

14.310x – Data Analysis for Social Scientists

MicroMasters en Statistics and Data Science (SDS) – MITx

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

Clase 2

Probabilidad Condicional, Independencia y Bayes

Soluciones

Módulos oficiales del MicroMaster:

M2 – Fundamentals of Probability (prob. condicional, independencia, Bayes); M3 – Joint/Marginal/Conditional (entrada)

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 (Calentamiento: probabilidad condicional con una tabla pequeña). En un salón de 40 estudiantes, 24 aprobaron matemática y, de *esos* aprobados, 18 también aprobaron lectura. Se elige un estudiante al azar.

- (a) ¿Cuál es la probabilidad de que haya aprobado matemática?
- (b) Dado que aprobó matemática, ¿cuál es la probabilidad de que haya aprobado lectura?

Solución

Sea M = “aprobó matemática” y L = “aprobó lectura”.

$$(a) \mathbb{P}(M) = \frac{24}{40} = 0.60.$$

(b) Al condicionar a M , el “nuevo” espacio muestral son los 24 que aprobaron matemática; de ellos 18 aprobaron lectura:

$$\mathbb{P}(L | M) = \frac{18}{24} = 0.75.$$

También por la fórmula: $\mathbb{P}(L \cap M) = 18/40 = 0.45$ y $\mathbb{P}(L | M) = \frac{\mathbb{P}(L \cap M)}{\mathbb{P}(M)} = \frac{0.45}{0.60} = 0.75$. Coincide.

Ejercicio 1.2 (Calentamiento: regla del producto con números redondos). En una comunidad, la probabilidad de que una familia tenga acceso a internet es $\mathbb{P}(I) = 0.50$. Dado que tiene internet, la probabilidad de que un hijo estudie en línea es $\mathbb{P}(E | I) = 0.80$. ¿Cuál es la probabilidad de que una familia *tenga internet y además* un hijo estudie en línea?

Solución

Por la regla del producto (multiplication rule):

$$\mathbb{P}(I \cap E) = \mathbb{P}(E | I) \mathbb{P}(I) = 0.80 \times 0.50 = 0.40.$$

El 40 por ciento de las familias cumple ambas condiciones.

Ejercicio 1.3 (Calentamiento: ¿independientes? compara dos números). Se sabe que $\mathbb{P}(A) = 0.4$, $\mathbb{P}(B) = 0.5$ y $\mathbb{P}(A \cap B) = 0.20$. ¿Son A y B independientes?

Solución

Comparamos $\mathbb{P}(A \cap B)$ con $\mathbb{P}(A)\mathbb{P}(B)$:

$$\mathbb{P}(A)\mathbb{P}(B) = 0.4 \times 0.5 = 0.20 = \mathbb{P}(A \cap B).$$

Como coinciden, A y B **sí** son independientes. La regla práctica: dos eventos son independientes exactamente cuando la probabilidad de que ocurran *ambos* es el producto de sus probabilidades.

Ejercicio 1.4 (Calentamiento: Bayes con dos cajas y números redondos). Hay dos cajas. La caja 1 tiene 1 bola roja y 1 azul; la caja 2 tiene 3 rojas y 1 azul. Eliges una caja al azar (cada una con probabilidad $1/2$) y sacas una bola.

- (a) ¿Cuál es la probabilidad de sacar una bola roja?
- (b) Si sacaste una roja, ¿cuál es la probabilidad de que viniera de la caja 2?

Solución

$$\mathbb{P}(C_1) = \mathbb{P}(C_2) = \frac{1}{2}, \mathbb{P}(\text{roja} | C_1) = \frac{1}{2}, \mathbb{P}(\text{roja} | C_2) = \frac{3}{4}.$$

(a) Ley de probabilidad total:

$$\mathbb{P}(\text{roja}) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{3}{4} = \frac{1}{4} + \frac{3}{8} = \frac{5}{8} = 0.625.$$

(b) Bayes:

$$\mathbb{P}(C_2 | \text{roja}) = \frac{\mathbb{P}(\text{roja} | C_2) \mathbb{P}(C_2)}{\mathbb{P}(\text{roja})} = \frac{\frac{3}{4} \cdot \frac{1}{2}}{\frac{5}{8}} = \frac{3/8}{5/8} = \frac{3}{5} = 0.6.$$

Ejercicio 1.5 (Definición de condicional). En un distrito, $\mathbb{P}(\text{un hogar es pobre y rural}) = 0.18$ y $\mathbb{P}(\text{un hogar es rural}) = 0.30$. Entre los hogares rurales, ¿qué fracción es pobre?

Solución

Por definición de probabilidad condicional, con R = rural y P = pobre:

$$\mathbb{P}(P | R) = \frac{\mathbb{P}(P \cap R)}{\mathbb{P}(R)} = \frac{0.18}{0.30} = 0.60.$$

Es decir, el 60 por ciento de los hogares rurales es pobre.

Ejercicio 1.6 (Regla del producto). La probabilidad de que un estudiante de una escuela rural termine la secundaria es $\mathbb{P}(S) = 0.40$. Dado que terminó la secundaria, la probabilidad de que acceda a la universidad es $\mathbb{P}(U | S) = 0.25$. ¿Cuál es la probabilidad de que un estudiante *termine la secundaria y además* acceda a la universidad?

Solución

Por la regla del producto (multiplication rule):

$$\mathbb{P}(S \cap U) = \mathbb{P}(U | S) \mathbb{P}(S) = 0.25 \times 0.40 = 0.10.$$

El 10 por ciento de los estudiantes logra ambas cosas.

Ejercicio 1.7 (¿Independientes? — caso sí). Se elige al azar a una persona de un padrón. Sea A = “nació en un mes par” con $\mathbb{P}(A) = 0.5$ y B = “su DNI termina en dígito par” con $\mathbb{P}(B) = 0.5$. Se sabe que $\mathbb{P}(A \cap B) = 0.25$. ¿Son A y B independientes?

Solución

Comparamos $\mathbb{P}(A \cap B)$ con $\mathbb{P}(A)\mathbb{P}(B)$:

$$\mathbb{P}(A)\mathbb{P}(B) = 0.5 \times 0.5 = 0.25 = \mathbb{P}(A \cap B).$$

Como coinciden, A y B **sí** son independientes: el mes de nacimiento no aporta información sobre el último dígito del DNI.

Ejercicio 1.8 (Complemento condicional). En cierta clínica, dado que un paciente tiene síntomas de gripe, la probabilidad de que sea realmente influenza es $\mathbb{P}(I | S) = 0.35$. ¿Cuál es la probabilidad de que *no* sea influenza, dado que tiene síntomas?

Solución

El complemento funciona dentro del mismo condicionamiento:

$$\mathbb{P}(I^c | S) = 1 - \mathbb{P}(I | S) = 1 - 0.35 = 0.65.$$

Ejercicio 1.9 (Tabla de contingencia 2x2 — lectura). Se encuestó a 200 personas sobre si participan en un programa social y si conocen sus derechos. Los conteos son:

	Conoce derechos	No conoce	Total
Participa	70	30	100
No participa	40	60	100
Total	110	90	200

Calcula $\mathbb{P}(\text{conoce derechos})$ y $\mathbb{P}(\text{conoce} | \text{participa})$.

Solución

$$\mathbb{P}(\text{conoce}) = \frac{110}{200} = 0.55.$$

$$\mathbb{P}(\text{conoce} | \text{participa}) = \frac{70}{100} = 0.70 \text{ (restringimos al “mundo” de los 100 participantes).}$$

Nivel 2 – Núcleo (core)

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

Ejercicio 1.10 (Operaciones de conjuntos y leyes). Sea el espacio muestral $S = \{1, 2, 3, 4, 5, 6\}$ y los eventos $A = \{1, 2\}$, $B = \{2, 4, 5\}$, $C = \{1, 5, 6\}$. Halla:

- (a) $A \cup B$; (b) $A \cap B$; (c) A^c ; (d) B^c .
 (e) Verifica una ley de De Morgan: $(A \cup B)^c = A^c \cap B^c$.
 (f) Verifica la ley distributiva: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Solución

- (a) $A \cup B = \{1, 2, 4, 5\}$. (b) $A \cap B = \{2\}$. (c) $A^c = \{3, 4, 5, 6\}$. (d) $B^c = \{1, 3, 6\}$.
 (e) $A \cup B = \{1, 2, 4, 5\} \Rightarrow (A \cup B)^c = \{3, 6\}$. Por otro lado $A^c \cap B^c = \{3, 4, 5, 6\} \cap \{1, 3, 6\} = \{3, 6\}$. Coinciden: De Morgan se cumple.
 (f) $B \cup C = \{1, 2, 4, 5, 6\} \Rightarrow A \cap (B \cup C) = \{1, 2\}$. Y $(A \cap B) \cup (A \cap C) = \{2\} \cup \{1\} = \{1, 2\}$. Coinciden: la distributiva se cumple.

Ejercicio 1.11 (Independencia con un número aleatorio). Se elige N al azar de manera uniforme en $\{1, 2, \dots, 10\}$. Sean los eventos $A = \{N < 7\}$ y $B = \{N \text{ es par}\}$. ¿Son A y B independientes, aunque ambos hablen del *mismo* número N ?

Solución

$$\begin{aligned} A = \{1, 2, 3, 4, 5, 6\} &\Rightarrow \mathbb{P}(A) = 6/10 = 0.6. \\ B = \{2, 4, 6, 8, 10\} &\Rightarrow \mathbb{P}(B) = 5/10 = 0.5. \\ A \cap B = \{2, 4, 6\} &\Rightarrow \mathbb{P}(A \cap B) = 3/10 = 0.3. \end{aligned}$$

Como $\mathbb{P}(A)\mathbb{P}(B) = 0.6 \times 0.5 = 0.3 = \mathbb{P}(A \cap B)$, los eventos **sí** son independientes. La independencia es una propiedad *numérica* de las probabilidades: no exige que los eventos provengan de experimentos distintos.

Ejercicio 1.12 (¿Independientes? — caso no). En una encuesta, $\mathbb{P}(\text{tiene seguro de salud}) = 0.60$, $\mathbb{P}(\text{tiene empleo formal}) = 0.50$ y $\mathbb{P}(\text{seguro y formal}) = 0.45$. ¿Son independientes “tener seguro” y “tener empleo formal”?

Solución

$\mathbb{P}(\text{seguro})\mathbb{P}(\text{formal}) = 0.60 \times 0.50 = 0.30$, pero el dato real es $\mathbb{P}(\text{seguro} \cap \text{formal}) = 0.45 \neq 0.30$. No coinciden, así que **no** son independientes. De hecho $0.45 > 0.30$: tener empleo formal hace *más* probable tener seguro.

Ejercicio 1.13 (Tabla 2x2 — ¿independencia?). Con la misma tabla del nivel anterior (200 personas; participar vs. conocer derechos: celdas 70, 30, 40, 60), ¿son independientes “participar” y “conocer derechos”?

Solución

$\mathbb{P}(\text{participa}) = 100/200 = 0.5$ y $\mathbb{P}(\text{conoce}) = 110/200 = 0.55$. Si fueran independientes, $\mathbb{P}(\text{participa} \cap \text{conoce})$ debería ser $0.5 \times 0.55 = 0.275$, es decir 55 personas. Pero la celda real es $70/200 = 0.35$. Como $0.35 \neq 0.275$, **no** son independientes: participar se asocia con conocer más los derechos.

Ejercicio 1.14 (Regla del producto en dos pasos). Una urna de un sorteo comunal tiene 5 balotas premiadas y 15 sin premio (20 en total). Se extraen 2 sin reemplazo. ¿Probabilidad de que *ambas* sean premiadas?

Solución

Sea P_1 premiada en la primera y P_2 en la segunda. Por la regla del producto:

$$\mathbb{P}(P_1 \cap P_2) = \mathbb{P}(P_1) \mathbb{P}(P_2 | P_1) = \frac{5}{20} \cdot \frac{4}{19} = \frac{20}{380} = \frac{1}{19} \approx 0.0526.$$

Ejercicio 1.15 (Dos encuestadores: al menos uno, exactamente uno). Dos encuestadores trabajan de forma independiente un mismo barrio. El encuestador 1 logra completar una entrevista en una vivienda con probabilidad 0.7 y el encuestador 2 con probabilidad 0.8. Para una vivienda dada:

- (a) ¿Probabilidad de que la vivienda sea entrevistada *al menos una vez*?
- (b) ¿Probabilidad de que sea entrevistada *exactamente una vez*?

Solución

Sean H_1 y H_2 los éxitos de cada encuestador, independientes, con $\mathbb{P}(H_1) = 0.7$, $\mathbb{P}(H_2) = 0.8$.

(a) Por complemento (que *ninguno* la entreviste):

$$\mathbb{P}(\geq 1) = 1 - (1 - 0.7)(1 - 0.8) = 1 - 0.3 \times 0.2 = 1 - 0.06 = 0.94.$$

(b) “Exactamente uno” es la suma de dos casos disjuntos:

$$\mathbb{P}(\text{exactamente uno}) = 0.7 \times 0.2 + 0.3 \times 0.8 = 0.14 + 0.24 = 0.38.$$

(Chequeo: $\mathbb{P}(\text{ambos}) = 0.7 \times 0.8 = 0.56$, y $0.56 + 0.38 = 0.94 = \mathbb{P}(\geq 1)$.) (*adaptado de gevorg-hunanyan/probability, two-shooters*)

Ejercicio 1.16 (Condicional con dos dados sociales). En un juego de feria comunal se lanzan dos dados justos. Dado que la suma de las caras es 8, ¿cuál es la probabilidad de que *al menos uno* de los dados muestre un 6?

Solución

Las parejas (ordenadas) con suma 8 son (2, 6), (3, 5), (4, 4), (5, 3), (6, 2): en total 5 resultados igualmente probables, que forman el “nuevo” espacio muestral al condicionar. De ellas, contienen un 6 exactamente dos: (2, 6) y (6, 2). Por tanto

$$\mathbb{P}(\text{algún 6} \mid \text{suma} = 8) = \frac{2}{5} = 0.4.$$

(*adaptado de gevorg-hunanyan/probability, twodice-6-given-sum-8*)

Ejercicio 1.17 (Moneda hasta la primera cruz). Se lanza una moneda justa repetidamente hasta obtener la primera cruz. Sea X el número total de lanzamientos realizados. Halla $\mathbb{P}(X = 5)$.

Solución

$X = 5$ exige una única secuencia: cara, cara, cara, cara y recién cruz (cuatro caras seguidas de una cruz). Como los lanzamientos son independientes y cada cara o cruz tiene probabilidad $1/2$:

$$\mathbb{P}(X = 5) = \left(\frac{1}{2}\right)^4 \cdot \frac{1}{2} = \left(\frac{1}{2}\right)^5 = \frac{1}{32} \approx 0.03125.$$

(X es una variable geométrica con parámetro $p = 1/2$: $\mathbb{P}(X = k) = (1/2)^k$.)

Ejercicio 1.18 (Tres monedas (una de dos caras)). Una caja tiene 3 monedas: dos normales (con cara y cruz) y una falsa con *dos caras* (siempre sale cara). Eliges una moneda al azar y la lanzas una vez.

- (a) ¿Cuál es la probabilidad de obtener cara?
- (b) Dado que saliste cara, ¿cuál es la probabilidad de que hayas tomado la moneda de dos caras?

Solución

Sea C_1 = “moneda normal” ($\mathbb{P}(C_1) = 2/3$, $\mathbb{P}(\text{cara} | C_1) = 1/2$) y C_2 = “moneda de dos caras” ($\mathbb{P}(C_2) = 1/3$, $\mathbb{P}(\text{cara} | C_2) = 1$).

(a) Ley de probabilidad total:

$$\mathbb{P}(\text{cara}) = \frac{1}{2} \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

(b) Bayes:

$$\mathbb{P}(C_2 | \text{cara}) = \frac{\mathbb{P}(\text{cara} | C_2) \mathbb{P}(C_2)}{\mathbb{P}(\text{cara})} = \frac{1 \cdot \frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}.$$

Observar una cara duplica la creencia de tener la moneda falsa: de $1/3$ (prior) a $1/2$.

Ejercicio 1.19 (Ley de probabilidad total con tres regiones). Un programa nacional opera en tres regiones. La región A atiende al 50 por ciento de los beneficiarios, la B al 30 por ciento y la C al 20 por ciento. La tasa de error administrativo es 4 por ciento en A, 6 por ciento en B y 9 por ciento en C. ¿Cuál es la probabilidad de que un beneficiario al azar tenga un error en su expediente?

Solución

Por la ley de probabilidad total sobre la partición $\{A, B, C\}$:

$$\mathbb{P}(\text{error}) = 0.04 \cdot 0.50 + 0.06 \cdot 0.30 + 0.09 \cdot 0.20 = 0.020 + 0.018 + 0.018 = 0.056.$$

La probabilidad es 0.056 (5.6 por ciento).

Ejercicio 1.20 (Bayes con partición de tres). Con los datos del ejercicio anterior, un expediente tiene error. ¿Cuál es la probabilidad de que provenga de la región C?

Solución

Por Bayes, usando $\mathbb{P}(\text{error}) = 0.056$ ya calculado:

$$\mathbb{P}(C | \text{error}) = \frac{\mathbb{P}(\text{error} | C) \mathbb{P}(C)}{\mathbb{P}(\text{error})} = \frac{0.09 \times 0.20}{0.056} = \frac{0.018}{0.056} \approx 0.321.$$

Aunque C aporta solo el 20 por ciento de los expedientes, aporta el 32 por ciento de los errores, por su alta tasa de error.

Nivel 3 – Aplicación

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

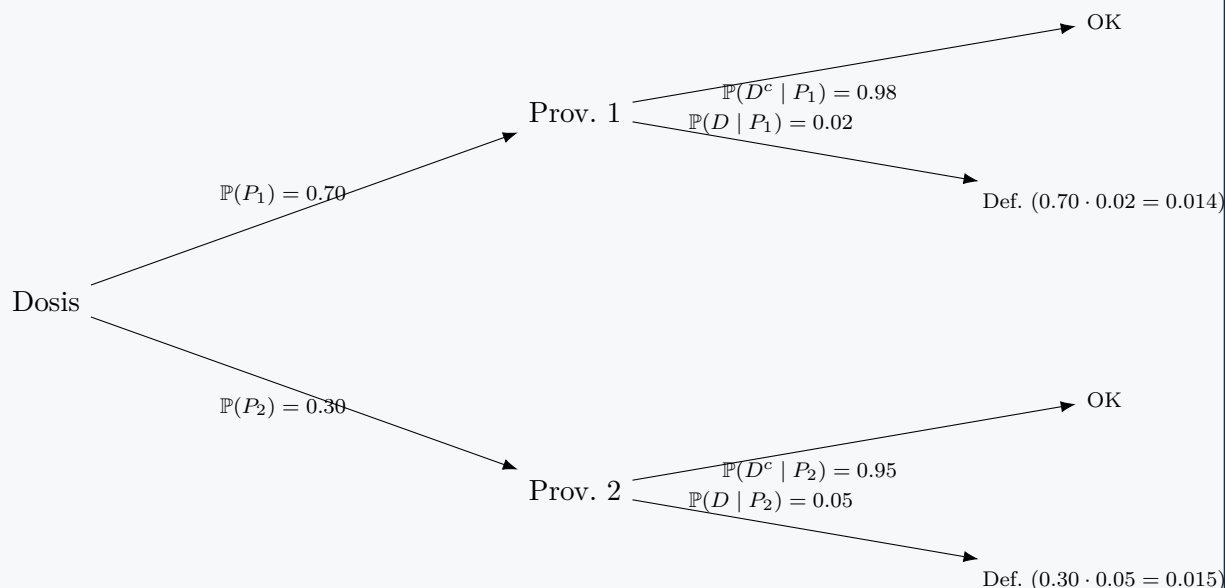
Ejercicio 1.21 (Conjuntos contables o incontables). Clasifica cada conjunto como *contable* (finito o numerable) o *incontable*:

- (a) $A = \{x \in \mathbb{Q} \mid -100 \leq x \leq 100\}$.
- (b) $B = \{(x, y) \mid x \in \mathbb{N}, y \in \mathbb{Z}\}$.
- (c) $C = (0, 0.1]$ (intervalo de números reales).
- (d) $D = \{1/n \mid n \in \mathbb{N}\}$.

Solución

- (a) **Contable**: es un subconjunto de los racionales $\mathbb{Q}$, que es numerable.
- (b) **Contable**: $\mathbb{N} \times \mathbb{Z}$ es un producto de dos conjuntos numerables, luego es numerable.
- (c) **Incontable**: todo intervalo real con interior no vacío tiene la cardinalidad del continuo.
- (d) **Contable**: se puede listar explícitamente $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$, luego es numerable.

Ejercicio 1.22 (Diagrama de árbol: dos proveedores de vacunas). Un centro de salud recibe vacunas de dos proveedores: el 70 por ciento del lote viene del proveedor 1 y el 30 por ciento del proveedor 2. El 2 por ciento de las dosis del proveedor 1 sale defectuosa y el 5 por ciento de las del proveedor 2. Dibuja el árbol y calcula la probabilidad de que una dosis al azar sea defectuosa.

Solución

Cada rama es una probabilidad condicional: depende del nodo del que sale; la probabilidad de un camino es el producto de sus ramas. Sumando los dos caminos que llegan a “defectuosa”:

$$\mathbb{P}(\text{def.}) = 0.70 \cdot 0.02 + 0.30 \cdot 0.05 = 0.014 + 0.015 = 0.029.$$

Ejercicio 1.23 (Bayes sobre el árbol anterior). Con el árbol del ejercicio previo, una dosis resultó defectuosa. ¿Cuál es la probabilidad de que provenga del proveedor 2?

Solución

Por Bayes, con $\mathbb{P}(\text{def.}) = 0.029$:

$$\mathbb{P}(\text{Prov.2} | \text{def.}) = \frac{0.30 \times 0.05}{0.029} = \frac{0.015}{0.029} \approx 0.517.$$

Aunque el proveedor 2 aporta solo el 30 por ciento de las dosis, aporta más de la mitad de las defectuosas.

Ejercicio 1.24 (Falsos positivos y tasa base). Una prueba rápida de anemia tiene sensibilidad 95 por ciento (detecta al enfermo) y da falso positivo en el 10 por ciento de los sanos. En una población donde el 5 por ciento tiene anemia, una persona da positivo. ¿Cuál es la probabilidad de que realmente tenga anemia?

Solución

Sea E enfermo ($\mathbb{P}(E) = 0.05$) y $+$ positivo, con $\mathbb{P}(+ | E) = 0.95$ y $\mathbb{P}(+ | E^c) = 0.10$. Por

Bayes:

$$\mathbb{P}(E | +) = \frac{0.95 \times 0.05}{0.95 \times 0.05 + 0.10 \times 0.95} = \frac{0.0475}{0.0475 + 0.095} = \frac{0.0475}{0.1425} \approx 0.333.$$

Solo un tercio de los positivos está realmente enfermo: la baja prevalencia (tasa base) infla los falsos positivos.

Ejercicio 1.25 (Dos detecciones independientes). Un sistema de auditoría usa dos revisores independientes. El revisor 1 detecta un fraude con probabilidad 0.7 y el revisor 2 con probabilidad 0.6, de forma independiente.

- (a) ¿Probabilidad de que *al menos uno* detecte el fraude?
- (b) ¿Probabilidad de que *ambos* lo detecten?

Solución

- (a) Por complemento (ninguno detecta): $\mathbb{P}(\geq 1) = 1 - (1 - 0.7)(1 - 0.6) = 1 - 0.3 \times 0.4 = 1 - 0.12 = 0.88$.
- (b) Por independencia: $\mathbb{P}(\text{ambos}) = 0.7 \times 0.6 = 0.42$.

Ejercicio 1.26 (Dos urnas y traslado de información). Hay dos cajas de expedientes. La caja I tiene 3 aprobados y 7 rechazados; la caja II tiene 6 aprobados y 4 rechazados. Se elige una caja al azar (mismas probabilidades) y de ella se extrae un expediente.

- (a) ¿Probabilidad de extraer un aprobado?
- (b) Si salió aprobado, ¿probabilidad de que viniera de la caja II?

Solución

- $\mathbb{P}(I) = \mathbb{P}(II) = 0.5$, $\mathbb{P}(\text{aprob} | I) = 3/10$, $\mathbb{P}(\text{aprob} | II) = 6/10$.
- (a) Ley de probabilidad total: $\mathbb{P}(\text{aprob}) = 0.5 \cdot \frac{3}{10} + 0.5 \cdot \frac{6}{10} = 0.15 + 0.30 = 0.45$.
 - (b) Bayes: $\mathbb{P}(II | \text{aprob}) = \frac{0.5 \cdot 0.6}{0.45} = \frac{0.30}{0.45} = \frac{2}{3} \approx 0.667$.

Ejercicio 1.27 (Dos padrones, mismo perfil). Una ONG tiene dos padrones de beneficiarios. El padrón I contiene 5 urbanos, 11 rurales y 8 periurbanos; el padrón II contiene 10 urbanos, 8 rurales y 6 periurbanos. Se extrae al azar una ficha de *cada* padrón (de forma independiente). ¿Cuál es la probabilidad de que ambas fichas correspondan al *mismo* tipo de zona?

Solución

Cada padrón tiene 24 fichas. Probabilidades por tipo:
 Padrón I: urbano 5/24, rural 11/24, periurbano 8/24.

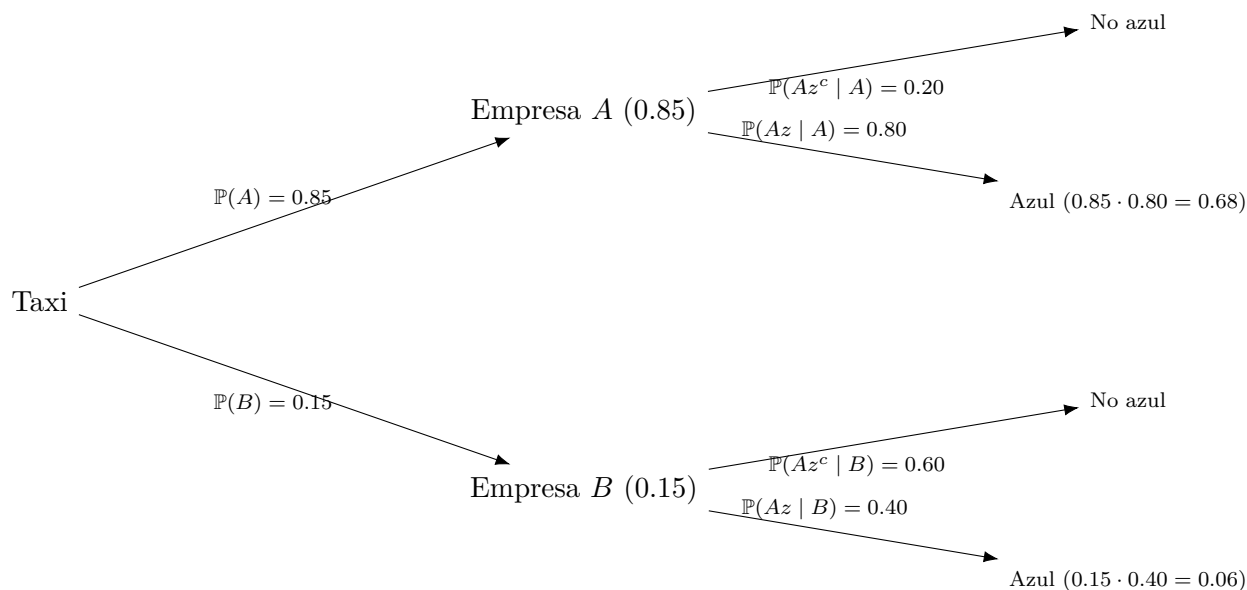
Padrón II: urbano 10/24, rural 8/24, periurbano 6/24.

Como las extracciones son independientes, “mismo tipo” es la suma de los tres productos:

$$\mathbb{P}(\text{mismo tipo}) = \frac{5 \cdot 10 + 11 \cdot 8 + 8 \cdot 6}{24 \cdot 24} = \frac{50 + 88 + 48}{576} = \frac{186}{576} \approx 0.3229.$$

(adaptado de *gevorg-hunanyan/probability, bowl-probabilities*)

Ejercicio 1.28 (Árbol y Bayes: el testigo del taxi). En una ciudad, el 85 por ciento de los taxis pertenece a la empresa A (de los cuales el 80 por ciento es azul) y el 15 por ciento a la empresa B (de los cuales el 40 por ciento es azul). Un testigo de un accidente reporta haber visto un taxi *azul*. Suponiendo que el reporte del color es fiable, construye el árbol y halla la probabilidad de que el taxi involucrado fuera de la empresa A .



Cada rama es una probabilidad condicional: depende del nodo del que sale; la probabilidad de un camino es el producto de sus ramas.

Solución

Ley de probabilidad total para “azul”, sumando los dos caminos azules del árbol:

$$\mathbb{P}(\text{azul}) = 0.85 \cdot 0.80 + 0.15 \cdot 0.40 = 0.68 + 0.06 = 0.74.$$

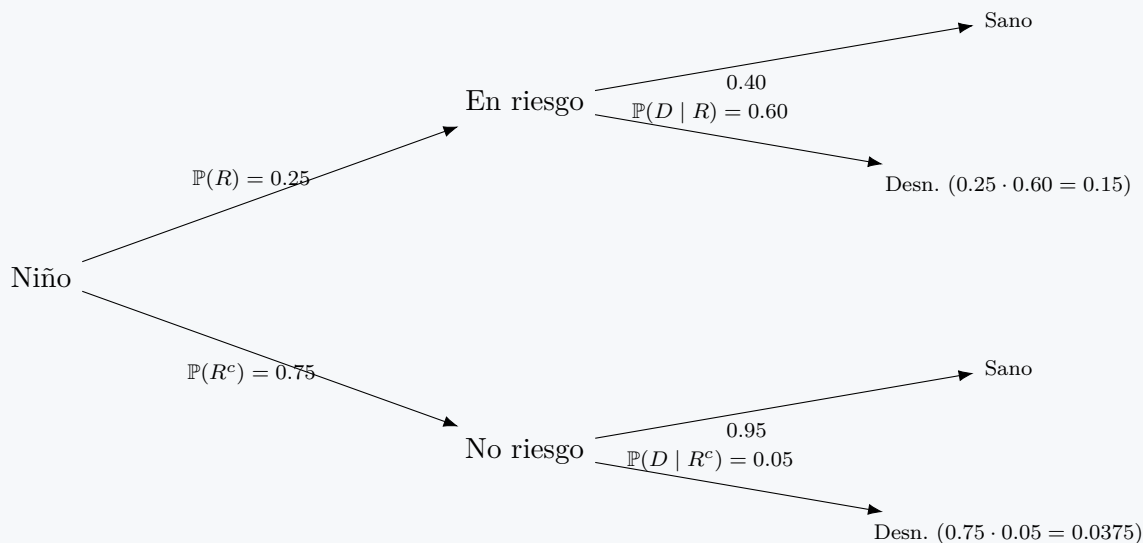
Por Bayes:

$$\mathbb{P}(A | \text{azul}) = \frac{0.85 \times 0.80}{0.74} = \frac{0.68}{0.74} \approx 0.919.$$

A pesar de que la empresa B tiene mayor *proporción* de taxis azules, su flota es tan pequeña que un taxi azul casi siempre proviene de la empresa A . (adaptado de *gevorg-hunanyan/probability, witness-taxi*)

Ejercicio 1.29 (Diagrama de árbol de dos etapas con Bayes). Un tamizaje de desnutrición tiene dos etapas. En la etapa 1, el 25 por ciento de los niños es marcado “en riesgo”. De los marcados en riesgo, el 60 por ciento resulta desnutrido en la etapa 2; de los *no* marcados, solo el 5 por ciento resulta desnutrido. Construye el árbol y halla: (a) la probabilidad de estar desnutrido; (b) dado que un niño está desnutrido, la probabilidad de que hubiera sido marcado “en riesgo”.

Solución



(a) $\mathbb{P}(\text{desn.}) = 0.25 \cdot 0.60 + 0.75 \cdot 0.05 = 0.15 + 0.0375 = 0.1875$.

(b) Por Bayes:

$$\mathbb{P}(\text{riesgo} | \text{desn.}) = \frac{0.15}{0.1875} = 0.80.$$

El 80 por ciento de los desnutridos había sido marcado “en riesgo”, lo que valida el tamizaje de la etapa 1.

Ejercicio 1.30 (Bayes secuencial: dos observaciones). Una ONG cree que el 30 por ciento de los hogares de una zona es elegible para un bono ($\mathbb{P}(E) = 0.3$). Aplica un cuestionario de tamizaje que, en hogares elegibles, da “apto” con probabilidad 0.9, y en no elegibles da “apto” con probabilidad 0.2. El mismo cuestionario se aplica *dos veces* de forma independiente (condicional al verdadero estado del hogar) y ambas veces sale “apto”. ¿Cuál es la probabilidad de que el hogar sea elegible tras las dos observaciones?

Solución

Condional al estado, las dos pruebas son independientes, así: $\mathbb{P}(\text{apto}, \text{apto} | E) = 0.9^2 = 0.81$ y $\mathbb{P}(\text{apto}, \text{apto} | E^c) = 0.2^2 = 0.04$. Por Bayes:

$$\mathbb{P}(E | \text{apto}, \text{apto}) = \frac{0.81 \times 0.30}{0.81 \times 0.30 + 0.04 \times 0.70} = \frac{0.243}{0.243 + 0.028} = \frac{0.243}{0.271} \approx 0.897.$$

La creencia sube de 0.30 a ≈ 0.90 .

Ejercicio 1.31 (Bayes con tres hipótesis y dos señales). Una agencia clasifica solicitudes en tres tipos: legítimas (L , 60 por ciento), dudosas (D , 30 por ciento) y fraudulentas (F , 10 por ciento). Un detector emite “alerta” con probabilidad 0.05 en legítimas, 0.40 en dudosas y 0.90 en fraudulentas. Una solicitud genera alerta. ¿Cuál es la probabilidad de que sea fraudulenta?

Solución

Ley de probabilidad total para la alerta:

$$\mathbb{P}(\text{alerta}) = 0.05 \cdot 0.60 + 0.40 \cdot 0.30 + 0.90 \cdot 0.10 = 0.03 + 0.12 + 0.09 = 0.24.$$

Por Bayes:

$$\mathbb{P}(F | \text{alerta}) = \frac{0.90 \times 0.10}{0.24} = \frac{0.09}{0.24} = 0.375.$$

También es útil notar $\mathbb{P}(D | \text{alerta}) = 0.12/0.24 = 0.50$ y $\mathbb{P}(L | \text{alerta}) = 0.03/0.24 = 0.125$ (suman 1).

Nivel 4 – Reto

Difícil: integra varios temas en un mismo problema.

Ejercicio 1.32 (Vida útil de un producto (condicional)). El tiempo de vida T (en años) de un producto cumple $\mathbb{P}(T \geq t) = e^{-t/5}$ para $t \geq 0$ (por ejemplo $\mathbb{P}(T \geq 2) = e^{-2/5} \approx 0.6703$). El producto ya funcionó 2 años sin fallar. ¿Cuál es la probabilidad de que falle *durante el tercer año*?

Solución

Sea $A = \{2 \leq T \leq 3\}$ (“falla en el tercer año”) y $B = \{T \geq 2\}$ (“no falló en los dos primeros años”). Como $A \subset B$, $A \cap B = A$, y:

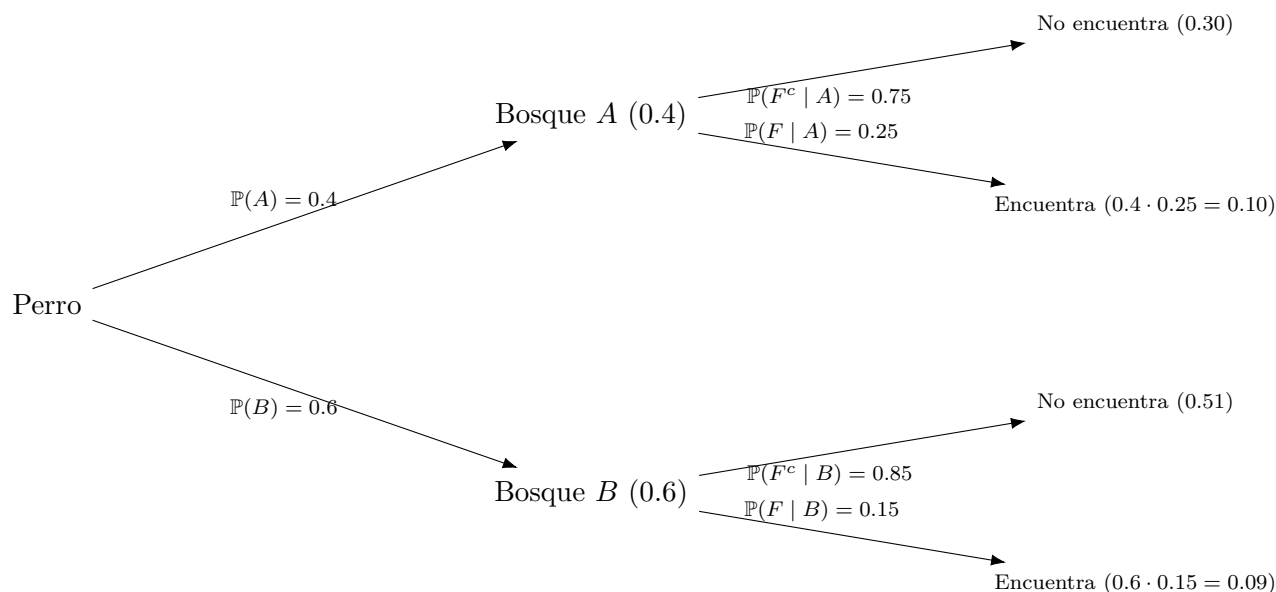
$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(T \geq 2) - \mathbb{P}(T \geq 3)}{\mathbb{P}(T \geq 2)} = \frac{e^{-2/5} - e^{-3/5}}{e^{-2/5}} = 1 - e^{-1/5} \approx 0.1813.$$

Notar que $1 - e^{-1/5} = \mathbb{P}(T < 1)$: por la *falta de memoria* de la ley exponencial, un producto que ya sobrevivió 2 años enfrenta el tercer año como si fuera nuevo.

Ejercicio 1.33 (El perro perdido de Oscar). El perro de Oscar se perdió: está en el bosque A con probabilidad 0.4 o en el bosque B con probabilidad 0.6. Si Oscar busca *un día* en el bosque correcto, lo encuentra con probabilidad 0.25 si está en A y 0.15 si está en B ; si busca en el bosque equivocado, la probabilidad de encontrarlo ese día es 0.

- (a) ¿En qué bosque debe buscar el primer día para maximizar la probabilidad de encontrarlo?

- (b) Buscó en A el primer día y *no* lo encontró. ¿Cuál es ahora la probabilidad de que el perro esté en A ?



Cada rama es una probabilidad condicional: depende del nodo del que sale; la probabilidad de un camino es el producto de sus ramas.

Solución

(a) Si busca en A el primer día, lo encuentra solo si el perro está en A y tiene éxito: $\mathbb{P}(\text{encontrar} \mid \text{busca } A) = 0.4 \times 0.25 = 0.10$. Si busca en B : $0.6 \times 0.15 = 0.09$. Como $0.10 > 0.09$, conviene buscar en el **bosque A**.

(b) Buscó en A y no encontró. Si el perro está en A , no encontrarlo tiene probabilidad 0.75; si está en B , no encontrarlo (buscando en A) tiene probabilidad 1. Por Bayes:

$$\mathbb{P}(A \mid \text{no enc. en } A) = \frac{0.4 \times 0.75}{0.4 \times 0.75 + 0.6 \times 1} = \frac{0.30}{0.30 + 0.60} = \frac{0.30}{0.90} = \frac{1}{3} \approx 0.3333.$$

La búsqueda fallida en A *baja* la creencia de que el perro esté en A de 0.4 a $1/3$.

Ejercicio 1.34 (Tamizaje masivo con tasa base baja: ¿sirve un positivo?). Un programa de salud pública aplica una prueba rápida de tuberculosis a *toda* la población de un distrito. La prevalencia real de la enfermedad es del 0.5 por ciento. La prueba tiene **sensibilidad** 99 por ciento ($\mathbb{P}(+ \mid \text{enfermo}) = 0.99$) y **tasa de falsos positivos** 5 por ciento ($\mathbb{P}(+ \mid \text{sano}) = 0.05$).

- (a) Imagina 100,000 personas tamizadas. Llena la tabla de cuatro celdas (enfermo/sano $\times$ positivo/negativo).
- (b) Una persona da positivo. ¿Cuál es el *valor predictivo positivo* $\mathbb{P}(\text{enfermo} \mid +)$? Interpreta por qué es tan bajo pese a una prueba “buena”.

- (c) A esa misma persona se le repite la prueba (independiente, condicional al estado) y vuelve a salir positiva. ¿Cuál es ahora $\mathbb{P}(\text{enfermo} \mid +, +)$?

(adaptado de MIT 18.05, Class 3 — the base rate fallacy)

Solución

(a) De 100,000 personas, 0.5 % son enfermas: 500 enfermas y 99,500 sanas.

	Positivo	Negativo	Total
Enfermo	495	5	500
Sano	4,975	94,525	99,500
Total	5,470	94,530	100,000

(Verdaderos positivos $0.99 \times 500 = 495$; falsos positivos $0.05 \times 99,500 = 4,975$.)

(b) Valor predictivo positivo:

$$\mathbb{P}(\text{enf.} \mid +) = \frac{495}{5,470} \approx 0.0905.$$

Solo el 9 por ciento de los positivos está realmente enfermo. La razón es la **tasa base** (base rate): como casi todos están sanos, el pequeño 5 por ciento de falsos positivos sobre una población enorme (4,975 personas) supera por mucho a los 495 verdaderos positivos. “99 % de las pruebas son acertadas” *no* implica “99 % de los positivos son acertados”.

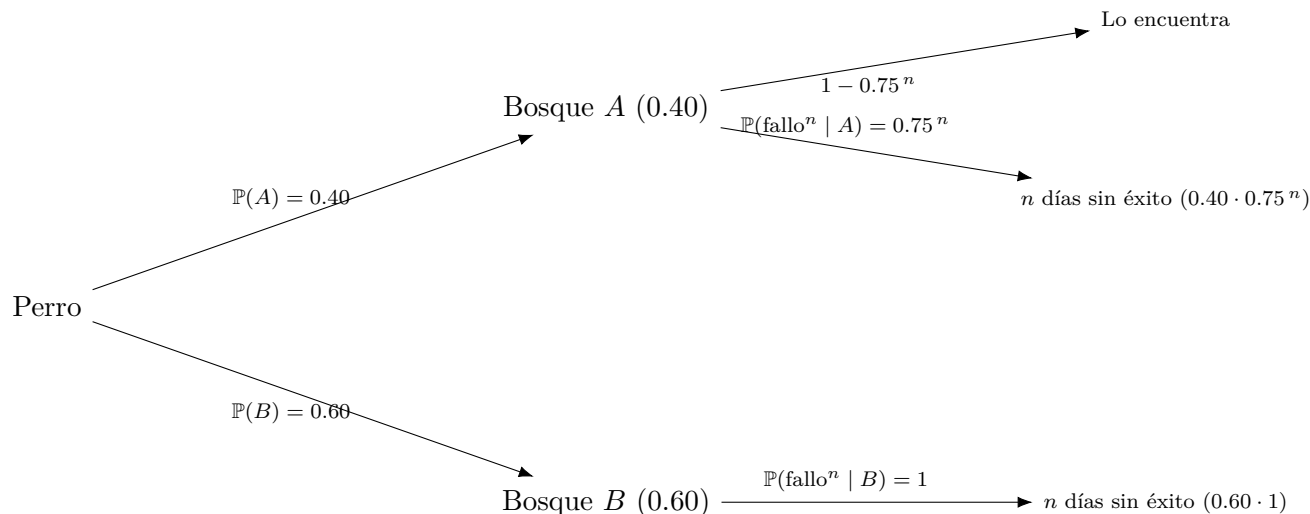
(c) Segundo positivo, usando 0.0905 como nuevo prior:

$$\mathbb{P}(\text{enf.} \mid +, +) = \frac{0.99 \times 0.0905}{0.99 \times 0.0905 + 0.05 \times 0.9095} = \frac{0.08960}{0.08960 + 0.04548} \approx 0.6633.$$

Un *segundo* positivo independiente eleva la creencia del 9 por ciento al 66 por ciento: por eso los tamizajes masivos confirman los positivos con una prueba adicional antes de tratar. (Equivalente: usar las apuestas. Prior odds 500:99,500; cada positivo multiplica por el factor de Bayes $0.99/0.05 = 19.8$.)

Ejercicio 1.35 (El perro de Oscar, versión extendida (varios días)). El perro de Oscar se perdió: está en el bosque A con probabilidad 0.40 o en el bosque B con probabilidad 0.60. Oscar decide buscar *cada día en el bosque A*. Si el perro está en A , cada búsqueda diaria lo encuentra con probabilidad 0.25, de forma independiente día a día; si está en B , nunca lo encuentra buscando en A .

- (a) Tras *un* día sin éxito en A , ¿cuál es la probabilidad de que el perro esté en A ?
- (b) Tras *tres* días sin éxito en A , ¿cuál es esa probabilidad?
- (c) Halla la fórmula general $\mathbb{P}(A \mid n \text{ días sin éxito})$ y describe su límite cuando $n \rightarrow \infty$.



Cada rama es una probabilidad condicional: la probabilidad de un camino es el producto de sus ramas; “ n días sin éxito en A ” tiene probabilidad 0.75^n si el perro está en A y probabilidad 1 si está en B .

Solución

Sea F_n = “ n días buscando en A sin encontrarlo”. Como cada día es independiente, $\mathbb{P}(F_n | A) = 0.75^n$ y $\mathbb{P}(F_n | B) = 1$ (buscando en A nunca se encuentra a un perro que está en B).

(a) Con $n = 1$:

$$\mathbb{P}(A | F_1) = \frac{0.40 \times 0.75}{0.40 \times 0.75 + 0.60 \times 1} = \frac{0.30}{0.30 + 0.60} = \frac{0.30}{0.90} = \frac{1}{3} \approx 0.3333.$$

(b) Con $n = 3$, $0.75^3 = 0.421875$:

$$\mathbb{P}(A | F_3) = \frac{0.40 \times 0.421875}{0.40 \times 0.421875 + 0.60 \times 1} = \frac{0.16875}{0.16875 + 0.60} = \frac{27}{123} = \frac{9}{41} \approx 0.2195.$$

(c) Fórmula general:

$$\mathbb{P}(A | F_n) = \frac{0.40 \cdot 0.75^n}{0.40 \cdot 0.75^n + 0.60} = \frac{2 \cdot 0.75^n}{2 \cdot 0.75^n + 3}.$$

Como $0.75^n \rightarrow 0$, el límite es 0: mientras más días busca en A sin encontrarlo, más se convence de que el perro nunca estuvo allí. Cada fracaso multiplica las apuestas de “estar en A ” por el factor $0.75 < 1$.

Nota (la fórmula clásica $N/(N+2)$). En una versión famosa de este problema la probabilidad de detección diaria no es fija sino *desconocida*, modelada como una variable θ con densidad $f(\theta) = 2\theta$ en $[0, 1]$ (los perros más fáciles de hallar pesan más). Si interpretamos “no hallado” como “el perro pudo perderse para siempre”, la probabilidad de seguir sin éxito tras N días resulta $\int_0^1 (1-\theta) 2\theta d\theta$ generalizado a $\int_0^1 \theta^N 2\theta d\theta = \frac{2}{N+2}$, de

modo que la probabilidad de un desenlace fatal (no recuperación) es $1 - \frac{2}{N+2} = \frac{N}{N+2}$. Para $N = 1, 2, 3$ da $\frac{1}{3}, \frac{1}{2}, \frac{3}{5}$: crece hacia 1 con el tiempo. (Este cálculo usa integración, fuera del alcance de la Clase 2; se incluye solo como cultura general.) (*adaptado de MIT 18.05, Class 3, y del clásico “Oscar’s lost dog”*)

Ejercicio 1.36 (Paradoja de Simpson con datos concretos). Se comparan dos modalidades de un programa de capacitación laboral, A (intensiva) y B (estándar), según su tasa de inserción (insertados / capacitados). Los datos, desglosados por tipo de participante (urbano = perfil “fácil”, rural = perfil “difícil”), son:

	Urbano (fácil)	Rural (difícil)	Total
Modalidad A	81/87 = 0.931	192/263 = 0.730	273/350 = 0.780
Modalidad B	234/270 = 0.867	55/80 = 0.688	289/350 = 0.826

- (a) ¿Qué modalidad inserta mejor en cada subgrupo por separado?
- (b) ¿Qué modalidad tiene mayor tasa *global*? Explica la aparente contradicción identificando la variable de confusión.

(*adaptado de MIT 18.05, Class 3 — conditional probability; estructura clásica de Simpson*)

Solución

(a) En *ambos* subgrupos gana la modalidad A: urbano $0.931 > 0.867$ y rural $0.730 > 0.688$.

(b) Globalmente parece mejor la modalidad B: $0.826 > 0.780$. ¡Se invierte la conclusión! Esta es la **paradoja de Simpson**. La causa es el *tipo de participante*, que actúa como **variable de confusión**:

- La modalidad A se aplicó sobre todo al grupo *rural* (difícil): 263 de sus 350 casos, donde las tasas son bajas para todos.
- La modalidad B se aplicó sobre todo al grupo *urbano* (fácil): 270 de sus 350 casos, donde las tasas son altas para todos.

El promedio global de A queda “arrastrado” por los muchos casos difíciles que atendió, y el de B “inflado” por los muchos casos fáciles. La comparación *justa* es la condicional (dentro de cada tipo), no la marginal: hay que condicionar por el tipo de participante antes de comparar.

Moraleja. $\mathbb{P}(\text{inserta} | A, \text{tipo}) > \mathbb{P}(\text{inserta} | B, \text{tipo})$ para cada tipo, y aún así $\mathbb{P}(\text{inserta} | A) < \mathbb{P}(\text{inserta} | B)$. Las proporciones no se pueden “sumar” ignorando los tamaños de los subgrupos.

Ejercicio 1.37 (Monty Hall en clave social: tres sobres). Un funcionario te ofrece tres sobres cerrados; uno contiene una beca y dos están vacíos. Eliges el sobre 1 (sin abrirlo). El

funcionario, que *sabe* dónde está la beca, abre uno de los otros dos sobres que *sabe* que está vacío (el sobre 3) y te ofrece cambiar tu elección al sobre 2. ¿Te conviene cambiar? Justifica con probabilidades.

Solución

Al inicio, $\mathbb{P}(\text{beca en } 1) = 1/3$ y $\mathbb{P}(\text{beca en } 2 \text{ o } 3) = 2/3$. La acción del funcionario (abrir un vacío entre 2 y 3) no cambia que tu sobre 1 tenía $1/3$, pero *concentra* todo el $2/3$ restante en el sobre 2 (porque el 3 ya se sabe vacío). Por tanto:

$$\mathbb{P}(\text{beca en } 2 \mid \text{abrió el } 3 \text{ vacío}) = \frac{2}{3}, \quad \mathbb{P}(\text{beca en } 1 \mid \cdot) = \frac{1}{3}.$$

Sí conviene cambiar: cambiar gana con probabilidad $2/3$. (Formalmente, con $H_i =$ “beca en i ” y $D =$ “abre el 3”: $\mathbb{P}(D \mid H_1) = 1/2$, $\mathbb{P}(D \mid H_2) = 1$, $\mathbb{P}(D \mid H_3) = 0$; Bayes da $\mathbb{P}(H_2 \mid D) = \frac{1 \cdot 1/3}{1/2 \cdot 1/3 + 1 \cdot 1/3} = \frac{1/3}{1/2} = 2/3$.)

Nivel 5 – Reto+

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

Ejercicio 1.38 (Reto+: independencia por pares pero no mutua). Se lanzan dos dados justos. Define los eventos

$$A_1 = \{\text{el primer dado es par}\}, \quad A_2 = \{\text{el segundo dado es impar}\}, \quad A_3 = \{\text{la suma es impar}\}.$$

Demuestra que A_1, A_2, A_3 son independientes *por pares* pero *no* mutuamente independientes. (Moraleja: la independencia mutua es estrictamente más fuerte que la independencia por pares.)

Solución

El espacio tiene 36 resultados equiprobables. $\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \frac{1}{2}$ (en cada caso la mitad de los 36 casos cumple).

Por pares.

- $A_1 \cap A_2$: primer dado par y segundo impar $\Rightarrow 3 \times 3 = 9$ casos, $\mathbb{P} = 9/36 = 1/4 = \frac{1}{2} \cdot \frac{1}{2}$. ✓
- $A_1 \cap A_3$: primer dado par y suma impar $\Rightarrow$ segundo dado impar $\Rightarrow 3 \times 3 = 9$ casos, $\mathbb{P} = 1/4$. ✓
- $A_2 \cap A_3$: segundo dado impar y suma impar $\Rightarrow$ primer dado par $\Rightarrow 3 \times 3 = 9$ casos, $\mathbb{P} = 1/4$. ✓

Las tres parejas factorizan: son independientes por pares.

Mutua. Si los tres ocurren: primer dado par y segundo impar, lo que ya *fuera* suma =

par+impar= impar; así $A_1 \cap A_2 \cap A_3 = A_1 \cap A_2$, con $\mathbb{P}(A_1 \cap A_2 \cap A_3) = 1/4$. Pero

$$\mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3) = \left(\frac{1}{2}\right)^3 = \frac{1}{8} \neq \frac{1}{4}.$$

No se cumple la factorización triple, luego **no** son mutuamente independientes. (*adaptado de gevorg-hunanyan/probability, pairwise-not-mutual*)

Ejercicio 1.39 (Reto+: independencia marginal sin independencia condicional). Se lanzan dos monedas justas. Define

$A = \{\text{la primera moneda es cara}\}, \quad B = \{\text{la segunda moneda es cara}\}, \quad C = \{\text{ambas monedas salen iguales}\}$

- Muestra que A y B son *independientes marginalmente*.
- Muestra que A y B **no** son independientes *condicionalmente dado* C . Da una explicación intuitiva.

(*adaptado de MIT 18.05, Class 3 — independence*)

Solución

Espacio muestral equiprobable: $\{HH, HT, TH, TT\}$, cada uno con probabilidad $1/4$.

(a) Independencia marginal. $\mathbb{P}(A) = \mathbb{P}(\{HH, HT\}) = \frac{1}{2}$, $\mathbb{P}(B) = \mathbb{P}(\{HH, TH\}) = \frac{1}{2}$, $\mathbb{P}(A \cap B) = \mathbb{P}(\{HH\}) = \frac{1}{4}$. Como $\mathbb{P}(A)\mathbb{P}(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = \mathbb{P}(A \cap B)$, **sí** son independientes marginalmente.

(b) No independencia condicional dado C . $C = \{HH, TT\}$ con $\mathbb{P}(C) = \frac{1}{2}$. Dentro de C ambos resultados son igualmente probables, así:

$$\mathbb{P}(A|C) = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} = \frac{\mathbb{P}(\{HH\})}{1/2} = \frac{1/4}{1/2} = \frac{1}{2}, \quad \mathbb{P}(B|C) = \frac{1}{2},$$

pero

$$\mathbb{P}(A \cap B|C) = \frac{\mathbb{P}(\{HH\})}{\mathbb{P}(C)} = \frac{1/4}{1/2} = \frac{1}{2} \neq \mathbb{P}(A|C)\mathbb{P}(B|C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Luego A y B **no** son condicionalmente independientes dado C .

Intuición. Saber que las dos monedas son *iguales* crea un vínculo: una vez que sabemos C , conocer la primera moneda determina por completo la segunda ($\mathbb{P}(B|A \cap C) = 1$, frente a $\mathbb{P}(B|C) = 1/2$). Acondicionar a una causa o restricción común puede *crear* dependencia entre eventos que eran independientes. (El fenómeno opuesto — dependencia marginal que desaparece al condicionar — es el de “controlar por una variable de confusión”.)

Ejercicio 1.40 (Reto+: el problema de los dos hijos con condicionamiento sutil). Una funcionaria del registro civil revisa familias con *exactamente dos hijos*. Supone que cada hijo es niño o niña con probabilidad $1/2$, de forma independiente, y que cada hijo nació en uno de los 7 días de la semana con igual probabilidad e independiente del sexo.

- (a) Sabiendo que la familia tiene *al menos un niño*, ¿cuál es la probabilidad de que *ambos* hijos sean niños?
- (b) Sabiendo que el hijo *mayor* es niño, ¿cuál es esa probabilidad? Compara con (a).
- (c) Sabiendo que la familia tiene *al menos un niño nacido un martes*, ¿cuál es la probabilidad de que ambos sean niños? (Resultado contraintuitivo.)

(adaptado de MIT 18.05, Class 3 — conditional probability; “the two children problem”)

Solución

(a) Espacio de sexos por orden de nacimiento: $\{BB, BG, GB, GG\}$, equiprobable. “Al menos un niño” descarta GG , dejando $\{BB, BG, GB\}$ (tres casos igualmente probables). “Ambos niños” es BB :

$$\mathbb{P}(BB \mid \text{al menos un niño}) = \frac{1}{3} \approx 0.3333.$$

(b) “El *mayor* es niño” fija el primer hijo: el espacio se reduce a $\{BB, BG\}$ (dos casos). “Ambos niños” es BB :

$$\mathbb{P}(BB \mid \text{mayor es niño}) = \frac{1}{2} = 0.5.$$

La diferencia con (a) es la clave: “al menos uno” *no* señala a un hijo en particular, mientras que “el mayor” sí. Identificar a un hijo concreto cambia el espacio muestral condicional y, con ello, la respuesta.

(c) Ahora cada hijo tiene un *atributo doble* (sexo, día): hay $2 \times 7 = 14$ tipos de hijo igualmente probables, y $14 \times 14 = 196$ familias equiprobables. Sea M = “al menos un niño nacido en martes”. Contamos las familias en M con el principio de inclusión–exclusión. “Hijo es niño–martes” tiene 1 de las 14 posibilidades.

- Familias con el *primer* hijo niño–martes: $1 \times 14 = 14$.
- Familias con el *segundo* hijo niño–martes: $14 \times 1 = 14$.
- Restamos la doble cuenta (ambos niño–martes): $1 \times 1 = 1$.

Entonces $|M| = 14 + 14 - 1 = 27$. De esas, las que tienen *ambos niños*: al menos uno es niño–martes y los dos son niños. El niño–martes puede ser el primero (el otro es cualquiera de los 7 días, niño): $1 \times 7 = 7$; o el segundo: $7 \times 1 = 7$; menos la doble cuenta (ambos niño–martes): 1. Total $7 + 7 - 1 = 13$. Por tanto:

$$\mathbb{P}(\text{ambos niños} \mid M) = \frac{13}{27} \approx 0.4815.$$

Lo sorprendente: añadir el dato aparentemente irrelevante “nació un martes” mueve la respuesta de $1/3$ (parte a) a $13/27 \approx 0.48$, acercándola a $1/2$. El motivo es sutil: especificar el día hace que el evento condicionante se parezca más a “señalar a un hijo concreto” (como en la parte b) que a “al menos uno”. La lección: al condicionar, la forma exacta en que se describe la información cambia el espacio muestral y la respuesta.

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