

Vectores aleatorios y transformaciones

Dr. Arturo Erdely *

Introducción

Esta sección es un repaso y resumen de algunos conceptos y resultados de un curso previo de Probabilidad, con el objeto de unificar notación y extender el concepto de variable aleatoria al de vector aleatorio.

A los *fenómenos* (procesos que observa el ser humano sin intervenir en ellos) y a los *experimentos* (procesos generados por el ser humano) se les añade el calificativo de *aleatorios* cuando no hay certeza sobre su resultado. Matemáticamente se les representa por medio de un *espacio de probabilidad* que consiste en una terna $(\Omega, \mathfrak{F}, \mathbb{P})$. El llamado *espacio muestral* Ω es un conjunto de símbolos que representan cada uno de los posibles resultados del fenómeno o experimento aleatorio. \mathfrak{F} es un conjunto de subconjuntos de Ω que representan eventos. Un *evento* \mathcal{E} es una proposición lógica cuya veracidad o falsedad queda determinada una vez que se conoce el resultado del fenómeno o experimento aleatorio, y el subconjunto $E \subseteq \Omega$ que representa a un evento \mathcal{E} tiene como elementos aquellos símbolos de Ω que representan resultados que hacen verdadera la proposición lógica que define al evento \mathcal{E} .

Ejemplo 1 Considere el experimento aleatorio consistente en lanzar 3 veces una moneda (a una cara la llamaremos “águila” y a la otra “sol”) diseñada de tal forma que la probabilidad de que la cara “águila” quede hacia arriba es un valor $0 < \theta < 1$ (es decir, no necesariamente se trata de una moneda equilibrada). El espacio muestral puede representarse mediante el conjunto:

$$\begin{aligned}\Omega &= \{(\omega_1, \omega_2, \omega_3) : \omega_j \in \{a, s\}\} \\ &= \{(a, a, a), (a, a, s), (a, s, a), (a, s, s), (s, a, a), (s, a, s), (s, s, a), (s, s, s)\}\end{aligned}$$

En este caso Ω resulta ser un espacio muestral finito con un total de $2^3 = 8$ elementos (ternas en este caso) distintos. Un evento, entre muchos que podrían definirse, es el siguiente:

$$\mathcal{E} \equiv \text{“se obtiene la misma cara de la moneda en los 3 lanzamientos.”}$$

El subconjunto $E \subseteq \Omega$ que representa a dicho evento tiene solo dos elementos:

$$E = \{(a, a, a), (s, s, s)\}.$$

□

Después de conocer un resultado de un fenómeno o experimento aleatorio se dice que un evento \mathcal{E} ocurrió si la proposición lógica que lo define resulta ser verdadera, o equivalentemente, si el símbolo que representa a dicho resultado pertenece al conjunto E que representa a dicho evento. En caso contrario, se dice que el evento no ocurrió. Dados dos eventos \mathcal{E}_1 y \mathcal{E}_2 representados por los subconjuntos E_1 y E_2 de Ω tenemos que la conjunción de proposiciones lógicas $\mathcal{E}_1 \wedge \mathcal{E}_2$ es también una proposición lógica cuyo valor de verdad o falsedad depende de la ocurrencia simultánea de dichos eventos, y por tanto es un evento también, que queda representado por el subconjunto $E_1 \cap E_2 \subseteq \Omega$. Comentario análogo para la disyunción lógica $\mathcal{E}_1 \vee \mathcal{E}_2$ que queda representada por el subconjunto $E_1 \cup E_2 \subseteq \Omega$. Más aún, nótese que lo anterior implica que si E_1 y E_2 pertenecen a \mathfrak{F} entonces $E_1 \cup E_2$, $E_1 \cap E_2$, $E_1 \setminus E_2$, $E_2 \setminus E_1$ también pertenecen a la colección \mathfrak{F} .

El tercer elemento de un espacio de probabilidad es una *medida de probabilidad* \mathbb{P} , que es una función que a cada subconjunto $E \subseteq \Omega$ que representa a un evento le asocia un número real en el intervalo $[0, 1]$, mismo que se denota $\mathbb{P}(E)$, y que cuantifica el *grado de incertidumbre* respecto a la ocurrencia del evento. $\mathbb{P}(E) = 1$ significa absoluta

*Profesor de Carrera Titular “C” de tiempo completo definitivo en la UNAM FES Acatlán. Email: aerdely@acatlan.unam.mx

certeza de que ocurre el evento, $\mathbb{P}(E) = 0$ significa absoluta certeza de que no ocurre el evento, y en cualquier otro caso $0 < \mathbb{P}(E) < 1$ significa que no hay certidumbre sobre la ocurrencia del evento. Si, por ejemplo, $\mathbb{P}(E) = 0.9$ es la probabilidad de que mañana llueva, aunque no hay certeza sobre que eso ocurra, el tener un valor más cercano a 1 que a 0 es información útil para tomar previsiones. En este sentido, el caso de mayor incertidumbre es cuando $\mathbb{P}(E) = \frac{1}{2}$ ya que es el punto más lejano a la certeza, esto es, de 0 y 1. La medida de probabilidad \mathbb{P} se deduce o aproxima con base en información que se tiene sobre el fenómeno o experimento aleatorio en cuestión. Pero en términos generales, cualquier medida de probabilidad satisface lo siguiente:

Definición 1 Una *medida de probabilidad* \mathbb{P} es una función que a todo subconjunto $E \subseteq \Omega$ que representa un evento le asocia un número real $\mathbb{P}(E) \in [0, 1]$ tal que:

- a) $\mathbb{P}(\Omega) = 1$.
- b) Si $\{E_1, E_2, \dots\}$ es una colección numerable de subconjuntos disjuntos de Ω que representan eventos entonces

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(E_n).$$

□

Si un evento resulta representado por Ω (esto es, la totalidad de resultados posibles) con toda certeza ocurre, y por ello es que se define $\mathbb{P}(\Omega) = 1$, y a dicho evento se le denomina *evento seguro*. Si dos eventos \mathcal{E}_1 y \mathcal{E}_2 están representados por subconjuntos disjuntos E_1 y E_2 , esto es, tales que $E_1 \cap E_2 = \emptyset$, al no tener elementos en común esto implica que la ocurrencia de un evento garantiza la no ocurrencia del otro, y en tal caso se les denomina *eventos mutuamente excluyentes* y, por lo tanto, la probabilidad de su unión (que representa la ocurrencia de uno o el otro evento) se define como la suma de sus probabilidades individuales: $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$. Más aún, nótese que lo anterior implica también que Ω y \emptyset son elementos de \mathfrak{F} , y que si $E \in \mathfrak{F}$ entonces su complemento $\Omega \setminus E \in \mathfrak{F}$.¹

A partir de la Definición 1 se deducen algunas otras propiedades que cumple cualquier medida de probabilidad que se utilice, y que se resumen en la siguiente proposición, cuya demostración se deja como ejercicio de repaso:

Proposición 1 Sea $(\Omega, \mathfrak{F}, \mathbb{P})$ un espacio de probabilidad que representa a un fenómeno o experimento aleatorio. Entonces, además de la Definición 1, la medida de probabilidad \mathbb{P} cumple las siguientes propiedades, dados E_1 y E_2 subconjuntos de Ω que representan eventos:

- a) $\mathbb{P}(\Omega \setminus E_1) = 1 - \mathbb{P}(E_1)$.
- b) $\mathbb{P}(\emptyset) = 0$.
- c) $\mathbb{P}(E_1 \setminus E_2) = \mathbb{P}(E_1) - \mathbb{P}(E_1 \cap E_2)$.
- d) Si $E_1 \subseteq E_2$ entonces $\mathbb{P}(E_1) \leq \mathbb{P}(E_2)$.
- e) $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_2)$.
- f) $\max\{\mathbb{P}(E_1) + \mathbb{P}(E_2) - 1, 0\} \leq \mathbb{P}(E_1 \cap E_2) \leq \min\{\mathbb{P}(E_1), \mathbb{P}(E_2)\}$

□

Cuando un evento resulta representado por el conjunto vacío, y dado que $\mathbb{P}(\emptyset) = 0$, entonces se le denomina *evento imposible* ya que con toda certeza no ocurre. Cuando el conjunto que representa a un evento tiene un solo elemento de Ω se le denomina *evento simple*. En el caso particular de que el espacio muestral sea un conjunto numerable (finito o infinito), lo denotamos de forma general $\Omega = \{\omega_1, \omega_2, \dots\}$ y en tal caso notemos que Ω puede expresarse como una unión numerable de conjuntos disjuntos que representan eventos simples:

$$\Omega = \bigcup_{n \geq 1} \{\omega_n\} \quad (1)$$

¹De hecho, \mathfrak{F} cumple con la definición de σ -álgebra de subconjuntos de Ω , consúltense las referencias [3] y [6].

y por tanto cualquier subconjunto $E \subseteq \Omega$ también puede expresarse como unión numerable de conjuntos disjuntos de un solo elemento cada uno, esto es:

$$E = \bigcup_{\omega \in E} \{\omega\}$$

y aplicando la Definición 1 b se obtiene que:

$$\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(\{\omega\}) \quad (2)$$

es decir, que en el caso de espacios muestrales numerables basta poder calcular la probabilidad de los eventos simples, y por medio de (2) se puede calcular la probabilidad de cualquier evento no simple.

Ejemplo 2 Considere un espacio muestral finito $\Omega = \{\omega_1, \dots, \omega_n\}$ donde todos los eventos simples representados por $\{\omega_j\}$ para $j \in \{1, \dots, n\}$ tienen la misma probabilidad. A este caso particular se le denomina **espacio muestral equiprobable**. Como consecuencia de la Definición 1 a y de (1) se obtiene:

$$1 = \mathbb{P}(\Omega) = \sum_{j=1}^n \mathbb{P}(\{\omega_j\}) = n\mathbb{P}(\{\omega_1\})$$

y por la condición de equiprobabilidad de todos los eventos simples se tiene que cualquier evento simple tiene probabilidad $\mathbb{P}(\{\omega_j\}) = \frac{1}{n}$. Finalmente, como consecuencia de (2), se obtiene que la fórmula general para una medida de probabilidad \mathbb{P} en espacios muestrales (finitos) y equiprobables es:

$$\mathbb{P}(E) = \sum_{\omega \in E} \frac{1}{n} = \frac{|E|}{|\Omega|} \quad (3)$$

donde la notación $|E|$ se refiere al número de elementos del conjunto.

□

La medida de probabilidad se deduce o aproxima con base en la información inicial con que se cuenta en un momento dado respecto al fenómeno o experimento aleatorio en cuestión. Pero posterior a ello, puede surgir nueva información que implique una actualización de dicha medida. Por ejemplo, consideremos un fenómeno o experimento aleatorio representado inicialmente por un espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ así como dos eventos representados por los subconjuntos A y B en \mathfrak{F} . Supongamos que inicialmente las probabilidades de dichos eventos son $0 < \mathbb{P}(A) < 1$ y $0 < \mathbb{P}(B) < 1$. Si posteriormente ocurre el evento representado por B , entonces a la medida de probabilidad inicial \mathbb{P} habría que actualizarla y reemplazarla por una medida de probabilidad que denotaremos \mathbb{P}_B , de modo que $\mathbb{P}_B(B) = 1$, y poder actualizar la probabilidad $\mathbb{P}(A)$ mediante $\mathbb{P}_B(A)$ (misma que puede cambiar o permanecer igual), que se denomina *probabilidad condicional* de A dado B , y que debe cuantificar la incertidumbre del evento representado por A pero ahora en términos relativos respecto a B :

Definición 2 Dado un espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ y un evento representado por el subconjunto $B \in \mathfrak{F}$ tal que $0 < \mathbb{P}(B) < 1$, la actualización de la medida de probabilidad \mathbb{P} ante la ocurrencia del evento representado por B se denota \mathbb{P}_B y cuantifica la incertidumbre de cualquier evento $E \in \mathfrak{F}$ mediante:

$$\mathbb{P}_B(E) := \frac{\mathbb{P}(E \cap B)}{\mathbb{P}(B)}$$

misma que se denomina *probabilidad condicional* de E dado B , quedando el espacio de probabilidad actualizado como $(\Omega, \mathfrak{F}, \mathbb{P}_B)$.

□

Inicialmente los posibles resultados del fenómeno o experimento aleatorio están descritos por Ω , pero después de la ocurrencia de un evento representado por $B \subseteq \Omega$ el conjunto de posibles resultados se reduce justamente a B , y por ello es que la cuantificación de la incertidumbre de todos los eventos deben ahora actualizarse en términos relativos, es decir, un evento que inicialmente estaba representado por $E \subseteq \Omega$ ahora queda representado por $E \cap B \subseteq B$ ya que

cualquier resultado fuera de B queda descartado. La notación \mathbb{P}_B deja claro que ha ocurrido un cambio de medida de probabilidad respecto a la inicial \mathbb{P} . Sin embargo, es más común la notación $\mathbb{P}(E|B)$ en lugar de $\mathbb{P}_B(E)$.

Una vez actualizado un espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ a $(\Omega, \mathfrak{F}, \mathbb{P}_B)$ lo anterior no significa que necesariamente todas la probabilidades de los eventos cambiarán, en algunos casos sí, en otros no. Es decir, para cada evento $E \in \mathfrak{F}$ el valor de $\mathbb{P}(E)$ puede o no ser distinto de $\mathbb{P}_B(E)$. Cuando la probabilidad de un evento no cambia ante la ocurrencia de otro evento, estamos ante el concepto de independencia de un evento respecto a otro:

Definición 3 Dado un espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ y un evento representado por $B \in \mathfrak{F}$ tal que $0 < \mathbb{P}(B) < 1$ se dice que un evento representado por $E \in \mathfrak{F}$ es un **evento independiente** de B si:

$$\mathbb{P}_B(E) = \mathbb{P}(E).$$

□

Dados dos eventos representados por A y B en \mathfrak{F} tales que $0 < \mathbb{P}(A) < 1$ y $0 < \mathbb{P}(B) < 1$ tenemos entonces que A es independiente de B si:

$$\mathbb{P}(A) = \mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (4)$$

Similarmente, B es independiente de A si:

$$\mathbb{P}(B) = \mathbb{P}_A(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}. \quad (5)$$

Notemos que de (4) y (5) se deduce en ambos casos que se debe cumplir la condición:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \quad (6)$$

por lo que podemos simplemente referir que dos eventos son independientes si y solo la probabilidad de su conjunción coincide con el producto de sus probabilidades individuales. En caso de que $\mathbb{P}(A \cap B) \neq \mathbb{P}(A)\mathbb{P}(B)$ se dice que los eventos representados por A y B son **eventos dependientes**. Recordando que cuando dos eventos, representados por subconjuntos A y B , son mutuamente excluyentes entonces esto implica $A \cap B = \emptyset$, si $\mathbb{P}(A) > 0$ y $\mathbb{P}(B) > 0$ entonces necesariamente son eventos dependientes ya que $\mathbb{P}(A)\mathbb{P}(B) > 0 = \mathbb{P}(\emptyset) = \mathbb{P}(A \cap B)$.

Para extender el concepto de independencia a tres o más eventos, no basta que se cumpla que la probabilidad de conjunción de todos los eventos sea igual al producto de las probabilidades individuales de los mismos, pues es factible que ocurra esto último a pesar de existir dependencia, como se ilustra en el siguiente:

Ejemplo 3 Sea un espacio muestral equiprobable $\Omega = \{a, b, c, d, e, f, g, h\}$. Aplicando (3) obtenemos que la medida de probabilidad correspondiente para cualquier evento representado por algún subconjunto $E \subseteq \Omega$ es, en este caso, $\mathbb{P}(E) = \frac{|E|}{8}$. Considere tres eventos representados por los siguientes subconjuntos de Ω :

$$A = \{a, b, c, d\}, \quad B = \{a, e, f, g\}, \quad C = \{a, b, e, h\}.$$

Claramente $\mathbb{P}(A) = \frac{1}{2} = \mathbb{P}(B) = \mathbb{P}(C)$, y por tanto su producto $\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{8}$. Por otro lado, notemos que $A \cap B \cap C = \{a\}$ y por tanto $\mathbb{P}(A \cap B \cap C) = \frac{1}{8}$, es decir, en este ejemplo particular tenemos que:

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

Sin embargo, lo anterior no implica independencia de eventos. Por ejemplo, $A \cap B = \{a\}$ y por tanto:

$$\mathbb{P}(A \cap B) = \frac{1}{8} \neq \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B),$$

es decir, en este caso A y B **no** representan eventos independientes.

□

También puede ocurrir el caso en que exista independencia por pares de eventos, pero no así independencia de todos los eventos simultáneamente, como se ilustra en el siguiente:

Ejemplo 4 Nuevamente el espacio muestral equiprobable del Ejemplo 3 anterior, pero ahora consideremos tres eventos representados por los siguientes subconjuntos de Ω :

$$A = \{a, b, c, d\} \quad , \quad D = \{a, b, e, f\} \quad , \quad E = \{a, b, g, h\} .$$

Claramente $\mathbb{P}(A) = \frac{1}{2} = \mathbb{P}(D) = \mathbb{P}(E)$. Más aún:

$$A \cap D = \{a, b\} = A \cap E = D \cap E ,$$

por lo que se cumple:

$$\mathbb{P}(A \cap D) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(D) , \quad \mathbb{P}(A \cap E) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(E) , \quad \mathbb{P}(D \cap E) = \frac{1}{4} = \mathbb{P}(D)\mathbb{P}(E) .$$

Es decir, los eventos representados por A , D y E resultan ser independientes cuando solo se consideran dos de ellos a la vez, pero al ser considerados los tres simultáneamente se tiene que $A \cap D \cap E = \{a, b\}$ y por lo tanto:

$$\mathbb{P}(A \cap D \cap E) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(A)\mathbb{P}(D)\mathbb{P}(E) .$$

Notemos que, por ejemplo:

$$\mathbb{P}_E(A \cap D) = \frac{\mathbb{P}(A \cap D \cap E)}{\mathbb{P}(E)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2} \neq \frac{1}{4} = \mathbb{P}(A \cap D) ,$$

esto es, la ocurrencia del evento representado por E modifica la probabilidad de ocurrencia del evento representado por $A \cap D$, a pesar de que A y D representan eventos independientes entre sí.

□

Para poder decir que una colección finita de eventos son independientes se requiere que lo sean por pares, por ternas, por cuartetos, etc. Esto es:

Definición 4 Dado un espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ y n eventos representados por E_1, \dots, E_n elementos de \mathfrak{F} , se les considera **eventos conjuntamente independientes**² si cualquier subcolección de ellos E_{i_1}, \dots, E_{i_m} , donde los subíndices i_k son números enteros positivos tales que $1 \leq i_1 < \dots < i_m \leq n$, se satisface lo siguiente:

$$\mathbb{P}\left(\bigcap_{k=1}^m A_{i_k}\right) = \prod_{k=1}^m \mathbb{P}(A_{i_k}) .$$

□

En los ejemplos 3 y 4 se ilustran casos en que los eventos considerados no son conjuntamente independientes, de acuerdo con la definición anterior. A continuación un ejemplo donde esto sí se cumple:

Ejemplo 5 Nuevamente el espacio muestral equiprobable del Ejemplo 3, pero ahora consideremos tres eventos representados por los siguientes subconjuntos de Ω :

$$E_1 = \{a, b, c, d\} \quad , \quad E_2 = \{a, b, e, f\} \quad , \quad E_3 = \{a, c, e, g\} ,$$

donde claramente $\mathbb{P}(E_j) = \frac{1}{2}$ para cada $j \in \{1, 2, 3\}$. Como en este caso se tienen $n = 3$ eventos en consideración, de acuerdo con la Definición 4 los casos no triviales a considerar corresponden a $m \in \{2, 3\}$. Para el caso $m = 2$ las posibilidades de subíndices son $(i_1, i_2) \in \{(1, 2), (1, 3), (2, 3)\}$ obteniendo lo siguiente:

$$E_1 \cap E_2 = \{a, b\} \quad \Rightarrow \quad \mathbb{P}(E_1 \cap E_2) = \frac{1}{4} = \mathbb{P}(E_1)\mathbb{P}(E_2) ,$$

$$E_1 \cap E_3 = \{a, c\} \quad \Rightarrow \quad \mathbb{P}(E_1 \cap E_3) = \frac{1}{4} = \mathbb{P}(E_1)\mathbb{P}(E_3) ,$$

$$E_2 \cap E_3 = \{a, e\} \quad \Rightarrow \quad \mathbb{P}(E_2 \cap E_3) = \frac{1}{4} = \mathbb{P}(E_2)\mathbb{P}(E_3) ,$$

²Usualmente se omite la palabra *conjuntamente*.

con lo que ya se tiene, hasta el momento, independencia por pares de eventos. Ahora falta el caso simultáneo $m = 3$, que en este caso la única posibilidad de subíndices es $(i_1, i_2, i_3) = (1, 2, 3)$ por lo que basta analizar el caso:

$$E_1 \cap E_2 \cap E_3 = \{a\} \Rightarrow \mathbb{P}(E_1 \cap E_2 \cap E_3) = \frac{1}{8} = \mathbb{P}(E_1)\mathbb{P}(E_2)\mathbb{P}(E_3),$$

con lo que puede concluirse que los tres eventos representados por E_1 , E_2 y E_3 son conjuntamente independientes. □

Si A y B son subconjuntos de Ω que representan eventos independientes, entonces los eventos representados por A y $\Omega \setminus B$ (complemento de B respecto a Ω) también son independientes ya que, como consecuencia de los incisos a) y c) de la Proposición 1 y de (6), se deduce que:

$$\begin{aligned} \mathbb{P}(A \cap (\Omega \setminus B)) &= \mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)[1 - \mathbb{P}(B)] \\ &= \mathbb{P}(A)\mathbb{P}(\Omega \setminus B), \end{aligned}$$

y de forma análoga se demuestra que los eventos representados por $\Omega \setminus A$ y $\Omega \setminus B$ también resultan ser independientes. Más aún, lo anterior puede generalizarse en la siguiente proposición, cuya demostración se deja como ejercicio:

Proposición 2 Sea $(\Omega, \mathfrak{F}, \mathbb{P})$ un espacio de probabilidad, y sean E_1, \dots, E_n subconjuntos de Ω que representan eventos conjuntamente independientes. Entonces los eventos representados por los subconjuntos A_1, \dots, A_n donde cada $A_j \in \{E_j, \Omega \setminus E_j\}$ también son conjuntamente independientes. □

Ejemplo 6 Experimento aleatorio binomial. Este tipo de experimento consiste en $n \in \{1, 2, \dots\}$ repeticiones de un procedimiento que en cada repetición solo arroja uno de dos resultados posibles, que denominaremos “éxito” (e) y “fracaso” (f). La probabilidad de éxito en cada repetición es un valor constante $0 < \theta < 1$, y el resultado de cualquier repetición no influye en las otras. En este caso, cualquier resultado de este experimento aleatorio consiste en una n -ada de éxitos y fracasos, por lo que el espacio muestral puede denotarse mediante:

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_j \in \{e, f\}\}.$$

Se trata de un espacio muestral finito donde $|\Omega| = 2^n$, pero no es en general equiprobable, salvo el caso particular $\theta = \frac{1}{2}$. Por tratarse de un espacio muestral numerable, para definir la medida de probabilidad correspondiente $\mathbb{P} : \mathfrak{F} \rightarrow [0, 1]$ para cualquier $E \in \mathfrak{F}$, de acuerdo con (2) basta poder deducir la probabilidad correspondiente a los eventos simples, esto es $\mathbb{P}(\{\omega_1, \dots, \omega_n\})$, con base en la descripción de este experimento aleatorio. Consideremos los eventos $\mathcal{E}_1, \dots, \mathcal{E}_n$ definidos como:

$$\mathcal{E}_k \equiv \text{“Se obtiene éxito en la } k\text{-ésima repetición”}, \quad k \in \{1, \dots, n\},$$

los cuales quedan representados por los subconjuntos de Ω siguientes:

$$\begin{aligned} E_1 &= \{(e, \omega_2, \omega_3, \dots, \omega_n) : \omega_j \in \{e, f\}\}, \\ E_2 &= \{(\omega_1, e, \omega_3, \dots, \omega_n) : \omega_j \in \{e, f\}\}, \\ &\vdots \\ E_n &= \{(\omega_1, \dots, \omega_{n-1}, e) : \omega_j \in \{e, f\}\}. \end{aligned}$$

Cada uno de estos subconjuntos tiene $|E_k| = 2^{n-1}$ elementos, por lo que con $n \geq 2$ no representan eventos simples, pero la descripción del experimento nos proporciona información sobre la probabilidad que les corresponde: $\mathbb{P}(E_k) = \theta$ para todo $k \in \{1, \dots, n\}$. Más aún, notemos que cualquier evento simple puede expresarse en términos de una conjunción de los eventos $\mathcal{E}_1, \dots, \mathcal{E}_n$ o algunas de sus negaciones. Por ejemplo, para un caso particular $n = 5$, el evento simple representado por $\{(e, f, e, e, f)\}$ es resultado de la siguiente conjunción de eventos:

$$\mathcal{E}_1 \wedge (\neg \mathcal{E}_2) \wedge \mathcal{E}_3 \wedge \mathcal{E}_4 \wedge (\neg \mathcal{E}_5) \longrightarrow E_1 \cap (\Omega \setminus E_2) \cap E_3 \cap E_4 \cap (\Omega \setminus E_5) = \{(e, f, e, e, f)\},$$

y también por la descripción del experimento, en el sentido de que el resultado de cualquier repetición no influye en las otras repeticiones, esto implica la independencia de los eventos $\mathcal{E}_1, \dots, \mathcal{E}_n$ y por tanto:

$$\begin{aligned}\mathbb{P}(\{(e, f, e, e, f)\}) &= \mathbb{P}(E_1 \cap (\Omega \setminus E_2) \cap E_3 \cap E_4 \cap (\Omega \setminus E_5)) \\ &= \mathbb{P}(E_1)\mathbb{P}(\Omega \setminus E_2)\mathbb{P}(E_3)\mathbb{P}(E_4)\mathbb{P}(\Omega \setminus E_5) \\ &= \theta \cdot (1 - \theta) \cdot \theta \cdot \theta \cdot (1 - \theta) = \theta^3(1 - \theta)^2,\end{aligned}$$

y en general, para cualquier valor de n la fórmula general para cualquier n -ada que representa un evento simple:

$$\mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) = \theta^r(1 - \theta)^{n-r}, \quad r = \sum_{j=1}^n \mathbf{1}_{\{\omega_j = e\}}, \quad (7)$$

donde r representa el número de éxitos (e) en la n -ada $(\omega_1, \dots, \omega_n)$, y por tanto necesariamente $n - r$ es el número de fracasos (f). Al tener una fórmula general (7) para los eventos simples en un espacio muestral numerable, la probabilidad de cualquier evento no simple se calcula mediante (2), y con esto el espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ para el experimento aleatorio binomial ha quedado especificado. □

Una vez especificado el espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ con base en la información que se tiene, en un momento dado, sobre un fenómeno o experimento aleatorio, es posible definir distintas funciones que cuantifiquen aspectos de interés. Se trata de funciones que tienen por dominio al espacio muestral Ω y por codominio al conjunto de los números reales \mathbb{R} , y cuya regla de correspondencia se define de acuerdo con un interés específico. A estas funciones se les conoce como *variables aleatorias*. En el caso del experimento aleatorio binomial, cuyo espacio de probabilidad se dedujo en el Ejemplo 6, podría ser de interés, por ejemplo, cuantificar el número de éxitos que se obtienen en las n repeticiones, por medio de una función $X : \Omega \rightarrow \mathbb{R}$ definida como:

$$X((\omega_1, \dots, \omega_n)) := \sum_{j=1}^n \mathbf{1}_{\{\omega_j = e\}}, \quad (8)$$

o bien una variable aleatoria $Y : \Omega \rightarrow \mathbb{R}$ que reporte el valor 1 si el resultado de las n repeticiones fue siempre el mismo, y 0 en caso contrario:

$$Y((\omega_1, \dots, \omega_n)) := \mathbf{1}_{\{\omega_1 = \omega_2 = \dots = \omega_n\}}. \quad (9)$$

Una vez definida una variable aleatoria X sobre un espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$, dado que existe incertidumbre sobre el resultado $\omega \in \Omega$ que se puede observar, esto se traduce en incertidumbre también sobre el número real $X(\omega)$ que puede reportar la variable aleatoria X , y por tanto interesa hacer cálculo de probabilidades sobre los valores que pueda reportar dicha variable aleatoria, ya sea en relación con un valor específico o bien sobre un conjunto de valores. Esto es, calcular la probabilidad de que $X(\omega) = x$ para un valor específico x , o bien la probabilidad de que $X(\omega) \in B$ para algún subconjunto $B \subseteq \mathbb{R}$.

Definición 5 Dado un espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ una **variable aleatoria** es una función $X : \Omega \rightarrow \mathbb{R}$ tal que para casi cualquier $B \subseteq \mathbb{R}$ la imagen inversa $X^{(-1)}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ representa a algún evento, es decir tal que $X^{(-1)}(B) \in \mathfrak{F}$. □

El uso de la expresión *casi cualquier* $B \subseteq \mathbb{R}$ se refiere a que existen subconjuntos de \mathbb{R} cuya imagen inversa no representa evento alguno, en *teoría de la medida* se les conoce como subconjuntos *no medibles*, probabilísticamente hablando en este caso particular. No abundaremos más en ello porque eso está más allá del objetivo del presente texto, así que de momento nos restringiremos a aquellos subconjuntos de \mathbb{R} que sí son probabilísticamente medibles, colección que se conoce como σ -álgebra de Borel³ $\mathfrak{B}(\mathbb{R})$, y a cuyos elementos se les conoce como *conjuntos borelianos*, que para efectos prácticos resulta más que suficiente para el cálculo de probabilidades que abordaremos.

Al definir una variable aleatoria X sobre un espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ se crea un nuevo espacio de probabilidad $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \mathbb{P}_X)$ donde $\mathbb{P}_X : \mathfrak{B}(\mathbb{R}) \rightarrow [0, 1]$ es la *medida de probabilidad inducida* por la variable aleatoria X , de

³Émile Borel (1871–1956) matemático y político francés.

modo que para cualquier $B \in \mathfrak{B}(\mathbb{R})$ el número real $\mathbb{P}_X(B)$ represente justamente la probabilidad de que la variable aleatoria X reporte un número real $X(\omega)$ que pertenezca a B . Esto último ocurrirá si el resultado del fenómeno o experimento aleatorio es alguno de los elementos del subconjunto $\{\omega \in \Omega : X(\omega) \in B\} = X^{(-1)}(B) \in \mathfrak{F}$, que al representar un evento es posible calcular su probabilidad mediante \mathbb{P} , es decir:

$$\mathbb{P}_X(B) = \mathbb{P}(X^{(-1)}(B)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}). \quad (10)$$

En lugar de escribir $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$ algunos autores acostumbran escribir simplemente $\mathbb{P}(X \in B)$. Se deja como ejercicio demostrar que \mathbb{P}_X cumple con la Definición 1 con $\Omega = \mathbb{R}$.

Ejemplo 7 Variable aleatoria binomial. Continuando con el experimento aleatorio binomial, cuyo espacio de probabilidad fue obtenido en el Ejemplo 6, consideremos la variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ definida en (8) que reporta el número de éxitos obtenido en las n repeticiones. Esto implica que el conjunto de valores distintos que puede reportar es el rango de dicha función, es decir:

$$\text{Ran } X = \{X(\omega) : \omega \in \Omega\} = \{0, 1, \dots, n\}.$$

Entonces para cualquier $x \in \{0, 1, \dots, n\}$ la probabilidad de que la variable aleatoria X reporte el valor $X(\omega) = x$ se puede calcular mediante (10) con $B = \{x\}$. Obtengamos primero su imagen inversa, que denotaremos:

$$E_x := X^{(-1)}(\{x\}) = \left\{ (\omega_1, \dots, \omega_n) \in \Omega : \sum_{j=1}^n \mathbf{1}_{\{\omega_j = e\}} = x \right\}.$$

Notemos que todos los elementos $(\omega_1, \dots, \omega_n) \in E_x$ son n -adas con exactamente x éxitos, y por tanto $n - x$ fracasos, así que aplicando (7) obtenemos que $\mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) = \theta^x(1 - \theta)^{n-x}$ para todo $(\omega_1, \dots, \omega_n) \in E_x$, y aplicando (2) obtenemos:

$$\mathbb{P}_X(\{x\}) = \mathbb{P}(E_x) = \sum_{(\omega_1, \dots, \omega_n) \in E_x} \mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) = |E_x| \theta^x (1 - \theta)^{n-x},$$

donde $|E_x|$ es el número de formas distintas en las que podemos acomodar x éxitos en los n lugares de cualquier n -ada $(\omega_1, \dots, \omega_n) \in E_x$ (los lugares restantes son automáticamente para los $n - x$ fracasos), y esto es justamente el número de combinaciones de n en x , denotado por el coeficiente binomial $\binom{n}{x} = \frac{n!}{x!(n-x)!}$, con lo que concluimos que la fórmula general para calcular la probabilidad de que la variable aleatoria X reporte exactamente x éxitos en n repeticiones, para cualquier $x \in \{0, 1, \dots, n\}$, es la siguiente:

$$\mathbb{P}_X(\{x\}) \equiv \mathbb{P}(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \mathbf{1}_{\{0, 1, \dots, n\}}(x). \quad (11)$$

Implícitamente $\mathbb{P}(X = x) = 0$ para toda $x \notin \{0, 1, \dots, n\}$ por lo que para cualquier boreliano $B \in \mathfrak{B}(\mathbb{R})$ tenemos que:

$$\mathbb{P}(X \in B) \equiv \mathbb{P}_X(B) = \mathbb{P}_X(B \cap \{0, 1, \dots, n\}) = \sum_{x=0}^n \mathbb{P}(X = x) \mathbf{1}_{\{x \in B\}}. \quad (12)$$

□

El cálculo de probabilidades con variables aleatorias $\mathbb{P}(X \in B) \equiv \mathbb{P}_X(B)$ más frecuente es aquél en el que B representa un intervalo en \mathbb{R} , que de forma general podríamos denotar como $B =]a, b[$ donde $-\infty < a \leq b < +\infty$, y donde el símbolo $|$ puede denotar abierto o cerrado. Por ejemplo, los intervalos acotados $[a, b]$, $]a, b[$, $]a, b]$, $[a, b[$, $[a, a] = \{a\}$, $]a, a[= \emptyset$, o bien intervalos no acotados como $] - \infty, b]$, $] - \infty, b[$, $[a, +\infty[$, $]a, +\infty[$, $] - \infty, +\infty[= \mathbb{R}$. Teniendo resuelto cómo calcular probabilidades de que una variable aleatoria X reporte un valor que pertenezca a este tipo de intervalos, el cálculo de probabilidades para uniones disjuntas numerables de intervalos es inmediato a partir de la Definición 1, esto es, si I_1, I_2, \dots son intervalos en \mathbb{R} tales que $I_j \cap I_k = \emptyset$ para todo $j \neq k$ entonces:

$$\mathbb{P}\left(X \in \bigcup_{n \geq 1} I_n\right) \equiv \mathbb{P}_X\left(\bigcup_{n \geq 1} I_n\right) = \sum_{n \geq 1} \mathbb{P}_X(I_n) \equiv \sum_{n \geq 1} \mathbb{P}(X \in I_n). \quad (13)$$

Todos los tipos de intervalos mencionados en el párrafo anterior pueden obtenerse mediante una cantidad numerable de operaciones de conjuntos de un solo tipo de intervalo, por ejemplo los intervalos del tipo $] - \infty, x]$ donde $x \in \mathbb{R}$.

Por ejemplo, $]a, b] =] - \infty, b] \setminus] - \infty, a]$ y por lo tanto aplicando propiedades de las medidas de probabilidad según la Definición 1 y la Proposición 1 y lo que de ellas se puede derivar, obtenemos:

$$\mathbb{P}(X \in]a, b]) \equiv \mathbb{P}_X(]a, b]) = \mathbb{P}_X(] - \infty, b] \setminus] - \infty, a]) = \mathbb{P}_X(] - \infty, b]) - \mathbb{P}_X(] - \infty, a]). \quad (14)$$

La demostración de los otros casos queda como ejercicio. La relevancia de lo anterior es que entonces basta saber calcular $\mathbb{P}_X(] - \infty, x])$ para cualquier $x \in \mathbb{R}$ y con ello será posible calcular la probabilidad de que la variable aleatoria X reporte un valor $X(\omega)$ que pertenezca a cualquier otro tipo de intervalo. Esto motiva la siguiente:

Definición 6 *La función de distribución de probabilidades de una variable aleatoria X es una función $F_X : \mathbb{R} \rightarrow [0, 1]$ definida como:*

$$F_X(x) := \mathbb{P}_X(] - \infty, x]) \equiv \mathbb{P}(X \leq x).$$

□

Consecuencia de la definición anterior, y de las propiedades de la medida de probabilidad P_X inducida por una variable aleatoria X , se obtiene el siguiente teorema, cuya demostración se deja como ejercicio:

Teorema 1 *Sea F_X una función de distribución de probabilidades de una variable aleatoria X .*

- a) F_X es una función monótona creciente, esto es que si $x_1 < x_2$ entonces $F_X(x_1) \leq F_X(x_2)$.
- b) F_X es continua por la derecha, es decir que $\lim_{\delta \rightarrow 0+} F_X(x + \delta) = F_X(x)$.
- c) $F_X(+\infty) \equiv \lim_{x \rightarrow +\infty} F_X(x) = 1$.
- d) $F_X(-\infty) \equiv \lim_{x \rightarrow -\infty} F_X(x) = 0$.
- e) $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$.
- f) $\mathbb{P}(X < a) = F_X(a-) \equiv \lim_{x \rightarrow a-} F_X(x)$.
- g) $\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a-)$.
- h) $\mathbb{P}(X > b) = 1 - F_X(b)$.
- i) $\mathbb{P}(X \geq b) = 1 - F_X(b-)$.
- j) $\mathbb{P}(a < X < b) = F_X(b-) - F_X(a)$.
- k) $\mathbb{P}(a \leq X < b) = F_X(b-) - F_X(a-)$.
- l) $\mathbb{P}(X = x) = F_X(x) - F_X(x-)$.

□

De hecho, si $F : \mathbb{R} \rightarrow [0, 1]$ es cualquier función que satisface las propiedades a), b), c) y d) del Teorema 1, se puede demostrar que siempre existe una variable aleatoria que tiene a F como función de distribución de probabilidades, y los incisos e) a l) son consecuencia de los cuatro primeros, demostración que puede consultarse en la referencia [3].

A las variables aleatorias se les puede clasificar en tres categorías, dependiendo de algunas propiedades particulares adicionales de sus funciones de distribución de probabilidades. Una primera categoría corresponde al caso en que $\mathbb{P}(X = x) = 0$ para todo valor $x \in \mathbb{R}$, lo cual implicaría que F_X es continua también por la izquierda, y por tanto continua en todo \mathbb{R} , véase la Figura 1a, razón por la cual en este caso particular se dice que X es una **variable aleatoria continua**. Consecuencia inmediata de lo anterior:

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b), \quad (15)$$

$$\mathbb{P}(X > b) = \mathbb{P}(X \geq b), \quad \mathbb{P}(X < a) = \mathbb{P}(X \leq a). \quad (16)$$

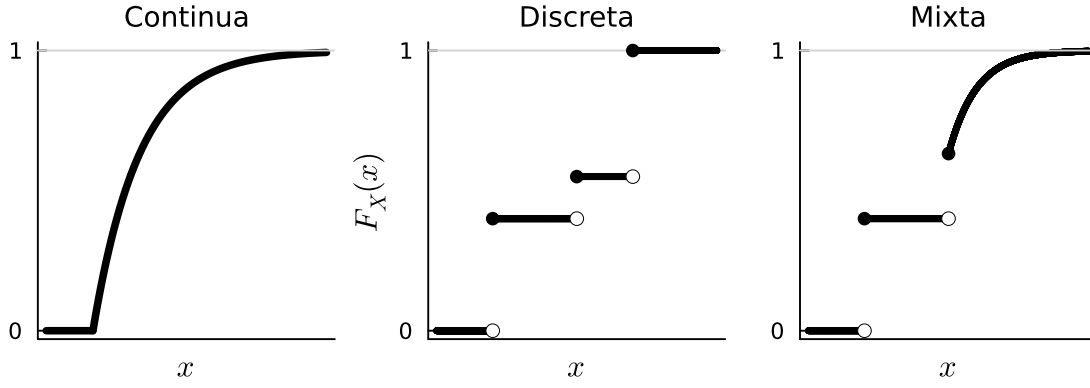


Figura 1: Tipos de funciones de distribución de probabilidades: a) continua, b) discreta, c) mixta.

Definición 7 Sea $|a, b|$ un intervalo cualquiera, abierto, cerrado o semi abierto, acotado o no, es decir $-\infty \leq a \leq b \leq +\infty$. Se dice que X es una **variable aleatoria absolutamente continua** si existe una función integrable $f_X : \mathbb{R} \rightarrow [0, +\infty[$ tal que:

$$\mathbb{P}(X \in |a, b|) = \int_a^b f_X(x) dx,$$

y a la función f_X se le denomina **función de densidad de probabilidades**.

□

Consecuencia inmediata de las Definiciones 6 y 7, en el caso de que X sea una variable aleatoria absolutamente continua se deduce la siguiente relación entre su función de distribución de probabilidades F_X y su función de densidad de probabilidades f_X :

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \in]-\infty, x]) = \int_{-\infty}^x f_X(t) dt, \quad \text{para todo valor } x \in \mathbb{R}, \quad (17)$$

y como $F_X(+\infty) = 1$ esto implica que necesariamente $\int_{-\infty}^{+\infty} f_X(x) dx = 1$. Más aún, en este caso F_X no solo es continua sino que además, como consecuencia del *Teorema Fundamental del Cálculo*, F_X es derivable en cada punto $x \in \mathbb{R}$ donde f_X sea continua, y en tales casos:

$$\frac{d}{dx} F_X(x) = f_X(x). \quad (18)$$

Ejemplo 8 Variable aleatoria continua uniforme. Sea X una variable aleatoria absolutamente continua con $\text{Ran } X = |a, b|$ un intervalo acotado y tal que subintervalos de $|a, b|$ de igual longitud tengan la misma probabilidad, es decir, para cualesquiera números reales c y d tales que $a < c < d < d + h < b$ con $h > 0$, se cumple que:

$$\mathbb{P}(X \in |c, c + h|) = \mathbb{P}(X \in |d, d + h|),$$

y por lo tanto:

$$\frac{F_X(c + h) - F_X(c)}{h} = \frac{F_X(d + h) - F_X(d)}{h},$$

en donde calculando límites cuando $h \rightarrow 0+$ se obtiene que $f_X(c) = f_X(d)$, lo que implica que f_X es una función constante sobre el intervalo $|a, b|$. Sea $k > 0$ dicha constante, cuyo valor se deduce a partir de lo siguiente:

$$1 = \int_{-\infty}^{+\infty} f_X(x) dx = \int_a^b k dx = (b - a)k,$$

es decir que $k = \frac{1}{b-a}$, por lo que la función de densidad de probabilidades resulta ser:

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

y su función de distribución de probabilidades:

$$F_X(x) = \int_{-\infty}^x f_X(t)dt = \begin{cases} 0, & \text{si } x \leq a, \\ \frac{x-a}{b-a}, & \text{si } a < x < b, \\ 1, & \text{si } x \geq b. \end{cases}$$

Nótese que F_X es continua en todo \mathbb{R} y derivable en cualquier punto excepto en $x = a$ y en $x = b$, ya que en dichos puntos f_X es discontinua. □

En caso de exista al menos un punto x_0 tal que $F_X(x_0-) \neq F_X(x_0)$, además de ser un punto de discontinuidad de F_X , también como consecuencia de los incisos a) y l) del Teorema 1 se obtiene que $P(X = x_0) > 0$. Una segunda categoría de variables aleatorias es aquella en la que una variable aleatoria X solo puede reportar una cantidad numerable de valores distintos, esto es $\text{Ran } X = \{x_1, x_2, \dots\}$, en cuyo caso se trata de una **variable aleatoria discreta**. Como:

$$\Omega = X^{(-1)}(\text{Ran } X) = \bigcup_{n \geq 1} X^{(-1)}(\{x_n\})$$

donde la unión es disjunta, entonces:

$$1 = \mathbb{P}(\Omega) = \sum_{n \geq 1} \mathbb{P}(X^{(-1)}(\{x_n\})) \equiv \sum_{n \geq 1} \mathbb{P}(X = x_n).$$

Lo anterior implica que en el caso de una variable aleatoria discreta X basta saber calcular $\mathbb{P}(X = x) > 0$ para cada $x \in \text{Ran } X = \{x_1, x_2, \dots\}$, porque para cualquier conjunto boreliano $B \in \mathfrak{B}(\mathbb{R})$ tenemos que:

$$\mathbb{P}(X \in B) \equiv \mathbb{P}_X(B) = \mathbb{P}_X(B \cap \{x_1, x_2, \dots\}) = \sum_{n \geq 1} \mathbb{P}(X = x_n) \mathbf{1}_{\{x_n \in B\}} \quad (19)$$

lo cual motiva la siguiente:

Definición 8 La **función de masa de probabilidades** de una variable aleatoria discreta X es una función $p_X : \mathbb{R} \rightarrow [0, 1]$ definida como:

$$p_X(x) := \mathbb{P}(X = x)$$

y donde $p_X(x) > 0$ si y solo si $x \in \text{Ran } X = \{x_1, x_2, \dots\}$. □

De la definición anterior tenemos entonces que, en general, para cualquier conjunto boreliano $B \in \mathfrak{B}(\mathbb{R})$ el cálculo de probabilidades con una variable aleatoria discreta es:

$$\mathbb{P}(X \in B) = \sum_{x \in B} p_X(x), \quad (20)$$

y en particular, si $B =] - \infty, t]$ para algún $t \in \mathbb{R}$ entonces:

$$F_X(t) = \sum_{x: x \leq t} p_X(x). \quad (21)$$

Nótese que la función de distribución de probabilidades $F_X : \mathbb{R} \rightarrow [0, 1]$ de una variable aleatoria discreta es una *función escalonada*, esto es, que siempre es constante por intervalos, y con saltos en cada valor $x \in \text{Ran } X = \{x_1, x_2, \dots\}$, véase la Figura 1b. El Ejemplo 7 es un caso de variable aleatoria discreta, cuya función de masa de probabilidades está dada por (11).

La tercera y última categoría a considerar es el caso la **variable aleatoria mixta**, que debe su nombre al hecho de que mezcla el comportamiento de una variable aleatoria discreta y una variable aleatoria continua. Esto es, su función de distribución de probabilidades no es continua pero tampoco es siempre constante por intervalos, véase por ejemplo la Figura 1c. Primero recordemos el siguiente resultado, cuya demostración se deja como ejercicio:

Proposición 3 Sean F_1 y F_2 dos funciones de distribución de probabilidades cualesquiera. Si para cualquier número real $0 \leq \alpha \leq 1$ se define una función $F : \mathbb{R} \rightarrow [0, 1]$ como la combinación lineal convexa de F_1 y F_2 , esto es:

$$F(x) := (1 - \alpha)F_1(x) + \alpha F_2(x),$$

entonces F cumple las propiedades de una función de distribución de probabilidades como en el Teorema 1.

□

El siguiente *teorema de descomposición* para variables aleatorias demuestra que la función de distribución de probabilidades de cualquier tipo de variable aleatoria siempre puede expresarse (descomponerse) como combinación lineal convexa de una función de distribución de probabilidades discreta y otra continua:

Teorema 2 Sea X una variable aleatoria y F_X su función de distribución de probabilidades. Entonces existen un número real $0 \leq \alpha \leq 1$, una función de distribución de probabilidades discreta F_D y una función de distribución de probabilidades continua F_C tales que:

$$F_X(x) = (1 - \alpha)F_D(x) + \alpha F_C(x).$$

□

La demostración del teorema anterior puede consultarse, por ejemplo, en la referencia [3].

1. Distribución conjunta y marginales

El concepto de variable aleatoria se generaliza al de *vector aleatorio* definiendo funciones que tienen por dominio al espacio muestral Ω y por codominio al conjunto \mathbb{R}^n para cualquier entero positivo $n \geq 2$.

Definición 9 Dado un espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ un *vector aleatorio* es una función $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$ tal que para casi cualquier $B \subseteq \mathbb{R}^n$ la imagen inversa $\mathbf{X}^{(-1)}(B) = \{\omega \in \Omega : \mathbf{X}(\omega) \in B\}$ representa a algún evento, es decir tal que $\mathbf{X}^{(-1)}(B) \in \mathfrak{F}$.

□

El uso de la expresión *casi cualquier* $B \subseteq \mathbb{R}^n$ es análogo a lo explicado respecto a la Definición 5.

Al definir un vector aleatorio \mathbf{X} sobre un espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ se crea un nuevo espacio de probabilidad $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n), \mathbb{P}_{\mathbf{X}})$ donde $\mathbb{P}_{\mathbf{X}} : \mathfrak{B}(\mathbb{R}^n) \rightarrow [0, 1]$ es la *medida de probabilidad inducida* por el vector aleatorio \mathbf{X} , de modo que para cualquier $B \in \mathfrak{B}(\mathbb{R}^n)$ el número real $\mathbb{P}_{\mathbf{X}}(B)$ represente justamente la probabilidad de que el vector aleatorio \mathbf{X} reporte un punto en \mathbb{R}^n que pertenezca a B . Esto último ocurrirá si el resultado del fenómeno o experimento aleatorio es alguno de los elementos del subconjunto $\{\omega \in \Omega : \mathbf{X}(\omega) \in B\} = \mathbf{X}^{(-1)}(B) \in \mathfrak{F}$, que al representar un evento es posible calcular su probabilidad mediante \mathbb{P} , es decir:

$$\mathbb{P}_{\mathbf{X}}(B) = \mathbb{P}(\mathbf{X}^{(-1)}(B)) = \mathbb{P}(\{\omega \in \Omega : \mathbf{X}(\omega) \in B\}). \quad (22)$$

En lugar de escribir $\mathbb{P}(\{\omega \in \Omega : \mathbf{X}(\omega) \in B\})$ algunos autores acostumbran escribir simplemente $\mathbb{P}(\mathbf{X} \in B)$. Se deja como ejercicio demostrar que $\mathbb{P}_{\mathbf{X}}$ cumple con la Definición 1 con $\Omega = \mathbb{R}^n$.

Para cada $\omega \in \Omega$ tenemos que $\mathbf{X}(\omega)$ es un punto en \mathbb{R}^n y por tanto existen n números reales tales que $\mathbf{X}(\omega) = (x_1, \dots, x_n)$. Más aún, al tratarse de una función $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$ existen n funciones componentes X_1, \dots, X_n tales que $X_j : \Omega \rightarrow \mathbb{R}$ para $j \in \{1, \dots, n\}$ de modo que:

$$\mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega)). \quad (23)$$

La demostración de lo siguiente se deja como ejercicio:

Proposición 4 Sea $(\Omega, \mathfrak{F}, \mathbb{P})$ un espacio de probabilidad. Una función $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$ es vector aleatorio si y solo si cada una de sus funciones componentes (23) son variables aleatorias.

□

Así como cualquier subconjunto $B \in \mathfrak{B}(\mathbb{R})$ puede obtenerse en términos de una cantidad numerable de operaciones de conjuntos con intervalos de la forma $] -\infty, x]$, de forma análoga cualquier subconjunto $B \in \mathfrak{B}(\mathbb{R}^n)$ con $n \geq 2$ puede obtenerse en términos de una cantidad numerable de operaciones de conjuntos con productos cartesianos de intervalos de la forma $] -\infty, x]$ también, es decir con subconjuntos de la clase:

$$\mathfrak{J}^n := \{] -\infty, x_1] \times \dots \times] -\infty, x_n] : (x_1, \dots, x_n) \in \mathbb{R}^n \}, \quad (24)$$

lo que implica que para definir la medida de probabilidad $\mathbb{P}_{\mathbf{X}}$ inducida por un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ basta que dicha medida esté definida sobre la clase \mathfrak{J}^n , es decir poder calcular para cualquier $(x_1, \dots, x_n) \in \mathbb{R}^n$:

$$\begin{aligned} \mathbb{P}_{\mathbf{X}}(] -\infty, x_1] \times \dots \times] -\infty, x_n]) &= \mathbb{P}(\mathbf{X}^{(-1)}(] -\infty, x_1] \times \dots \times] -\infty, x_n])) \\ &= \mathbb{P}(\{ \omega \in \Omega : (X_1(\omega), \dots, X_n(\omega)) \in] -\infty, x_1] \times \dots \times] -\infty, x_n] \}) \\ &= \mathbb{P}(\{ \omega \in \Omega : X_1(\omega) \leq x_1 \wedge \dots \wedge X_n(\omega) \leq x_n \}) \\ &= \mathbb{P}(\{ \omega \in \Omega : X_1(\omega) \leq x_1 \} \cap \dots \cap \{ \omega \in \Omega : X_n(\omega) \leq x_n \}) \end{aligned} \quad (25)$$

en donde es costumbre escribir (25) de forma simplificada como $\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$, lo cual motiva la siguiente definición, que resulta análoga a la Definición 6:

Definición 10 La *función de distribución conjunta de probabilidades* de un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ es una función $F_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, 1]$ definida como:

$$F_{\mathbf{X}}(x_1, \dots, x_n) := \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n).$$

□

A partir de la función de distribución conjunta de probabilidades $F_{\mathbf{X}}$ de un vector aleatorio de probabilidades $\mathbf{X} = (X_1, \dots, X_n)$ es posible deducir las funciones de distribución de probabilidades individuales F_{X_j} de cada una de las $j \in \{1, \dots, n\}$ variables aleatorias componentes. A cada F_{X_j} se le denomina *función de distribución marginal* y al procedimiento para deducirla a partir de la función de distribución conjunta $F_{\mathbf{X}}$ se le denomina *marginalización*:

$$\begin{aligned} F_{X_j}(x_j) &= \mathbb{P}(X_j \leq x_j) = \mathbb{P}(\{\omega \in \Omega : X_j(\omega) \leq x_j\}) \\ &= \mathbb{P}(D_1 \cap \dots \cap D_{j-1} \cap \{\omega \in \Omega : X_j(\omega) \leq x_j\} \cap D_{j+1} \cap \dots \cap D_n) \end{aligned}$$

donde $D_k = \{\omega \in \Omega : X_k(\omega) \leq +\infty\} = \Omega$ para $k \neq j$ y por lo tanto:

$$F_{X_j}(x_j) = F_{\mathbf{X}}(+\infty, \dots, +\infty, x_j, +\infty, \dots, +\infty). \quad (26)$$

Sin embargo, a partir de las funciones de distribución marginales F_{X_1}, \dots, F_{X_n} no es posible deducir la función de distribución conjunta del vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ ya que pueden existir distintas funciones de distribución conjunta que tengan esas mismas funciones de distribución marginales, como se ilustra en el siguiente:

Ejemplo 9 Considérese el caso $n = 2$, primero un vector aleatorio $\mathbf{X} = (X_1, X_2)$ donde X_1 es una variable aleatoria continua uniforme, véase el Ejemplo 8, sobre el intervalo abierto $]0, 1[$, y sea la variable aleatoria $X_2 = X_1$, lo que implica que las funciones de distribución marginales son iguales, y dadas por $F_{X_1}(x) = F_{X_2}(x) = F(x) = x\mathbf{1}_{\{0 < x < 1\}} + \mathbf{1}_{\{x \geq 1\}}$. En cuanto a su función de distribución conjunta:

$$\begin{aligned} F_{\mathbf{X}}(x_1, x_2) &= \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) = \mathbb{P}(X_1 \leq x_1, X_1 \leq x_2) = \mathbb{P}(X_1 \leq \min\{x_1, x_2\}), \\ &= F(\min\{x_1, x_2\}) = F(x_1)\mathbf{1}_{\{x_1 \leq x_2\}} + F(x_2)\mathbf{1}_{\{x_1 > x_2\}} = \min\{F(x_1), F(x_2)\}. \end{aligned}$$

Por otro lado, consideremos ahora el vector aleatorio $\mathbf{Y} = (Y_1, Y_2)$ donde Y_1 es una variable aleatoria continua uniforme sobre el intervalo abierto $]0, 1[$, y donde $Y_2 = 1 - Y_1$. Es inmediato verificar que Y_2 resulta ser también una variable aleatoria continua uniforme sobre el intervalo abierto $]0, 1[$, lo que implica que las funciones de distribución marginales son iguales, y dadas por $F_{Y_1}(y) = F_{Y_2}(y) = F(y) = y\mathbf{1}_{\{0 < y < 1\}} + \mathbf{1}_{\{y \geq 1\}}$. En cuanto a su función de distribución conjunta:

$$\begin{aligned} F_{\mathbf{Y}}(y_1, y_2) &= \mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2) = \mathbb{P}(Y_1 \leq y_1, 1 - Y_1 \leq y_2) = \mathbb{P}(1 - y_2 \leq Y_1 \leq y_1), \\ &= [F(y_1) - F(1 - y_2)]\mathbf{1}_{\{1 - y_2 < y_1\}} = \max\{F(y_1) + F(y_2) - 1, 0\}. \end{aligned}$$

Tenemos entonces dos vectores aleatorios $\mathbf{X} = (X_1, X_2)$ e $\mathbf{Y} = (Y_1, Y_2)$ cuyas funciones de distribución marginales son iguales, pero con funciones de distribución conjunta diferentes ya que $F_{\mathbf{X}} \neq F_{\mathbf{Y}}$.

□

En el caso de una variable aleatoria X , por medio de su función de distribución de probabilidades $F_X(x) = \mathbb{P}(X \leq x)$ es relativamente accesible hacer cálculo de probabilidades sobre cualquier tipo de intervalo en \mathbb{R} , como consecuencia del Teorema 1. Pero en el caso de un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ con $n \geq 2$, reexpresar un subconjunto de interés $B \in \mathfrak{B}(\mathbb{R}^n)$ por medio de operaciones de conjuntos de la clase \mathfrak{J}^n es, por lo general, demasiado complejo y hasta inaccesible en términos prácticos, lo que dificulta el cálculo de probabilidades por medio de su función de distribución conjunta, salvo en el caso muy particular de que B sea un producto cartesiano de intervalos semiabiertos de la forma $]a, b]$. Por ejemplo, en el caso $n = 2$ se tendría:

$$\mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) = F_{X_1, X_2}(b_1, b_2) - F_{X_1, X_2}(b_1, a_2) - F_{X_1, X_2}(a_1, b_2) + F_{X_1, X_2}(a_1, a_2). \quad (27)$$

Por simplicidad de notación, a partir de este momento nos restringiremos al caso de vectores aleatorios bivariados, y utilizaremos la notación (X, Y) para el vector aleatorio, y $F_{X, Y}$ para su función de distribución conjunta (bivariada), pero los conceptos y resultados que se abordarán son fácilmente extendibles a dimensiones mayores que $n = 2$.

1.1. Vectores aleatorios discretos

Si las variables aleatorias componentes de un vector aleatorio (X, Y) son discretas, entonces tanto $\text{Ran } X$ como $\text{Ran } Y$ son conjuntos numerables, lo que implica que $\text{Ran}(X, Y)$ es también un conjunto numerable, por ser un subconjunto de un producto cartesiano de conjuntos numerables, esto es que $\text{Ran}(X, Y) \subseteq \text{Ran } X \times \text{Ran } Y$, donde la igualdad puede o no cumplirse, porque aun teniendo que $\mathbb{P}(X = x) > 0$ y $\mathbb{P}(Y = y) > 0$ puede ocurrir que simultáneamente esto no sea posible, es decir que $\mathbb{P}((X, Y) = (x, y)) = 0$ para algún (o algunos) par(es) de números reales (x, y) , como se ilustra en el siguiente:

Ejemplo 10 Considere una variable aleatoria X con distribución de probabilidad Bernoulli con parámetro $0 < \theta < 1$, lo que implica que $\text{Ran } X = \{0, 1\}$, $\mathbb{P}(X = 1) = \theta$ y $\mathbb{P}(X = 0) = 1 - \theta$. Si definimos ahora la variable aleatoria $Y = 1 - X$ es inmediato deducir que $\text{Ran } Y = \{0, 1\}$, $\mathbb{P}(Y = 1) = 1 - \theta$, y que $\mathbb{P}(Y = 0) = \theta$, es decir que Y resulta ser también una variable aleatoria Bernoulli, pero con parámetro $1 - \theta$. Aunque el producto cartesiano:

$$\text{Ran } X \times \text{Ran } Y = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$$

notemos que, por ejemplo, $(1, 1) \notin \text{Ran}(X, Y)$ ya que:

$$\begin{aligned} \mathbb{P}((X, Y) = (1, 1)) &= \mathbb{P}(\{\omega \in \Omega : X(\omega) = 1\} \cap \{\omega \in \Omega : Y(\omega) = 1\}) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) = 1\} \cap \{\omega \in \Omega : 1 - X(\omega) = 1\}) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) = 1\} \cap \{\omega \in \Omega : X(\omega) = 0\}) \\ &= \mathbb{P}(\emptyset) = 0, \end{aligned}$$

y de forma análoga se concluye que $(0, 0) \notin \text{Ran}(X, Y)$. Por otro lado:

$$\begin{aligned} \mathbb{P}((X, Y) = (1, 0)) &= \mathbb{P}(\{\omega \in \Omega : X(\omega) = 1\} \cap \{\omega \in \Omega : Y(\omega) = 0\}) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) = 1\} \cap \{\omega \in \Omega : 1 - X(\omega) = 0\}) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) = 1\} \cap \{\omega \in \Omega : X(\omega) = 1\}) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) = 1\}) = \theta, \end{aligned}$$

y de forma análoga se deduce que $\mathbb{P}((X, Y) = (0, 1)) = 1 - \theta$, por lo que en este ejemplo:

$$\text{Ran}(X, Y) = \{(0, 1), (1, 0)\} \subsetneq \text{Ran } X \times \text{Ran } Y.$$

□

Definición 11 Un **vector aleatorio** (X, Y) se denomina **discreto** si sus variables aleatorias componentes son discretas.

□

Lo anterior implica que $\text{Ran}(X, Y) \subseteq \text{Ran } X \times \text{Ran } Y$ es un conjunto numerable y, de forma análoga al caso de variables aleatorias discretas, basta con poder calcular $\mathbb{P}((X, Y) = (x, y))$ para cualquier $(x, y) \in \text{Ran}(X, Y)$, ya que para cualquier $B \in \mathfrak{B}(\mathbb{R}^2)$ la probabilidad correspondiente se calcula mediante:

$$\mathbb{P}((X, Y) \in B) = \sum_{(x, y) \in B \cap \text{Ran}(X, Y)} \mathbb{P}((X, Y) = (x, y)), \quad (28)$$

lo cual motiva la siguiente:

Definición 12 La **función de masa de probabilidades conjunta** de un vector aleatorio discreto (X, Y) se denota y define mediante:

$$p_{X, Y}(x, y) := \mathbb{P}((X, Y) = (x, y)).$$

□

Entonces $p_{X,Y}(x, y) > 0$ si solo si $(x, y) \in \text{Ran}(X, Y)$. Más aún:

$$p_{X,Y}(x, y) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\} \cap \{\omega \in \Omega : Y(\omega) = y\}) \equiv \mathbb{P}(X = x, Y = y), \quad (29)$$

donde $\{\omega \in \Omega : X(\omega) = x\}$ y $\{\omega \in \Omega : Y(\omega) = y\}$ son subconjuntos del espacio muestral Ω que representan eventos en un mismo espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$, y que por separado pudiera ocurrir que tengan probabilidad positiva, pero al intersectarlos, si son disjuntos, su intersección resultaría vacía y, por tanto, con probabilidad cero, es decir, aún teniendo $p_X(x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}) > 0$, y $p_Y(y) = \mathbb{P}(\{\omega \in \Omega : Y(\omega) = y\}) > 0$, si $\{\omega \in \Omega : X(\omega) = x\} \cap \{\omega \in \Omega : Y(\omega) = y\} = \emptyset$ entonces $p_{X,Y}(x, y) = 0$.

Si para cada $(x, y) \in \text{Ran}(X, Y)$ definimos la imagen inversa:

$$E_{x,y} := (X, Y)^{(-1)}(\{(x, y)\}) = \{\omega \in \Omega : (X(\omega), Y(\omega)) = (x, y)\} \in \mathfrak{F}$$

entonces $p_{X,Y}(x, y) = \mathbb{P}(E_{x,y}) > 0$ y además la colección $\{E_{x,y} : (x, y) \in \text{Ran}(X, Y)\}$ constituye una partición de Ω , y por lo tanto:

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(\bigcup_{(x,y) \in \text{Ran}(X,Y)} E_{x,y}\right) = \sum_{(x,y) \in \text{Ran}(X,Y)} \mathbb{P}(E_{x,y}) = \sum_{(x,y) \in \text{Ran}(X,Y)} p_{X,Y}(x, y)$$

Lo anterior implica que la función de masa de probabilidades (conjunta) de un vector aleatorio discreto (X, Y) siempre cumple las siguientes propiedades:

- $p_{X,Y}(x, y) > 0$ para todo $(x, y) \in \text{Ran}(X, Y)$,
- $\sum_{(x,y) \in \text{Ran}(X,Y)} p_{X,Y}(x, y) = 1$,

A partir de $p_{X,Y}$ es posible deducir las funciones de masa de probabilidades marginales (o individuales) de cada una de las variables aleatorias discretas componentes:

$$\begin{aligned} p_X(x) &= \mathbb{P}(X = x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\} \cap \Omega) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\} \cap \bigcup_{y \in \text{Ran } Y} \{\omega \in \Omega : Y(\omega) = y\}) \\ &= \mathbb{P}\left(\bigcup_{y \in \text{Ran } Y} \{\omega \in \Omega : X(\omega) = x\} \cap \{\omega \in \Omega : Y(\omega) = y\}\right) \\ &= \sum_{y \in \text{Ran } Y} \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\} \cap \{\omega \in \Omega : Y(\omega) = y\}) \\ &= \sum_{y \in \text{Ran } Y} p_{X,Y}(x, y) \end{aligned}$$

Al procedimiento de deducir las funciones de masa de probabilidades marginales de un vector aleatorio discreto (X, Y) a partir de su función de masa de probabilidades conjunta se le denomina *marginalización*:

$$p_X(x) = \sum_{y \in \text{Ran } Y} p_{X,Y}(x, y), \quad p_Y(y) = \sum_{x \in \text{Ran } X} p_{X,Y}(x, y). \quad (30)$$

Como consecuencia inmediata a partir de aplicar la Proposición 1 f con $E_1 = \{\omega \in \Omega : X(\omega) = x\}$ y $E_2 = \{\omega \in \Omega : Y(\omega) = y\}$, obtenemos la siguiente relación entre la función de masa de probabilidades conjunta y sus marginales:

$$\max\{p_X(x) + p_Y(y) - 1, 0\} \leq p_{X,Y}(x, y) \leq \min\{p_X(x), p_Y(y)\}. \quad (31)$$

Sin embargo, el solo conocimiento de las funciones de masa de probabilidades marginales p_X y p_Y no permite reconstruir, en general, la función de masa de probabilidades conjunta $p_{X,Y}$, como se verá en el siguiente:

Ejemplo 11 Distribución Bernoulli bivariada. Obtendremos la fórmula general para la función de masa de probabilidades conjunta de un vector aleatorio discreto (X, Y) , donde X sea una variable aleatoria Bernoulli con parámetro $0 < \theta_1 < 1$ e Y una variable aleatoria Bernoulli con parámetro $0 < \theta_2 < 1$. Como en este caso $\text{Ran}(X, Y) \subseteq \text{Ran } X \times \text{Ran } Y$, el caso general debe considerar el rango más amplio posible, es decir:

$$\text{Ran}(X, Y) = \{(0, 0), (0, 1), (1, 0), (1, 1)\}.$$

Denotemos:

$$p_{ij} := p_{X,Y}(i, j) = \mathbb{P}(X = i, Y = j), \quad (i, j) \in \text{Ran}(X, Y),$$

y por lo tanto debemos encontrar números reales $\{p_{00}, p_{01}, p_{10}, p_{11}\} \subset [0, 1]$ tales que $p_{00} + p_{01} + p_{10} + p_{11} = 1$, que además satisfagan la características marginales (individuales) de X e Y , esto es, tales que:

$$\begin{aligned} p_X(0) &= 1 - \theta_1 = p_{00} + p_{01}, & p_X(1) &= \theta_1 = p_{10} + p_{11}, \\ p_Y(0) &= 1 - \theta_2 = p_{00} + p_{10}, & p_Y(1) &= \theta_2 = p_{01} + p_{11}. \end{aligned}$$

Es decir, tenemos que resolver el siguiente sistema de ecuaciones con restricciones:

$$0 \leq p_{ij} \leq 1, \quad (i, j) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}, \quad (32)$$

$$p_{00} + p_{01} + p_{10} + p_{11} = 1, \quad (33)$$

$$p_{00} + p_{01} = 1 - \theta_1, \quad (34)$$

$$p_{10} + p_{11} = \theta_1, \quad (35)$$

$$p_{00} + p_{10} = 1 - \theta_2, \quad (36)$$

$$p_{01} + p_{11} = \theta_2. \quad (37)$$

La ecuación (33) implica que basta determinar tres valores p_{ij} y el cuarto se deduce, por ejemplo:

$$p_{00} = 1 - p_{01} - p_{10} - p_{11}. \quad (38)$$

De hecho, al sustituir (38) en (34) obtenemos (35), y al sustituir (38) en (36) obtenemos (37), por lo que el sistema de ecuaciones se reduce a lo siguiente:

$$p_{10} = \theta_1 - p_{11}, \quad (39)$$

$$p_{01} = \theta_2 - p_{11}, \quad (40)$$

$$p_{00} = 1 - \theta_1 - \theta_2 + p_{11}, \quad (41)$$

siendo lo anterior un sistema de 3 ecuaciones con 4 incógnitas, lo que en este caso conduce a una infinidad de soluciones que pueden caracterizarse, por ejemplo, en términos de un valor admisible que se asigne a p_{11} , que en principio tenemos que $0 \leq p_{11} \leq 1$, pero la propiedad (31) impone una restricción mayor:

$$\max\{\theta_1 + \theta_2 - 1, 0\} \leq p_{11} \leq \min\{\theta_1, \theta_2\}. \quad (42)$$

Esto es, existe una infinidad no numerable de valores admisibles para p_{11} , y una vez escogido alguno entonces aplicando (39), (40) y (41) se deducen los valores correspondientes para p_{10} , p_{01} y p_{00} . Con esto queda demostrado que existe una infinidad no numerable de funciones de masa de probabilidad conjunta $p_{X,Y}$ que tienen exactamente las mismas marginales, Bernoulli con parámetro θ_1 para X , y Bernoulli con parámetro θ_2 para Y , y por lo tanto el solo conocimiento de las funciones de masa de probabilidad marginales no permite identificar la función de masa de probabilidad conjunta. \square

De forma análoga, para el caso de dimensión 3, en el caso de un vector aleatorio discreto (X, Y, Z) su función de masa de probabilidades conjunta estaría dada por:

$$\begin{aligned} p_{X,Y,Z}(x, y, z) &:= \mathbb{P}((X, Y, Z) = (x, y, z)) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\} \cap \{\omega \in \Omega : Y(\omega) = y\} \cap \{\omega \in \Omega : Z(\omega) = z\}) \\ &\equiv \mathbb{P}(X = x, Y = y, Z = z), \end{aligned}$$

y mediante una argumentación análoga a lo obtenido en (30) en este caso se pueden obtener tres funciones de masa de probabilidad conjunta de dimensión 2, para los vectores aleatorios (X, Y) , (X, Z) e (Y, Z) :

$$p_{X,Y}(x, y) = \sum_{z \in \text{Ran } Z} p_{X,Y,Z}(x, y, z), \quad p_{X,Z}(x, z) = \sum_{y \in \text{Ran } Y} p_{X,Y,Z}(x, y, z), \quad p_{Y,Z}(y, z) = \sum_{x \in \text{Ran } X} p_{X,Y,Z}(x, y, z),$$

a partir de las cuales, a su vez, se pueden deducir las funciones de masa de probabilidad marginal p_X , p_Y y p_Z .

1.2. Vectores aleatorios absolutamente continuos

Ahora consideraremos el caso de un vector aleatorio (X, Y) donde cada una de las variables aleatorias componentes es absolutamente continua (ver Definición 7), con funciones de densidad de probabilidades f_X y f_Y , respectivamente.

Definición 13 Un **vector aleatorio** (X, Y) se denomina **absolutamente continuo** si existe una función $f_{X,Y} : \mathbb{R}^2 \rightarrow [0, +\infty[$ tal que para todo $B \in \mathfrak{B}(\mathbb{R}^2)$:

$$\mathbb{P}((X, Y) \in B) = \iint_B f_{X,Y}(x, y) \, dx dy,$$

y a la función $f_{X,Y}$ se le denomina **función de densidad conjunta de probabilidades**. □

Como necesariamente $\mathbb{P}((X, Y) \in \mathbb{R}^2) = 1$ entonces la definición anterior implica en particular que:

$$\iint_{\mathbb{R}^2} f_{X,Y}(x, y) \, dx dy = 1.$$

A partir de $f_{X,Y}$ es posible deducir las funciones de densidad marginales de cada una de las variables aleatorias, procediendo al que denominamos *marginalización*, aplicando (17) y (18):

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\} \cap \Omega) \\ &= \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\} \cap \{\omega \in \Omega : -\infty < Y(\omega) < +\infty\}) \\ &= \mathbb{P}((X, Y) \in]-\infty, x] \times]-\infty, +\infty[) \\ &= \int_{-\infty}^x \int_{-\infty}^{+\infty} f_{X,Y}(s, y) \, dy ds = \int_{-\infty}^x \psi(s) \, ds, \end{aligned}$$

donde $\psi(s) := \int_{-\infty}^{+\infty} f_{X,Y}(s, y) \, dy$, y luego aplicando el Teorema Fundamental del Cálculo obtenemos:

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) = \frac{d}{dx} \int_{-\infty}^x \psi(s) \, ds = \psi(x) \\ &= \int_{-\infty}^{+\infty} f_{X,Y}(x, y) \, dy, \end{aligned} \tag{43}$$

y de manera análoga se obtiene $f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) \, dx$. También, a partir de $f_{X,Y}$ podemos obtener la función de distribución conjunta mediante:

$$F_{X,Y}(x, y) = \mathbb{P}((X, Y) \in]-\infty, x] \times]-\infty, y]) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(s, t) \, ds dt, \tag{44}$$

y por tanto, como consecuencia del Teorema Fundamental del Cálculo, tenemos que para todo $(x, y) \in \mathbb{R}^2$ donde $f_{X,Y}$ sea continua entonces $F_{X,Y}$ es derivable:

$$\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) = f_{X,Y}(x, y). \tag{45}$$

Pero el que (X, Y) sea un vector de variables aleatorias absolutamente continuas no implica necesariamente que el vector sea absolutamente continuo, es decir, la existencia de una función de densidad conjunta no está garantizada. Por ejemplo, sea X una variable aleatoria absolutamente continua, $g : \mathbb{R} \rightarrow \mathbb{R}$ una función continua y estrictamente creciente, y defínase una variable aleatoria $Y := g(X)$. Es inmediato verificar que $F_{X,Y}(x, y) = \min\{F_X(x), F_Y(y)\}$ y por tanto $\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) = 0$, donde claramente la constante cero no es una función de densidad conjunta de probabilidades.

Ejemplo 12 Densidad conjunta uniforme sobre un conjunto acotado. Sea $C \subset \mathbb{R}^2$ un conjunto acotado y con área positiva $\Delta(C) > 0$. Si definimos una función $f : \mathbb{R}^2 \rightarrow [0, +\infty[$ mediante:

$$f(x, y) := \frac{1}{\Delta(C)} \mathbf{1}_C(x, y)$$

entonces:

$$\iint_{\mathbb{R}^2} f(x, y) dx dy = \frac{1}{\Delta(C)} \iint_C dx dy = \frac{1}{\Delta(C)} \Delta(C) = 1.$$

Si (X, Y) es un vector absolutamente continuo con la función de densidad $f(x, y)$ anterior, entonces para cualquier $B \in \mathfrak{B}(\mathbb{R}^2)$ tenemos que:

$$\mathbb{P}((X, Y) \in B) = \frac{1}{\Delta(C)} \iint_B \mathbf{1}_C(x, y) dx dy = \frac{1}{\Delta(C)} \iint_{B \cap C} dx dy = \frac{\Delta(B \cap C)}{\Delta(C)}.$$

Y para deducir sus funciones de densidad marginal, por ejemplo la de X sería mediante:

$$f_X(x) = \frac{1}{\Delta(C)} \int_{-\infty}^{+\infty} \mathbf{1}_C(x, y) dy.$$

Y si en particular el conjunto C puede expresarse como $C = \{(x, y) \in \mathbb{R}^2 : x \in A, g_1(x) \leq y \leq g_2(x)\}$ entonces:

$$f_X(x) = \frac{1}{\Delta(C)} \int_{-\infty}^{+\infty} \mathbf{1}_A(x) \mathbf{1}_{\{g_1(x) \leq y \leq g_2(x)\}} dy = \frac{\mathbf{1}_A(x)}{\Delta(C)} \int_{g_1(x)}^{g_2(x)} dy = \frac{g_2(x) - g_1(x)}{\Delta(C)} \mathbf{1}_A(x),$$

y lo análogo para la densidad marginal f_Y .

□

Si (X, Y) es un vector aleatorio absolutamente continuo, la ecuación (43) implica que sus variables aleatorias componentes son también absolutamente continuas. Pero al revés, no necesariamente: puede ocurrir que X e Y sean marginalmente (individualmente) absolutamente continuas, con funciones de densidad de probabilidades f_X y f_Y , respectivamente, pero que el vector aleatorio (X, Y) no sea absolutamente continuo, es decir, que no exista su función de densidad conjunta. Esto último se ilustra en el siguiente:

Ejemplo 13 Sea X una variable aleatoria absolutamente continua con función de densidad de probabilidades uniforme sobre el intervalo abierto $]0, 1[$, esto es $f_X(x) = \mathbf{1}_{\{0 < x < 1\}}$, y por lo tanto con función de distribución $F_X(t) = \int_{-\infty}^t f_X(x) dx = x \mathbf{1}_{\{0 < x < 1\}} + \mathbf{1}_{\{x \geq 1\}}$. Definamos la variable aleatoria $Y := X^2$. Como $\text{Ran } X =]0, 1[$ entonces $\text{Ran } Y =]0, 1[$ también, ya que si $0 < x < 1$ entonces $0 < x^2 < 1$. Deducimos la función de distribución de probabilidades de Y :

$$F_Y(y) = \mathbb{P}(Y \leq y) = \begin{cases} 0, & \text{si } y \leq 0, \\ \mathbb{P}(X^2 \leq y) = \mathbb{P}(X \leq \sqrt{y}) = F_X(\sqrt{y}) = \sqrt{y}, & \text{si } 0 < y < 1, \\ 1, & \text{si } y \geq 1. \end{cases}$$

Como $F_Y(y)$ es derivable para todo $y \in \mathbb{R} \setminus \{0, 1\}$ (es decir, en todo \mathbb{R} excepto en una cantidad numerable de puntos aislados, en este caso $y = 0$ e $y = 1$) entonces Y resulta ser también una variable aleatoria absolutamente continua, con función de densidad de probabilidades:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{\sqrt{y}} f_X(\sqrt{y}) = \frac{1}{\sqrt{y}} \mathbf{1}_{\{0 < y < 1\}}.$$

Ahora obtengamos la función de distribución conjunta $F_{X,Y} : \mathbb{R}^2 \rightarrow [0, 1]$ del vector aleatorio (X, Y) :

$$\begin{aligned} F_{X,Y}(x, y) &= \mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x, X^2 \leq y) = \mathbb{P}(X \leq x, X \leq \sqrt{y}) \\ &= \mathbb{P}(X \leq \min\{x, \sqrt{y}\}) = F_X(\min\{x, \sqrt{y}\}) \\ &= \begin{cases} 1, & x \geq 1 \wedge y \geq 1, \\ x, & 0 < x < 1 \wedge y \geq x^2, \\ \sqrt{y}, & 0 < y < 1 \wedge x > \sqrt{y}, \\ 0, & x \leq 0 \vee y \leq 0. \end{cases} \end{aligned}$$

$F_{X,Y}$ es dos veces diferenciable sobre todo \mathbb{R}^2 excepto sobre el conjunto de área cero $C := C_1 \cup C_2 \cup C_3$ donde:

$$C_1 := \{(x, y) \in \mathbb{R}^2 : x \geq 0, y \in \{0, 1\}\}$$

$$C_2 := \{(0, y) \in \mathbb{R}^2 : x \in \{0, 1\}, y \geq 0\}$$

$$C_3 := \{(x, x^2) \in \mathbb{R}^2 : 0 \leq x \leq 1\}$$

pero para todo $(x, y) \in \mathbb{R}^2 \setminus C$:

$$\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) = 0,$$

lo cual evidentemente no cumple con ser una función de densidad conjunta ya que $\iint_{\mathbb{R}^2} 0 \, dx dy = 0 \neq 1$, y tenemos así el caso de un vector aleatorio (X, Y) de variables aleatorias absolutamente continuas, cuya función de densidad conjunta no existe.

□

2. Distribuciones condicionales e independencia

Dado un vector aleatorio (X, Y) definido sobre un espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ buscaremos ahora analizar la influencia que tiene una de las variables aleatorias componentes en el cálculo de probabilidades de la otra. Por ejemplo ¿qué tanto cambiaría $\mathbb{P}(Y \in B)$ si ante un valor observado $\omega \in \Omega$ se sabe que $X(\omega) \in A$? ¿Cómo incorporar o reflejar esta información en el cálculo de $\mathbb{P}(Y \in B)$?

Como $Y^{(-1)}(B) = \{\omega \in \Omega : Y(\omega) \in B\} \in \mathfrak{F}$ y $X^{(-1)}(A) = \{\omega \in \Omega : X(\omega) \in A\} \in \mathfrak{F}$ representan eventos sobre el mismo espacio de probabilidad, la pregunta anterior nos conduce a calcular la siguiente probabilidad condicional sobre dicho espacio mediante la Definición 2:

$$\begin{aligned} \mathbb{P}(\{\omega \in \Omega : Y(\omega) \in B\} | \{\omega \in \Omega : X(\omega) \in A\}) &= \frac{\mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\} \cap \{\omega \in \Omega : Y(\omega) \in B\})}{\mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})} \\ &= \frac{\mathbb{P}(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A \times B\})}{\mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})} \\ \mathbb{P}(Y \in B | X \in A) &= \frac{\mathbb{P}((X, Y) \in A \times B)}{\mathbb{P}(X \in A)}. \end{aligned} \quad (46)$$

Es decir, sin tener información sobre el valor que reporta la variable aleatoria X calcularíamos $\mathbb{P}(Y \in B)$ mediante la distribución de probabilidad marginal de Y únicamente (mediante su función de distribución F_Y o bien mediante su función de masa o densidad marginal, según sea el caso). Pero ante el conocimiento de que el evento $\{\omega \in \Omega : X(\omega) \in A\}$ ha ocurrido, la probabilidad del evento $\{\omega \in \Omega : Y(\omega) \in B\}$ se actualiza a $\mathbb{P}(Y \in B | X \in A)$ dada por (46), siempre que $\mathbb{P}(X \in A) > 0$.

2.1. Caso discreto

En el caso particular de que (X, Y) sea un vector aleatorio discreto con función de masa de probabilidades conjunta $p_{X,Y}$ y funciones de masa de probabilidades marginales p_X y p_Y , respectivamente, si $x \in \text{Ran } X$ y definimos $A := \{x\}$ y $B := \{y\}$ en (46) obtenemos:

$$\mathbb{P}(Y = y | X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)} = \frac{p_{X,Y}(x, y)}{p_X(x)} \quad (47)$$

donde está garantizado que $p_X(x) > 0$ si $x \in \text{Ran } X$. Esto es, la probabilidad $\mathbb{P}(Y = y) = p_Y(y)$ se actualiza a $\mathbb{P}(Y = y | X = x)$ ante la información de que ocurrió (o así lo suponemos) el evento $\{\omega \in \Omega : X(\omega) = x\} \in \mathfrak{F}$. De (47) es inmediato notar que $\mathbb{P}(Y = y | X = x) \geq 0$ ya que $p_{X,Y}(x, y) \geq 0$ y $p_X(x) > 0$, y además que $\mathbb{P}(Y = y | X = x) \leq 1$ ya que, como consecuencia de la Proposición 1 d, se deduce que:

$$p_{X,Y}(x, y) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\} \cap \{\omega \in \Omega : Y(\omega) = y\}) \leq \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}) = p_X(x).$$

También, aplicando (30) obtenemos:

$$\sum_{y \in \text{Ran } Y} \mathbb{P}(Y = y | X = x) = \frac{1}{p_X(x)} \sum_{y \in \text{Ran } Y} p_{X,Y}(x, y) = \frac{1}{p_X(x)} p_X(x) = 1,$$

lo que implica que (47) cumple con las propiedades de una función de masa de probabilidades para la variable aleatoria discreta Y , posiblemente distinta a la marginal p_Y ya que incorpora la ocurrencia del evento $\{X = x\}$, motivando la siguiente:

Definición 14 Sea (X, Y) un vector aleatorio discreto con función de masa de probabilidades conjunta $p_{X,Y}$, y con funciones de masa de probabilidades marginales p_X y p_Y , respectivamente. Para toda $x \in \text{Ran } X$ (y por tanto $p_X(x) > 0$) se denota y define la **función de masa de probabilidades condicional** para Y dado el evento $\{X = x\}$ como:

$$p_{Y|X}(y | x) := \frac{p_{X,Y}(x, y)}{p_X(x)}.$$

Análogamente, para toda $y \in \text{Ran } Y$ (y por tanto $p_Y(y) > 0$) se denota y define la función de masa de probabilidades condicional para X dado el evento $\{Y = y\}$ como:

$$p_{X|Y}(x | y) := \frac{p_{X,Y}(x, y)}{p_Y(y)}.$$

□

Tanto $p_Y(y)$ como $p_{Y|X}(y|x)$ son funciones de masa de probabilidades univariadas para la misma variable aleatoria discreta Y , con la diferencia que $p_{Y|X}(y|x)$ toma en cuenta (o supone) que el evento $\{X = x\}$ ha ocurrido, mientras que $p_Y(y)$ no lo toma en cuenta. Como $x \in \text{Ran } X$, el valor x se convierte en un *parámetro* de la función de masa de probabilidades $p_{Y|X}(y|x)$, ya que valores distintos de x pudieran arrojar distintas probabilidades condicionales, es decir si $x_1 \neq x_2$ podría (o no) resultar que $p_{Y|X}(y|x_1) \neq p_{Y|X}(y|x_2)$.

Utilizamos $p_{Y|X}(y|x)$ en lugar de $p_Y(y)$ cuando requerimos hacer cálculo de probabilidades con la variable aleatoria discreta Y tomando en cuenta la posible influencia del evento $\{X = x\}$. Si para todo $x \in \text{Ran } X$ el evento $\{X = x\}$ no influye en el cálculo de probabilidades con la variable aleatoria Y tendríamos lo siguiente:

$$\mathbb{P}(Y = y | X = x) = \mathbb{P}(Y = y) \Leftrightarrow \frac{p_{X,Y}(x, y)}{p_X(x)} = p_Y(y) \Leftrightarrow p_{X,Y}(x, y) = p_X(x)p_Y(y), \quad (48)$$

y a la misma conclusión (48) llegaríamos partiendo de $\mathbb{P}(X = x | Y = y) = \mathbb{P}(X = x)$ para todo $y \in \text{Ran } Y$. Dado un vector aleatorio discreto, el que los valores de una de las variables no influya en el cálculo de probabilidades de otras nos conduce al concepto de *independencia* entre variables aleatorias:

Definición 15 Dado un vector aleatorio discreto (X, Y) con función de masa de probabilidades conjunta $p_{X,Y}$, y funciones de masa de probabilidades marginales p_X y p_Y , respectivamente, decimos que X e Y son **variables aleatorias independientes** si y solo si:

$$p_{X,Y}(x, y) = p_X(x)p_Y(y).$$

□

Si de un vector aleatorio discreto (X, Y) conocemos su función de masa de probabilidades conjunta $p_{X,Y}$, a partir de esta última podemos deducir las funciones de masa de probabilidades marginales p_X y p_Y mediante (30), y solo en caso de que $p_X(x)p_Y(y) = p_{X,Y}(x, y)$ podemos entonces afirmar que X e Y son variables aleatorias independientes. En otro sentido, si solo conocemos las funciones de masa de probabilidades marginales p_X y p_Y , solo en el caso de que sepamos además que X e Y son independientes podríamos entonces deducir que la función de masa de probabilidades conjunta de (X, Y) es $p_{X,Y}(x, y) = p_X(x)p_Y(y)$.

Ejemplo 14 Distribución Bernoulli bivariada. Si en el Ejemplo 11 definimos como $\theta_{12} := p_{11}$ tenemos una función de masa de probabilidades conjunta $p_{X,Y}$ con parámetros $0 < \theta_1 < 1$, $0 < \theta_2 < 1$ y $\max\{\theta_1 + \theta_2 - 1, 0\} \leq \theta_{12} \leq \min\{\theta_1, \theta_2\}$, donde $p_{ij} := p_{X,Y}(i, j)$, con $(i, j) \in \{0, 1\}^2$ y:

$$\begin{aligned} p_{00} &= 1 - \theta_1 - \theta_2 + \theta_{12}, \\ p_{10} &= \theta_1 - \theta_{12}, \\ p_{01} &= \theta_2 - \theta_{12}, \\ p_{11} &= \theta_{12}. \end{aligned}$$

Aplicando (30) deducimos las marginales $p_X(x) = \theta_1^x(1 - \theta_1)^{1-x} \mathbf{1}_{\{0,1\}}(x)$ y $p_Y(y) = \theta_2^y(1 - \theta_2)^{1-y} \mathbf{1}_{\{0,1\}}(y)$. De acuerdo con la Definición 15 X e Y son independientes si y solo si $p_{X,Y}(x, y) = p_X(x)p_Y(y)$, que en este caso equivale a que se cumplan las siguientes cuatro desigualdades:

$$\begin{aligned} 1 - \theta_1 - \theta_2 + \theta_{12} &= (1 - \theta_1)(1 - \theta_2), \\ \theta_1 - \theta_{12} &= \theta_1(1 - \theta_2), \\ \theta_2 - \theta_{12} &= (1 - \theta_1)\theta_2, \\ \theta_{12} &= \theta_1\theta_2 \end{aligned}$$

de lo cual se deduce que X e Y son independientes si y solo si $\theta_{12} = \theta_1\theta_2$, valor admisible ya que es inmediato verificar que $\max\{\theta_1 + \theta_2 - 1, 0\} \leq \theta_1\theta_2 \leq \min\{\theta_1, \theta_2\}$. Si el parámetro $\theta_{12} \neq \theta_1\theta_2$ tenemos el caso de variables aleatorias dependientes, y procedemos a deducir las funciones de masa de probabilidades condicionales. Para el caso

de Y condicional en $X = x$, como $x \in \{0, 1\}$ tenemos los casos $p_{Y|X}(y|0)$ y $p_{Y|X}(y|1)$, cuyas fórmulas deducimos aplicando la Definición 14:

$$\begin{aligned} p_{Y|X}(y|0) &= \frac{(\theta_2 - \theta_{12})^y (1 - \theta_1 - \theta_2 + \theta_{12})^{1-y}}{1 - \theta_1} \mathbf{1}_{\{0,1\}}(y) = \left(\frac{\theta_2 - \theta_{12}}{1 - \theta_1} \right)^y \left(1 - \frac{\theta_2 - \theta_{12}}{1 - \theta_1} \right)^{1-y} \mathbf{1}_{\{0,1\}}(y), \\ p_{Y|X}(y|1) &= \frac{\theta_{12}^y (\theta_1 - \theta_{12})^{1-y}}{\theta_1} \mathbf{1}_{\{0,1\}}(y) = \left(\frac{\theta_{12}}{\theta_1} \right)^y \left(1 - \frac{\theta_{12}}{\theta_1} \right)^{1-y} \mathbf{1}_{\{0,1\}}(y), \end{aligned}$$

esto es, la distribución de probabilidad de Y condicional en $X = 0$ es Bernoulli con parámetro $\frac{\theta_2 - \theta_{12}}{1 - \theta_1}$, pero la distribución de probabilidad de Y condicional en $X = 1$ es Bernoulli con parámetro $\frac{\theta_{12}}{\theta_1}$, en contraste con la distribución marginal (no condicional) de Y que es Bernoulli con parámetro θ_2 . De forma análoga se deduce que la distribución de probabilidad de X condicional en $Y = 0$ es Bernoulli con parámetro $\frac{\theta_1 - \theta_{12}}{1 - \theta_2}$, y que la distribución de probabilidad de X condicional en $Y = 1$ es Bernoulli con parámetro $\frac{\theta_{12}}{\theta_2}$, en contraste con la distribución marginal (no condicional) de X que es Bernoulli con parámetro θ_1 . □

Para tener idea del caso de vectores aleatorios discretos de dimensión mayor que 2, basta con considerar, por ejemplo, el caso de un vector aleatorio discreto (X, Y, Z) con función de masa de probabilidades conjunta $p_{X,Y,Z}$, en cuyo caso se pueden deducir funciones de masa de probabilidades condicionales de dimensión 2, por ejemplo:

$$p_{X,Z|Y}(x, z | y) = \mathbb{P}(X = x, Z = z | Y = y) = \frac{\mathbb{P}(X = x, Y = y, Z = z)}{\mathbb{P}(Y = y)} = \frac{p_{X,Y,Z}(x, y, z)}{p_Y(y)},$$

o bien de dimensión 1, por ejemplo:

$$p_{Z|X,Y}(z | x, y) = \mathbb{P}(Z = z | X = x, Y = y) = \frac{\mathbb{P}(X = x, Y = y, Z = z)}{\mathbb{P}(X = x, Y = y)} = \frac{p_{X,Y,Z}(x, y, z)}{p_{X,Y}(x, y)}.$$

Y también, análogamente, las variables aleatorias X , Y y Z son conjuntamente independientes si y solo si $p_{X,Y,Z}(x, y, z) = p_X(x)p_Y(y)p_Z(z)$. Es importante notar que independencia por pares de variables no implica necesariamente la independencia conjunta de las tres:

Ejemplo 15 Considere un vector aleatorio discreto (X, Y, Z) cuya función de masa de probabilidades conjunta $p_{X,Y,Z}$ está definida de la siguiente forma:

(x, y, z)	$p_{X,Y,Z}(x, y, z)$
(0, 0, 0)	1/4
(0, 0, 1)	0
(0, 1, 0)	0
(1, 0, 0)	0
(0, 1, 1)	1/4
(1, 0, 1)	1/4
(1, 1, 0)	1/4
(1, 1, 1)	0

A partir de $p_{X,Y,Z}$ se deducen tres funciones de masa de probabilidades bivariadas:

$$p_{X,Y}(x, y) = \sum_z p_{X,Y,Z}(x, y, z), \quad p_{X,Z}(x, z) = \sum_y p_{X,Y,Z}(x, y, z), \quad p_{Y,Z}(y, z) = \sum_x p_{X,Y,Z}(x, y, z),$$

	$p_{X,Y}(x, y)$	$p_{X,Z}(x, z)$	$p_{Y,Z}(y, z)$
(0, 0)	1/4	1/4	1/4
(0, 1)	1/4	1/4	1/4
(1, 0)	1/4	1/4	1/4
(1, 1)	1/4	1/4	1/4

y marginalizando las funciones de masa de probabilidades bivariadas anteriores obtenemos:

	$p_X(x)$	$p_Y(y)$	$p_Z(z)$
0	1/2	1/2	1/2
1	1/2	1/2	1/2

Esto es, (X, Y, Z) es un vector aleatorio discreto de variables aleatorias Bernoulli con parámetro igual a $\frac{1}{2}$ en los tres casos, y con lo anterior es inmediato verificar que:

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

lo que implica que X e Y son independientes, X y Z son independientes, Y y Z son independientes, por pares en los tres casos, pero no son conjuntamente independientes ya que, por ejemplo:

$$p_{X,Y,Z}(1, 1, 1) = 0 \neq \frac{1}{8} = p_X(1)p_Y(1)p_Z(1).$$

□

Con vectores aleatorios de dimensión 3 en adelante surge el concepto de *independencia condicional*. Si (X, Y, Z) es un vector aleatorio discreto con función de masa de probabilidades conjunta $p_{X,Y,Z}$, y a partir de ella deducimos las funciones de masa de probabilidades condicionales $p_{X,Y|Z}$, $p_{X|Z}$ y $p_{Y|Z}$, se dice, por ejemplo, que las variables aleatorias X e Y son *condicionalmente independientes* dado $Z = z$ si:

$$\mathbb{P}(X = x, Y = y | Z = z) = \mathbb{P}(X = x | Z = z)\mathbb{P}(Y = y | Z = z), \text{ es decir si } p_{X,Y|Z}(x, y | z) = p_{X|Z}(x | z)p_{Y|Z}(y | z).$$

En general, independencia condicional no necesariamente implica independencia no condicional, ni viceversa, como se ilustra en los siguientes ejemplos.

Ejemplo 16 Continuando con el Ejemplo 15, donde en particular se obtuvo que Y y Z son independientes. Obtenemos primero las funciones de masa de probabilidades condicionales univariadas siguientes:

	$p_{Y X}(y 0)$	$p_{Y X}(y 1)$	$p_{Z X}(z 0)$	$p_{Z X}(z 1)$
0	1/2	1/2	1/2	1/2
1	1/2	1/2	1/2	1/2

Luego las funciones de masa de probabilidades condicionales bivariadas siguientes, y las comparamos versus los productos de las condicionales univariadas anteriores:

(y, z)	$p_{Y,Z X}(y, z 0)$	$p_{Y X}(y 0)p_{Z X}(z 0)$	$p_{Y,Z X}(y, z 1)$	$p_{Y X}(y 1)p_{Z X}(z 1)$
(0, 0)	1/2	1/4	0	1/4
(0, 1)	0	1/4	1/2	1/4
(1, 0)	0	1/4	1/2	1/4
(1, 1)	1/2	1/4	0	1/4

de donde claramente $p_{Y,Z|X}(y, z | x) \neq p_{Y|X}(y | x)p_{Z|X}(z | x)$, y por tanto Y y Z no son condicionalmente independientes dado $X = x$, a pesar de que son independientes de forma no condicional.

□

Ejemplo 17 Considere un vector aleatorio discreto (X, Y, Z) cuya función de masa de probabilidades conjunta $p_{X,Y,Z}$ está definida de la siguiente forma, para algún $0 < \theta < 1$:

(x, y, z)	$p_{X,Y,Z}(x, y, z)$
(0, 0, 0)	$1 - \theta$
(0, 0, 1)	0
(0, 1, 0)	0
(1, 0, 0)	0
(0, 1, 1)	0
(1, 0, 1)	0
(1, 1, 0)	0
(1, 1, 1)	θ

A partir de $p_{X,Y,Z}$ se deducen tres funciones de masa de probabilidades bivariadas:

$$p_{X,Y}(x, y) = \sum_z p_{X,Y,Z}(x, y, z), \quad p_{X,Z}(x, z) = \sum_y p_{X,Y,Z}(x, y, z), \quad p_{Y,Z}(y, z) = \sum_x p_{X,Y,Z}(x, y, z),$$

	$p_{X,Y}(x, y)$	$p_{X,Z}(x, z)$	$p_{Y,Z}(y, z)$
(0, 0)	$1 - \theta$	$1 - \theta$	$1 - \theta$
(0, 1)	0	0	0
(1, 0)	0	0	0
(1, 1)	θ	θ	θ

y marginalizando las funciones de masa de probabilidades bivariadas anteriores obtenemos:

	$p_X(x)$	$p_Y(y)$	$p_Z(z)$
0	$1 - \theta$	$1 - \theta$	$1 - \theta$
1	θ	θ	θ

Esto es, (X, Y, Z) es un vector aleatorio discreto de variables aleatorias Bernoulli con parámetro igual a θ en los tres casos. A partir de lo anterior podemos deducir las funciones de masa de probabilidades condicionales, univariadas y bivariadas, y luego comparar el producto de las condicionales univariadas con las condicionales bivariadas:

	$p_{X Z}(x 0)$	$p_{Y Z}(y 0)$	$p_{X Z}(x 1)$	$p_{Y Z}(y 1)$
0	1	1	0	0
1	0	0	1	1

(x, y)	$p_{X,Y Z}(x, y 0)$	$p_{X Z}(x 0)p_{Y Z}(y 0)$	$p_{X,Y Z}(x, y 1)$	$p_{X Z}(x 1)p_{Y Z}(y 1)$
(0, 0)	1	1	0	0
(0, 1)	0	0	0	0
(1, 0)	0	0	0	0
(1, 1)	0	0	1	1

De esto último concluimos que $p_{X,Y|Z}(x, y|z) = p_{X|Z}(x|z)p_{Y|Z}(y|z)$, y por tanto X e Y son condicionalmente independientes dado $Z = z$. Sin embargo, al comparar la función de masa de probabilidades conjunta (no condicional) $p_{X,Y}(x, y)$ con el producto de las funciones de masa de probabilidades (no condicionales) $p_X(x)p_Y(y)$ obtenemos lo siguiente:

(x, y)	$p_{X,Y}(x, y)$	$p_X(x)p_Y(y)$
(0, 0)	$1 - \theta$	$(1 - \theta)^2$
(0, 1)	0	$(1 - \theta)\theta$
(1, 0)	0	$\theta(1 - \theta)$
(1, 1)	θ	θ^2

de donde queda claro que $p_{X,Y}(x, y) \neq p_X(x)p_Y(y)$ y, por lo tanto, X e Y no son independientes, a pesar de que sí son condicionalmente independientes.

□

2.2. Caso absolutamente continuo

Consideremos ahora el caso en que (X, Y) es un vector aleatorio absolutamente continuo con función de densidad conjunta $f_{X,Y}$, y por tanto las variables aleatorias componentes X e Y son absolutamente continuas, con funciones de densidad marginal $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dy$ y $f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dx$, respectivamente. Si para algún $A \in \mathfrak{B}(\mathbb{R})$ tenemos que $\mathbb{P}(X \in A) > 0$, entonces para cualquier $B \in \mathfrak{B}(\mathbb{R})$ la probabilidad condicional obtenida en (46) se calcula mediante:

$$\mathbb{P}(Y \in B | X \in A) = \frac{\mathbb{P}((X, Y) \in A \times B)}{\mathbb{P}(X \in A)} = \frac{\iint_{A \times B} f_{X,Y}(x, y) dy dx}{\int_A f_X(x) dx}. \quad (49)$$

En lo anterior, si en particular utilizamos $B =] - \infty, t]$ para cualquier $t \in \mathbb{R}$ entonces obtenemos la función de distribución de probabilidades de Y condicional en el evento $\{X \in A\}$:

$$F_{Y|X}(t|A) := \mathbb{P}(Y \leq t | X \in A) = \frac{1}{\mathbb{P}(X \in A)} \int_{-\infty}^t \int_A f_{X,Y}(x, y) dx dy = \frac{1}{\mathbb{P}(X \in A)} \int_{-\infty}^t g(y) dy, \quad (50)$$

donde $g(y) := \int_A f_{X,Y}(x, y) dx$. Aplicando el Teorema Fundamental de Cálculo podemos obtener la derivada de $F_{Y|X}(t|A)$:

$$\frac{d}{dt} F_{Y|X}(t|A) = \frac{1}{\mathbb{P}(X \in A)} \frac{d}{dt} \int_{-\infty}^t g(y) dy = \frac{1}{\mathbb{P}(X \in A)} g(t) = \frac{1}{\mathbb{P}(X \in A)} \int_A f_{X,Y}(x, t) dx,$$

y con ello motivar la siguiente:

Definición 16 Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad conjunta $f_{X,Y}$. Para todo $A \in \mathfrak{B}(\mathbb{R})$ tal que $\mathbb{P}(X \in A) > 0$ se denota y define la **función de densidad de probabilidades condicional** para Y dado el evento $\{X \in A\}$ como:

$$f_{Y|X}(y|A) := \frac{1}{\mathbb{P}(X \in A)} \int_A f_{X,Y}(x, y) dx,$$

y análogamente, para todo $B \in \mathfrak{B}(\mathbb{R})$ tal que $\mathbb{P}(Y \in B) > 0$ la función de densidad de probabilidades condicional para X dado el evento $\{Y \in B\}$ como:

$$f_{X|Y}(x|B) := \frac{1}{\mathbb{P}(Y \in B)} \int_B f_{X,Y}(x, y) dy.$$

□

Tanto $f_Y(y)$ como $f_{Y|X}(y|A)$ son funciones de densidad de probabilidades univariadas para la misma variable aleatoria absolutamente continua Y , con la diferencia que $f_{Y|X}(y|A)$ toma en cuenta (o supone) que el evento $\{X \in A\}$ ha ocurrido, mientras que $f_Y(y)$ no lo toma en cuenta. Nótese que en particular $f_{Y|X}(y|\mathbb{R}) = f_Y(y)$, y análogamente $f_{X|Y}(x|\mathbb{R}) = f_X(x)$.

Para el caso $A = \{x\}$ con $x \in \mathbb{R}$, el evento $\{X = x\}$ tiene probabilidad cero en el caso de variables aleatorias continuas, y por tanto la ecuación (46) y la consecuente Definición 16 no son directamente aplicables, pero puede aplicarse (46) a un conjunto $A_x(\varepsilon) := [x, x + \varepsilon]$ siempre que $\mathbb{P}(X \in A_x(\varepsilon)) > 0$ para todo $\varepsilon > 0$, y que $f_X(x) > 0$, y luego intentar calcular el límite por la derecha $\varepsilon \rightarrow 0+$ para obtener, primero, la función de distribución condicional de Y dado que $X = x$, y de ser posible, su derivada:

$$\begin{aligned} F_{Y|X}(y|x) &:= \mathbb{P}(Y \leq y | X = x) = \lim_{\varepsilon \rightarrow 0+} \mathbb{P}(Y \leq y | X \in A_x(\varepsilon)) = \lim_{\varepsilon \rightarrow 0+} \frac{\mathbb{P}(x \leq X \leq x + \varepsilon, Y \leq y)}{\mathbb{P}(x \leq X \leq x + \varepsilon)} \\ &= \lim_{\varepsilon \rightarrow 0+} \frac{F_{X,Y}(x + \varepsilon, y) - F_{X,Y}(x, y)}{F_X(x + \varepsilon) - F_X(x)} = \lim_{\varepsilon \rightarrow 0+} \frac{[F_{X,Y}(x + \varepsilon, y) - F_{X,Y}(x, y)] / \varepsilon}{[F_X(x + \varepsilon) - F_X(x)] / \varepsilon} \\ &= \frac{\frac{\partial}{\partial x} F_{X,Y}(x, y)}{\frac{d}{dx} F_X(x)} = \frac{1}{f_X(x)} \frac{\partial}{\partial x} \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(s, t) dt ds \stackrel{\text{TFC}}{=} \frac{1}{f_X(x)} \int_{-\infty}^y f_{X,Y}(x, t) dt \\ \frac{\partial}{\partial y} F_{Y|X}(y|x) &= \frac{1}{f_X(x)} \frac{\partial}{\partial y} \int_{-\infty}^y f_{X,Y}(x, t) dt \stackrel{\text{TFC}}{=} \frac{f_{X,Y}(x, y)}{f_X(x)} =: \varphi_x(y), \end{aligned} \quad (51)$$

donde las siglas TFC indican que se aplicó el Teorema Fundamental del Cálculo. Notemos que el miembro derecho de (51) es solo función de la variable y porque x es un valor dado (fijo), y que cumple las propiedades de una función de densidad de probabilidades univariada, ya que $\varphi_x(y) \geq 0$ y:

$$\int_{-\infty}^{+\infty} \varphi_x(y) dy = \frac{1}{f_X(x)} \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy = \frac{1}{f_X(x)} f_X(x) = 1,$$

lo que motiva la siguiente:

Definición 17 Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad conjunta $f_{X,Y}$. Para todo $x \in \mathbb{R}$ tal que función de densidad marginal $f_X(x) > 0$ se denota y define la **función de densidad de probabilidades condicional** para Y dado el evento $\{X = x\}$ como:

$$f_{Y|X}(y|x) := \frac{f_{X,Y}(x,y)}{f_X(x)},$$

y análogamente, para todo $y \in \mathbb{R}$ tal que la función de densidad marginal $f_Y(y) > 0$ la función de densidad de probabilidades condicional para X dado el evento $\{Y = y\}$ como:

$$f_{X|Y}(x|y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

□

Tanto $f_Y(y)$ como $f_{Y|X}(y|x)$ son funciones de densidad de probabilidades univariadas para la misma variable aleatoria absolutamente continua Y , con la diferencia que $f_{Y|X}(y|x)$ toma en cuenta (o supone) que el evento $\{X = x\}$ ha ocurrido, mientras que $f_Y(y)$ no lo toma en cuenta. Como $x \in \text{Ran } X$, el valor x se convierte en un *parámetro* de la función de densidad de probabilidades $f_{Y|X}(y|x)$, ya que valores distintos de x pudieran arrojar distintas funciones de densidad condicionales, es decir si $x_1 \neq x_2$ podría (o no) resultar que $f_{Y|X}(y|x_1) \neq f_{Y|X}(y|x_2)$.

Utilizamos $f_{Y|X}(y|x)$ en lugar de $f_Y(y)$ cuando requerimos hacer cálculo de probabilidades con la variable aleatoria absolutamente continua Y tomando en cuenta la posible influencia del evento $\{X = x\}$. Si para todo $x \in \text{Ran } X$ el evento $\{X = x\}$ no influye en el cálculo de probabilidades con la variable aleatoria Y tendríamos lo siguiente:

$$f_{Y|X}(y|x) = f_Y(y) \Leftrightarrow \frac{f_{X,Y}(x,y)}{f_X(x)} = f_Y(y) \Leftrightarrow f_{X,Y}(x,y) = f_X(x)f_Y(y), \quad (52)$$

y a la misma conclusión (52) llegaríamos partiendo de $f_{X|Y}(x|y) = f_X(x)$ para todo $y \in \text{Ran } Y$. Dado un vector aleatorio absolutamente continuo, el que los valores de una de las variables no influya en el cálculo de probabilidades de otras nos conduce al concepto de *independencia* entre variables aleatorias absolutamente continuas:

Definición 18 Dado un vector aleatorio absolutamente continuo (X, Y) con función de densidad de probabilidades conjunta $f_{X,Y}$, y funciones de densidad de probabilidades marginales f_X y f_Y , respectivamente, decimos que X e Y son **variables aleatorias independientes** si y solo si:

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

□

Si de un vector aleatorio absolutamente continuo (X, Y) conocemos su función de densidad de probabilidades conjunta $f_{X,Y}$, a partir de esta última podemos deducir las funciones de densidad de probabilidades marginales f_X y f_Y mediante (43), y solo en caso de que $f_X(x)f_Y(y) = f_{X,Y}(x,y)$ podemos entonces afirmar que X e Y son variables aleatorias independientes. En otro sentido, si solo conocemos las funciones de densidad de probabilidades marginales f_X y f_Y , solo en el caso de que sepamos además que X e Y son independientes podríamos entonces deducir que la función de densidad de probabilidades conjunta de (X, Y) es $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Ejemplo 18 Consideremos la función $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ definida como:

$$f(x,y) := e^{-y} \mathbf{1}_{\{0 < x < y\}}$$

Claramente se cumple que $f(x,y) \geq 0$ y además:

$$\iint_{\mathbb{R}^2} f(x,y) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-y} \mathbf{1}_{\{0 < x < y\}} dx dy = \int_{-\infty}^{+\infty} e^{-y} \mathbf{1}_{\{y > 0\}} \int_0^y dx = \int_0^{+\infty} ye^{-y} dy = 1,$$

donde la última igualdad se justifica porque el integrando es justamente la función de densidad de probabilidades de una variable aleatoria absolutamente continua $\text{Gamma}(2, 1)$.⁴ Entonces f cumple con las características de una

⁴La función de densidad de probabilidades Gamma con parámetros $\alpha > 0$ y $\beta > 0$ está dada por $f(z|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\beta z} \mathbf{1}_{\{z > 0\}}$.

función de densidad de probabilidades conjunta. Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad conjunta de probabilidades $f_{X,Y} = f$. Procedemos primero a deducir las funciones de densidad marginales de X e Y :

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dy = \int_{-\infty}^{+\infty} e^{-y} \mathbf{1}_{\{0 < x\}} \mathbf{1}_{\{x < y\}} dy = \mathbf{1}_{\{x > 0\}} \int_x^{+\infty} e^{-y} dy = e^{-x} \mathbf{1}_{\{x > 0\}},$$

que es la función de densidad de probabilidades de una variable aleatoria absolutamente continua Exponencial estándar (con parámetro 1).⁵

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) dx = \int_{-\infty}^{+\infty} e^{-y} \mathbf{1}_{\{0 < x < y\}} dx = e^{-y} \mathbf{1}_{\{y > 0\}} \int_0^y dx = ye^{-y} \mathbf{1}_{\{y > 0\}},$$

que es la función de densidad de probabilidades de una variable aleatoria absolutamente continua Gamma(2, 1).

$$f_X(x)f_Y(y) = ye^{-(x+y)} \mathbf{1}_{\{x > 0\}} \mathbf{1}_{\{y > 0\}} \neq e^{-y} \mathbf{1}_{\{0 < x < y\}} = f_{X,Y}(x, y)$$

de donde concluimos que X e Y no son variables aleatorias independientes. Si y es un valor dado tal que $f_Y(y) > 0$, que en este caso equivale a que $y > 0$, entonces la densidad de probabilidades de X condicional en que $Y = y$ está dada por:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{e^{-y} \mathbf{1}_{\{0 < x < y\}}}{ye^{-y}} = \frac{1}{y} \mathbf{1}_{\{0 < x < y\}},$$

que es la función de densidad de probabilidades de una variable aleatoria absolutamente continua Uniforme(0, y).⁶ Si x es un valor dado tal que $f_X(x) > 0$, que en este caso equivale a que $x > 0$, entonces la densidad de probabilidades de Y condicional en que $X = x$ está dada por:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{e^{-y} \mathbf{1}_{\{y > x\}}}{e^{-x}} = e^{-(y-x)} \mathbf{1}_{\{y > x\}},$$

que es la función de densidad de probabilidades de una variable aleatoria absolutamente continua Exponencial($x, 1$).⁷

□

Para tener idea del caso de vectores aleatorios absolutamente continuos de dimensión mayor que 2, basta con considerar, por ejemplo, el caso de un vector aleatorio (X, Y, Z) con función de densidad de probabilidades conjunta $f_{X,Y,Z}$, en cuyo caso se pueden deducir funciones de densidad de probabilidades condicionales de dimensión 2, por ejemplo:

$$f_{X,Z|Y}(x, z|y) = \frac{f_{X,Y,Z}(x, y, z)}{f_Y(y)}, \quad \forall y : f_Y(y) > 0,$$

o bien de dimensión 1, por ejemplo:

$$f_{Z|X,Y}(z|x, y) = \frac{f_{X,Y,Z}(x, y, z)}{f_{X,Y}(x, y)}, \quad \forall (x, y) : f_{X,Y}(x, y) > 0.$$

Y también, análogamente al caso discreto, las variables aleatorias X , Y y Z son conjuntamente independientes si y solo si $f_{X,Y,Z}(x, y, z) = f_X(x)f_Y(y)f_Z(z)$. De forma análoga al caso discreto, se pueden construir ejemplos de vectores aleatorios absolutamente continuos donde la independencia por pares de variables no implica necesariamente la independencia conjunta de tres.

Con vectores aleatorios de dimensión 3 en adelante surge el concepto de *independencia condicional*. Si (X, Y, Z) es un vector aleatorio absolutamente continuo con función de densidad de probabilidades conjunta $f_{X,Y,Z}$, y a partir de ella deducimos las funciones de densidad de probabilidades condicionales $f_{X,Y|Z}$, $f_{X|Z}$ y $f_{Y|Z}$, se dice, por ejemplo, que las variables aleatorias X e Y son *condicionalmente independientes* dado $Z = z$ si:

$$f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z)f_{Y|Z}(y|z).$$

De forma análoga al caso discreto, se pueden construir ejemplos que demuestran que, en general, independencia condicional no necesariamente implica independencia no condicional, ni viceversa.

⁵La función de densidad de probabilidades Exponencial con parámetro $\beta > 0$ es un caso particular Gamma(1, β), esto es: $f(z|\beta) = \beta e^{-\beta z} \mathbf{1}_{\{z > 0\}}$.

⁶La función de densidad de probabilidades Uniforme sobre un intervalo (abierto, cerrado, semiabierto) $|a, b|$ está dada por: $f(z|a, b) = \frac{1}{b-a} \mathbf{1}_{\{a < x < b\}}$.

⁷La función de densidad de probabilidades Exponencial con parámetro de localización $\mu \in \mathbb{R}$ y parámetro de escala $\theta > 0$ está dada por: $f(z|\mu, \theta) = \theta e^{-\theta(z-\mu)} \mathbf{1}_{\{z > \mu\}}$, y es resultado de la transformación $Z := \mu + W/\theta$, donde W es Exponencial estándar, esto es con función de densidad $f_W(w) = e^{-w} \mathbf{1}_{\{w > 0\}}$.

3. Transformaciones

Recordemos primero el procedimiento para transformar una variable aleatoria en otra. Dada una variable aleatoria X con determinada distribución de probabilidad (ya sea su función de distribución F_X , o bien su función de densidad o masa de probabilidades f_X , cuando sea el caso) se define otra variable aleatoria $Y := g(X)$ mediante alguna función g tal que su dominio contenga a $\text{Ran } X$. ¿Cómo deducir la distribución de probabilidad de Y ? Ya sea su función de distribución F_Y , o bien, de ser el caso, su función de densidad o masa de probabilidades f_Y . Primero, deducir el conjunto de valores que puede reportar la variable aleatoria Y , esto es:

$$\text{Ran } Y = \{g(x) : x \in \text{Ran } X\}.$$

Si $\text{Ran } Y$ resulta ser un conjunto numerable, entonces Y es evidentemente una variable aleatoria discreta, y por tanto en este caso bastará deducir su función de masa de probabilidades, para cualquier $y \in \text{Ran } Y$:

$$p_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}(g(X) = y) = \mathbb{P}(X \in B_y),$$

donde $B_y = \{x \in \text{Ran } X : g(x) = y\}$, y $\mathbb{P}(X \in B_y)$ se calcula mediante la distribución de probabilidades de X .

Ejemplo 19 Sea X una variable aleatoria discreta uniforme sobre el conjunto $\{-1, 0, 1\}$, es decir con función de masa de probabilidades:

$$p_X(x) = \mathbb{P}(X = x) = \frac{1}{3} \mathbf{1}_{\{-1, 0, 1\}}(x).$$

Si utilizamos la función $g(x) = x^2$ para definir la variable aleatoria $Y := g(X) = X^2$ entonces:

$$\text{Ran } Y = \{x^2 : x \in \{-1, 0, 1\}\} = \{0, 1\},$$

lo que implica, en este caso particular, que Y es una variable aleatoria discreta Bernoulli.

$$p_Y(1) = \mathbb{P}(Y = 1) = \mathbb{P}(X^2 = 1) = \mathbb{P}(X \in \{-1, 1\}) = p_X(-1) + p_X(1) = \frac{2}{3},$$

y por complemento $p_Y(0) = 1 - p_Y(1) = \frac{1}{3}$. Esto es, Y resulta ser una variable aleatoria Bernoulli con parámetro $\theta = \frac{2}{3}$.

□

Si, en cambio, $\text{Ran } Y$ resulta ser un conjunto no numerable, entonces queda descartado que pueda tratarse de una variable aleatoria discreta, y por tanto es continua o mixta. En este caso procedemos a deducir su función de distribución $F_Y(y) = \mathbb{P}(Y \leq y)$, y si F_Y resulta ser una función continua en todo \mathbb{R} entonces Y es variable continua, pero si existe al menos un punto de discontinuidad entonces se trata de una variable aleatoria mixta:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \in C_y), \quad y \in \mathbb{R},$$

donde $C_y := \{x \in \text{Ran } X : g(x) \leq y\}$, y $\mathbb{P}(X \in C_y)$ se calcula mediante la distribución de probabilidades de X .

Ejemplo 20 Sea X una variable aleatoria continua uniforme sobre el intervalo abierto $] -1, 1[$, es decir con función de densidad de probabilidades:

$$f_X(x) = \frac{1}{2} \mathbf{1}_{\{-1 < x < 1\}},$$

y por lo tanto con función de distribución de probabilidades:

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(t) dt = \frac{x+1}{2} \mathbf{1}_{\{-1 < x < 1\}} + \mathbf{1}_{\{x \geq 1\}}.$$

Si se define ahora la variable aleatoria $Y := X^2$ entonces:

$$\text{Ran } Y = \{x^2 : -1 < x < 1\} = [0, 1[,$$

que al ser un conjunto no numerable implica que Y es una variable aleatoria continua o mixta. Como $\text{Ran } Y = [0, 1[$, claramente si $y < 0$ entonces $F_Y(y) = 0$, y si $y \geq 1$ entonces $F_Y(y) = 1$, por lo que solo falta analizar el caso cuando $0 \leq y < 1$:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}),$$

esto es, para cualquier $y \in \mathbb{R}$:

$$F_Y(y) = \begin{cases} 0, & \text{si } y < 0, \\ F_X(\sqrt{y}) - F_X(-\sqrt{y}), & \text{si } 0 \leq y < 1, \\ 1, & \text{si } y \geq 1. \end{cases}$$

Como F_X es una función continua, y la función $h(y) = \sqrt{y}$ también, recordemos que composición de funciones continuas es continua, y por tanto cuando $0 < y < 1$ tenemos que $F_Y(y)$ es continua. Si $y < 0$ entonces $F_Y(y) = 0$ es la función constante cero, que es continua, y adicionalmente el límite por la izquierda $F_Y(0-) = 0$, mismo que coincide con $F_Y(0) = 0$, por lo que también en el punto $y = 0$ la función F_Y es continua. Si $y \geq 1$ entonces $F_Y(y) = 1$ es la función constante 1, que es continua, y adicionalmente tiene límite por la izquierda $F_Y(1-) = 1$, por lo que también en el punto $y = 1$ la función F_Y es continua. Con lo anterior concluimos que F_Y es continua en todo \mathbb{R} y por tanto Y es una variable aleatoria continua.

¿Es Y una variable aleatoria absolutamente continua? Es decir ¿es F_X derivable en todo \mathbb{R} , excepto quizás en una colección numerable de puntos aislados? Si $y < 0$ o bien $y > 1$ entonces $F_Y(y)$ es claramente derivable, con derivada cero por ser función constante en esos casos. La derivada por la izquierda en $y = 0$ claramente es cero, pero la derivada de $F_Y(y)$ por la derecha no existe porque la función $h(y) = \sqrt{y}$ no es derivable en $y = 0$. Si $0 < y < 1$ tenemos que $F_Y(y)$ sí es derivable:

$$\frac{d}{dy} F_Y(y) = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})).$$

$F_Y(y)$ no es derivable en $y = 1$ porque:

$$\frac{d}{dy} F_Y(1-) = \frac{1}{2} (f_X(1) + f_X(-1)) = \frac{1}{2} \neq 0 = \frac{d}{dy} F_Y(1+).$$

En resumen, $F_Y(y)$ es derivable en todo \mathbb{R} excepto cuando $y \in \{0, 1\}$, es decir excepto sobre una cantidad numerable de puntos aislados, por lo que se concluye, además, que Y es una variable aleatoria absolutamente continua, con función de densidad de probabilidades:

$$f_Y(y) = \mathbf{1}_{\{0 < y < 1\}} \frac{d}{dy} F_Y(y) = \frac{1}{2\sqrt{y}} \mathbf{1}_{\{0 < y < 1\}}.$$

□

Para transformaciones que involucran vectores aleatorios, primero consideraremos el caso en que se transforma un vector aleatorio (X, Y) en una variable aleatoria Z mediante alguna función $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$, esto es:

$$Z := \varphi(X, Y). \quad (53)$$

Nuevamente, el primer paso es deducir el conjunto de valores posibles para Z , esto es:

$$\text{Ran } Z = \{\varphi(x, y) : (x, y) \in \text{Ran } (X, Y)\}. \quad (54)$$

Si $\text{Ran } Z$ resulta ser un conjunto numerable, entonces Z es una variable aleatoria discreta, y en este caso basta deducir su función de masa de probabilidades:

$$p_Z(z) = \mathbb{P}(Z = z) = \mathbb{P}(\varphi(X, Y) = z) = \mathbb{P}((X, Y) \in A_z), \quad (55)$$

donde $A_z = \{(x, y) \in \text{Ran } (X, Y) : \varphi(x, y) = z\}$, y $\mathbb{P}((X, Y) \in A_z)$ se calcula mediante la distribución de probabilidad conjunta del vector aleatorio (X, Y) , ya sea su función de distribución conjunta $F_{X,Y}$, o bien, de existir, su función de densidad o de masa de probabilidad conjunta $f_{X,Y}$.

Si $\text{Ran } Z$ no es numerable, entonces Z es variable aleatoria continua o mixta. Procedemos en este caso a deducir su función de distribución de probabilidades:

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(\varphi(X, Y) \leq z) = \mathbb{P}((X, Y) \in D_z), \quad (56)$$

donde $D_z = \{(x, y) \in \text{Ran}(X, Y) : \varphi(x, y) \leq z\}$, y $\mathbb{P}((X, Y) \in D_z)$ se calcula mediante la distribución de probabilidad conjunta del vector aleatorio (X, Y) , ya sea su función de distribución conjunta $F_{X,Y}$, o bien, de existir, su función de densidad o de masa de probabilidad conjunta $f_{X,Y}$. Si F_Z resulta ser una función continua en todo \mathbb{R} entonces Z es una variable aleatoria continua, en caso contrario, existe al menos un punto de discontinuidad, y por tanto se trataría de una variable aleatoria mixta.

Ejemplo 21 Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad de probabilidades conjunta:

$$f_{X,Y}(x, y) = e^{-y} \mathbf{1}_{\{0 < x < y\}}.$$

(Se recomienda como ejercicio verificar que efectivamente $f_{X,Y}$ es una función de densidad conjunta, esto es que $f_{X,Y}(x, y) \geq 0$ para todo $(x, y) \in \mathbb{R}^2$ y que $\iint_{\mathbb{R}^2} f_{X,Y}(x, y) dx dy = 1$).

Utilizando una función $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ especificada por $\varphi(x, y) := x + y$ definimos una variable aleatoria $Z := \varphi(X, Y) = X + Y$. Como $\text{Ran}(X, Y) = \{(x, y) \in \mathbb{R}^2 : 0 < x < y\}$ entonces:

$$\text{Ran } Z = \{x + y : (x, y) \in \text{Ran}(X, Y)\} =]0, +\infty[,$$

que al resultar ser un conjunto no numerable implica que Z es una variable aleatoria continua o mixta, por lo que procedemos a deducir su función de distribución de probabilidades $F_Z : \mathbb{R} \rightarrow [0, 1]$. Como $\text{Ran } Z =]0, +\infty[$ es inmediato notar que si $z \leq 0$ entonces $F_Z(z) = 0$. Si $z > 0$ entonces:

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(X + Y \leq z) = \mathbb{P}((X, Y) \in B_z) = \iint_{B_z} f_{X,Y}(x, y) dx dy,$$

en donde:

$$B_z = \{(x, y) \in \text{Ran}(X, Y) : x + y \leq z\} = \left\{ (x, y) : 0 < x < \frac{z}{2}, x < y \leq z - x \right\},$$

por lo que:

$$F_Z(z) = \int_0^{z/2} \int_x^{z-x} e^{-y} dy dx = (1 - 2e^{-z/2} + e^{-z}) \mathbf{1}_{\{z > 0\}},$$

que claramente resulta ser una función continua, incluso en $z = 0$ porque el límite por la derecha $F_Z(0+) = 0$, de lo que se concluye que Z es una variable aleatoria continua. Más aún, si $z < 0$ claramente $F_Z(z)$ es derivable por ser la función constante cero, con derivada cero. Si $z > 0$ también es derivable:

$$\frac{d}{dz} F_Z(z) = e^{-z/2} - e^{-z} \rightarrow 0 \text{ cuando } z \rightarrow 0+$$

por lo que también es derivable en $z = 0$. Por lo anterior, deducimos además que Z es una variable aleatoria absolutamente continua, con función de densidad de probabilidades:

$$f_Z(z) = (e^{-z/2} - e^{-z}) \mathbf{1}_{\{z > 0\}}.$$

□

Aunque la explicación y ejemplo anterior fue para ilustrar la transformación de un vector aleatorio de dimensión 2 en una variable aleatoria, el procedimiento para transformar un vector aleatorio de dimensión $n > 2$ en una variable aleatoria es totalmente análogo, aunque los cálculos serán un tanto más complicados por las sumas o integrales múltiples que involucrarían.

Ahora abordaremos el caso en que se desea transformar un vector aleatorio en otro vector aleatorio. Partimos de un vector aleatorio m -dimensional $\mathbf{X} = (X_1, \dots, X_m)$, con $m \geq 2$, con función de distribución de probabilidades conjunta $F_{\mathbf{X}}(x_1, \dots, x_m)$, y de ser el caso, con función de densidad o de masa de probabilidades conjunta $f_{\mathbf{X}}(x_1, \dots, x_m)$. Por medio de una función $\psi : \mathbb{R}^m \rightarrow \mathbb{R}^n$, donde $n \geq 2$, definimos un vector aleatorio n -dimensional:

$$\mathbf{Y} = (Y_1, \dots, Y_n) := \psi(\mathbf{X}) = (\psi_1(X_1, \dots, X_m), \dots, \psi_n(X_1, \dots, X_m)), \quad (57)$$

donde ψ_1, \dots, ψ_n son las funciones componentes de ψ tales que $\psi_j : \mathbb{R}^m \rightarrow \mathbb{R}$. El objetivo es deducir la distribución de probabilidades de \mathbf{Y} . Nuevamente, lo primero es deducir el conjunto de valores que puede reportar el vector aleatorio \mathbf{Y} , esto es:

$$\text{Ran } \mathbf{Y} = \{\psi(\mathbf{x}) : \mathbf{x} \in \text{Ran } \mathbf{X}\}.$$

Si $\text{Ran } \mathbf{Y}$ resulta ser un conjunto numerable, entonces \mathbf{Y} es evidentemente un vector aleatorio discreto, y por tanto en este caso bastará deducir su función de masa de probabilidades conjunta, para cualquier $\mathbf{y} \in \text{Ran } \mathbf{Y}$:

$$p_{\mathbf{Y}}(\mathbf{y}) = \mathbb{P}(\mathbf{Y} = \mathbf{y}) = \mathbb{P}(\psi(\mathbf{X}) = \mathbf{y}) = \mathbb{P}(\mathbf{X} \in B_{\mathbf{y}}), \quad (58)$$

donde $B_{\mathbf{y}} = \{\mathbf{x} \in \text{Ran } \mathbf{X} : \psi(\mathbf{x}) = \mathbf{y}\}$, y $\mathbb{P}(\mathbf{X} \in B_{\mathbf{y}})$ se calcula mediante la distribución de probabilidades de \mathbf{X} . Si $\text{Ran } \mathbf{Y}$ no es un conjunto numerable, entonces habría que deducir su función de distribución conjunta:

$$F_{\mathbf{Y}}(\mathbf{y}) = \mathbb{P}(Y_1 \leq y_1, \dots, Y_n \leq y_n) = \mathbb{P}(\psi_1(X_1, \dots, X_m) \leq y_1, \dots, \psi_n(X_1, \dots, X_m) \leq y_n) = \mathbb{P}(\mathbf{X} \in C_{y_1, \dots, y_n}), \quad (59)$$

donde $C_{y_1, \dots, y_n} = \{\mathbf{x} \in \text{Ran } \mathbf{X} : \psi_1(\mathbf{x}) \leq y_1, \dots, \psi_n(\mathbf{x}) \leq y_n\}$, y $\mathbb{P}(\mathbf{X} \in C_{y_1, \dots, y_n})$ se calcula por medio de la distribución de probabilidades de \mathbf{X} . En el caso particular de que \mathbf{Y} resulte ser un vector aleatorio absolutamente continuo entonces su función de densidad de probabilidades conjunta se obtiene mediante:

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = \frac{\partial^n}{\partial y_1 \dots \partial y_n} F_{Y_1, \dots, Y_n}(y_1, \dots, y_n). \quad (60)$$

Ejemplo 22 Sea (X_1, X_2) un vector aleatorio absolutamente continuo con función de densidad conjunta f_{X_1, X_2} y se define un vector aleatorio:

$$(Y_1, Y_2) := (\alpha X_1 + \beta, \gamma X_2 + \delta),$$

donde las constantes $\beta, \delta \in \mathbb{R}$ pero $\alpha > 0$ y $\gamma > 0$. Tenemos entonces que:

$$\text{Ran}(Y_1, Y_2) = \{(\alpha x_1 + \beta, \gamma x_2 + \delta) : (x_1, x_2) \in \text{Ran}(X_1, X_2)\}.$$

Como (X_1, X_2) es absolutamente continuo entonces $\text{Ran}(X_1, X_2)$ es un conjunto no numerable, y en consecuencia $\text{Ran}(Y_1, Y_2)$ también, por lo que procedemos a deducir su función de distribución conjunta de probabilidades:

$$\begin{aligned} F_{Y_1, Y_2}(y_1, y_2) &= \mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2) = \mathbb{P}(\alpha X_1 + \beta \leq y_1, \gamma X_2 + \delta \leq y_2) = \mathbb{P}\left(X_1 \leq \frac{y_1 - \beta}{\alpha}, X_2 \leq \frac{y_2 - \delta}{\gamma}\right) \\ &= F_{X_1, X_2}\left(\frac{y_1 - \beta}{\alpha}, \frac{y_2 - \delta}{\gamma}\right). \end{aligned}$$

Como $f_{X_1, X_2}(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_{X_1, X_2}(x_1, x_2)$ entonces existe la función de densidad conjunta de (Y_1, Y_2) , es decir, resulta ser un vector aleatorio absolutamente continuo también:

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_{X_1, X_2}\left(\frac{y_1 - \beta}{\alpha}, \frac{y_2 - \delta}{\gamma}\right) = \frac{1}{\alpha \gamma} f_{X_1, X_2}\left(\frac{y_1 - \beta}{\alpha}, \frac{y_2 - \delta}{\gamma}\right).$$

□

Aunque en el ejemplo anterior la dimensión del vector a transformar coincide con la del vector resultante, el proceso descrito en (57), (58), (59) y (60) no requiere que m y n sean iguales, ni impone restricción alguna sobre la transformación ψ , se trata de un procedimiento totalmente general.

4. Distribuciones multivariadas

Existen diversas distribuciones de probabilidad multivariadas de uso común, discretas y absolutamente continuas, de las cuales, por su relevancia, se abordarán dos: una discreta (Multinomial) y otra absolutamente continua (Normal bivariada).

4.1. Multinomial

Un *experimento aleatorio multinomial* es una generalización del caso abordado en el Ejemplo 6. Consiste en n repeticiones independientes de un procedimiento que en cada repetición arroja alguno de entre $k \geq 2$ resultados posibles $\{a_1, \dots, a_k\}$, cada uno con probabilidad $0 < \theta_j < 1$ de que en cada repetición salga el resultado a_j , de modo que $\sum_{j=1}^k \theta_j = 1$, necesariamente. Identificaremos el espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ que representa a dicho experimento. El espacio muestral puede representarse de la siguiente forma:

$$\Omega = \{(\omega_1, \dots, \omega_n) : \omega_j \in \{a_1, \dots, a_k\}, j \in \{1, \dots, n\}\}, \quad (61)$$

con un total de $|\Omega| = k^n$ elementos, en general no equiprobable, salvo en el muy particular caso en que $\theta_j = \frac{1}{k}$ para todo $j \in \{1, \dots, k\}$. Dentro de la colección $\mathfrak{F} \subseteq \wp(\Omega)$ consideremos los siguientes nk eventos:

$$\mathcal{E}_{ij} \equiv \text{“En la } i\text{-ésima repetición sale el resultado } a_j\text{”}, \quad i \in \{1, \dots, n\}, \quad j \in \{1, \dots, k\}. \quad (62)$$

Los subconjuntos $E_{ij} \subset \Omega$ que representan a dichos eventos son:

$$E_{ij} = \{(\omega_1, \dots, \omega_n) \in \Omega : \omega_i = a_j\}, \quad (63)$$

y por tanto tienen $|E_{ij}| = k^{n-1}$ elementos cada uno. Aunque todavía no hemos deducido una fórmula general para la medida de probabilidad \mathbb{P} , por hipótesis del experimento aleatorio se deduce que debe satisfacer $\mathbb{P}(E_{ij}) = \theta_j$ para cualquier repetición $i \in \{1, \dots, n\}$. Veremos a continuación que cualquier evento simple representado como $\{\omega\}$, donde $\omega = (\omega_1, \dots, \omega_n) \in \Omega$, puede reexpresarse en términos de los subconjuntos E_{ij} . Sea la función φ tal que $\varphi(\omega_i) = j$ si y solo si $\omega_i = a_j$. Entonces:

$$\{\omega\} \equiv \{(\omega_1, \dots, \omega_n)\} = \bigcap_{i=1}^n E_{i\varphi(\omega_i)}. \quad (64)$$

Como por hipótesis se trata de repeticiones independientes, entonces los eventos representados por $E_{1\varphi(1)}, \dots, E_{n\varphi(n)}$ son independientes y, por lo tanto, la medida de probabilidad \mathbb{P} satisface:

$$\mathbb{P}(\{\omega\}) = \mathbb{P}\left(\bigcap_{i=1}^n E_{i\varphi(\omega_i)}\right) = \prod_{i=1}^n \mathbb{P}(E_{i\varphi(\omega_i)}) = \prod_{i=1}^n \theta_{\varphi(\omega_i)}. \quad (65)$$

Si denotamos como $r_j(\omega) \equiv r_j((\omega_1, \dots, \omega_n))$ el número de veces que aparece el resultado a_j en $(\omega_1, \dots, \omega_n)$, entonces $\sum_{j=1}^k r_j(\omega) = n$ y además (65) puede reexpresarse como:

$$\mathbb{P}(\{\omega\}) = \mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) = \prod_{j=1}^k \prod_{i: \varphi(\omega_i)=j} \theta_j = \prod_{j=1}^k \theta_j^{r_j(\omega)}. \quad (66)$$

Como Ω es numerable, el tener una fórmula general (66) para calcular la probabilidad de cualquier evento simple, representado por $\{\omega\}$ para cualquier $\omega \in \Omega$, implica que para cualquier evento representado por un subconjunto $A \in \mathfrak{F}$ podemos calcular su probabilidad mediante (2), es decir:

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}\{\omega\} = \sum_{\omega \in A} \prod_{j=1}^k \theta_j^{r_j(\omega)}, \quad (67)$$

y así ha quedado completamente especificado el espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ que corresponde al experimento aleatorio multinomial con parámetros $n \in \{1, 2, \dots\}$ y $(\theta_1, \dots, \theta_k)$, donde $0 < \theta_j < 1$, con $j \in \{1, \dots, k\}$ y $k \geq 2$, y con la condición $\theta_1 + \dots + \theta_k = 1$. Nótese que el caso particular $k = 2$ corresponde al experimento aleatorio binomial analizado en el Ejemplo 6.

Sobre el espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ del experimento aleatorio multinomial se pueden definir diversas variables o vectores aleatorios distintos, y cada uno inducirá su propia distribución de probabilidades, pero hay un vector aleatorio en particular que resulta con frecuencia de interés en este caso, y que es el que da origen a lo que usualmente se conoce como la *distribución de probabilidad multinomial*, misma que desarrollaremos a continuación.

Definimos el vector aleatorio $\mathbf{X} := (X_1, \dots, X_k) : \Omega \rightarrow \mathbb{R}^k$ de forma que para $j \in \{1, \dots, k\}$ cada una de las variables aleatorias componentes se define mediante:

$$X_j(\omega) := \sum_{i=1}^n \mathbf{1}_{\{\omega_i = a_j\}}, \quad (68)$$

es decir, la variable aleatoria X_j reporta el número de ocurrencias del resultado a_j en las n repeticiones independientes del experimento aleatorio multinomial. Esto implica que X_j es, de hecho, una variable aleatoria binomial (véase Ejemplo 7) con parámetros n y θ_j . Esto a su vez implica que el vector aleatorio (X_1, \dots, X_k) es discreto y, por tanto, su distribución de probabilidad queda bien representada mediante la deducción de su función de masa de probabilidades conjunta. Primero identificamos el rango de este vector aleatorio:

$$\text{Ran}(X_1, \dots, X_k) = \{(x_1, \dots, x_k) : x_j \in \{0, \dots, n\}, x_1 + \dots + x_k = n\}, \quad (69)$$

lo cual implica que $\mathbb{P}(X_1 + \dots + X_k = n) = 1$. Esta última característica implica que X_1, \dots, X_k no son variables aleatorias independientes ya que su suma siempre debe ser igual a n . Buscamos una fórmula general para la función de masa de probabilidades conjunta, dado cualquier $(x_1, \dots, x_k) \in \text{Ran}(X_1, \dots, X_k)$:

$$\begin{aligned} p_{X_1, \dots, X_k}(x_1, \dots, x_k) &= \mathbb{P}(X_1 = x_1, \dots, X_k = x_k), \\ &= \mathbb{P}(\{\omega \in \Omega : X_1(\omega) = x_1\} \cap \dots \cap \{\omega \in \Omega : X_k(\omega) = x_k\}), \\ &= \mathbb{P}(E_{x_1, \dots, x_k}), \end{aligned} \quad (70)$$

donde E_{x_1, \dots, x_k} representa una intersección de subconjuntos de \mathfrak{F} que también pertenece a \mathfrak{F} . Nótese que para cualquier $\omega \in E_{x_1, \dots, x_k}$ la probabilidad es la misma, y es $\mathbb{P}(\{\omega\}) = \theta_1^{x_1} \dots \theta_k^{x_k}$, por lo que:

$$p_{X_1, \dots, X_k}(x_1, \dots, x_k) = \sum_{\omega \in E_{x_1, \dots, x_k}} \theta_1^{x_1} \dots \theta_k^{x_k} = |E_{x_1, \dots, x_k}| \theta_1^{x_1} \dots \theta_k^{x_k}. \quad (71)$$

El número de elementos de E_{x_1, \dots, x_k} es igual al número de formas distintas en que se pueden acomodar x_1 resultados a_1 , x_2 resultados a_2 , ..., y x_k resultados a_k , en las n posiciones de cada $\omega = (\omega_1, \dots, \omega_n) \in E_{x_1, \dots, x_k}$, esto es:

$$\begin{aligned} |E_{x_1, \dots, x_k}| &= \binom{n}{x_1} \binom{n-x_1}{x_2} \binom{n-x_1-x_2}{x_3} \dots \binom{n-x_1-x_2-\dots-x_{k-1}}{x_k} \\ &= \frac{n!}{x_1!(n-x_1)!} \cdot \frac{(n-x_1)!}{x_2!(n-x_1-x_2)!} \cdot \frac{(n-x_1-x_2)!}{x_3!(n-x_1-x_2-x_3)!} \dots \\ &= \frac{n!}{x_1! x_2! \dots x_k!}. \end{aligned} \quad (72)$$

Sustituyendo (72) en (71) obtenemos la función de masa de probabilidades conjunta Multinomial($n, \theta_1, \dots, \theta_k$):

$$p_{X_1, \dots, X_k}(x_1, \dots, x_k | n, \theta_1, \dots, \theta_k) = \frac{n!}{x_1! x_2! \dots x_k!} \theta_1^{x_1} \dots \theta_k^{x_k}, \quad (73)$$

en donde para $j \in \{1, \dots, k\}$ se deben cumplir las condiciones $0 < \theta_j < 1$, $\theta_1 + \dots + \theta_k = 1$, $x_j \in \{0, \dots, n\}$, y $x_1 + \dots + x_k = n$.

Finalmente, para el caso particular $k = 2$ notemos que se obtiene:

$$p_{X_1, X_2}(x_1, x_2 | n, \theta_1, \theta_2) = \frac{n!}{x_1! x_2!} \theta_1^{x_1} \theta_2^{x_2}, \quad (74)$$

sujeta a las restricciones $\theta_1 + \theta_2 = 1$ y $x_1 + x_2 = n$, equivalentes a que $\theta_2 = 1 - \theta_1$ y $x_2 = n - x_1$, por lo que la fórmula (74) puede expresarse solo en términos de x_1 y θ_1 , es decir:

$$p_{X_1}(x_1 | n, \theta_1) = \frac{n!}{x_1! (n-x_1)!} \theta_1^{x_1} (1 - \theta_1)^{n-x_1} \mathbf{1}_{\{0, \dots, n\}}(x_1), \quad (75)$$

que es precisamente la función de masa de probabilidades de una variable aleatoria Binomial(n, θ_1), como en el Ejemplo 7.

4.2. Normal bivariada

Recordemos que una variable aleatoria absolutamente continua Z tiene distribución de probabilidad *Normal estándar* si su función de densidad es:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad z \in \mathbb{R}. \quad (76)$$

La gráfica en \mathbb{R}^2 de la función f_Z es de forma acampanada, alcanza su máximo en el punto $z = 0$, y de hecho es simétrica respecto a la recta $z = 0$. Entre algunas otras características, su esperanza es igual a cero y su varianza igual a 1 :

$$\mathbb{E}(Z) = \int_{-\infty}^{+\infty} z f_Z(z) dz = 0, \quad \mathbb{V}(Z) = \mathbb{E}[(Z - \mathbb{E}(Z))^2] = \int_{-\infty}^{+\infty} z^2 f_Z(z) dz = 1.$$

Lo anterior puede extenderse a un modelo paramétrico más general, conocido como *familia de localización y dispersión*, agregando justamente un parámetro de localización $\mu \in \mathbb{R}$ y otro de dispersión $\sigma > 0$ mediante la transformación lineal:

$$X := \mu + \sigma Z, \quad (77)$$

de donde es inmediato deducir que $\mathbb{E}(X) = \mu$ y $\mathbb{V}(X) = \sigma^2$. La función de distribución de probabilidades de Z no tiene fórmula explícita ya que no existe una fórmula explícita para la antiderivada de $e^{-z^2/2}$, pero ello no impide deducir la función de densidad de probabilidades de X :

$$F_X(x | \mu, \sigma) = \mathbb{P}(X \leq x) = \mathbb{P}(\mu + \sigma Z \leq x) = \mathbb{P}\left(Z \leq \frac{x - \mu}{\sigma}\right) = F_Z\left(\frac{x - \mu}{\sigma}\right),$$

a partir de lo cual deducimos:

$$\begin{aligned} f_X(x | \mu, \sigma) &= \frac{d}{dx} F_X(x | \mu, \sigma) = \frac{d}{dx} F_Z\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sigma} f_Z\left(\frac{x - \mu}{\sigma}\right) \\ &= \frac{1}{\sigma \sqrt{2\pi}} e^{-(x - \mu)^2 / 2\sigma^2}, \end{aligned} \quad (78)$$

a la que se le denomina función de densidad de probabilidades *Normal* con parámetros $\mu \in \mathbb{R}$ y $\sigma > 0$. Si ahora buscamos construir una función de densidad conjunta para un vector aleatorio bivariado, que tenga un comportamiento análogo al de una función de densidad Normal univariada, un caso particular y sencillo sería considerar dos variables aleatorias Z_1 y Z_2 independientes, cada una con distribución de probabilidad Normal estándar como en (76), y aplicando la Definición 18 obtenemos:

$$f_{Z_1, Z_2}(z_1, z_2) = f_{Z_1}(z_1) \cdot f_{Z_2}(z_2) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(z_1^2 + z_2^2)\right\}, \quad (79)$$

misma que graficada como superficie en \mathbb{R}^3 tiene forma acampanada, alcanzando su máximo en el punto $(z_1, z_2) = (0, 0)$, donde $f_{Z_1, Z_2}(0, 0) = \frac{1}{2\pi} < 1$. Más aún, tiene simetría radial respecto a dicho punto, ya que, para cualquier $0 < t < \frac{1}{2\pi}$, sus conjuntos de nivel son de la forma:

$$\mathcal{N}_t := \{(z_1, z_2) \in \mathbb{R}^2 : f_{Z_1, Z_2}(z_1, z_2) = t\} = \{(z_1, z_2) \in \mathbb{R}^2 : z_1^2 + z_2^2 = c_t^2\}, \quad (80)$$

donde $c_t^2 = -2 \log 2\pi t$, lo que significa que los conjuntos de nivel t son *circunferencias* con radio c_t y con centro en $(z_1, z_2) = (0, 0)$. Los resultados (79) y (80) solo son válidos si Z_1 y Z_2 son independientes, pero para un caso más general que incluya algún tipo de dependencia, se requiere añadir algún tipo de estructura y parámetro adicional. La forma más general de hacerlo es por medio de *funciones cópula*,⁸ tema que escapa al propósito del presente texto. Abordaremos una estructura particular, que tiene como consecuencia que los conjuntos de nivel sean *elipses*, recordando que una circunferencia puede verse como caso particular de una elipse.

Despejando en la Definición 17 podemos expresar cualquier función de densidad conjunta bivariada como el producto de una de sus funciones de densidad marginal por una condicional, por ejemplo:

$$f_{Z_1, Z_2}(z_1, z_2) = f_{Z_1}(z_1) \cdot f_{Z_2|Z_1}(z_2 | z_1). \quad (81)$$

⁸Para una introducción al respecto, se recomienda consultar: A. Erdelyi (2009) *Cómulas y variables aleatorias: Una introducción*. *Misceleánea Matemática* **48**, 7–28.

Si escogemos para Z_1 una distribución de probabilidad Normal estándar como en (76), y como distribución condicional para Z_2 dado $Z_1 = z_1$ una Normal($\rho z_1, \sqrt{1 - \rho^2}$) como en (78), donde $-1 < \rho < 1$ fungirá como parámetro de la función de densidad conjunta, obtenemos:

$$\begin{aligned} f_{Z_1, Z_2}(z_1, z_2 | \rho) &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z_1^2}{2}\right\} \cdot \frac{1}{\sqrt{2\pi(1 - \rho^2)}} \exp\left\{-\frac{(z_2 - \rho z_1)^2}{2(1 - \rho^2)}\right\} \\ &= \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2(1 - \rho^2)}(z_1^2 + z_2^2 - 2\rho z_1 z_2)\right\}. \end{aligned} \quad (82)$$

Por construcción, la función de densidad conjunta (82) para el vector aleatorio (Z_1, Z_2) tiene como densidad marginal para Z_1 una distribución Normal estándar univariada. Pero en (81) podríamos haber elegido también utilizar la densidad marginal de Z_2 como Normal estándar univariada, especificar de manera análoga la densidad condicional de Z_1 dado $Z_2 = z_2$, y se llegaría a la misma fórmula (82), por lo que necesariamente, y en cualquier caso, en la densidad conjunta (82) ambas densidades marginales tienen distribución Normal estándar univariada.

Notemos además que (82) es igual a (79) si y solo si $\rho = 0$, lo que implica que para un vector aleatorio (Z_1, Z_2) con función de densidad (82) las variables aleatorias Z_1 y Z_2 son independientes si y solo si $\rho = 0$. En el otro extremo, conforme $\rho \rightarrow \pm 1$ notemos que $\mathbb{V}(Z_2 | Z_1 = z_1) \rightarrow 0$ y que $\mathbb{E}(Z_2 | Z_1 = z_1) \rightarrow \pm z_1$, y viceversa si se analiza la condicional de Z_1 dado $Z_2 = z_2$. A la función de densidad conjunta (82) se le denomina función de densidad *Normal bivariada estándar* con parámetro $-1 < \rho < 1$. Es un sencillo ejercicio de geometría analítica demostrar que sus conjuntos de nivel son curvas, que en particular son elipses con ángulo de rotación de 45 grados.

Finalmente, la densidad Normal bivariada estándar (82) se puede extender al caso en que las densidades marginales tengan distribución de probabilidad Normal con distintos parámetros, aplicando la transformación (77) a ambas marginales, como en el Ejemplo 22, y así poder definir lo siguiente:

Definición 19 La función de densidad conjunta **Normal bivariada** con parámetros $\mu_X \in \mathbb{R}$, $\mu_Y \in \mathbb{R}$, $\sigma_X > 0$, $\sigma_Y > 0$, $-1 < \rho < 1$, se define para todo $(x, y) \in \mathbb{R}^2$ mediante:

$$f(x, y | \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) := \frac{\exp\left\{-\frac{1}{2(1 - \rho^2)}\left(\left(\frac{x - \mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x - \mu_X}{\sigma_X}\right)\left(\frac{y - \mu_Y}{\sigma_Y}\right) + \left(\frac{y - \mu_Y}{\sigma_Y}\right)^2\right)\right\}}{2\pi\sigma_X\sigma_Y\sqrt{1 - \rho^2}}$$

□

Con base en lo anteriormente discutido, si un vector aleatorio (X, Y) tiene función de densidad conjunta Normal bivariada con vector de parámetros $(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$ entonces la distribución marginal de X es Normal(μ_X, σ_X) y la marginal de Y es Normal(μ_Y, σ_Y). X e Y son independientes si y solo si $\rho = 0$. Si $\rho \neq 0$ es inmediato deducir que la densidad condicional de Y dado $X = x$ es Normal($\mu_Y + \rho(\sigma_Y/\sigma_X)(x - \mu_X)$, $\sigma_Y\sqrt{1 - \rho^2}$), y resultado análogo para la condicional de X dado $Y = y$. Conforme ρ se aproxima a ± 1 , las varianzas condicionales se aproximan a cero, y las densidades condicionales de $Y|X = x$ y $X|Y = y$ se concentran cada vez más cerca de sus medias, que en este caso serían las rectas $y = \mu_Y + \rho(\sigma_Y/\sigma_X)(x - \mu_X)$ y $x = \mu_X + \rho(\sigma_X/\sigma_Y)(y - \mu_Y)$, respectivamente. Es posible también extender las ideas anteriores a vectores aleatorios de dimensiones mayores que 2, y obtener una función de densidad conjunta *Normal Multivariada*, asunto que escapa al objetivo del presente texto, por lo que se sugiere consultar la bibliografía pertinente.⁹

⁹Por ejemplo: S. Kotz, N. Balakrishnan, N.L Johnson (2000) *Continuous Multivariate Distributions*. Editorial: Wiley (Nueva York).

Referencias

- [1] Bernardo JM, Smith AFM (1994) *Bayesian Theory*. Wiley.
- [2] Casella G, Berger RL (2002) *Statistical Inference*. Duxbury.
- [3] Domínguez-Martínez JI (2001) *Diseño y análisis de modelos de probabilidad*. Grupo Editorial Iberoamérica.
- [4] García-Álvarez MA (2005) *Introducción a la teoría de la probabilidad. Primer curso*. Fondo de Cultura Económica.
- [5] García-Álvarez MA (2005) *Introducción a la teoría de la probabilidad. Segundo curso*. Fondo de Cultura Económica.
- [6] Grimmett G, Stirzaker D (2004) *Probability and Random Processes*. Oxford University Press.
- [7] Mood AM, Graybill FA, Boes DC (1974). *Introduction to the Theory of Statistics*. McGraw-Hill.