

TEMA 1

VARIABLES ALEATORIAS MULTIDIMENSIONALES

Recordemos que el concepto de variable aleatoria surge de la necesidad de transformar el espacio muestral asociado a un experimento aleatorio en un espacio cuantitativo, lo que se consigue asignando un valor real a cada resultado elemental del experimento (a cada elemento del espacio muestral). Este valor se obtiene midiendo una determinada característica numérica de los resultados del experimento que describa alguna propiedad de interés.

En muchas ocasiones, para describir las propiedades de interés de los resultados de un experimento es preciso considerar varias características. Por ejemplo, en el experimento consistente en la elección de un individuo de una determinada población, se consideran las variables “altura” y “peso”. Si se extraen cinco bolas de una urna con bolas blancas, negras y rojas, podemos estar interesados en el número de bolas de cada color.

Es evidente que al considerar diversas características para describir los resultados de un experimento aleatorio (o sea, diversas variables aleatorias), estas estarán a menudo relacionadas, por lo que será conveniente realizar un estudio conjunto de ellas que refleje dichas relaciones, más que analizarlas individualmente.

De esta forma aparece el concepto de variable aleatoria multidimensional o vector aleatorio que, en términos generales, puede definirse como una función que asigna a cada elemento del espacio muestral un conjunto finito de números reales que describen el valor de cada una de las características bajo estudio en dicho elemento.

Como en el caso unidimensional, al considerar un vector aleatorio definido en relación a un experimento, los conjuntos de interés estarán definidos en términos de dicho vector y, para poder calcular sus probabilidades, es preciso exigir determinadas propiedades. La definición formal y el tratamiento de las variables aleatorias multidimensionales son análogos a los de unidimensionales.

Espacio de Borel multidimensional

Sobre el conjunto \mathbb{R}^n se define la σ -álgebra de borel \mathcal{B}^n , como la mínima clase, con estructura de σ -álgebra, que contiene a todos los intervalos de \mathbb{R}^n .

- Un intervalo de \mathbb{R}^n es un subconjunto de la forma $I_1 \times \cdots \times I_n$, donde $I_i \in \mathcal{Y} = \{\text{intervalos de } \mathbb{R}\}$.
- Notando por \mathcal{Y}^n a la clase formada por los intervalos de \mathbb{R}^n , se tiene

$$\mathcal{B}^n = \sigma(\mathcal{Y}^n)$$

$(\mathbb{R}^n, \mathcal{B}^n) \longrightarrow$ Espacio de Borel n-dimensional

$B \in \mathcal{B}^n \longrightarrow B :$ Conjunto de Borel

- Igual que en el caso unidimensional, todo intervalo, y en particular, todo punto y, por tanto, todo conjunto numerable de \mathbb{R}^n es un conjunto de Borel.

- Puede probarse que, como en el caso unidimensional, la σ -álgebra de Borel \mathcal{B}^n es la generada por intervalos de cualquier tipo y, en particular, por intervalos del tipo $(-\infty, x_1] \times \dots \times (-\infty, x_n]$, que notaremos $(-\infty, x]$, siendo $x = (x_1, \dots, x_n) \in \mathbb{R}^n$

Teorema: Caracterización de \mathcal{B}^n

\mathcal{B}^n coincide con la σ -álgebra generada por los intervalos de la forma $(-\infty, x]$, $x \in \mathbb{R}^n$.

VARIABLES ALEATORIAS MULTIDIMENSIONALES (VECTORES ALEATORIOS)

Definición: Una variable aleatoria n-dimensional sobre un espacio de probabilidad (Ω, \mathcal{A}, P) es una función $X : \Omega \rightarrow \mathbb{R}^n$ medible ($X^{-1}(B) \in \mathcal{A}$, $\forall B \in \mathcal{B}^n$)

Notación: $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$

Caracterizaciones de vectores aleatorios:

- $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$ es un vector aleatorio si y sólo si

$$X^{-1}((-\infty, x]) = \{\omega \in \Omega / X(\omega) \leq x\} = \{\omega \in \Omega / X_1(\omega) \leq x_1, \dots, X_n(\omega) \leq x_n\} \in \mathcal{A}$$

$$\forall x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

- $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$ es un vector aleatorio si y sólo si

$$\forall i = 1, \dots, n \quad X_i : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}) \text{ es variable aleatoria}$$

Dem

$$\begin{aligned} \implies] \text{ Sea } x_i \in \mathbb{R} : \quad X_i^{-1}((-\infty, x_i]) &= \{\omega \in \Omega / X_i(\omega) \leq x_i, X_j(\omega) \in \mathbb{R} \forall j \neq i\} = \\ &= X^{-1}(\mathbb{R} \times \dots \times \mathbb{R} \times (-\infty, x_i] \times \dots \times \mathbb{R}) \in \mathcal{A} \end{aligned}$$

$$\impliedby] \text{ Sea } x = (x_1, \dots, x_n) \in \mathbb{R}^n :$$

$$X^{-1}((-\infty, x]) = \{\omega \in \Omega / X_i(\omega) \leq x_i \quad \forall i = 1, \dots, n\} = \bigcap_{i=1}^n X_i^{-1}((-\infty, x_i]) \in \mathcal{A}$$

DISTRIBUCIÓN DE PROBABILIDAD DE UN VECTOR ALEATORIO

Como ocurría en el caso unidimensional, al considerar un vector aleatorio $X = (X_1, \dots, X_n)$ sobre un espacio de probabilidad, los únicos sucesos de interés son los que se expresan en términos de dicho vector:

$$\{\omega \in \Omega / X(\omega) \in B\}, \quad B \in \mathcal{B}^n$$

$$\{\omega \in \Omega / X(\omega) \in B\} = X^{-1}(B) = \{X \in B\} \in \mathcal{A}$$

y las probabilidades de estos sucesos describen completamente el comportamiento del vector X .

Definición: Dado un vector aleatorio $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B}^n)$, se denomina **distribución de probabilidad** de X o probabilidad inducida por X en $(\mathbb{R}^n, \mathcal{B}^n)$ a la función de conjunto

$$\begin{aligned} P_X = P \circ X^{-1} : \mathcal{B}^n &\longrightarrow [0, 1] \\ B &\longmapsto P_X(B) = P\{X^{-1}(B)\} = P\{X \in B\} \end{aligned}$$

Justificación: $X^{-1}(B) \in \mathcal{A}$ y se puede aplicar P .

Teorema: P_X es una medida de probabilidad sobre $(\mathbb{R}^n, \mathcal{B}^n)$

Por tanto, X transforma el espacio probabilístico original (Ω, \mathcal{A}, P) en un nuevo espacio probabilístico

$$X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B}^n, P_X)$$

y el interés se centra exclusivamente en el estudio de este nuevo espacio, o sea, en el estudio de P_X .

Este estudio se llevará a cabo, como en \mathbb{R} , a través de la función de distribución.

Función de distribución de un vector aleatorio

Definición: Dado $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B}^n, P_X)$, se define la denominada función de distribución de X como la función

$$\begin{aligned} F_X : \mathbb{R}^n &\longrightarrow [0, 1] \\ x &\longmapsto F_X(x) = P_X((-\infty, x]) = P(X^{-1}((-\infty, x])) = P\{X \leq x\} \end{aligned}$$

Dado que $x = (x_1, \dots, x_n)$, se escribe también $F_X(x_1, \dots, x_n) = P\{X_1 \leq x_1, \dots, X_n \leq x_n\}$

Teorema: Propiedades de F_X

La función de distribución de un vector aleatorio X satisface:

- 1) Es monótona no decreciente en cada argumento.
- 2) Es continua a la derecha en cada argumento.
- 3) $\exists \lim_{x_1, \dots, x_n \rightarrow +\infty} F_X(x_1, \dots, x_n) = F_X(+\infty, \dots, +\infty) = 1$
 $\exists \lim_{x_i \rightarrow -\infty} F_X(x_1, \dots, x_i, \dots, x_n) = F_X(x_1, \dots, x_{i-1}, -\infty, x_{i+1}, \dots, x_n) = 0$
 $\forall x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$
- 4) $\forall (x_1, \dots, x_n) \in \mathbb{R}^n, \forall (\epsilon_1, \dots, \epsilon_n) \in \mathbb{R}^{n+}$:

$$F_X(x_1 + \epsilon_1, \dots, x_n + \epsilon_n) - \sum_{i=1}^n F_X(x_1 + \epsilon_1, \dots, x_{i-1} + \epsilon_{i-1}, x_i, x_{i+1} + \epsilon_{i+1}, \dots, x_n + \epsilon_n) +$$

$$\begin{aligned}
 & + \sum_{\substack{i,j=1 \\ i < j}}^n F_X(x_1 + \epsilon_1, \dots, x_{i-1} + \epsilon_{i-1}, x_i, x_{i+1} + \epsilon_{i+1}, \dots, x_{j-1} + \epsilon_{j-1}, x_j, x_{j+1} + \epsilon_{j+1}, \dots, x_n + \epsilon_n) + \\
 & \qquad \qquad \qquad + \dots + (-1)^n F_X(x_1, \dots, x_n) \geq 0
 \end{aligned}$$

La demostración son análogas al caso unidimensional. Vamos a probar la 4) para el caso $n = 2$ y en general se probaría por inducción.

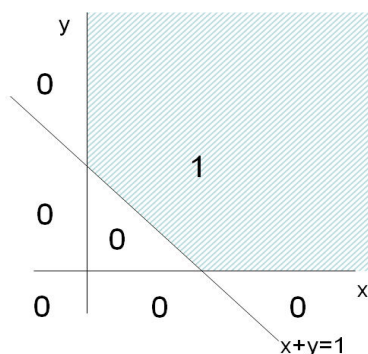
$$\begin{aligned}
 & F_X(x_1 + \epsilon_1, x_2 + \epsilon_2) - F_X(x_1, x_2 + \epsilon_2) - F_X(x_1 + \epsilon_1, x_2) + F_X(x_1, x_2) = \\
 & \quad P[X_1 \leq x_1 + \epsilon_1, X_2 \leq x_2 + \epsilon_2] - P[X_1 \leq x_1, X_2 \leq x_2 + \epsilon_2] - \\
 & \quad - (P[X_1 \leq x_1 + \epsilon_1, X_2 \leq x_2] - P[X_1 \leq x_1, X_2 \leq x_2]) = \\
 & \quad P[x_1 < X_1 \leq x_1 + \epsilon_1, X_2 \leq x_2 + \epsilon_2] - P[x_1 < X_1 \leq x_1 + \epsilon_1, X_2 \leq x_2] = \\
 & \quad P[x_1 < X_1 \leq x_1 + \epsilon_1, x_2 < X_2 \leq x_2 + \epsilon_2]
 \end{aligned}$$

Nota: Observar que, para $n = 1$, esta última propiedad es $F_X(x_1 + \epsilon_1) - F_X(x_1) \geq 0, \forall x_1 \in \mathbb{R}, \epsilon_1 \in \mathbb{R}^+$ y es, por tanto, junto con 1) generalización del no decrecimiento de funciones de distribución en \mathbb{R} .

De hecho, como probamos en el siguiente ejemplo, las propiedades 1), 2) y 3) no bastan para que F_X sea la función de distribución de un vector aleatorio.

Ejemplo

$$F(x, y) = \begin{cases} 0 & x < 0 \text{ o } y < 0 \text{ o } x + y < 1 \\ 1 & x \geq 0, y \geq 0, x + y \geq 1 \end{cases}$$



- Es no decreciente en x y y .
- Es continua a la derecha en x y y .
- $F(+\infty, +\infty) = 1$ y $F(x, -\infty) = F(-\infty, y) = 0, \forall x, y \in \mathbb{R}$

$$\blacksquare F(1, 1) - F(1, 0) - F(0, 1) + F(0, 0) = 1 - 1 - 1 + 0 = -1 < 0$$

Luego no es función de distribución, porque de ser la función de distribución de (X, Y) , la última expresión sería $P[0 < X \leq 1, 0 < Y \leq 1]$.

Nota: Estas propiedades caracterizan a las funciones de distribución multidimensionales en el sentido de que toda función de \mathbb{R}^n en \mathbb{R} que las cumpla es la función de distribución de algún vector aleatorio

Otras propiedades de la función de distribución

a) $\forall (x_1, \dots, x_n) \in \mathbb{R}^n,$

$$\exists \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} F_X(x_1, \dots, x_{i-1}, x_i - \epsilon, x_{i+1}, \dots, x_n) =$$

$$P\{X_1 \leq x_1, \dots, X_{i-1} \leq x_{i-1}, X_i < x_i, X_{i+1} \leq x_{i+1}, \dots, X_n \leq x_n\}$$

y se nota $F_X(x_1, \dots, x_{i-1}, x_i^-, x_{i+1}, \dots, x_n)$

b) $\forall (x_1, \dots, x_n) \in \mathbb{R}^n,$

$$P\{X_1 \leq x_1, \dots, X_{i-1} \leq x_{i-1}, X_i = x_i, X_{i+1} \leq x_{i+1}, \dots, X_n \leq x_n\} =$$

$$F_X(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) - F_X(x_1, \dots, x_{i-1}, x_i^-, x_{i+1}, \dots, x_n)$$

c) F_X es continua en el argumento i -ésimo en el punto $x_i \in \mathbb{R}$ si y sólo si

$$P\{X_1 \leq x_1, \dots, X_{i-1} \leq x_{i-1}, X_i = x_i, X_{i+1} \leq x_{i+1}, \dots, X_n \leq x_n\} = 0$$

La importancia de la función de distribución (al igual que en el caso unidimensional) es que determina la distribución de probabilidad y, por tanto, el comportamiento del vector aleatorio; esto es, conocida F_X se puede calcular $P(X \in B)$, $\forall B \in \mathcal{B}$. La forma de hacerlo para conjuntos de Borel arbitrarios requiere usar técnicas de integración que escapan a nuestro nivel. Sin embargo si se trabajan con intervalos, lo que generalmente ocurre en la práctica, estas probabilidades se calculan de forma simple.

Variables bidimensionales: Cálculo de probabilidades de intervalos

$$\blacksquare P\{a < X_1 \leq b, X_2 \in I\} = P\{X_1 \leq b, X_2 \in I\} - P\{X_1 \leq a, X_2 \in I\}$$

$$\blacksquare P\{a < X_1 < b, X_2 \in I\} = P\{X_1 < b, X_2 \in I\} - P\{X_1 \leq a, X_2 \in I\}$$

$$\blacksquare P\{a \leq X_1 < b, X_2 \in I\} = P\{X_1 < b, X_2 \in I\} - P\{X_1 < a, X_2 \in I\}$$

$$\blacksquare P\{a \leq X_1 \leq b, X_2 \in I\} = P\{X_1 \leq b, X_2 \in I\} - P\{X_1 < a, X_2 \in I\}$$

$$\blacksquare P\{X_1 \leq b, c < X_2 \leq d\} = P\{X_1 \leq b, X_2 \leq d\} - P\{X_1 \leq b, X_2 \leq c\} = F(b, d) - F(b, c)$$

- $P\{X_1 \leq b, c < X_2 < d\} = P\{X_1 \leq b, X_2 < d\} - P\{X_1 \leq b, X_2 \leq c\} = F(b, d^-) - F(b, c)$
- $P\{X_1 \leq b, c \leq X_2 \leq d\} = F(b, d) - F(b, c^-)$
- $P\{X_1 \leq b, c \leq X_2 < d\} = F(b, d^-) - F(b, c^-)$

- $P\{X_1 < b, c < X_2 \leq d\} = F(b^-, d) - F(b^-, c)$
- $P\{X_1 < b, c < X_2 < d\} = F(b^-, d^-) - F(b^-, c)$
- $P\{X_1 < b, c \leq X_2 \leq d\} = F(b^-, d) - F(b^-, c^-)$
- $P\{X_1 < b, c \leq X_2 < d\} = F(b^-, d^-) - F(b^-, c^-)$

Ejercicio propuesto. Dar la expresión de las siguientes probabilidades en términos de la función de distribución del vector aleatorio $X = (X_1, X_2)$:

- $P[a < X_1 < b, c < X_2 < d]$
- $P[a \leq X_1 < b, c < X_2 < d]$
- $P[a < X_1 \leq b, c < X_2 < d]$
- $P[a \leq X_1 \leq b, c < X_2 < d]$
- $P[a < X_1 < b, c \leq X_2 < d]$
- $P[a \leq X_1 < b, c \leq X_2 < d]$
- $P[a < X_1 \leq b, c \leq X_2 < d]$
- $P[a \leq X_1 \leq b, c \leq X_2 < d]$
- $P[a < X_1 < b, c < X_2 \leq d]$
- $P[a \leq X_1 < b, c < X_2 \leq d]$
- $P[a < X_1 \leq b, c < X_2 \leq d]$
- $P[a \leq X_1 \leq b, c < X_2 \leq d]$
- $P[a < X_1 < b, c \leq X_2 \leq d]$
- $P[a \leq X_1 < b, c \leq X_2 \leq d]$
- $P[a < X_1 \leq b, c \leq X_2 \leq d]$
- $P[a \leq X_1 \leq b, c \leq X_2 \leq d]$

Vectores aleatorios discretos

Definición: $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B}^n, P_X)$ es de tipo discreto si $\exists E_X \subset \mathbb{R}^n$ numerable tal que $P\{X \in E_X\} = 1$

El tratamiento de este tipo de vectores es totalmente análogo al de las variables discretas, mediante la función masa de probabilidad

$$p : E_X \longrightarrow [0, 1]$$

$$x_n \longmapsto P\{X = x_n\} = p_n$$

que satisface $p_n \geq 0$ y $\sum_{x \in E_X} p_n = 1$ y determina completamente la distribución del vector y, por tanto, su función de distribución:

- $P_X(B) = P\{X \in B\} = \sum_{x \in B \cap E_X} P\{X = x\} \quad \forall B \in \mathcal{B}^n$
- $F_X(x) = \sum_{\substack{x_j \in E_X \\ x_j \leq x}} P\{X = x_j\} \quad \forall x \in \mathbb{R}^n$

Teorema.- Toda colección numerable de números no negativos y de suma la unidad constituyen los valores de la f.m.p. de algún vector aleatorio n-dimensional de tipo discreto.

Ya hemos visto que un vector aleatorio no es más que un conjunto de variables aleatorias unidimensionales. Vemos a continuación que los vectores discretos están formados por variables aleatorias discretas y recíprocamente.

Teorema: Caracterización de vectores aleatorios discretos por sus componentes

$X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B}^n, P_X)$ es discreto si y sólo si $\forall i = 1, \dots, n$ las componentes $X_i : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ son discretas

\implies] Sea E_X el conjunto de valores de X : $P\{X \in E_X\} = 1$ y sea

$$E_X^i = \{x_i \in \mathbb{R} / \exists x \in E_X : (x)_i = x_i\} \rightarrow \text{PROYECCIÓN DE } E_x \text{ sobre el lado } i$$

Es evidente que, puesto que E_X es numerable, E_X^i también lo es y

$$P\{X_i \in E_X^i\} = P\{X_1 \in \mathbb{R}, \dots, X_{i-1} \in \mathbb{R}, X_i \in E_X^i, X_{i+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R}\} \geq P\{X \in E_X\} = 1$$

Por tanto, X_i es también de tipo discreto.

\implies] Supongamos ahora que X_i es discreta $\forall i = 1, \dots, n$ y $P\{X_i \in E_i\} = 1$.

Entonces, puesto que E_1, \dots, E_n son numerables, es claro que $E_1 \times \dots \times E_n \subset \mathbb{R}^n$ es también numerable y

$$P\{X \in E_1 \times \dots \times E_n\} \geq P\{X_1 \in E_1, \dots, X_n \in E_n\} = 1$$

¹Si A_1, \dots, A_n son sucesos seguros $\implies P(\bigcap_{i=1}^n A_i) = 1$

$$P\left(\bigcap_{i=1}^n A_i\right) = 1 - P\left(\bigcup_{i=1}^n A_i^c\right) \underset{\text{subadit}}{\geq} 1 - \sum_{i=1}^n P(A_i^c) = 1 - 0 = 1$$

Se deduce entonces que un vector aleatorio n -dimensional de tipo discreto no es más que un conjunto de n variables aleatorias unidimensionales de tipo discreto.

Ejemplo: Se lanza un dado equilibrado y se observa la cara superior. Se definen las variables aleatorias

$$X_1 = \begin{cases} -1 & \text{si el resultado es impar} \\ 1 & \text{si el resultado es par} \end{cases} \quad X_2 = \begin{cases} -2 & \text{si sale } 1, 2, 3 \\ 0 & \text{si sale } 4 \\ 3 & \text{si sale } 5, 6 \end{cases}$$

Determinar la función masa de probabilidad y la función de distribución de (X_1, X_2) .

$$E_X = \{(-1, -2), (-1, 3), (1, -2), (1, 0), (1, 3)\}$$

Función masa de probabilidad

X_2	-2	0	3
X_1	-1	2/6	0
	1	1/6	1/6

Función de distribución

$$F_X(x_1, x_2) = \begin{cases} 0 & x_1 < -1 \text{ o } x_2 < -2 \\ 2/6 & -1 \leq x_1 < 1, -2 \leq x_2 < 3 \\ 3/6 & -1 \leq x_1 < 1, x_2 \geq 3 \\ 3/6 & x_1 \geq 1, -2 \leq x_2 < 0 \\ 4/6 & x_1 \geq 1, 0 \leq x_2 < 3 \\ 1 & x_1 \geq 1, x_2 \geq 3 \end{cases}$$

Vectores aleatorios continuos

Definición: $X = (X_1, \dots, X_n) : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B}^n, P_X)$ es de tipo continuo si $\exists f_X : \mathbb{R}^n \longrightarrow \mathbb{R}$ no negativa e integrable tal que

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \quad \forall x \in \mathbb{R}^n$$

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \int_{-\infty}^{x_{n-1}} \dots \int_{-\infty}^{x_1} f_X(t_1, \dots, t_n) dt_1 \dots dt_n \quad \forall x_1, \dots, x_n \in \mathbb{R}$$

La función f_X se denomina función de densidad del vector aleatorio X y, como vemos en la definición, f_X determina F_X y, por tanto, P_X :

$$P_X(B) = P\{X \in B\} = \int_B f_X(t) dt$$

lo cual implica, que si E es numerable, $P\{X \in E\} = 0$.

Además, f_X tiene propiedades análogas a las del caso unidimensional.

Propiedades de f_X

- 1) f_X es no negativa, integrable y $\int_{-\infty}^{+\infty} f_X(t)dt = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_X(t_1, \dots, t_n)dt_1 \cdots dt_n = 1$.
- 2) f_X es continua salvo en un conjunto con medida de Lebesgue nula, y en los puntos de continuidad de f_X , F_X es derivable y

$$\frac{\partial^n F_X(x_1, \dots, x_n)}{\partial x_1 \cdots \partial x_n} = f_X(x_1, \dots, x_n)$$

- 3) f_X puede ser cambiada en conjuntos de medida nula sin afectar a la integral, o sea, a F_X . Por tanto, notamos que f_X no es única. Elegiremos siempre una versión de ella que sea, en la medida de lo posible, continua.

Teorema. Toda función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ no negativa, integrable y tal que $\int_{\mathbb{R}^n} f(t)dt = 1$ es la función de densidad de algún vector aleatorio n-dimensional de tipo continuo.

En el caso discreto hemos visto que los vectores pueden caracterizarse como conjuntos de variables aleatorias unidimensionales discretas. Sin embargo, en el caso continuo esto no es cierto.

Si bien las componentes de un vector aleatorio continuo son también de tipo continuo, el recíproco no es cierto. Probamos lo primero en la siguiente proposición y lo segundo con un contraejemplo.

Proposición: Si $X = (X_1, \dots, X_n) : (\Omega; \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}^n, P_X)$ es un vector aleatorio de tipo continuo, entonces $X_i : (\Omega; \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}, P_{X_i})$ es de tipo continuo $\forall i = 1, \dots, n$.

Dem.- Fijado $i = 1, \dots, n$, $F_{X_i}(x_i) = P[X_i \leq x_i] = P[X_1 \in \mathbb{R}, \dots, X_{i-1} \in \mathbb{R}, X_i \leq x_i, X_{i+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R}] =$

$$\lim_{\substack{x_j \rightarrow +\infty \\ j \neq i}} P[X_1 \leq x_1, \dots, X_{i-1} \leq x_{i-1}, X_i \leq x_i, X_{i+1} \leq x_{i+1}, \dots, X_n \leq x_n] =$$

$$\begin{aligned} & F_X(+\infty, \dots, +\infty, x_i, +\infty, \dots, +\infty) = \\ & \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \int_{-\infty}^{x_i} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(t_1, \dots, t_i, \dots, t_n) dt_1 \cdots dt_n = \\ & \int_{-\infty}^{x_i} \left[\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(t_1, \dots, t_i, \dots, t_n) dt_1 \cdots dt_{i-1} dt_{i+1} \cdots dt_n \right] dt_i \end{aligned}$$

Entonces si definimos

$$f_i(t_i) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(t_1, \dots, t_i, \dots, t_n) dt_1 \cdots dt_{i-1} dt_{i+1} \cdots dt_n$$

se cumple que f_i es una función real de variable real, no negativa, integrable y

$$F_{X_i}(x_i) = \int_{-\infty}^{x_i} f_i(t_i) dt_i$$

Por tanto, X_i es de tipo continuo, con función de densidad f_i .

Ejemplo: Variables continuas no tienen por qué componer un vector continuo.

Sea X_1 una variable aleatoria unidimensional de tipo continuo y $X_2 = X_1$. Supongamos que el vector $X = (X_1, X_2)$ es de tipo continuo. En tal caso, tendrá una función de densidad $f(x_1, x_2)$ y podemos calcular la probabilidad de que X pertenezca a cualquier conjunto integrando en él.

$$P\{X_1 = X_2\} = \int_{-\infty}^{+\infty} \int_{x_2}^{x_2} f(x_1, x_2) dx_1 dx_2 = 0$$

Sin embargo, es evidente que $P\{X_1 = X_2\} = 1$. Por tanto, deducimos que (X_1, X_2) no puede ser de tipo continuo.

Ejemplo 1: Calcular la función de distribución de un vector aleatorio (X, Y) con función de densidad

$$f(x, y) = e^{-x-y}, \quad x > 0, y > 0$$

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) dt ds = \begin{cases} 0 & x < 0 \text{ o } y < 0 \\ \int_0^x \int_0^y e^{-s-t} dt ds = \left(\int_0^x e^{-s} ds\right) \left(\int_0^y e^{-t} dt\right) = (1 - e^{-x})(1 - e^{-y}) & x, y \geq 0 \end{cases}$$

Ejemplo 2: Sea (X, Y) un vector aleatorio con función de densidad

$$f(x, y) = 1 \quad 0 \leq x, y \leq 1$$

Calcular:

- $P\{X + Y \leq 1\}$
- $P\{1/3 \leq X + Y \leq 3/2\}$
- $P\{X \geq 2Y\}$

a)

$$P\{X + Y \leq 1\} = \int_0^1 \int_0^{1-y} dx dy = \int_0^1 (1 - y) dy = (y - y^2/2) \Big|_0^1 = 1/2$$

o bien

$$\int_0^1 \int_0^{1-x} dy dx = \int_0^1 (1 - x) dx = (x - x^2/2) \Big|_0^1 = 1/2$$

b)

$$P\{1/3 \leq X + Y \leq 3/2\} = \int_0^{1/3} \int_{1/3-x}^1 dydx + \int_{1/3}^{1/2} \int_0^1 dydx + \int_{1/2}^1 \int_0^{3/2-y} dydx = \frac{59}{72}$$

o bien

$$1 - \int_0^{1/3} \int_0^{1/3-y} dx dy - \int_{1/2}^1 \int_{3/2-y}^1 dx dy = \frac{59}{72}$$

c)

$$P\{X \geq 2Y\} = \int_0^{1/2} \int_{2y}^1 dx dy = \int_0^1 \int_0^{x/2} dy dx = \frac{1}{4}$$

Distribuciones marginales

Al considerar un vector aleatorio como conjunto de variables aleatorias unidimensionales la distribución del vector se denomina distribución conjunta (función de distribución conjunta, función masa de probabilidad conjunta o función de densidad conjunta) y a la distribución de cada componente se le denomina distribución marginal de dicha componente.

Veamos a continuación cómo pueden obtenerse las distribuciones marginales de cada componente a partir de la conjunta.

Función de distribución marginal

Si $X = (X_1, \dots, X_n)$ es un vector aleatorio con función de distribución F_X

$$\forall i = 1, \dots, n \quad F_{X_i}(x_i) = F_X(+\infty, \dots, +\infty, x_i, +\infty, \dots, +\infty) \quad \forall x_i \in \mathbb{R}$$

Análogamente,

$$F_{X_{i_1}, \dots, X_{i_k}}(x_{i_1}, \dots, x_{i_k}) = F_X(+\infty, \dots, +\infty, x_{i_1}, +\infty, \dots, +\infty, x_{i_k}, +\infty, \dots, +\infty)$$

Sin embargo, ya que nosotros trabajamos con vectores discretos o continuos cuyas componentes, como hemos visto, son variables aleatorias discretas o continuas, lo que nos interesa es el cálculo de las f.m.p o funciones de densidad marginales a partir de la conjunta.

Distribuciones marginales de vectores discretos

Sea $X = (X_1, \dots, X_n)$ un vector aleatorio discreto con $P\{X \in E_X\} = 1$ y función masa de probabilidad conocida: $P\{X = x\} \quad \forall x \in E_X$.

Si X_i es una componente arbitraria (ya sabemos que es discreta y que toma valores en el conjunto $E_{X_i} = \{x_i \in \mathbb{R} / \exists x \in E_X : (x)_i = x_i\}$) su función masa de probabilidad puede obtenerse a partir de la de X por la siguiente relación

$$P\{X_i = x_i\} = \sum_{\substack{x \in E_X \\ (x)_i = x_i}} P\{X = x\} = \sum_{\substack{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n \\ (x_1, \dots, x_i, \dots, x_n) \in E_X}} P\{X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i = x_i, X_{i+1} = x_{i+1}, \dots, X_n = x_n\}$$

relación que se obtiene inmediatamente de $P\{X_i = x_i\} = P\{X_i = x_i, X \in E_X\}$.

Análogamente se obtiene la función masa de probabilidad de cualquier subvector que será también de tipo discreto puesto que todas sus componentes los son.

$$P\{X_{i_1} = x_{i_1}, \dots, X_{i_k} = x_{i_k}\} = \sum_{\substack{x \in E_X \\ (x)_{i_j} = x_{i_j}, j=1, \dots, k}} P\{X = x\} = \sum_{\substack{x_1, \dots, x_{i_1-1}, x_{i_1+1}, \dots, x_{i_k-1}, x_{i_k+1}, \dots, x_n \\ \dots, x_{i_k-1}, x_{i_k+1}, \dots, x_n}} P\{X_1 = x_1, \dots, X_{i_1-1} = x_{i_1-1}, X_{i_1} = x_{i_1}, X_{i_1+1} = x_{i_1+1}, \dots, X_{i_k-1} = x_{i_k-1}, X_{i_k} = x_{i_k}, X_{i_k+1} = x_{i_k+1}, \dots, X_n = x_n\}$$

Ejemplo: Se lanza una moneda tres veces y se define X : Número de caras e Y : Diferencia, en valor absoluto, de número de caras y cruces. Calcular las distribuciones marginales a partir de la conjunta.

La función masa de probabilidad conjunta y marginales son (sin más que sumar por filas y por columnas)

	X	0	1	2	3	
Y						
	1	0	3/8	3/8	0	6/8
	3	1/8	0	0	1/8	2/8
		1/8	3/8	3/8	1/8	1

Distribuciones marginales de vectores continuos

Sea $X = (X_1, \dots, X_n)$ un vector aleatorio de tipo continuo con función de densidad f_X conocida. En el apartado anterior vimos que cada componente X_i es de tipo continuo y su función de distribución es

$$F_{X_i}(x_i) = \int_{-\infty}^{x_i} f_i(t_i) dt_i$$

con

$$f_i(t_i) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_X(t_1, \dots, t_i, \dots, t_n) dt_1 \dots dt_{i-1} dt_{i+1} \dots dt_n \quad \forall t_i \in \mathbb{R}$$

Por tanto, esta es la función de densidad de X_i que, como observamos se obtiene integrando la conjunta en el resto de las componentes, fijando en la componente i -ésima el punto de interés t_i .

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

$$f(x, y) = 10x^2y \quad 0 \leq y \leq x \leq 1$$

Calcular las marginales y $P\{Y \leq 1/2\}$

$$f_1(x) = \int_0^x 10x^2y dy = 10x^2 \frac{y^2}{2} \Big|_0^x = 5x^4 \quad 0 \leq x \leq 1$$

$$f_2(y) = \int_y^1 10x^2 y dx = 10y \frac{x^3}{3} \Big|_y^1 = \frac{10}{3} y(1 - y^3) \quad 0 \leq y \leq 1$$

La probabilidad $P\{Y \leq 1/2\}$ se puede calcular de dos formas, con la marginal

$$P\{Y \leq 1/2\} = \int_0^{1/2} \frac{10}{3} y(1 - y^3) dy = \frac{19}{48}$$

o con la conjunta,

$$P\{Y \leq 1/2\} = \int_0^{1/2} \int_0^x 10x^2 y dy dx + \int_{1/2}^1 \int_0^{1/2} 10x^2 y dy dx = \frac{19}{48}$$

o bien,

$$P\{Y \leq 1/2\} = \int_0^{1/2} \int_y^1 10x^2 y dx dy = \frac{19}{48}$$

Con un razonamiento similar se obtiene la distribución marginal de un subvector (de dimensión mayor que uno) de un vector continuo

$$F_{X_{i_1}, \dots, X_{i_k}}(x_{i_1}, \dots, x_{i_k}) = F_X(+\infty, \dots, +\infty, x_{i_1}, \dots, x_{i_k}, +\infty, \dots, +\infty) = \int_{-\infty}^{x_{i_1}} \dots \int_{-\infty}^{x_{i_k}} \left[\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_X(t_1, \dots, t_n) dt_n \dots dt_{i_k+1} dt_{i_k-1} \dots dt_{i_1+1} dt_{i_1-1}, \dots, dt_1 \right]$$

de lo que se deduce

$$f_{i_1, \dots, i_k}(t_{i_1}, \dots, t_{i_k}) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_X(t_1, \dots, t_n) dt_n \dots dt_{i_k+1} dt_{i_k-1} \dots dt_{i_1+1} dt_{i_1-1}, \dots, dt_1$$

Distribuciones condicionadas

En el análisis conjunto de variables aleatorias surge una importante cuestión, como es el estudio del comportamiento de un subconjunto de ellas, cuando el resto está sujeto a alguna condición y, en particular, cuando se conoce su valor. Aparece así, el concepto de distribución condicionada.

Analizamos a continuación cómo pueden obtenerse las distribuciones condicionadas a partir de la conjunta distinguiendo entre vectores discretos y continuos.

Distribuciones condicionadas de vectores discretos

Caso bidimensional: $X = (X_1, X_2)$

Si tenemos un vector aleatorio $X = (X_1, X_2)$ definido en relación a un experimento aleatorio y se sabe que, por ejemplo, la variable X_1 ha tomado un determinado valor, x_1^* , ($P\{X_1 = x_1^*\} \neq 0$), este conocimiento afecta a la distribución de la variable X_2 . De hecho, las probabilidades

de los distintos valores de X_2 habrá que calcularlas teniendo en cuenta que $X_1 = x_1^*$ y serán, por tanto, distribuciones condicionadas:

$$P\{X_2 = x_2 \mid X_1 = x_1^*\} = \begin{cases} \frac{P\{X_1 = x_1^*, X_2 = x_2\}}{P\{X_1 = x_1^*\}} & \text{si } x_2 \in E_{X_2} \\ 0 & \text{si } x_2 \notin E_{X_2} \end{cases}$$

Notemos que $P\{X_2 = x_2 \mid X_1 = x_1^*\} \geq 0$ y $\sum_{x_2 \in E_{X_2}} P\{X_2 = x_2 \mid X_1 = x_1^*\} = 1$, por lo que estos valores proporcionan una función masa de probabilidad sobre E_{X_2} .

La distribución determinada por esta f.m.p. se denomina **distribución condicionada** de X_2 a $X_1 = x_1^*$ (distribución de X_2 dado $X_1 = x_1^*$).

Notemos que $\forall x_1^* / P\{X_1 = x_1^*\} > 0$ puede definirse la distribución de X_2 dado $X_1 = x_1^*$ y, de forma análoga, $\forall x_2^* / P\{X_2 = x_2^*\} > 0$ se define la distribución condicionada de X_1 a $X_2 = x_2^*$.

Caso general: $X = (X_1, \dots, X_n)$

La definición de distribuciones condicionadas en el caso bidimensional se generaliza de forma inmediata al caso de vectores de dimensión arbitraria, caso en el que puede condicionarse bien al valor de una sola variable o de varias.

Definición: Distribución condicionada al valor de una variable

Sea $X = (X_1, \dots, X_n)$ un vector aleatorio discreto. Sea X_i una componente arbitraria y $x_i^* \in \mathbb{R} / P\{X_i = x_i^*\} > 0$. Se define la distribución condicionada de $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ a $\{X_i = x_i^*\}$ (distribución de $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ dado $X_i = x_i^*$) como la determinada por la función masa de probabilidad:

$$P\{X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \dots, X_n = x_n \mid X_i = x_i^*\} = \frac{P\{X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i = x_i^*, X_{i+1} = x_{i+1}, \dots, X_n = x_n\}}{P\{X_i = x_i^*\}} \\ \forall (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) / (x_1, \dots, x_{i-1}, x_i^*, x_{i+1}, \dots, x_n) \in E_X$$

EJEMPLO: Se lanza una moneda tres veces y se define X : número de caras e Y : diferencia, en valor absoluto, entre el número de caras y cruces. Calcular las distribuciones condicionadas de X a $Y = 1$, de X a $Y = 3$ y de Y a $X = 0$.

Y	X				
	0	1	2	3	
1	0	3/8	3/8	0	6/8
3	1/8	0	0	1/8	2/8
	1/8	3/8	3/8	1/8	1

- $X|Y = 1$

$P\{X = 0|Y = 1\} = 0$, $P\{X = 1|Y = 1\} = 1/2$, $P\{X = 2|Y = 1\} = 1/2$, $P\{X = 3|Y = 1\} = 0$

- $X|Y = 3$

$$P\{X = 0|Y = 3\} = 1/2, P\{X = 1|Y = 3\} = 0, P\{X = 2|Y = 3\} = 0, P\{X = 3|Y = 3\} = 1/2$$

- $Y|X = 0$

$$P\{Y = 1|X = 0\} = 0, P\{Y = 3|X = 0\} = 1 \rightarrow \text{Degenerada en } 3$$

Definición: Distribución condicionada a valores de varias variables

Sea $X = (X_1, \dots, X_n)$ un vector aleatorio discreto y sea $(X_{i_1}, \dots, X_{i_k})$ un subvector arbitrario y $(x_{i_1}^*, \dots, x_{i_k}^*) \in \mathbb{R}^k / P\{X_{i_1} = x_{i_1}^*, \dots, X_{i_k} = x_{i_k}^*\} > 0$. Se define la distribución condicionada de $(X_1, \dots, X_{i_1-1}, X_{i_1+1}, \dots, X_{i_k-1}, X_{i_k+1}, \dots, X_n)$ a $\{X_{i_1} = x_{i_1}^*, \dots, X_{i_k} = x_{i_k}^*\}$ como la determinada por la función masa de probabilidad:

$$\begin{aligned} & P\{X_1 = x_1, \dots, X_{i_1-1} = x_{i_1-1}, X_{i_1+1} = x_{i_1+1}, \dots, \\ & X_{i_k-1} = x_{i_k-1}, X_{i_k+1} = x_{i_k+1}, \dots, X_n = x_n / X_{i_1} = x_{i_1}^*, \dots, X_{i_k} = x_{i_k}^*\} = \\ & \frac{P\{X_1 = x_1, \dots, X_{i_1} = x_{i_1}^*, \dots, X_{i_k} = x_{i_k}^*, \dots, X_n = x_n\}}{P\{X_{i_1} = x_{i_1}^*, \dots, X_{i_k} = x_{i_k}^*\}} \\ & \forall (x_1, \dots, x_{i_1-1}, x_{i_1+1}, \dots, x_{i_k-1}, x_{i_k+1}, \dots, x_n) / (x_1, \dots, x_{i_1}^*, \dots, x_{i_k}^*, \dots, x_n) \in E_X \end{aligned}$$

Distribuciones condicionadas de vectores continuos

Si $X = (X_1, \dots, X_n)$ es un vector aleatorio continuo, $\forall i = 1, \dots, n$, X_i es continua y $P\{X_i = x_i^*\} = 0 \forall x_i^* \in \mathbb{R}$. No es posible, por tanto, definir la probabilidad condicionada al suceso $\{X_i = x_i^*\}$ como en el caso discreto.

Para obtener de forma rigurosa las distribuciones condicionadas debe realizarse un procedimiento de paso al límite que escapa a los contenidos de este curso. Veamos simplemente las definiciones.

Caso bidimensional

Sea $X = (X_1, X_2)$ un vector aleatorio de tipo continuo con función de densidad f_X . Sea $x_2^* \in \mathbb{R}$ tal que $f_{X_2}(x_2^*) > 0$. Se define la distribución condicionada de X_1 a $\{X_2 = x_2^*\}$ (distribución de X_1 dado $X_2 = x_2^*$) como la determinada por la función de densidad

$$f_{X_1|X_2=x_2^*}(x_1/x_2^*) = \frac{f_X(x_1, x_2^*)}{f_{X_2}(x_2^*)}$$

Caso general: $X = (X_1, \dots, X_n)$

Definición 1: Distribución condicionada al valor de una variable

Sea $X = (X_1, \dots, X_n)$ un vector aleatorio de tipo continuo con función de densidad f_X . Sea X_i una componente arbitraria y $x_i^* \in \mathbb{R}$ tal que $f_{X_i}(x_i^*) > 0$. Se define la distribución condicionada de $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ a $\{X_i = x_i^*\}$ (distribución de $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ dado $X_i = x_i^*$) como la determinada por la función de densidad

$$f_{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n | X_i = x_i^*}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n | x_i^*) = \frac{f_X(x_1, \dots, x_i^*, \dots, x_n)}{f_{X_i}(x_i^*)}$$

Definición 2: Distribución condicionada a valores de varias variables

Sea $X = (X_1, \dots, X_n)$ un vector aleatorio de tipo continuo con función de densidad f_X y sea $(X_{i_1}, \dots, X_{i_k})$ un subvector arbitrario y $(x_{i_1}^*, \dots, x_{i_k}^*) \in \mathbb{R}^k / f_{X_{i_1}, \dots, X_{i_k}}(x_{i_1}^*, \dots, x_{i_k}^*) > 0$. Se define la distribución condicionada de $(X_1, \dots, X_{i_1-1}, X_{i_1+1}, \dots, X_{i_k-1}, X_{i_k+1}, \dots, X_n)$ a $\{X_{i_1} = x_{i_1}^*, \dots, X_{i_k} = x_{i_k}^*\}$ como la determinada por la función de densidad:

$$f_{X_1, \dots, X_{i_1-1}, X_{i_1+1}, \dots, X_{i_k-1}, X_{i_k+1}, \dots, X_n | X_{i_1} = x_{i_1}^*, \dots, X_{i_k} = x_{i_k}^*}(x_1, \dots, x_{i_1-1}, x_{i_1+1}, \dots, x_{i_k-1}, x_{i_k+1}, \dots, x_n / x_{i_1}^*, \dots, x_{i_k}^*) = \\ = \frac{f_X(x_1, \dots, x_{i_1}^*, \dots, x_{i_k}^*, \dots, x_n)}{f_{X_{i_1}, \dots, X_{i_k}}(x_{i_1}^*, \dots, x_{i_k}^*)}$$

EJEMPLO: Sea (X, Y) un vector aleatorio con función de densidad

$$f(x, y) = 2, \quad 0 < x < y < 1$$

Calcular las funciones de densidad y de distribución condicionadas. Calcular también las probabilidades

$$P\{Y \geq 1/2 / X = 1/2\} \quad P\{X \geq 1/3 / Y = 2/3\}$$

DISTRIBUCIÓN DE $X/Y = y_0$

$$f_2(y) = \int_0^y 2dx = 2y \quad 0 < y < 1$$

Si $0 < y_0 < 1$

$$f_{X/Y=y_0}(x/y_0) = f_{X/Y=y_0}(x) = \frac{f(x, y)}{f_Y(y_0)} = \frac{2}{2y_0} = \frac{1}{y_0} \quad 0 < x < y_0$$

es decir,

$$X/Y = y_0 \sim U(0, y_0)$$

A partir de la función de densidad condicionada podemos calcular la función de distribución condicionada

$$F_{X/Y=y_0}(x/y_0) = F_{X/Y=y_0}(x) = \begin{cases} 0 & x < 0 \\ \int_0^x \frac{1}{y_0} dt = \frac{x}{y_0} & 0 \leq x < y_0 \\ 1 & x \geq y_0 \end{cases}$$

DISTRIBUCIÓN DE $Y/X = x_0$

$$f_1(x) = \int_x^1 2dy = 2 - 2x \quad 0 < x < 1$$

Si $0 < x_0 < 1$

$$f_{Y/X=x_0}(y/x_0) = f_{Y/X=x_0}(y) = \frac{f(x_0, y)}{f_1(x_0)} = \frac{2}{2-2x_0} = \frac{1}{1-x_0} \quad x_0 < y < 1$$

es decir,

$$Y/X = x_0 \sim U(x_0, 1)$$

A partir de la función de densidad condicionada podemos calcular la función de distribución condicionada

$$F_{Y/X=x_0}(y/x_0) = F_{Y/X=x_0}(y) = \begin{cases} 0 & y < x_0 \\ \frac{y-x_0}{1-x_0} & x_0 \leq y < 1 \\ 1 & y \geq 1 \end{cases}$$

Por otra parte

$$P\{Y \geq 1/2 / X = 1/2\} = \int_{\{y \geq 1/2\}} f_{Y/X=1/2}(y) dy = 1 \quad \text{dado que } Y/X = 1/2 \sim U(1/2, 1)$$

$$P\{X \geq 1/3 / Y = 2/3\} = \int_{\{x \geq 1/3\}} f_{X/Y=2/3}(x) dx = \int_{1/3}^{2/3} \frac{3}{2} dx = \frac{3}{2} \frac{1}{3} = \frac{1}{2}$$

dado que $X/Y = 2/3 \sim U(0, 2/3)$. También se podían haber calculado a partir de las correspondientes funciones de distribución.

Funciones de vectores aleatorios: Cambio de variable

Como en el caso unidimensional, si X es un vector aleatorio n -dimensional y $g : (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\mathbb{R}^m, \mathcal{B}^m)$ es una transformación medible, $Y = g(X)$ es un vector aleatorio m -dimensional (composición de funciones medibles) cuya distribución puede hallarse a partir de la de X .

Fórmula general de cambio de variable

$$F_Y(y) = P[Y \leq y] = P[g(X) \leq y] = P[X \in g^{-1}((-\infty, y])]$$

Cambio de variable de discreto a discreto

Si X es un vector aleatorio n -dimensional discreto con valores en E_X y $g : (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\mathbb{R}^m, \mathcal{B}^m)$ es una transformación medible, $Y = g(X)$ es un vector aleatorio m -dimensional discreto con valores en $g(E_X)$ y su función masa de probabilidad se puede obtener a partir de la de X como

$$P[Y = y] = P[X \in g^{-1}(y)] = \sum_{x \in g^{-1}(y)} P[X = x], \quad y \in g(E_X).$$

Ejemplo 1

Sea $X = (X_1, X_2)$ un vector aleatorio discreto con función masa de probabilidad

x_1	-1	0	1
x_2	-2	1/6	1/12
	1	1/6	1/12
	2	1/12	0

Calcular la función masa de probabilidad de $Y = (|X_1|, X_2^2)$

Y toma valores en $\{(1, 4), (1, 1), (0, 4), (0, 1)\}$

$$P[Y = (1, 4)] = P[X_1 = \pm 1, X_2 = \pm 2] = \frac{1}{6} + \frac{1}{6} + \frac{1}{12} + \frac{1}{12} = \frac{6}{12}$$

$$P[Y = (1, 1)] = P[X_1 = \pm 1, X_2 = 1] = \frac{1}{6} + \frac{1}{6} = \frac{4}{12}$$

$$P[Y = (0, 4)] = P[X_1 = 0, X_2 = \pm 2] = \frac{1}{12}$$

$$P[Y = (0, 1)] = P[X_1 = 0, X_2 = 1] = \frac{1}{12}$$

Cambio de variable de continuo a discreto

Si X es un vector aleatorio n -dimensional continuo con función de densidad f_X y $g : (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\mathbb{R}^m, \mathcal{B}^m)$ es una transformación medible tal que $Y = g(X)$ es un vector aleatorio m -dimensional discreto, su función masa de probabilidad se puede obtener a partir de la función de densidad de X como

$$P[Y = y] = P[X \in g^{-1}(y)] = \int_{g^{-1}(y)} f_X(x) dx.$$

Ejemplo 2

Sea $X = (X_1, X_2)$ con función de densidad

$$f(x_1, x_2) = \lambda \mu e^{-\lambda x_1} e^{-\mu x_2}, x_1, x_2 > 0 \quad (\lambda, \mu > 0)$$

Sea

$$Y = \begin{cases} 0 & X_1 > X_2 \\ 1 & X_1 < X_2 \end{cases}$$

Calcular su distribución.

$$\begin{aligned}
P[Y = 0] &= P[X_1 > X_2] = \int_0^\infty \int_{x_2}^\infty \lambda \mu e^{-\lambda x_1} e^{-\mu x_2} dx_1 dx_2 \\
&= \int_0^\infty \mu e^{-\mu x_2} \left[\int_{x_2}^\infty \lambda e^{-\lambda x_1} dx_1 \right] dx_2 = \int_0^\infty \mu e^{-\mu x_2} [-e^{-\lambda x_1}]_{x_2}^\infty dx_2 \\
&= \int_0^\infty \mu e^{-(\mu+\lambda)x_2} dx_2 = \mu \frac{e^{-(\mu+\lambda)x_2}}{-(\mu+\lambda)} \Big|_0^\infty = \frac{\mu}{\mu+\lambda}
\end{aligned}$$

$$P[Y = 1] = \frac{\lambda}{\mu + \lambda}$$

Cambio de variable de continuo a continuo

Si X es un vector aleatorio n -dimensional continuo con función de densidad f_X y $g : (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\mathbb{R}^m, \mathcal{B}^m)$ es una transformación medible tal que $Y = g(X)$ es un vector aleatorio m -dimensional continuo, su función de densidad se puede obtener derivando su función de distribución que se obtiene como

$$F_Y(y) = P[Y \leq y] = P[g(X) \leq y] = P[X \in g^{-1}((-\infty, y])] = \int_{g^{-1}((-\infty, y])} f_X(x) dx$$

Ejemplo 3

Sea $X = (X, Y)$ con función de densidad

$$f(x, y) = e^{-x-y}, \quad x, y > 0$$

Calcular la función de densidad de $U = \frac{X}{X+Y}$.

$$U = \frac{X}{X+Y} \in (0, 1) \longrightarrow F_U(u) = \begin{cases} 0 & u < 0 \\ (*) & 0 \leq u < 1 \\ 1 & u \geq 1 \end{cases}$$

$$\begin{aligned}
P\left[\frac{X}{X+Y} \leq u\right] &= (X+Y > 0) = P[X \leq uX + uY] = (1-u > 0) = P\left[X \leq \frac{u}{1-u}Y\right] \\
&= \int_0^\infty \int_0^{\frac{u}{1-u}y} e^{-x} e^{-y} dx dy = \int_0^\infty e^{-y} \left[1 - e^{-\frac{u}{1-u}y}\right] dy \\
&= \int_0^\infty e^{-y} dy - \int_0^\infty e^{-y(1+\frac{u}{1-u})} dy = u
\end{aligned}$$

Así,

$$F_U(u) = \begin{cases} 0 & u < 0 \\ u & 0 \leq u < 1 \\ 1 & u \geq 1 \end{cases}$$

y, por tanto,

$$f_U(u) = 1, \quad 0 < u < 1,$$

es decir, $\frac{X}{X+Y} \sim U(0,1)$.

Teorema: Cambio de variable de continuo a continuo

Sea $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}^n, P_X)$ un vector aleatorio con función de densidad f_X y $g : (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$ una función medible tal que

1) $g = (g_1, \dots, g_n)$ admite inversa $g^{-1} = (g_1^*, \dots, g_n^*)$.

2) $\forall i, j = 1, \dots, n, \exists \frac{\partial g_i^*(y_1, \dots, y_n)}{\partial y_j}$.

3) $J = \left| \left(\left(\frac{\partial g_i^*(y_1, \dots, y_n)}{\partial y_j} \right) \right)_{i,j} \right| \neq 0$.

En estas condiciones, el vector aleatorio n -dimensional $Y = g(X)$ es de tipo continuo y su función de densidad es

$$f_Y(y) = f_X(g^{-1}(y))|J|, \quad \forall y \in \mathbb{R}^n$$

NOTA 1: Si el vector transformado es de dimensión menor que el original se consideran variables auxiliares y luego se calcula la marginal correspondiente.

NOTA 2: Si g no admite inversa pero cada $y \in \mathbb{R}^n$ tiene un número finito de antiimágenes, $g_1^{-1}(y), \dots, g_{k(y)}^{-1}(y)$, y cada una de las antiimágenes satisface las hipótesis del Teorema, entonces

$$f_Y(y) = \sum_{i=1}^{k(y)} |J_i| f_X(g_i^{-1}(y))$$

siendo J_i el jacobiano correspondiente a g_i^{-1} .

Ejemplo 3

Sea $X = (X, Y)$ con función de densidad

$$f(x, y) = e^{-x-y}, \quad x, y > 0$$

Calcular la función de densidad de $U = \frac{X}{X+Y}$.

$$U = \frac{X}{X+Y}$$

$$V = X+Y$$

de forma que $0 < u < 1, v > 0$. La transformación inversa es

$$\begin{aligned} X &= UV \\ Y &= V(1 - U) \end{aligned}$$

con jacobiano

$$J = \begin{vmatrix} v & u \\ -v & 1 - u \end{vmatrix} = v$$

Así,

$$f_{(U,V)}(u, v) = f(uv, v(1-u))|v| = e^{-uv-v(1-u)}|v| = ve^{-v}, \quad uv > 0, v(1-u) > 0 \quad (0 < u < 1, v > 0),$$

de donde

$$f_U(u) = \int_0^\infty ve^{-v} dv = 1, \quad 0 < u < 1$$

Distribución del máximo y del mínimo de variables aleatorias

Un tipo de transformaciones que aparecen comúnmente en la práctica son el máximo y el mínimo de variables aleatorias.

Sea $X = (X_1, \dots, X_n)$ un vector aleatorio, se define $M = \max(X_1, \dots, X_n)$ como una variable aleatoria tal que

$$M(\omega) = \max\{X_1(\omega), X_2(\omega), \dots, X_n(\omega)\}$$

y $N = \min(X_1, \dots, X_n)$ como una variable aleatoria tal que

$$N(\omega) = \min\{X_1(\omega), X_2(\omega), \dots, X_n(\omega)\}$$

Dado que estas transformaciones no satisfacen, en general, las condiciones de los Teoremas de Cambio de variable, para obtener su distribución usamos la fórmula general.

Distribución del máximo: $M = \max(X_1, \dots, X_n)$

Sea $X = (X_1, \dots, X_n)$ un vector aleatorio con función de distribución F_X . Entonces

$$F_M(x) = P[M \leq x] = P[X_1 \leq x, \dots, X_n \leq x] = F_X(x, \dots, x)$$

y, a partir de aquí se obtiene su función masa de probabilidad, en el caso discreto, o su función de densidad, en el caso continuo.

Distribución del mínimo: $N = \min(X_1, \dots, X_n)$

$$F_N(x) = P[N \leq x] = 1 - P[N > x] = 1 - P[X_1 > x, \dots, X_n > x]$$

y, a partir de aquí se obtiene su función masa de probabilidad, en el caso discreto, o su función de densidad, en el caso continuo.

Distribución conjunta: (M, N)

$$F_{(M,N)}(x, y) = P[M \leq x, N \leq y] = \begin{cases} P[M \leq x] = F_X(x, \dots, x) & x < y \\ P[M \leq x] - P[M \leq x, N > y] \\ = P[M \leq x] - P[y < X_i \leq x, \forall i = 1, \dots, n] & x \geq y \end{cases}$$

Ejemplo

Sea $X = (X_1, X_2)$ un vector aleatorio discreto con función masa de probabilidad

$$P[X_1 = x_1, X_2 = x_2] = p^2(1-p)^{x_1+x_2}, \quad x_i = 0, 1, 2, \dots, \quad i = 1, 2 \quad (0 < p < 1).$$

Calcular la función masa de probabilidad de $M = \max(X_1, X_2)$, $N = \min(X_1, X_2)$ y la conjunta de M y N .

- $M = \max(X_1, X_2) : 0, 1, 2, \dots$

$$\begin{aligned} P[M \leq x] &= P[X_1 \leq x, X_2 \leq x] = \sum_{x_1, x_2=0}^x P[X_1 = x_1, X_2 = x_2] = p^2 \sum_{x_1, x_2=0}^x (1-p)^{x_1+x_2} \\ &= p^2 \left(\sum_{x_1=0}^x (1-p)^{x_1} \right)^2 = p^2 \left(\frac{(1-p)^{x+1} - 1}{-p} \right)^2 = (1 - (1-p)^{x+1})^2 \end{aligned}$$

de donde se deduce

$$P[M = x] = P[M \leq x] - P[M \leq x-1] = (1 - (1-p)^{x+1})^2 - (1 - (1-p)^x)^2$$

También se podría haber hecho directamente

$$\begin{aligned} P[M = m] &= P[X_1 = m, X_2 < m] + P[X_1 < m, X_2 = m] + P[X_1 = m, X_2 = m] \\ &= \sum_{x_2=0}^{m-1} p^2(1-p)^{m+x_2} + \sum_{x_1=0}^{m-1} p^2(1-p)^{m+x_1} + p^2(1-p)^{2m} \\ &= 2p^2(1-p)^m \sum_{x=0}^{m-1} (1-p)^x + p^2(1-p)^{2m} \\ &= 2p^2(1-p)^m \frac{(1-p)^m - 1}{-p} + p^2(1-p)^{2m} \\ &= 2p(1-p)^m(1 - (1-p)^m) + p^2(1-p)^{2m}, \quad m = 0, 1, 2, \dots \end{aligned}$$

- $M = \min(X_1, X_2) : 0, 1, 2, \dots$

$$\begin{aligned} P[N \leq x] &= 1 - P[X_1 > x, X_2 > x] = 1 - \sum_{x_1, x_2=x+1}^{\infty} P[X_1 = x_1, X_2 = x_2] = 1 - \sum_{x_1, x_2=x+1}^{\infty} p^2(1-p)^{x_1+x_2} \\ &= 1 - p^2 \left(\sum_{x_1=x+1}^{\infty} (1-p)^{x_1} \right)^2 = 1 - p^2 \left(\frac{(1-p)^{x+1}}{p} \right)^2 = 1 - (1-p)^{2x+2} \end{aligned}$$

de donde se deduce

$$P[N = x] = P[N \leq x] - P[N \leq x - 1] = (1 - p)^{2x} - (1 - p)^{2x+2}$$

También se podría haber hecho directamente

$$\begin{aligned} P[N = n] &= P[X_1 = n, X_2 > n] + P[X_1 > n, X_2 = n] + P[X_1 = n, X_2 = n] \\ &= \sum_{x_2=n+1}^{\infty} p^2(1-p)^{n+x_2} + \sum_{x_1=n+1}^{\infty} p^2(1-p)^{n+x_1} + p^2(1-p)^{2n} \\ &= 2p^2 \sum_{x=n+1}^{\infty} (1-p)^{n+x} + p^2(1-p)^{2n} = 2p^2(1-p)^n \frac{(1-p)^{n+1}}{-p} + p^2(1-p)^{2n} \\ &= -2p(1-p)^n(1-p)^{n+1} + p^2(1-p)^{2n}, \quad n = 0, 1, 2, \dots \end{aligned}$$

- (M, N) toma valores en $\{(m, n) / m, n = 0, 1, \dots, m \geq n\}$

Calculamos directamente su función masa de probabilidad

$$\begin{aligned} P[M = m, N = m] &= P[X_1 = m, X_2 = m] = p^2(1-p)^{2m}, \quad m = n \\ P[M = m, N = n] &= P[X_1 = m, X_2 = n] + P[X_1 = n, X_2 = m] = 2p^2(1-p)^{m+n}, \quad m \neq n \end{aligned}$$

Independencia de variables aleatorias

El concepto de independencia de variables aleatorias es esencial en el Cálculo de Probabilidades. La distribución conjunta de un conjunto de variables aleatorias determina de forma única las distribuciones marginales. Sin embargo, el recíproco no es cierto en general.

A continuación estudiamos una situación en la que si se cumple este hecho: las marginales determinan de forma única la distribución conjunta.

Definición y caracterizaciones de independencia

Definición

Sean X_1, \dots, X_n variables aleatorias unidimensionales definidas sobre el mismo espacio de probabilidad y sea $X = (X_1, \dots, X_n)$. Decimos que las variables X_1, \dots, X_n son mutuamente independientes (completamente independientes) o, simplemente, independientes si

$$F_X(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n), \quad \forall x_1, \dots, x_n \in \mathbb{R}$$

Caracterización para variables aleatorias discretas

Sean X_1, \dots, X_n variables aleatorias unidimensionales discretas definidas sobre el mismo espacio de probabilidad

$$X_1, \dots, X_n \text{ son independientes} \Leftrightarrow P[X_1 = x_1, \dots, X_n = x_n] = P[X_1 = x_1] \cdots P[X_n = x_n],$$

$$\forall x_1, \dots, x_n \in \mathbb{R}$$

Caracterización para variables aleatorias continuas

Sean X_1, \dots, X_n variables aleatorias unidimensionales continuas definidas sobre el mismo espacio de probabilidad

$$X_1, \dots, X_n \text{ son independientes} \Leftrightarrow \begin{cases} X = (X_1, \dots, X_n) \text{ es de tipo continuo} \\ f_X(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n), \quad \forall x_1, \dots, x_n \in \mathbb{R} \end{cases}$$

El siguiente teorema de caracterización de independencia establece la relación entre la independencia de variables aleatorias y la independencia de sucesos expresados en términos de dichas variables.

Caracterización de independencia por conjuntos de Borel

Sean X_1, \dots, X_n variables aleatorias unidimensionales definidas sobre el mismo espacio de probabilidad

$$X_1, \dots, X_n \text{ son independientes} \Leftrightarrow P[X_1 \in B_1, \dots, X_n \in B_n] = P[X_1 \in B_1] \cdots P[X_n \in B_n], \\ \forall B_1, \dots, B_n \in \mathcal{B}$$

Dem

\Leftarrow] Si la relación es cierta $\forall B_1, \dots, B_n \in \mathcal{B}$, tomando $B_i = (-\infty, x_i]$ se tiene la definición de independencia.

\Rightarrow] Caso discreto: X_1, \dots, X_n V.A. discretas

$$\begin{aligned} P[X_1 \in B_1, \dots, X_n \in B_n] &= \sum_{\substack{x_1 \in B_1 \\ \dots \\ x_n \in B_n}} P[X_1 = x_1, \dots, X_n = x_n] = \sum_{\substack{x_1 \in B_1 \\ \dots \\ x_n \in B_n}} P[X_1 = x_1] \cdots P[X_n = x_n] \\ &= \left(\sum_{x_1 \in B_1} P[X_1 = x_1] \right) \cdots \left(\sum_{x_n \in B_n} P[X_n = x_n] \right) \\ &= P[X_1 \in B_1] \cdots P[X_n \in B_n] \end{aligned}$$

\Rightarrow] Caso continuo: X_1, \dots, X_n V.A. continuas (si las variables son independientes y continuas, el vector aleatorio formado por ellas es continuo)

$$\begin{aligned} P[X_1 \in B_1, \dots, X_n \in B_n] &= \int_{B_1} \cdots \int_{B_n} f_X(x_1, \dots, x_n) dx_n \cdots dx_1 \\ &= \int_{B_1} \cdots \int_{B_n} f_{X_1}(x_1) \cdots f_{X_n}(x_n) dx_n \cdots dx_1 \\ &= \left(\int_{B_1} f_{X_1}(x_1) dx_1 \right) \cdots \left(\int_{B_n} f_{X_n}(x_n) dx_n \right) \\ &= P[X_1 \in B_1] \cdots P[X_n \in B_n] \end{aligned}$$

A continuación damos otra caracterización, de gran interés práctico, para variables discretas y continuas. Su interés radica en que no es preciso calcular las funciones masa de probabilidad o funciones de densidad marginales para comprobar la independencia.

Caracterización de independencia por factorización

a) Sean X_1, \dots, X_n variables aleatorias unidimensionales discretas definidas sobre el mismo espacio de probabilidad

X_1, \dots, X_n son independientes $\Leftrightarrow P[X_1 = x_1, \dots, X_n = x_n] = h_1(x_1) \cdots h_n(x_n), \forall x_1, \dots, x_n \in \mathbb{R}$

b) Sean X_1, \dots, X_n variables aleatorias unidimensionales continuas definidas sobre el mismo espacio de probabilidad

$$X_1, \dots, X_n \text{ son independientes} \Leftrightarrow \begin{cases} X = (X_1, \dots, X_n) \text{ es continuo} \\ f_X(x_1, \dots, x_n) = h_1(x_1) \cdots h_n(x_n), \forall x_1, \dots, x_n \in \mathbb{R} \end{cases}$$

Ejemplos

- $f(x, y) = 1, 0 < x < 1, 0 < y < 1 \rightarrow$ Independientes
- $f(x, y) = 2, 0 < x < y < 1 \rightarrow$ No son independientes

$$f(x, y) = 2\epsilon(x)\epsilon(y-x)\epsilon(1-y) \text{ con } \epsilon(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

De hecho las marginales son

$$f_1(x) = 2(1-x), 0 < x < 1, \quad f_2(y) = 2y, 0 < y < 1$$

Propiedades de independencia

Teorema 1

Una variable aleatoria degenerada es independiente de cualquier otra.

Dem

$X_1 \equiv c, X_2$ arbitraria

$$F(x_1, x_2) = P[X_1 \leq x_1, X_2 \leq x_2] = \begin{cases} 0 = P[X_1 \leq x_1]P[X_2 \leq x_2] & \text{si } x_1 < c \\ P[X_2 \leq x_2] = P[X_1 \leq x_1]P[X_2 \leq x_2] & \text{si } x_1 \geq c \end{cases}$$

Teorema 2

Si X_1, \dots, X_n son independientes y $g_i : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$ son funciones medibles ($i = 1, \dots, n$), entonces las variables $g_1(X_1), \dots, g_n(X_n)$ son independientes.

Dem

$$\begin{aligned} P[g_1(X_1) \leq y_1, \dots, g_n(X_n) \leq y_n] &= P[X_1 \in g_1^{-1}((-\infty, y_1]), \dots, X_n \in g_n^{-1}((-\infty, y_n])] \\ &= P[X_1 \in g_1^{-1}((-\infty, y_1])] \cdots P[X_n \in g_n^{-1}((-\infty, y_n])] \\ &= P[g_1(X_1) \leq y_1] \cdots P[g_n(X_n) \leq y_n] \end{aligned}$$

Teorema 3

Si X_1, \dots, X_n son independientes, cualquier subcolección X_{i_1}, \dots, X_{i_k} también lo son.

Dem

$$\begin{aligned} &P[X_{i_1} \in B_{i_1}, \dots, X_{i_k} \in B_{i_k}] = \\ &P[X_1 \in \mathbb{R}, \dots, X_{i_1-1} \in \mathbb{R}, X_{i_1} \in B_{i_1}, X_{i_1+1} \in \mathbb{R}, \dots, X_{i_k-1} \in \mathbb{R}, X_{i_k} \in B_{i_k}, X_{i_k+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R}] \\ &\quad \text{(por la caracterización de conjuntos de Borel)} \\ &= P[X_1 \in \mathbb{R}] \cdots P[X_{i_1-1} \in \mathbb{R}] P[X_{i_1} \in B_{i_1}] P[X_{i_1+1} \in \mathbb{R}] \cdots P[X_{i_k-1} \in \mathbb{R}] P[X_{i_k} \in B_{i_k}] P[X_{i_k+1} \in \mathbb{R}] \cdots P[X_n \in \mathbb{R}] \\ &= P[X_{i_1} \in B_{i_1}] \cdots P[X_{i_k} \in B_{i_k}] \end{aligned}$$

Teorema 4

X_1, \dots, X_n son independientes \Leftrightarrow las distribuciones condicionadas de cualquier subvector a cualquier otro coinciden con la marginal del primero.

Independencia dos a dosDefinición

Sean X_1, \dots, X_n variables aleatorias unidimensionales definidas sobre el mismo espacio de probabilidad son independientes dos a dos si

$$\forall i, j = 1, \dots, n, i \neq j, X_i \text{ y } X_j \text{ son independientes.}$$

Puesto que cualquier subcolección de variables aleatorias independientes lo son, es claro que
INDEPENDENCIA MUTUA \Rightarrow INDEPENDENCIA DOS A DOS

Sin embargo, el recíproco no es cierto como se prueba en el siguiente ejemplo.

Ejemplo. Sean X_1 y X_2 variables aleatorias independientes y con idéntica distribución $P[X_i = \pm 1] = 1/2$, $i = 1, 2$. Sea $X_3 = X_1 \cdot X_2$. Vamos a probar que X_1 , X_2 y X_3 son independientes dos a dos pero no mutuamente independientes.

Independencia dos a dos.

- X_1 y X_2 son independientes por hipótesis.
- X_1 y X_3 son independientes:

$$P[X_3 = \pm 1] = 1/2$$

- $P[X_1 = 1, X_3 = 1] = P[X_1 = 1, X_2 = 1] = (\text{por indep.}) = \frac{1}{4} = P[X_1 = 1]P[X_3 = 1]$
- $P[X_1 = 1, X_3 = -1] = P[X_1 = 1, X_2 = -1] = (\text{por indep.}) = \frac{1}{4} = P[X_1 = 1]P[X_3 = -1]$
- $P[X_1 = -1, X_3 = 1] = P[X_1 = -1, X_2 = -1] = (\text{por indep.}) = \frac{1}{4} = P[X_1 = -1]P[X_3 = 1]$
- $P[X_1 = -1, X_3 = -1] = P[X_1 = -1, X_2 = 1] = (\text{por indep.}) = \frac{1}{4} = P[X_1 = -1]P[X_3 = -1]$

- X_2 y X_3 son independientes (igual que el anterior)

Sin embargo, no son mutuamente independientes ya que

$$P[X_1 = 1, X_2 = 1, X_3 = -1] = 0 \neq \frac{1}{8} = P[X_1 = 1]P[X_2 = 1]P[X_3 = -1]$$

Familias arbitrarias de variables aleatorias independientes

La noción de independencia de variables aleatoria, que se ha definido para colecciones finitas, se puede extender a familias arbitrarias.

Definición

Dada una familia arbitraria de variables aleatorias definidas sobre el mismo espacio de probabilidad, se dice que dichas variables son independientes si cualquier subcolección finita de ellas son VA independientes.

Sucesiones de variables aleatorias independientes

Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias definidas sobre el mismo espacio de probabilidad. Se dice que $\{X_n\}_{n \in \mathbb{N}}$ es una sucesión de variables aleatorias independientes si cualquier subcolección finita de variables de la sucesión son independientes o, equivalentemente, si

$$\forall n \in \mathbb{N}, X_1, \dots, X_n \text{ son independientes.}$$

Independencia de vectores aleatorios

La definición de independencia de variables aleatoria se extiende de forma inmediata a vectores aleatorios.

Definición

Si X^1, \dots, X^m son vectores aleatorios definidos sobre un mismo espacio de probabilidad, con dimensiones n_1, \dots, n_m , respectivamente,

X^1, \dots, X^m son independientes $\Leftrightarrow F_X(x^1, \dots, x^m) = F_{X^1}(x^1) \cdots F_{X^m}(x^m), \forall x^1 \in \mathbb{R}^{n_1}, \dots, x^m \in \mathbb{R}^{n_m}$

siendo $X = (X^1, \dots, X^m)$.

Las caracterizaciones y propiedades se extienden al caso multidimensional.