Vectores aleatorios: marginales e independencia (Borradores, Curso 23)

Sebastian Grynberg

25 de marzo 2013

Um coup de dés jamais n'abolira le hasard

(Stéphane Mallarmé)

Índice

1. Vectores aleatorios

Notación. Para simplificar la escritura usaremos las siguientes notaciones. Los puntos del espacio n-dimensional $\mathbb{R}^n$, $n \geq 2$, 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 se puede interpretar 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 $J = \{i_1, \dots, i_m\} \subset \{1, \dots, n\}$ denotaremos mediante $\mathbf{x}_J \in \mathbb{R}^m$ al punto m-dimensional que se obtiene de $\mathbf{x}$ quitándole todas las coordenadas que tengan índices fuera de J. Por ejemplo, si $J = \{1, 2\}$, entonces $\mathbf{x}_J = (x_1, x_2)$.

Definición 1.1. Un *vector aleatorio* sobre un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$ es una función $\mathbf{X} = (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ tal que para todo $\mathbf{x} \in \mathbb{R}^n$

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

1.1. Distribución conjunta

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

$$F_{\mathbf{X}}(\mathbf{x}) := \mathbb{P}(\mathbf{X} \in S_{\mathbf{x}}) \tag{1}$$

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 más simple: $n = 2$. Si $a_1 < b_1$ y $a_2 < b_2$ vale que[1]

$$\mathbb{P}(a_1 < X_1 \leq b_1,\, a_2 < X_2 \leq b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2). \tag{2}$$

La identidad (2) 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 (2) 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.)

[1] Ver la Figura 1.

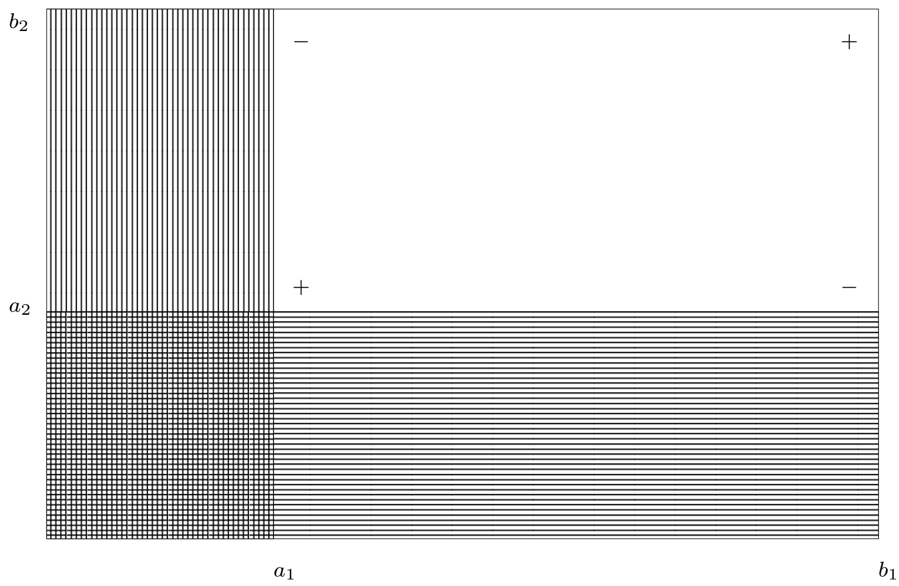

Figura 1: Esquema de la demostración de la identidad (2). El rectángulo $(a_1, b_1] \times (a_2, b_2]$ se puede representar en la forma $S_{(b_1,b_2)} \setminus \big(S_{(a_1,b_2)} \cup S_{(b_1,a_2)}\big)$.

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 $p_{\mathbf{X}} : \mathbb{R}^n \to [0, 1]$ definida por

$$p_{\mathbf{X}}(\mathbf{x}) := \mathbb{P}(\mathbf{X} = \mathbf{x}) \tag{3}$$

   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 $f_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{R}^+$, 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 aleatorio $\mathbf{X}$. 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 1.2. Sea (X, Y) un vector aleatorio continuo con densidad conjunta $f_{X,Y}(x, y)$. Si $a < b$ y $c < d$, entonces

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x, y) dx dy. \tag{4}$$

Ejemplo 1.3 (Distribución uniforme)**.** Sea $\Lambda \subset \mathbb{R}^2$ una región acotada de área $|\Lambda|$. Si la densidad conjunta de un vector aleatorio continuo (X, Y) es de la forma

$$f_{X,Y}(x, y) = \frac{1}{|\Lambda|} \mathbf{1}\{(x, y) \in \Lambda\}, \tag{5}$$

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

$$\mathbb{P}((X, Y) \in \mathcal{B}) = \iint_{\mathcal{B}} f_{X,Y}(x, y) dx dy = \iint_{\mathcal{B}} \frac{1}{|\Lambda|} dx dy = \frac{|\mathcal{B}|}{|\Lambda|}. \tag{6}$$

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

Ejemplo 1.4. 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,

$$(x, y) \in \mathcal{B} \iff y > 1/2x \tag{7}$$

y como $y \leq 1$, la desigualdad del lado derecho de (7) 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,

$$\begin{aligned} \mathbb{P}(XY > 1/2) &= |\mathcal{B}| = \iint_{\mathcal{B}} 1 \, dx dy = \int_{\frac{1}{2}}^1 \left(\int_{\frac{1}{2x}}^1 1 \, dy \right) dx = \int_{\frac{1}{2}}^1 \left(1 - \frac{1}{2x} \right) dx \\ &= \frac{1}{2} + \frac{1}{2} \log \left(\frac{1}{2} \right) = \frac{1}{2}(1 - \log 2) \approx 01534.... \end{aligned}$$

□

1.2. 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

$$F_{X_i}(x_i) = \mathbb{P}(X_i \leq x_i). \tag{8}$$

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 $x_j \to \infty$ para toda $j \neq i$. □

1.2.1. 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 $p_{X,Y}(x, y)$, $(x, y) \in X(\Omega) \times Y(\Omega) = \{(X(\omega), Y(\omega)) : \omega \in \Omega\}$, se pueden representar en la forma de una matriz con las siguientes propiedades

$$p_{X,Y}(x, y) \geq 0, \qquad \text{y} \qquad \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} p_{X,Y}(x, y) = 1. \tag{9}$$

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

$$\sum_{y \in Y(\Omega)} p_{X,Y}(x, y) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) = p_X(x). \tag{10}$$

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

$$\sum_{x \in X(\Omega)} p_{X,Y}(x, y) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x, Y = y) = \mathbb{P}(Y = y) = p_Y(y). \tag{11}$$

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 (10) y (11) en una tabla de doble entrada.

Ejemplo 1.5. 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 y poniendo $q = 1/105$ obtenemos

$x \setminus y$	0	1	2	p_X
0	$6q$	$20q$	$10q$	$36q$
1	$24q$	$30q$	0	$54q$
2	$15q$	0	0	$15q$
p_Y	$45q$	$50q$	$10q$	

Figura 2: 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

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

1.2.2. 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

$$F_X(x) = \mathbb{P}(X \leq x) = \lim_{y \to \infty} F_{X,Y}(x, y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f_{X,Y}(s, y)\, dy \right) ds, \tag{12}$$

$$F_Y(y) = \mathbb{P}(Y \leq y) = \lim_{x \to \infty} F_{X,Y}(x, y) = \int_{-\infty}^{y} \left(\int_{-\infty}^{\infty} f_{X,Y}(x, t)\, dx \right) dt. \tag{13}$$

Aplicando en (12) y en (13) 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

$$f_X(x) = \frac{d}{dx} F_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy, \tag{14}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx. \tag{15}$$

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

Ejemplo 1.6 (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 Λ. La densidad marginal de X en la abscisa x es igual al cociente entre el ancho de Λ en x y el área de Λ. $\square$

Ejemplo 1.7 (Dardos)**.** Consideramos un juego de dardos de blanco circular Λ 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 Λ y se clava en un punto de coordenadas (X, Y). El punto (X, Y) está uniformemente distribuido sobre Λ. Debido a que el área de Λ es igual a π, la densidad conjunta de X e Y es

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

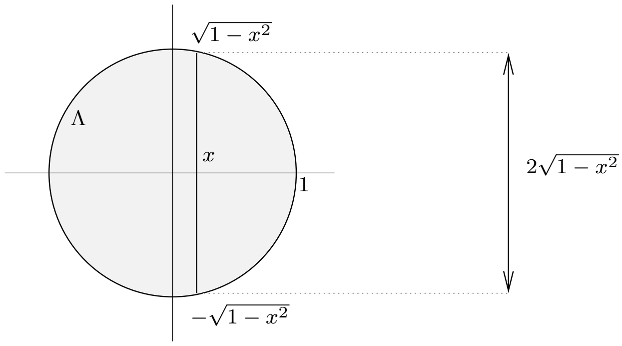

Figura 3: Para cada $x \in [-1, 1]$ se observa que el ancho del círculo en x es $2\sqrt{1 - x^2}$.

Si se observa la Figura 3 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

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

□

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 sub-colección 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 $x_j \to \infty$ 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_Λ, $f_\Lambda(\mathbf{x}_\Lambda)$, 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 Λ.

1.3. 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

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \leq x_i\}\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x_i) = \prod_{i=1}^{n} F_{X_i}(x_i). \tag{16}$$

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 (16), 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 1.8 (Independencia de una cantidad finita de variables aleatorias)**.** Diremos que las variables aleatorias $X_1, \dots, X_n$ son *independientes* si la ecuación (16) se verifica en todo $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$.

Definición 1.9 (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

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

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

Ejemplo 1.10 (Números al azar)**.** Se elige al azar un número U del intervalo $[0,1)$. Sea $U = 0.X_1X_2X_3\cdots$ el desarrollo decimal de U. Mostraremos que los dígitos de U son independientes entre sí y que cada uno de ellos se distribuye uniformemente sobre el conjunto $\{0,1,\dots,9\}$.

El problema se reduce a mostrar que para cada $n \geq 2$ las variables aleatorias $X_1, X_2, \dots, X_n$ son independientes entre sí y que para cada $k \geq 1$ y todo $x_k \in \{0,1,\dots,9\}$, $\mathbb{P}(X_k = x_k) = 1/10$.

Primero observamos que para cada $n \geq 1$ y para todo $(x_1,\dots,x_n) \in \{0,1,\dots,9\}^n$ vale que

$$\bigcap_{i=1}^{n}\{X_i = x_i\} \iff U \in \left[\sum_{i=1}^{n}\frac{x_i}{10^i}, \sum_{i=1}^{n}\frac{x_i}{10^i} + \frac{1}{10^n}\right).$$

En consecuencia,

$$\mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i = x_i\}\right) = \frac{1}{10^n}. \tag{17}$$

Para calcular las marginales de los dígitos observamos que para cada $x_k \in \{0,1,\dots,9\}$ vale que

$$\{X_k = x_k\} = \bigcup_{(x_1,\dots,x_{k-1})\in\{0,1,\dots,9\}^{k-1}}\left[\left(\bigcap_{i=1}^{k-1}\{X_i = x_i\}\right)\cap\{X_k = x_k\}\right].$$

De acuerdo con (17) cada uno de los 10^{k-1} eventos que aparecen en la unión del lado derecho de la igualdad tiene probabilidad $1/10^k$ y como son disjuntos dos a dos obtenemos que

$$\mathbb{P}(X_k = x_k) = 10^{k-1}\frac{1}{10^k} = \frac{1}{10}. \tag{18}$$

De (17) y (18) se deduce que para todo $(x_1,\dots,x_n) \in \{0,1,\dots,9\}^n$ vale que

$$\mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i = x_i\}\right) = \prod_{i=1}^{n}\mathbb{P}(X_i = x_i).$$

Por lo tanto, las variables aleatorias $X_1, X_2, \dots, X_n$ son independientes entre sí y cada una de ellas se distribuye uniformemente sobre el conjunto $\{0,1,\dots,9\}$. $\square$

1.3.1. 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

$$p_{X,Y}(x,y) = p_X(x)\,p_Y(y) \tag{19}$$

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

Ejemplo 1.11. Se arrojan dos dados equilibrados y se observan las variables aleatorias X e Y definidas por $X =$“el resultado del primer dado” e $Y =$“el mayor de los dos resultados”.

El espacio de muestral asociado al experimento se puede representar en la forma $\Omega = \{1, 2, \dots, 6\}^2$, cada punto $(i, j) \in \Omega$ indica que el resultado del primer dado es i y el resultado del segundo es j. Para reflejar que arrojamos *dos dados equilibrados*, todos los puntos de Ω serán equiprobables, i.e., para cada $(i, j) \in \Omega$ se tiene $\mathbb{P}(i, j) = 1/36$. Formalmente las variables aleatorias X e Y están definidas por

$$X(i, j) := i, \qquad Y(i, j) := \text{máx}\{i, j\}. \tag{20}$$

Distribución conjunta y distribuciones marginales de X e Y. En primer lugar vamos a representar el espacio muestral Ω en la forma de una matriz para poder observar más claramente los resultados posibles

$$\begin{pmatrix}
(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\
(2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\
(3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\
(4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\
(5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\
(6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)
\end{pmatrix}$$

Figura 4: Resultados posibles del experimento aleatorio que consiste en arrojar dos dados.

Debido a que $Y \geq X$, tenemos que $p_{X,Y}(x, y) = 0$ para todo $1 \leq y < x \leq 6$. En los otros casos, i.e., $1 \leq x \leq y \leq 6$, para calcular el valor de $p_{X,Y}(x, y)$ hay que contar la cantidad de elementos de la fila x, de la matriz representada en la Figura 4, que contengan alguna coordenada igual a y. Multiplicando por $q = \frac{1}{36}$ la cantidad encontrada se obtiene $p_{X,Y}(x, y)$. En la figura 5 representamos la distribución conjunta $p_{X,Y}(x, y)$ y las distribuciones marginales p_X y p_Y.

$x \setminus y$	1	2	3	4	5	6	p_X
1	q	q	q	q	q	q	$6q$
2	0	$2q$	q	q	q	q	$6q$
3	0	0	$3q$	q	q	q	$6q$
4	0	0	0	$4q$	q	q	$6q$
5	0	0	0	0	$5q$	q	$6q$
6	0	0	0	0	0	$6q$	$6q$
p_Y	q	$3q$	$5q$	$7q$	$9q$	$11q$	

Figura 5: Distribución conjunta de (X, Y). En el margen derecho se encuentra la distribución marginal de X y en el margen inferior, la marginal de Y. Para abreviar hemos puesto $q = \frac{1}{36}$.

De acuerdo con los resultados expuestos en la tabla que aparece en la Figura 5, las distribuciones marginales son

$$p_X(x) = \frac{1}{6}, \qquad p_Y(y) = \frac{2y - 1}{36}.$$

Debido a que no se trata de una tabla de multiplicar las variables X e Y no son independientes. Lo que, por otra parte, constituye una obviedad.

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 del Ejemplo 1.5 no son independientes.) $\square$

1.3.2. 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

$$f_{X,Y}(x,y) = f_X(x)f_Y(y). \tag{21}$$

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

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 (21) 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 1.7.)

Ejemplo 1.12. Sean X e Y variables aleatorias independientes con distribución uniforme sobre el intervalo $(0, L)$. Una vara de longitud L metros se quiebra en dos puntos cuyas distancias a una de sus puntas son X e Y metros. Calcular la probabilidad de que las tres piezas se puedan usar para construir un triángulo.

Primero designamos mediante L_1, L_2 y L_3 a las longitudes de las tres piezas. Las tres piezas se pueden usar para construir un triángulo si y solamente si se satisfacen las desigualdades triangulares

$$L_1 + L_2 > L_3, \quad L_1 + L_3 > L_2 \quad \text{y } L_2 + L_3 > L_1. \tag{22}$$

Vamos a distinguir dos casos: el caso en que $X \leq Y$ y el caso en que $Y < X$. En el primer caso, $X \leq Y$, tenemos que $L_1 = X$, $L_2 = Y - X$ y $L_3 = L - Y$ y las desigualdades triangulares (22) son equivalentes a las siguientes

$$Y > L/2, \quad X + L/2 > Y \quad \text{y } L/2 > X. \tag{23}$$

En el segundo caso, $Y < X$, tenemos que $L_1 = Y$, $L_2 = X - Y$ y $L_3 = L - X$ y las desigualdades triangulares (22) son equivalentes a las siguientes

$$X > L/2, \quad Y > X - L/2 \quad \text{y } L/2 > Y. \tag{24}$$

Por lo tanto, las tres piezas se pueden usar para construir un triángulo si y solamente si $(X, Y) \in \mathcal{B}$, donde

$$\begin{aligned} \mathcal{B} &= \{(x, y) \in (0, L) \times (0, L) : 0 < x < L/2, L/2 < y < x + L/2\} \\ &\cup \{(x, y) \in (0, L) \times (0, L) : L/2 < x < L, x - L/2 < y < L/2\}. \end{aligned} \tag{25}$$

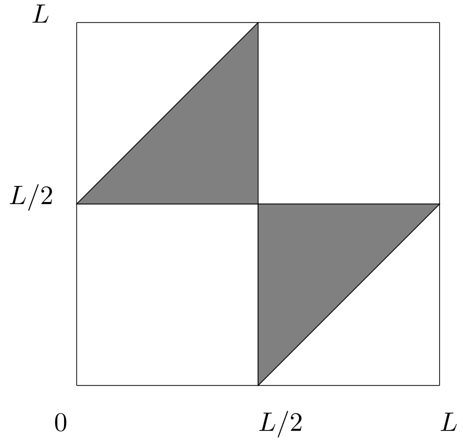

Figura 6: La región sombreada representa al conjunto $\mathcal{B}$ que es la unión de dos triángulos disjuntos cada uno de área $L^2/8$.

La hipótesis de que X e Y son independientes con distribución uniforme sobre el intervalo $(0, L)$ significa que $(X, Y) \sim \mathcal{U}(\Lambda)$, donde Λ es el cuadrado de lado $(0, L)$

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \left(\frac{1}{L}\mathbf{1}\{0 < x < L\}\right)\left(\frac{1}{L}\mathbf{1}\{0 < y < L\}\right) = \frac{1}{L^2}\mathbf{1}\{(x, y) \in \Lambda\}.$$

De (6) se deduce que

$$\mathbb{P}((X, Y) \in \mathcal{B}) = \frac{|\mathcal{B}|}{|\Lambda|} = \frac{(2/8)L^2}{L^2} = \frac{1}{4}. \tag{26}$$

$\square$

2. 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. Feller, W.: An introduction to Probability Theory and Its Applications. Vol. 1. John Wiley & Sons, New York. (1968)
3. Feller, W.: An introduction to Probability Theory and Its Applications. Vol. 2. John Wiley & Sons, New York. (1971)
4. Ross, S.: Introduction to Probability Models. Academic Press, San Diego. (2007)