

14.310x – Data Analysis for Social Scientists

MicroMasters en Statistics and Data Science (SDS) – MITx

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

Clase 1

Fundamentos de Probabilidad y Conteo

Soluciones

Módulos oficiales del MicroMaster:

M1 – Introduction (setup de R); M2 – Fundamentals of Probability (axiomas, conteo)

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 (Contar resultados favorables (guiado)). Una bolsa tiene 4 fichas iguales numeradas 1, 2, 3, 4. Se saca *una* ficha al azar (todas igualmente probables). Sigue los pasos.

- (a) Escribe el espacio muestral S y di cuántos resultados tiene, $n(S)$.
- (b) Lista el evento $A =$ “sale un número par”. ¿Cuántos resultados favorables, $n(A)$?
- (c) Aplica la regla de Laplace $\mathbb{P}(A) = \frac{n(A)}{n(S)}$ para hallar $\mathbb{P}(A)$.

Solución

- (a) $S = \{1, 2, 3, 4\}$, así $n(S) = 4$.
- (b) Los pares en S son $\{2, 4\}$, luego $n(A) = 2$.
- (c) $\mathbb{P}(A) = \frac{n(A)}{n(S)} = \frac{2}{4} = \frac{1}{2} = 0.5$.

Idea clave: cuando todos los resultados son igualmente probables, la probabilidad es simplemente “casos favorables sobre casos totales”.

Ejercicio 1.2 (Regla del complemento, paso a paso). La probabilidad de que mañana llueva en una ciudad es $\mathbb{P}(\text{lluvia}) = 0.30$.

- (a) ¿Qué significa el evento “complemento” de “lluvia”?
- (b) Usando $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, calcula $\mathbb{P}(\text{no lluvia})$.

Solución

(a) El complemento de “lluvia” es “no lluvia”: todo lo que ocurre cuando el evento no sucede. Juntos cubren todas las posibilidades.

(b) $\mathbb{P}(\text{no lluvia}) = 1 - \mathbb{P}(\text{lluvia}) = 1 - 0.30 = 0.70$.

Idea clave: como un evento y su complemento agotan el espacio muestral y son disjuntos, sus probabilidades *suman* 1.

Ejercicio 1.3 (Regla del producto para contar (guiado)). Una cafetería ofrece 3 tipos de pan y 4 rellenos. Un sándwich se arma eligiendo *un* pan y *un* relleno.

(a) Para el primer pan, ¿cuántos sándwiches distintos puedes formar variando solo el relleno?

(b) Repite el razonamiento para los 3 panes y suma. ¿Cuántos sándwiches hay en total?

(c) Comprueba tu respuesta con la *regla del producto* (multiplicar las opciones de cada paso).

Solución

(a) Con un pan fijo hay 4 rellenos posibles, así 4 sándwiches.

(b) Lo mismo para cada uno de los 3 panes: $4 + 4 + 4 = 12$.

(c) Regla del producto: $3 \text{ panes} \times 4 \text{ rellenos} = 12$. Coincide.

Idea clave: si una elección se hace en pasos independientes, se *multiplican* las opciones de cada paso.

Ejercicio 1.4 (Probabilidad de un evento simple en un dado (guiado)). Se lanza un dado justo de 6 caras una vez.

(a) ¿Cuál es $n(S)$?

(b) Halla $\mathbb{P}(\text{sale un número} \leq 4)$ listando los casos favorables.

(c) Halla $\mathbb{P}(\text{sale un número} > 4)$ usando el complemento de (b) y verifica que ambos sumen 1.

Solución

(a) $S = \{1, 2, 3, 4, 5, 6\}$, $n(S) = 6$.

(b) Favorables: $\{1, 2, 3, 4\}$, son 4, así $\mathbb{P}(\leq 4) = \frac{4}{6} = \frac{2}{3} \approx 0.667$.

(c) $\mathbb{P}(> 4) = 1 - \frac{2}{3} = \frac{1}{3}$ (corresponde a $\{5, 6\}$, 2 casos: $2/6 = 1/3$). Y en efecto $\frac{2}{3} + \frac{1}{3} = 1$.

Ejercicio 1.5 (Eventos al lanzar una moneda 3 veces). Se lanza una moneda justa 3 veces y se anota cara (C) o sello (S). El espacio muestral tiene 8 resultados:

$$S = \{CCC, CCS, CSC, SCC, CSS, SCS, SSC, SSS\}.$$

- (a) ¿Cuál de estos conjuntos es exactamente el evento “salieron *exactamente* dos caras”?
 $A = \{SCC, CSC, CCS, CCC\}$, $B = \{SCC, CSC, CCS\}$, $C = \{CSC, SCC\}$.
- (b) En palabras, ¿qué evento describe el conjunto $\{SCC, CSC, CCS\}$?
- (c) ¿Son disjuntos los eventos “exactamente 2 caras” y “exactamente 2 sellos”?
- (d) ¿“Salir al menos 2 caras” implica “salir exactamente 2 caras”?

Solución

- (a) La respuesta es $B = \{SCC, CSC, CCS\}$. El conjunto A incluye de más a CCC (que tiene *tres* caras), y C olvida el resultado CCS .
- (b) Es el evento “sale *exactamente* un sello” (one tail): en cada resultado de $\{SCC, CSC, CCS\}$ hay un solo sello y dos caras.
- (c) **Verdadero:** son disjuntos. “Exactamente 2 caras” = $\{CCS, CSC, SCC\}$ y “exactamente 2 sellos” = $\{SSC, SCS, CSS\}$ no comparten ningún resultado, así que su intersección es vacía.
- (d) **Falso.** El resultado CCC tiene “al menos 2 caras” (tiene 3), pero *no* tiene exactamente 2. Tener al menos 2 es un evento más grande que tener exactamente 2.

Ejercicio 1.6 (Probabilidad con un dado). Se lanza un dado justo de 6 caras.

- (a) ¿Cuál es la probabilidad de que *no* salga un 1?
- (b) Si se lanza el dado 3 veces de forma independiente, ¿cuál es la probabilidad de que en *ninguno* de los tres salga un 1?

Solución

- (a) Cinco de las seis caras (2, 3, 4, 5, 6) no son 1, así que $\mathbb{P}(\text{no sale 1}) = \frac{5}{6} \approx 0.8333$.
- (b) Como los lanzamientos son independientes, multiplicamos:

$$\mathbb{P}(\text{en ninguno sale 1}) = \left(\frac{5}{6}\right)^3 = \frac{125}{216} \approx 0.5787.$$

En R

```
prob <- 5/6 prob^3 # 0.5787037
```

Ejercicio 1.7 (Canicas: una extracción). Un recipiente tiene 5 canicas azules, 3 amarillas y 4 grises (12 en total). Se extrae una al azar. ¿Cuál es la probabilidad de sacar una canica azul?

Solución

Hay $n(S) = 5 + 3 + 4 = 12$ canicas equiprobables, de las cuales 5 son azules:

$$\mathbb{P}(\text{azul}) = \frac{5}{12} \approx 0.4167.$$

En R

```
azules <- 5; total <- 12 azules / total # 0.4166667
```

Ejercicio 1.8 (Canicas: complemento). Del mismo recipiente (5 azules, 3 amarillas, 4 grises; 12 en total) se extrae una canica al azar. ¿Cuál es la probabilidad de que *no* sea azul?

Solución

Por la regla del complemento,

$$\mathbb{P}(\text{no azul}) = 1 - \mathbb{P}(\text{azul}) = 1 - \frac{5}{12} = \frac{7}{12} \approx 0.5833.$$

(Equivale a contar directamente las $3 + 4 = 7$ canicas no azules.)

Ejercicio 1.9 (Espacio muestral de una encuesta). Una encuestadora visita un hogar y registra si tiene acceso a internet (Sí / No) y la zona donde vive (Urbana / Rural).

- (a) Escribe el espacio muestral S completo.
- (b) ¿Cuántos resultados tiene $n(S)$?
- (c) Escribe el evento $A = \text{“el hogar es rural”}$ como subconjunto de S .

Solución

- (a) Cada hogar es un par (internet, zona):

$$S = \{(\text{Sí}, \text{Urbana}), (\text{Sí}, \text{Rural}), (\text{No}, \text{Urbana}), (\text{No}, \text{Rural})\}.$$

- (b) $n(S) = 2 \times 2 = 4$ (regla del producto).
(c) $A = \{(\text{Sí}, \text{Rural}), (\text{No}, \text{Rural})\}.$

Ejercicio 1.10 (Regla de la suma). En un distrito, $\mathbb{P}(\text{un adulto tiene empleo}) = 0.55$ y $\mathbb{P}(\text{tiene educación superior}) = 0.30$. Además $\mathbb{P}(\text{empleo y educación superior}) = 0.20$.

- (a) ¿Probabilidad de tener empleo *o* educación superior (al menos uno)?
- (b) ¿Probabilidad de no tener ninguno de los dos?

Solución

- (a) Regla de la suma: $\mathbb{P}(E \cup U) = 0.55 + 0.30 - 0.20 = 0.65$.
- (b) Complemento: $1 - 0.65 = 0.35$.

Ejercicio 1.11 (Factorial y conteo básico). Cinco ponentes (entre ellos Ana y Beto) se sientan en una fila de 5 sillas, en orden.

- (a) ¿De cuántas maneras pueden ordenarse?
- (b) ¿En cuántas Ana queda en la primera silla?

Solución

- (a) $5! = 120$ ordenamientos (permutaciones de 5 objetos).
- (b) Si Ana está fija en la silla 1, los otros 4 se ordenan de $4! = 24$ maneras.

Ejercicio 1.12 (Complemento e independencia). En una población, la probabilidad de que una persona esté afiliada a algún seguro de salud es 0.78. Se elige una persona al azar.

- (a) ¿Probabilidad de que *no* esté afiliada?
- (b) Si se eligen dos personas de forma independiente, ¿probabilidad de que ninguna esté afiliada? (Usa que para independientes $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.)

Solución

- (a) $1 - 0.78 = 0.22$.
- (b) $0.22 \times 0.22 = 0.0484$.

Ejercicio 1.13 (Verificar un axioma). Alguien propone asignar a tres resultados $\{a, b, c\}$ las “probabilidades” $\mathbb{P}(a) = 0.5$, $\mathbb{P}(b) = 0.3$, $\mathbb{P}(c) = 0.4$.

- (a) ¿Es esto una probabilidad válida? Justifica con los axiomas.
- (b) Propone una corrección que mantenga las proporciones relativas.

Solución

- (a) No: $0.5 + 0.3 + 0.4 = 1.2 \neq 1$, viola el axioma de normalización $\mathbb{P}(S) = 1$.
- (b) Normalizamos dividiendo entre 1.2: $\mathbb{P}(a) = 0.5/1.2 \approx 0.417$, $\mathbb{P}(b) = 0.25$, $\mathbb{P}(c) \approx 0.333$, que suman 1.

Nivel 2 – Núcleo (core)

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

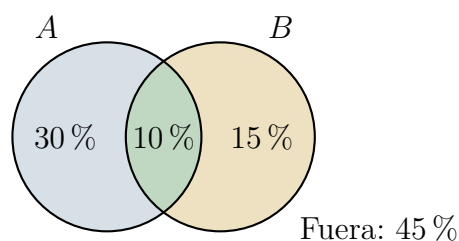
Ejercicio 1.14 (Regla del producto en conteo). Una clave de acceso a un sistema de encuestas tiene 3 caracteres: el primero una letra (de 26), los otros dos dígitos (de 0 a 9), y se permite repetir.

- (a) ¿Cuántas claves distintas hay?
- (b) ¿Probabilidad de que una clave al azar empiece con la letra A?

Solución

- (a) $26 \times 10 \times 10 = 2600$ (regla del producto).
- (b) Claves que empiezan con A: $1 \times 10 \times 10 = 100$, así $\mathbb{P} = 100/2600 = 1/26 \approx 0.0385$.

Ejercicio 1.15 (Diagrama de Venn: dos programas sociales). En un distrito, el 40 % de los hogares recibe el programa *A* (apoyo alimentario) y el 25 % recibe el programa *B* (beca escolar). Un 10 % recibe *ambos*. Se elige un hogar al azar. El diagrama de Venn resume la situación (en porcentajes del total):



- (a) ¿Probabilidad de que el hogar reciba *solo* el programa *A* (y no *B*)?
- (b) ¿Probabilidad de que reciba *al menos uno* de los dos?
- (c) ¿Probabilidad de que no reciba *ninguno*?

Solución

Trabajamos con proporciones. La parte central (ambos) es 0.10.

- (a) Solo *A*: $\mathbb{P}(A) - \mathbb{P}(A \cap B) = 0.40 - 0.10 = 0.30$ (la región izquierda del diagrama).
- (b) Al menos uno (regla de la suma): $\mathbb{P}(A \cup B) = 0.40 + 0.25 - 0.10 = 0.55$.
- (c) Ninguno (complemento): $1 - 0.55 = 0.45$ (la región de afuera). (adaptado de gevorg-hunanyan/probability)

Ejercicio 1.16 (Suma de dos dados). Se lanzan dos dados justos de 6 caras y se anota el par ordenado (d_1, d_2) , de modo que el espacio muestral tiene $n(S) = 6 \times 6 = 36$ resultados equiprobables.

- (a) ¿Probabilidad de que la suma sea exactamente 7?
- (b) ¿Probabilidad de que la suma sea 7 o 11?

Solución

- (a) La suma 7 se logra con $\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$: 6 casos, así $\mathbb{P}(\text{suma} = 7) = 6/36 = 1/6 \approx 0.1667$.
- (b) La suma 11 se logra con $\{(5, 6), (6, 5)\}$: 2 casos. Como “suma 7” y “suma 11” son disjuntos, $\mathbb{P} = \frac{6+2}{36} = \frac{8}{36} = \frac{2}{9} \approx 0.2222$. (adaptado de gevorg-hunanyan/probability)

Ejercicio 1.17 (Canicas sin reemplazo). Del mismo recipiente del nivel anterior (5 azules, 3 amarillas, 4 grises; 12 en total) se extrae una canica, se deja *fuera* (no se devuelve), y luego se extrae otra. ¿Cuál es la probabilidad de que la *primera* sea azul y la *segunda* no sea azul?

Solución

Como no hay reemplazo, la segunda extracción depende de la primera (sus probabilidades cambian). Usamos la regla del producto para eventos en secuencia:

$$\mathbb{P}(1^{\text{a}} \text{ azul}) = \frac{5}{12}.$$

Si la primera fue azul, quedan 11 canicas de las cuales 7 no son azules:

$$\mathbb{P}(2^{\text{a}} \text{ no azul} \mid 1^{\text{a}} \text{ azul}) = \frac{7}{11}.$$

Por lo tanto

$$\mathbb{P}(\text{azul, luego no azul}) = \frac{5}{12} \cdot \frac{7}{11} = \frac{35}{132} \approx 0.2651.$$

En R

```
(5/12) * (7/11) # 0.2651515
```

Ejercicio 1.18 (Ranking de modelos (permutaciones)). Diez data scientists (Jenny, Freddy, Pamela, Enrique, ...) compiten por construir el modelo de Machine Learning más preciso. Al final se publica un *ranking*, así que el orden importa.

- (a) ¿De cuántas formas distintas puede quedar el ranking completo, del 1.º al 10.º lugar?
- (b) ¿De cuántas formas distintas puede quedar el *podio* (solo 1.º, 2.º y 3.º)?

Solución

(a) Es una permutación de los 10 participantes (el orden importa y se usan todos):

$$10! = 3\,628\,800.$$

(b) Es una permutación de 10 tomados de 3 en 3 (el orden del podio importa):

$$P_3^{10} = \frac{10!}{(10-3)!} = 10 \cdot 9 \cdot 8 = 720.$$

En R

```
factorial(10) # 3628800 factorial(10)/factorial(7) # 720
```

Ejercicio 1.19 (Con vs. sin reemplazo (orden)). Un comité de transparencia tiene 10 miembros. Se eligen 3 para formar una terna.

- (a) Si la terna es (presidente, secretario, vocal) — es decir, los cargos importan — ¿cuántas ternas distintas hay?
- (b) Si la terna es solo “tres personas” sin cargos, ¿cuántas hay?
- (c) Explica en una frase por qué (a) es mayor que (b).

Solución

(a) El orden importa: permutaciones $P_3^{10} = \frac{10!}{7!} = 10 \cdot 9 \cdot 8 = 720$.

(b) El orden no importa: combinaciones $\binom{10}{3} = \frac{10!}{7!3!} = 120$.

(c) Cada grupo de 3 personas se puede ordenar de $3! = 6$ formas; por eso $720 = 120 \times 6$. Las permutaciones cuentan esos 6 órdenes como distintos.

Ejercicio 1.20 (Canicas: al menos una). Una caja tiene 7 balotas de un programa social marcadas “A” y 5 marcadas “B” (12 en total). Se extraen 4 al azar (sin orden, sin reemplazo).

- (a) ¿Cuántos grupos de 4 hay en total?
- (b) ¿Cuántos grupos no contienen ninguna “A”?
- (c) ¿Cuál es la probabilidad de que el grupo tenga *al menos una* “A”?

Solución

(a) $\binom{12}{4} = 495$.

(b) Sin ninguna “A” significa elegir las 4 entre las 5 “B”: $\binom{5}{4} = 5$.

$$(c) \text{ Por complemento: } \mathbb{P}(\geq 1 A) = 1 - \frac{5}{495} = \frac{490}{495} = \frac{98}{99} \approx 0.990.$$

Ejercicio 1.21 (Composición de una muestra). De una población de 6 mujeres y 5 hombres se elige al azar un focus group de 2 personas (sin orden).

- (a) ¿Probabilidad de que el grupo tenga una mujer y un hombre?
- (b) ¿Probabilidad de que sean dos mujeres?

Solución

Total de grupos: $\binom{11}{2} = 55$.

(a) Una mujer y un hombre: $\binom{6}{1}\binom{5}{1} = 6 \cdot 5 = 30$, así $\mathbb{P} = \frac{30}{55} = \frac{6}{11}$.

(b) Dos mujeres: $\binom{6}{2} = 15$, así $\mathbb{P} = \frac{15}{55} = \frac{3}{11}$.

Ejercicio 1.22 (Inclusión–exclusión con conteo). En una rifa con boletos numerados del 1 al 30 (uno se extrae al azar), define $A =$ “múltiplo de 3” y $B =$ “múltiplo de 5”.

- (a) Calcula $\mathbb{P}(A)$, $\mathbb{P}(B)$ y $\mathbb{P}(A \cap B)$.
- (b) Calcula $\mathbb{P}(A \cup B)$.

Solución

$n(S) = 30$.

(a) Múltiplos de 3: $\{3, 6, \dots, 30\}$, 10 casos $\Rightarrow \mathbb{P}(A) = 10/30 = 1/3$. Múltiplos de 5: $\{5, 10, \dots, 30\}$, 6 casos $\Rightarrow \mathbb{P}(B) = 6/30 = 1/5$. Múltiplos de 15: $\{15, 30\}$, 2 casos $\Rightarrow \mathbb{P}(A \cap B) = 2/30 = 1/15$.

(b) $\mathbb{P}(A \cup B) = \frac{1}{3} + \frac{1}{5} - \frac{1}{15} = \frac{5+3-1}{15} = \frac{7}{15}$.

Ejercicio 1.23 (Partición de una muestra por estratos). De 4 encuestadores de Lima y 3 de Cusco se forma un equipo de 3 (sin orden).

- (a) ¿Probabilidad de que el equipo tenga exactamente 2 de Lima y 1 de Cusco?
- (b) ¿Probabilidad de que el equipo sea todo de Lima?

Solución

Total: $\binom{7}{3} = 35$.

- (a) $\binom{4}{2}\binom{3}{1} = 6 \cdot 3 = 18$, así $\mathbb{P} = 18/35 \approx 0.514$.
 (b) $\binom{4}{3} = 4$, así $\mathbb{P} = 4/35 \approx 0.114$.

Ejercicio 1.24 (Anagramas con repetición). ¿Cuántas “palabras” distintas (secuencias de letras) se pueden formar reordenando las letras de ENCUESTA? ¿Y de CENSO?

Solución

ENCUESTA tiene 8 letras con la E repetida 2 veces; el resto únicas. Número de arreglos:
 $\frac{8!}{2!} = \frac{40320}{2} = 20160$.
 CENSO tiene 5 letras todas distintas: $5! = 120$.

Nivel 3 – Aplicación

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

Ejercicio 1.25 (Cumpleaños por simulación Monte Carlo). En una clase de Data Science hay 50 estudiantes. Queremos la probabilidad de que *al menos dos* cumplan años el mismo día (suponiendo 365 días equiprobables, cumpleaños independientes, ignorando el 29 de febrero).

- (a) Escribe la expresión teórica exacta y da su valor aproximado.
 (b) Estima la probabilidad por simulación de Monte Carlo en R y di a qué valor debería acercarse.

Solución

(a) Por complemento (“todos distintos”): el primero ocupa cualquier día, el segundo debe evitar 1, y así sucesivamente. Con $n = 50$ personas y $d = 365$ días,

$$\mathbb{P}(\text{al menos dos coinciden}) = 1 - \frac{365!}{(365 - 50)! 365^{50}} = 1 - \prod_{k=0}^{49} \frac{365 - k}{365} \approx 0.9704.$$

Es decir, con 50 estudiantes es casi seguro que haya una coincidencia.

(b) Repetimos el experimento muchas veces y medimos la proporción de veces que hay un día repetido:

En R

```
set.seed(1) sim <- replicate(10000, { b <- sample(1:365, 50, replace =
TRUE) any(duplicated(b)) }) mean(sim) # ~0.970, cerca del valor teorico
```

La proporción `mean(sim)` estima la probabilidad y debería acercarse a ≈ 0.970 . (`any(duplicated(b))` es TRUE si hay algún cumpleaños repetido; `mean` de un vector lógico es la proporción de TRUE.)

Ejercicio 1.26 (Adyacencia). En una oficina renovada hay 8 escritorios en línea. Se asignan al azar a 8 analistas. ¿Cuál es la probabilidad de que dos de ellos, Carla y Diego, queden *juntos* (en escritorios contiguos)?

Solución

Tratamos a “Carla–Diego” como un solo bloque, que puede ordenarse internamente de 2 maneras. Así hay 7 “objetos” (el bloque + los otros 6) que se ordenan de $7!$ formas, y el bloque internamente de $2!$:

$$\mathbb{P} = \frac{2 \cdot 7!}{8!} = \frac{2}{8} = \frac{1}{4}.$$

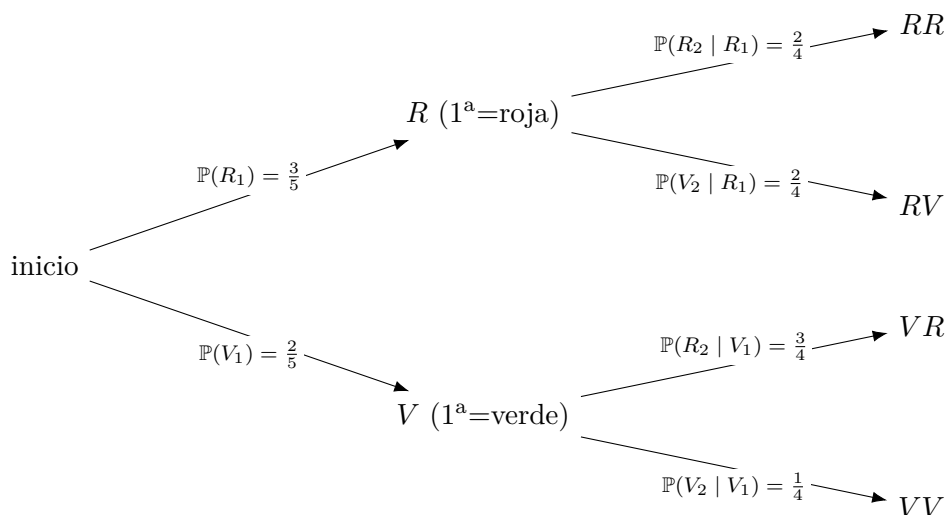
Ejercicio 1.27 (Cumpleaños en un mes). En una oficina de 5 personas, cada una nació en uno de los 12 meses (igualmente probable, independiente).

- (a) ¿Probabilidad de que las 5 hayan nacido en meses *distintos*?
- (b) ¿Probabilidad de que al menos dos compartan mes?

Solución

- (a) $\frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{12^5} = \frac{95040}{248832} \approx 0.382$.
- (b) Complemento: $1 - 0.382 = 0.618$.

Ejercicio 1.28 (Árbol de probabilidad: dos extracciones). Una urna tiene 3 balotas rojas (R) y 2 verdes (V). Se extraen 2 balotas *sin reemplazo*, una tras otra. El árbol de probabilidad organiza los caminos posibles (las probabilidades de la segunda rama dependen de lo que salió en la primera):



Cada rama es una probabilidad condicional: depende del nodo del que sale (V_1 = la 1ª es verde, R_2 = la 2ª es roja, etc.). La probabilidad de un camino es el producto de sus ramas.

- (a) ¿Probabilidad de que ambas sean rojas?
- (b) ¿Probabilidad de que salga exactamente una roja (en cualquier orden)?

Solución

Multiplicamos a lo largo de cada rama del árbol y sumamos las hojas que cumplen.

(a) Rama $R \rightarrow R$: $\frac{3}{5} \cdot \frac{2}{4} = \frac{6}{20} = \frac{3}{10} = 0.3$.

(b) Exactamente una roja = (R luego V) + (V luego R) = $\frac{3}{5} \cdot \frac{2}{4} + \frac{2}{5} \cdot \frac{3}{4} = \frac{6}{20} + \frac{6}{20} = \frac{12}{20} = \frac{3}{5} = 0.6$.

En R

```
(3/5)*(2/4) # 0.3 (3/5)*(2/4) + (2/5)*(3/4) # 0.6
```

(adaptado de gevorg-hunanyan/probability)

Ejercicio 1.29 (Cartas de un comité (combinaciones múltiples)). Una junta directiva tiene 4 economistas, 5 sociólogos y 3 estadísticos (12 personas). Se elige al azar un subcomité de 4 personas (sin orden).

- (a) ¿Cuántos subcomités distintos hay?
- (b) ¿Probabilidad de que el subcomité tenga exactamente 2 economistas, 1 sociólogo y 1 estadístico?

Solución

(a) $\binom{12}{4} = 495$ subcomités.

(b) Casos favorables: $\binom{4}{2} \binom{5}{1} \binom{3}{1} = 6 \cdot 5 \cdot 3 = 90$. Por lo tanto

$$\mathbb{P} = \frac{90}{495} = \frac{2}{11} \approx 0.1818.$$

En R

```
choose(4,2)*choose(5,1)*choose(3,1) / choose(12,4) # 0.1818182
```

(adaptado de mikem5/5080-5090-Solutions)

Ejercicio 1.30 (El problema del cumpleaños, versión aula). En tu clase hay 23 estudiantes. Suponiendo que cada cumpleaños cae con igual probabilidad en cualquiera de 365 días (sin años bisiestos, cumpleaños independientes), ¿cuál es la probabilidad de que *al menos dos* compartan cumpleaños? Plantea la expresión exacta y da el valor aproximado.

Solución

Usamos el complemento: “todos distintos”. El primero puede nacer cualquier día, el segundo en 364 días libres, etc.:

$$\mathbb{P}(\text{todos distintos}) = \frac{365 \cdot 364 \cdots (365 - 22)}{365^{23}} = \frac{365!}{(365 - 23)! 365^{23}} = \prod_{k=0}^{22} \frac{365 - k}{365}.$$

Numéricamente ≈ 0.4927 . Por lo tanto

$$\mathbb{P}(\text{al menos dos coinciden}) = 1 - 0.4927 \approx 0.507.$$

Más del 50 % —resultado contraintuitivo: bastan 23 personas.

Ejercicio 1.31 (Conteo con restricción, en R). Un equipo de campo de 4 personas se forma a partir de 5 encuestadores y 3 supervisores (8 en total). Se exige que el equipo tenga *al menos un* supervisor.

- (a) ¿Cuántos equipos cumplen la condición? Resuélvelo por complemento.
- (b) Escribe una línea de R que devuelva ese número.

Solución

- (a) Total de equipos: $\binom{8}{4} = 70$. Equipos *sin* supervisor (los 4 salen de los 5 encuestadores): $\binom{5}{4} = 5$. Con al menos un supervisor: $70 - 5 = 65$.
- (b) En R:

En R

```
choose(8, 4) - choose(5, 4) # 65
```

Ejercicio 1.32 (Conteo de caminos en una cuadrícula). Un encuestador parte de la esquina inferior izquierda de una cuadrícula de 4×3 manzanas y debe llegar a la esquina superior derecha moviéndose solo a la derecha (D) o hacia arriba (U).

- (a) ¿Cuántas rutas más cortas distintas existen?
- (b) Si elige una ruta al azar, ¿probabilidad de que sus dos primeros pasos sean ambos “D”?

Solución

- (a) Toda ruta corta usa 4 pasos D y 3 pasos U; es una secuencia de 7 pasos donde elegimos cuáles 3 son U: $\binom{7}{3} = 35$.
- (b) Si los dos primeros son D, quedan 2 D y 3 U en los últimos 5 pasos: $\binom{5}{3} = 10$ rutas. Así $\mathbb{P} = 10/35 = 2/7$.

Ejercicio 1.33 (Falacia del jugador vs. independencia). Una moneda justa se lanza 5 veces de forma independiente.

- (a) ¿Probabilidad de la secuencia exacta CCCCC?
- (b) ¿Probabilidad de la secuencia exacta SCSCS?
- (c) Tras observar cuatro caras seguidas, ¿cuál es la probabilidad de que el quinto lanzamiento sea cara? Comenta la “falacia del jugador”.

Solución

- (a) y (b): cada secuencia específica de 5 lanzamientos tiene probabilidad $(1/2)^5 = 1/32$; ambas son igual de probables.
- (c) Como los lanzamientos son independientes, el quinto es cara con probabilidad $1/2$, sin importar los anteriores. Creer que “ya toca sello” es la *falacia del jugador*: confunde independencia con compensación.

Ejercicio 1.34 (Principio del palomar). En una encuesta nacional, cada hogar se clasifica por uno de 4 niveles socioeconómicos.

- (a) ¿Cuál es el mínimo número de hogares que garantiza que al menos 2 comparten nivel?
- (b) ¿Y para garantizar que al menos 3 comparten nivel?

Solución

- (a) Con 4 niveles, 4 hogares podrían caer en niveles distintos; el quinto fuerza una repetición: mínimo 5.
- (b) Para forzar un nivel con 3, en el peor caso hay 2 por nivel ($2 \times 4 = 8$) sin lograrlo; el siguiente lo fuerza: mínimo 9. (Generaliza: $4(k - 1) + 1$ para garantizar k .)

Nivel 4 – Reto

Difícil: integra varios temas en un mismo problema.

Ejercicio 1.35 (Distribución hipergeométrica, sin nombrarla). Un padrón de 50 votantes contiene 20 que efectivamente votaron y 30 que no. Un auditor toma una muestra aleatoria de 8 registros (sin reemplazo).

- (a) Escribe, en términos de combinaciones, la probabilidad de que la muestra contenga *exactamente* 3 votantes.
- (b) Generaliza: probabilidad de exactamente k votantes en la muestra de 8.

Solución

Total de muestras: $\binom{50}{8}$.

(a) Elegir 3 de los 20 votantes y 5 de los 30 no votantes:

$$\mathbb{P}(3 \text{ votantes}) = \frac{\binom{20}{3} \binom{30}{5}}{\binom{50}{8}}.$$

(b) En general, para $k = 0, 1, \dots, 8$ (con $\binom{20}{k} = 0$ si $k > 20$, etc.):

$$\mathbb{P}(k \text{ votantes}) = \frac{\binom{20}{k} \binom{30}{8-k}}{\binom{50}{8}}.$$

Esta es la **distribución hipergeométrica (hypergeometric)**, que reaparecerá más adelante en el curso.

Ejercicio 1.36 (Cota de Bonferroni / subaditividad). Sean A_1, A_2, A_3 tres eventos cualesquiera (no necesariamente disjuntos).

- (a) Demuestra, usando solo los axiomas y la regla de la suma, que $\mathbb{P}(A_1 \cup A_2) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2)$.
- (b) Extiende el argumento para mostrar que $\mathbb{P}\left(\bigcup_{i=1}^3 A_i\right) \leq \sum_{i=1}^3 \mathbb{P}(A_i)$.
- (c) Si un programa tiene tres riesgos, cada uno con probabilidad 0.05 de materializarse, ¿qué cota superior te da esto para “ocurre al menos un riesgo”? ¿Por qué es solo una cota y no el valor exacto?

Solución

(a) Por la regla de la suma, $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$. Como por el axioma 1 se tiene $\mathbb{P}(A_1 \cap A_2) \geq 0$, restar algo no negativo da $\mathbb{P}(A_1 \cup A_2) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2)$.

(b) Aplica (a) a $A_1 \cup A_2$ y luego a $(A_1 \cup A_2) \cup A_3$:

$$\mathbb{P}\left((A_1 \cup A_2) \cup A_3\right) \leq \mathbb{P}(A_1 \cup A_2) + \mathbb{P}(A_3) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3).$$

(c) Cota: $\mathbb{P}(\text{al menos un riesgo}) \leq 0.05 \times 3 = 0.15$. Es solo una cota porque ignora los solapamientos (cuando dos riesgos ocurren juntos), que la regla de la suma restaría. El valor exacto es ≤ 0.15 .

Ejercicio 1.37 (Monotonía de la probabilidad). Sean A y B dos eventos de un mismo espacio muestral con $A \subseteq B$ (siempre que ocurre A , también ocurre B).

- (a) Demuestra, usando solo los axiomas, que $\mathbb{P}(A) \leq \mathbb{P}(B)$.

- (b) Como aplicación, si “el hogar es pobre extremo” $\subseteq$ “el hogar es pobre”, ¿qué desigualdad deben cumplir sus probabilidades? Da un ejemplo numérico coherente y otro *incoherente*.

Solución

(a) Si $A \subseteq B$, podemos partir B en dos pedazos disjuntos: $B = A \cup (B \setminus A)$, con $A \cap (B \setminus A) = \emptyset$. Por aditividad (axioma 3),

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

Como por el axioma 1 se tiene $\mathbb{P}(B \setminus A) \geq 0$, se concluye $\mathbb{P}(B) \geq \mathbb{P}(A)$, es decir $\mathbb{P}(A) \leq \mathbb{P}(B)$.

(b) Debe cumplirse $\mathbb{P}(\text{pobre extremo}) \leq \mathbb{P}(\text{pobre})$. Ejemplo *coherente*: $0.08 \leq 0.20$. Ejemplo *incoherente*: afirmar $\mathbb{P}(\text{pobre extremo}) = 0.30$ y $\mathbb{P}(\text{pobre}) = 0.20$ viola la monotonicidad, porque un subconjunto no puede ser más probable que el conjunto que lo contiene. (adaptado de gevorg-hunanyan/probability, “monotonicity”)

Ejercicio 1.38 (Sobres mal repartidos: desordenamientos pequeños). 4 encuestados entregan su tablet personal y, al terminar, un asistente distraído se las devuelve al azar (una permutación uniforme de las 4). Nos interesa la probabilidad de que *nadie* reciba su propia tablet (un *desordenamiento* o *derangement*).

- (a) Por inclusión-exclusión, el número de desordenamientos de n objetos es $D_n = n! \sum_{k=0}^n \frac{(-1)^k}{k!}$.
Calcula D_4 y la probabilidad pedida.
- (b) Compara tu resultado con el límite $e^{-1} \approx 0.3679$ cuando $n \rightarrow \infty$.

Solución

(a) Hay $4! = 24$ permutaciones equiprobables. El número de desordenamientos es

$$D_4 = 4! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} \right) = 24 \left(1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} \right) = 24 \cdot \frac{9}{24} = 9.$$

Por lo tanto $\mathbb{P}(\text{nadie recibe la suya}) = \frac{D_4}{4!} = \frac{9}{24} = \frac{3}{8} = 0.375$.

(b) Ya con $n = 4$ el valor 0.375 está muy cerca del límite $e^{-1} \approx 0.3679$: la probabilidad converge rapidísimo y casi no depende de n .

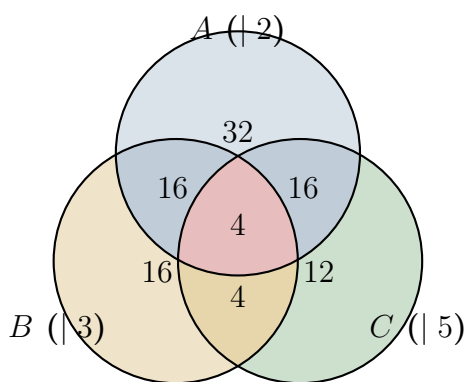
En R

```
n <- 4
D4 <- factorial(n) * sum((-1)^(0:n) / factorial(0:n))
D4 / factorial(n) # 0.375
```

(adaptado de gevorg-hunanyan/probability, “five-envelopes”)

Ejercicio 1.39 (Inclusión–exclusión con tres criterios). Un padrón tiene los registros numerados del 1 al 120. Se extrae un registro al azar (equiprobable). Define los eventos según el divisor del número de registro: A = “divisible por 2”, B = “divisible por 3”, C = “divisible por 5”.

- Calcula $n(A)$, $n(B)$, $n(C)$ y las tres intersecciones de a pares $n(A \cap B)$, $n(A \cap C)$, $n(B \cap C)$, y la triple $n(A \cap B \cap C)$.
- Usa la *inclusión–exclusión* para tres conjuntos para hallar $\mathbb{P}(A \cup B \cup C)$.
- Interpreta: ¿qué proporción de registros *no* es divisible por ninguno de 2, 3 ni 5? Relaciona el conteo con el diagrama de Venn de abajo.



Solución

Aquí $n(S) = 120$. Cada divisibilidad se cuenta con la parte entera de 120 entre el divisor.

(a) Conteos básicos:

$$n(A) = \left\lfloor \frac{120}{2} \right\rfloor = 60, \quad n(B) = \left\lfloor \frac{120}{3} \right\rfloor = 40, \quad n(C) = \left\lfloor \frac{120}{5} \right\rfloor = 24.$$

“Divisible por 2 y por 3” equivale a “divisible por 6”, etc.:

$$n(A \cap B) = \left\lfloor \frac{120}{6} \right\rfloor = 20, \quad n(A \cap C) = \left\lfloor \frac{120}{10} \right\rfloor = 12, \quad n(B \cap C) = \left\lfloor \frac{120}{15} \right\rfloor = 8,$$

y la triple es “divisible por 30”: $n(A \cap B \cap C) = \lfloor 120/30 \rfloor = 4$.

(b) Inclusión–exclusión para tres conjuntos (se suman los individuales, se restan los pares y se vuelve a sumar la triple, porque al restar los pares el centro se descontó de más):

$$n(A \cup B \cup C) = 60 + 40 + 24 - 20 - 12 - 8 + 4 = 88.$$

Por lo tanto $\mathbb{P}(A \cup B \cup C) = \frac{88}{120} = \frac{11}{15} \approx 0.7333$.

(c) Por el complemento, los que *no* son divisibles por ninguno son $120 - 88 = 32$, así $\mathbb{P}(\text{ninguno}) = \frac{32}{120} = \frac{4}{15} \approx 0.2667$. En el diagrama, las cifras son los conteos *exclusivos*

de cada región: las siete regiones internas suman $32 + 16 + 12 + 16 + 16 + 4 + 4 = 100$, pero el conteo correcto de $n(A \cup B \cup C)$ tras la inclusión-exclusión es 88, y “Fuera” es 32.

En R

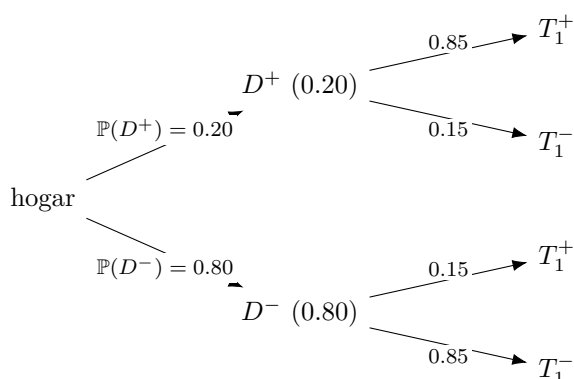
```
N <- 120 union <- sum(c(60,40,24)) - sum(c(20,12,8)) + 4 union # 88 (N - union)/N # 0.2666667
```

(adaptado de MIT 18.05, Class 1 — inclusion-exclusion principle)

Ejercicio 1.40 (Bayes secuencial: focalización con dos filtros). Un programa social quiere identificar hogares en pobreza. En el distrito, la proporción real de hogares pobres es $\mathbb{P}(D^+) = 0.20$. Se aplican *dos* filtros, que se suponen condicionalmente independientes dado el estado real del hogar:

- **Filtro 1** (encuesta de variables proxy): sensibilidad $\mathbb{P}(T_1^+ | D^+) = 0.85$ y falsos positivos $\mathbb{P}(T_1^+ | D^-) = 0.15$.
- **Filtro 2** (visita domiciliaria, solo si el filtro 1 dio positivo): $\mathbb{P}(T_2^+ | D^+) = 0.90$ y $\mathbb{P}(T_2^+ | D^-) = 0.10$.

- (a) Un hogar da positivo en el filtro 1. Usa Bayes para hallar $\mathbb{P}(D^+ | T_1^+)$.
- (b) Ese mismo hogar también da positivo en el filtro 2. Actualiza *secuencialmente* (usando el resultado de (a) como nueva probabilidad previa) para hallar $\mathbb{P}(D^+ | T_1^+, T_2^+)$.
- (c) Verifica que el resultado coincide con aplicar Bayes una sola vez al evento conjunto “ambos positivos”.



Solución

(a) Por la ley de probabilidad total, el denominador (probabilidad de positivo en el filtro

1) es

$$\mathbb{P}(T_1^+) = \underbrace{0.85 \cdot 0.20}_{D^+} + \underbrace{0.15 \cdot 0.80}_{D^-} = 0.17 + 0.12 = 0.29.$$

Bayes:

$$\mathbb{P}(D^+ | T_1^+) = \frac{0.85 \cdot 0.20}{0.29} = \frac{0.17}{0.29} = \frac{17}{29} \approx 0.5862.$$

La probabilidad de pobreza sube del 20 % al ≈ 58.6 %, pero *aún no* es mayoritaria: el filtro 1 por sí solo no basta para focalizar.

(b) Tomamos $\mathbb{P}(D^+)_{\text{nueva}} = \frac{17}{29}$ como previa y aplicamos Bayes con el filtro 2 (condicionalmente independiente):

$$\mathbb{P}(D^+ | T_1^+, T_2^+) = \frac{0.90 \cdot \frac{17}{29}}{0.90 \cdot \frac{17}{29} + 0.10 \cdot \frac{12}{29}} = \frac{0.90 \cdot 17}{0.90 \cdot 17 + 0.10 \cdot 12} = \frac{15.3}{15.3 + 1.2} = \frac{51}{55} \approx 0.9273.$$

(c) Aplicando Bayes de una vez al evento “ambos positivos” (usando independencia condicional, $\mathbb{P}(T_1^+ T_2^+ | D^\pm)$ es el producto):

$$\mathbb{P}(D^+ | T_1^+ T_2^+) = \frac{(0.85 \cdot 0.90)(0.20)}{(0.85 \cdot 0.90)(0.20) + (0.15 \cdot 0.10)(0.80)} = \frac{0.153}{0.153 + 0.012} = \frac{51}{55} \approx 0.9273,$$

idéntico a (b). *Actualizar de a poco o todo junto da el mismo resultado*: esa es la coherencia de la regla de Bayes.

En R

```
PDp <- 0.20; PDn <- 0.80 post1 <- (0.85*PDp)/(0.85*PDp + 0.15*PDn) #
0.5862069 post2 <- (0.90*post1)/(0.90*post1 + 0.10*(1-post1)) post2 #
0.9272727
```

(adaptado de MIT 18.05, Class 3 — Bayes’ theorem y the base rate fallacy)

Ejercicio 1.41 (Estrellas y barras: repartir cupos idénticos). Un ministerio reparte 10 cupos idénticos de un programa de becas entre 4 regiones (R_1, R_2, R_3, R_4). Solo importa *cuántos* cupos recibe cada región (los cupos no se distinguen entre sí).

- ¿De cuántas formas se pueden repartir los 10 cupos si una región puede quedar con 0 cupos? (Modelo de *estrellas y barras*.)
- ¿Y si se exige que *cada* región reciba al menos 1 cupo?
- Si el reparto se elige al azar de forma uniforme entre todos los del inciso (a), ¿cuál es la probabilidad de que la región R_1 no reciba ningún cupo?

Solución

Idea (estrellas y barras). Repartir n objetos idénticos en r grupos equivale a colocar n “estrellas” en fila e insertar $r - 1$ “barras” separadoras; cada disposición de estrellas y

barras es un reparto. El número de soluciones enteras ≥ 0 de $x_1 + \dots + x_r = n$ es $\binom{n+r-1}{r-1}$.

(a) Con $n = 10$, $r = 4$:

$$\binom{10+4-1}{4-1} = \binom{13}{3} = 286.$$

(b) “Cada región al menos 1” se reduce dando primero 1 cupo a cada una (gasta 4) y repartiendo los $10 - 4 = 6$ restantes sin restricción:

$$\binom{6+4-1}{4-1} = \binom{9}{3} = 84.$$

(c) Si R_1 recibe 0, repartimos los 10 cupos entre las otras 3 regiones: $\binom{10+3-1}{3-1} = \binom{12}{2} = 66$. Por lo tanto

$$\mathbb{P}(R_1 = 0) = \frac{66}{286} = \frac{3}{13} \approx 0.2308.$$

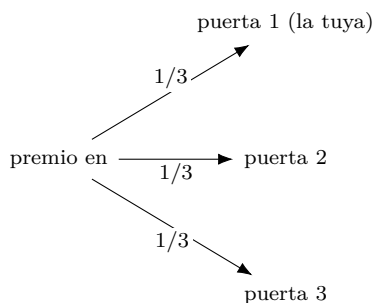
En R

```
choose(13,3) # 286 choose(9,3) # 84 choose(12,2)/choose(13,3) # 0.2307692
```

(adaptado de MIT 18.05, Class 1 — counting con objetos idénticos)

Ejercicio 1.42 (Monty Hall riguroso). En un concurso hay 3 puertas; detrás de una hay un premio y detrás de las otras dos no hay nada. Eliges la puerta 1. El presentador, que *sabe* dónde está el premio, abre una de las otras dos puertas que con seguridad *no* tiene premio y te ofrece cambiar a la puerta cerrada restante.

- Define con cuidado el espacio muestral y las probabilidades. Calcula, con Bayes o con la ley de probabilidad total, la probabilidad de ganar si *te quedas* y la de ganar si *cambias*.
- Explica en una frase por qué la información del presentador rompe la simetría y por qué la respuesta es contraintuitiva.



Solución

(a) Como el premio se colocó al azar, antes de abrir nada $\mathbb{P}(\text{premio en 1}) = \mathbb{P}(\text{premio en 2}) = \mathbb{P}(\text{premio en 3}) = \frac{1}{3}$. Elegiste la puerta 1.

- **Si te quedas** en la 1, ganas exactamente cuando el premio estaba en la 1: $\mathbb{P}(\text{ganar} \mid \text{quedarse}) = \frac{1}{3}$.
- **Si cambias:** ganas exactamente cuando tu elección inicial fue *incorrecta* (premio en 2 o en 3), porque entonces el presentador se ve obligado a abrir la única puerta vacía restante y la que queda cerrada es la del premio. Eso ocurre con probabilidad $\frac{2}{3}$:

$$\mathbb{P}(\text{ganar} \mid \text{cambiar}) = \mathbb{P}(\text{elección inicial incorrecta}) = \frac{2}{3}.$$

Formalmente con Bayes: sea que abre la puerta 3. Si el premio está en 1, abre 2 o 3 con prob. $\frac{1}{2}$; si está en 2, abre 3 con prob. 1; si está en 3, nunca la abre. Entonces

$$\mathbb{P}(\text{premio en 2} \mid \text{abre 3}) = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0} = \frac{\frac{1}{3}}{\frac{1}{2} + \frac{1}{3}} = \frac{2}{5},$$

que es la probabilidad de ganar cambiando.

(b) El presentador *no* abre al azar: usa su conocimiento para abrir siempre una puerta vacía, concentrando en la otra puerta cerrada toda la probabilidad ($\frac{2}{3}$) que tenían las dos puertas que no elegiste. Por eso conviene cambiar, aunque intuitivamente parezca que quedan “dos puertas al 50 %”.

En R

```
set.seed(1) N <- 1e5 premio <- sample(1:3, N, replace = TRUE) cambiar
<- mean(premio != 1) # ganas cambiando si tu eleccion (1) era incorrecta
cambiar # 0.667 1 - cambiar # 0.333 (quedarse)
```

(adaptado de MIT 18.05, Class 3 — conditional probability / Bayes)

Nivel 5 – Reto+

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

Ejercicio 1.43 (Problema del cumpleaños generalizado). Demuestra que, con n personas y d días igualmente probables, la probabilidad de que *todas* tengan días distintos es

$$p(n, d) = \prod_{k=0}^{n-1} \frac{d-k}{d},$$

y úsalo para argumentar por qué $p(n, d)$ decrece a medida que n crece.

Solución

El primer individuo ocupa cualquiera de los d días ($d/d = 1$). El segundo debe evitar 1 día ocupado: $(d-1)/d$. El $k+1$ -ésimo debe evitar k días: $(d-k)/d$. Como los nacimientos son independientes, multiplicamos: $p(n, d) = \prod_{k=0}^{n-1} (d-k)/d$. Cada factor adicional (para $k \geq 1$) es < 1 , así que agregar una persona multiplica por un número menor que 1: p es decreciente en n .

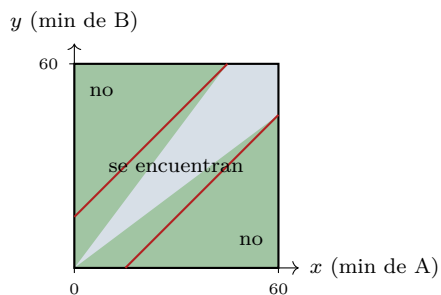
Ejercicio 1.44 (Coincidencias: el problema de los sombreros). n asistentes dejan su DNI y al salir se les devuelve uno al azar (una permutación uniforme). Sea A_i el evento “el asistente i recibe su propio DNI”.

- (a) Argumenta que $\mathbb{P}(A_i) = 1/n$ para cada i .
- (b) Se puede mostrar que $\mathbb{P}(\text{al menos una coincidencia}) = 1 - \sum_{k=0}^n \frac{(-1)^k}{k!}$. Calcula su valor límite cuando $n \rightarrow \infty$.

Solución

- (a) Por simetría, el DNI del asistente i es igualmente probable de ir a cualquiera de las n posiciones; la probabilidad de que vuelva a él es $1/n$. (Formalmente: fijado que i recibe el suyo, las otras $(n-1)!$ permutaciones sobre $n!$ dan $1/n$.)
- (b) $\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = e^{-1}$, así $\mathbb{P}(\text{al menos una}) \rightarrow 1 - e^{-1} \approx 0.632$. Curiosamente, casi no depende de n .

Ejercicio 1.45 (Probabilidad geométrica: el problema de la cita). Dos analistas de campo acuerdan reunirse en la oficina entre las 14:00 y las 15:00. Cada uno llega en un instante *uniforme e independiente* dentro de esa hora y espera a lo más 15 minutos al otro antes de irse. ¿Cuál es la probabilidad de que efectivamente se encuentren?

**Solución**

Midamos en minutos: x, y son las llegadas (en $[0, 60]$) de A y B, uniformes e independientes, así el par (x, y) es uniforme en el cuadrado $[0, 60]^2$ de área $60^2 = 3600$. Se encuentran si y solo si $|x - y| \leq 15$.

El *complemento* ($|x - y| > 15$) son los dos triángulos de las esquinas, cada uno con catetos

de longitud $60 - 15 = 45$:

$$\text{área}(\text{no se encuentran}) = 2 \cdot \frac{45^2}{2} = 45^2 = 2025.$$

Por probabilidad geométrica (probabilidad = área favorable / área total),

$$\mathbb{P}(\text{se encuentran}) = 1 - \frac{2025}{3600} = 1 - \left(\frac{45}{60}\right)^2 = 1 - \frac{9}{16} = \frac{7}{16} = 0.4375.$$

La regla clásica de Laplace (casos favorables / casos totales) sigue valiendo, pero ahora “contar” significa *medir áreas* porque el espacio muestral es continuo.

En R

```
set.seed(1) N <- 1e6 x <- runif(N, 0, 60); y <- runif(N, 0, 60) mean(abs(x
- y) <= 15) # 0.4375
```

(adaptado de MIT 18.05, Class 2 — probability como medida; extensión al caso continuo)

Ejercicio 1.46 (Simetría elegante: ¿qué carta sale primero?). Se baraja un mazo estándar de 52 cartas en orden uniformemente aleatorio.

- ¿Cuál es la probabilidad de que el *as de corazones* aparezca *antes* que el *as de picas* en el mazo? Resuélvelo con un argumento de simetría, *sin* calcular factoriales.
- Generaliza: entre m cartas marcadas específicas, ¿cuál es la probabilidad de que una de ellas elegida de antemano sea la *primera* de las m en aparecer?
- Aplica (b): probabilidad de que, entre los cuatro ases, el as de corazones sea el *primero* de los cuatro en salir.

Solución

(a) Mira solo las posiciones relativas de las dos cartas de interés. En una permutación uniforme, las dos posiciones que ocupan el as de corazones y el as de picas son igualmente probables en cualquiera de los dos órdenes (“corazones antes que picas” o “picas antes que corazones”); ningún orden está favorecido. Por simetría,

$$\mathbb{P}(\text{corazones antes que picas}) = \frac{1}{2}.$$

No hace falta contar las $52!$ permutaciones: basta notar la simetría entre las dos cartas.

(b) Por el mismo argumento, las m cartas marcadas aparecen entre sí en uno de $m!$ órdenes relativos, todos igualmente probables. La carta fijada es la primera en exactamente $(m - 1)!$ de ellos, así

$$\mathbb{P}(\text{la fijada es la primera de las } m) = \frac{(m - 1)!}{m!} = \frac{1}{m}.$$

(c) Con $m = 4$ ases, la probabilidad de que el as de corazones sea el primero de los cuatro es $\frac{1}{4}$.

En R

```
set.seed(1) N <- 1e5 sim <- replicate(N, { pos <- sample(52, 4) #  
posiciones de los 4 ases which.min(pos) == 1 # TRUE si el "as de corazones"  
(1er marcado) sale primero }) mean(sim) # 0.25
```

(adaptado de MIT 18.05, Class 1 — permutations y argumentos de simetría)

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