“Introducción a la Estadística, Probabilidad e Inferencia”

Maestría en Estadística Aplicada

Facultad de Ciencias Económicas y Estadística

UNR

Unidad 4 – Parte 3

- Vectores aleatorios multidimensionales. Casos discreto y continuo.
- Funciones de probabilidad y de densidad. Función de distribución.
- Distribuciones marginales y condicionales. Independencia.
- Esperanza y variancia de vectores aleatorios.

Introducción

Hasta aquí hemos considerado sólo el caso en el que interesa una única variable como resultado del experimento aleatorio, y trabajado entonces con variables aleatorias unidimensionales.

Sin embargo, en muchos casos interesa observar dos o incluso más características simultáneamente. Por ejemplo, diámetro y longitud de una pieza de manufactura, altura y peso de una persona, intensidad de la luz en un punto y distancia del punto a la fuente de luz, etc.

En estos casos, a cada elemento de Ω se le asignan, digamos k números reales que representan cada una de las k características de interés, mediante la definición de k variables aleatorias.

Cuando estas variables se consideran como las componentes de un vector k-dimensional, éste recibe el nombre de ***vector aleatorio***.

Vectores aleatorios

<u>Definición:</u> Sea E un experimento aleatorio y Ω su espacio muestral. Si para cada $i = 1, \ldots, k$, X_i es una función que asigna a cada uno de los elementos $\omega \in \Omega$ un número real $X_i(\omega)$, el vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_k)$ se llama ***vector aleatorio k-dimensional***.

Llamaremos recorrido de $\boldsymbol{X}$ al conjunto de todos los valores posibles del vector, y lo notaremos como $R_{\boldsymbol{X}}$. Notar que cada uno de los elementos en $R_{\boldsymbol{X}}$ es un punto en el espacio k-dimensional: $\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_k}) \in \mathbb{R}^{\boldsymbol{k}}$.

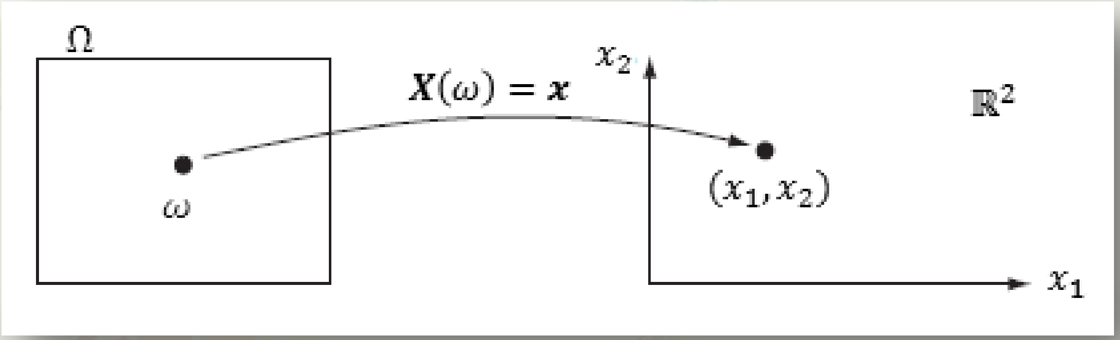
Teniendo en cuenta todos los conceptos vistos para variables aleatorias unidimensionales, veremos como hacer la extensión de los mismos al caso k-dimensional.

Vectores aleatorios

Al igual que en el caso unidimensional, podemos pensar que los vectores aleatorios “traducen” los resultados del espacio muestral a *k-uplas* ordenadas de números reales, mediante la función:

$$\boldsymbol{X}(\omega)\colon \Omega \to \mathbb{R}^k \mid \boldsymbol{X}(\omega) = \boldsymbol{x} = (x_1, x_2, \dots, x_k)$$

Es decir, a cada elemento de Ω le corresponde un punto en el espacio $\mathbb{R}^k$. En el caso bidimensional por ejemplo, cada $\omega \in \Omega$ es traducido a un punto del plano $\mathbb{R}^2$.



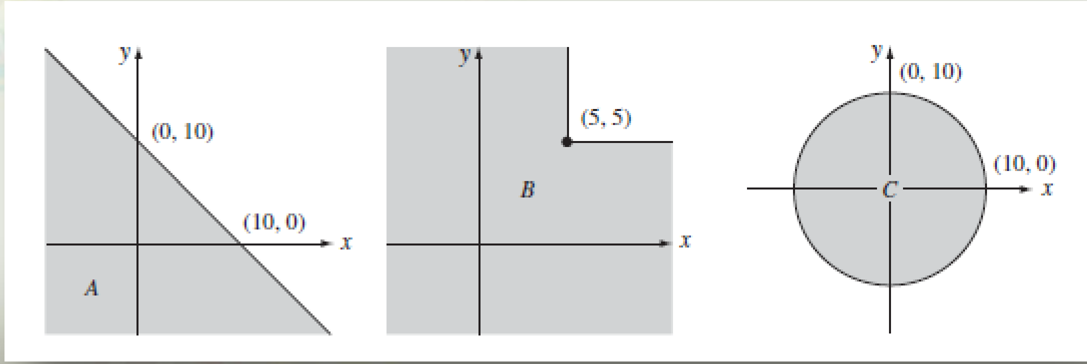
Vectores aleatorios - Sucesos

La noción de ***suceso*** con respecto al recorrido de una variable aleatoria, puede también ser extendido fácilmente al caso de vectores aleatorios. Estos sucesos vendrán dados por las condiciones que deben cumplir las variables que componen el vector, y pueden ser interpretados como regiones del espacio $\mathbb{R}^k$.

Por ejemplo, para el caso bidimensional, podríamos pensar en los eventos:

$$A = \{X + Y \leq 10\}, B = \{\min(X, Y) \leq 5\}, C = \{X^2 + Y^2 \leq 10\}$$

Luego, si por ejemplo, el recorrido del vector aleatorio (X, Y) es todo el plano, estos eventos estarían representados por las siguientes regiones.

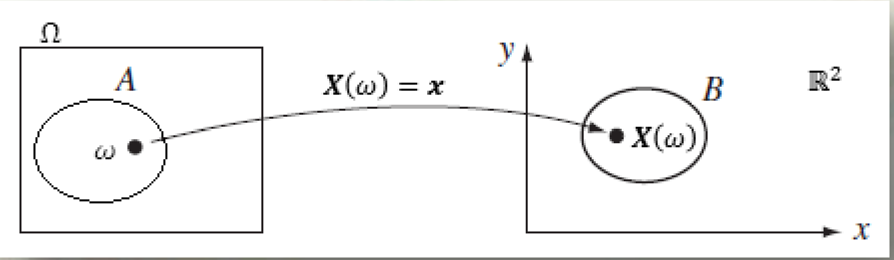


Vectores aleatorios - Sucesos

La asignación de probabilidades a estos eventos en el recorrido del vector aleatorio se lleva a cabo de la misma manera que en el caso unidimensional, teniendo en cuenta la correspondencia de estos sucesos con sus equivalentes en el espacio muestral Ω.

Definición: sea E un experimento aleatorio, Ω su espacio muestral, $\boldsymbol{X}$ un vector aleatorio en Ω y $R_{\boldsymbol{X}}$ su recorrido. Sea B un suceso respecto a $R_{\boldsymbol{X}}$ ($B \subset R_X$).
Si consideramos el conjunto $A = \{\omega \in \Omega | \boldsymbol{X}(\omega) \in B\}$, decimos que A y B son ***sucesos equivalentes***.

En ocasiones se utiliza la notación $A = \boldsymbol{X}^{-1}(B)$.



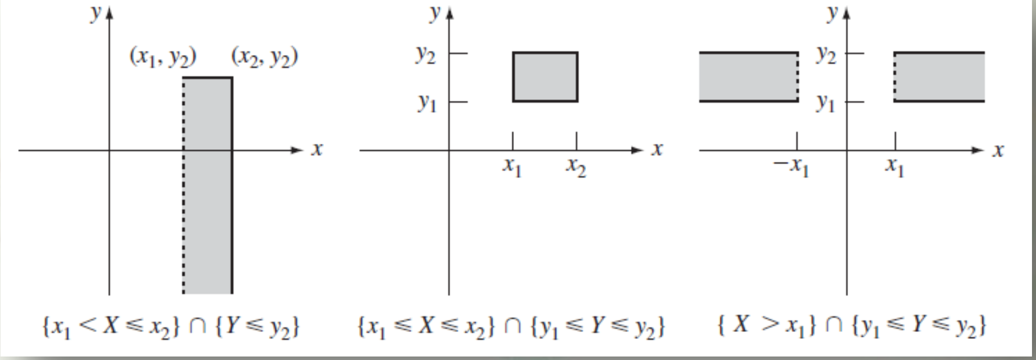
Vectores aleatorios - Sucesos

Luego, la ***probabilidad del evento B en R_X*** se obtiene calculando la probabilidad del evento equivalente A en Ω:

$$P(B) = P(A) \text{ con } A = \{\omega \in \Omega \mid \boldsymbol{X}(\omega) \in B\} = X^{-1}(B)$$

El tipo más simple de evento que se puede considerar es el llamado ***evento de tipo producto***, representado en $\mathbb{R}^k$ por el producto cartesiano de intervalos unidimensionales en $\mathbb{R}$:

$$B = \left\{\{X_1 \in I_1\} \cap \{X_2 \in I_2\} \cap \cdots \cap \{X_k \in I_k\}\right\}$$



Para el que resulta:

$$P(B) = P(A_1 \cap A_2 \cap \cdots \cap A_k), \text{ con } A_i = \boldsymbol{X_i^{-1}}(I_i) \;\; \forall i = 1, \dots, k$$

Vectores aleatorios - Sucesos

<u>Ejemplo:</u> Un experimento aleatorio consiste en lanzar un dado de cuatro caras al aire, dos veces. Sean X_1 y X_2 las variables que representan el número obtenido en cada lanzamiento.

¿Cuál es la probabilidad que el mínimo de los resultados obtenidos sea 3?

Sea el suceso $B = \{\min(X_1, X_2) = 3\}$. Este evento ocurre si y solo si ocurre alguno de los siguientes eventos: $\omega = (3,3)$ ó $\omega = (3,4)$ ó $\omega = (4,3)$.

Luego $B = \{\min(X_1, X_2) = 3\} \subset R_{X_1,X_2}$ y $A = \{(3,3),(3,4),(4,3)\} \subset \Omega$ son equivalentes. De este modo,

$$P(B) = P(A) = \frac{3}{16}$$

Vectores aleatorios

Como en el caso unidimensional, distinguiremos dos tipos de vectores aleatorios: ***vectores aleatorios discretos*** y ***vectores aleatorios continuos***.

Para cada uno de ellos extenderemos los conceptos ya vistos: funciones de probabilidad y densidad, función de distribución acumulada, valor esperado, variancia, etc.

Asimismo, introduciremos otros conceptos propios del caso multidimensional: distribuciones marginales, distribuciones condicionales, independencia de variables aleatorias.

Vectores aleatorios discretos

Vectores aleatorios discretos

Definición: Sea $\boldsymbol{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio. Si el número de valores posibles $\boldsymbol{x} = (x_1, x_2, \dots, x_k)$ de $\boldsymbol{X}$, es decir, $R_{\boldsymbol{X}}$ es finito o infinito numerable, diremos que $\boldsymbol{X}$ es un *vector aleatorio discreto*.

Ejemplo 1: Sobre el mismo experimento del lanzamiento de los dos dados de cuatro caras. Sean las variables X_1: indicadora de resultado doble (mismo número en ambos lanzamientos) y X_2: suma de los resultados obtenidos.

¿Cuál es el recorrido del vector aleatorio?

Claramente $\boldsymbol{X} = (X_1, X_2)$ es un *vector aleatorio discreto*. Notar que tanto X_1 como X_2 son variables aleatorias discretas.

Vectores aleatorios discretos

¿Cómo resumir o describir el comportamiento en probabilidad de un vector aleatorio discreto?

<u>Definición:</u> Sea $\boldsymbol{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio discreto. A cada resultado posible $\boldsymbol{x} = (x_1, x_2, \dots, x_k)$ se asocia un número $p(\boldsymbol{x})$ que representa la $P(\boldsymbol{X} = \boldsymbol{x})$. Si estos números $p(\boldsymbol{x})$ satisfacen, $\forall \boldsymbol{x} \in R_{\boldsymbol{X}}$, las condiciones:

- $p(\boldsymbol{x}) \geq 0 \ \forall \ \boldsymbol{x} \in R_{\boldsymbol{X}}$
- $\sum_{x_1} \dots \sum_{x_k} p(\boldsymbol{x}) = 1$

La función $p(\boldsymbol{x}) = P(\boldsymbol{X} = \boldsymbol{x})$ se llama ***función de probabilidad <u>conjunta</u>*** del vector aleatorio $\boldsymbol{X}$, y lo notamos como: $\boldsymbol{X} \sim \boldsymbol{p}(\boldsymbol{x})$

Vectores aleatorios discretos

Para los datos del ejemplo, tenemos un espacio muestral consistente en 16 resultados equiprobables. La función de probabilidad conjunta puede hallarse identificando, para cada resultado elemental del recorrido del vector, su suceso equivalente en el espacio muestral.

Así, por ejemplo, el suceso $\mathrm{B}_1 = \{(0,4)\} \subset R_{\boldsymbol{X}}$ tiene por suceso equivalente al conjunto $A_1 = \{(1,3),(3,1)\} \subset \Omega$.

Así, $P_{\boldsymbol{X}}(B) = P(A) = \frac{2}{16}$.

De la misma manera:

$B_2 = \{(1,4)\}$ es equivalente a $A_2 = \{(2,2)\}$, y entonces $P_{\boldsymbol{X}}(B_2) = \frac{1}{16}$.

$B_3 = \{(0,5)\}$ es equivalente a $A_3 = \{(2,3),(3,2),(4,1),(1,4)\}$, y entonces $P_{\boldsymbol{X}}(B_3) = \frac{4}{16}$.

Vectores aleatorios discretos

Una vez inducidas las probabilidades de todos los resultados en el recorrido del vector, la función de probabilidad conjunta puede representarse de ***manera tabular*** de la siguiente manera:

		X_2						
	$p(x_1, x_2)$	**2**	**3**	**4**	**5**	**6**	**7**	**8**
X_1	**0**	0	2/16	2/16	4/16	2/16	2/16	0
	1	1/16	0	1/16	0	1/16	0	1/16

¿Cuál sería la representación gráfica?

Vectores aleatorios discretos

Ejemplo 2: un lote de 100 artículos artesanales de decoración está compuesto, según el artesano productor, por 50 artículos de excelente calidad, 30 de calidad media y los 20 restantes de menor calidad por poseer algún defecto de terminación o pintura. El comprador decide sacar una muestra de 10 artículos al azar, a partir de la cual decidirá si comprar el lote o no. Sean X_1 y X_2 las variables que representan el número de artículos de excelente y buena calidad en la muestra, respectivamente.

¿Cuál es el recorrido del vector $\boldsymbol{X} = (\boldsymbol{X_1}, \boldsymbol{X_2})$?

Nuevamente $\boldsymbol{X} = (X_1, X_2)$ es un *vector aleatorio discreto*.

Vectores aleatorios discretos

¿Cómo resumir o describir el comportamiento en probabilidad de un vector aleatorio discreto?

La probabilidad que en la muestra haya exactamente k artículos de excelente calidad y h de buena calidad:

$$P(X_1 = k, X_2 = h) = p(k,h) = \frac{\binom{50}{k}\binom{30}{h}\binom{20}{10-k-h}}{\binom{100}{10}}$$

Esta es la ***función de probabilidad puntual conjunta*** del vector $\boldsymbol{X} = (X_1, X_2)$, y vale siempre que $k \geq 0, h \geq 0$, tales que $k + h \leq 10$.

Vectores aleatorios discretos

A partir de la función de probabilidad conjunta del vector $\boldsymbol{X}$ es posible calcular la probabilidad de cualquier evento de interés, simplemente sumando las probabilidades puntuales de los casos favorables al evento de interés:

Sea $B \subset R_{\boldsymbol{X}}$, entonces:

$$P(B) = P(\omega \in \Omega | \boldsymbol{X}(\omega) \in B) = \sum_{\boldsymbol{x} \in B} \dots \sum p(\boldsymbol{x})$$

Donde la suma se extiende sobre todos los puntos $\boldsymbol{x}$ favorables al suceso de interés, B.

Por ejemplo, para los datos del lote de artesanías, el comprador anuncia que comprará el lote si en la muestra se obtienen al menos 8 productos de excelente y buena calidad.

¿Cuál es la probabilidad de comprar el lote?

Vectores aleatorios discretos

Debería obtenerse la probabilidad del evento $B = \{X_1 + X_2 \geq 8\}$, para lo cual deberían sumarse las probabilidades puntuales de los resultados señalados en la figura. Podemos escribir esta suma como:

$$P(B) = \sum_{k=0}^{10} \sum_{h=8-k}^{10-k} p(k,h) = 0.6812$$

X_1 | X_2	0	1	2	3	4	5	6	7	8	9	10
0									X	X	X
1								X	X	X	
2							X	X	X		
3						X	X	X			
4					X	X	X				
5				X	X	X					
6			X	X	X						
7		X	X	X							
8	X	X	X								
9	X	X									
10	X										

Vectores aleatorios continuos

Vectores aleatorios continuos

<u>Definición:</u> Diremos que $\boldsymbol{X} = (X_1, X_2, \dots, X_k)$ es un ***vector aleatorio continuo*** si existe una función f definida sobre $\mathbb{R}^k$, llamada ***función de densidad de probabilidad <u>conjunta</u> (fdp),*** que satisface las siguientes condiciones:

- $f(x_1, x_2, \dots, x_k) \geq 0 \quad \forall\, \boldsymbol{x} = (x_1, x_2, \dots, x_k) \in \mathbb{R}^k$
- $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_k)\, dx_1 dx_2 \dots dx_k = 1$

En ese caso lo notaremos como $\boldsymbol{X} \sim \boldsymbol{f}(\boldsymbol{x})$.

Si $\boldsymbol{X}$ es un vector aleatorio continuo, puede tomar todos los valores posibles en un conjunto no numerable del espacio $\mathbb{R}^k$.

Como en el caso unidimensional, se adopta la convención de que $f(\boldsymbol{x}) = 0$ si $\boldsymbol{x} \notin R_{\boldsymbol{X}}$, de modo que la función de densidad se considera definida en todo el espacio $\mathbb{R}^k$.

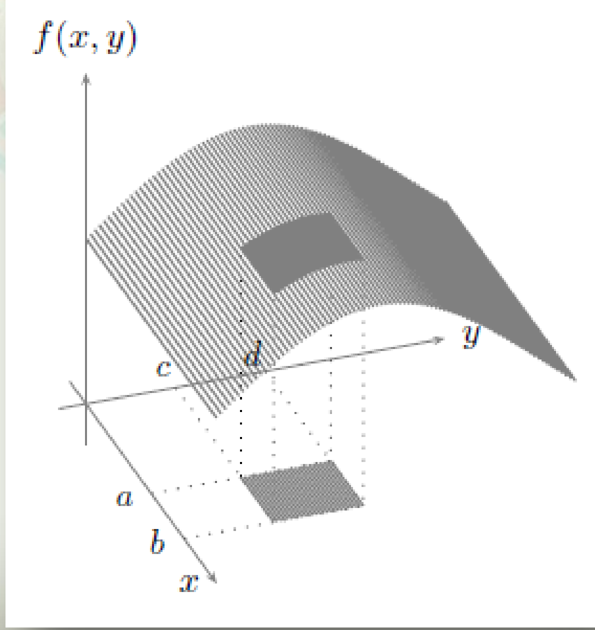
Vectores aleatorios continuos

Observaciones:

- Dado $\boldsymbol{X}$ vector aleatorio continuo con fdp $f(\boldsymbol{x})$, la probabilidad de cualquier evento k-dimensional A, viene dada por la integral múltiple de $f(\boldsymbol{x})$ sobre la región A:

$$P(\boldsymbol{X} \in A) = \int \dots \int_{\boldsymbol{x} \in \boldsymbol{A}} f(\boldsymbol{x})\, d\boldsymbol{x}$$

- Este valor representa el volumen de la región debajo de la superficie definida por la función de densidad y acotada por la región A.



Vectores aleatorios continuos

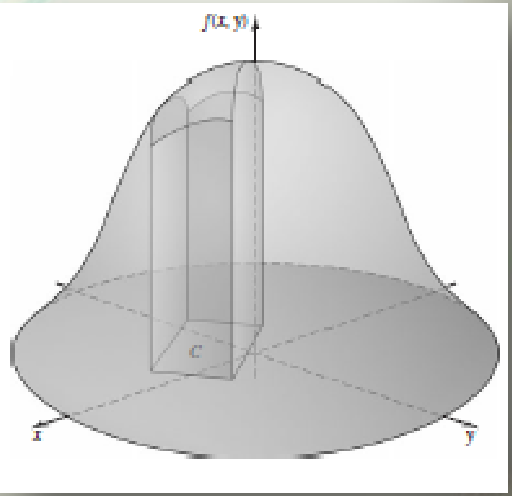
<u>Observaciones:</u>

- Al igual que en el caso de variables aleatorias continuas, de la definición se desprende que para cualquier valor puntual de $\boldsymbol{X}$, digamos $\boldsymbol{x_0}$, se tiene: $P(\boldsymbol{X} = \boldsymbol{x_0}) = 0$

- Nuevamente, **el valor de $\boldsymbol{f(x)}$ NO ES UNA PROBABILIDAD.** Sólo cuando la fdp se integra sobre una región produce una probabilidad. El valor de $f(\boldsymbol{x})$ puede interpretarse como la densidad de probabilidad asociada al punto $\boldsymbol{x}$, en el sentido que para valores Δx_i pequeños, resulta:

$$f(x_1, x_2, \dots, x_k)\Delta x_1 \Delta x_2 \dots \Delta x_k \cong$$

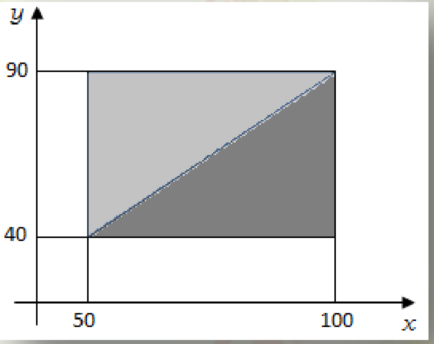
$$\cong P(x_1 \leq X_1 \leq x_1 + \Delta x_1, \dots., x_k \leq X_k \leq x_k + \Delta x_k)$$



Vectores aleatorios continuos

Ejemplo: un empresario desea analizar las ventas de su negocio en los meses de enero y febrero. Sean X e Y variables aleatorias que representan el volumen de ventas (en miles de pesos) en cada mes, respectivamente. Se supone que (X, Y) es vector aleatorio continuo con fdp dada por:

$$f(x,y) = \begin{cases} 1/2500, & 50 < x < 100; 40 < y < 90 \\ 0, & en\ otro\ caso \end{cases}$$



Veamos que $f(x,y)$ es función de densidad:

- $f(x,y) \geq 0 \quad \forall (x,y) \in \mathbb{R}^2$
- $\int_{50}^{100} \int_{40}^{90} 1/2500\, dy\, dx = 1.$

Entonces podemos calcular la probabilidad de eventos en $R_{X,Y}$, integrando esta fdp. Por ejemplo:

$$P(X \geq Y) = \int_{50}^{100} \int_{40}^{x} 1/2500\, dy\, dx = 0.70$$

Función de distribución acumulada de vectores aleatorios

Función de distribución acumulada

Definición: Sea $\boldsymbol{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio k-dimensional. La ***función de distribución acumulada (fda)*** del vector aleatorio $\boldsymbol{X}$ se define, para cada $\boldsymbol{x} = (x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ como la probabilidad del evento $\left\{\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\} \cap \cdots \cap \{X_k \leq x_k\}\right\}$. Esto es:

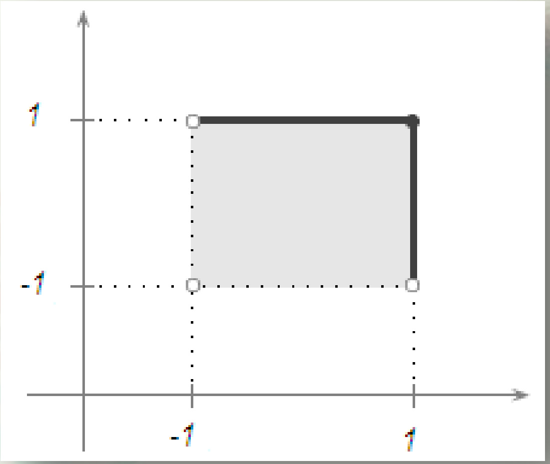
$$F_{\boldsymbol{X}}(\boldsymbol{x}) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_k \leq x_k) \; \forall \boldsymbol{x} \in \mathbb{R}^k$$

Una fda así definida goza de las siguientes ***propiedades***:

- Es monótona no decreciente en cada $X_i, i = \overline{1, k}$. Es decir, si $x_i < x_i^*$ entonces: $F_{\boldsymbol{X}}(x_1, x_2, \dots, x_i, \dots, x_k) \leq F_{\boldsymbol{X}}(x_1, x_2, \dots, x_i^*, \dots, x_k)$.
- Es continua por derecha en cada $X_i, i = \overline{1, k}$. Es decir, dado $h > 0$, $F_{\boldsymbol{X}}(x_1, x_2, \dots, x_i, \dots, x_k) = \lim_{h \to 0} F_X(x_1, x_2, \dots, x_i, +h \; \dots, x_k)$.
- $F_{\boldsymbol{X}}(x)$ está acotada entre 0 y 1.
- $F_{\boldsymbol{X}}(x)$ asigna probabilidades mayores o iguales a cero a los eventos de tipo producto.

Función de distribución acumulada

En el caso multidimensional, las primeras 3 propiedades no alcanzan para asegurar que una función $F(x,y)$ sea una función de distribución.



<u>Ejemplo:</u>

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

y siendo $B=(-1,1]\times(-1,1]$, resulta:

$$
\begin{aligned}
P(B) &= P(-1<X\leq 1, -1<Y\leq 1) \\
&= F(1,1)-F(1,-1)-F(-1,1)+F(-1,-1) = 1-1-1+0 \\
&= -1 < 0
\end{aligned}
$$

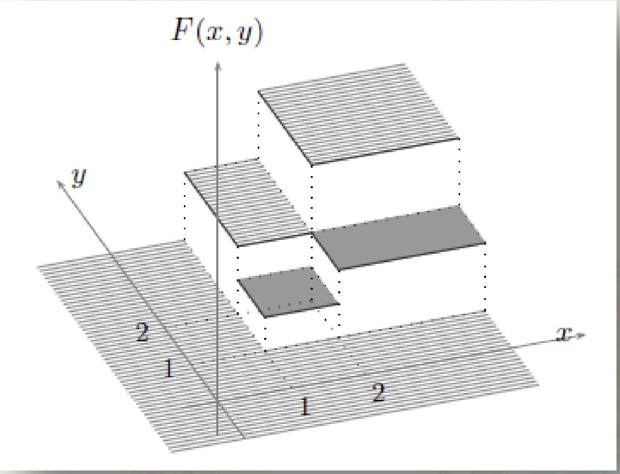
Por lo tanto, aunque esta función cumple las tres primeras propiedades, no puede ser función de distribución, pues no asegura la asignación de probabilidades a los eventos en concordancia con los axiomas.

Función de distribución acumulada

Si X es un vector aleatorio discreto, entonces:

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_k) = \sum_{x_{1i} \le x_1} \dots \sum_{x_{ki} \le x_k} p(x_i)$$

Si X es variable aleatoria discreta, su función de distribución tendrá una apariencia *"escalonada"*.



Función de distribución acumulada

Ejemplo: Para los datos del ejemplo 1 sobre el lanzamiento del dado, habíamos obtenido la función de probabilidad puntual del vector aleatorio $\boldsymbol{X} = (X_1, X_2)$, a partir del cual se puede obtener la función de distribución acumulada:

$\boldsymbol{p_X}(x_1, x_2)$

$X_1 \vert X_2$	2	3	4	5	6	7	8
0	0	2/16	2/16	4/16	2/16	2/16	0
1	1/16	0	1/16	0	1/16	0	1/16

$\boldsymbol{F_X}(x_1, x_2)$

$X_1 \vert X_2$	2	3	4	5	6	7	8
0	0	2/16	4/16	8/16	10/16	12/16	12/16
1	1/16	3/16	6/16	10/16	13/16	15/16	16/16

$$F(0,2) = \sum_{x_1 \le 0} \sum_{x_2 \le 2} p(x, y) = p(0,2) = 0$$

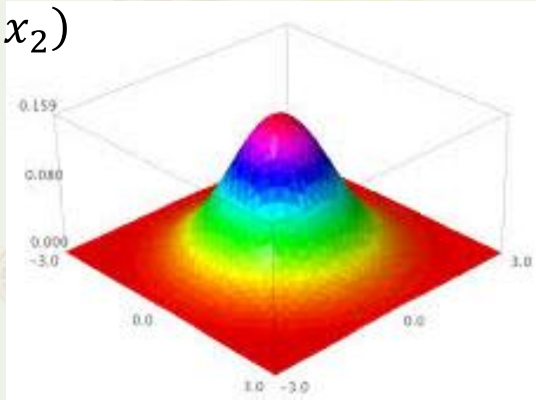
$$F(1, 4) = \sum_{x_1 \le 1} \sum_{x_2 \le 4} p(x, y) = p(0,2) + p(1,2) + p(0,3) + p(1,3) + p(0,4) + p(1,4) = 6/16$$

Función de distribución acumulada

Si X es una variable aleatoria continua, entonces:

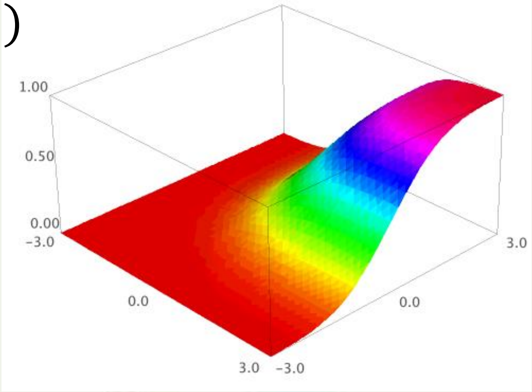
$$F_{\boldsymbol{X}}(x_1, x_2, \ldots, x_k) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_k} f(x_1, x_2, \ldots, x_k) dx_k \ldots dx_1$$

$\boldsymbol{f_X}(x_1, x_2)$

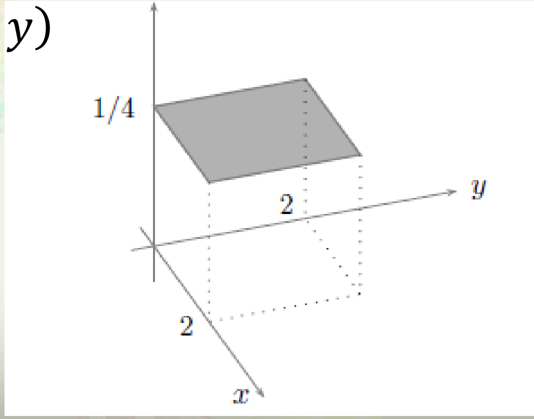


⇨

$\boldsymbol{F_X}(x_1, x_2)$

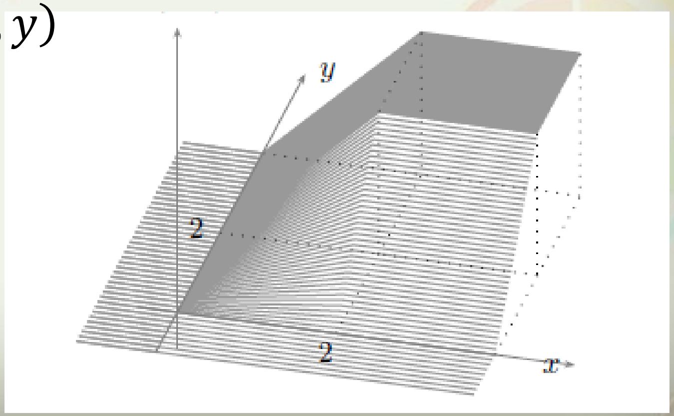


$\boldsymbol{f_{X,Y}}(x, y)$



⇨

$\boldsymbol{F_{X,Y}}(x, y)$



Función de distribución acumulada

Ejemplo 3: Sea (X_1, X_2, X_3) un vector aleatorio continuo que representa los tiempos (en miles de horas) de falla de tres componentes eléctricos, cuyo comportamiento puede describirse mediante la siguiente densidad:

$$f(x_1, x_2, x_3) = 6e^{-x_1-2x_2-3x_3} \quad \text{si } x_1, x_2, x_3 > 0$$

Luego,

$$F_{\boldsymbol{X}}(x_1, x_2, x_3) = \int_0^{x_1}\int_0^{x_2}\int_0^{x_3} f(x_1, x_2, x_3)\, dx_3 dx_2 dx_1 =$$

$$= (1 - e^{-x_1})(1 - e^{-2x_2})(1 - e^{-3x_3}) \quad \text{para } x_1, x_2, x_3 > 0$$

Ahora, es posible calcular por ejemplo, la probabilidad que los componentes duren a lo sumo 2000 horas:

$$\begin{aligned} P(X_1 \leq 2, X_2 \leq 2, X_3 \leq 2) &= F(2,2,2) \\ &= (1 - e^{-2})(1 - e^{-2\cdot 2})(1 - e^{-3\cdot 2}) = 0.8467 \end{aligned}$$

Función de distribución acumulada

Ejemplo 4: Sea (X, Y) un vector aleatorio continuo con función de densidad dada por:

$$f(x,y) = \frac{3(x^2+y^2)}{16} \quad \text{si } 0 < x < y < 2$$

Campos de variación relacionados

Luego,

$$F_{X,Y}(x_0, y_0) = \int_0^{x_0} \int_x^{y_0} f(x,y)dy\,dx = \cdots = \frac{x^3y + xy^3 - x^4}{16}$$

siempre que $0 < x < y < 2$.

Función de distribución acumulada

En este ejemplo, hemos obtenido la función de distribución acumulada a partir de la función de densidad conjunta. También es posible hacer el camino inverso:

Si $\boldsymbol{X} = (X_1, X_2, \ldots, X_k)$ es un vector aleatorio continuo, entonces su función de densidad puede obtenerse a partir de la distribución acumulada, en todos los puntos donde la derivada exista, mediante la relación:

$$f(x_1, x_2, \ldots, x_k) = \frac{\partial^k F(x_1, x_2, \ldots, x_k)}{\partial x_1 \ldots \partial x_k}$$

En el ejemplo anterior:

$$f(x, y) = \frac{\partial^2}{\partial x, \partial y}\left(\frac{x^3 y + x y^3 - x^4}{16}\right) = \frac{3}{16}(x^2 + y^2)$$

Distribuciones marginales

Distribuciones marginales

Con cada vector aleatorio k-dimensional se asocian k variables aleatorias unidimensionales, y podría interesar conocer la ***distribución marginal*** de cada una de ellas.

¿Cómo obtener por ejemplo la función de probabilidad puntual de $\boldsymbol{X}$, conociendo la función de probabilidad conjunta del vector $(\boldsymbol{X}, \boldsymbol{Y})$?

<u>Ejemplo:</u> $\mathbf{X} = (X_1, X_2)$ vector aleat. discreto con función de probabilidad:

$\boldsymbol{X_1 \vert X_2}$	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0	0	2/16	2/16	4/16	2/16	2/16	0
1	1/16	0	1/16	0	1/16	0	1/16

Si interesa la distribución marginal de X_1, debemos calcular $P(X_1 = x_1)$ para $x_1 = 0, 1$

$$p_{X_1}(0) = P(X_1 = 0) = p_{\boldsymbol{X}}(0{,}2) + p_{\boldsymbol{X}}(0{,}3) + \cdots + p_{\boldsymbol{X}}(0{,}8) = 12/16$$
$$p_{X_1}(1) = P(X_1 = 1) = p_{\boldsymbol{X}}(1{,}2) + p_{\boldsymbol{X}}(1{,}3) + \cdots + p_{\boldsymbol{X}}(1{,}8) = 4/16$$

Distribuciones marginales

CASO DISCRETO:

Si (X, Y) es un vector aleatorio discreto bidimensional, la distribución marginal de cada una de las 2 variables aleatorias que componen el vector puede obtenerse como:

$$p_X(x) = \sum_{y \in R_Y} p_{X,Y}(x, y) \text{ para cada valor de } x \in R_X$$

Análogamente,

$$p_Y(y) = \sum_{x \in R_X} p_{X,Y}(x, y) \text{ para cada valor de } y \in R_Y$$

Estas definiciones pueden extenderse de forma natural al caso k-dimensional, sumando a través de las k-1 variables restantes.

Distribuciones marginales

CASO DISCRETO:

Volviendo al ejemplo, se tendría:

$X_1 \mid X_2$	2	3	4	5	6	7	8	$p_{X_1}(x_1)$
0	0	2/16	2/16	4/16	2/16	2/16	0	**12/16**
1	1/16	0	1/16	0	1/16	0	1/16	**4/16**
$p_{X_2}(x_2)$	**1/16**	**2/16**	**3/16**	**4/16**	**3/16**	**2/16**	**1/16**	**1**

En el caso k-dimensional, puede pensarse incluso en la **distribución marginal** no de una, sino **de un subgrupo de las variables** que componen el vector. Sea por ejemplo, el vector $\boldsymbol{X} = (X_1, X_2, X_3)$. La distribución marginal del vector (X_1, X_2), por ejemplo, se obtiene como:

$$p_{X_1,X_2}(x_1, x_2) = \sum_{x_3 \in R_{X_3}} p(x_1, x_2, x_3), \qquad \text{para cada } x_1, x_2 \in R_{X_1,X_2}$$

Distribuciones marginales

CASO CONTINUO:

Si (X, Y) es un vector aleatorio continuo bidimensional, la ***distribución marginal de la variable aleatoria X*** puede obtenerse como:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}\,(x, y)\,dy, \quad \text{con } x \in R_X$$

Análogamente, la ***distribución marginal de Y***:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}\,(x, y)\,dx, \quad \text{con y} \in R_Y$$

Estas definiciones pueden extenderse de forma natural al caso k-dimensional, integrando a través de las k-1 variables restantes, o también al caso en que interesa obtener la distribución marginal de un subgrupo de las variables que componen el vector, integrando a través del resto de las variables en el vector.

Distribuciones marginales

CASO CONTINUO:

Ejemplo: retomamos el problema de los tiempos de falla de los componentes eléctricos, donde se tenía:

$$f(x_1, x_2, x_3) = 6e^{-x_1-2x_2-3x_3} \quad \text{si } x_1, x_2, x_3 > 0$$

Entonces, por ejemplo, la distribución marginal de X_1 viene dada por:

$$f_{X_1}(x_1) = \int_0^\infty \int_0^\infty 6e^{-x_1-2x_2-3x_3}\, dx_2\, dx_3 = e^{-x_1}, \quad \text{con } x_1 > 0$$

O bien, la distribución marginal de (X_1, X_3):

$$f_{X_1,X_3}(x_1, x_3) = \int_0^\infty 6e^{-x_1-2x_2-3x_3}\, dx_2 = 3e^{-x_1-3x_3}, \quad \text{con } x_1, x_3 > 0$$

Variables aleatorias independientes

Variables aleatorias independientes

En unidades anteriores definimos que dos sucesos A y B son independientes si $P(A \cap B) = P(A)P(B)$, es decir, si la probabilidad de que ocurran A y B simultáneamente puede factorizarse como el producto de las probabilidades individuales o marginales.

Este mismo concepto puede trasladarse a las variables aleatorias. Intuitivamente, diremos que ***dos variables aleatorias X e Y son independientes*** si el resultado de X no influye sobre el resultado de Y.

Ejemplos:

- En cierto juego de azar se lanzan simultáneamente una moneda y un dado. Si X representa el resultado del dado ($R_X = \{1,2,3,4,5,6\}$) e Y representa el resultado de la moneda ($X = 1$ para cara, $X = 0$ para cruz), ¿podrían X e Y ser independientes?
- Un material radioactiva emite partículas α. Sea X el número de partículas emitidas durante la primera hora de observación e Y el número de partículas emitidas en la segunda hora de hora de observación, ¿podrían X e Y ser independientes?

Variables aleatorias independientes

Diremos que las variables aleatorias del vector $\boldsymbol{X} = (X_1, X_2, \dots, X_k)$ son ***independientes*** si y sólo si:

$$F_{X_1,X_2,\dots,X_k}(x_1, x_2, \dots, x_k) = F_{X_1}(x_1) \dots F_{X_k}(x_k) \quad \forall\, \boldsymbol{x} \in R_{\boldsymbol{X}}$$

Si el vector aleatorio es discreto, esta condición es equivalente a plantear la igualdad en términos de las funciones de probabilidad:

$$p_{X_1,X_2,\dots,X_k}(x_1, x_2, \dots, x_k) = p_{X_1}(x_1) \dots p_{X_k}(x_k) \quad \forall\, \boldsymbol{x} \in R_{\boldsymbol{X}}$$

Si el vector aleatorio es continuo, la condición puede plantearse sobre las funciones de densidad:

$$f_{X_1,X_2,\dots,X_k}(x_1, x_2, \dots, x_k) = f_{X_1}(x_1) \dots f_{X_k}(x_k) \quad \forall\, \boldsymbol{x} \in R_{\boldsymbol{X}}$$

Variables aleatorias independientes

Ejemplos en el caso discreto:

$X_1 \mid X_2$	2	3	4	5	6	7	8	$p_{X_1}(x_1)$
0	0	2/16	2/16	4/16	2/16	2/16	0	**12/16**
1	1/16	0	1/16	0	1/16	0	1/16	**4/16**
$p_{X_2}(x_2)$	**1/16**	**2/16**	**3/16**	**4/16**	**3/16**	**2/16**	**1/16**	**1**

En todas las celdas se observa que: $p_X(x_1, x_2) \neq p_{X_1}(x_1)p_{X_2}(x_2)$
Por lo tanto, X_1 **y** X_2 **no son variables aleatorias independientes**.

$X \mid Y$	0	1	2	$p_X(x)$
0	1/16	1/8	1/16	**1/4**
1	1/16	1/8	1/16	**1/4**
2	1/8	1/4	1/8	**1/2**
$p_Y(y)$	**1/4**	**1/2**	**1/4**	**1**

En este caso, la función de probabilidad conjunta es igual al producto de las marginales para todo punto del recorrido del vector (es decir, para todas las celdas de la tabla). Por lo tanto, X **e** Y **son variables aleatorias independientes**.

Variables aleatorias independientes

Ejemplos en el caso continuo:

$$\boldsymbol{f(x_1, x_2, x_3) = 6e^{-x_1-2x_2-3x_3} \quad si\ x_1, x_2, x_3 > 0}$$

$$f_{X_1}(x_1) = e^{-x_1} \quad \text{si } x_1 > 0$$
$$f_{X_2}(x_2) = 2e^{-2x_2} \quad \text{si } x_2 > 0$$
$$f_{X_3}(x_3) = 3e^{-3x_3} \quad \text{si } x_3 > 0$$

$$f(x_1, x_2, x_3) = \prod_{i=1}^{3} f_{x_i}(x_i)$$

X_1 y X_2 son independientes

$$\boldsymbol{f(x, y) = \frac{3(x^2 + y^2)}{16} \quad \text{si } 0 < x < y < 2}$$

$$f_X(x) = \frac{3}{8}\left(x^2 - \frac{2}{3x^3} + \frac{4}{3}\right) \quad \text{si } 0 < x < 2$$
$$f_Y(y) = \frac{1}{4}y^3 \quad \text{si } 0 < y < 2$$

$$f(x, y) \neq f_X(x) f_Y(y)$$

X e Y no son independientes

Distribuciones condicionales

Distribuciones condicionales

El concepto de ***probabilidad condicional*** puede extenderse también de manera natural a las distribuciones de variables aleatorias.

<u>Para vectores bidimensionales:</u>

Sea (X, Y) un **vector aleatorio discreto** con función de probabilidad conjunta $p(x, y)$, y sean $p_X(x)$ y $p_Y(y)$ las distribuciones marginales de X e Y.

La distribución condicional de X dado $Y = y$ se define como:

$$p_{X|Y=y}(x) = P(X = x|Y = y) = \frac{p(x, y)}{p_Y(y)} \qquad \text{si } p_Y(y) \neq 0$$

Y la de Y dado $X = x$ como:

$$p_{Y|X=x}(y) = P(Y = y|X = x) = \frac{p(x, y)}{p_X(x)} \qquad \text{si } p_X(x) \neq 0$$

Distribuciones condicionales

<u>Ejemplo:</u> en el experimento del lanzamiento del dado teníamos las siguientes funciones de probabilidad conjunta y marginales:

$X_1 \mid X_2$	2	3	4	5	6	7	8	$p_{X_1}(x_1)$
0	0	2/16	2/16	4/16	2/16	2/16	0	**12/16**
1	1/16	0	1/16	0	1/16	0	1/16	**4/16**
$p_{X_2}(x_2)$	**1/16**	**2/16**	**3/16**	**4/16**	**3/16**	**2/16**	**1/16**	**1**

Podría interesar por ejemplo, la probabilidad de lograr resultado doble, sabiendo que se obtuvieron 6 puntos entre ambos lanzamientos…

$$P(X_1 = 1 \mid X_2 = 6) = \frac{P(X_1 = 1 \cap X_2 = 6)}{P(X_2 = 6)} = \frac{p_{\mathbf{X}}(1,6)}{p_{X_2}(6)} = \frac{1/16}{3/16} = \frac{1}{3}$$

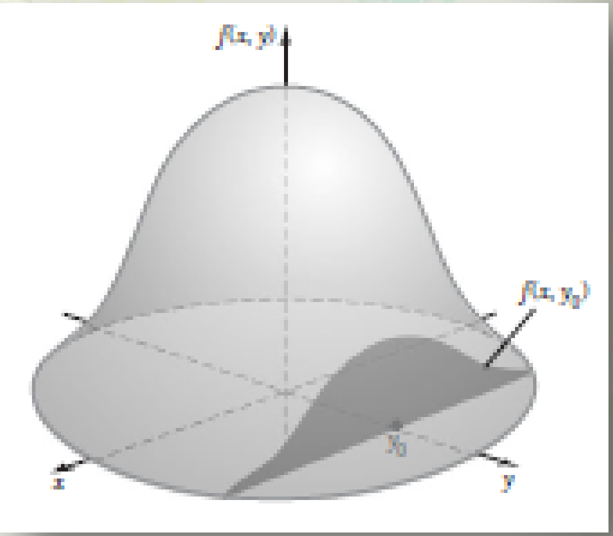
Y si obtuvieron 7 puntos en lugar de 6…?

Distribuciones condicionales

Análogamente, siendo (X, Y) un **vector aleatorio continuo** con función de densidad conjunta $f(x, y)$, y $f_X(x)$ y $f_Y(y)$ las distribuciones marginales de X e Y, se define la función de densidad condicional de X dado Y, o de Y dado X, respectivamente como:

$$f_{X|Y=y}(x) = \frac{f(x,y)}{f_Y(y)} \quad \text{si } f_Y(y) > 0; \qquad f_{Y|X=x}(y) = \frac{f(x,y)}{f_X(x)} \quad \text{si } f_X(x) > 0$$

<u>**OBS:**</u> En el caso discreto, la distribución condicional representa realmente una probabilidad. <u>**No es así en el caso continuo**</u>, la función de densidad condicional no es una probabilidad, sino que es una densidad unidimensional que resulta de intersectar la densidad conjunta con un plano de la forma $X = x$ o $Y = y$ según cada caso.



Distribuciones condicionales

Ejemplo: para el vector aleatorio con función de densidad dada por:

$$f(x,y) = \frac{3(x^2+y^2)}{16} \quad si \ \ 0 < x < y < 2$$

puede obtenerse, por ejemplo, la distribución condicional de X dado Y, la cual vendría dada por:

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\frac{3(x^2+y^2)}{16}}{\frac{1}{4}y^3} = \frac{3(x^2+y^2)}{4y^3}, \qquad \text{con } \ 0 < x < 2$$

Distribuciones condicionales

Para vectores k-dimensionales, puede darse una definición general:

Sea $\boldsymbol{X}$ un vector aleatorio (continuo o discreto) k-dimensional, dividido en dos subvectores $\boldsymbol{X}^{(1)}$ y $\boldsymbol{X}^{(2)}$. La distribución condicional de $\boldsymbol{X}^{(1)}$ dado $\boldsymbol{X}^{(2)} = \boldsymbol{x}^{(2)}$, puede escribirse, siempre que los denominadores sean positivos, como:

$$\boldsymbol{f}_{\boldsymbol{X}^{(1)}|\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}}\left(\boldsymbol{x}^{(1)}\right) = \frac{f(\boldsymbol{x})}{f_{\boldsymbol{X}^{(2)}}(\boldsymbol{x}^{(2)})}$$

O bien:

$$p_{\boldsymbol{X}^{(1)}|\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}}\left(\boldsymbol{x}^{(1)}\right) = \frac{p(\boldsymbol{x})}{p_{\boldsymbol{X}^{(2)}}(\boldsymbol{x}^{(2)})}$$

OBS: las distribuciones marginales se obtienen de las distribuciones conjuntas, pero no es cierto que con las distribuciones marginales pueda reconstruirse una distribución conjunta, a menos que las variables sean independientes. Esto sí puede hacerse a partir de las distribuciones condicionales.

Distribuciones condicionales

<u>**Ejemplo:**</u> en el problema de los tiempos de falla de los componentes eléctricos, se tenía la siguiente densidad conjunta:

$$f(x_1, x_2, x_3) = 6e^{-x_1-2x_2-3x_3} \quad \text{si } x_1, x_2, x_3 > 0$$

Entonces, por ejemplo, puede obtenerse la distribución condicional de X_2 dado X_1 y X_3, la cual vendría dada por:

$$f_{X_2|(X_1,X_3)}(x_1) = \frac{f_X(x_1, x_2, x_3)}{f_{X_1,X_3(x_1,x_3)}} = \frac{6e^{-x_1-2x_2-3x_3}}{3e^{-x_1-3x_3}} = 2e^{-2x_2}, \qquad \text{con } x_2 > 0$$

¿Puede explicar el resultado obtenido?

Caracterización de vectores aleatorios

Caracterización de los vectores aleatorios

Al igual que en el caso unidimensional, la distribución conjunta de un vector aleatorio puede caracterizarse a través de ciertos parámetros. Veremos en particular el valor esperado y la matriz de variancias y covariancias.

Esperanza de un vector aleatorio.

Sea $\boldsymbol{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio (discreto o continuo). Su ***esperanza*** se define como el vector numérico que tiene por elementos a los correspondientes valores esperados de las variables X_i:

$$\boldsymbol{E}(\boldsymbol{X}) = E(X_1, X_2, \dots, X_k) = \big(E(X_1), E(X_2), \dots, E(X_K)\big)$$

La interpretación de esta característica es análoga al caso univariado, sólo que ahora el centro de gravedad de la distribución es un punto en el espacio k-dimensional.

Caracterización de los vectores aleatorios

Propiedades de la esperanza:

Algunas propiedades de la esperanza de variables aleatorias, vistas anteriormente, eran:

- $E(c) = c$, con c constante real.
- $E(aX + b) = aE(X) + b$, con a y b constantes reales.
- $E(X + Y) = E(X) + E(Y)$, siendo X e Y variables aleatorias.

Combinando éstas, teníamos en forma general:

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

Con los conceptos vistos ahora podemos agregar:

- $E(XY) = E(X)E(Y)$, siempre que X e Y sean variables aleatorias independientes.

Caracterización de los vectores aleatorios

Variancia de un vector aleatorio.

Sea $\boldsymbol{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio (discreto o continuo). Su ***variancia*** se define como la matriz cuadrada de dimensión $k \times k$, cuyos elementos diagonales son las respectivas variancias de cada variable X_i en el vector, y los elementos no diagonales son las covariancias entre los distintos pares de variables. Recibe el nombre de ***matriz de variancias y covariancias***:

$$\boldsymbol{V}(\boldsymbol{X}) = \begin{pmatrix} V(X_1) & \cdots & Cov(X_1, X_k) \\ \vdots & \ddots & \vdots \\ Cov(X_k, X_1) & \cdots & V(X_k) \end{pmatrix}$$

donde la covariancia entre dos variables aleatorias es una medida del nivel de asociación lineal entre las variables. En efecto, es el numerador del coeficiente de correlación lineal que hemos visto en unidades anteriores.

Caracterización de los vectores aleatorios

Propiedades de la varianza:

Algunas propiedades de la varianza de variables aleatorias, vistas anteriormente, eran:

- $V(c) = 0$, con c constante real.
- $V(aX + b) = a^2V(X)$, con a y b constantes reales.

A diferencia de la esperanza, la varianza de la suma de variables no siempre es la suma de las varianzas individuales. En el caso general, resulta:

- $V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$

Pero si X e Y son independientes, entonces resulta:

- $V(X + Y) = V(X) + V(Y)$

Caracterización de los vectores aleatorios

Covariancia entre dos variables aleatorias.

Sean X e Y dos variables aleatorias. La ***covariancia*** entre ellas se define como:

$$Cov(X,Y) = E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big] = E(XY) - E(X)E(Y)$$

Algunas propiedades de esta medida son:

- Si X e Y son variables aleatorias independientes $\Rightarrow Cov(X,Y) = 0$
  No vale su recíproco.

Si a y b son constantes reales, entonces:

- $Cov(aX,bY) = ab\, Cov(X,Y)$
- $Cov(X + a, Y + b) = Cov(X,Y)$

Como ya hemos visto, el ***coeficientes de correlación*** se define como:

$$\rho_{XY} = \frac{E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big]}{\sqrt{V(X)V(Y)}} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

Caracterización de los vectores aleatorios

Esperanza de una función de un vector aleatorio.

Sea $\boldsymbol{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio (discreto o continuo). Sea $Y = g(\boldsymbol{X})$ una variable unidimensional, obtenida como función de las variables que componen el vector $\boldsymbol{X}$. La ***esperanza de la nueva variable Y*** se define en función de la distribución conjunta del vector $\boldsymbol{X}$:

Si $\boldsymbol{X}$ es discreto:

$$E(Y) = E\big(g(X_1, \dots, X_k)\big) = \sum_{x_1} \dots \sum_{x_k} g(x_1, \dots, x_k) p(x_1, \dots, x_k)$$

Si $\boldsymbol{X}$ es continuo:

$$E(Y) = E\big(g(X_1, \dots, X_k)\big) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_k) f(x_1, \dots, x_k)\, dx_1 \dots dx_k$$

Caracterización de los vectores aleatorios

Esperanza condicional

De la misma manera que hemos definido el valor esperado de una variable aleatoria *marginalmente*, podemos definir su valor esperado en términos de la distribución condicional. Es decir, el valor esperado de una variable para valores fijos de otra variable.

Sea (X, Y) un vector aleatorio bidimensional (discreto o continuo). Entonces la ***esperanza condicional de $\boldsymbol{X}$ dado $\boldsymbol{Y = y}$*** está dada por:
Si (X, Y) es discreto:

$$E(X|Y = y) = \sum_{x \in R_X} x \, p_{X|Y=y}(x)$$

Si (X, Y) es continuo:

$$E(X|Y = y) = \int_{-\infty}^{\infty} x \, f_{X|Y=y}(x) \, dx$$

Bibliografía

- Agresti A., Franklin C. (2009) The art and science of learning from data. 2º Ed. New Jersey. Pearson Prentice Hall.
- Meyer P. (1998). Probabilidad y Aplicaciones Estadísticas. Edición revisada. Addison-Wesley Iberoamericana.
- Leon-Garcia A. (2008). Probability, Statistics and Random Processes for Electrical Engineering. 3° Ed. Pearson Prentice Hall.
- DeGroot M, Schervish M. (2012). Probability and Statistics. 4° Ed. Pearson Education Inc.
- Casella G., Berger R. (2002). Statistical Inference. 2° Ed. Duxbury Press.