Probabilidad condicional, Bayes e independencia

Clase 3: *conditional probability*, *Bayes' rule* e *independence*

6.431x — Probability: The Science of Uncertainty and Data

B.Sc. Cristian Lazo Quispe

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

Programa SDS (Statistics and Data Science, MITx) · **BREIT** (Brescia Institute of Technology), iniciativa de Aporta —Grupo Breca— en alianza con MIT IDSS.
Basado en MIT 6.431x / MIT OCW RES.6-012 (CC BY-NC-SA 4.0).

clazoq@uni.pe · linkedin.com/in/cristian-lazo-quispe · github.com/CristianLazoQuispe

Dónde quedamos (Clases 1 y 2)

La Unit 1, en una línea

Un **modelo de probabilidad** = un espacio muestral Ω + una ley $\mathbb{P}(\cdot)$ que cumple los tres axiomas. De ahí dedujimos todo lo demás.

Lo que ya tenemos

- Los tres axiomas: $\mathbb{P}(A) \geq 0$, $\mathbb{P}(\Omega) = 1$ y aditividad para eventos **disjuntos**.
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, monotonía, inclusión–exclusión, cota de Boole.
- Modelos **discretos** (contar masa) y **continuos** (medir área).

La idea que resume toda la clase

Condicionar es encoger el universo. Si ocurrió B, lo que está fuera de B deja de importar y la masa se reparte —en la misma proporción— dentro de B. Nada más.

¿Qué veremos hoy? (Unit 2 — Lectures 2 y 3)

1. Conditioning: el universo se encoge
2. Multiplication rule y árboles
3. Total probability theorem y Bayes' rule
4. El false-positive puzzle
5. Independence
6. Reliability: series y parallel
7. Cierre

Contenido

1. Conditioning: el universo se encoge
2. Multiplication rule y árboles
3. Total probability theorem y Bayes' rule
4. El false-positive puzzle
5. Independence
6. Reliability: series y parallel
7. Cierre

Conditional probability: la definición y por qué es esa

Definición

Si $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

El numerador es **lo que sobrevive** de A dentro del nuevo universo; el denominador **renormaliza** para que $\mathbb{P}(B \mid B) = 1$.

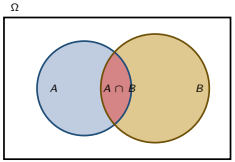
La consecuencia práctica

$\mathbb{P}(\cdot \mid B)$ **es** una ley de probabilidad: cumple los tres axiomas. Así que *todo* lo de las Clases 1 y 2 vale con la barra puesta:

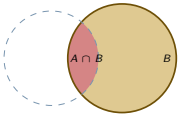
$$\mathbb{P}(A^c \mid B) = 1 - \mathbb{P}(A \mid B),$$

$$\mathbb{P}(A \cup C \mid B) = \mathbb{P}(A \mid B) + \mathbb{P}(C \mid B) - \mathbb{P}(A \cap C \mid B).$$

No hay reglas nuevas que memorizar.



modelo original



ocurrió B: el universo es B

Demo: dos tiradas de un dado de 4 caras

Problema

Se tira dos veces un dado justo de 4 caras; los 16 pares son igualmente probables. Sea $M = \max(X, Y)$ y sea $B = \{\min(X, Y) = 2\}$. Calcula $\mathbb{P}(M = 1 \mid B)$ y $\mathbb{P}(M = 3 \mid B)$.

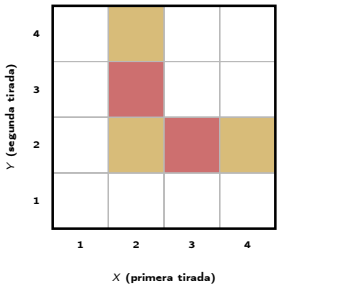
Solución

B tiene 5 celdas: $(2,2), (2,3), (2,4), (3,2), (4,2)$, así que $\mathbb{P}(B) = 5/16$.

- $\mathbb{P}(M = 1 \mid B) = 0$: si el mínimo es 2, el máximo no puede ser 1.
- $M = 3$ ocurre en B solo con $(2,3)$ y $(3,2)$:

$$\mathbb{P}(M = 3 \mid B) = \frac{2/16}{5/16} = \frac{2}{5} = 0.4.$$

El 1/16 se cancela: dentro de B basta **contar celdas**.



Contenido

1. Conditioning: el universo se encoge
2. Multiplication rule y árboles
3. Total probability theorem y Bayes' rule
4. El false-positive puzzle
5. Independence
6. Reliability: series y parallel
7. Cierre

Demo: tres extracciones sin reemplazo

Problema

Despejando la definición sale la ***multiplication rule***:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B \mid A),$$

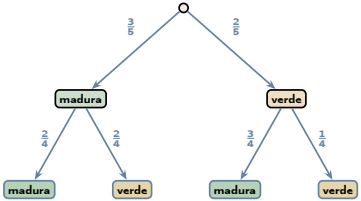
$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \prod_{i=2}^{n} \mathbb{P}(A_i \mid A_1 \cap \cdots \cap A_{i-1}).$$

Una caja tiene 3 paltas maduras y 2 verdes. Se sacan 3 sin reemplazo: ¿probabilidad de que las tres sean maduras?

Solución

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} = \frac{6}{60} = \tfrac{1}{10} = 0.1.$$

Cada denominador baja porque quedan menos paltas; cada numerador, porque ya se sacaron maduras. **Eso es exactamente lo que codifica el condicionamiento.**



La hoja es el **producto del camino**; las ramas de un mismo nodo suman 1.

Contenido

1. Conditioning: el universo se encoge
2. Multiplication rule y árboles
3. Total probability theorem y Bayes' rule
4. El false-positive puzzle
5. Independence
6. Reliability: series y parallel
7. Cierre

Las dos herramientas que sostienen el resto del curso

Total probability theorem

Si $A_1, \ldots, A_n$ es una **partición** de Ω:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i)\,\mathbb{P}(B \mid A_i).$$

Es un **promedio ponderado** con pesos que suman $1 \Rightarrow \mathbb{P}(B)$ siempre cae **entre el menor y el mayor** de los $\mathbb{P}(B \mid A_i)$. Chequeo gratis.

Bayes' rule

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(A_i)\,\mathbb{P}(B \mid A_i)}{\sum_j \mathbb{P}(A_j)\,\mathbb{P}(B \mid A_j)}.$$

El vocabulario de todo el resto del curso

- $\mathbb{P}(A_i)$ — el ***prior***: lo que creíamos antes.
- $\mathbb{P}(B \mid A_i)$ — la ***likelihood***: qué tan compatible es lo observado con la causa A_i.
- $\mathbb{P}(A_i \mid B)$ — el ***posterior***: la creencia revisada.
- $\mathbb{P}(B)$ — solo la constante que hace que los *posteriors* sumen 1.

Por qué importa tanto

Esta es la estructura completa de la **inferencia bayesiana** (Unit 7), donde caen **3 de los 6 problemas del examen final**. Todo lo que viene después es esta misma fórmula con variables aleatorias en lugar de eventos.

Demo: *radar detection* (el ejemplo canónico de MIT)

Problema

A = "hay un avión", B = "el radar registra algo".

$$\mathbb{P}(A) = 0.05, \quad \mathbb{P}(B \mid A) = 0.99, \quad \mathbb{P}(B \mid A^c) = 0.10.$$

El radar se activó. ¿Cuál es la probabilidad de que haya un avión?

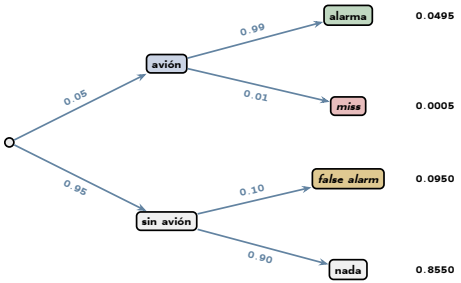
Solución

$$\mathbb{P}(B) = 0.05 \cdot 0.99 + 0.95 \cdot 0.10 = 0.0495 + 0.095 = 0.1445$$

$$\mathbb{P}(A \mid B) = \frac{0.0495}{0.1445} = \frac{99}{289} \approx 0.34$$

La lectura

El radar acierta el **99 por ciento** de las veces y aun así, tras una alarma, solo hay un **34 por ciento** de probabilidad de que haya avión: los aviones son raros y el cielo vacío es constante, así que hay muchas más oportunidades de dar una *false alarm*.



Las dos hojas con alarma suman 0.1445; de ellas, solo 0.0495 tienen avión.

Contenido

1. Conditioning: el universo se encoge
2. Multiplication rule y árboles
3. Total probability theorem y Bayes' rule
4. El false-positive puzzle
5. Independence
6. Reliability: series y parallel
7. Cierre

Demo: un test del 99 por ciento que deja la duda en 50–50

Problema

Una enfermedad afecta al **1 por ciento** de la población. Un test acierta el **99 por ciento** de las veces en ambos sentidos. Alguien da positivo. ¿Probabilidad de estar enfermo?

Solución

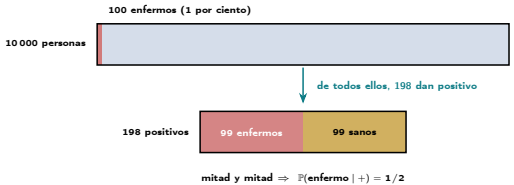
$$\mathbb{P}(+) = 0.01 \cdot 0.99 + 0.99 \cdot 0.01 = 0.0099 + 0.0099 = 0.0198$$

$$\mathbb{P}(E \mid +) = \frac{0.0099}{0.0198} = \tfrac{1}{2}$$

No es casualidad: hay **tantos enfermos que dan positivo como sanos que dan positivo**.

Base rate fallacy — el error más caro de la clase

Responder "99 por ciento" confunde $\mathbb{P}(+ \mid E)$ con $\mathbb{P}(E \mid +)$. Falta el ***prior***: cuánta gente está enferma *antes* del test. En tribunales, la misma confusión se llama *prosecutor's fallacy*.



Con prevalencia 1/1000 y test al 95 por ciento, el mismo cálculo da ≈ 0.019: menos del 2 por ciento.

Contenido

1. Conditioning: el universo se encoge
2. Multiplication rule y árboles
3. Total probability theorem y Bayes' rule
4. El false-positive puzzle
5. Independence
6. Reliability: series y parallel
7. Cierre

Independence: la definición y la trampa de siempre

Definición

$$A \perp\!\!\!\perp B \iff \mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B).$$

Si $\mathbb{P}(B) > 0$, equivale a $\mathbb{P}(A \mid B) = \mathbb{P}(A)$: saber B **no cambia nada**. La forma del producto es la definición porque es **simétrica** y funciona aunque $\mathbb{P}(B) = 0$.

Se hereda al complemento

Si $A \perp\!\!\!\perp B$ entonces $A \perp\!\!\!\perp B^c$:

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B^c).$$

Lógico: si B no informa sobre A, B^c tampoco.

Disjuntos $\neq$ independientes

Si A y B son **disjuntos** con probabilidad positiva:

$$\mathbb{P}(A \cap B) = 0 \quad \text{pero} \quad \mathbb{P}(A)\mathbb{P}(B) > 0.$$

Son **dependientes**, y de la forma más fuerte: si ocurre B, A queda descartado. Disjunto habla de la **geometría** de los conjuntos; independiente, de la **información**.

Demo: un dado justo de 6 caras

$A = \{\text{par}\}$, $B = \{5, 6\}$: **no** son disjuntos ($A \cap B = \{6\}$) y **sí** son independientes: $\mathbb{P}(A \cap B) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3}$. ✓ $C = \{1, 2\}$, $D = \{3, 4\}$: **sí** son disjuntos y por eso mismo **no** son independientes ($0 \neq \frac{1}{9}$).

Demo: *pairwise* no implica *mutual*

Problema

Se lanza dos veces una moneda justa. Sean

$$H_1 = \{1.^{\text{a}} \text{ cara}\}, \quad H_2 = \{2.^{\text{a}} \text{ cara}\},$$

$$C = \{\text{las dos salen igual}\} = \{HH, TT\}.$$

¿Son los tres independientes?

Solución

Los tres tienen probabilidad 1/2. **Por pares, sí:**

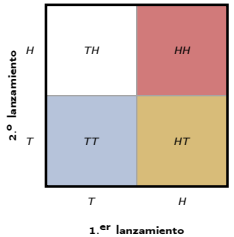
$$\mathbb{P}(H_1 \cap C) = \mathbb{P}(\{HH\}) = \tfrac{1}{4} = \tfrac{1}{2} \cdot \tfrac{1}{2} \checkmark$$

(y lo mismo para H_2, C y para H_1, H_2). **Pero los tres juntos, no:**

$$\mathbb{P}(H_1 \cap H_2 \cap C) = \tfrac{1}{4} \neq \tfrac{1}{8} = \mathbb{P}(H_1)\mathbb{P}(H_2)\mathbb{P}(C).$$

Por qué

H_1 sola no dice nada de C; pero H_1 *y* H_2 juntas **determinan** C: $\mathbb{P}(C \mid H_1 \cap H_2) = 1$. Para n eventos hacen falta **todas** las condiciones, no solo las de a pares.



■ y ■ forman C (las dos iguales).
H_1 = la columna derecha = ■+■.
$H_1 \cap C = \{HH\}$: una celda de cuatro.

Demo: condicionar puede *romper* la independencia

Problema

Hay dos monedas: la A sale cara con probabilidad 0.9, la B con probabilidad 0.1. Se elige una al azar y se lanza muchas veces. ¿Son independientes los lanzamientos?

Solución

Cada lanzamiento por separado:

$$\mathbb{P}(\text{cara}) = \tfrac{1}{2}(0.9) + \tfrac{1}{2}(0.1) = \tfrac{1}{2}.$$

Pero si los **10 primeros** salieron cara, eso es 9^{10} veces más probable con la moneda A, así que $\mathbb{P}(A \mid 10 \text{ caras}) \approx 1$ y

$$\mathbb{P}(\text{cara}_{11} \mid 10 \text{ caras}) \approx 0.9 \neq \tfrac{1}{2}.$$

No son independientes.

Conditional independence

$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\,\mathbb{P}(B \mid C)$$

Los lanzamientos **sí** son independientes *dada la moneda*. Están correlacionados justamente porque comparten una **causa oculta**.

Las dos implicaciones son falsas

- $A \perp\!\!\!\perp B$ **no** implica independencia condicional dado C.
- Independencia dado C **no** implica $A \perp\!\!\!\perp B$, ni independencia dado C^c.

Son propiedades **distintas**. Este es el mecanismo que explica por qué repetir una medición solo informa si los errores son independientes.

Contenido

1. Conditioning: el universo se encoge
2. Multiplication rule y árboles
3. Total probability theorem y Bayes' rule
4. El false-positive puzzle
5. Independence
6. Reliability: series y parallel
7. Cierre

Demo: tres componentes, dos topologías

Problema

Tres componentes funcionan de forma independiente, cada uno con probabilidad $p = 0.9$. Calcula la fiabilidad del sistema en *series* y en *parallel*.

Solución

Series — deben funcionar **todos**, así que se multiplica:

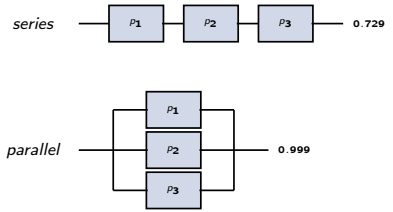
$$\mathbb{P} = p^3 = 0.9^3 = 0.729.$$

Parallel — basta **uno**; conviene ir por el **complemento** (que fallen todos):

$$\mathbb{P} = 1 - (1-p)^3 = 1 - 0.1^3 = 0.999.$$

La lectura

Los **mismos** tres componentes dan 72.9 por ciento en serie y 99.9 por ciento en paralelo. **La topología importa más que la calidad de las piezas**, y redundancia es, literalmente, poner cosas en paralelo.



En serie el sistema es peor que su peor componente; en paralelo, mejor que el mejor.

Contenido

1. Conditioning: el universo se encoge
2. Multiplication rule y árboles
3. Total probability theorem y Bayes' rule
4. El false-positive puzzle
5. Independence
6. Reliability: series y parallel
7. Cierre

Kahoot de repaso (Clase 3)

Jugamos ahora

30 preguntas sobre **toda la Unit 2**: *conditional probability*, *multiplication rule*, *total probability theorem*, *Bayes' rule*, *independence*, *conditional independence* y *reliability*. Las **15 primeras son de calentamiento** (30 s cada una); el resto, de 1 a 1,5 min.

- Cada pregunta tiene una **trampa**: el error clásico está entre las opciones.
- Antes de responder, decide **si toca sumar o multiplicar** —y por qué—.
- Si el enunciado invierte el condicional, pregúntate: ¿me están pidiendo $\mathbb{P}(A \mid B)$ o $\mathbb{P}(B \mid A)$?
- En los problemas con etapas, **piensa el árbol** aunque no te dé tiempo a dibujarlo.

Resumen de la Clase 3

Lo que hay que saber hacer

1. Calcular $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$ y usar que la ley condicional **cumple los tres axiomas** (todo lo de las Clases 1 y 2 vale con la barra).
2. Montar un **árbol** y aplicar la *multiplication rule*: la hoja es el producto del camino.
3. Reconocer una **partición** y aplicar el *total probability theorem* como promedio ponderado (y usarlo de chequeo).
4. Invertir un condicional con *Bayes' rule*, nombrando *prior*, *likelihood* y *posterior*.
5. Distinguir **disjuntos** de **independientes**, y saber que la independencia **no** sobrevive necesariamente al condicionamiento.
6. Verificar las **cuatro** condiciones de independencia para tres eventos, no solo las de a pares.
7. Calcular *reliability* en *series* (producto) y en *parallel* (complemento).

Siguiente clase (Clase 4)

Unit 3 — *Counting* y Unit 4 — *Discrete random variables*: primero técnicas para contar Ω sin enumerarlo, y luego el salto conceptual grande del curso, de **eventos** a **variables aleatorias** (PMFs y esperanza).

¡Gracias!

Clase 3 — Probabilidad condicional, Bayes e independencia (6.431x, Unit 2)

clazoq@uni.pe · linkedin.com/in/cristian-lazo-quispe · github.com/CristianLazoQuispe

BREIT (Brescia Institute of Technology) · Aporta — Grupo Breca · en alianza con MIT IDSS

Basado en MIT 6.431x / MIT OCW RES.6-012 (Bertsekas & Tsitsiklis), CC BY-NC-SA 4.0.