



Universidad de Chile
Facultad de Ciencias Físicas y Matemáticas
Departamento de Ingeniería Matemática
MA3403 - Probabilidades y Estadística

Vectores Aleatorios

Alberto Vera Azócar, albvera@ing.uchile.cl

1. Vectores Aleatorios bi-variados

Nuestro interés está en variables aleatorias X e Y que no son necesariamente independientes, trabajaremos con vectores bi-variados, las generalizaciones a n variables no son difíciles de extender y por tanto quedan propuestas.

Definición 1.1. Diremos que el par (X, Y) es un vector aleatorio si X e Y representan variables aleatorias.

La función que caracteriza el comportamiento del vector aleatorio es la distribución conjunta, ya que entrega la información de como se comporta el vector en cada componente y además el comportamiento cuando interactúan las componentes (uno pensaría que es más difícil obtener $P(X > 2 \wedge Y < 1)$ que $P(X > 2)$).

Definición 1.2. Sean $X, Y : \Omega \rightarrow \mathbb{R}$ variables aleatorias, se define la distribución conjunta como

$$F_{XY}(x, y) := P(X \leq x, Y \leq y)$$

Las leyes marginales se refieren siempre a como se comporta una v.a. sin tener en cuenta el resultado de las demás, así tenemos:

$$\begin{aligned} P(X \leq x) &= P(X \leq x, Y \leq \infty) \\ &= \lim_{n \rightarrow \infty} P(X \leq x, Y \leq n) \\ &= \lim_{n \rightarrow \infty} F_{XY}(x, n) \end{aligned}$$

Definición 1.3. Sean $X, Y : \Omega \rightarrow \mathbb{R}$ v.a., definimos las distribuciones marginales como

$$F_X(x) := \lim_{y \rightarrow \infty} F_{XY}(x, y) \quad F_Y(y) := \lim_{x \rightarrow \infty} F_{XY}(x, y)$$

1.1. Caso bi-discreto

Definición 1.4. Diremos que (X, Y) es un vector discreto si X e Y son v.a. discretas, es decir si existen $A_x, A_y \subseteq \mathbb{R}$ a lo más numerables tales que $P(X \in A_x) = 1$ y $P(Y \in A_y) = 1$.

A los conjuntos de la definición los llamamos “soportes” de la v.a. y representan “donde toman valores las v.a. con probabilidad 1”, denotamos $S(X) = A_x$ y $S(Y) = A_y$.

De la definición es claro que también existe un conjunto $A \subseteq \mathbb{R}^2$ a lo más numerable tal que $P((X, Y) \in A) = 1$, pues basta tomar por ejemplo $A = A_x \times A_y$.

Se busca caracterizar al vector (X, Y) , en general $F_{XY}(x, y)$ es suficiente, pero no es útil para los cálculos, no así como la ley conjunta:

Definición 1.5. Sea (X, Y) un vector aleatorio discreto, se define la ley conjunta como

$$p_{ij} := P(X = i, Y = j)$$

1.1.1. Leyes Marginales

Las leyes marginales que denotaremos $p_i := \mathbb{P}(X = i)$ se obtienen particionando el espacio: dado que los eventos $Y = j$ e $Y = j'$ son disjuntos $\forall j \neq j'$ (una v.a. no puede tomar dos valores a la vez) y recordando que dada una partición $\{A_i\}_{i \in \mathbb{N}}$ del conjunto A se tiene

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \mathbb{P}(A_i)$$

Proposición 1.1. *Sea (X, Y) un vector aleatorio bi-discreto con ley conjunta p_{ij} , entonces las leyes marginales están dadas por*

$$\begin{aligned} p_i &= \sum_{j \in S(Y)} \mathbb{P}(X = i, Y = j) = \sum_j p_{ij} \\ p_j &= \sum_{i \in S(X)} \mathbb{P}(X = i, Y = j) = \sum_i p_{ij} \end{aligned} \quad (1)$$

La fórmula (1) debe interpretarse como “para un i fijo, sumo p_{ij} para todos los posibles valores de j ”, o bien “la probabilidad de que $X = i$ es la probabilidad de que $X = i$ e Y tome cualquier valor”.

Ejemplo 1.1. *Una pareja va al cine y deciden al azar, con prob $p \in (0, 1)$, si ver una película que cuesta 1. Suponemos que ambos entran a ver la película o ninguno lo hace. Sea X el dinero que gasta el hombre e Y el que gasta la mujer, se pide calcular las leyes marginales.*

Solución: Tenemos que

$$\begin{aligned} \mathbb{P}(X = 1, Y = 0) &= \mathbb{P}(X = 0, Y = 1) = 0 \\ \mathbb{P}(X = 1, Y = 1) &= p \\ \mathbb{P}(X = 0, Y = 0) &= 1 - p \end{aligned}$$

Por simetría es claro que la ley marginal de X es la misma que la de Y , gracias a la identidad (1) tenemos que:

$$\mathbb{P}(X = 1) = \mathbb{P}(X = 1, Y = 0) + \mathbb{P}(X = 1, Y = 1) = p$$

Sigue que tanto X como Y son v.a. Bernoulli de parámetro p . En este caso observamos que es totalmente falso que $\mathbb{P}(X = i, Y = j) = \mathbb{P}(X = i)\mathbb{P}(Y = j)$.

Propuesto: En el ejemplo anterior mostrar que la distribución conjunta está dada por

$$F_{XY}(x, y) = \begin{cases} 1 & \text{si } x \geq 1 \wedge y \geq 1 \\ 0 & \text{si } x < 0 \vee y < 0 \\ p & \text{si } x, y \in [0, 1) \end{cases}$$

1.1.2. Distribuciones Marginales

Dadas las leyes marginales, se obtienen directamente las distribuciones marginales:

$$\begin{aligned} F_X(x) &\triangleq \mathbb{P}(X \leq x) \\ &= \sum_{i \leq x} \mathbb{P}(X = i) \\ &= \sum_{i \leq x} \sum_{j \in S(Y)} p_{ij} \end{aligned} \quad (2)$$

donde en (2) se utilizó (1). Análogamente se tiene

$$F_Y(y) = \sum_{j \leq y} \sum_{i \in S(X)} p_{ij}$$

Siguiendo con el razonamiento de particionar, la función de distribución conjunta se calcula como

$$F_{XY}(x, y) = \sum_{i \leq x} \sum_{j \leq y} p_{ij}$$

1.2. Caso bi-continuo

Aceptaremos la siguiente definición poco formal, pues para una versión completa habría que introducir conceptos que escapan al alcance del curso.

Definición 1.6. Diremos que $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ es un vector conjuntamente continuo si existe una función $f_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ que llamaremos densidad conjunta tal que $\forall A \subseteq \mathbb{R}^2$ se tiene

$$\mathbb{P}((X, Y) \in A) = \int \int_A f_{XY}(x, y) dx dy$$

Observación: En la literatura se puede encontrar la definición anterior con el nombre de “variable aleatoria absolutamente continua” y definen v.a. continua equivalente a que F_{XY} sea una función continua, estas definiciones NO son equivalentes.

Las deducciones en el caso discreto las realizamos con argumentos de partición, usando implícitamente la σ -aditividad, que es válida para uniones numerables. Como en el caso continuo no podemos hacer ésto, las deducciones se harán vía derivación.

Proposición 1.2. Sea (X, Y) vector aleatorio bi-continuo con densidad conjunta f_{XY}

1. Claramente la integral en todo el soporte da 1, de la definición tomando $A = \mathbb{R}^2$:

$$\int \int_{\mathbb{R}^2} f_{XY}(x, y) dx dy = 1$$

2. De la definición, eligiendo $A = (-\infty, x) \times (-\infty, y)$

$$F_{XY}(x, y) \triangleq \mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(x, y) dx dy$$

3. Derivando F_{XY} se puede obtener la densidad conjunta

$$\frac{\partial^2}{\partial x \partial y} F_{XY}(x, y) = f_{XY}(x, y)$$

1.2.1. Densidades Marginales

Sabemos que

$$\begin{aligned} F_X(x) &\triangleq \lim_{y \rightarrow \infty} F_{XY}(x, y) \\ &= \lim_{y \rightarrow \infty} \int_{-\infty}^x \int_{-\infty}^y f_{XY}(x, y) dx dy \\ &= \int_{-\infty}^x \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy \end{aligned}$$

Usando el Teorema Fundamental del Cálculo

$$\frac{d}{dx}F_X(x) = \frac{d}{dx} \int_{-\infty}^x \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

Por lo anterior parece natural definir

Definición 1.7. Sea (X, Y) vector aleatorio bi-continuo con densidad f_{XY} , definimos las densidades marginales como

$$f_X(x) := \int_{\mathbb{R}} f_{XY}(x, y) dy \quad f_Y(y) := \int_{\mathbb{R}} f_{XY}(x, y) dx$$

Antes de ver un ejemplo introducimos el concepto de indicatriz, esta función es en extremo útil para los cálculos, además se utiliza también en otros ámbitos, como en lógica difusa, estadística y optimización.

Definición 1.8. Sea $A \subseteq \mathbb{R}^n$, la función indicatriz $\mathbb{1}_{\{x \in A\}}$ se define como

$$\mathbb{1}_{\{x \in A\}} := \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

Ejemplo 1.2. Consideremos el vector aleatorio (X, Y) con densidad $f_{XY}(x, y) = c \frac{1}{x^2 y^2} \mathbb{1}_{\{x \geq y, y \geq 1\}}$. Se pide calcular las densidades marginales.

Solución: Para resolver el problema hay que calcular c , siempre es mejor obtener las marginales primero y después obtener c , pues así no tendremos que trabajar dos veces.

Primero calculamos f_X , para esto debemos tomar $x \geq 1$ fijo e integrar respecto a y :

$$\begin{aligned} f_X(x) &\triangleq \int_{\mathbb{R}} f_{XY}(x, y) dy \\ &= \int_{\mathbb{R}} c \frac{1}{x^2 y^2} \mathbb{1}_{\{x \geq y, y \geq 1\}} dy \\ &= \int_1^x c \frac{1}{x^2 y^2} dy \\ &= c \frac{1}{x^2} \left(\frac{-1}{y} \right)_{y=1}^x \\ \Rightarrow f_X(x) &= c \frac{1}{x^2} \left(1 - \frac{1}{x} \right) \quad x \geq 1 \end{aligned}$$

Para calcular la densidad marginal f_Y nos damos un $y \geq 1$ fijo e integramos respecto a x :

$$\begin{aligned} f_Y(y) &= \int_{\mathbb{R}} c \frac{1}{x^2 y^2} \mathbb{1}_{\{x \geq y, y \geq 1\}} dx \\ &= c \frac{1}{y^2} \int_y^{\infty} \frac{1}{x^2} dx \\ &= c \frac{1}{y^2} \left(\frac{-1}{x} \right)_{x=y}^{\infty} \\ \Rightarrow f_Y(y) &= c \frac{1}{y^3} \quad y \geq 1 \end{aligned}$$

Finalmente

$$f_Y(y) = \frac{c}{y^3} \mathbb{1}_{\{y \geq 1\}} \quad f_X(x) = \frac{c}{x^2} \left(1 - \frac{1}{x}\right) \mathbb{1}_{\{x \geq 1\}}$$

Para calcular c se utiliza el hecho de que la integral en $\mathbb{R}$ de una densidad vale 1:

$$\begin{aligned} 1 &= \int_{\mathbb{R}} f_Y(y) dy \\ &= c \int_1^{\infty} \frac{1}{y^3} dy \\ &= c \frac{1}{2} \end{aligned}$$

Concluimos que $c = 2$. Si hubiéramos integrado f_X sobre $\mathbb{R}$ en vez de f_Y nos hubiera resultado mágicamente¹ lo mismo.

Propuesto: En el ejemplo anterior, muestre que $F_{XY}(x, y) = \frac{xy^2 - 2y^2 - x + 2y}{xy^2}$ cuando $x \geq y \geq 1$.

2. Independencia

Definición 2.1. Diremos que X e Y v.a. son independientes, denotado $X \perp Y$, si cumplen que para cualquier par de conjuntos A y B se tiene

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B)$$

La definición es difícil de trabajar, por lo que se hace necesario tener más caracterizaciones de independencia. Aceptaremos el siguiente Teorema sin demostración.

Teorema 2.1. Sea (X, Y) un vector aleatorio, las v.a. X e Y son independientes ssi

- 1) $F_{XY}(x, y) = F_X(x)F_Y(y)$
- 2) Es vector bi-discreto y $p_{ij} = p_i p_j$
- 3) Es vector bi-continuo y $f_{XY}(x, y) = f_X(x)f_Y(y)$

Las caracterizaciones 2 y 3 son las más útiles, pero la 1 es válida para casos generales, por ejemplo X continua e Y discreta.

Ejemplo 2.1. Estudiar la independencia de (X, Y) con ley $p_{ij} = \frac{1}{N^2}$ para $i, j \in \{1, 2, \dots, N\}$.

Solución: Calculamos la marginal de X :

$$p_i = \sum_{j=1}^N p_{ij} = \sum_{j=1}^N \frac{1}{N^2} = \frac{1}{N} \quad i = 1, 2, \dots, N$$

Resulta entonces que X tiene ley uniforme discreta entre 1 y N , de forma análoga $p_j = \frac{1}{N}$ para $j = 1, \dots, N$. En virtud del Teorema 2.1 tenemos que $X \perp Y$ por cuanto $p_{ij} = p_i p_j$.

¹Por supuesto que no es coincidencia, se puede demostrar la equivalencia

3. Esperanza de Una Función

Definición 3.1. Sea (X, Y) un vector aleatorio y sea $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ función a valores reales. Definimos la esperanza de g como

$$\mathbb{E}(g(X, Y)) := \sum_{i \in S(X)} \sum_{j \in S(Y)} g(i, j) p_{ij} \quad \text{si } (X, Y) \text{ es bi-discreto} \quad (3)$$

$$\mathbb{E}(g(X, Y)) := \int \int_{\mathbb{R}^2} g(x, y) f_{XY}(x, y) dx dy \quad \text{si } (X, Y) \text{ es bi-continuo} \quad (4)$$

El siguiente teorema es una implicancia de la independencia, cuidado con la recíproca: es falso que $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) \Rightarrow X \perp Y$.

Teorema 3.1. Sean X, Y v.a. independientes, entonces $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

DEMOSTRACIÓN. Haremos la demostración para el caso continuo, el caso discreto es análogo y queda propuesto.

Sean f_X, f_Y las densidades, sabemos por independencia que $f_X f_Y$ es la densidad conjunta. Tenemos que calcular la esperanza de la función $g(x, y) = xy$, por definición:

$$\begin{aligned} \mathbb{E}(XY) &\triangleq \int \int_{\mathbb{R}^2} xy f_{XY}(x, y) dx dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} xy f_X(x) f_Y(y) dx dy \\ &= \int_{\mathbb{R}} x f_X(x) dx \int_{\mathbb{R}} y f_Y(y) dy \\ &\triangleq \mathbb{E}(X)\mathbb{E}(Y) \end{aligned}$$

□

Ahora tomamos una función que llega a $\mathbb{R}^n$, es simplemente una extensión de la definición que ya vimos:

Definición 3.2. Sea (X, Y) un vector aleatorio, y sea $g : \mathbb{R}^2 \rightarrow \mathbb{R}^m$, se define

$$\mathbb{E}(g(X, Y)) := (\mathbb{E}(g_1(X, Y)), \mathbb{E}(g_2(X, Y)), \dots, \mathbb{E}(g_m(X, Y)))$$

La esperanza (“a secas”) del vector (X, Y) es mediante la función $g(x, y) = (x, y)$, con lo que

$$\mathbb{E}((X, Y)) := (\mathbb{E}(X), \mathbb{E}(Y))$$

Ejemplo 3.1. A un negocio llegan dos tipos de clientes en un día, denotemos X la cantidad de clientes de tipo 1, que reportan utilidad π_1 , Y la cantidad de clientes de tipo 2 que reportan utilidad π_2 . Sabiendo que $\mathbb{P}(X = i, Y = j) = c(i + j)$ para $i, j = 0, 1, 2$ se pide calcular la utilidad esperada del negocio.

Solución: Definamos para esto $g(x, y)$ como la utilidad dado que llegan x clientes del tipo 1 e y

clientes del tipo 2. Se tiene que $g(x, y) = x\pi_1 + y\pi_2$, sigue que la utilidad esperada es según (3)

$$\begin{aligned}
\mathbb{E}(g(X, Y)) &\triangleq \sum_{i=0}^2 \sum_{j=0}^2 g(i, j) p_{ij} \\
&= c \sum_{i=0}^2 \sum_{j=0}^2 (i\pi_1 + j\pi_2)(i + j) \\
\mathbb{E}(g(X, Y)) \frac{1}{c} &= \sum_{i=0}^2 \sum_{j=0}^2 (i^2\pi_1 + j^2\pi_2 + ji(\pi_1 + \pi_2)) \\
&= \sum_{i=0}^2 \sum_{j=0}^2 i^2\pi_1 + \sum_{i=0}^2 \sum_{j=0}^2 j^2\pi_2 + \sum_{i=0}^2 \sum_{j=0}^2 ji(\pi_1 + \pi_2) \\
&= 3\pi_1 \sum_{i=0}^2 i^2 + 3\pi_2 \sum_{j=0}^2 j^2 + (\pi_1 + \pi_2) \sum_{i=0}^2 i \sum_{j=0}^2 j \\
&= 15\pi_1 + 15\pi_2 + (\pi_1 + \pi_2) \cdot 3 \cdot 3 \\
\Rightarrow \mathbb{E}(g(X, Y)) &= 24c(\pi_1 + \pi_2)
\end{aligned}$$

Sólo para fijar ideas calculemos la ley de X , intuitivamente la esperanza de la utilidad asociada a los clientes de tipo 1 es $\pi_1 \mathbb{E}(X)$.

$$\begin{aligned}
\mathbb{P}(X = i) &= \sum_{j=0}^2 c(i + j) \\
&= c(i + (i + 1) + (i + 2)) \\
\Rightarrow \mathbb{P}(X = i) &= 3c(i + 1)
\end{aligned}$$

Luego

$$\mathbb{E}(X) = 1 \cdot \mathbb{P}(X = 1) + 2 \cdot \mathbb{P}(X = 2) = 1 \cdot 6c + 2 \cdot 9c = 24c$$

Vemos que ésto es consecuente con el hecho de que la esperanza de la utilidad es $24c\pi_1 + 24c\pi_2$

Propuesto: Muestre que en el ejemplo anterior las v.a. no son independientes, que $c = \frac{1}{18}$ y en el caso $\pi_1 = 1, \pi_2 = 2$ se tiene que la varianza de la utilidad es $\frac{7}{3}$.

4. Función Generadora de Momentos

La función generadora de momentos nos será útil como herramienta para identificar distribuciones, se encuentran tabuladas las funciones generadoras de muchas funciones conocidas (Poisson, exponencial, etc), entonces para encontrar la distribución de una nueva variable aleatoria a veces puede ser conveniente calcular su función generadora para concluir utilizando las que son conocidas.

Definición 4.1. Sea X una v.a. y $n \in \mathbb{N}$, se define el momento de orden n de X como $\mathbb{E}(X^n)$.

Definición 4.2. Sea X una v.a., se define la función generadora de momentos de X como

$$\psi_X(t) := \mathbb{E}(e^{tX}) \quad (5)$$

el dominio de ψ_X es t tal que la esperanza en (5) converge.

La mayor gracia de la función generadora de momentos es el siguiente teorema, que aceptaremos sin demostración.

Teorema 4.1. Sean X, Y v.a., si $\psi_Y(t) = \psi_X(t)$ entonces X e Y tienen la misma distribución.

Las siguientes propiedades son muy útiles para los cálculos

Proposición 4.1. Sean X, Y v.a.

1. Si ψ_X es k veces derivable, entonces

$$\left. \frac{d^k}{dt^k} \psi_X(t) \right|_{t=0} = \mathbb{E}(X^k)$$

2. $X \perp Y \Rightarrow \psi_{X+Y}(t) = \psi_X(t) \cdot \psi_Y(t)$

DEMOSTRACIÓN. (1) queda propuesto, demostraremos (2):

Sean X, Y independientes, luego las v.a. $U = e^{tX}$ y $V = e^{tY}$ son independientes, por lo tanto por Teorema 3.1 $\mathbb{E}(UV) = \mathbb{E}(U)\mathbb{E}(V)$, tenemos entonces:

$$\psi_{X+Y}(t) \triangleq \mathbb{E}(e^{t(X+Y)}) = \mathbb{E}(UV) = \mathbb{E}(U)\mathbb{E}(V) = \psi_X(t)\psi_Y(t)$$

□

Ejemplo 4.1. Sean $X, Y \sim \mathcal{N}(0, 1)$, con $X \perp Y$, encuentre la distribución de $Z = X + Y$.

Solución: Se vio en clases que si $R \sim \mathcal{N}(\mu, \sigma^2)$, entonces su función generadora de momentos es $\psi_R(t) = e^{t\mu + \frac{\sigma^2 t^2}{2}}$. Luego en nuestro caso $\psi_X(t) = e^{\frac{t^2}{2}}$, que es igual a ψ_Y , utilizando la proposición anterior:

$$\psi_Z(t) = \psi_X(t)\psi_Y(t) = e^{\frac{t^2}{2}} e^{\frac{t^2}{2}} = e^{\frac{2t^2}{2}}$$

Identificamos que ψ_Z es la función generadora de momentos de una v.a. normal de media 0 y varianza 2, en virtud del Teorema 4.1 sigue que $Z \sim \mathcal{N}(0, 2)$.

5. Teorema del Cambio de Variables

El siguiente teorema es muy útil cuando queremos calcular la densidad de una función de una v.a., la demostración se vio en clases, por lo que no se repetirá acá.

Teorema 5.1. Sea (X, Y) un par conjuntamente continuo con densidad f_{XY} , sea $\mathcal{U} \subseteq \mathbb{R}^2$ abierto tal que $\mathbb{P}((X, Y) \in \mathcal{U}) = 1$. Sea $\varphi : \mathcal{U} \rightarrow \varphi(\mathcal{U})$ función $\in \mathcal{C}^1(\mathcal{U})$ biyectiva con $\det(J(\varphi)) \neq 0$. Entonces si denotamos $(Z, W) = \varphi(X, Y)$

$$f_{ZW}(z, w) = f_{XY}(\varphi^{-1}(z, w)) |\det(J(\varphi^{-1}))| \mathbb{1}_{\{(z, w) \in \varphi(\mathcal{U})\}}$$

Ejemplo 5.1. Sean $X, Y \sim \exp(\lambda)$ con $X \perp Y$. Se pide calcular la densidad del vector aleatorio (X^3, Y) .

Solución: Para lo anterior usaremos el teorema del cambio de variables, por lo que reconocemos todos los términos y comprobaremos las hipótesis.

Sea el abierto $\mathcal{U} := (0, \infty) \times (0, \infty)$, claramente $\mathbb{P}((X, Y) \in \mathcal{U}) = 1$. Definamos $\varphi(x, y) := (x^3, y)$, esta función es continuamente derivable, y es claramente biyectiva, más aún si denotamos $\varphi(x, y) = (z, w)$:

$$\varphi^{-1}(z, w) = (\sqrt[3]{z}, w)$$

Calculemos el jacobiano de φ^{-1} , si está bien definido en $\mathcal{U}$ quiere decir que $\det(J(\varphi)) \neq 0$.

$$J(\varphi^{-1}) = \begin{pmatrix} \frac{1}{3} z^{-\frac{2}{3}} & 0 \\ 0 & 1 \end{pmatrix}$$

Luego, como $(0, 0) \notin \mathcal{U}$ el jacobiano está bien definido y su determinante es $\frac{1}{3}z^{-\frac{2}{3}}$.
Tenemos también que $\varphi(\mathcal{U}) = (0, \infty) \times (0, \infty)$, luego $(z, w) \in \varphi(\mathcal{U}) \iff z, w > 0$.
Como $X \perp\!\!\!\perp Y$ su densidad conjunta es

$$f_{XY}(x, y) = f_X(x)f_Y(y) = \lambda e^{-\lambda x} \mathbf{1}_{\{x>0\}} \lambda e^{-\lambda y} \mathbf{1}_{\{y>0\}}$$

por Teorema del cambio de variable:

$$\begin{aligned} f_{ZW}(z, w) &= f_{XY}(\varphi^{-1}(z, w)) |det(J(\varphi^{-1}))| \mathbf{1}_{\{(z, w) \in \varphi(\mathcal{U})\}} \\ &= f_{XY}(\sqrt[3]{z}, w) \frac{1}{3} z^{-\frac{2}{3}} \mathbf{1}_{\{(z, w) \in \varphi(\mathcal{U})\}} \\ &= \lambda e^{-\lambda \sqrt[3]{z}} \mathbf{1}_{\{\sqrt[3]{z}>0\}} \lambda e^{-\lambda w} \mathbf{1}_{\{w>0\}} \frac{1}{3} z^{-\frac{2}{3}} \mathbf{1}_{\{(z, w) \in \varphi(\mathcal{U})\}} \\ &= \frac{\lambda^2}{3} z^{-\frac{2}{3}} e^{-\lambda \sqrt[3]{z}} e^{-\lambda w} \mathbf{1}_{\{\sqrt[3]{z}>0\}} \mathbf{1}_{\{w>0\}} \mathbf{1}_{\{(z, w) \in \varphi(\mathcal{U})\}} \\ &= \frac{\lambda^2}{3} z^{-\frac{2}{3}} e^{-\lambda \sqrt[3]{z}} e^{-\lambda w} \mathbf{1}_{\{z>0, w>0\}} \end{aligned}$$

Propuesto: En el ejemplo anterior mostrar que $Z \perp\!\!\!\perp W$ y que $\mathbb{E}(Z) = \frac{6}{\lambda^3}$.

6. Covarianza y Correlación

No siempre es fácil estudiar la independencia, por eso usaremos conceptos que tomarán mayor relevancia en la parte estadística, que además en sus versiones empíricas son los verdaderos objetos de estudio al momento de ver independencia.

Definición 6.1. Sean $X, Y : \Omega \rightarrow \mathbb{R}$ v.a. que cumplen $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$, se define la covarianza entre X e Y como

$$\text{cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

Bajo los supuestos de la definición la covarianza está bien definida ($\text{cov}(X, Y) < \infty$), ahora se puede calcular como esperanza de una función, por ejemplo si (X, Y) es un vector conjuntamente continuo, entonces de la identidad (4) tomando $g(x, y) = (x - \mathbb{E}(X))(y - \mathbb{E}(Y))$:

$$\text{cov}(X, Y) = \int \int_{\mathbb{R}^2} (x - \mathbb{E}(X))(y - \mathbb{E}(Y)) f_{XY}(x, y) dx dy$$

La covarianza es una medida de relación entre las variables, es adecuado pensar que si la covarianza es muy grande entonces las v.a. están fuertemente relacionadas; si es muy negativa por ejemplo se intuye que las v.a. están muy relacionadas y que además cuando una aumenta el valor la otra disminuye.

Para medir distintas variables, la covarianza puede ser engañosa por cuanto depende de la unidad de medida de las variables, por esto se “estandariza” como el coeficiente de correlación (lineal), este sí es comparable entre distintos pares de variables.

Definición 6.2. Sean $X, Y : \Omega \rightarrow \mathbb{R}$ v.a. con $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$, se define el coeficiente de correlación como

$$\rho(X, Y) := \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}}$$

El coeficiente de correlación cumple $-1 \leq \rho(X, Y) \leq 1$.

La propiedad más famosa de la covarianza es

$$X \perp\!\!\!\perp Y \Rightarrow \text{cov}(X, Y) = 0$$

Es importante recalcar que la implicancia es sólo hacia la derecha, $\text{cov}(X, Y) = 0 \not\Rightarrow X \perp\!\!\!\perp Y$.

7. Condicionalidad

Ya se ha visto que la probabilidad condicional induce una medida de probabilidad, es decir, dado un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ si tomamos a $B \subseteq \Omega$ con $\mathbb{P}(B) > 0$, entonces $\mathbb{P}(\cdot|B)$ es medida de probabilidad, donde:

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Lo que ahora nos motiva es reemplazar el evento B por el evento $Y = y$ donde Y es una v.a., en el caso discreto esto es más natural que en el caso continuo, pero luego uno se convence.

7.1. Caso Discreto

Definición 7.1. Sean $X, Y : \Omega \rightarrow \mathbb{R}$ v.a. discretas, se define la probabilidad de que $X = i$ dado que $Y = j$ cuando $\mathbb{P}(Y = j) > 0$ como

$$\mathbb{P}(X = i|Y = j) := \frac{\mathbb{P}(X = i, Y = j)}{\mathbb{P}(Y = j)}$$

Dada la definición anterior consideraremos como ley condicional a la colección de todos los valores de probabilidad condicional, denotaremos para $\mathbb{P}(Y = y) > 0$ a la ley condicional de X dado Y como

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

Observamos que para calcular la ley condicional necesitamos la ley conjunta p_{ij} , en efecto es fácil ver que

$$p_{X|Y}(i|j) = \frac{p_{ij}}{\sum_k p_{kj}}$$

Si tenemos la ley marginal de Y y la ley condicional de X dado Y , entonces tenemos toda la información que necesitamos sobre el par (X, Y) , pues de la definición:

$$\mathbb{P}(X = i, Y = j) = \mathbb{P}(X = i|Y = j)\mathbb{P}(Y = j) \quad (6)$$

si en ambos lados de (6) sumamos sobre j y aplicamos en el lado izquierdo la identidad (1):

$$\Rightarrow \mathbb{P}(X = i) = \sum_j \mathbb{P}(X = i|Y = j)\mathbb{P}(Y = j)$$

7.2. Caso Continuo

En este caso es un poco raro pensar en el observable $Y = y$, ya que si Y es v.a. continua, entonces $\forall y \in \mathbb{R}, \mathbb{P}(Y = y) = 0$, la intuición es que si bien eso es cierto, cuando uno echa a correr una v.a. igual se obtiene un resultado (alguna temperatura se observa por ejemplo) y este hecho da información sobre la ocurrencia de la v.a. X .

Definición 7.2. Sean $X, Y : \Omega \rightarrow \mathbb{R}$ v.a. conjuntamente continuas, es decir, que tienen densidad conjunta f_{XY} , se define la densidad condicional de X dado el observable $Y = y$ como

$$f_{X|Y}(x|y) := \frac{f_{XY}(x, y)}{f_Y(y)} \mathbb{1}_{\{f_Y(y) > 0\}}$$

Además definiremos para $A \subseteq \mathbb{R}$ la probabilidad de que $X \in A$ dado un observable como

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

La principal herramienta que tendremos para cálculo mediante probabilidades condicionales es el siguiente Teorema (la demostración se vio en clases, por tanto queda propuesta):

Teorema 7.1. Sean $X, Y : \Omega \rightarrow \mathbb{R}$ v.a. conjuntamente continuas, para $A \subseteq \mathbb{R}$ se tiene que

$$\mathbb{P}(X \in A) = \int_{\mathbb{R}} \mathbb{P}(X \in A|Y = y)f_Y(y)dy$$

7.3. Esperanza Condicional

Tendremos la esperanza definida análogamente que para el caso conjunto, solo que se reemplaza ley (densidad) conjunta por la condicional, así dada una función $g : \mathbb{R} \rightarrow \mathbb{R}$ se tiene en el caso discreto

$$\mathbb{E}(g(X)|y) := \sum_{i \in \mathbb{N}} g(i)p_{X|Y}(i|y)$$

y en el caso continuo

$$\mathbb{E}(g(X)|Y) := \int_{\mathbb{R}} g(x)f_{X|Y}(x|y)dx$$

La esperanza condicional debe entenderse como “el mejor estimador dado lo que estamos observando”, esto lo muestra el siguiente ejemplo:

Ejemplo 7.1. Hay dos máquinas, que fallan un tiempo X y un tiempo Y respectivamente, los cuales se distribuyen de acuerdo a la densidad $f_{XY}(x, y) = ce^{-xy}\mathbf{1}_{\{1 \geq x, y \geq 0\}}$. Se observa que la máquina 2 falló durante un tiempo de 1, ¿cuál es el tiempo esperado que fallará la máquina 1?

Solución: Procedemos calculando la ley marginal de Y , tomamos un $y > 0$ fijo:

$$f_Y(y) \triangleq \int_0^1 ce^{-xy}dx = c \frac{1 - e^{-y}}{y}$$

Por definición calculamos la densidad condicional para $1 > x, y > 0$:

$$f_{X|Y}(x|y) \triangleq ce^{-xy} \frac{y}{c(1 - e^{-y})} = \frac{ye^{-xy}}{1 - e^{-y}}$$

Luego la esperanza condicional dado un $y > 0$ es

$$\mathbb{E}(X|Y) \triangleq \int_0^1 xf_{X|Y}(x|y)dx = \int_0^1 x \frac{ye^{-xy}}{1 - e^{-y}}dx = \frac{1 + y - e^y}{y(1 - e^y)}$$

Por lo que, dado que se observó $Y = 1$, $\mathbb{E}(X|Y = 1) = \frac{2-e}{1-e} \approx 0,41$.

Propuesto: En el ejemplo anterior, muestre que $\mathbb{E}(X) = ce^{-1}$ (no calcule c). Muestre que las v.a. no son independientes, dado este hecho ¿esperaría que $\mathbb{E}(X) = \mathbb{E}(X|Y = 1)$?

Ya hemos visto varias herramientas, en particular se vio que dada la ley condicional se puede obtener la marginal, así para calcular $\mathbb{E}(X)$ se puede primero obtener la marginal y luego aplicar definición de esperanza, a veces esto complica, por lo que se tiene el siguiente Teorema que nos ayuda a calcular la esperanza de X dada la condicional (válido para v.a. continuas y discretas).

Teorema 7.2 (Ley de esperanzas iteradas). Sean $X, Y : \Omega \rightarrow \mathbb{R}$ v.a., se tiene que

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))$$

DEMOSTRACIÓN. Lo haremos para el caso continuo, el caso discreto queda propuesto.

Tomemos X, Y v.a. con densidad conjunta f_{XY} , tenemos que

$$\mathbb{E}(X|Y) \triangleq \int_{\mathbb{R}} x f_{X|Y}(x|y) dx = \int_{\mathbb{R}} x \frac{f_{XY}(x, y)}{f_Y(y)} \mathbb{1}_{\{f_Y(y) > 0\}} dx$$

sigue que

$$\begin{aligned} \mathbb{E}(\mathbb{E}(X|Y)) &\triangleq \int_{\mathbb{R}} \mathbb{E}(X|Y) f_Y(y) dy \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} x \frac{f_{XY}(x, y)}{f_Y(y)} \mathbb{1}_{\{f_Y(y) > 0\}} dx \right) f_Y(y) dy \\ &= \int \int_{\mathbb{R}^2} x \frac{f_{XY}(x, y)}{f_Y(y)} \mathbb{1}_{\{f_Y(y) > 0\}} f_Y(y) dx dy \\ &= \int \int_{\mathbb{R}^2} x f_{XY}(x, y) dx dy \\ &= \int_{\mathbb{R}} x \left(\int_{\mathbb{R}} f_{XY}(x, y) dy \right) dx \\ &= \int_{\mathbb{R}} x f_X(x) dx \\ &= \mathbb{E}(X) \end{aligned}$$

□