

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

Teoría

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

Índice

1. De qué trata esta clase	2
2. Probabilidad condicional	2
3. Independencia	2
4. Ley de probabilidad total	3
5. Regla de Bayes	3
6. Diagramas de árbol	4
7. Conjuntos contables y no contables	4
8. En R: simular probabilidad condicional y Bayes	5
Resumen de la clase	5

1. De qué trata esta clase

La Clase 1 nos dio el lenguaje del azar (espacio muestral, eventos, axiomas, conteo). Hoy respondemos la pregunta que mueve casi toda la inferencia: **¿cómo cambia la probabilidad de algo cuando recibimos información nueva?** Esa es la **probabilidad condicional (conditional probability)**.

De ahí salen tres herramientas que usaremos todo el curso: la **independencia (independence)**, la **ley de probabilidad total (law of total probability)** y la **regla de Bayes (Bayes' rule)**, que es la base de cómo actualizamos creencias con datos. Cerramos con una nota sobre **conjuntos contables y no contables (countable / uncountable sets)**, que distingue las variables discretas de las continuas (Clase 3).

Condicionar es *redefinir el mundo*: al saber que ocurrió B , nuestro nuevo espacio muestral ya no es S sino B . La probabilidad condicional simplemente mide qué fracción de ese mundo más pequeño corresponde a A .

2. Probabilidad condicional

Definición 2.1 (Probabilidad condicional). La probabilidad de A **dado** B (conditional on B), con $\mathbb{P}(B) > 0$, es

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

Reordenando se obtiene la **regla del producto (multiplication rule)**:

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

Ejemplo 2.1 (Reescalando el espacio). En una encuesta, $\mathbb{P}(\text{tiene empleo formal y seguro}) = 0.35$ y $\mathbb{P}(\text{tiene empleo formal}) = 0.50$. Entre los que tienen empleo formal, ¿qué fracción tiene seguro?

$$\mathbb{P}(\text{seguro} | \text{formal}) = \frac{0.35}{0.50} = 0.70.$$

Pasamos de mirar toda la población a mirar solo el “mundo” de los formales.

3. Independencia

Definición 3.1 (Independencia). Dos eventos A y B son **independientes (independent)** si

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

Equivalentemente, si $\mathbb{P}(B) > 0$, son independientes si y solo si $\mathbb{P}(A | B) = \mathbb{P}(A)$: saber que ocurrió B no cambia la probabilidad de A .

Son ideas *opuestas*, no sinónimos. Si A y B son disjuntos y ambos tienen probabilidad positiva, entonces saber que ocurrió B *garantiza* que A no ocurrió: ¡eso es máxima dependencia! Disjuntos con prob. positiva nunca son independientes.

Ejemplo 3.1 (Verificar independencia). Se lanza un dado justo. Sea A = “sale menor que 5” y B = “sale par”. $\mathbb{P}(A) = 4/6 = 2/3$, $\mathbb{P}(B) = 3/6 = 1/2$. La intersección $A \cap B = \{2, 4\}$, así $\mathbb{P}(A \cap B) = 2/6 = 1/3$. Como $\mathbb{P}(A)\mathbb{P}(B) = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}$ coincide, A y B *son* independientes (aunque dependan del mismo lanzamiento).

4. Ley de probabilidad total

Ley de probabilidad total

Si $B_1, B_2, \dots, B_k$ *particionan* el espacio (disjuntos y exhaustivos, cada uno con prob. positiva), entonces para cualquier evento A :

$$\mathbb{P}(A) = \sum_{i=1}^k \mathbb{P}(A | B_i) \mathbb{P}(B_i).$$

Para hallar $\mathbb{P}(A)$ partimos el mundo en escenarios B_i que no se solapan, calculamos A dentro de cada escenario y promediamos *ponderando* por cuán probable es cada escenario. Es “divide y vencerás” para probabilidades.

Ejemplo 4.1 (Dos urnas). Un programa entrega bonos desde dos oficinas. La oficina 1 procesa el 60% de los casos y comete error en el 5%; la oficina 2 procesa el 40% y se equivoca en el 10%. ¿Probabilidad de que un caso al azar tenga error?

$$\mathbb{P}(\text{error}) = 0.05 \cdot 0.60 + 0.10 \cdot 0.40 = 0.03 + 0.04 = 0.07.$$

5. Regla de Bayes

Regla de Bayes

Para eventos con probabilidad positiva,

$$\mathbb{P}(B | A) = \frac{\mathbb{P}(A | B) \mathbb{P}(B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A | B) \mathbb{P}(B)}{\sum_i \mathbb{P}(A | B_i) \mathbb{P}(B_i)},$$

donde el denominador usa la ley de probabilidad total sobre una partición $\{B_i\}$.

Bayes: con partición y con evidencia adicional

Las dos formas que más usarás, con una partición $\{A_1, A_2, \dots\}$ del espacio:

$$\mathbb{P}(A_i | B) = \frac{\mathbb{P}(A_i) \mathbb{P}(B | A_i)}{\sum_j \mathbb{P}(A_j) \mathbb{P}(B | A_j)}, \quad \mathbb{P}(A | B \cap C) = \frac{\mathbb{P}(B | A \cap C) \mathbb{P}(A | C)}{\mathbb{P}(B | C)}.$$

La primera es la regla de Bayes estándar (la causa A_i dado el dato B). La segunda es *el mismo* Bayes pero **dentro del mundo donde ya sabemos C** : todo se condiciona además a C .

Bayes *invierte* el condicionamiento: convierte “probabilidad del síntoma dado el estado” en “probabilidad del estado dado el síntoma”. Es exactamente lo que hacemos al inferir causas a partir de datos observados: partimos de una creencia previa (*prior*) $\mathbb{P}(B)$ y la actualizamos a una creencia posterior (*posterior*) $\mathbb{P}(B | A)$ tras observar A .

Ejemplo 5.1 (Prueba de tamizaje). Una enfermedad afecta al 1 % de la población. Un test detecta correctamente al enfermo el 90 % de las veces (sensibilidad) y da falso positivo el 8 % en sanos. Si una persona da positivo, ¿cuál es la probabilidad de que esté realmente enferma? Sea E enfermo, $+$ positivo. Por Bayes:

$$\mathbb{P}(E | +) = \frac{\mathbb{P}(+ | E) \mathbb{P}(E)}{\mathbb{P}(+ | E) \mathbb{P}(E) + \mathbb{P}(+ | E^c) \mathbb{P}(E^c)} = \frac{0.90 \cdot 0.01}{0.90 \cdot 0.01 + 0.08 \cdot 0.99} = \frac{0.009}{0.0882} \approx 0.102.$$

Aunque el test “acierta” mucho, ¡solo el 10 % de los positivos está enfermo! La baja prevalencia (1 %) domina. Este es el famoso problema de la *tasa base* (*base rate*).

6. Diagramas de árbol

Un **diagrama de árbol** (**tree diagram**) organiza experimentos por etapas: cada rama lleva una probabilidad condicional, y la probabilidad de un camino completo es el *producto* de sus ramas (regla del producto). Sumar caminos que terminan en el mismo evento es, en el fondo, la ley de probabilidad total.

El árbol es la versión visual de la regla del producto y la ley de prob. total juntas: multiplicas a lo largo de un camino y sumas entre caminos. Ideal para problemas de Bayes con dos o tres etapas.

7. Conjuntos contables y no contables

Definición 7.1 (Contable / no contable). Un conjunto es **contable** (**countable**) si es finito o si sus elementos se pueden poner en correspondencia uno a uno con los naturales $\mathbb{N}$ (“se pueden listar”). Si no se puede, es **no contable** (**uncountable**).

Ejemplo 7.1 (¿Cuál es cuál?). ■ $\mathbb{Z}$ (enteros) y $\mathbb{Q}$ (rationales) son **contables**.

- $\{1, \frac{1}{2}, \frac{1}{3}, \dots\} = \{\frac{1}{n} : n \in \mathbb{N}\}$ es contable (se lista).
- Cualquier intervalo real como $(0, 1]$ es **no contable**.

Esta distinción separa los dos mundos de la Clase 3: las variables aleatorias *discretas* toman valores en un conjunto contable (y usan sumas $\sum$), mientras las *continuas* toman valores en un intervalo no contable (y usan integrales $\int$).

8. En R: simular probabilidad condicional y Bayes

Cuando un cálculo es enredado, conviene *simular*. Verifiquemos el problema de las tres monedas (dos normales y una de dos caras):

Simular Bayes en R

```
set.seed(1)
N <- 1e5
moneda <- sample(c("normal", "normal", "doscaras"), N, replace = TRUE)
cara <- ifelse(moneda == "doscaras", TRUE, runif(N) < 0.5)
mean(cara) # P(cara) ~ 0.667
mean(moneda[cara] == "doscaras") # P(dos caras | cara) ~ 0.5
```

Las frecuencias se acercan a $\mathbb{P}(\text{cara}) = 2/3$ y $\mathbb{P}(\text{dos caras} \mid \text{cara}) = 1/2$.

Resumen de la clase

Lo esencial de la Clase 2

- **Condiciona**l: $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B) / \mathbb{P}(B)$; reescala el mundo a B .
- **Producto**: $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$.
- **Independencia**: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \Leftrightarrow \mathbb{P}(A \mid B) = \mathbb{P}(A)$. No confundir con disjuntos.
- **Prob. total**: $\mathbb{P}(A) = \sum_i \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)$ sobre una partición.
- **Bayes**: invierte el condicionamiento; prior $\rightarrow$ posterior. Cuidado con la tasa base.
- **Árbol**: multiplicar por camino, sumar entre caminos.
- **Contable vs no contable**: discreto ($\sum$) vs continuo ($\int$).

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