

14.310x – Data Analysis for Social Scientists

MicroMasters en Statistics and Data Science (SDS) – MITx

BREIT (Brescia Institute of Technology) · Aporta — Grupo Breca

Clase 3

Variables Aleatorias y Distribuciones Conjuntas

Soluciones

Módulos oficiales del MicroMaster:

M2 – Joint Distributions; M3 – Joint, Marginal & Conditional Distributions; Summarizing & Describing Data

B.Sc. Cristian Lazo Quispe

✉ clazoq@uni.pe

[in linkedin.com/in/cristian-lazo-quispe](https://www.linkedin.com/in/cristian-lazo-quispe) · github.com/CristianLazoQuispe

Material del *Advanced Program in Data Science & Global Skills*, diseñado y desarrollado por **BREIT (Brescia Institute of Technology)**, iniciativa de **Aporta** —plataforma de innovación e impacto social del **Grupo Breca**— en alianza con el **MIT IDSS** (Institute for Data, Systems, and Society).

Basado en MIT 14.310x (E. Duflo & S. Ellison) y en diversas fuentes abiertas (ver bibliografía al final), bajo licencia CC BY-NC-SA 4.0.

11 de agosto de 2026

Niveles de dificultad (*difficulty levels*)

Cada problema está rotulado por color según su nivel, de lo más accesible (Nivel 1) a lo más difícil (Nivel 5). Empieza por donde te sientas cómodo y avanza a tu ritmo; los niveles 4–5 son extensiones para profundizar, no un requisito.

- **Nivel 1 – Fundamentos:** calentamiento muy guiado; para quien parte de casi cero.
- **Nivel 2 – Núcleo:** el nivel objetivo del curso/examen.
- **Nivel 3 – Aplicación:** combina varias ideas en contexto realista.
- **Nivel 4 – Reto:** difícil, integra varios temas.
- **Nivel 5 – Reto+:** muy difícil / fuera del scope típico.

1. Problemas y soluciones

Nivel 1 – Fundamentos (warm-up)

Para quien parte de casi cero: cálculos directos y guiados.

Ejercicio 1.1 (¿Discreta o continua?). Para cada variable aleatoria, indica si es *discreta* o *continua*.

- (a) X = número de hijos de un hogar.
- (b) Y = ingreso mensual (en soles) de una persona.
- (c) Z = años de educación completados.
- (d) T = tiempo (en meses) hasta conseguir empleo.
- (e) N = número de intentos hasta aprobar un examen.

Solución

Una v.a. es **discreta** si toma valores aislados (se pueden contar) y **continua** si puede tomar cualquier valor en un intervalo.

- (a) Discreta (0, 1, 2, ...). (b) Continua (cualquier monto ≥ 0 , en principio con decimales). (c) Discreta (conteo de años). (d) Continua (un tiempo, medible con cualquier precisión). (e) Discreta (1, 2, 3, ...).

Ejercicio 1.2 (Reconocer el tipo y su función de probabilidad). Une cada v.a. con el objeto que la describe (*pmf* para discretas, *pdf* para continuas) y di qué valores puede tomar.

- (a) X = resultado de lanzar un dado equilibrado.
- (b) Y = estatura (en metros) de un adulto elegido al azar.
- (c) Z = número de caras en 3 lanzamientos de una moneda.

Solución

- (a) X es *discreta*; se describe con una **pmf**. Valores $\{1, 2, 3, 4, 5, 6\}$, cada uno con probabilidad $\frac{1}{6}$.
- (b) Y es *continua*; se describe con una **pdf** (densidad). Valores en un intervalo, p.ej. $[1.4, 2.1]$ m (cualquier valor real del rango).
- (c) Z es *discreta*; se describe con una **pmf**. Valores $\{0, 1, 2, 3\}$. *Idea clave*: una pmf asigna probabilidad a *puntos*; una pdf asigna probabilidad a *intervalos* (como área bajo la curva).

Ejercicio 1.3 (¿Es una pmf válida?). Una v.a. discreta X (número de adultos por hogar que tienen seguro) tiene supuesta pmf

$$p(0) = 0.2, \quad p(1) = 0.5, \quad p(2) = 0.3.$$

¿Es una pmf válida? Justifica con las dos condiciones que debe cumplir toda pmf.

Solución

Una pmf debe cumplir: (i) $p(x) \geq 0$ para todo x , y (ii) $\sum_x p(x) = 1$.
 (i) Los tres valores son ≥ 0 . (ii) $0.2 + 0.5 + 0.3 = 1$.
 Se cumplen ambas, así que **sí** es una pmf válida.

Ejercicio 1.4 (Leer una pmf en tabla). El número de personas mayores por hogar, X , tiene la siguiente pmf:

x	0	1	2	3
$P(X = x)$	0.4	0.3	0.2	0.1

- (a) ¿Cuánto vale $\mathbb{P}(X = 2)$?
- (b) ¿Cuánto vale $\mathbb{P}(X \geq 2)$?
- (c) ¿Cuánto vale $\mathbb{P}(X \leq 1)$?

Solución

- (a) $\mathbb{P}(X = 2) = 0.2$ (lectura directa).
- (b) $\mathbb{P}(X \geq 2) = \mathbb{P}(X = 2) + \mathbb{P}(X = 3) = 0.2 + 0.1 = 0.3$.
- (c) $\mathbb{P}(X \leq 1) = 0.4 + 0.3 = 0.7$ (o bien $1 - 0.3 = 0.7$ por complemento).

Ejercicio 1.5 (Completar una pmf con un valor desconocido). El número de quejas que recibe una oficina al día, X , tiene pmf

$$p(0) = 0.1, \quad p(1) = 0.3, \quad p(2) = k, \quad p(3) = 0.2.$$

- (a) Halla el valor de k .
- (b) Calcula $\mathbb{P}(X \geq 2)$.

Solución

- (a) Como $\sum_x p(x) = 1$: $0.1 + 0.3 + k + 0.2 = 1 \Rightarrow k = 0.4$ (es ≥ 0 , válido).
 (b) $\mathbb{P}(X \geq 2) = p(2) + p(3) = 0.4 + 0.2 = 0.6$ (o bien $1 - (0.1 + 0.3) = 0.6$ por complemento).

Ejercicio 1.6 (Hallar c en una densidad uniforme). El tiempo de espera (en minutos) en una ventanilla se modela con densidad constante $f(x) = c$ para $x \in [2, 6]$ (y 0 fuera). Halla c .

Solución

Toda densidad debe integrar 1:

$$\int_2^6 c \, dx = c(6 - 2) = 4c = 1 \implies c = \frac{1}{4}.$$

Es una uniforme en $[2, 6]$: $f(x) = \frac{1}{4}$ en ese intervalo.

Ejercicio 1.7 (Hallar c en una densidad lineal $f = cx$). Sea $f(x) = cx$ para $x \in [0, 2]$ (y 0 fuera). Halla c para que f sea una densidad y luego calcula $\mathbb{P}(X \leq 1)$.

Solución

$$\int_0^2 cx \, dx = c \left. \frac{x^2}{2} \right|_0^2 = c \cdot 2 = 2c = 1 \implies c = \frac{1}{2}.$$

Con $f(x) = \frac{1}{2}x$: $\mathbb{P}(X \leq 1) = \int_0^1 \frac{1}{2}x \, dx = \left. \frac{1}{4}x^2 \right|_0^1 = \frac{1}{4}.$

Ejercicio 1.8 (Densidad $f = 2x$ y una probabilidad). Sea $f(x) = 2x$ para $x \in [0, 1]$ (y 0 fuera).

- (a) Verifica que f es una densidad válida.
 (b) Calcula $\mathbb{P}(X \leq 0.5)$.

Solución

(a) $f(x) = 2x \geq 0$ en $[0, 1]$ y $\int_0^1 2x \, dx = x^2 \Big|_0^1 = 1$. Válida.

(b) $\mathbb{P}(X \leq 0.5) = \int_0^{0.5} 2x \, dx = x^2 \Big|_0^{0.5} = 0.25.$

Ejercicio 1.9 (Probabilidad como área). El tiempo de atención (en minutos) sigue una uniforme en $[0, 10]$, con densidad $f(x) = \frac{1}{10}$ en ese intervalo.

- (a) ¿Cuál es la probabilidad de que la atención dure entre 3 y 7 minutos?
 (b) ¿Cuánto vale $\mathbb{P}(X = 5)$ *exactamente*? Explica.

Solución

- (a) Es el área del rectángulo de base $7 - 3 = 4$ y altura $\frac{1}{10}$: $\mathbb{P}(3 \leq X \leq 7) = 4 \cdot \frac{1}{10} = 0.4$.
 (b) En una v.a. *continua* la probabilidad de un punto aislado es 0: el área bajo un solo punto (ancho cero) es nula. $\mathbb{P}(X = 5) = 0$.

Ejercicio 1.10 (Interpretar una densidad mayor que 1). Un estudiante afirma: “la densidad $f(2.5) = 2$ es imposible porque una probabilidad no puede ser mayor que 1”. ¿Tiene razón?

Solución

No tiene razón. Una *densidad* $f(x)$ no es una probabilidad; puede valer más que 1. Lo que nunca puede pasar de 1 es el *área* bajo la curva (que es la probabilidad). Por ejemplo, $f(x) = 2x$ en $[0, 1]$ alcanza $f(1) = 2$ y aún así integra 1. La densidad solo debe cumplir $f(x) \geq 0$ y $\int f = 1$.

Nivel 2 – Núcleo (core)

El nivel objetivo del curso/examen; todos deberían llegar aquí.

Ejercicio 1.11 (De la cdf escalonada a la pmf). Una v.a. discreta X (número de créditos reprobados) tiene cdf escalonada

$$F(x) = \begin{cases} 0 & x < 0, \\ 0.5 & 0 \leq x < 1, \\ 0.8 & 1 \leq x < 2, \\ 1 & x \geq 2. \end{cases}$$

Recupera la pmf de X .

Solución

La pmf es el *salto* de la cdf en cada punto: $p(x) = F(x) - F(x^-)$.

$$p(0) = 0.5 - 0 = 0.5, \quad p(1) = 0.8 - 0.5 = 0.3, \quad p(2) = 1 - 0.8 = 0.2.$$

Verificación: $0.5 + 0.3 + 0.2 = 1$.

Ejercicio 1.12 (Hallar la constante de una pmf con fórmula). El número de talleres a los que asiste un participante en un mes, $X \in \{1, 2, 3, 4\}$, tiene pmf de la forma $p(x) = cx$.

- (a) Halla la constante c para que sea una pmf válida.
 (b) Calcula $\mathbb{P}(X \leq 2)$.

Solución

(a) La suma de las probabilidades debe ser 1:

$$\sum_{x=1}^4 c x = c(1 + 2 + 3 + 4) = 10c = 1 \implies c = \frac{1}{10} = 0.1.$$

Así $p(1) = 0.1$, $p(2) = 0.2$, $p(3) = 0.3$, $p(4) = 0.4$ (todas ≥ 0).

(b) $\mathbb{P}(X \leq 2) = p(1) + p(2) = 0.1 + 0.2 = 0.3$.

Ejercicio 1.13 (De la cdf continua a la pdf). Una v.a. continua X tiene cdf $F(x) = x^2$ para $x \in [0, 1]$ (con $F(x) = 0$ si $x < 0$ y $F(x) = 1$ si $x > 1$).

(a) Halla la densidad $f(x)$.

(b) Calcula $\mathbb{P}(X \leq 0.5)$ de dos maneras (con F y con f).

Solución

(a) La densidad es la derivada de la cdf: $f(x) = F'(x) = 2x$ en $[0, 1]$ (y 0 fuera).

(b) Con la cdf: $\mathbb{P}(X \leq 0.5) = F(0.5) = 0.25$.

Con la densidad: $\int_0^{0.5} 2x \, dx = 0.25$. Coinciden.

Ejercicio 1.14 (De la densidad a la cdf). Sea $f(x) = \frac{1}{2}x$ para $x \in [0, 2]$ (y 0 fuera).

(a) Halla la cdf $F(x)$ para $x \in [0, 2]$.

(b) Evalúa $F(1)$.

Solución

(a) Para $x \in [0, 2]$: $F(x) = \int_0^x \frac{1}{2}t \, dt = \frac{1}{4}t^2 \Big|_0^x = \frac{x^2}{4}$. (Y $F(x) = 0$ si $x < 0$, $F(x) = 1$ si $x > 2$.)

(b) $F(1) = \frac{1}{4}$.

Ejercicio 1.15 (Probabilidad con densidad cuadrática). Sea $f(x) = \frac{3}{8}x^2$ para $x \in [0, 2]$ (y 0 fuera).

(a) Verifica que es una densidad válida.

(b) Calcula $\mathbb{P}(1 \leq X \leq 2)$.

Solución

(a) $f \geq 0$ y $\int_0^2 \frac{3}{8}x^2 dx = \frac{3}{8} \cdot \frac{x^3}{3} \Big|_0^2 = \frac{1}{8} \cdot 8 = 1$. Válida.

(b) $\mathbb{P}(1 \leq X \leq 2) = \int_1^2 \frac{3}{8}x^2 dx = \frac{1}{8}x^3 \Big|_1^2 = \frac{1}{8}(8 - 1) = \frac{7}{8}$.

Ejercicio 1.16 (Esperanza de una v.a. discreta). El número de personas por hogar atendidas en un programa, X , tiene pmf $p(1) = 0.5$, $p(2) = 0.3$, $p(3) = 0.2$.

(a) Calcula $\mathbb{E}[X]$.

(b) Calcula $\text{Var}(X)$.

Solución

(a) $\mathbb{E}[X] = 1(0.5) + 2(0.3) + 3(0.2) = 0.5 + 0.6 + 0.6 = 1.7$.

(b) $\mathbb{E}[X^2] = 1(0.5) + 4(0.3) + 9(0.2) = 0.5 + 1.2 + 1.8 = 3.5$; $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 3.5 - 1.7^2 = 3.5 - 2.89 = 0.61$.

Ejercicio 1.17 (Esperanza con densidad lineal). Sea X con densidad $f(x) = \frac{1}{2}x$ en $[0, 2]$.

(a) Calcula $\mathbb{E}[X]$.

(b) Calcula $\mathbb{P}(X > 1)$.

Solución

(a) $\mathbb{E}[X] = \int_0^2 x \cdot \frac{1}{2}x dx = \frac{1}{2} \int_0^2 x^2 dx = \frac{1}{2} \cdot \frac{8}{3} = \frac{4}{3}$.

(b) $\mathbb{P}(X > 1) = \int_1^2 \frac{1}{2}x dx = \frac{1}{4}x^2 \Big|_1^2 = \frac{1}{4}(4 - 1) = \frac{3}{4}$.

Ejercicio 1.18 (PMF conjunta en tabla). Dos v.a. discretas $X \in \{1, 2, 3\}$ (p.ej. nivel de estudios) y $Y \in \{2, 4, 5\}$ (p.ej. estrato) tienen la siguiente pmf conjunta:

$f(x, y)$	$Y = 2$	$Y = 4$	$Y = 5$	$f_X(x)$
$X = 1$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{24}$	
$X = 2$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{8}$	
$X = 3$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{12}$	
$f_Y(y)$				

(a) Calcula $\mathbb{P}(X \leq 2, Y \leq 4)$.

(b) Halla las marginales f_X y f_Y (completa la tabla).

(c) Calcula $\mathbb{P}(Y = 2 | X = 1)$.

(d) ¿Son X e Y independientes?

Solución

(a) Sumamos las celdas con $X \in \{1, 2\}$ y $Y \in \{2, 4\}$:

$$\mathbb{P}(X \leq 2, Y \leq 4) = \frac{1}{12} + \frac{1}{24} + \frac{1}{6} + \frac{1}{12} = \frac{2+1+4+2}{24} = \frac{9}{24} = \frac{3}{8}.$$

(b) Marginal de X = suma por filas; marginal de Y = suma por columnas:

$$f_X(1) = \frac{1}{6}, \quad f_X(2) = \frac{3}{8}, \quad f_X(3) = \frac{11}{24}; \quad f_Y(2) = \frac{1}{2}, \quad f_Y(4) = \frac{1}{4}, \quad f_Y(5) = \frac{1}{4}.$$

(Comprobación: $\frac{1}{6} + \frac{3}{8} + \frac{11}{24} = \frac{4+9+11}{24} = 1$.)

$$(c) \mathbb{P}(Y = 2 | X = 1) = \frac{f(1, 2)}{f_X(1)} = \frac{1/12}{1/6} = \frac{1}{2}.$$

(d) **No** son independientes: por ejemplo $f(2, 2) = \frac{1}{6}$ pero $f_X(2) f_Y(2) = \frac{3}{8} \cdot \frac{1}{2} = \frac{3}{16} \neq \frac{1}{6}$.

Ejercicio 1.19 (Densidad conjunta $f = cxy$ y marginal). Sea $f(x, y) = cxy$ para $(x, y) \in [0, 1] \times [0, 1]$ (y 0 fuera).

(a) Halla c .

(b) Halla la marginal $f_X(x)$.

(c) ¿Son X e Y independientes?

Solución

$$(a) \int_0^1 \int_0^1 cxy \, dx \, dy = c \left(\int_0^1 x \, dx \right) \left(\int_0^1 y \, dy \right) = c \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{c}{4} = 1 \Rightarrow c = 4.$$

$$(b) f_X(x) = \int_0^1 4xy \, dy = 4x \cdot \frac{1}{2} = 2x \text{ en } [0, 1].$$

(c) Por simetría $f_Y(y) = 2y$. Como $f(x, y) = 4xy = (2x)(2y) = f_X(x)f_Y(y)$, **sí** son independientes.

Ejercicio 1.20 (Densidad conjunta $f = c(x + y)$ y condicional). Sea $f(x, y) = c(x + y)$ para $(x, y) \in [0, 1] \times [0, 1]$ (y 0 fuera).

(a) Halla c .

(b) Halla la marginal $f_X(x)$.

(c) Halla la condicional $f_{Y|X}(y | x)$.

Solución

$$(a) \int_0^1 \int_0^1 c(x + y) \, dx \, dy = c \int_0^1 \left(\frac{1}{2} + y \right) dy = c \left(\frac{1}{2} + \frac{1}{2} \right) = c = 1. \text{ Así } c = 1.$$

$$(b) f_X(x) = \int_0^1 (x+y) dy = x + \frac{1}{2} \text{ en } [0, 1].$$

$$(c) f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{x+y}{x+\frac{1}{2}}, \text{ para } y \in [0, 1]. \text{ (Depende de } x \Rightarrow \text{ no independientes.)}$$

Nivel 3 – Aplicación

Combina dos o más ideas en contexto realista; consolida lo aprendido.

Ejercicio 1.21 (Verificar una densidad conjunta). Sea $f(x, y) = 6xy^2$ para $(x, y) \in [0, 1] \times [0, 1]$ (y 0 fuera).

(a) Verifica que integra a 1.

(b) Halla la marginal $f_X(x)$.

Solución

$$(a) \int_0^1 \int_0^1 6xy^2 dx dy = 6 \left(\int_0^1 x dx \right) \left(\int_0^1 y^2 dy \right) = 6 \cdot \frac{1}{2} \cdot \frac{1}{3} = 1. \text{ Válida.}$$

$$(b) f_X(x) = \int_0^1 6xy^2 dy = 6x \cdot \frac{1}{3} = 2x \text{ en } [0, 1].$$

(Además $f_Y(y) = 3y^2$ y $f = f_X f_Y$: son independientes.)

Ejercicio 1.22 (Marginal sobre el cuadrado: $f = \frac{3}{2}(x^2 + y^2)$). Sea $f(x, y) = \frac{3}{2}(x^2 + y^2)$ para $(x, y) \in [0, 1] \times [0, 1]$ (y 0 fuera).

(a) Verifica que integra a 1.

(b) Halla la marginal $f_X(x)$.

(c) ¿Son independientes X e Y ?

Solución

$$(a) \int_0^1 \int_0^1 \frac{3}{2}(x^2 + y^2) dx dy = \frac{3}{2} \int_0^1 \left(\frac{1}{3} + y^2 \right) dy = \frac{3}{2} \left(\frac{1}{3} + \frac{1}{3} \right) = \frac{3}{2} \cdot \frac{2}{3} = 1. \text{ Válida.}$$

$$(b) f_X(x) = \int_0^1 \frac{3}{2}(x^2 + y^2) dy = \frac{3}{2} \left(x^2 + \frac{1}{3} \right) = \frac{3}{2}x^2 + \frac{1}{2} \text{ en } [0, 1].$$

(c) Por simetría $f_Y(y) = \frac{3}{2}y^2 + \frac{1}{2}$. El producto $f_X(x)f_Y(y) \neq \frac{3}{2}(x^2 + y^2) = f(x, y)$ en general, así que **no** son independientes.

Ejercicio 1.23 (Probabilidad conjunta uniforme sobre un cuadrado). Un punto (X, Y) se elige uniformemente en el cuadrado $[0, 1] \times [0, 1]$, es decir $f(x, y) = 1$ allí.

(a) Calcula $\mathbb{P}(X + Y \leq 1)$.

(b) Calcula $\mathbb{P}(X > Y)$.

Solución

Como $f = 1$, las probabilidades son áreas dentro del cuadrado unitario.

(a) La región $X + Y \leq 1$ es el triángulo inferior de vértices $(0, 0)$, $(1, 0)$, $(0, 1)$, de área $\frac{1}{2}$. Así $\mathbb{P}(X + Y \leq 1) = \frac{1}{2}$.

(b) La región $X > Y$ es media diagonal del cuadrado, área $\frac{1}{2}$: $\mathbb{P}(X > Y) = \frac{1}{2}$.

Ejercicio 1.24 (Independencia por factorización y soporte). Decide en cada caso si X e Y son independientes (asume densidad 0 fuera del soporte indicado).

(a) $f(x, y) = e^{-x-y}$ para $x \geq 0$, $y \geq 0$.

(b) $f(x, y) = 2$ para $0 \leq y \leq x \leq 1$.

Solución

(a) Se factoriza como $f = (e^{-x})(e^{-y})$ y el soporte es un rectángulo $([0, \infty) \times [0, \infty))$. Ambas condiciones se cumplen: **sí** son independientes ($X, Y \sim \text{Exp}(1)$ i.i.d.).

(b) Aunque $f = 2$ “parece” factorizable como constante, el soporte $\{0 \leq y \leq x \leq 1\}$ **no** es un rectángulo: el rango de Y depende de X . Por lo tanto **no** son independientes. (En efecto $f_X(x) = 2x$, $f_Y(y) = 2(1 - y)$, y $f_X f_Y \neq 2$.)

Ejercicio 1.25 (Triángulo $0 \leq y \leq x \leq 1$: marginales y condicional). Sea $f(x, y) = 3x$ para $0 \leq y \leq x \leq 1$ (y 0 fuera).

(a) Verifica que integra 1 y halla la marginal $f_X(x)$.

(b) Halla la marginal $f_Y(y)$.

(c) Halla la condicional $f_{Y|X}(y|x)$ e interprétala.

Solución

(a) $\int_0^1 \int_0^x 3x \, dy \, dx = \int_0^1 3x \cdot x \, dx = \int_0^1 3x^2 \, dx = 1$. ✓ Marginal: $f_X(x) = \int_0^x 3x \, dy = 3x^2$ en $[0, 1]$.

(b) Para $y \in [0, 1]$, x va de y a 1: $f_Y(y) = \int_y^1 3x \, dx = \frac{3}{2}(1 - y^2)$.

(c) $f_{Y|X}(y|x) = \frac{3x}{3x^2} = \frac{1}{x}$ para $0 \leq y \leq x$. Es *constante* en y : dado $X = x$, la variable Y es **uniforme** en $[0, x]$.

Ejercicio 1.26 (Esperanza con densidad conjunta). Sea $f(x, y) = x + y$ para $(x, y) \in [0, 1] \times [0, 1]$ (y 0 fuera). Comprueba primero que integra 1 y luego calcula $\mathbb{E}[XY]$ y $\mathbb{E}[X]$.

Solución

Integra 1: $\int_0^1 \int_0^1 (x+y) dx dy = \int_0^1 \left(\frac{1}{2} + y\right) dy = \frac{1}{2} + \frac{1}{2} = 1. \checkmark$

$$\mathbb{E}[XY] = \int_0^1 \int_0^1 xy(x+y) dx dy = \int_0^1 \int_0^1 (x^2y + xy^2) dx dy = \int_0^1 \left(\frac{y}{3} + \frac{y^2}{2}\right) dy = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

La marginal es $f_X(x) = \int_0^1 (x+y) dy = x + \frac{1}{2}$, así

$$\mathbb{E}[X] = \int_0^1 x\left(x + \frac{1}{2}\right) dx = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}.$$

(Como $\mathbb{E}[XY] = \frac{1}{3} \neq \mathbb{E}[X]\mathbb{E}[Y] = \left(\frac{7}{12}\right)^2$, X e Y no son independientes.)

Ejercicio 1.27 (Función de una v.a. por el método de la cdf: $Y = X^2$). Sea X con densidad $f(x) = 2x$ en $[0, 1]$ (y 0 fuera), e.g. una nota normalizada. Define $Y = X^2$.

- (a) Halla la cdf $F_Y(y)$ para $y \in [0, 1]$ usando el método de la cdf.
- (b) Deriva la densidad $f_Y(y)$ e identifica la distribución de Y .

Solución

(a) Para $y \in [0, 1]$, como $X \geq 0$ se tiene $X^2 \leq y \iff X \leq \sqrt{y}$:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X \leq \sqrt{y}) = \int_0^{\sqrt{y}} 2x dx = x^2 \Big|_0^{\sqrt{y}} = y.$$

(b) $f_Y(y) = F'_Y(y) = 1$ para $y \in [0, 1]$: $Y \sim \text{Unif}[0, 1]$. (Es un caso de la *transformada integral de probabilidad*: aquí $Y = X^2 = F_X(X)$, que siempre es uniforme.) *(adaptado de mikem5/5080-5090-Solutions)*

Ejercicio 1.28 (Función de una v.a.: $Y = \sqrt{X}$). Sea $X \sim \text{Unif}[0, 1]$ y define $Y = \sqrt{X}$.

- (a) Halla la cdf $F_Y(y)$ para $y \in [0, 1]$.
- (b) Deriva la densidad $f_Y(y)$.

Solución

Para $y \in [0, 1]$, usando que $F_X(x) = x$ en $[0, 1]$ y que $\sqrt{X} \leq y \iff X \leq y^2$:

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

Derivando: $f_Y(y) = F'_Y(y) = 2y$ para $y \in [0, 1]$ (y 0 fuera).

Ejercicio 1.29 (Transformación $Y = -\ln X$ (inversa de la cdf)). Sea $X \sim \text{Unif}[0, 1]$ y define $Y = -\ln X$.

- (a) Halla la cdf $F_Y(y)$ para $y \geq 0$.
- (b) Deriva la densidad $f_Y(y)$ e identifica la distribución de Y .

Solución

(a) Para $y \geq 0$, como $-\ln X \leq y \iff \ln X \geq -y \iff X \geq e^{-y}$:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X \geq e^{-y}) = 1 - F_X(e^{-y}) = 1 - e^{-y}.$$

(b) $f_Y(y) = F'_Y(y) = e^{-y}$ para $y \geq 0$: $Y \sim \text{Exponencial}(1)$. (Este es el método de la *transformada inversa* para simular una exponencial a partir de una uniforme.)

Ejercicio 1.30 (Densidad conjunta exponencial: dos colas independientes). El tiempo (en horas) de atención en dos ventanillas distintas se modela con la densidad conjunta $f(x, y) = 4e^{-2(x+y)}$ para $x \geq 0, y \geq 0$ (y 0 fuera).

- (a) Muestra que X e Y son independientes e identifica sus distribuciones.
- (b) Calcula $\mathbb{P}(X > 1, Y > 1)$.
- (c) Calcula $\mathbb{P}(X < Y)$.

Solución

(a) Factorizamos: $f(x, y) = (2e^{-2x})(2e^{-2y}) = f_X(x)f_Y(y)$, con el soporte rectangular $[0, \infty) \times [0, \infty)$. Luego **sí** son independientes, con $X, Y \sim \text{Exponencial}(2)$ (cada marginal integra 1, lo que confirma que f es densidad).

(b) Por independencia y como $\mathbb{P}(X > 1) = \int_1^\infty 2e^{-2x} dx = e^{-2}$:

$$\mathbb{P}(X > 1, Y > 1) = \mathbb{P}(X > 1)\mathbb{P}(Y > 1) = e^{-2} \cdot e^{-2} = e^{-4} \approx 0.0183.$$

(c) Por simetría (misma distribución e independientes) $\mathbb{P}(X < Y) = \mathbb{P}(Y < X)$ y $\mathbb{P}(X = Y) = 0$, así que $\mathbb{P}(X < Y) = \frac{1}{2}$. (*adaptado de mikem5/5080-5090-Solutions*)

Nivel 4 – Reto

Difícil: integra varios temas en un mismo problema.

Ejercicio 1.31 (MEGA-RETO — densidad sobre un triángulo: de c a $\text{Corr}(X, Y)$). Estudiamos a la vez dos indicadores estandarizados de un mismo hogar: X = nivel de *consumo* y Y = nivel de *ahorro* (ambos reescalados al intervalo $[0, 1]$). Como un hogar nunca ahorra más de lo que consume en este índice, el soporte es el triángulo $0 \leq y \leq x \leq 1$. Modelamos la densidad conjunta

$$f(x, y) = cxy, \quad 0 \leq y \leq x \leq 1, \quad (\text{y } 0 \text{ fuera}).$$

- Halla la constante c .
- Dibuja** la región R_{XY} donde $f > 0$ y sombrealala.
- Halla las marginales $f_X(x)$ y $f_Y(y)$.
- Halla la condicional $f_{Y|X}(y|x)$.
- ¿Son X e Y independientes? Justifica.
- Calcula $\mathbb{E}[X]$ y $\mathbb{E}[Y]$.
- Calcula $\mathbb{E}[XY]$.
- Calcula $\text{Cov}(X, Y)$ y $\text{Corr}(X, Y)$.
- Calcula $\mathbb{P}\left(Y > \frac{X}{2}\right)$.
- Interpreta** en palabras el *signo* y la *magnitud* de la covarianza/correlación en este contexto.

(adaptado de MIT 18.05, Class 7: Joint Distributions, Covariance and Correlation)

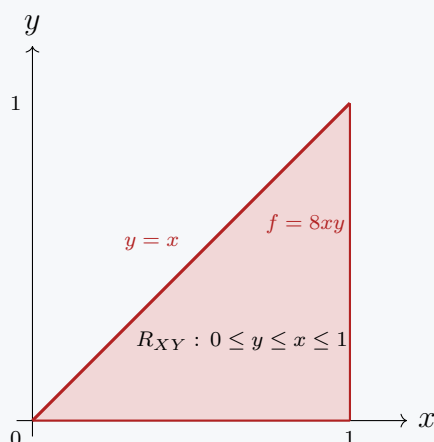
Solución

(a) Constante c . Integramos sobre el triángulo (y de 0 a x , luego x de 0 a 1):

$$\int_0^1 \int_0^x cxy \, dy \, dx = \int_0^1 cx \cdot \frac{x^2}{2} \, dx = \frac{c}{2} \int_0^1 x^3 \, dx = \frac{c}{2} \cdot \frac{1}{4} = \frac{c}{8} = 1 \implies \boxed{c = 8}.$$

Así $f(x, y) = 8xy$ en el triángulo.

(b) Región R_{XY} .



El soporte es el triángulo bajo la diagonal $y = x$ (la zona donde el ahorro Y no supera al consumo X).

(c) Marginales. Para $x \in [0, 1]$, y recorre $[0, x]$:

$$f_X(x) = \int_0^x 8xy \, dy = 8x \cdot \frac{x^2}{2} = 4x^3, \quad 0 \leq x \leq 1.$$

Para $y \in [0, 1]$, fijado y se necesita $x \geq y$ (y $x \leq 1$):

$$f_Y(y) = \int_y^1 8xy \, dx = 8y \cdot \frac{1 - y^2}{2} = 4y(1 - y^2), \quad 0 \leq y \leq 1.$$

(Comprobación: $\int_0^1 4x^3 \, dx = 1$ y $\int_0^1 4y(1 - y^2) \, dy = 1$.)

(d) Condicional. Para x fijo, $y \in [0, x]$:

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{8xy}{4x^3} = \frac{2y}{x^2}, \quad 0 \leq y \leq x.$$

(Es una densidad triangular en $[0, x]$; integra 1: $\int_0^x \frac{2y}{x^2} \, dy = 1$.)

(e) ¿Independientes? No. Dos razones, cualquiera basta: (i) El soporte $\{0 \leq y \leq x \leq 1\}$ *no es un rectángulo*: el rango de Y depende de X . (ii) $f_X(x)f_Y(y) = 4x^3 \cdot 4y(1 - y^2) \neq 8xy = f(x, y)$.

(f) Esperanzas.

$$\begin{aligned} \mathbb{E}[X] &= \int_0^1 x \cdot 4x^3 \, dx = \int_0^1 4x^4 \, dx = \frac{4}{5}, \\ \mathbb{E}[Y] &= \int_0^1 y \cdot 4y(1 - y^2) \, dy = \int_0^1 (4y^2 - 4y^4) \, dy = \frac{4}{3} - \frac{4}{5} = \frac{8}{15}. \end{aligned}$$

(g) $\mathbb{E}[XY]$.

$$\mathbb{E}[XY] = \int_0^1 \int_0^x xy \cdot 8xy \, dy \, dx = \int_0^1 8x^2 \cdot \frac{x^3}{3} \, dx = \frac{8}{3} \int_0^1 x^5 \, dx = \frac{8}{3} \cdot \frac{1}{6} = \frac{4}{9}.$$

(h) Covarianza y correlación.

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y] = \frac{4}{9} - \frac{4}{5} \cdot \frac{8}{15} = \frac{4}{9} - \frac{32}{75} = \frac{100 - 96}{225} = \frac{4}{225} \approx 0.0178.$$

Para la correlación necesitamos las varianzas marginales:

$$\mathbb{E}[X^2] = \int_0^1 x^2 \cdot 4x^3 dx = \frac{4}{6} = \frac{2}{3}, \quad \text{Var}(X) = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{3} - \frac{16}{25} = \frac{2}{75},$$

$$\begin{aligned} \mathbb{E}[Y^2] &= \int_0^1 y^2 \cdot 4y(1 - y^2) dy = \int_0^1 (4y^3 - 4y^5) dy = 1 - \frac{2}{3} = \frac{1}{3}, \\ \text{Var}(Y) &= \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{1}{3} - \frac{64}{225} = \frac{11}{225}. \end{aligned}$$

Entonces

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} = \frac{4/225}{\sqrt{\frac{2}{75} \cdot \frac{11}{225}}} = \frac{2\sqrt{66}}{33} \approx 0.492.$$

(i) $\mathbb{P}(Y > X/2)$. Sobre el soporte, y va de $x/2$ a x :

$$\mathbb{P}\left(Y > \frac{X}{2}\right) = \int_0^1 \int_{x/2}^x 8xy dy dx = \int_0^1 8x \cdot \frac{x^2 - x^2/4}{2} dx = \int_0^1 4x \cdot \frac{3}{4} x^2 dx = 3 \int_0^1 x^3 dx = \frac{3}{4}.$$

(j) Interpretación. La covarianza es *positiva* ($+\frac{4}{225}$) y la correlación ≈ 0.49 es *moderada-alta*. En palabras: los hogares con mayor consumo X tienden también a tener mayor ahorro Y (varían en el *mismo* sentido). Tiene sentido: la propia restricción $0 \leq y \leq x$ obliga a que un ahorro alto solo sea posible cuando el consumo es alto, lo que “empuja” a las dos variables a moverse juntas. La magnitud ≈ 0.49 (lejos de 0, lejos de 1) dice que la asociación es real pero no determinista: conocer X ayuda a predecir Y , sin fijarlo por completo. **Cuidado:** Corr mide solo asociación *lineal* y *no* implica causalidad.

Ejercicio 1.32 (Densidad $f = c(x + y)$ en el cuadrado: $\mathbb{E}[XY]$, Cov y condicional). Dos índices estandarizados $X, Y \in [0, 1]$ (p.ej. horas de estudio y horas de sueño, ambas normalizadas) tienen densidad conjunta

$$f(x, y) = c(x + y), \quad (x, y) \in [0, 1] \times [0, 1] \quad (\text{y } 0 \text{ fuera}).$$

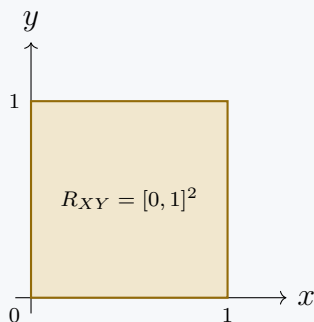
- Halla c y dibuja el soporte R_{XY} .
- Halla la marginal $f_X(x)$ y la condicional $f_{Y|X}(y|x)$.
- Calcula $\mathbb{E}[X]$, $\mathbb{E}[XY]$, $\text{Cov}(X, Y)$ y $\text{Corr}(X, Y)$.
- Interpreta el signo de la covarianza.

(adaptado de MIT 18.05, Class 7)

Solución**(a) Constante y soporte.**

$$\int_0^1 \int_0^1 c(x+y) dx dy = c \int_0^1 \left(\frac{1}{2} + y\right) dy = c \left(\frac{1}{2} + \frac{1}{2}\right) = c = 1,$$

luego $c = 1$ y $f(x, y) = x + y$. El soporte es el cuadrado unitario:

**(b) Marginal y condicional.**

$$f_X(x) = \int_0^1 (x+y) dy = x + \frac{1}{2}, \quad x \in [0, 1].$$

$$f_{Y|X}(y|x) = \frac{x+y}{x+\frac{1}{2}} = \frac{2(x+y)}{2x+1}, \quad y \in [0, 1].$$

(Depende de $x \Rightarrow$ no independientes, aunque el soporte sí sea rectangular.)

(c) Momentos.

$$\mathbb{E}[X] = \int_0^1 x \left(x + \frac{1}{2}\right) dx = \frac{1}{3} + \frac{1}{4} = \frac{7}{12} \quad (= \mathbb{E}[Y] \text{ por simetría}).$$

$$\mathbb{E}[XY] = \int_0^1 \int_0^1 xy(x+y) dx dy = \int_0^1 \left(\frac{y}{3} + \frac{y^2}{2}\right) dy = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

$$\text{Cov}(X, Y) = \frac{1}{3} - \left(\frac{7}{12}\right)^2 = \frac{1}{3} - \frac{49}{144} = -\frac{1}{144} \approx -0.0069.$$

Para Corr: $\mathbb{E}[X^2] = \int_0^1 x^2 \left(x + \frac{1}{2}\right) dx = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}$, así $\text{Var}(X) = \frac{5}{12} - \frac{49}{144} = \frac{11}{144}$ ($= \text{Var}(Y)$).
Entonces

$$\text{Corr}(X, Y) = \frac{-1/144}{11/144} = -\frac{1}{11} \approx -0.091.$$

(d) Interpretación. La covarianza es *negativa* pero *muy pequeña* ($\text{Corr} \approx -0.09$): hay una leve tendencia a que cuando una variable es alta la otra sea algo más baja. El efecto es débil porque la densidad $f = x + y$ se aparta poco de la uniforme. El signo negativo proviene de que la masa se concentra cerca de las esquinas $(1, 0)$ y $(0, 1)$ algo más que en la diagonal.

Ejercicio 1.33 (Densidad conjunta polinomial: hallar K). Sea $f(x, y) = \frac{2}{81} x^2 y$ para $0 < x < K, 0 < y < K$ (y 0 fuera).

- Halla el valor de K .
- Con ese K , halla las marginales y decide si X e Y son independientes.
- Calcula $\mathbb{P}(X > 3Y)$.

Solución

(a) Imponemos que integre 1:

$$\int_0^K \int_0^K \frac{2}{81} x^2 y \, dx \, dy = \frac{2}{81} \left(\frac{K^3}{3} \right) \left(\frac{K^2}{2} \right) = \frac{K^5}{243} = 1 \implies K^5 = 243 \implies K = 3.$$

(b) Con $K = 3$:

$$f_X(x) = \int_0^3 \frac{2}{81} x^2 y \, dy = \frac{2}{81} x^2 \cdot \frac{9}{2} = \frac{x^2}{9}, \quad 0 < x < 3;$$

$$f_Y(y) = \int_0^3 \frac{2}{81} x^2 y \, dx = \frac{2}{81} y \cdot \frac{27}{3} = \frac{2y}{9}, \quad 0 < y < 3.$$

Como $f_X(x)f_Y(y) = \frac{x^2}{9} \cdot \frac{2y}{9} = \frac{2x^2 y}{81} = f(x, y)$, **sí** son independientes.

(c) La condición $x > 3y$ (con $0 < y, 0 < x < 3$) equivale a $0 < y < x/3$:

$$\mathbb{P}(X > 3Y) = \int_0^3 \int_0^{x/3} \frac{2}{81} x^2 y \, dy \, dx = \int_0^3 \frac{2}{81} x^2 \cdot \frac{1}{2} \left(\frac{x}{3} \right)^2 dx = \int_0^3 \frac{x^4}{729} dx = \frac{1}{729} \cdot \frac{3^5}{5} = \frac{243}{3645} = \frac{1}{15}.$$

Es decir $\mathbb{P}(X > 3Y) = \frac{1}{15} \approx 0.0667$.

Ejercicio 1.34 (Densidad conjunta exponencial). Sea $f(x, y) = 6e^{-(2x+3y)}$ para $x \geq 0, y \geq 0$ (y 0 fuera). Piensa en X como el tiempo de espera en un trámite e Y el tiempo en otro.

- ¿Son X e Y independientes? Identifica sus distribuciones.
- Calcula $\mathbb{E}[Y | X > 2]$.
- Calcula $\mathbb{P}(X > Y)$.

Solución

(1) Factorizamos:

$$f(x, y) = (2e^{-2x})(3e^{-3y}) = f_X(x)f_Y(y),$$

con $f_X(x) = 2e^{-2x}$ ($X \sim \text{Exponencial}(2)$) y $f_Y(y) = 3e^{-3y}$ ($Y \sim \text{Exponencial}(3)$). Como

f se factoriza, **sí** son independientes.

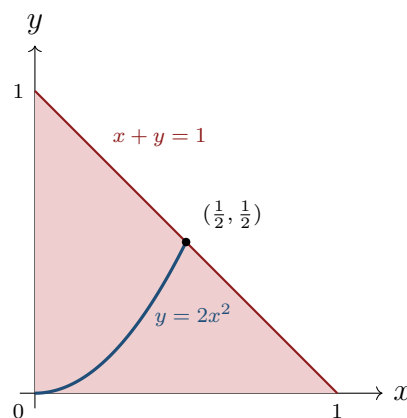
(2) Por independencia, condicionar en $\{X > 2\}$ no cambia la ley de Y :

$$\mathbb{E}[Y | X > 2] = \mathbb{E}[Y] = \frac{1}{3}.$$

(3)

$$\mathbb{P}(X > Y) = \int_0^\infty \int_y^\infty 6e^{-(2x+3y)} dx dy = \int_0^\infty 3e^{-3y} e^{-2y} dy = \int_0^\infty 3e^{-5y} dy = \frac{3}{5}.$$

Ejercicio 1.35 (Densidad sobre un triángulo: $\mathbb{P}(Y < 2X^2)$). Sea $f(x, y) = cx + 1$ para $x \geq 0$, $y \geq 0$, $x + y < 1$ (y 0 fuera). La región R_{XY} es el triángulo de vértices $(0, 0)$, $(1, 0)$, $(0, 1)$.



(1) Halla la constante c .

(2) Halla las marginales $f_X(x)$ y $f_Y(y)$.

(3) Calcula $\mathbb{P}(Y < 2X^2)$.

Solución

(1) Integramos sobre el triángulo (y de 0 a $1-x$, luego x de 0 a 1):

$$\int_0^1 \int_0^{1-x} (cx+1) dy dx = \int_0^1 (cx+1)(1-x) dx = \int_0^1 (cx - cx^2 + 1 - x) dx = \frac{c}{2} - \frac{c}{3} + 1 - \frac{1}{2} = \frac{c}{6} + \frac{1}{2}.$$

Igualando a 1: $\frac{c}{6} + \frac{1}{2} = 1 \Rightarrow c = 3$. Así $f(x, y) = 3x + 1$.

(2) Marginal de X (para $x \in [0, 1]$):

$$f_X(x) = \int_0^{1-x} (3x+1) dy = (3x+1)(1-x).$$

Marginal de Y (para $y \in [0, 1]$, con x de 0 a $1-y$):

$$f_Y(y) = \int_0^{1-y} (3x+1) dx = \frac{3}{2}(1-y)^2 + (1-y) = \frac{1}{2}(1-y)(5-3y).$$

(3) La parábola $y = 2x^2$ corta a la recta $x + y = 1$ donde $2x^2 = 1 - x$, i.e. $2x^2 + x - 1 = 0$, $x = \frac{1}{2}$. Para $x < \frac{1}{2}$ el techo es $2x^2$; para $x > \frac{1}{2}$ el techo es $1 - x$.

$$\mathbb{P}(Y < 2X^2) = \int_0^{1/2} \int_0^{2x^2} (3x + 1) dy dx + \int_{1/2}^1 \int_0^{1-x} (3x + 1) dy dx.$$

Primer trozo:

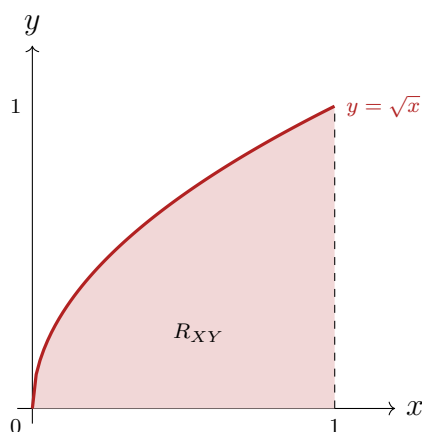
$$\int_0^{1/2} 2x^2(3x + 1) dx = \int_0^{1/2} (6x^3 + 2x^2) dx = \left. \frac{3}{2}x^4 + \frac{2}{3}x^3 \right|_0^{1/2} = \frac{3}{32} + \frac{1}{12} = \frac{17}{96}.$$

Segundo trozo:

$$\int_{1/2}^1 (3x + 1)(1 - x) dx = \int_{1/2}^1 (-3x^2 + 2x + 1) dx = \left. \left(-x^3 + x^2 + x \right) \right|_{1/2}^1 = 1 - \frac{5}{8} = \frac{3}{8} = \frac{36}{96}.$$

Sumando: $\mathbb{P}(Y < 2X^2) = \frac{17}{96} + \frac{36}{96} = \frac{53}{96} \approx 0.552$.

Ejercicio 1.36 (Densidad sobre la región $y \leq \sqrt{x}$). Sea $f(x, y) = 6xy$ para $0 \leq x \leq 1$, $0 \leq y \leq \sqrt{x}$ (y 0 fuera).



- (1) Halla la marginal $f_X(x)$ y la marginal $f_Y(y)$.
- (2) ¿Son independientes X e Y ?
- (3) Halla la densidad condicional $f_{X|Y}(x|y)$.
- (4) Calcula $\mathbb{E}[X|Y = y]$.
- (5) Calcula $\text{Var}(X|Y = y)$.

Solución

Primero, comprobemos que integra 1: $\int_0^1 \int_0^{\sqrt{x}} 6xy \, dy \, dx = \int_0^1 3x \cdot x \, dx = \int_0^1 3x^2 \, dx = 1$. ✓

(1) Marginal de X (para $x \in [0, 1]$):

$$f_X(x) = \int_0^{\sqrt{x}} 6xy \, dy = 6x \cdot \frac{(\sqrt{x})^2}{2} = 3x^2.$$

Marginal de Y : dado $y \in [0, 1]$, la condición $y \leq \sqrt{x}$ es $x \geq y^2$, con x de y^2 a 1:

$$f_Y(y) = \int_{y^2}^1 6xy \, dx = 6y \cdot \frac{1-y^4}{2} = 3y(1-y^4), \quad y \in [0, 1].$$

(2) $f_X(x)f_Y(y) = 3x^2 \cdot 3y(1-y^4) \neq 6xy = f(x, y)$ (además el soporte $y \leq \sqrt{x}$ no es un rectángulo). **No** son independientes.

(3) Para y fijo, x vive en $[y^2, 1]$:

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{6xy}{3y(1-y^4)} = \frac{2x}{1-y^4}, \quad y^2 \leq x \leq 1.$$

(4)

$$\mathbb{E}[X | Y = y] = \int_{y^2}^1 x \cdot \frac{2x}{1-y^4} \, dx = \frac{2}{1-y^4} \cdot \frac{1-y^6}{3} = \frac{2(1-y^6)}{3(1-y^4)}.$$

(5) Primero $\mathbb{E}[X^2 | Y = y] = \int_{y^2}^1 x^2 \frac{2x}{1-y^4} \, dx = \frac{2}{1-y^4} \cdot \frac{1-y^8}{4} = \frac{1-y^8}{2(1-y^4)}$. Entonces

$$\text{Var}(X | Y = y) = \frac{1-y^8}{2(1-y^4)} - \left(\frac{2(1-y^6)}{3(1-y^4)} \right)^2.$$

Ejercicio 1.37 (Densidad clásica $f = cx^2y$ sobre $x^2 \leq y \leq 1$). Sea $f(x, y) = cx^2y$ para $x^2 \leq y \leq 1$ (es decir, $-1 \leq x \leq 1$ y $x^2 \leq y \leq 1$), y 0 fuera.

(1) Halla c .

(2) Calcula $\mathbb{P}(X > Y)$.

Solución

(1) Integraremos (y de x^2 a 1, x de -1 a 1):

$$\int_{-1}^1 \int_{x^2}^1 cx^2y \, dy \, dx = \int_{-1}^1 cx^2 \cdot \frac{1-x^4}{2} \, dx = \frac{c}{2} \int_{-1}^1 (x^2 - x^6) \, dx = \frac{c}{2} \left(\frac{2}{3} - \frac{2}{7} \right) = \frac{c}{2} \cdot \frac{8}{21} = \frac{4c}{21}.$$

Igualando a 1: $c = \frac{21}{4}$.

(2) La región $\{x > y\}$ dentro del soporte exige $x > y \geq x^2$, lo que obliga $x > x^2$, es decir

$0 < x < 1$; ahí y va de x^2 a x :

$$\begin{aligned}\mathbb{P}(X > Y) &= \int_0^1 \int_{x^2}^x \frac{21}{4} x^2 y \, dy \, dx = \int_0^1 \frac{21}{4} x^2 \cdot \frac{x^2 - x^4}{2} \, dx \\ &= \frac{21}{8} \int_0^1 (x^4 - x^6) \, dx = \frac{21}{8} \left(\frac{1}{5} - \frac{1}{7} \right) = \frac{21}{8} \cdot \frac{2}{35} = \frac{3}{20}.\end{aligned}$$

Así $\mathbb{P}(X > Y) = \frac{3}{20} = 0.15$.

Ejercicio 1.38 (Función de v.a.: suma $Z = X + Y$ (convolución) y máximo (order statistics)). Tres sensores miden *independientemente* un mismo fenómeno; cada lectura es $\text{Unif}[0, 1]$.

- (a) Sean $X, Y \sim \text{Unif}[0, 1]$ independientes y $Z = X + Y$. Halla la densidad $f_Z(z)$ por *convolución* y dibújala.
- (b) Sean $X_1, X_2, X_3 \sim \text{Unif}[0, 1]$ i.i.d. y $M = \max(X_1, X_2, X_3)$. Halla F_M , f_M y $\mathbb{E}[M]$.
- (c) Para el mínimo $L = \min(X_1, X_2, X_3)$, halla $\mathbb{E}[L]$ y comenta la simetría con M .

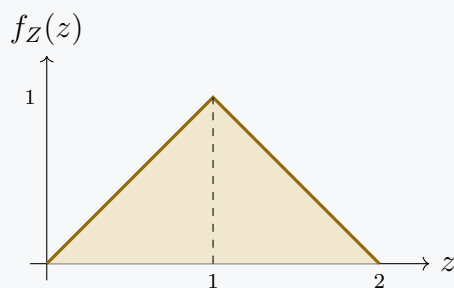
(adaptado de MIT 18.05, Class 7: functions of random variables)

Solución

(a) Convolución. Por independencia, $f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) \, dx$, con $f_X = f_Y = 1$ en $[0, 1]$. El integrando vale 1 cuando $0 \leq x \leq 1$ y $0 \leq z-x \leq 1$.

$$0 \leq z \leq 1: \quad f_Z(z) = \int_0^z 1 \, dx = z; \quad 1 \leq z \leq 2: \quad f_Z(z) = \int_{z-1}^1 1 \, dx = 2 - z.$$

Es la densidad “triangular” (carpa) sobre $[0, 2]$, con pico 1 en $z = 1$:



(Comprobación: área = $\frac{1}{2} \cdot 2 \cdot 1 = 1$. Además $\mathbb{E}[Z] = \mathbb{E}[X] + \mathbb{E}[Y] = 1$, el centro de la carpa.)

(b) Máximo. Como los X_i son i.i.d. con $F_X(t) = t$ en $[0, 1]$ y $\{M \leq m\} = \{\text{todos } X_i \leq m\}$:

$$F_M(m) = \mathbb{P}(M \leq m) = \left(F_X(m)\right)^3 = m^3, \quad 0 \leq m \leq 1.$$

$$f_M(m) = F'_M(m) = 3m^2, \quad \mathbb{E}[M] = \int_0^1 m \cdot 3m^2 \, dm = \frac{3}{4}.$$

(c) **Mínimo.** $\mathbb{P}(L > m) = (1 - m)^3$, luego $F_L(m) = 1 - (1 - m)^3$ y $f_L(m) = 3(1 - m)^2$:

$$\mathbb{E}[L] = \int_0^1 m \cdot 3(1 - m)^2 dm = \frac{1}{4}.$$

Por simetría de la uniforme respecto del centro $\frac{1}{2}$, $\mathbb{E}[L] = 1 - \mathbb{E}[M]$; en efecto $\frac{1}{4} = 1 - \frac{3}{4}$. El máximo se “empuja” hacia 1 y el mínimo hacia 0.

Ejercicio 1.39 (Mínimo de dos tiempos exponenciales). En una oficina, el tiempo (en horas) hasta que se desocupa la ventanilla A es $X \sim \text{Exponencial}(2)$ y el de la ventanilla B es $Y \sim \text{Exponencial}(3)$, con X e Y independientes. Te atienden en cuanto se libera *alguna*, así que tu espera es $T = \min(X, Y)$.

- (1) Halla $\mathbb{P}(T > t)$ para $t \geq 0$ e identifica la distribución de T .
- (2) Calcula la espera media $\mathbb{E}[T]$.
- (3) ¿Cuál es la probabilidad de que te atienda primero la ventanilla A , es decir $\mathbb{P}(X < Y)$?

Solución

Recuerda que para una $\text{Exponencial}(\lambda)$ se tiene $\mathbb{P}(W > w) = e^{-\lambda w}$.

(1) Como $T > t$ exige que *ambas* esperas superen t , y por independencia:

$$\mathbb{P}(T > t) = \mathbb{P}(X > t, Y > t) = \mathbb{P}(X > t) \mathbb{P}(Y > t) = e^{-2t} e^{-3t} = e^{-5t}.$$

Esa es la cola de una $\text{Exponencial}(5)$, luego $T \sim \text{Exponencial}(2 + 3) = \text{Exponencial}(5)$. (En general, el mínimo de exponenciales independientes es exponencial con tasa $\sum \lambda_i$.)

(2) $\mathbb{E}[T] = \frac{1}{5} = 0.2$ horas (es decir, 12 minutos).

(3) Condicionando en Y (o integrando la conjunta $f(x, y) = 6e^{-2x}e^{-3y}$ sobre $x < y$):

$$\mathbb{P}(X < Y) = \int_0^\infty 2e^{-2x} \mathbb{P}(Y > x) dx = \int_0^\infty 2e^{-2x} e^{-3x} dx = 2 \int_0^\infty e^{-5x} dx = \frac{2}{5}.$$

En general $\mathbb{P}(X < Y) = \frac{\lambda_X}{\lambda_X + \lambda_Y} = \frac{2}{5} = 0.4$. (adaptado de *hanmochen/Probability-Theory-2020-Autumn, HW4*)

Ejercicio 1.40 (El perro de Oscar (búsqueda bayesiana)). Oscar perdió a su perro en el bosque A (con probabilidad a priori 0.4) o en el bosque B (con probabilidad a priori 0.6). Si el perro sigue vivo y no es hallado al N -ésimo día de búsqueda, morirá esa noche con probabilidad $N/(N + 2)$. Si el perro está en A y Oscar busca un día en A , lo encuentra ese día con probabilidad 0.25; si está en B y busca en B , lo encuentra con probabilidad 0.15 (si busca en el bosque equivocado, no lo encuentra). El perro no cambia de bosque; Oscar busca de día y solo puede viajar de noche. Trata cada parte por separado.

- (a) Dibuja un diagrama de árbol.
- (b) ¿En qué bosque debe buscar Oscar para maximizar la probabilidad de encontrar a su perro el primer día?
- (c) Dado que Oscar buscó en A el primer día pero no lo encontró, ¿cuál es la probabilidad de que el perro esté en A ?

Notación. Sea $A = \{\text{perro en } A\}$ con $\mathbb{P}(A) = 0.4$ y $A^c = \{\text{perro en } B\}$ con $\mathbb{P}(A^c) = 0.6$; $F_1 = \{\text{lo encuentra el día 1}\}$; $S_1 = \{\text{Oscar busca en } A \text{ el día 1}\}$ y $S_1^c = \{\text{busca en } B\}$.

Solución

(a) Diagrama de árbol. Ver la página horizontal al final de esta solución. Las ramas llevan *probabilidades condicionales*; la probabilidad de un camino es el *producto* de las ramas que lo componen. El nivel de muerte nocturna $N/(N+2)$ solo entra en juego a partir de la *noche* del día 1, por lo que **no** afecta a las preguntas (b) ni (c), que viven enteramente dentro del día 1.

(b) ¿Dónde buscar el día 1? Comparamos la probabilidad (incondicional, antes de saber el resultado) de hallar al perro el día 1 según dónde busque. Si busca en A , solo puede encontrarlo si el perro estaba en A :

$$\mathbb{P}(F_1 | S_1) = \mathbb{P}(F_1 | A, S_1) \mathbb{P}(A) = 0.25 \cdot 0.4 = 0.10.$$

Si busca en B , solo puede encontrarlo si el perro estaba en B :

$$\mathbb{P}(F_1 | S_1^c) = \mathbb{P}(F_1 | A^c, S_1^c) \mathbb{P}(A^c) = 0.15 \cdot 0.6 = 0.09.$$

Como $0.10 > 0.09$, conviene **buscar en el bosque A** el primer día.

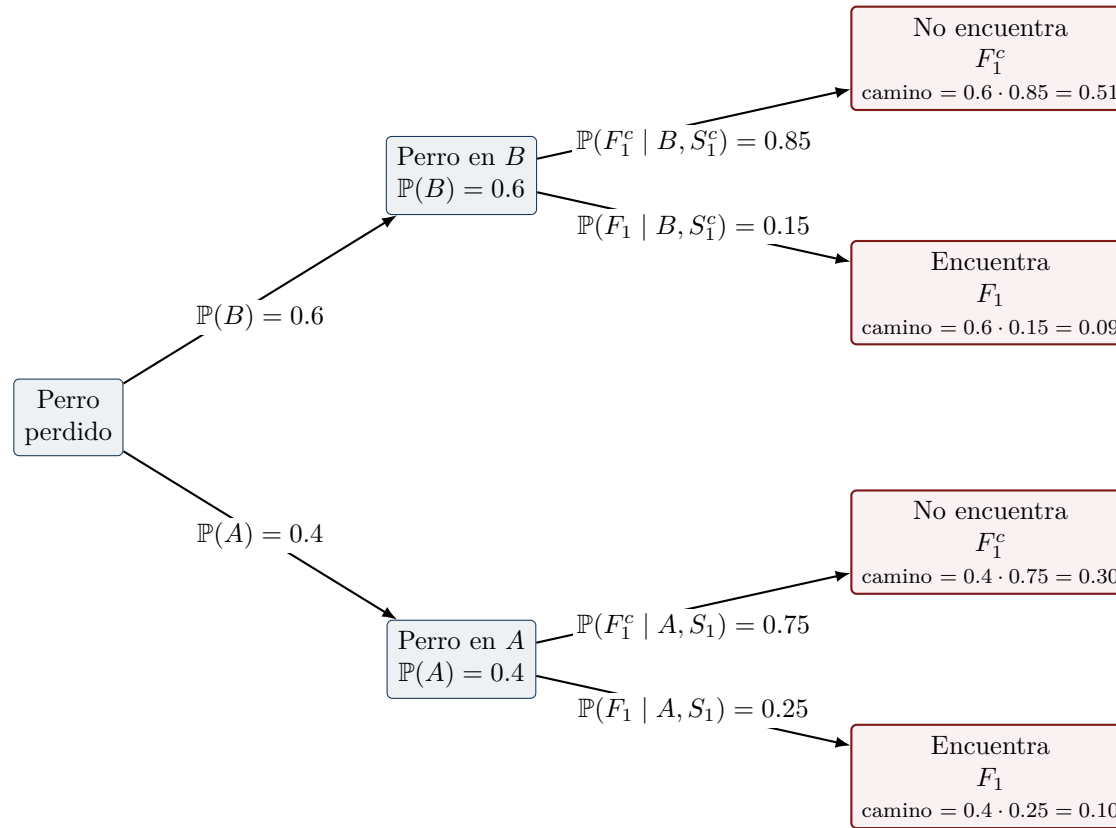
(c) Probabilidad posterior de que esté en A tras no hallarlo en A . Por la regla de Bayes, condicionando en $\{S_1, F_1^c\}$ (buscó en A y no lo encontró). Si el perro está en A , no encontrarlo tiene probabilidad $\mathbb{P}(F_1^c | A, S_1) = 0.75$; si está en B , como buscó en el bosque equivocado es *seguro* que no lo encuentra, $\mathbb{P}(F_1^c | A^c, S_1) = 1$. Entonces

$$\mathbb{P}(A | S_1, F_1^c) = \frac{\mathbb{P}(A) \mathbb{P}(F_1^c | A, S_1)}{\mathbb{P}(A) \mathbb{P}(F_1^c | A, S_1) + \mathbb{P}(A^c) \mathbb{P}(F_1^c | A^c, S_1)} = \frac{0.4 \cdot 0.75}{0.4 \cdot 0.75 + 0.6 \cdot 1} = \frac{0.30}{0.90} = \frac{1}{3}.$$

La probabilidad de que el perro esté en A baja de 0.4 (a priori) a $\frac{1}{3} \approx 0.33$ (posterior), porque buscar sin éxito en A es evidencia leve de que no estaba ahí.

Árbol de probabilidades — El perro de Oscar (día 1)

Las ramas son *probabilidades condicionales*; cada camino completo es el *producto* de sus ramas. S_1 = buscar en A , S_1^c = buscar en B .



Cómo leer el árbol. El segundo nivel muestra dos *escenarios de búsqueda* distintos: la rama superior asume que Oscar busca en A (S_1) y la inferior que busca en B (S_1^c); por eso las probabilidades de hallazgo difieren (0.25 frente a 0.15). Para la parte (b) se comparan los caminos “Encuentra”: 0.10 (buscar en A) frente a 0.09 (buscar en B). Para la parte (c) se usan los dos caminos bajo S_1 (buscar en A): el perro está en A y no lo encuentra ($0.4 \cdot 0.75 = 0.30$) frente a que esté en B y no lo encuentre ($0.6 \cdot 1 = 0.60$), de donde $\mathbb{P}(A | S_1, F_1^c) = 0.30 / (0.30 + 0.60) = \frac{1}{3}$. La muerte nocturna $N/(N+2)$ ocurriría después de este día y no interviene aquí.

Nivel 5 – Reto+

Muy difícil / fuera del scope típico: demostraciones y casos límite.

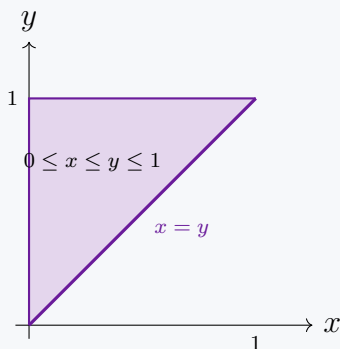
Ejercicio 1.41 (Reto+ — esperanza condicional $\mathbb{E}[X | Y]$ y ley de esperanza iterada). Un punto (X, Y) se elige *uniformemente* en el triángulo $0 \leq x \leq y \leq 1$, es decir $f(x, y) = 2$ en esa región (y 0 fuera). Piensa en Y = presupuesto disponible y X = gasto efectivo (nunca mayor que el presupuesto).

- Verifica que f integra 1 y halla las marginales f_X y f_Y .
- Halla la condicional $f_{X|Y}(x | y)$ e identifícala.
- Calcula $\mathbb{E}[X | Y = y]$ (como función de y).
- Verifica la **ley de esperanza iterada** $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | Y]]$ calculando ambos lados.
- Calcula $\text{Var}(\mathbb{E}[X | Y])$ y compáralo con $\text{Var}(X)$. ¿Qué dice la *ley de varianza total*?

(adaptado de MIT 18.05, Class 7; ley iterada estilo Blitzstein–Hwang)

Solución

(a) Área del triángulo = $\frac{1}{2}$, y $2 \cdot \frac{1}{2} = 1$. ✓



Marginales: para $x \in [0, 1]$, y va de x a 1, $f_X(x) = \int_x^1 2 dy = 2(1 - x)$; para $y \in [0, 1]$, x va de 0 a y , $f_Y(y) = \int_0^y 2 dx = 2y$.

(b) $f_{X|Y}(x | y) = \frac{2}{2y} = \frac{1}{y}$ para $0 \leq x \leq y$: dado $Y = y$, X es **uniforme** en $[0, y]$.

(c) Como $X | Y = y \sim \text{Unif}[0, y]$, su media es el punto medio:

$$\mathbb{E}[X | Y = y] = \frac{y}{2}.$$

(d) **Ley iterada.** Lado derecho, promediando $\mathbb{E}[X | Y] = \frac{Y}{2}$ con la ley de Y ($f_Y(y) = 2y$):

$$\mathbb{E}[\mathbb{E}[X | Y]] = \int_0^1 \frac{y}{2} \cdot 2y dy = \int_0^1 y^2 dy = \frac{1}{3}.$$

Lado izquierdo, directo con $f_X(x) = 2(1 - x)$:

$$\mathbb{E}[X] = \int_0^1 x \cdot 2(1 - x) dx = 2\left(\frac{1}{2} - \frac{1}{3}\right) = \frac{1}{3}.$$

Coinciden: $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | Y]] = \frac{1}{3}$. ✓

(e) Ley de varianza total. La ley dice $\text{Var}(X) = \mathbb{E}[\text{Var}(X | Y)] + \text{Var}(\mathbb{E}[X | Y])$.

■ $\text{Var}(\mathbb{E}[X | Y]) = \text{Var}\left(\frac{Y}{2}\right) = \frac{1}{4} \text{Var}(Y)$. Con $f_Y = 2y$: $\mathbb{E}[Y] = \frac{2}{3}$, $\mathbb{E}[Y^2] = \frac{1}{2}$, $\text{Var}(Y) = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$, luego $\text{Var}(\mathbb{E}[X | Y]) = \frac{1}{72}$.

■ $\text{Var}(X | Y = y) = \frac{y^2}{12}$ (varianza de $\text{Unif}[0, y]$), así $\mathbb{E}[\text{Var}(X | Y)] = \int_0^1 \frac{y^2}{12} \cdot 2y dy = \frac{1}{6} \cdot \frac{1}{4} = \frac{1}{24}$.

Suma: $\frac{1}{24} + \frac{1}{72} = \frac{3}{72} + \frac{1}{72} = \frac{4}{72} = \frac{1}{18}$. Y directamente $\text{Var}(X) = \mathbb{E}[X^2] - (\frac{1}{3})^2$ con $\mathbb{E}[X^2] = \int_0^1 x^2 \cdot 2(1 - x) dx = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$, da $\text{Var}(X) = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}$. ✓ **Interpretación:** de la varianza total de X ($\frac{1}{18}$), una parte ($\frac{1}{24}$) es la variabilidad *dentro* de cada presupuesto fijo y otra ($\frac{1}{72}$) viene de que el presupuesto Y mismo varía.

Ejercicio 1.42 (Reto+ — cambio de variables con jacobiano: $U = X + Y$, $V = \frac{X}{X+Y}$). Dos tiempos de espera independientes $X, Y \sim \text{Exponencial}(1)$ (densidad e^{-x} , e^{-y}). Define el tiempo total $U = X + Y$ y la proporción $V = \frac{X}{X+Y}$.

- Escribe la transformación inversa $x(u, v)$, $y(u, v)$ y calcula el jacobiano.
- Halla la densidad conjunta $f_{U,V}(u, v)$ sobre su soporte.
- Muestra que U y V son *independientes*, identifica sus distribuciones y di qué distribución tiene V .

(Reto+: el cambio de variables 2D con jacobiano excede el MicroMaster; adaptado de MIT 18.05 / Blitzstein–Hwang)

Solución

(a) Inversa y jacobiano. De $u = x + y$, $v = x/(x + y) = x/u$ se despeja

$$x = uv, \quad y = u(1 - v).$$

El jacobiano de (x, y) respecto de (u, v) :

$$J = \det \begin{pmatrix} \partial x / \partial u & \partial x / \partial v \\ \partial y / \partial u & \partial y / \partial v \end{pmatrix} = \det \begin{pmatrix} v & u \\ 1 - v & -u \end{pmatrix} = -uv - u(1 - v) = -u, \quad |J| = u.$$

(b) Densidad transformada. La conjunta de (X, Y) es $f_{X,Y}(x, y) = e^{-x}e^{-y} = e^{-(x+y)} = e^{-u}$. Por la fórmula de cambio de variables,

$$f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v)) |J| = e^{-u} \cdot u = u e^{-u},$$

válida en el soporte $u > 0$ y $0 < v < 1$ (pues $x, y > 0$ obligan $u > 0$ y $0 < v < 1$).

(c) Independencia y distribuciones. La densidad se *factoriza* y el soporte es un rectángulo $(0, \infty) \times (0, 1)$:

$$f_{U,V}(u, v) = \underbrace{(u e^{-u})}_{f_U(u)} \cdot \underbrace{(1)}_{f_V(v)}.$$

Luego U y V son independientes, con

$$U \sim \text{Gamma}(2, 1) \quad (f_U(u) = u e^{-u}), \quad V \sim \text{Unif}[0, 1] \quad (f_V(v) = 1).$$

(Verificación: $\int_0^\infty u e^{-u} du = 1$ y $\int_0^1 1 dv = 1$.) Es un resultado clásico: para exponenciales i.i.d., el *total* y la *proporción* son independientes, y la proporción es uniforme.

Ejercicio 1.43 (Reto+ — uniforme sobre un disco: $\text{Cov} = 0$ pero *no* independientes). Un punto (X, Y) se elige uniformemente en el disco unitario $D = \{(x, y) : x^2 + y^2 \leq 1\}$, de área π , así que $f(x, y) = \frac{1}{\pi}$ en D (y 0 fuera).

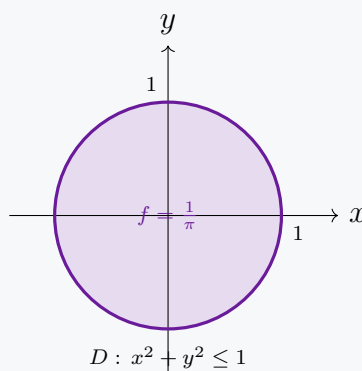
- Halla la marginal $f_X(x)$ y dibuja el soporte.
- Calcula $\mathbb{E}[X]$, $\mathbb{E}[Y]$ y $\mathbb{E}[XY]$; deduce $\text{Cov}(X, Y)$.
- ¿Son X e Y independientes? Justifica con cuidado (este es el punto clave).
- Calcula $\mathbb{P}(X^2 + Y^2 \leq \frac{1}{4})$.

(adaptado de MIT 18.05, Class 7: “zero covariance does not imply independence”)

Solución

(a) Marginal. Para $x \in [-1, 1]$, y recorre $[-\sqrt{1-x^2}, \sqrt{1-x^2}]$:

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2\sqrt{1-x^2}}{\pi}, \quad -1 \leq x \leq 1.$$



(b) Momentos y covarianza. Por simetría de D respecto de ambos ejes, $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. Para $\mathbb{E}[XY]$, la función xy es impar bajo $x \mapsto -x$ mientras el disco es

simétrico, así que

$$\mathbb{E}[XY] = \frac{1}{\pi} \iint_D xy \, dx \, dy = 0.$$

Por tanto $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$.

(c) ¿Independientes? NO. A pesar de $\text{Cov} = 0$, no son independientes: el soporte D *no es un rectángulo*, el rango de Y depende de X . Concretamente, si X está cerca de 1, entonces Y está forzado a estar cerca de 0 (la franja vertical es muy corta), mientras que si $X = 0$, Y puede ir de -1 a 1 . Además $f_X(x)f_Y(y) = \frac{4}{\pi^2}\sqrt{1-x^2}\sqrt{1-y^2} \neq \frac{1}{\pi} = f(x, y)$. Este es el contraejemplo clásico: **covarianza cero no implica independencia** (la covarianza solo detecta dependencia *lineal*, y aquí la dependencia es puramente geométrica/no lineal).

(d) Probabilidad radial. Es la razón de áreas (densidad constante):

$$\mathbb{P}\left(X^2 + Y^2 \leq \frac{1}{4}\right) = \frac{\text{área del disco de radio } 1/2}{\text{área de } D} = \frac{\pi(1/2)^2}{\pi(1)^2} = \frac{1}{4}.$$

Ejercicio 1.44 (Reto+ — de la tabla conjunta a Cov, Corr e interpretación). Una encuesta clasifica hogares por X = número de adultos que trabajan $(0, 1, 2)$ y Y = número de hijos en edad escolar $(0, 1, 2)$. La pmf conjunta (frecuencias relativas) es:

$f(x, y)$	$Y = 0$	$Y = 1$	$Y = 2$	$f_X(x)$
$X = 0$	0.10	0.04	0.01	
$X = 1$	0.08	0.20	0.12	
$X = 2$	0.02	0.16	0.27	
$f_Y(y)$				

- Completa las marginales f_X y f_Y .
- Calcula $\mathbb{E}[X]$, $\mathbb{E}[Y]$, $\mathbb{E}[XY]$ y $\text{Cov}(X, Y)$.
- Calcula $\text{Corr}(X, Y)$.
- ¿Son independientes? Da una celda que lo refute.
- Interpreta el signo y la magnitud de la correlación en este contexto social.

Solución

(a) Marginales. Sumando filas y columnas:

$$f_X(0) = 0.15, f_X(1) = 0.40, f_X(2) = 0.45; \quad f_Y(0) = 0.20, f_Y(1) = 0.40, f_Y(2) = 0.40.$$

(Comprobación: cada conjunto suma 1.)

(b) Momentos.

$$\mathbb{E}[X] = 0(0.15) + 1(0.40) + 2(0.45) = 1.30, \quad \mathbb{E}[Y] = 0(0.20) + 1(0.40) + 2(0.40) = 1.20.$$

$\mathbb{E}[XY] = \sum_{x,y} xy f(x, y)$; solo aportan las celdas con $x, y \geq 1$:

$$\mathbb{E}[XY] = 1 \cdot 1(0.20) + 1 \cdot 2(0.12) + 2 \cdot 1(0.16) + 2 \cdot 2(0.27) = 0.20 + 0.24 + 0.32 + 1.08 = 1.84.$$

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 1.84 - 1.30 \cdot 1.20 = 1.84 - 1.56 = 0.28.$$

(c) Correlación. $\mathbb{E}[X^2] = 0 + 1(0.40) + 4(0.45) = 2.20$, $\text{Var}(X) = 2.20 - 1.30^2 = 0.51$; $\mathbb{E}[Y^2] = 0 + 1(0.40) + 4(0.40) = 2.00$, $\text{Var}(Y) = 2.00 - 1.20^2 = 0.56$.

$$\text{Corr}(X, Y) = \frac{0.28}{\sqrt{0.51 \cdot 0.56}} = \frac{0.28}{\sqrt{0.2856}} \approx \frac{0.28}{0.5344} \approx 0.524.$$

(d) Independencia. No: por ejemplo $f(0, 0) = 0.10$ pero $f_X(0)f_Y(0) = 0.15 \cdot 0.20 = 0.03 \neq 0.10$.

(e) Interpretación. La correlación ≈ 0.52 es *positiva y moderada*: los hogares con más adultos trabajando tienden también a tener más hijos en edad escolar. Esto es *asociación*, no causalidad: ambas variables pueden subir juntas por un factor común (p.ej. hogares más grandes o en cierta etapa del ciclo de vida). La magnitud intermedia indica una relación real pero con bastante dispersión: hay hogares con dos trabajadores y sin hijos, y viceversa.

Fuentes y atribución

Material de repaso **basado en MIT 14.310x – Data Analysis for Social Scientists** (Esther Duflo & Sara Fisher Ellison), MIT OpenCourseWare, licencia **CC BY-NC-SA 4.0** (uso no comercial, con atribución, compartir bajo la misma licencia), y en las fuentes de la bibliografía. Los enunciados se basan en el material del curso y en diversas fuentes abiertas; cuando un problema se adapta de una fuente específica, se cita en el enunciado.

Bibliografía.

- Duflo, E. & Ellison, S. F. *14.310x: Data Analysis for Social Scientists*. MIT OpenCourseWare (slides, lecture notes y problem sets). ocw.mit.edu.
- Larsen, R. J. & Marx, M. L. *An Introduction to Mathematical Statistics and Its Applications*, 6.^a ed., Pearson, 2018.
- DeGroot, M. H. & Schervish, M. J. *Probability and Statistics*, 4.^a ed., Pearson, 2012.
- Blitzstein, J. & Hwang, J. *Introduction to Probability* (Harvard Stat 110); W. Chen, *Probability Cheatsheet*.
- Diez, D., Çetinkaya-Rundel, M. & Barr, C. *OpenIntro Statistics*. openintro.org.
- Materiales abiertos de la comunidad del curso en GitHub: [gevorg-hunanyan/probability](https://github.com/gevorg-hunanyan/probability); [hanmochen/Probability-Theory-2020](https://github.com/hanmochen/Probability-Theory-2020); [mynameisjanus/14310xDataAnalysis](https://github.com/mynameisjanus/14310xDataAnalysis).

creativecommons.org/licenses/by-nc-sa/4.0