

Probabilidad y Estadística (Borradores, Curso 23)  
Variables aleatorias (1)  
Nociones básicas

Sebastian Grynberg

23-30 de marzo de 2011



*Alea jacta est*  
(Julio César)

# Índice

<b>1. Variables aleatorias</b>	<b>2</b>
1.1. Propiedades de la función de distribución . . . . .	6
1.2. Clasificación de variables aleatorias . . . . .	7
1.3. Cuantiles . . . . .	12
1.4. Construcción de variables aleatorias . . . . .	15
1.5. Función de distribución empírica e histogramas . . . . .	18
<b>2. Vectores aleatorios</b>	<b>22</b>
2.1. Distribución conjunta . . . . .	23
2.2. Distribuciones marginales . . . . .	25
2.3. Independencia . . . . .	29
<b>3. Variables truncadas</b>	<b>31</b>
3.1. Dividir y conquistar . . . . .	32
3.2. Memoria . . . . .	33
<b>4. Bibliografía consultada</b>	<b>34</b>

## 1. Variables aleatorias

Sea  $(\Omega, \mathcal{A}, \mathbb{P})$  un espacio de probabilidad. Una *variable aleatoria* sobre  $\Omega$  es una función  $X : \Omega \rightarrow \mathbb{R}$  tal que para cada  $x \in \mathbb{R}$

$$\{X \leq x\} := \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A},$$

i.e., para cada  $x \in \mathbb{R}$  el evento  $\{X \leq x\}$  tiene asignada probabilidad. La *función de distribución*  $F_X : \mathbb{R} \rightarrow [0, 1]$  de la variable aleatoria  $X$  se define mediante

$$F_X(x) := \mathbb{P}(X \leq x).$$

**Cálculo de probabilidades.** La función de distribución resume (y contiene) toda la información relevante sobre de la variable aleatoria. Para ser más precisos, para cada pareja de números reales  $a < b$  vale que <sup>1</sup>

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a). \tag{1}$$

□

---

<sup>1</sup>Basta observar que  $\{X \leq a\} \subset \{X \leq b\}$  y usar las propiedades de la probabilidad: de la igualdad  $\{a < X \leq b\} = \{X \leq b\} \setminus \{X \leq a\}$  se deduce que  $\mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F_X(b) - F_X(a)$ .

## Ejemplos

**Ejemplo 1.1** (Dado equilibrado). Sea  $X$  el resultado del lanzamiento de un dado equilibrado. Los posibles valores de  $X$  son 1, 2, 3, 4, 5, 6. Para cada  $k = 1, 2, 3, 4, 5, 6$ , la probabilidad de que  $X$  tome el valor  $k$  es  $1/6$ .

Sea  $x \in \mathbb{R}$ . Si  $x < 1$  es evidente que  $\mathbb{P}(X \leq x) = 0$ . Si  $k \leq x < k + 1$  para algún  $k = 1, 2, 3, 4, 5$ , la probabilidad del evento  $\{X \leq x\}$  es la probabilidad de observar un valor menor o igual que  $k$  y en consecuencia,  $\mathbb{P}(X \leq x) = k/6$ . Finalmente, si  $x \geq 6$  es evidente que  $\mathbb{P}(X \leq x) = 1$ .

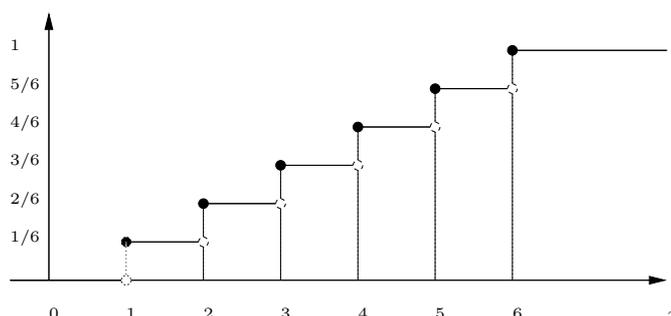


Figura 1: Gráfico de la función de distribución de un dado equilibrado.

Por lo tanto, la función de distribución de  $X$  se puede expresar del siguiente modo

$$F_X(x) = \sum_{k=1}^6 \frac{1}{6} \mathbf{1}\{x \geq k\}.$$

□

**Ejemplo 1.2** (Fiabilidad). Un problema fundamental de la ingeniería es el problema de la *fiabilidad*. Informalmente, la fiabilidad de un sistema se define como su capacidad para cumplir ciertas funciones prefijadas. Esta propiedad se conserva durante un período de tiempo hasta que ocurre una *falla* que altera la capacidad de trabajo del sistema. Por ejemplo: rupturas y cortocircuitos; fracturas, deformaciones y atascamientos de piezas mecánicas; el fundido o la combustión de las componentes de un circuito.

Debido a que las fallas pueden ocurrir como hechos casuales, podemos considerar que el *tiempo de funcionamiento*,  $T$ , hasta la aparición de la primer falla es una variable aleatoria a valores no negativos.

La fiabilidad de un sistema se caracteriza por su *función intensidad de fallas*  $\lambda(t)$ . Esta función temporal tiene la siguiente propiedad: cuando se la multiplica por  $dt$  se obtiene la probabilidad condicional de que el sistema sufra una falla durante el intervalo de tiempo  $(t, t + dt]$  sabiendo que hasta el momento  $t$  funcionaba normalmente. Si se conoce la función  $\lambda(t)$  se puede hallar la ley de distribución de probabilidades de  $T$ .

Para calcular la función de distribución de  $T$  estudiaremos dos eventos:  $A := \{T > t\}$  (el sistema funciona hasta el momento  $t$ ) y  $B := \{t < T \leq t + dt\}$  (el sistema sufre una

falla en el intervalo de tiempo  $(t, t + dt]$ . Como  $B \subset A$ , tenemos que  $\mathbb{P}(B) = \mathbb{P}(B \cap A)$  y de la regla del producto se deduce que

$$\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A). \quad (2)$$

Si la función de distribución de  $T$  admite derivada continua, salvo términos de segundo orden que se pueden despreciar, la probabilidad del evento  $B$  se puede expresar en la forma

$$\mathbb{P}(B) = \mathbb{P}(t < T \leq t + dt) = F_T(t + dt) - F_T(t) = F'_T(t)dt. \quad (3)$$

La probabilidad del evento  $A$  se puede expresar en la forma

$$\mathbb{P}(A) = \mathbb{P}(T > t) = 1 - \mathbb{P}(T \leq t) = 1 - F_T(t). \quad (4)$$

Finalmente, la probabilidad condicional  $\mathbb{P}(B|A)$  se expresa mediante la función intensidad de fallas  $\lambda(t)$ :

$$\mathbb{P}(B|A) = \lambda(t)dt \quad (5)$$

Sustituyendo las expresiones (3)-(5) en la fórmula (2) obtenemos, después de dividir ambos miembros por  $dt$ , una ecuación diferencial de primer orden para  $F_T(t)$

$$F'_T(t) = \lambda(t)(1 - F_T(t)). \quad (6)$$

Debido a que la duración del servicio del sistema no puede ser negativa, el evento  $\{T \leq 0\}$  es imposible. En consecuencia,  $F_T(0) = 0$ . Integrando la ecuación diferencial (6) con la condición inicial  $F(0) = 0$ , obtenemos <sup>2</sup>

$$F_T(t) = 1 - \exp\left(-\int_0^t \lambda(s)ds\right). \quad (7)$$

---

2

$$\begin{aligned} F'_T(t) = \lambda(t)(1 - F_T(t)) &\iff \frac{F'_T(t)}{1 - F_T(t)} = \lambda(t) \iff -\frac{d}{dt} \log(1 - F_T(t)) = \lambda(t) \\ &\iff \frac{d}{dt} \log(1 - F_T(t)) = -\lambda(t) \iff \log(1 - F_T(t)) = -\int_0^t \lambda(s)ds + C \\ &\iff 1 - F_T(t) = \exp\left(-\int_0^t \lambda(s)ds + C\right) \\ &\iff F_T(t) = 1 - \exp\left(-\int_0^t \lambda(s)ds + C\right). \end{aligned}$$

Usando que  $F_T(0) = 0$  se deduce que  $C = 0$ .

**Nota Bene.** El desarrollo anterior presupone que la función intensidad de fallas  $\lambda(t)$  verifica las siguientes condiciones: (1)  $\lambda(t) \geq 0$  para todo  $t > 0$  y (2)  $\int_0^\infty \lambda(t)dt = +\infty$ .  $\square$

**Ejemplo 1.3** (Fiabilidad). Se estipula que la duración de servicio de un sistema automático debe ser  $t_0$ . Si durante ese período el sistema falla, se lo repara y se lo utiliza hasta que sirva el plazo estipulado. Sea  $S$  el tiempo de funcionamiento del sistema después de la primera reparación. Se quiere hallar la ley de distribución de  $S$ .

En primer lugar observamos que la variable aleatoria  $S$  está ligada con el instante  $T$  en que ocurre la primera falla del sistema mediante la siguiente fórmula

$$S = \text{máx}(t_0 - T, 0) = \begin{cases} t_0 - T & \text{si } T \leq t_0, \\ 0 & \text{si } T > t_0. \end{cases}$$

Sea  $F_S(s)$  la función de distribución de la variable  $S$ . Es claro que para  $s < 0$ ,  $F_S(s) = 0$  y que para  $s \geq t_0$ ,  $F_S(s) = 1$ . Lo que resta es analizar el comportamiento de  $F_S$  sobre el intervalo  $0 \leq s < t_0$ . Sea  $s \in [0, t_0)$

$$\begin{aligned} F_S(s) &= \mathbb{P}(S \leq s) = \mathbb{P}(\text{máx}(t_0 - T, 0) \leq s) = \mathbb{P}(t_0 - T \leq s, 0 \leq s) \\ &= \mathbb{P}(t_0 - T \leq s) = \mathbb{P}(t_0 - s \leq T) = \exp\left(-\int_0^{t_0-s} \lambda(t)dt\right), \end{aligned}$$

donde  $\lambda(t)$  es la función intensidad de fallas del sistema.

Por lo tanto,

$$F_S(s) = \exp\left(-\int_0^{t_0-s} \lambda(t)dt\right) \mathbf{1}\{0 \leq s < t_0\} + \mathbf{1}\{s \geq t_0\}.$$

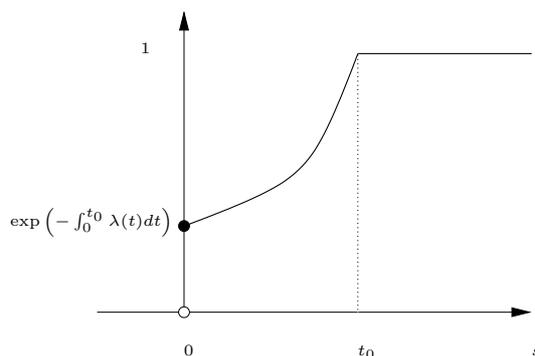


Figura 2: Gráfico de la función de distribución de la variable aleatoria  $S$ .

$\square$

---

## Ejercicios adicionales

1. Sea  $X$  una variable aleatoria con función de distribución  $F_X(x)$ . Mostrar que para cada pareja de números reales  $a < b$  vale que:

$$\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a) + \mathbb{P}(X = a) \quad (8)$$

$$\mathbb{P}(a \leq X < b) = F_X(b) - \mathbb{P}(X = b) - F_X(a) + \mathbb{P}(X = a) \quad (9)$$

$$\mathbb{P}(a < X < b) = F_X(b) - \mathbb{P}(X = b) - F_X(a) \quad (10)$$

Notar que las fórmulas (8)-(10), junto con (1), muestran como calcular la probabilidad de que la variable aleatoria  $X$  tome valores en un intervalo de extremos  $a$  y  $b$  y contienen una advertencia sobre la acumulación de masa positiva en alguno de los dos extremos.

---

### 1.1. Propiedades de la función de distribución

**Lema 1.4.** Sea  $X : \Omega \rightarrow \mathbb{R}$  una variable aleatoria. La función de distribución de  $X$ ,  $F_X(x) = \mathbb{P}(X \leq x)$ , tiene las siguientes propiedades:

(F1) es *no decreciente*: si  $x_1 \leq x_2$ , entonces  $F_X(x_1) \leq F_X(x_2)$ ;

(F2) es *continua a derecha*: para todo  $x_0 \in \mathbb{R}$  vale que  $\lim_{x \downarrow x_0} F_X(x) = F_X(x_0)$ ;

(F3)  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  y  $\lim_{x \rightarrow \infty} F_X(x) = 1$ .

#### **Demostración.**

La propiedad (F1) se deduce de la fórmula (1).

La propiedad (F2) es consecuencia del axioma de continuidad de la medida de probabilidad  $\mathbb{P}$ . Se considera una sucesión decreciente de números positivos que converge a 0,  $\epsilon_1 > \epsilon_2 > \dots > 0$ , arbitraria, pero fija y se definen eventos  $A_n = \{x_0 < X \leq x_0 + \epsilon_n\}$ . Se observa que  $A_1 \supset A_2 \supset \dots$  y  $\bigcap_{n \in \mathbb{N}} A_n = \emptyset$ :

$$\begin{aligned} 0 &= \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \lim_{n \rightarrow \infty} \mathbb{P}(x_0 < X \leq x_0 + \epsilon_n) \\ &= \lim_{n \rightarrow \infty} F(x_0 + \epsilon_n) - F(x_0). \end{aligned}$$

Por lo tanto,

$$F(x_0) = \lim_{n \rightarrow \infty} F(x_0 + \epsilon_n).$$

Las propiedades (F3) se demuestran de manera similar. □

**Observación 1.5.** Si se define

$$F_X(x_0^-) := \lim_{x \uparrow x_0} F_X(x),$$

entonces  $F_X(x_0^-) = \mathbb{P}(X < x_0)$ . Por lo tanto,  $\mathbb{P}(X = x_0) = F_X(x_0) - F_X(x_0^-)$ . En particular, si  $F_X(x)$  es continua en  $x_0$ , entonces  $\mathbb{P}(X = x_0) = 0$ . Si  $\mathbb{P}(X = x_0) > 0$ , entonces  $F_X(x)$  es discontinua en  $x_0$  y su discontinuidad es un salto de altura  $\mathbb{P}(X = x_0) > 0$ .

---

### Ejercicios adicionales

**2.** Sea  $(\Omega, \mathcal{A}, \mathbb{P})$  un espacio de probabilidad y  $X : \Omega \rightarrow \mathbb{R}$  una variable aleatoria con función de distribución  $F_X(x)$ .

(a) Mostrar que

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad \text{y} \quad \lim_{x \rightarrow \infty} F_X(x) = 1.$$

(*Sugerencia.* Considerar sucesiones de eventos  $B_n = \{X \leq -n\}$  y  $C_n = \{X \leq n\}$ ,  $n \in \mathbb{N}$ , y utilizar el axioma de continuidad de la medida de probabilidad  $\mathbb{P}$ .)

(b) Mostrar que

$$\lim_{x \uparrow x_0} F_X(x) = \mathbb{P}(X < x_0).$$

(*Sugerencia.* Observar que si  $x \uparrow x_0$ , entonces  $\{X \leq x\} \uparrow \{X < x_0\}$ , y utilizar el axioma de continuidad de la medida de probabilidad  $\mathbb{P}$ .)

---

## 1.2. Clasificación de variables aleatorias

En todo lo que sigue,  $X$  designa una variable aleatoria definida sobre un espacio de probabilidad  $(\Omega, \mathcal{A}, \mathbb{P})$  y  $F_X(x) := \mathbb{P}(X \leq x)$  su función de distribución.

**Nota Bene.** Al observar el gráfico de una función de distribución lo primero que llama la atención son sus saltos y sus escalones.

**Átomos.** Diremos que  $a \in \mathbb{R}$  es un *átomo* de  $F_X(x)$  si su peso es positivo:  $\mathbb{P}(X = a) = F_X(a) - F_X(a-) > 0$ .

El conjunto de todos los átomos de  $F_X(x)$ :  $\mathbb{A} = \{a \in \mathbb{R} : F_X(a) - F_X(a-) > 0\}$ , coincide con el conjunto de todos los puntos de discontinuidad de  $F_X(x)$ . El peso de cada átomo coincide con la longitud del salto dado por la función de distribución en dicho átomo. En consecuencia, existen a lo sumo un átomo de probabilidad  $> \frac{1}{2}$ , a lo sumo dos átomos de probabilidad  $> \frac{1}{3}$ , etcétera. Por lo tanto, es posible reordenar los átomos en una sucesión  $a_1, a_2, \dots$  tal que  $\mathbb{P}(X = a_1) \geq \mathbb{P}(X = a_2) \geq \dots$ . En otras palabras, *existen a lo sumo numerables átomos*.

La propiedad de  $\sigma$ -aditividad de la medida de probabilidad  $\mathbb{P}$  implica que el peso total del conjunto  $\mathbb{A}$  no puede exceder la unidad:

$$\sum_{a \in \mathbb{A}} \mathbb{P}(X = a) \leq 1.$$

**Definición 1.6** (Variables discretas). Diremos que  $X$  es una variable aleatoria *discreta* si

$$\sum_{a \in \mathbb{A}} \mathbb{P}(X = a) = 1.$$

En tal caso, la función  $p_X : \mathbb{A} \rightarrow \mathbb{R}$  definida por  $p_X(x) = \mathbb{P}(X = x)$  se denomina la *función de probabilidad* de  $X$ .

**Escalones.** Sea  $X$  una variable aleatoria discreta. Si  $a_1 < a_2$  son dos átomos consecutivos, entonces  $F_X(x) = F_X(a_1)$  para todo  $x \in (a_1, a_2)$ . En otras palabras, *la función de distribución de una variable aleatoria discreta debe ser constante entre saltos consecutivos.*

Si no lo fuera, deberían existir dos números  $x_1 < x_2$  contenidos en el intervalo  $(a_1, a_2)$  tales que  $F_X(x_1) < F_X(x_2)$ . En tal caso,

$$\begin{aligned} \mathbb{P}(X \in \mathbb{A} \cup (x_1, x_2]) &= \mathbb{P}(X \in \mathbb{A}) + \mathbb{P}(x_1 < X \leq x_2) \\ &= \sum_{a \in \mathbb{A}} \mathbb{P}(X = a) + F_X(x_2) - F_X(x_1) \\ &= 1 + F_X(x_2) - F_X(x_1) > 1. \end{aligned}$$

lo que constituye un absurdo. □

**Definición 1.7** (Variables continuas). Diremos que  $X$  es una variable aleatoria *continua* si su función de distribución es continua.

**Definición 1.8** (Variables mixtas). Diremos que  $X$  es una variable aleatoria *mixta* si no es continua ni discreta.

**Definición 1.9** (Variables absolutamente continuas). Diremos que  $X$  es *absolutamente continua* si existe una función (medible)  $f_X : \mathbb{R} \rightarrow \mathbb{R}^+$ , llamada *densidad* de  $X$ , tal que cualesquiera sean  $-\infty \leq a < b < \infty$  vale que

$$\mathbb{P}(a < X \leq b) = \int_a^b f_X(x) dx. \tag{11}$$

En particular, para cada  $x \in \mathbb{R}$ , vale que

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(t) dt. \tag{12}$$

**Nota Bene.** Notar que de (12) se deduce que

$$\int_{-\infty}^{\infty} f_X(x) dx = 1.$$

Aplicando en (12) el teorema Fundamental del Cálculo Integral, se obtiene que si  $X$  es *absolutamente continua*,  $F_X(x)$  es una función continua para todo  $x$ , y su derivada es  $f_X(x)$  en todos los  $x$  donde  $f_X$  es continua.

Como la expresión “absolutamente continua” es demasiado larga, se suele hablar simplemente de “distribuciones continuas”. Sin embargo, hay que tener en cuenta que el hecho de que  $F_X$  sea una *función* continua, *no* implica que la distribución de  $X$  sea *absolutamente* continua: hay funciones monótonas y continuas, que sin embargo no son la primitiva de ninguna función. (Para más detalles consultar el ejemplo sobre *distribuciones tipo Cantor* que está en Feller Vol II, p.35-36).  $\square$

**Interpretación intuitiva de la densidad de probabilidad.** Sea  $X$  una variable aleatoria absolutamente continua con función densidad  $f_X(x)$  continua. Para cada  $\epsilon > 0$  pequeño y para  $x \in \mathbb{R}$  vale que

$$\mathbb{P}(x - \epsilon/2 < X \leq x + \epsilon/2) = \int_{x-\epsilon/2}^{x+\epsilon/2} f_X(t) dt \approx f_X(x)\epsilon.$$

Dicho en palabras, la probabilidad de que el valor de  $X$  se encuentre en un intervalo de longitud  $\epsilon$  centrado en  $x$  es aproximadamente  $f_X(x)\epsilon$ .  $\square$

## Ejemplos

**Ejemplo 1.10.** El resultado,  $X$ , del lanzamiento de un dado equilibrado (ver Ejemplo 1.1) es una variable aleatoria discreta. Esto resulta evidente de observar que el gráfico de la función de distribución de  $X$  (ver Figura 1) tiene la forma de una escalera con saltos de altura  $1/6$  en los puntos  $1, 2, 3, 4, 5, 6$ . Dicho en otras palabras, toda la masa de la variable aleatoria  $X$  está concentrada en el conjunto de los átomos de  $F_X$ ,  $\mathbb{A} = \{1, 2, 3, 4, 5, 6\}$ .  $\square$

**Ejemplo 1.11** (Números al azar). El resultado de “sortear” un número al azar sobre el intervalo  $(0, 1)$  es una variable aleatoria absolutamente continua. La probabilidad del evento  $U \leq u$  es igual a la longitud del intervalo  $(-\infty, u] \cap (0, 1)$ .

Notar que cuando  $u \leq 0$  el intervalo  $(-\infty, u] \cap (0, 1)$  se reduce al conjunto vacío que por definición tiene longitud 0. Por otra parte, para cualquier  $u \in (0, 1)$  se tiene que  $(-\infty, u] \cap (0, 1) = (0, u)$  y en consecuencia  $\mathbb{P}(U \leq u) = u$ ; mientras que si  $u \geq 1$ ,  $(-\infty, u] \cap (0, 1) = (0, 1)$  de donde sigue que  $\mathbb{P}(U \leq u) = 1$ . Por lo tanto, la función de distribución de  $U$  es

$$F_U(u) = u\mathbf{1}\{0 \leq u < 1\} + \mathbf{1}\{u \geq 1\}.$$

Derivando, respecto de  $u$ , la función de distribución  $F_U(u)$  se obtiene una función densidad para  $U$ :

$$f_U(u) = \mathbf{1}\{0 < u < 1\}.$$

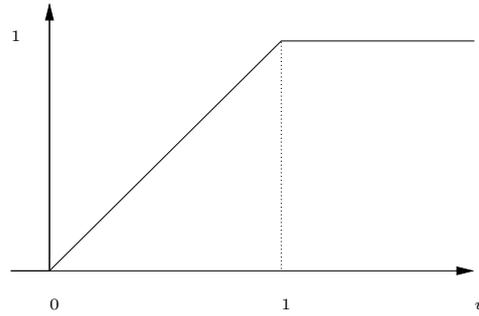


Figura 3: Gráfico de la función de distribución del resultado de “sortear” un número al azar.

□

**Nota Bene.** Sortear un número al azar sobre el intervalo  $(0, 1)$  es un caso particular de una familia de variables aleatorias denominadas *uniformes*. Una variable aleatoria  $X$ , definida sobre un espacio de probabilidad  $(\Omega, \mathcal{A}, \mathbb{P})$ , se denomina *uniformemente distribuida sobre el intervalo*  $(a, b)$ , donde  $a < b$ , si  $X$  es absolutamente continua y admite una función densidad de la forma

$$f_X(x) = \frac{1}{b-a} \mathbf{1}\{x \in (a, b)\}.$$

En tal caso escribiremos  $X \sim \mathcal{U}(a, b)$ . Es fácil ver que

$$\mathbb{P}(X \in (c, d]) = \frac{d-c}{b-a},$$

para todo intervalo  $(c, d] \subset (a, b)$  y que

$$F_X(x) = \frac{x-a}{b-a} \mathbf{1}\{x \in (a, b)\} + \mathbf{1}\{x \geq b\}.$$

□

**Comentario.** En la Sección 1.4 mostraremos que todas las variables aleatorias se pueden construir utilizando variables aleatorias uniformemente distribuidas sobre el intervalo  $(0, 1)$ .

**Ejemplo 1.12.** El tiempo,  $T$ , de funcionamiento hasta la aparición de la primera falla para un sistema con función intensidad de fallas continua  $\lambda(t)$  (ver Ejemplo 1.2) es una variable aleatoria absolutamente continua que admite una densidad de la forma

$$f_T(t) = \lambda(t) \exp\left(-\int_0^t \lambda(s) ds\right) \mathbf{1}\{t > 0\}. \quad (13)$$

**Nota Bene: algunos casos particulares del Ejemplo 1.12.** El comportamiento de la densidad (13) depende de la forma particular de la función intensidad de fallas  $\lambda(t)$ . En lo que sigue mostraremos algunos casos particulares.

- *Exponencial de intensidad  $\lambda$ .* Se obtiene poniendo  $\lambda(t) = \lambda \mathbf{1}\{t \geq 0\}$ , donde  $\lambda$  es una constante positiva, arbitraria pero fija.

$$f_T(t) = \lambda \exp(-\lambda t) \mathbf{1}\{t > 0\}. \quad (14)$$

- *Weibull de parámetros  $c$  y  $\alpha$ .* Se obtiene poniendo  $\lambda(t) = \frac{c}{\alpha} \left(\frac{t}{\alpha}\right)^{c-1} \mathbf{1}\{t \geq 0\}$ , donde  $c > 0$  y  $\alpha > 0$ . En este caso, la densidad (13) adopta la forma

$$f_T(t) = \frac{c}{\alpha} \left(\frac{t}{\alpha}\right)^{c-1} \exp\left(-\left(\frac{t}{\alpha}\right)^c\right). \quad (15)$$

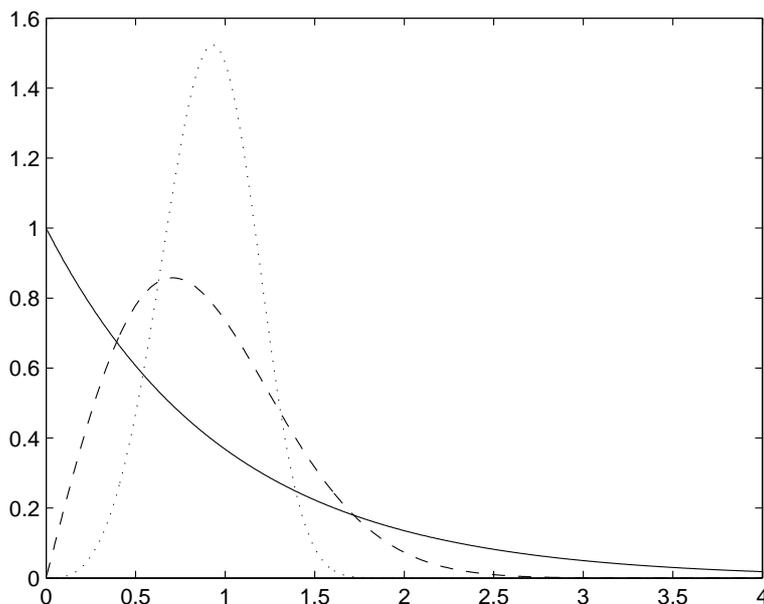


Figura 4: Gráficos de las densidades Weibull de parámetro de escala  $\alpha = 1$  y parámetro de forma:  $c = 1, 2, 4$ : en línea sólida  $c = 1$ ; en línea quebrada  $c = 2$  y en línea punteada  $c = 4$ .

Notar que la exponencial de intensidad  $\lambda$  es un caso especial de la Weibull puesto que (14) se obtiene de (15) poniendo  $c = 1$  y  $\alpha = \lambda^{-1}$ .  $\square$

**Ejemplo 1.13.** La variable aleatoria,  $S$ , considerada en el Ejemplo 1.3 es una variable aleatoria mixta (ver Figura 2) porque no es discreta ni continua. Tiene un único átomo en  $s = 0$  y su peso es  $\exp\left(-\int_0^{t_0} \lambda(x)dx\right)$ .  $\square$

### 1.3. Cuantiles

**Definición 1.14.** Sea  $\alpha \in (0, 1)$ . Un cuantil- $\alpha$  de  $X$  es cualquier número  $x_\alpha \in \mathbb{R}$  tal que

$$\mathbb{P}(X < x_\alpha) \leq \alpha \quad \text{y} \quad \alpha \leq \mathbb{P}(X \leq x_\alpha). \quad (16)$$

**Observación 1.15.** Notar que las desigualdades que caracterizan a los cuantiles- $\alpha$  se pueden reescribir de la siguiente manera

$$F_X(x_\alpha) - \mathbb{P}(X = x_\alpha) \leq \alpha \quad \text{y} \quad \alpha \leq F_X(x_\alpha). \quad (17)$$

Por lo tanto, si  $F_X(x)$  es continua,  $x_\alpha$  es un cuantil  $\alpha$  si y sólo si

$$F_X(x_\alpha) = \alpha. \quad (18)$$

**Interpretación “geométrica” del cuantil- $\alpha$ .** Si  $X$  es una variable aleatoria absolutamente continua con función de densidad  $f_X(x)$  el cuantil- $\alpha$  de  $X$  es la única solución de la ecuación

$$\int_{-\infty}^{x_\alpha} f_X(x) dx = \alpha.$$

Esto significa que el cuantil- $\alpha$  de  $X$  es el único punto sobre el eje de las abscisas a cuya izquierda el área bajo la función de densidad  $f_X(x)$  es igual a  $\alpha$ .

**Nota Bene.** Sea  $x \in \mathbb{R}$ . Las desigualdades (17) significan que  $x$  es un cuantil- $\alpha$  si y sólo si  $\alpha \in [F(x) - \mathbb{P}(X = x), F(x)]$

**Nota Bene.** El cuantil- $\alpha$  siempre existe. Sea  $\alpha \in (0, 1)$ , la existencia del cuantil  $\alpha$  se deduce analizando el conjunto  $R_X^\alpha = \{x \in \mathbb{R} : \alpha \leq F_X(x)\}$ .

1.  $R_X^\alpha$  es no vacío porque  $\lim_{x \rightarrow \infty} F_X(x) = 1$ .
2.  $R_X^\alpha$  es acotado inferiormente porque  $\lim_{x \rightarrow -\infty} F_X(x) = 0$ .
3. Si  $x_0 \in R_X^\alpha$ , entonces  $[x_0, +\infty) \subset R_X^\alpha$  porque  $F_X(x)$  es no decreciente.
4.  $\inf R_X^\alpha \in R_X^\alpha$  porque existe una sucesión  $\{x_n : n \in \mathbb{N}\} \subset R_X^\alpha$  tal que  $x_n \downarrow \inf R_X^\alpha$  y  $F_X(x)$  es una función continua a derecha:

$$\alpha \leq \lim_{n \rightarrow \infty} F_X(x_n) = F_X\left(\lim_{n \rightarrow \infty} x_n\right) = F_X(\inf R_X^\alpha).$$

De las propiedades anteriores se deduce que

$$R_X^\alpha = [\inf R_X^\alpha, +\infty) = [\min R_X^\alpha, +\infty).$$

Hay dos casos posibles: (a)  $F_X(\min R_X^\alpha) = \alpha$  o (b)  $F_X(\min R_X^\alpha) > \alpha$ .

(a) Si  $F_X(\text{mín } R_X^\alpha) = \alpha$ , entonces  $\mathbb{P}(X < \text{mín } R_X^\alpha) = \alpha - \mathbb{P}(X = \text{mín } R_X^\alpha) \leq \alpha$ .

(b) Si  $F_X(\text{mín } R_X^\alpha) > \alpha$ , entonces

$$\mathbb{P}(X < x) < \alpha \quad \forall x < \text{mín } R_X^\alpha \quad (19)$$

porque sino existe un  $x < \text{mín } R_X^\alpha$  tal que  $\alpha \leq \mathbb{P}(X < x) \leq F_X(x)$  y por lo tanto,  $x \in R_X^\alpha$  lo que constituye un absurdo.

De (19) se deduce que  $\mathbb{P}(X < \text{mín } R_X^\alpha) = \lim_{x \uparrow \text{mín } R_X^\alpha} F_X(x) \leq \alpha$ .

En consecuencia, en cualquiera de los dos casos

$$x_\alpha = \text{mín } \{x \in \mathbb{R} : F_X(x) \geq \alpha\} \quad (20)$$

es un cuantil- $\alpha$ . □

**Nota Bene.** Si  $F_X$  es discontinua, (18) no tiene siempre solución; y por eso es mejor tomar (16) como definición. Si  $F_X$  es estrictamente creciente, los cuantiles son únicos. Pero si no, los valores que satisfacen (18) forman un intervalo.

**Cuartiles y mediana.** Los cuantiles correspondientes a  $\alpha = 0.25, 0.50$  y  $0.75$  son respectivamente el primer, el segundo y tercer *cuartiles*. El segundo cuartil es la *mediana*.

## Ejemplos

**Ejemplo 1.16.** En el Ejemplo 1.1 hemos visto que la función de distribución del resultado del lanzamiento de un dado equilibrado es una escalera con saltos de altura  $1/6$  en los puntos  $1, 2, 3, 4, 5, 6$ :

$$F_X(x) = \sum_{i=1}^5 \frac{i}{6} \mathbf{1}\{i \leq x < i+1\} + \mathbf{1}\{6 \leq x\}.$$

Notar que la imagen de  $F_X$  es

$$\{\alpha \in [0, 1] : F_X(x) = \alpha \text{ para algún } x \in \mathbb{R}\} = \{0, 1/6, 2/6, 3/6, 4/6, 5/6, 1\}.$$

Esto significa que la ecuación (18) solo tiene solución para  $\alpha \in \{1/6, 2/6, 3/6, 4/6, 5/6\}$ . Más aún, para cada  $i = 1, \dots, 5$

$$F_X(x) = \frac{i}{6} \iff x \in [i, i+1).$$

En otras palabras, para cada  $i = 1, \dots, 5$  los cuantiles- $i/6$  de  $X$  son el intervalo  $[i, i+1)$ . En particular, “la” mediana de  $X$  es cualquier punto del intervalo  $[3, 4)$ .

Para cada  $\alpha \in (\frac{i-1}{6}, \frac{i}{6})$ ,  $i = 1, \dots, 6$ , el cuantil  $\alpha$  de  $X$  es  $x_\alpha = i$ . □

**Ejemplo 1.17.** Sea  $T$  el tiempo de funcionamiento hasta la aparición de la primera falla para un sistema con función intensidad de fallas  $\lambda(t) = 2t\mathbf{1}\{t \geq 0\}$  (ver Ejemplo 1.2). La función de distribución de  $T$  es

$$F_T(t) = \left(1 - \exp\left(-\int_0^t 2s ds\right)\right) \mathbf{1}\{t > 0\} = (1 - \exp(-t^2)) \mathbf{1}\{t > 0\}. \quad (21)$$

. Debido a que  $F_T(t)$  es continua los cuantiles- $\alpha$ ,  $\alpha \in (0, 1)$  se obtienen resolviendo la ecuación (18):

$$\begin{aligned} F_T(t) = \alpha &\iff 1 - \exp(-t^2) = \alpha \iff 1 - \alpha = \exp(-t^2) \iff \log(1 - \alpha) = -t^2 \\ &\iff t = \sqrt{-\log(1 - \alpha)}. \end{aligned}$$

Por lo tanto, para cada  $\alpha \in (0, 1)$  el cuantil- $\alpha$  de  $T$  es

$$t_\alpha = \sqrt{-\log(1 - \alpha)}. \quad (22)$$

En particular, la mediana de  $T$  es  $t_{0.5} = \sqrt{-\log(1 - 0.5)} \approx 0.8325$ .  $\square$

**Ejemplo 1.18.** Se considera un sistema con función intensidad de fallas  $\lambda(t) = 2t\mathbf{1}\{t \geq 0\}$ . El sistema debe prestar servicios durante 1 hora. Si durante ese período el sistema falla, se lo repara y se lo vuelve a utilizar hasta que cumpla con el el plazo estipulado. Sea  $S$  el tiempo de funcionamiento (medido en horas) del sistema después de la primera reparación.

En el Ejemplo 1.3 hemos visto que la función de distribución de  $S$  es

$$\begin{aligned} F_S(s) &= \exp\left(-\int_0^{1-s} 2t dt\right) \mathbf{1}\{0 \leq s < 1\} + \mathbf{1}\{s \geq 1\} \\ &= \exp(-(1-s)^2) \mathbf{1}\{0 \leq s < 1\} + \mathbf{1}\{s \geq 1\}, \end{aligned}$$

y que  $S$  es una variable aleatoria mixta (ver Figura 2) con un único átomo en  $s = 0$  cuyo peso es  $e^{-1}$ . En consecuencia,  $s = 0$  es un cuantil- $\alpha$  de  $S$  para todo  $\alpha \in (0, e^{-1}]$ . Restringida al intervalo  $(0, 1)$  la función  $F_S(s)$  es continua y su imagen es el intervalo  $(e^{-1}, 1)$ . Por ende, para cada  $\alpha \in (e^{-1}, 1)$  el cuantil- $\alpha$  de  $S$  se obtiene resolviendo la ecuación  $F_S(s) = \alpha$ :

$$\begin{aligned} F_S(s) = \alpha &\iff \exp(-(1-s)^2) = \alpha \iff -(1-s)^2 = \log(\alpha) \\ &\iff (1-s)^2 = -\log(\alpha) \iff |1-s| = \sqrt{-\log(\alpha)} \\ &\iff 1-s = \sqrt{-\log(\alpha)} \iff 1 - \sqrt{-\log(\alpha)} = s. \end{aligned}$$

Por lo tanto, para cada  $\alpha \in (e^{-1}, 1)$  el cuantil- $\alpha$  de  $S$  es

$$s_\alpha = 1 - \sqrt{-\log(\alpha)}.$$

En particular, la mediana de  $S$  es  $s_{0.5} = 1 - \sqrt{-\log(0.5)} \approx 0.1674$ .  $\square$

## 1.4. Construcción de variables aleatorias

**Teorema 1.19** (Simulación). Sea  $F : \mathbb{R} \rightarrow [0, 1]$  una función con las siguientes propiedades

(F1) es *no decreciente*: si  $x_1 \leq x_2$ , entonces  $F(x_1) \leq F(x_2)$ ;

(F2) es *continua a derecha*: para todo  $x_0 \in \mathbb{R}$  vale que  $\lim_{x \downarrow x_0} F(x) = F(x_0)$ ;

(F3)  $\lim_{x \rightarrow -\infty} F(x) = 0$  y  $\lim_{x \rightarrow \infty} F(x) = 1$ .

Existe una variable aleatoria  $X$  tal que  $F(x) = \mathbb{P}(X \leq x)$ .

### Esquema de la demostración.

1º) Definir la *inversa generalizada* de  $F$  mediante

$$F^{-1}(u) := \min\{x \in \mathbb{R} : u \leq F(x)\}, \quad u \in (0, 1).$$

2º) Definir  $X$  mediante

$$X := F^{-1}(U), \quad \text{donde } U \sim \mathcal{U}(0, 1).$$

3º) Observar que vale la equivalencia (inmediata)  $F^{-1}(u) \leq x \Leftrightarrow u \leq F(x)$  y deducir que  $\mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$ .  $\square$

**Observación 1.20.** Si la función  $F$  del enunciado del Teorema 1.19 es continua, la inversa generalizada es simplemente la inversa.  $\square$

**Nota Bene.** El esquema de la demostración del Teorema 1.19 muestra *cómo se construye una variable aleatoria  $X$  con función de distribución  $F_X(x)$* . La construcción es clave para simular variables aleatorias en una computadora: algoritmos estándar generan variables aleatorias  $U$  con distribución uniforme sobre el intervalo  $(0, 1)$ , aplicando la inversa generalizada de la función de distribución se obtiene la variable aleatoria  $F_X^{-1}(U)$  cuya función de distribución es  $F_X(x)$ .  $\square$

**Método gráfico para calcular inversas generalizadas.** Sea  $u \in (0, 1)$ , por definición,  $F^{-1}(u) := \min\{x \in \mathbb{R} : u \leq F(x)\}$ ,  $0 < u < 1$ . Gráficamente esto significa que para calcular  $F^{-1}(u)$  hay que determinar el conjunto de todos los puntos del gráfico de  $F(x)$  que están sobre o por encima de la recta horizontal de altura  $u$  y proyectarlo sobre el eje de las abscisas. El resultado de la proyección es una semirecta sobre el eje de las abscisas y el valor de la abscisa que la cierra por izquierda es el valor de  $F^{-1}(u)$ .  $\square$

**Ejemplo 1.21** (Moneda cargada). *Se quiere simular el lanzamiento de una moneda “cargada” con probabilidad  $p \in (0, 1)$  de salir cara.* El problema se resuelve construyendo una variable aleatoria  $X$  a valores  $\{0, 1\}$  tal que  $\mathbb{P}(X = 1) = p$  y  $\mathbb{P}(X = 0) = 1 - p$ , ( $X = 1$  representa el evento “la moneda sale cara” y  $X = 0$  “la moneda sale ceca”). La función de distribución de  $X$  debe ser  $F(x) = (1 - p)\mathbf{1}\{0 \leq x < 1\} + \mathbf{1}\{x \geq 1\}$  y su gráfico se muestra en la Figura 5.

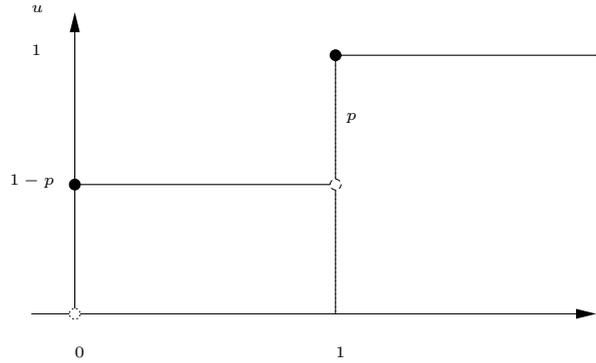


Figura 5: Gráfico de la función  $F(x) = (1 - p)\mathbf{1}\{0 \leq x < 1\} + \mathbf{1}\{x \geq 1\}$ .

La demostración del Teorema 1.19 indica que para construir la variable aleatoria  $X$  lo primero que hay que hacer es determinar la expresión de la inversa generalizada de  $F(x)$ . Para ello usaremos el método gráfico.

En la Figura 5 se puede ver que para cada  $0 < u \leq 1 - p$  el conjunto  $\{x \in \mathbb{R} : u \leq F(x)\}$  es la semirecta  $[0, \infty)$  y el punto que la cierra por izquierda es  $x = 0$ . En consecuencia  $F^{-1}(u) = 0$  para todo  $0 < u \leq 1 - p$ . Del mismo modo se puede ver que  $F^{-1}(u) = 1$  para todo  $1 - p < u < 1$ . Por lo tanto,  $F^{-1}(u) = \mathbf{1}\{1 - p < u < 1\}$ .

Definiendo  $X := \mathbf{1}\{1 - p < U < 1\}$ , donde  $U \sim \mathcal{U}(0, 1)$  se obtiene la variable aleatoria deseada.

**Ejemplo 1.22** (Moneda cargada). *Simular diez lanzamientos de una moneda “cargada” con probabilidad 0.6 de salir cara en cada lanzamiento.*

De acuerdo con el resultado obtenido en el Ejemplo 1.21, para simular el lanzamiento de una moneda cargada con probabilidad 0.6 de salir cara se construye la variable aleatoria  $X := \mathbf{1}\{0.4 < U < 1\}$ , donde  $U \sim \mathcal{U}(0, 1)$ .

Para simular 10 valores de  $X$  se simulan 10 valores de  $U$ . Si en 10 simulaciones de  $U$  se obtuviesen los valores 0.578, 0.295, 0.885, 0.726, 0.548, 0.048, 0.474, 0.722, 0.786, 0.598, los valores de la variable  $X$  serían 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, respectivamente, y en tal caso, los resultados de los 10 lanzamientos de la moneda serían  $H, T, H, H, H, T, H, H, H, H$ .  $\square$

**Ejemplo 1.23** (Fiabilidad). *Se considera un sistema electrónico con función intensidad de fallas de la forma  $\lambda(t) = 2t\mathbf{1}\{t > 0\}$ . Se quiere estimar la función de probabilidad de la cantidad de fallas ocurridas durante la primer unidad de tiempo de funcionamiento.*

Para simplificar el problema vamos a suponer que cada vez que se produce una falla, el sistema se repara instantáneamente renovándose sus condiciones iniciales de funcionamiento. Según el Ejemplo 1.2, la función de distribución del tiempo de funcionamiento hasta la aparición de la primer falla es

$$F(t) = (1 - \exp(-t^2)) \mathbf{1}\{t > 0\}. \quad (23)$$

Debido a que la función de distribución  $F(t)$  es continua, su inversa generalizada es simplemente su inversa y se obtiene despejando  $t$  de la ecuación  $1 - \exp(-t^2) = u$ . En consecuencia,  $F^{-1}(u) = \sqrt{-\log(1-u)}$ ,  $u \in (0, 1)$ . Para construir la variable  $T$  usamos un número aleatorio  $U$ , uniformemente distribuido sobre el intervalo  $(0, 1)$  y definimos

$$T := F^{-1}(U) = \sqrt{-\log(1-U)}. \quad (24)$$

La ventaja de la construcción es que puede implementarse casi de inmediato en una computadora. Por ejemplo, una rutina en Octave para simular  $T$  es la siguiente

---

```
U=rand;
T=sqrt(-log(1-rand))
```

---

Sobre la base de esa rutina podemos simular valores de  $T$ . Por ejemplo, en diez simulaciones de  $T$  obtuvimos los valores siguientes: 0.3577, 1.7233, 1.1623, 0.3988, 1.4417, 0.3052, 1.1532, 0.3875, 0.8493, 0.9888.

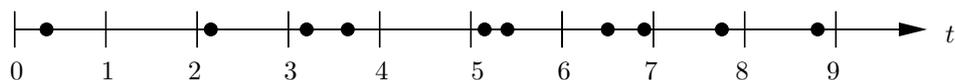


Figura 6: Simulación de los tiempos de ocurrencia de las fallas de un sistema electrónico con función intensidad de fallas de la forma  $\lambda(t) = 2t\mathbf{1}\{t \geq 0\}$ . Las fallas ocurren los instantes 0.3577, 2.0811, 3.2434, 3.6422, 5.0839, 5.3892, 6.5423, 6.9298, 7.7791, 8.7679.

La rutina puede utilizarse para simular cien mil realizaciones del experimento que consiste en observar la cantidad de fallas durante la primer unidad de tiempo de funcionamiento del sistema electrónico bajo consideración:  $N[0, 1] := \min\{n \geq 1 : \sum_{i=1}^n T_i > 1\} - 1$ , donde  $T_1, T_2, \dots$  son realizaciones independientes de los tiempos de funcionamiento del sistema hasta la ocurrencia de una falla.

Por ejemplo, repitiendo la simulación 100000 veces obtuvimos la siguiente tabla que contiene la cantidad de veces que fué simulado cada valor de la variable  $N[0, 1]$ :

valor simulado	0	1	2	3	4	(25)
frecuencia	36995	51792	10438	743	32	

obteniéndose las siguientes estimaciones

$$\begin{aligned} \mathbb{P}(N[0, 1] = 0) &\approx 0.36995, & \mathbb{P}(N[0, 1] = 1) &\approx 0.51792, & \mathbb{P}(N[0, 1] = 2) &\approx 0.10438, \\ \mathbb{P}(N[0, 1] = 3) &\approx 0.00743, & \mathbb{P}(N[0, 1] = 4) &\approx 0.00032. \end{aligned}$$

Para finalizar este ejemplo, presentamos una rutina en Octave que simula cien mil veces la cantidad de fallas en la primer unidad de tiempo y que al final produce los resultados para construir una tabla similar a la tabla (25).

---

```

for i=1:100000
    n=-1;
    S=0;
    while S<=1;
        T=sqrt(-log(1-rand));
        S=S+T;
        n=n+1;
    end
    f(i)=n;
end
M=max(f);
for i=1:M+1;
    N(i)=length(find(f==i-1));
end
N

```

**Ejemplo 1.24** (*Saltando, saltando, sa, sa, sa, saltando,...*  $\hat{\mathbb{Q}}$ ). La función

$$F(x) = \sum_{n=1}^{\infty} \frac{1}{2^n} \mathbf{1}\{x \geq r_n\}, \quad (26)$$

donde  $r_1, r_2, \dots$  es un reordenamiento de los números racionales del intervalo  $(0, 1)$  con denominadores crecientes:  $\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \dots$ , tiene las siguientes propiedades es creciente, continua a derecha,  $\lim_{x \rightarrow -\infty} F(x) = 0$  y  $\lim_{x \rightarrow \infty} F(x) = 1$ ; tiene saltos en todos los números racionales del  $(0, 1)$  y es continua en los irracionales del  $(0, 1)$ .

*Pero no! Mejor no hablar de ciertas cosas ...* □

### Ejercicios adicionales

**3.** Sea  $X$  una variable aleatoria con función de distribución  $F_X(x)$ . Mostrar que para cada  $\alpha \in (0, 1)$  vale que

$$\sup\{x \in \mathbb{R} : F_X(x) < \alpha\} = \min\{x \in \mathbb{R} : F_X(x) \geq \alpha\}.$$

## 1.5. Función de distribución empírica e histogramas

### Distribución empírica

La *función de distribución empírica*  $F_n(x)$  de  $n$  puntos sobre la recta  $x_1, \dots, x_n$  es la función escalera con saltos de altura  $1/n$  en los puntos  $x_1, \dots, x_n$ . En otras palabras,  $nF_n(x)$

es igual a la cantidad de puntos  $x_k$  en  $(-\infty, x]$  y  $F_n(x)$  es una función de distribución:

$$F_n(x) = \frac{1}{n} |\{i = 1, \dots, n : x_i \leq x\}| = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x_i \leq x\}. \quad (27)$$

**Nota Bene.** En la práctica, disponemos de conjuntos de observaciones (“muestras”) correspondientes a un experimento considerado aleatorio y queremos extraer de ellas conclusiones sobre los modelos que podrían cumplir. Dada una muestra  $x_1, \dots, x_n$ , la función de distribución empírica  $F_n(x)$  coincide con la función de distribución de una variable aleatoria discreta que concentra toda la masa en los valores  $x_1, \dots, x_n$ , dando a cada uno probabilidad  $1/n$ .

**Observación 1.25.** Sea  $F_n(x)$  la función de distribución empírica correspondiente a una muestra de  $n$  valores  $x_1, \dots, x_n$ . Sean  $a$  y  $b$  dos números reales tales que  $a < b$ . Notar que

$$F_n(b) - F_n(a) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x_i \in (a, b]\} = \frac{1}{n} |\{i = 1, \dots, n : x_i \in (a, b]\}|.$$

En consecuencia, el cociente incremental de  $F_n(x)$  sobre el intervalo  $[a, b]$  es la frecuencia relativa de los valores de la muestra  $x_1, \dots, x_n$  contenidos en el intervalo  $(a, b]$  “normalizada” por la longitud de dicho intervalo:

$$\frac{F_n(b) - F_n(a)}{b - a} = \left( \frac{1}{b - a} \right) \left( \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x_i \in (a, b]\} \right). \quad (28)$$

Notar que si los  $n$  valores,  $x_1, \dots, x_n$ , corresponden a  $n$  observaciones independientes de los valores de una variable aleatoria  $X$ , la interpretación intuitiva de la probabilidad indica que el cociente incremental (28) debería estar próximo del cociente incremental de la función de distribución,  $F_X(x)$ , de la variable aleatoria  $X$  sobre el intervalo  $[a, b]$ :

$$\frac{F_n(b) - F_n(a)}{b - a} \approx \frac{\mathbb{P}(a < X \leq b)}{b - a} = \frac{F_X(b) - F_X(a)}{b - a}. \quad (29)$$

Cuando  $X$  es una variable aleatoria absolutamente continua con función densidad continua  $f_X(x)$  la aproximación (28) adopta la forma

$$\frac{F_n(b) - F_n(a)}{b - a} \approx \frac{1}{b - a} \int_a^b f_X(x) dx = f_X(x), \quad (30)$$

donde  $x$  es algún punto perteneciente al intervalo  $(a, b)$ . □

## Histogramas

Un *histograma* de una muestra  $x_1, \dots, x_n$  se obtiene eligiendo una partición en  $m$  intervalos de extremos  $a_0 < \dots < a_m$ , con longitudes  $L_j = a_j - a_{j-1}$ ; calculando las *frecuencias relativas*

$$p_j = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{a_{j-1} < x_i < a_j\}$$

y graficando la función igual a  $p_j/L_j$  en el intervalo  $(a_{j-1}, a_j]$  y a 0 fuera de los intervalos:

$$f_{x_1, \dots, x_n; a_0, \dots, a_m}(x) := \sum_{j=1}^m \frac{p_j}{L_j} \mathbf{1}\{x \in (a_{j-1}, a_j]\}. \quad (31)$$

O sea, un conjunto de rectángulos con área  $p_j$ .

Cuando la muestra  $x_1, \dots, x_n$  corresponde a  $n$  observaciones independientes de una variable aleatoria  $X$  absolutamente continua la función definida en (31) es una versión discreta de la densidad de  $X$  en la que las áreas miden frecuencias relativas.

**Ejemplo 1.26.** Sea  $T$  una variable aleatoria con distribución exponencial de intensidad 1 (ver (14)). Esto es,  $T$  es una variable aleatoria absolutamente continua con función densidad de probabilidad

$$f_T(t) = e^{-t} \mathbf{1}\{t > 0\}$$

y función de distribución

$$F_T(t) = (1 - e^{-t}) \mathbf{1}\{t \geq 0\}.$$

De acuerdo con el esquema de la demostración del Teorema 1.19 podemos simular muestras de  $T$  utilizando un generador de números aleatorios uniformemente distribuidos sobre el intervalo  $(0, 1)$ . Concretamente, si  $U \sim \mathcal{U}(0, 1)$ , entonces

$$\hat{T} = -\log(1 - U)$$

es una variable con distribución exponencial de intensidad 1.

Para obtener una muestra de 10 valores  $t_1, \dots, t_{10}$  de una variable con distribución exponencial de intensidad 1 generamos 10 números aleatorios  $u_1, \dots, u_{10}$  y los transformamos poniendo  $t_i = -\log(1 - u_i)$ . Por ejemplo, si los valores  $u_1, \dots, u_{10}$  son, respectivamente,

$$0.1406, 0.3159, 0.8613, 0.4334, 0.0595, 0.8859, 0.2560, 0.2876, 0.2239, 0.5912,$$

los valores de la muestra obtenida,  $t_1, \dots, t_{10}$ , son, respectivamente,

$$0.1515, 0.3797, 1.9753, 0.5682, 0.0613, 2.1703, 0.2957, 0.3390, 0.2535, 0.8946. \quad (32)$$

La función de distribución empírica de la muestra observada,  $F_{10}(t)$ , es una función escalera con saltos de altura  $1/10$  en los siguientes puntos del eje  $t$ :

$$0.0613, 0.1515, 0.2535, 0.2957, 0.3390, 0.3797, 0.5682, 0.8946, 1.9753, 2.1703.$$

Para construir un histograma usaremos la partición que se obtiene dividiendo en dos intervalos de igual longitud el intervalo comprendido entre los valores mínimos y máximos observados: 0.0613, 1.1158, 2.1703. La longitud  $L$  de cada intervalo es 1.0545. La frecuencia relativa de la muestra sobre el primer intervalo es  $p_1 = 8/10$  y sobre el segundo  $p_2 = 2/10$  y la correspondiente altura de cada rectángulo es  $p_1/L = 0.75865$  y  $p_2/L = 0.18966$ .

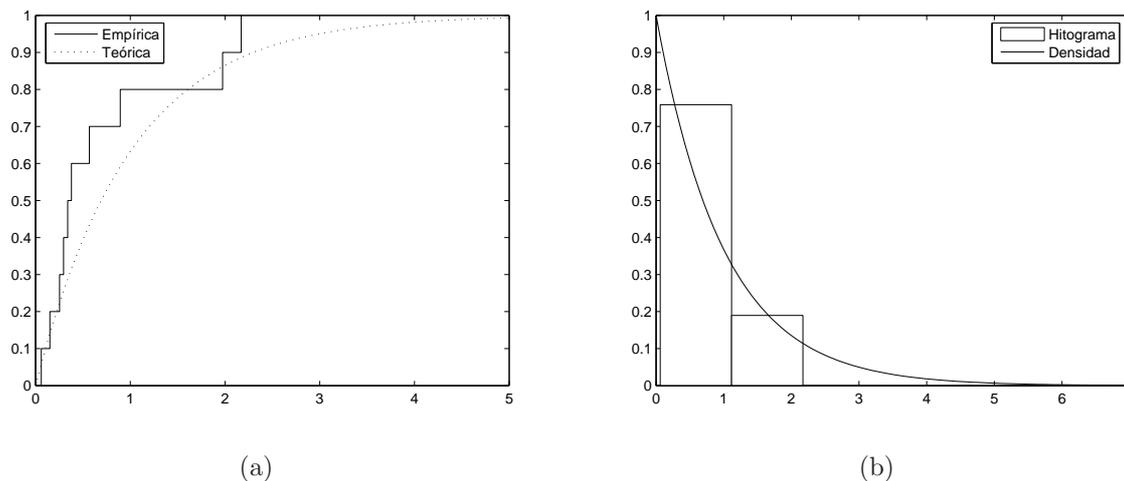


Figura 7: (a) Gráficos de la función de distribución empírica  $F_{10}(t)$  correspondiente a la muestra dada en (32) y de la función de distribución de  $T$ . (b) Histograma correspondiente a la misma muestra y gráfico de la densidad de  $T$ .

Para producir los gráficos de la Figura 7 usamos las siguientes rutinas en Octave.

---

#### Rutina para simular 10 valores de una exponencial de intensidad 1

---

```
U=rand(1,10);
T=-log(1-U);
```

---



---

#### Rutina para graficar la función de distribución empírica de la muestra $T$

---

```
t=sort(T);
s=empirical_cdf(t,t);
stairs([t(1),t],[0 s])
```

---



---

#### Rutina para graficar un histograma de la muestra $T$

---

```
[f,c]=hist(T,2);
p=f/10;
L=c(2)-c(1);
bar(c,p/L,1,'w')
```

---

Usando rutinas similares para muestras de tamaño 100 se obtienen los siguientes gráficos.

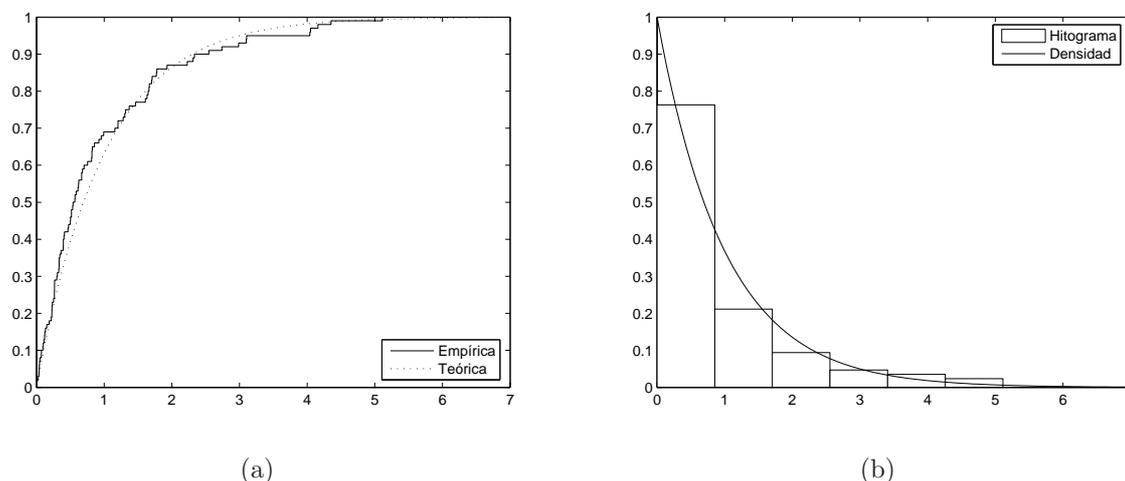


Figura 8: (a) Gráficos de la función de distribución empírica  $F_{100}(t)$  correspondiente a una muestra de tamaño 100 de una variable  $T$  con distribución exponencial de intensidad 1 y de la función de distribución de  $T$ . (b) Histograma correspondiente a la misma muestra y gráfico de la densidad de  $T$ .

□

## 2. Vectores aleatorios

**Notación.** Para simplificar la escritura, en todo lo que sigue, usaremos las siguientes notaciones. Los puntos del espacio  $n$ -dimensional ( $n \geq 2$ )  $\mathbb{R}^n$  se denotan en negrita,  $\mathbf{x} = (x_1, \dots, x_n)$ . La desigualdad  $\mathbf{y} \leq \mathbf{x}$  significa que  $y_i \leq x_i$  para todo  $i = 1, \dots, n$  y puede interpretarse diciendo que  $\mathbf{y}$  está al “sudoeste” de  $\mathbf{x}$ . El conjunto de todos los puntos al “sudoeste” de  $\mathbf{x}$  será denotado mediante  $S_{\mathbf{x}} := \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y} \leq \mathbf{x}\}$ . Finalmente, cualquiera sea el subconjunto de índices  $\Lambda = \{i_1, \dots, i_m\} \subset \{1, 2, \dots, n\}$  denotaremos mediante  $\mathbf{x}_{\Lambda} \in \mathbb{R}^m$  al punto  $m$ -dimensional que se obtiene de  $\mathbf{x}$  quitándole todas las coordenadas que tengan índices fuera de  $\Lambda$ . Por ejemplo, si  $\Lambda = \{1, 2\}$ , entonces  $\mathbf{x}_{\Lambda} = (x_1, x_2)$ .

**Definición 2.1.** Un *vector aleatorio* sobre un espacio de probabilidad  $(\Omega, \mathcal{A}, \mathbb{P})$  es una función  $\mathbf{X} = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$  tal que para cada  $\mathbf{x} \in \mathbb{R}^n$

$$\{\mathbf{X} \in S_{\mathbf{x}}\} = \{\omega \in \Omega : \mathbf{X}(\omega) \leq \mathbf{x}\} \in \mathcal{A}.$$

## 2.1. Distribución conjunta

La función de distribución (*conjunta*)  $F_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, 1]$  del vector aleatorio  $\mathbf{X}$  se define por

$$F_{\mathbf{X}}(\mathbf{x}) := \mathbb{P}(\mathbf{X} \in S_{\mathbf{x}}) = \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \leq x_i\}\right). \quad (33)$$

**Cálculo de probabilidades.** La función de distribución conjunta resume toda la información relevante sobre el comportamiento de las variables aleatorias  $X_1, \dots, X_n$ . Para fijar ideas, consideremos el caso multivariado más simple,  $n = 2$ . Si  $a_1 < b_1$  y  $a_2 < b_2$  vale que, ver la Figura 9,

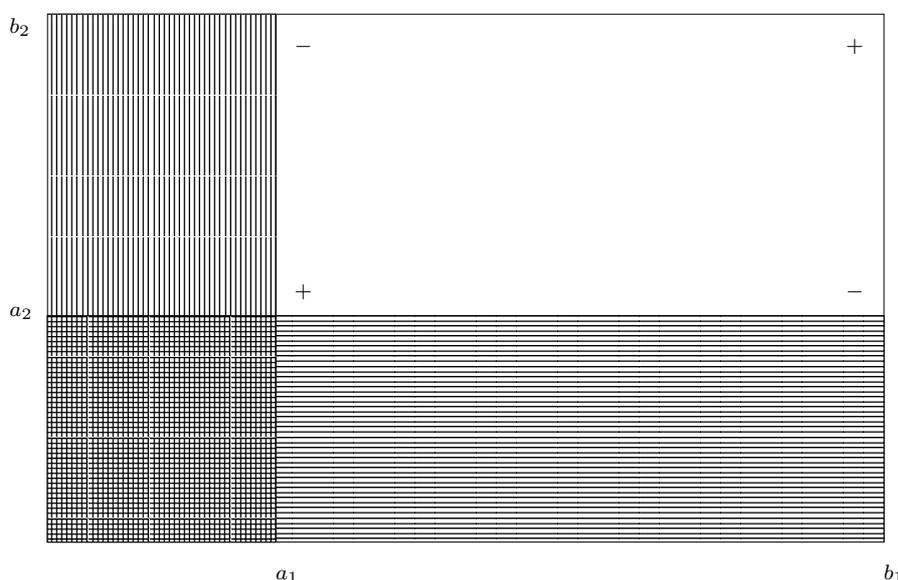


Figura 9: Esquema de la demostración de la identidad (34). El rectángulo  $(a_1, b_1] \times (a_2, b_2]$  se puede representar en la forma  $S_{(b_1, b_2)} \setminus (S_{(a_1, b_2)} \cup S_{(b_1, a_2)})$ .

$$\mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2). \quad (34)$$

La identidad (34) permite calcular la probabilidad de observar al vector  $(X_1, X_2)$  en el rectángulo  $(a_1, b_1] \times (a_2, b_2]$ .

La fórmula  $n$ -dimensional análoga de (34) es complicada y no es relevante para el desarrollo posterior. (Se obtiene aplicando la fórmula de inclusión-exclusión para calcular la probabilidad de la unión de eventos.)

## Clasificación

1. *Vectores aleatorios discretos.* El vector aleatorio  $\mathbf{X}$  se dice *discreto* cuando existe un conjunto numerable  $\mathbb{A} \subset \mathbb{R}^n$  tal que  $\mathbb{P}(\mathbf{X} \in \mathbb{A}) = 1$ . En tal caso, las variables aleatorias  $X_1, \dots, X_n$  son discretas y la función  $p_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, 1]$  definida por

$$p_{\mathbf{X}}(\mathbf{x}) := \mathbb{P}(\mathbf{X} = \mathbf{x}) \quad (35)$$

se llama la *función de probabilidad conjunta* de  $\mathbf{X}$ . Su relación con la función de distribución conjunta es la siguiente

$$F_{\mathbf{X}}(\mathbf{x}) = \sum_{\mathbf{y} \in S_{\mathbf{x}}} p_{\mathbf{X}}(\mathbf{y}).$$

2. *Vectores aleatorios continuos.* El vector aleatorio  $\mathbf{X} = (X_1, \dots, X_n)$  se dice *continuo* cuando existe una función  $f_{\mathbf{X}} : \mathbb{R}^n \rightarrow \mathbb{R}^+$ , llamada *densidad de probabilidades conjunta* de  $X_1, \dots, X_n$  tal que

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{S_{\mathbf{x}}} f_{\mathbf{X}}(\mathbf{y}) d\mathbf{y}.$$

(Para evitar dificultades relacionadas con el concepto de integración supondremos que las densidades son seccionalmente continuas.)

3. *Vectores aleatorios mixtos.* El vector aleatorio  $\mathbf{X}$  se dice *mixto* si no es continuo ni discreto.

**Cálculo de probabilidades** Dependiendo del caso, la función de probabilidad conjunta  $p_{\mathbf{X}}(\mathbf{x})$ , o la densidad conjunta  $f_{\mathbf{X}}(\mathbf{x})$ , resume toda la información relevante sobre el comportamiento del vector  $\mathbf{X}$ . Más precisamente, para todo conjunto  $A \subset \mathbb{R}^n$  “suficientemente regular”, vale que

$$\mathbb{P}(\mathbf{X} \in A) = \begin{cases} \sum_{\mathbf{x} \in A} p_{\mathbf{X}}(\mathbf{x}) & \text{en el caso discreto,} \\ \int_A f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} & \text{en el caso continuo.} \end{cases}$$

**Ejemplo 2.2.** Sea  $(X, Y)$  un vector aleatorio continuo con densidad conjunta  $f_{X,Y}(x, y)$ . Si  $a < b$  y  $c < d$ , entonces

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x, y) dx dy. \quad (36)$$

**Ejemplo 2.3** (Distribución uniforme). Sea  $\Lambda \subset \mathbb{R}^2$  una región del plano acotada de área  $|\Lambda|$ . Si la densidad conjunta de un vector aleatorio continuo  $(X, Y)$  es de la forma

$$f_{X,Y}(x, y) = \frac{1}{|\Lambda|} \mathbf{1}\{(x, y) \in \Lambda\}, \quad (37)$$

diremos que  $(X, Y)$  está *uniformemente distribuido sobre*  $\Lambda$  y escribiremos  $(X, Y) \sim \mathcal{U}(\Lambda)$ . Sea  $\mathcal{B} \subset \Lambda$  una sub-región de  $\Lambda$  de área  $|\mathcal{B}|$ . La probabilidad de que  $(X, Y) \in \mathcal{B}$  se calcula del siguiente modo

$$\mathbb{P}((X, Y) \in \mathcal{B}) = \iint_{\mathcal{B}} f_{X,Y}(x, y) dx dy = \iint_{\mathcal{B}} \frac{1}{|\Lambda|} dx dy = \frac{|\mathcal{B}|}{|\Lambda|}. \quad (38)$$

En otras palabras, la probabilidad de que  $(X, Y) \in \mathcal{B}$  es la proporción del área de la región  $\Lambda$  contenida en la sub-región  $\mathcal{B}$ .  $\square$

**Ejemplo 2.4.** Sea  $(X, Y)$  un vector aleatorio uniformemente distribuido sobre el cuadrado  $[0, 1] \times [0, 1]$ . ¿Cuánto vale  $\mathbb{P}(XY > 1/2)$ ?

Debido a que el cuadrado  $[0, 1] \times [0, 1]$  tiene área 1 la probabilidad requerida es el área de la región  $\mathcal{B} = \{(x, y) \in [0, 1] \times [0, 1] : xy > 1/2\}$ . Ahora bien,

$$(x, y) \in \mathcal{B} \iff y > 1/2x \quad (39)$$

y como  $y \leq 1$ , la desigualdad del lado derecho de (39) sólo es posible si  $1/2 \leq x$ . Vale decir,

$$\mathcal{B} = \{(x, y) : 1/2 \leq x \leq 1, 1/2x < y \leq 1\}.$$

En consecuencia,

$$\begin{aligned} \mathbb{P}(XY > 1/2) &= |\mathcal{B}| = \iint_{\mathcal{B}} 1 dx dy = \int_{1/2}^1 \left( \int_{1/2x}^1 1 dy \right) dx = \int_{1/2}^1 \left( 1 - \frac{1}{2x} \right) dx \\ &= \frac{1}{2} + \frac{1}{2} \log \left( \frac{1}{2} \right) = \frac{1}{2} (1 - \log 2) \approx 0.1534\dots \end{aligned}$$

$\square$

## 2.2. Distribuciones marginales

Sea  $\mathbf{X} = (X_1, \dots, X_n)$  un vector aleatorio  $n$ -dimensional y sea  $F_{\mathbf{X}}(\mathbf{x})$  su función de distribución conjunta. La coordenadas de  $\mathbf{X}$  son variables aleatorias. Cada variable individual  $X_i$  tiene su correspondiente función de distribución

$$F_{X_i}(x_i) = \mathbb{P}(X_i \leq x_i). \quad (40)$$

Para enfatizar la relación entre  $X_i$  y el vector  $\mathbf{X} = (X_1, \dots, X_n)$  se dice que  $F_{X_i}(x_i)$  es la *función de distribución marginal de  $X_i$*  o la  *$i$ -ésima marginal de  $\mathbf{X}$* .

**Nota Bene.** Observar que, para cada  $i = 1, \dots, n$ , la función de distribución marginal de  $X_i$ ,  $F_{X_i}(x_i)$ , se obtiene de la función de distribución conjunta  $F_{\mathbf{X}}(x_1, \dots, x_n)$  fijando el valor de  $x_i$  y haciendo  $x_j \rightarrow \infty$  para toda  $j \neq i$ .  $\square$

## Marginales discretas

**Caso bidimensional.** Sea  $(X, Y)$  un vector aleatorio discreto definido sobre un espacio de probabilidad  $(\Omega, \mathcal{A}, \mathbb{P})$  con función de probabilidad conjunta  $p_{X,Y}(x, y)$ . Los números  $p_{X,Y}(x, y)$ ,  $x \in X(\Omega) = \{X(\omega) : \omega \in \Omega\}$ ,  $y \in Y(\Omega) = \{Y(\omega) : \omega \in \Omega\}$  se pueden representar en la forma de una matriz con las siguientes propiedades

$$p_{X,Y}(x, y) \geq 0, \quad y \quad \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} p_{X,Y}(x, y) = 1. \quad (41)$$

Fijando  $x \in X(\Omega)$  y sumando las probabilidades que aparecen en la fila  $x$  de la matriz  $p_{X,Y}(x, y)$  se obtiene

$$\sum_{y \in Y(\Omega)} p_{X,Y}(x, y) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) = p_X(x). \quad (42)$$

Fijando  $y \in Y(\Omega)$  y sumando las probabilidades que aparecen en la columna  $y$  de la matriz  $p_{X,Y}(x, y)$  se obtiene

$$\sum_{x \in X(\Omega)} p_{X,Y}(x, y) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x, Y = y) = \mathbb{P}(Y = y) = p_Y(y). \quad (43)$$

En otras palabras, sumando las probabilidades por filas obtenemos la función de probabilidad *marginal* de la variable aleatoria  $X$  y sumando las probabilidades por columnas obtenemos la función de probabilidad *marginal* de la variable aleatoria  $Y$ . El adjetivo “marginal” que reciben las funciones de probabilidad  $p_X(x)$  y  $p_Y(y)$  refiere a la apariencia externa que adoptan (42) y (43) en una tabla de doble entrada.

**Ejemplo 2.5.** Supongamos que  $X$  e  $Y$  tengan distribución conjunta dada por la siguiente tabla:

$X \setminus Y$	1	2	3
1	0	1/5	0
2	1/5	1/5	1/5
3	0	1/5	0

Para determinar las distribuciones marginales de  $X$  e  $Y$  hay que sumar los valores de cada fila y los valores de cada columna, respectivamente:

$$p_X(1) = 1/5, \quad p_X(2) = 3/5, \quad p_X(3) = 1/5, \\ p_Y(1) = 1/5, \quad p_Y(2) = 3/5, \quad p_Y(3) = 1/5.$$

□

**Ejemplo 2.6.** En una urna hay 6 bolas rojas, 5 azules y 4 verdes. Se extraen dos. Sean  $X$  la cantidad de bolas rojas extraídas e  $Y$  la cantidad de azules.

Existen  $\binom{15}{2} = 105$  resultados posibles. La cantidad de resultados con  $x$  rojas,  $y$  azules y  $2 - (x + y)$  verdes es

$$\binom{6}{x} \binom{5}{y} \binom{4}{2 - (x + y)}$$

Usando esa fórmula obtenemos

$X \setminus Y$	0	1	2	$p_X$
0	6/105	20/105	10/105	36/105
1	24/105	30/105	0	54/105
2	15/105	0	0	15/105
$p_Y$	45/105	50/105	10/105	

Cuadro 1: Distribución conjunta de  $(X, Y)$ . En el margen derecho de la tabla se encuentra la distribución marginal de  $X$  y en el margen inferior, la marginal de  $Y$ .  $\square$

**Caso general.** Para cada  $i = 1, \dots, n$ , la función de probabilidad marginal de  $X_i$ ,  $p_{X_i}(x_i)$ , se puede obtener fijando la variable  $x_i$  y sumando la función de probabilidad conjunta  $p_{\mathbf{X}}(\mathbf{x})$  respecto de las demás variables

$$p_{X_i}(x_i) = \sum_{\mathbf{x}_{\{i\}^c}} p_{\mathbf{X}}(\mathbf{x}).$$

## Marginales continuas

Sea  $(X, Y)$  un vector aleatorio continuo con función densidad conjunta  $f_{X,Y}(x, y)$ .

Las funciones de distribución *marginales* de las variables individuales  $X$  e  $Y$  se obtienen de la distribución conjunta haciendo lo siguiente

$$F_X(x) = \mathbb{P}(X \leq x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y) = \int_{-\infty}^x \left( \int_{-\infty}^{\infty} f_{X,Y}(s, y) dy \right) ds, \quad (44)$$

$$F_Y(y) = \mathbb{P}(Y \leq y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y) = \int_{-\infty}^y \left( \int_{-\infty}^{\infty} f_{X,Y}(x, t) dx \right) dt. \quad (45)$$

Aplicando en (44) y en (45) el Teorema Fundamental del Cálculo Integral se obtiene que las funciones de distribución marginales  $F_X(x)$  y  $F_Y(y)$  son derivables (salvo quizás en un conjunto despreciable de puntos) y vale que

$$f_X(x) = \frac{d}{dx}F_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad (46)$$

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx. \quad (47)$$

En consecuencia, las variables aleatorias  $X$  e  $Y$  son *individualmente* (absolutamente) *continuas* con densidades “marginales”  $f_X(x)$  y  $f_Y(y)$ , respectivamente.

**Ejemplo 2.7** (Distribución uniforme). Sea  $\Lambda \subset \mathbb{R}^2$  una región del plano acotada, que para simplificar supondremos convexa, y sea  $(X, Y)$  un vector aleatorio uniformemente distribuido sobre  $\Lambda$ . La densidad marginal de  $X$  en la abscisa  $x$  es igual al cociente entre el ancho de  $\Lambda$  en  $x$  y el área de  $\Lambda$ .  $\square$

**Ejemplo 2.8** (Dardos). Consideramos un juego de dardos de blanco circular  $\Lambda$  de radio 1 centrado en el origen del plano:  $\Lambda = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ . Un tirador lanza un dardo al azar sobre  $\Lambda$  y se clava en un punto de coordenadas  $(X, Y)$ . El punto  $(X, Y)$  está uniformemente distribuido sobre  $\Lambda$ . Debido a que el área de  $\Lambda$  es igual a  $\pi$ , la densidad conjunta de  $X$  e  $Y$  es

$$f_{X,Y}(x, y) = \frac{1}{\pi} \mathbf{1}\{x^2 + y^2 \leq 1\}.$$

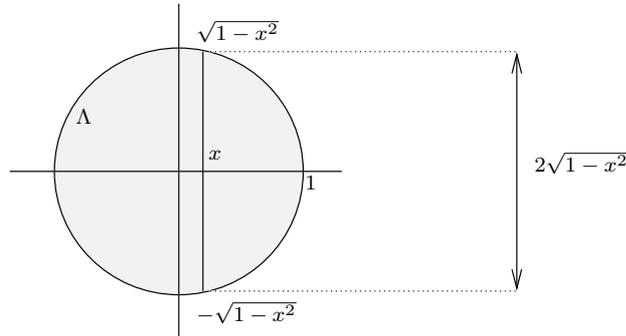


Figura 10: Para cada  $x \in [-1, 1]$  se observa que el ancho del círculo en  $x$  es  $2\sqrt{1-x^2}$ .

Si se observa la Figura 10 es claro que la densidad marginal de  $X$  es

$$f_X(x) = \frac{2\sqrt{1-x^2}}{\pi} \mathbf{1}\{x \in [-1, 1]\},$$

y por razones de simetría la densidad marginal de  $Y$  debe ser

$$f_Y(y) = \frac{2\sqrt{1-y^2}}{\pi} \mathbf{1}\{y \in [-1, 1]\}.$$

$\square$

**Caso general.** Para cada  $i = 1, \dots, n$ , la densidad marginal de  $X_i$ ,  $f_{X_i}(x_i)$ , se puede obtener fijando la variable  $x_i$  e integrando la densidad conjunta  $f_{\mathbf{X}}(x)$  respecto de las demás variables

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{n-1}} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}_{\{i\}^c}.$$

**Nota Bene: Conjuntas y marginales.** A veces, es necesario conocer la distribución de una subcolección de variables aleatorias. En el caso bidimensional este problema no se manifiesta porque se reduce al cálculo de las marginales. Para cada subconjunto de índices  $\Lambda \subset \{1, 2, \dots, n\}$  la función de distribución conjunta de las variables  $X_i : i \in \Lambda$ ,  $F_{\Lambda}(\mathbf{x}_{\Lambda})$ , se obtiene fijando los valores de las coordenadas  $x_i : i \in \Lambda$  y haciendo  $x_j \rightarrow \infty$  para toda  $j \notin \Lambda$ .

En el caso discreto, la función de probabilidad conjunta de las variables  $X_i : i \in \Lambda$ ,  $p_{\Lambda}(x_{\Lambda})$ , se obtiene fijando la variables  $x_i : i \in \Lambda$  y sumando la función de probabilidad conjunta  $p(\mathbf{x})$  respecto de las demás variables

$$p_{\Lambda}(\mathbf{x}_{\Lambda}) = \sum_{\mathbf{x}_{\Lambda^c}} p_{\mathbf{X}}(\mathbf{x}).$$

En el caso continuo, la densidad conjunta de las variables  $X_{\Lambda}$ ,  $f_{\Lambda}(\mathbf{x}_{\Lambda})$ , se obtiene fijando los valores de las variables  $x_i : i \in \Lambda$  e integrando la densidad conjunta  $f(x)$  respecto de las demás variables

$$f_{\Lambda}(\mathbf{x}_{\Lambda}) = \int_{\mathbb{R}^{n-m}} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}_{\Lambda^c}.$$

donde  $m$  es la cantidad de índices contenidos en el conjunto  $\Lambda$ .

### 2.3. Independencia

Las variables  $X_1, \dots, X_n$  son *independientes* si para cualquier colección de conjuntos (medibles)  $A_1, \dots, A_n \subset \mathbb{R}$ , los eventos  $\{X_1 \in A_1\}, \dots, \{X_n \in A_n\}$  son independientes.

Tomando conjuntos de la forma  $A_i = (-\infty, x_i]$  se deduce que la independencia de  $X_1, \dots, X_n$  implica

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P} \left( \bigcap_{i=1}^n \{X_i \leq x_i\} \right) = \prod_{i=1}^n \mathbb{P}(X_i \leq x_i) = \prod_{i=1}^n F_{X_i}(x_i). \quad (48)$$

Dicho en palabras, la independencia de las variables implica que *su función de distribución conjunta se factoriza como el producto de todas las marginales*.

Recíprocamente, se puede demostrar que si para cada  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$  se verifica la ecuación (48), las variables aleatorias  $X_1, \dots, X_n$  son independientes. (La demostración es técnica y no viene al caso). Esta equivalencia reduce al mínimo las condiciones que permiten caracterizar la independencia de variables aleatorias y motivan la siguiente definición más simple.

**Definición 2.9** (Independencia de una cantidad finita de variables aleatorias). Diremos que las variables aleatorias  $X_1, \dots, X_n$  son *independientes* si la ecuación (48) se verifica en todo  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ .

**Definición 2.10** (Independencia). Dada una familia de variables aleatorias  $(X_i : i \in \mathbb{I})$  definidas sobre un mismo espacio de probabilidad  $(\Omega, \mathcal{A}, \mathbb{P})$ , diremos que sus variables son (*conjuntamente*) *independientes* si para cualquier subconjunto finito de índices  $J \subset \mathbb{I}$  las variables  $X_i, i \in J$  son independientes.

**Nota Bene.** La independencia de las variables aleatorias  $X_1, \dots, X_n$  es equivalente a la factorización de la distribución conjunta como producto de sus distribuciones marginales. Más aún, esta propiedad se manifiesta a nivel de la función de probabilidad,  $p_{\mathbf{X}}(\mathbf{x})$  o de la densidad conjunta,  $f_{\mathbf{X}}(\mathbf{x})$ , del vector aleatorio  $\mathbf{X} = (X_1, \dots, X_n)$ , según sea el caso. Para ser más precisos,  $X_1, \dots, X_n$  son independientes si y solo si

$$p_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n p_{X_i}(x_i) \quad \text{en el caso discreto,}$$

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n f_{X_i}(x_i) \quad \text{en el caso continuo.}$$

### Caso bidimensional discreto

Sea  $(X, Y)$  un vector aleatorio discreto con función de probabilidad conjunta  $p_{X,Y}(x, y)$  y marginales  $p_X(x)$  y  $p_Y(y)$ . Las variables  $X, Y$  son *independientes* si para cada pareja de valores  $x \in X(\Omega)$ ,  $y \in Y(\Omega)$  vale que

$$p_{X,Y}(x, y) = p_X(x) p_Y(y) \quad (49)$$

En otras palabras, la matriz  $p_{X,Y}(x, y)$  es la tabla de multiplicar de las marginales  $p_X(x)$  y  $p_Y(y)$ .

**Criterio para detectar dependencia.** Cuando en la tabla de la distribución conjunta de dos variables hay un 0 ubicado en la intersección de una fila y una columna de sumas positivas, las variables no pueden ser independientes. (Las variables de los Ejemplos 2.5 y 2.6 no son independientes.)  $\square$

### Caso bidimensional continuo

Sean  $X$  e  $Y$  variables aleatorias con densidad conjunta  $f_{X,Y}(x, y)$  y marginales  $f_X(x)$  y  $f_Y(y)$ . Las variables aleatorias  $X$  e  $Y$  son *independientes* si y solo si

$$f_{X,Y}(x, y) = f_X(x) f_Y(y). \quad (50)$$

En otras palabras,  $X$  e  $Y$  son independientes si y solo si su densidad conjunta se factoriza como el producto de las marginales.

**Ejemplo 2.11** (Variables independientes). Sean  $X$  e  $Y$  variables aleatorias continuas, definidas sobre un mismo espacio de probabilidad, con densidades  $f_1(x)$  y  $f_2(y)$ . Entonces, podemos definir una densidad de probabilidades  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^+$  mediante

$$f(x, y) = f_1(x)f_2(y),$$

y considerar que  $f(x, y)$  es la densidad conjunta de las variables  $X$  e  $Y$ . En tal caso,  $X$  e  $Y$  resultan independientes y las funciones  $f_1$  y  $f_2$  representan las densidades marginales de  $X$  e  $Y$ , respectivamente.  $\square$

### Criterios para detectar (in)dependencia.

**1.** La independencia de  $X$  e  $Y$  equivale a la existencia de dos funciones  $f_1(x)$  y  $f_2(y)$  tales que  $f_{X,Y}(x, y) = f_1(x)f_2(y)$ . Por lo tanto, para verificar independencia basta comprobar que la densidad conjunta se puede factorizar como *alguna* función de  $x$  por *alguna* función de  $y$ , siendo innecesario verificar que se trata de las densidades marginales. (*Ejercicio*)

**2.** La factorización (50) implica que, si  $X$  e  $Y$  son independientes, el recinto del plano  $Sop(f_{X,Y}) := \{(x, y) \in \mathbb{R}^2 : f_{X,Y}(x, y) > 0\}$ , llamado *el soporte de la densidad conjunta*  $f_{X,Y}$ , debe coincidir con el producto cartesiano de los soportes de sus densidades marginales:  $Sop(f_X) \times Sop(f_Y) = \{x \in \mathbb{R} : f_X(x) > 0\} \times \{y \in \mathbb{R} : f_Y(y) > 0\}$ . Por ejemplo, si el soporte de la densidad conjunta es *conexo* y no es un rectángulo las variables  $X$  e  $Y$  no pueden ser independientes. (Ver el Ejemplo 2.8.)

## 3. Variables truncadas

Sea  $X$  una variable aleatoria definida sobre un espacio de probabilidad  $(\Omega, \mathcal{A}, \mathbb{P})$ . Sea  $B \subset \mathbb{R}$  un conjunto tal que  $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}$  y tal que  $\mathbb{P}(X \in B) > 0$ .

*Truncar la variable aleatoria  $X$  al conjunto  $B$*  significa condicionarla a tomar valores en el conjunto  $B$ .

Mediante  $X|B$  designaremos la variable aleatoria obtenida por truncar  $X$  al conjunto  $B$ . Por definición, la función de distribución de  $X|B$  es

$$F_{X|B}(x) = \mathbb{P}(X \leq x | X \in B) = \frac{\mathbb{P}(X \leq x, X \in B)}{\mathbb{P}(X \in B)}. \quad (51)$$

**Caso absolutamente continuo.** Si la variable aleatoria  $X$  es absolutamente continua con densidad de probabilidades  $f_X(x)$ , la función de distribución de  $X|B$  adopta la forma

$$F_{X|B}(x) = \frac{\int_{\{X \leq x\} \cap \{X \in B\}} f_X(x) dx}{\mathbb{P}(X \in B)} = \frac{\int_{-\infty}^x f_X(x) \mathbf{1}\{x \in B\} dx}{\mathbb{P}(X \in B)}. \quad (52)$$

Por lo tanto  $X|B$  es una variable aleatoria absolutamente continua con densidad de probabilidades

$$f_{X|B}(x) = \frac{f_X(x)}{\mathbb{P}(X \in B)} \mathbf{1}\{x \in B\}. \quad (53)$$

**Nota Bene.** La densidad condicional  $f_{X|B}(x)$  es cero fuera del conjunto condicionante  $B$ . Dentro del conjunto condicionante la densidad condicional tiene exactamente la misma forma que la densidad incondicional, salvo que está escalada por el factor de normalización  $1/\mathbb{P}(X \in A)$  que asegura que  $f_{X|A}(x)$  integra 1.  $\square$

**Caso discreto.** El caso discreto se trata en forma análoga a la anterior. La función de probabilidad de  $X|B$  adopta la forma

$$p_{X|B}(x) = \frac{\mathbb{P}(X = x)}{\mathbb{P}(X \in B)} \mathbf{1}\{x \in B\}. \quad (54)$$

**Ejemplo 3.1** (Dado equilibrado). Sea  $X$  el resultado del tiro de un dado equilibrado y sea  $A$  el evento “el resultado del tiro es un número par”. Aplicando la fórmula anterior obtenemos

$$p_{X|A}(x) = \frac{1/6}{1/2} \mathbf{1}\{x = 2, 4, 6\} = \frac{1}{3} \mathbf{1}\{x = 2, 4, 6\}. \quad (55)$$

$\square$

### 3.1. Dividir y conquistar

**Teorema 3.2** (Fórmula de probabilidad total). Sea  $X$  una variable aleatoria absolutamente continua con densidad de probabilidades  $f_X(x)$  y sea  $A_1, \dots, A_n$  una partición en eventos disjuntos dos a dos del espacio muestral  $\Omega$  tal que  $\mathbb{P}(A_i) > 0$  para cada  $i$ . Entonces,

$$f_X(x) = \sum_{i=1}^n \mathbb{P}(A_i) f_{X|A_i}(x). \quad (56)$$

**Demostración.** La prueba se basa en la fórmula de probabilidad total.

$$\mathbb{P}(X \leq x) = \sum_{i=1}^n \mathbb{P}(A_i) \mathbb{P}(X \leq x | A_i).$$

Esta fórmula se puede reescribir en la forma

$$\int_{-\infty}^x f_X(t) dt = \sum_{i=1}^n \mathbb{P}(A_i) \int_{-\infty}^x f_{X|A_i}(t) dt.$$

La fórmula (56) se obtiene derivando ambos lados de la última igualdad.  $\square$

**Ejemplo 3.3** (Dividir y conquistar). Todas las mañanas Lucas llega a la estación del subte entre las 7:10 y las 7:30 (con distribución uniforme en el intervalo). El subte llega a la estación cada quince minutos comenzando a las 6:00. Cuál es la densidad de probabilidades del tiempo que tiene que esperar Lucas hasta subirse al subte?

Sea  $X$  el tiempo de llegada de Lucas a la estación del subte,  $X \sim \mathcal{U}[7:10, 7:30]$ . Sea  $Y$  el tiempo de espera. Consideramos los eventos  $A = \{7:10 \leq X \leq 7:15\} = \text{"Lucas sube en el subte de las 7:15"}$ ;  $B = \{7:15 < X \leq 7:30\} = \text{"Lucas sube en el subte de las 7:30"}$ .

Condicionado al evento  $A$ , el tiempo de llegada de Lucas a la estación del subte es uniforme entre las 7:10 y las 7:15. En ese caso, el tiempo de espera  $Y$  también es uniforme entre 0 y 5 minutos. Análogamente, condicionado al evento  $B$ ,  $Y$  es uniforme entre 0 y 15 minutos. La densidad de probabilidades de  $Y$  se obtiene usando la fórmula de probabilidad total

$$\begin{aligned} f_Y(y) &= \left(\frac{5}{20}\right) \frac{1}{5} \mathbf{1}\{0 \leq y \leq 5\} + \left(\frac{15}{20}\right) \frac{1}{15} \mathbf{1}\{0 \leq y \leq 15\} \\ &= \frac{1}{10} \mathbf{1}\{0 \leq y \leq 5\} + \frac{1}{20} \mathbf{1}\{5 \leq y \leq 15\}. \end{aligned}$$

□

### 3.2. Memoria

**Ejemplo 3.4** (La exponencial no tiene memoria). Lucas camina hacia la parada del colectivo donde el tiempo,  $T$ , entre llegadas de colectivos tiene distribución exponencial de intensidad  $\lambda$ . Supongamos que Lucas llega  $t$  minutos después de la llegada de un colectivo. Sea  $X$  el tiempo que Lucas tendrá que esperar hasta que llegue el próximo colectivo. Cuál es la distribución del tiempo de espera  $X$ ?

Designamos mediante  $A = \{T > t\}$  el evento "*Lucas llegó  $t$  minutos después de la llegada de un colectivo*". Tenemos que

$$\begin{aligned} \mathbb{P}(X > x|A) &= \mathbb{P}(T > t + x|T > t) = \frac{\mathbb{P}(T > t + x, T > t)}{\mathbb{P}(T > t)} \\ &= \frac{\mathbb{P}(T > t + x)}{\mathbb{P}(T > t)} = \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x}. \end{aligned}$$

□

**Nota Bene.** En general, si modelamos el tiempo para completar cierta operación por una variable aleatoria  $T$  con distribución exponencial, la propiedad de pérdida de memoria implica que mientras la operación no haya sido completada, el tiempo restante para completarla tiene la misma función de distribución, no importa cuando haya empezado la operación. □

**Ejemplo 3.5** (Exponencial truncada a la derecha). Sea  $T$  una variable aleatoria con distribución exponencial de intensidad  $\lambda > 0$  y sea  $t_0 > 0$ . Según la fórmula (53) la variable aleatoria  $T$  truncada a la semirecta  $(t, +\infty)$ ,  $T|\{T > t_0\}$ , tiene la siguiente densidad de probabilidades

$$f_{T|\{T > t_0\}}(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t_0}} \mathbf{1}\{t > t_0\} = e^{-\lambda(t-t_0)} \mathbf{1}\{t - t_0 > 0\} = f_T(t - t_0).$$

En otros términos, si  $T \sim \text{Exp}(\lambda)$ , entonces  $T|\{T > t_0\} \sim t_0 + \text{Exp}(\lambda)$ . □

## 4. Bibliografía consultada

Para redactar estas notas se consultaron los siguientes libros:

1. Bertsekas, D. P., Tsitsiklis, J. N.: Introduction to Probability. M.I.T. Lecture Notes. (2000)
2. Chung, K. L.: A Course in Probability Theory. Academic Press, San Diego. (2001)
3. Durrett R.: Probability. Theory and Examples. Duxbury Press, Belmont. (1996)
4. Feller, W.: An introduction to Probability Theory and Its Applications. Vol. 1. John Wiley & Sons, New York. (1968)
5. Feller, W.: An introduction to Probability Theory and Its Applications. Vol. 2. John Wiley & Sons, New York. (1971)
6. Grimmett, G. R., Stirzaker, D. R.: Probability and Random Processes. Oxford University Press, New York. (2001)
7. Johnson, N. L., Kotz, S., Balakrishnan, N.: Continuous Univariate Distributions. Vol. 1. John Wiley & Sons, New York. (1995)
8. Kolmogorov, A. N.: Foundations of the Theory of Probability. Chelsea Publishing Co., New York. (1956)
9. Maronna R.: Probabilidad y Estadística Elementales para Estudiantes de Ciencias. Editorial Exacta, La Plata. (1995).
10. Pugachev, V. S.: Introducción a la Teoría de las Probabilidades. Mir, Moscú. (1973)
11. Ross, S.: Introduction to Probability Models. Academic Press, San Diego. (2007)