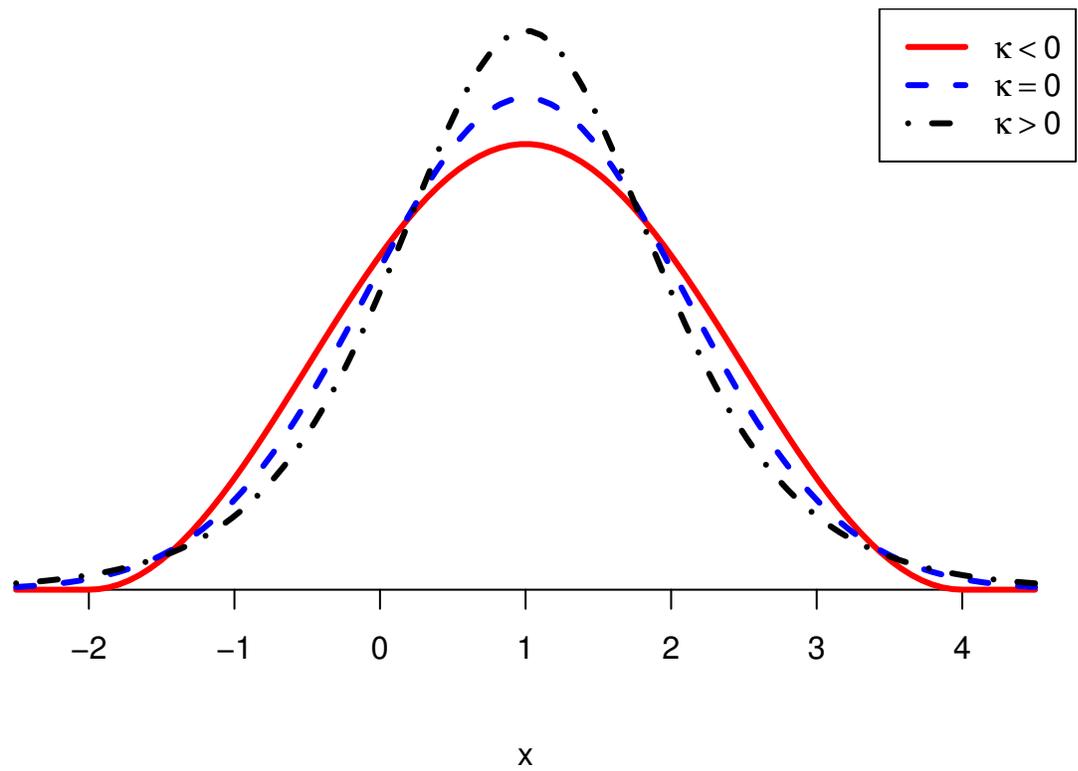


Figura 2.15: Funciones de densidad de tres variables aleatorias con distintos grados de apuntamiento. Las tres variables tienen distribución simétrica y las mismas esperanza y varianza.

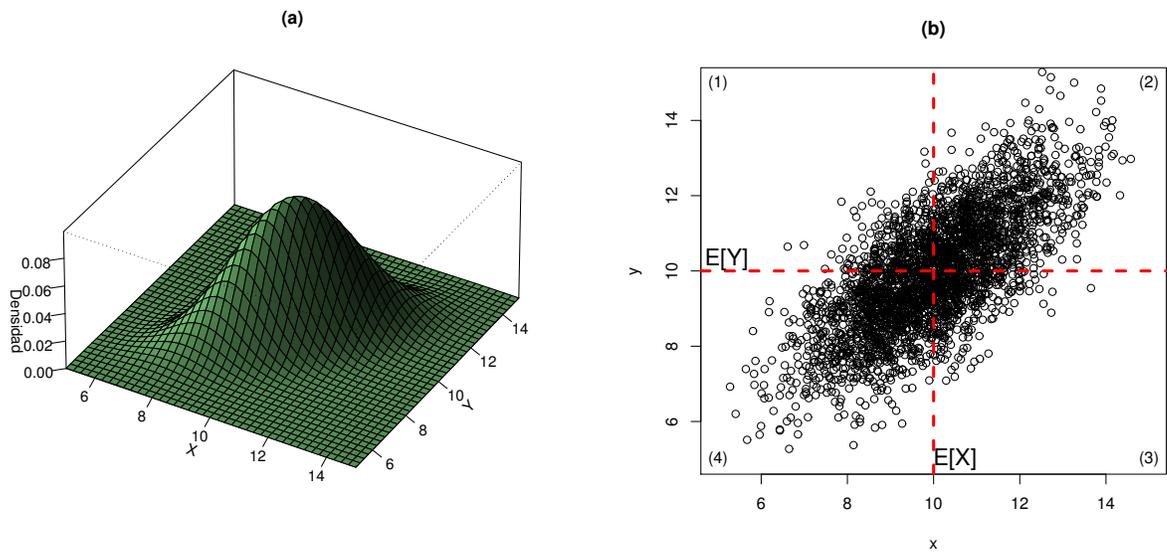


Figura 2.16: (a) Función de densidad de un vector aleatorio (X, Y) para el que $\text{cov}(X, Y) > 0$. (b) Nube de puntos generada por la función de densidad anterior.

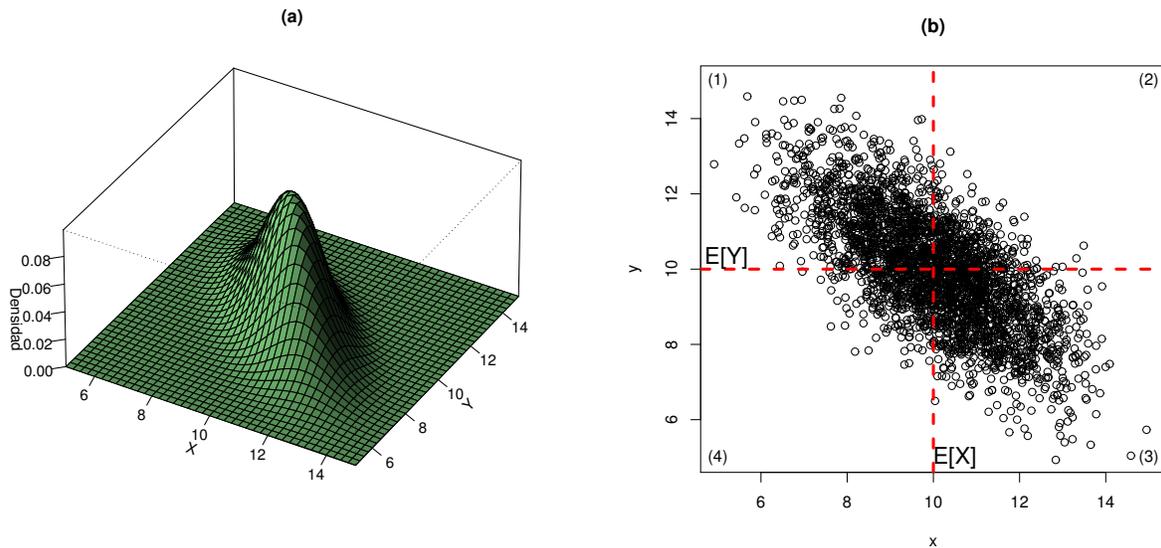


Figura 2.17: (a) Función de densidad de un vector aleatorio (X, Y) para el que $\text{cov}(X, Y) < 0$. (b) Nube de puntos generada por la función de densidad anterior.

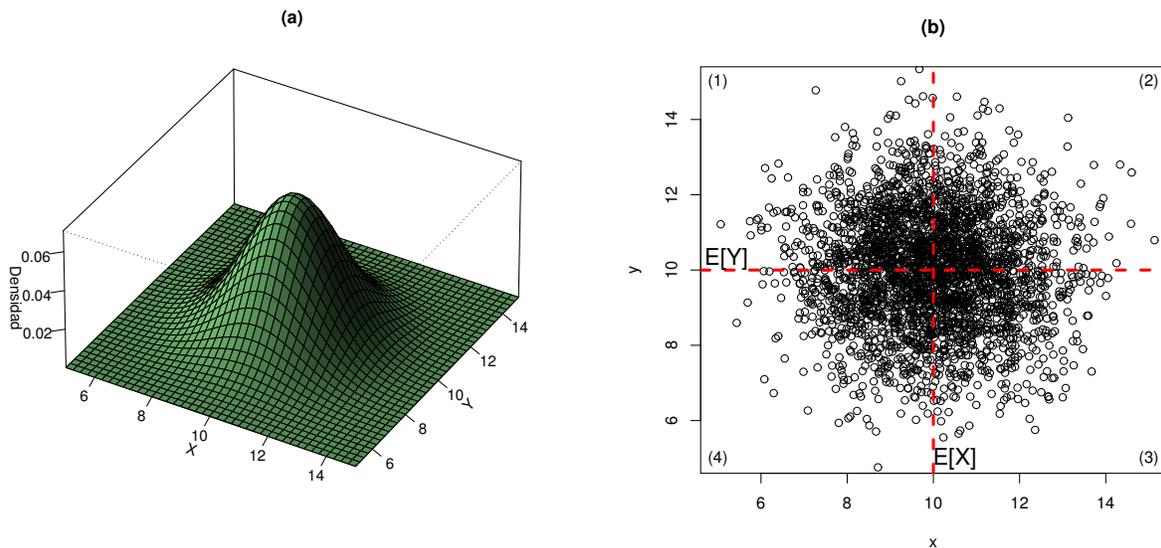


Figura 2.18: (a) Función de densidad de un vector aleatorio (X, Y) para el que $\text{cov}(X, Y) = 0$. (b) Nube de puntos generada por la función de densidad anterior. No se aprecia asociación entre las variables.

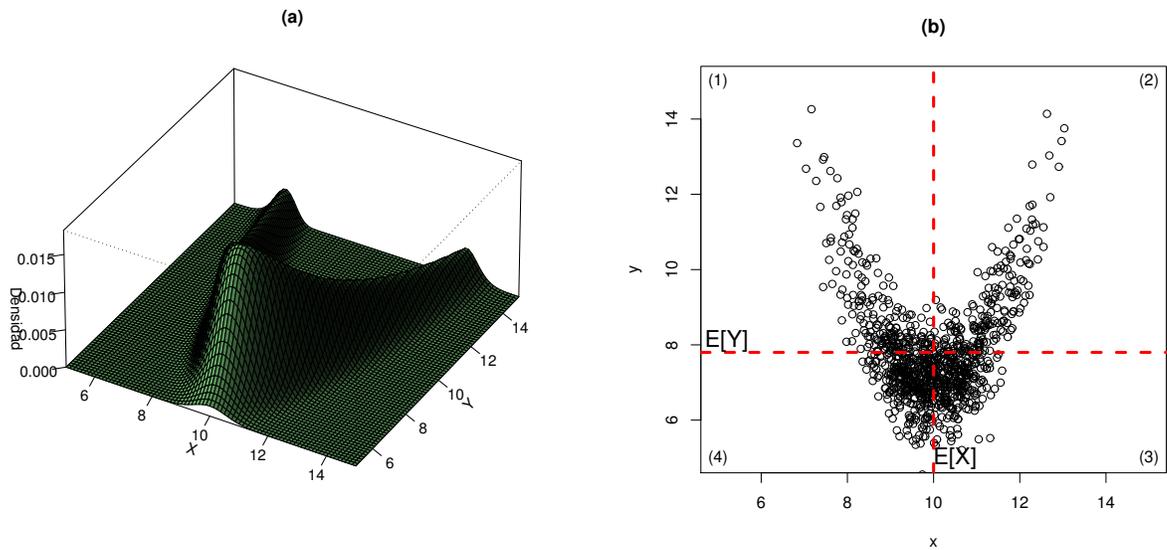


Figura 2.19: (a) Función de densidad de un vector aleatorio (X, Y) para el que $\text{cov}(X, Y) = 0$. (b) Nube de puntos generada por la función de densidad anterior. Entre X e Y se aprecia la existencia de una asociación no lineal.

Apéndice A

Demostraciones

Propiedades de la distribución conjunta de variables aleatorias independientes.

En el caso de que dos variables aleatorias X e Y sean independientes se cumplen las siguientes propiedades:

1. $F(x, y) = F_X(x) \cdot F_Y(y)$
2. $f(x, y) = f_X(x) \cdot f_Y(y)$

Demostración.

$$\begin{aligned} 1. \quad F(x, y) &= P(\{X \leq x\} \cap \{Y \leq y\}) = P(\{-\infty < X \leq x\} \cap \{-\infty < Y \leq y\}) = \\ &= P(-\infty < X \leq x) \cdot P(-\infty < Y \leq y) = P(X \leq x) \cdot P(Y \leq y) = F_X(x) \cdot F_Y(y) \end{aligned}$$

2. a) Si X e Y son discretas:

$$\begin{aligned} f(x, y) &= P(\{X = x\} \cap \{Y = y\}) = P(\{x - 1 < X \leq x\} \cap \{y - 1 < Y \leq y\}) = \\ &= P(x - 1 < X \leq x) \cdot P(y - 1 < Y \leq y) = P(X = x) \cdot P(Y = y) = f_X(x) \cdot f_Y(y) \end{aligned}$$

b) Si (X, Y) tiene distribución absolutamente continua:

$$\begin{aligned} f(x, y) &= \lim_{\Delta x \rightarrow 0, \Delta y \rightarrow 0} \frac{P(\{x < X \leq x + \Delta x\}) \cdot P(\{y < Y \leq y + \Delta y\})}{\Delta x \Delta y} = \\ &= \lim_{\Delta x \rightarrow 0} \frac{P(\{x < X \leq x + \Delta x\})}{\Delta x} \lim_{\Delta y \rightarrow 0} \frac{P(\{y < Y \leq y + \Delta y\})}{\Delta y} = f_X(x) f_Y(y) \end{aligned}$$

□

Propiedades de la esperanza matemática.

1. Para cualquier constante arbitraria c :

$$E[c] = c$$

2. Dadas una variable aleatoria X , y una constante arbitraria c :

$$E[cX] = cE[X]$$

3. Dadas dos variables aleatorias X e Y :

$$E[X + Y] = E[X] + E[Y]$$

4. Si X e Y son independientes, entonces:

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

Demostración.

1. Una constante c puede considerarse equivalente a una variable aleatoria I_c que toma el valor c con probabilidad 1. De esta forma, la función de probabilidad de esta variable es:

$$P(I_c = c) = 1$$

$$P(I_c = x) = 0 \quad \forall x \neq c$$

Su esperanza es entonces $E[c] = E[I_c] = \sum_x x \cdot P(I_c = x) = c \cdot P(I_c = c) = c$

2. La demostración de esta propiedad es trivial y se deja como ejercicio.
3. Demostraremos este resultado sólo en el caso de que ambas variables sean discretas. Como $X + Y$ es una función de dos variables aleatorias, para calcular su esperanza

hemos de utilizar su función de probabilidad conjunta:

$$\begin{aligned}
 E[X + Y] &= \sum_x \sum_y (x + y) \cdot P(X = x, Y = y) = \\
 &= \sum_x \sum_y x \cdot P(X = x, Y = y) + \sum_x \sum_y y \cdot P(X = x, Y = y) = \\
 &= \sum_x x \cdot \sum_y P(X = x, Y = y) + \sum_y y \cdot \sum_x P(X = x, Y = y) = \\
 &= \sum_x x \cdot P(X = x) + \sum_y y \cdot P(Y = y) = \\
 &= E[X] + E[Y]
 \end{aligned}$$

Aquí hemos utilizado que

$$\sum_y P(X = x, Y = y) = P(X = x) \quad \text{y que} \quad \sum_x P(X = x, Y = y) = P(Y = y)$$

Ambos resultados son triviales: los sucesos de la forma $\{Y = y\}$ forman un sistema completo de sucesos (el espacio muestral es $E = \cup_y \{Y = y\}$ y son incompatibles dos a dos, $\{Y = y_i\} \cap \{Y = y_j\} = \emptyset$ para cualesquiera $y_i \neq y_j$). Por tanto:

$$\begin{aligned}
 P(X = x) &= P(\{X = x\} \cap E) = P(\{X = x\} \cap (\cup_y \{Y = y\})) = \\
 &= P(\cup_y (\{X = x\} \cap \{Y = y\})) = \sum_y P(\{X = x\} \cap \{Y = y\})
 \end{aligned}$$

La demostración para el caso continuo es análoga, sustituyendo sumatorias por integrales y la función de probabilidad conjunta por la función de densidad conjunta.

4. En el caso discreto es $E[X \cdot Y] = \sum_x \sum_y x \cdot y \cdot P(X = x, Y = y) = \sum_i \sum_j x \cdot y \cdot f(x, y)$. Como X e Y son independientes $f(x, y) = f_X(x) f_Y(y)$, y por tanto:

$$E[X \cdot Y] = \sum_x \sum_y x \cdot y \cdot f_X(x) f_Y(y) = \left(\sum_x x \cdot f_X(x) \right) \left(\sum_y y \cdot f_Y(y) \right) = E[X] E[Y]$$

La demostración en el caso continuo es análoga cambiando sumatoria por integral.

□

Propiedades de la varianza.

1. Dadas una variable aleatoria X , y una constante arbitraria c :

$$\begin{aligned}\text{var}(cX) &= c^2 \text{var}(X) \\ \text{var}(c + X) &= \text{var}(X)\end{aligned}$$

2. $\text{var}(X) = E[X^2] - (E[X])^2$

3. Si X e Y son variables aleatorias independientes, $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$

Demostración.

- La primera propiedad se sigue directamente de la linealidad de la esperanza. En efecto, si $E[X] = \mu$ se tiene que $E[cX] = c\mu$, y de aquí: $\text{var}(cX) = E[(cX - c\mu)^2] = E[c^2(X - \mu)^2] = c^2 E[(X - \mu)^2] = c^2 \text{var}(X)$. Asimismo $\text{var}(c + X) = E[((c + X) - E(c + X))^2] = E[(c + X - E[c] - E[X])^2] = E[(X - E[X])^2] = \text{var}(X)$ ya que $E[c] = c$.
- La segunda propiedad se sigue desarrollando el cuadrado $(X - \mu)^2$ y aplicando la linealidad de la esperanza: $\text{var}(X) = E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2\mu E[X] + \mu^2 = E[X^2] - \mu^2$
- Para demostrar la tercera propiedad, llamando $\mu_X = E[X]$ y $\mu_Y = E[Y]$ y teniendo en cuenta que, por la segunda propiedad de la esperanza, $E[X + Y] = E[X] + E[Y] = \mu_X + \mu_Y$:

$$\begin{aligned}\text{var}(X + Y) &= E[(X + Y - (\mu_X + \mu_Y))^2] = E[((X - \mu_X) - (Y - \mu_Y))^2] = \\ &= E[(X - \mu_X)^2 - 2(X - \mu_X)(Y - \mu_Y) + (Y - \mu_Y)^2] = \\ &= E[(X - \mu_X)^2] - 2E[(X - \mu_X)(Y - \mu_Y)] + E[(Y - \mu_Y)^2] = \\ &= \text{var}(X) + \text{var}(Y) - 2E[(X - \mu_X)(Y - \mu_Y)]\end{aligned}$$

Veamos ahora que $E[(X - \mu_X)(Y - \mu_Y)] = 0$ cuando X e Y son independientes; en efecto:

$$\begin{aligned}E[(X - \mu_X)(Y - \mu_Y)] &= E[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y] = \\ &= E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y = \\ &= E[XY] - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y = E[XY] - \mu_X \mu_Y\end{aligned}$$

De acuerdo con la tercera propiedad de la esperanza, para variables independientes se tiene $E[XY] = E[X]E[Y] = \mu_X \mu_Y$, por lo que $E[X]E[Y] - \mu_X \mu_Y = 0$

□

Desigualdad de Chebyshev.

Si X es una variable aleatoria tal que $E[X] = \mu$ y $\text{var}(X) = \sigma^2$, entonces para todo $k \geq 1$:

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Presentamos la demostración para el caso de variables aleatorias discretas. El caso continuo es análogo sustituyendo sumatorias por integrales.

Demostración. Consideremos el suceso:

$$A = \{x : |x - \mu| \geq k\sigma\}$$

De la definición de varianza se tiene:

$$\sigma^2 = E[(X - \mu)^2] = \sum_x (x - \mu)^2 P(X = x) = \sum_{x \in A} (x - \mu)^2 P(X = x) + \sum_{x \in \bar{A}} (x - \mu)^2 P(X = x)$$

Como ambos sumandos son positivos:

$$\sigma^2 \geq \sum_{x \in A} (x - \mu)^2 P(X = x)$$

Ahora bien, tal como se ha definido el suceso A , para todos los $x \in A$ se tiene que $|x - \mu| \geq k\sigma$.

Por tanto:

$$\sigma^2 \geq \sum_{x \in A} (x - \mu)^2 P(X = x) \geq \sum_{x \in A} (k\sigma)^2 P(X = x) = (k\sigma)^2 \sum_{x \in A} P(X = x) = (k\sigma)^2 P(A)$$

De aquí se sigue que

$$P(A) \leq \frac{1}{k^2}$$

y por tanto

□

$$P(|X - \mu| < k\sigma) = 1 - P(A) \geq 1 - \frac{1}{k^2}$$

Propiedades de la covarianza.

1. $\text{cov}(aX, bY) = ab \text{cov}(X, Y)$
2. $\text{cov}(X, X) = \text{var}(X)$
3. $\text{cov}(X, Y) = E[XY] - E[X]E[Y]$
4. Si X e Y son independientes, $\text{cov}(X, Y) = 0$

Demostración. La demostración de las tres primeras propiedades es inmediata. La cuarta se sigue de la tercera y de que, como hemos visto en 2.7.1, si X e Y son independientes entonces $E[XY] = E[X]E[Y]$. \square

Propiedades del coeficiente de correlación.

1. Si X e Y son independientes, entonces $\rho_{X,Y} = 0$
2. $-1 \leq \rho \leq 1$
3. Si $|\rho| = 1$ entonces $Y = aX + b$ (los valores (X, Y) se disponen exactamente a lo largo de una recta)

Demostración.

1. La demostración de la primera propiedad es inmediata a partir de la propiedad 4 de la covarianza.
2. Para la segunda propiedad observemos que para cualquier constante a , y para cualesquiera variables aleatorias U y V se tiene que $E[(aU + V)^2] \geq 0$. Desarrollando el cuadrado y aplicando las propiedades de la esperanza resulta:

$$a^2 E[U^2] + 2aE[UV] + E[V^2] \geq 0$$

Esta ecuación representa una parábola que a lo sumo toca al eje de abcisas en un punto; por tanto, la ecuación $a^2 E[U^2] + 2aE[UV] + E[V^2] = 0$ tiene como mucho una solución, lo que significa que su discriminante debe ser menor o igual que cero, esto es, $4(E[UV])^2 - 4E[U^2]E[V^2] \leq 0$, o lo que es lo mismo:

$$(E[UV])^2 \leq E[U^2]E[V^2]$$

Si consideramos $U = (X - E[X])$ y $V = (Y - E[Y])$ se obtiene de inmediato la propiedad 2.

3. Por último, si $|\rho| = 1$ entonces siguiendo hacia atrás el argumento que acabamos de emplear, concluimos que existe una constante a tal que $E[(aU + V)^2] = 0$. Como los términos $(aU + V)^2$ son siempre mayores o iguales que 0 (por ser un cuadrado), la única forma de que su esperanza sea 0, es que $aU + V = 0$. Luego $a(X - E[X]) + (Y - E[Y]) = 0$, de donde $Y = aX - aE[X] + E[Y]$. Llamando $b = -aE[X] + E[Y]$ resulta la propiedad 3.

□