

14.310x – Data Analysis for Social Scientists

MicroMasters en Statistics and Data Science (SDS) – MITx

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

Clase 5

Distribuciones Especiales, Media Muestral y TLC

Soluciones

Módulos oficiales del MicroMaster:

M5 – Special Distributions, The Sample Mean, the Central Limit Theorem and Estimation

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 (¿Qué distribución modela cada caso?). Para cada situación, indica qué distribución especial la modela mejor (Bernoulli, Binomial, Geométrica, Poisson, Uniforme o Exponencial) y sus parámetros.

- Se encuesta a 1 persona y se anota si tiene (1) o no (0) acceso a internet, con probabilidad 0.6 de tenerlo.
- Número de hogares con internet en una muestra de 50 hogares, cada uno con probabilidad 0.6 e independientes.
- Número de llamadas que llegan a una central de emergencias en 1 hora, con un promedio de 8 por hora.
- Número de personas que hay que entrevistar hasta encontrar la primera que está desempleada, si la probabilidad de desempleo es 0.07.

Solución

- Bernoulli**($p = 0.6$): un solo ensayo con dos resultados.
- Binomial**($n = 50, p = 0.6$): número de éxitos en n ensayos independientes.
- Poisson**($\lambda = 8$): conteo de eventos raros en un intervalo fijo.
- Geométrica**($p = 0.07$): número de ensayos hasta el primer éxito.

Ejercicio 1.2 (Receta guiada: identificar y nombrar la distribución). Sigue los pasos para clasificar cada caso. *Pregunta clave:* ¿cuentas éxitos en un número *fijo* de ensayos (Binomial), eventos en un *intervalo* (Poisson), o ensayos *hasta* el primer éxito (Geométrica)?

- (a) En un salón de 30 estudiantes, cada uno aprueba con probabilidad 0.8 (independientes). Cuentas cuántos aprueban.
- (b) En una biblioteca llegan en promedio 5 usuarios cada 10 minutos. Cuentas cuántos llegan en los próximos 10 minutos.
- (c) Marcas números al azar hasta que alguien contesta; cada llamada se contesta con probabilidad 0.25. Cuentas cuántas llamadas haces hasta la primera contestada.

Solución

Paso 1: ¿conteo en ensayos fijos, en un intervalo, o hasta el primer éxito?

- (a) Ensayos *fijos* ($n = 30$) y cuentas éxitos $\Rightarrow$ **Binomial**($n = 30, p = 0.8$).
- (b) Conteo de eventos en un *intervalo* de tiempo con tasa media $\Rightarrow$ **Poisson**($\lambda = 5$) (por cada 10 minutos).
- (c) Ensayos *hasta* el primer éxito $\Rightarrow$ **Geométrica**($p = 0.25$).

Ejercicio 1.3 (pmf de una Binomial). En una región, el 30% de los hogares recibe un programa social. Se eligen al azar $n = 5$ hogares (independientes). Sea $X \sim \text{Bin}(5, 0.3)$ el número de hogares beneficiarios.

- (a) Calcula $\mathbb{P}(X = 2)$.
- (b) Calcula $\mathbb{P}(X = 0)$.

(Recuerda la fórmula $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$.)

Solución

- (a) $\mathbb{P}(X = 2) = \binom{5}{2} (0.3)^2 (0.7)^3 = 10 \cdot 0.09 \cdot 0.343 = 10 \cdot 0.03087 = 0.3087$.
- (b) $\mathbb{P}(X = 0) = \binom{5}{0} (0.3)^0 (0.7)^5 = (0.7)^5 = 0.16807$.

Ejercicio 1.4 (pmf de una Poisson). A una posta médica llegan en promedio $\lambda = 3$ pacientes por hora. Sea $X \sim \text{Pois}(3)$ el número de pacientes que llegan en una hora dada.

- (a) Calcula $\mathbb{P}(X = 0)$.
- (b) Calcula $\mathbb{P}(X = 2)$.

(Usa $\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ y $e^{-3} \approx 0.0498$.)

Solución

$$(a) \mathbb{P}(X = 0) = \frac{3^0 e^{-3}}{0!} = e^{-3} \approx 0.0498.$$

$$(b) \mathbb{P}(X = 2) = \frac{3^2 e^{-3}}{2!} = \frac{9}{2} e^{-3} = 4.5 \cdot 0.0498 \approx 0.2240.$$

Ejercicio 1.5 ($\mathbb{E}[X]$ y $\text{Var}(X)$ de distribuciones nombradas). Da $\mathbb{E}[X]$ y $\text{Var}(X)$ para cada distribución, usando las fórmulas estándar.

$$(a) X \sim \text{Bin}(20, 0.25) \quad (\mathbb{E}[X] = np, \text{Var}(X) = np(1-p)).$$

$$(b) X \sim \text{Pois}(4) \quad (\mathbb{E}[X] = \text{Var}(X) = \lambda).$$

$$(c) X \sim \text{Geom}(0.2) \quad (\mathbb{E}[X] = 1/p, \text{Var}(X) = (1-p)/p^2).$$

Solución

$$(a) \mathbb{E}[X] = np = 20 \cdot 0.25 = 5; \text{Var}(X) = np(1-p) = 20 \cdot 0.25 \cdot 0.75 = 3.75.$$

$$(b) \mathbb{E}[X] = \lambda = 4; \text{Var}(X) = \lambda = 4 \text{ (en la Poisson coinciden media y varianza).}$$

$$(c) \mathbb{E}[X] = 1/0.2 = 5; \text{Var}(X) = 0.8/0.04 = 20.$$

Ejercicio 1.6 (Bernoulli: media y varianza de un indicador). Sea $X = 1$ si una persona vota en una elección y $X = 0$ si no, con $\mathbb{P}(X = 1) = p = 0.65$ (Bernoulli).

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

$$(b) \text{Calcula } \text{Var}(X) = p(1-p) \text{ y } \text{SD}(X).$$

Solución

$$(a) \mathbb{E}[X] = p = 0.65.$$

$$(b) \text{Var}(X) = p(1-p) = 0.65 \cdot 0.35 = 0.2275; \text{SD}(X) = \sqrt{0.2275} \approx 0.477.$$

Ejercicio 1.7 (Estandarizar a Z paso a paso). Los puntajes X de una prueba nacional se distribuyen aproximadamente $\mathcal{N}(\mu = 500, \sigma^2 = 100^2)$, es decir $\mu = 500, \sigma = 100$. Estandarizar significa restar la media y dividir por la desviación: $Z = \frac{x - \mu}{\sigma}$.

$$(a) \text{Estandariza el puntaje } x = 600.$$

$$(b) \text{Estandariza el puntaje } x = 350.$$

$$(c) \text{Interpreta el signo de cada } z: \text{¿el puntaje está por encima o por debajo del promedio?}$$

Solución

- (a) $z = \frac{600 - 500}{\frac{100}{100}} = \frac{100}{100} = 1$ (un σ por encima de la media).
- (b) $z = \frac{350 - 500}{\frac{100}{100}} = \frac{-150}{100} = -1.5$ (un 1.5σ por debajo).
- (c) $z > 0$ indica un valor por encima del promedio; $z < 0$, por debajo. El valor de z dice cuántas desviaciones estándar se aleja de μ .

Ejercicio 1.8 (Leer la Normal con la regla 68–95–99.7). Las tallas X de adultos en una población son $\mathcal{N}(\mu = 165, \sigma = 8)$ (cm). Recuerda la regla empírica: $\approx 68\%$ cae en $\mu \pm \sigma$, $\approx 95\%$ en $\mu \pm 2\sigma$ y $\approx 99.7\%$ en $\mu \pm 3\sigma$.

- (a) ¿Qué intervalo centrado en 165 contiene aproximadamente el 95 % de las tallas?
- (b) ¿Qué porcentaje aproximado de personas mide entre 157 y 173 cm?
- (c) ¿Qué porcentaje aproximado mide más de 181 cm?

Solución

- (a) El 95 % cae en $\mu \pm 2\sigma = 165 \pm 2(8) = 165 \pm 16$, es decir $[149, 181]$ cm.
- (b) $157 = \mu - \sigma$ y $173 = \mu + \sigma$, así que es $\mu \pm \sigma \Rightarrow$ cerca del **68 %**.
- (c) $181 = \mu + 2\sigma$. Fuera de $\pm 2\sigma$ queda 0.05 repartido en dos colas; la cola superior es ≈ 0.025 . Así, cerca del **2.5 %** mide más de 181 cm.

Ejercicio 1.9 (Media muestral con error estándar simple). El ingreso mensual en una población tiene $\mu = 1\,800$ y $\sigma = 600$ soles. Se toma una muestra aleatoria de $n = 100$ personas y se calcula la media muestral $\bar{X}$.

- (a) Calcula $\mathbb{E}[\bar{X}]$.
- (b) Calcula el error estándar $\text{SD}(\bar{X}) = \sigma/\sqrt{n}$.

Solución

- (a) $\mathbb{E}[\bar{X}] = \mu = 1\,800$ soles (la media muestral es insesgada para μ).
- (b) $\text{SD}(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{600}{\sqrt{100}} = \frac{600}{10} = 60$ soles.

Nivel 2 – Núcleo (core)

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

Ejercicio 1.10 (pmf de una Geométrica). La probabilidad de que una persona encuestada acepte responder es $p = 0.4$. Las personas son independientes. Sea X el número de personas contactadas hasta la primera que acepta (incluyéndola), $X \sim \text{Geom}(0.4)$ con $f(k) = (1 - p)^{k-1}p$.

- (a) Calcula $\mathbb{P}(X = 1)$.
- (b) Calcula $\mathbb{P}(X = 3)$.

Solución

- (a) $\mathbb{P}(X = 1) = (0.6)^0(0.4) = 0.4$.
- (b) $\mathbb{P}(X = 3) = (0.6)^2(0.4) = 0.36 \cdot 0.4 = 0.144$.

Ejercicio 1.11 (Uniforme continua). El tiempo de espera X (en minutos) para ser atendido en un trámite municipal es uniforme en $[0, 30]$, $X \sim \text{Unif}[0, 30]$.

- (a) Calcula $\mathbb{E}[X]$ y $\text{Var}(X)$.
- (b) Calcula $\mathbb{P}(X \leq 10)$.

Solución

- (a) $\mathbb{E}[X] = \frac{a+b}{2} = \frac{0+30}{2} = 15$ minutos; $\text{Var}(X) = \frac{(b-a)^2}{12} = \frac{900}{12} = 75$.
- (b) $\mathbb{P}(X \leq 10) = \frac{10-0}{30} = \frac{1}{3} \approx 0.333$.

Ejercicio 1.12 (Exponencial). El tiempo X (en años) que dura un cargo público antes de rotar es exponencial con tasa $\lambda = \frac{1}{2}$ por año, $f(x) = \frac{1}{2}e^{-x/2}$ para $x \geq 0$. Recuerda que $\mathbb{P}(X > t) = e^{-\lambda t}$.

- (a) Calcula $\mathbb{E}[X]$ y $\text{Var}(X)$.
 - (b) Calcula $\mathbb{P}(X > 4)$.
- (Usa $e^{-2} \approx 0.135$.)

Solución

- (a) $\mathbb{E}[X] = 1/\lambda = 2$ años; $\text{Var}(X) = 1/\lambda^2 = 4$.
- (b) $\mathbb{P}(X > 4) = e^{-\lambda \cdot 4} = e^{-(1/2) \cdot 4} = e^{-2} \approx 0.135$.

Ejercicio 1.13 (Probabilidad Binomial acumulada). El 20% de los estudiantes de una escuela trabaja además de estudiar. Se eligen al azar $n = 6$ estudiantes independientes; $X \sim \text{Bin}(6, 0.2)$ es el número que trabaja.

- (a) Calcula $\mathbb{P}(X \geq 1)$.
- (b) Calcula $\mathbb{P}(X \leq 1)$.

Solución

- (a) $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - (0.8)^6 = 1 - 0.262144 = 0.737856 \approx 0.738$.
- (b) $\mathbb{P}(X = 1) = \binom{6}{1}(0.2)(0.8)^5 = 6 \cdot 0.2 \cdot 0.32768 = 0.393216$. Entonces $\mathbb{P}(X \leq 1) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = 0.262144 + 0.393216 = 0.65536 \approx 0.655$.

Ejercicio 1.14 (Probabilidad Poisson acumulada). Los accidentes de tránsito en un cruce siguen $X \sim \text{Pois}(\lambda = 2)$ por semana.

- (a) Calcula $\mathbb{P}(X \leq 2)$.
- (b) Calcula $\mathbb{P}(X \geq 3)$.

(Usa $e^{-2} \approx 0.1353$.)

Solución

- (a) $\mathbb{P}(X = 0) = e^{-2} = 0.1353$, $\mathbb{P}(X = 1) = 2e^{-2} = 0.2707$, $\mathbb{P}(X = 2) = \frac{4}{2}e^{-2} = 2e^{-2} = 0.2707$. Sumando: $\mathbb{P}(X \leq 2) = 0.1353 + 0.2707 + 0.2707 = 0.6767 \approx 0.677$.
- (b) $\mathbb{P}(X \geq 3) = 1 - \mathbb{P}(X \leq 2) = 1 - 0.6767 = 0.3233 \approx 0.323$.

Ejercicio 1.15 (Exponencial: falta de memoria). El tiempo X (en años) hasta que una pequeña empresa cierra es exponencial con $\mathbb{E}[X] = 5$ años, es decir $\lambda = \frac{1}{5}$.

- (a) Calcula $\mathbb{P}(X > 5)$.
- (b) Una empresa ya lleva 3 años abierta. Usando la falta de memoria, calcula la probabilidad de que dure al menos 5 años más, es decir $\mathbb{P}(X > 3 + 5 \mid X > 3)$.

(Usa $e^{-1} \approx 0.368$.)

Solución

- (a) $\mathbb{P}(X > 5) = e^{-(1/5) \cdot 5} = e^{-1} \approx 0.368$.
- (b) Por la propiedad de *falta de memoria* de la exponencial, $\mathbb{P}(X > 3 + 5 \mid X > 3) = \mathbb{P}(X > 5) = e^{-1} \approx 0.368$. El pasado no importa: la espera adicional “empieza de cero”.

Ejercicio 1.16 (Media muestral: $\mathbb{E}[\bar{X}]$ y $\text{Var}(\bar{X})$). Sea $X_1, \dots, X_n$ una muestra aleatoria i.i.d. con media μ y varianza σ^2 , y sea $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

- (a) Demuestra que $\mathbb{E}[\bar{X}] = \mu$.
- (b) Demuestra que $\text{Var}(\bar{X}) = \sigma^2/n$.

Solución

(a) Por linealidad, $\mathbb{E}[\bar{X}] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n} \cdot n\mu = \mu$.

(b) Por independencia la varianza de la suma es la suma de varianzas, y $\text{Var}(aX) = a^2 \text{Var}(X)$:

$$\text{Var}(\bar{X}) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

Ejercicio 1.17 (Error estándar y efecto de n). El gasto diario en transporte tiene $\mu = 12$ y $\sigma = 4$ soles.

(a) Calcula el error estándar de $\bar{X}$ con $n = 16$.

(b) ¿Cuánto vale con $n = 64$? ¿Qué factor cambió respecto de (a) y por qué?

Solución

(a) $\text{SD}(\bar{X}) = \frac{4}{\sqrt{16}} = \frac{4}{4} = 1$ sol.

(b) $\text{SD}(\bar{X}) = \frac{4}{\sqrt{64}} = \frac{4}{8} = 0.5$ soles. Se redujo a la *mitad*: al cuadruplicar n ($16 \rightarrow 64$), el error estándar baja en factor $\sqrt{4} = 2$, porque decrece como $1/\sqrt{n}$.

Ejercicio 1.18 (Cálculo con Z y la Normal). Los tiempos X de respuesta de una ambulancia son $\mathcal{N}(\mu = 10, \sigma = 2)$ minutos.

(a) Calcula $\mathbb{P}(X < 12)$. (Usa $\mathbb{P}(Z < 1) \approx 0.84$.)

(b) Calcula $\mathbb{P}(8 < X < 14)$. (Usa $\mathbb{P}(Z < 1) \approx 0.84$, $\mathbb{P}(Z < 2) \approx 0.975$.)

Solución

(a) $z = \frac{12 - 10}{2} = 1$, así $\mathbb{P}(X < 12) = \mathbb{P}(Z < 1) \approx 0.84$.

(b) $\mathbb{P}(8 < X < 14) = \mathbb{P}\left(\frac{8-10}{2} < Z < \frac{14-10}{2}\right) = \mathbb{P}(-1 < Z < 2) = \mathbb{P}(Z < 2) - \mathbb{P}(Z < -1) = 0.975 - (1 - 0.84) = 0.975 - 0.16 = 0.815$.

Ejercicio 1.19 (Suma de variables y media muestral en conteo). En una encuesta, cada persona responde X_i horas de estudio semanal, i.i.d. con $\mu = 10$ y $\sigma = 5$ horas. Se entrevista a $n = 25$ personas.

(a) Calcula $\mathbb{E}[\bar{X}]$ y $\text{SD}(\bar{X})$.

(b) Calcula $\mathbb{E}[S]$ y $\text{SD}(S)$ para la suma total $S = \sum_{i=1}^{25} X_i$.

Solución

- (a) $\mathbb{E}[\bar{X}] = \mu = 10$; $\text{SD}(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{25}} = \frac{5}{5} = 1$ hora.
- (b) $\mathbb{E}[S] = n\mu = 25 \cdot 10 = 250$; $\text{Var}(S) = n\sigma^2 = 25 \cdot 25 = 625$, así $\text{SD}(S) = \sqrt{625} = 25$ horas. (Nota: la suma crece su SD como $\sqrt{n}$, la media la reduce como $1/\sqrt{n}$.)

Nivel 3 – Aplicación

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

Ejercicio 1.20 (Aproximación Normal a una proporción muestral). En una población, la proporción que apoya una política es $p = 0.5$. Se encuesta a $n = 100$ personas; sea $\hat{p}$ la proporción muestral que apoya. Para n grande, $\hat{p} \approx \mathcal{N}(p, p(1-p)/n)$.

- (a) Calcula el error estándar de $\hat{p}$.
- (b) Usando la regla empírica, da un intervalo que contenga $\hat{p}$ con probabilidad ≈ 0.95 .

Solución

- (a) $\text{SD}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.5 \cdot 0.5}{100}} = \sqrt{0.0025} = 0.05$.
- (b) $\approx 95\%$ cae en $p \pm 2 \text{SD}(\hat{p}) = 0.5 \pm 2(0.05) = 0.5 \pm 0.10$, es decir $[0.40, 0.60]$.

Ejercicio 1.21 (Varianza de una suma de independientes). Tres encuestadoras independientes reportan errores aleatorios X_1, X_2, X_3 , cada uno con $\mathbb{E}[X_i] = 0$ y $\text{Var}(X_i) = 9$. Se promedia: $\bar{X} = \frac{1}{3}(X_1 + X_2 + X_3)$ y se considera la suma $S = X_1 + X_2 + X_3$.

- (a) Calcula $\text{Var}(S)$.
- (b) Calcula $\text{Var}(\bar{X})$ y compáralo con $\text{Var}(X_i)$.

Solución

- (a) Por independencia, $\text{Var}(S) = \text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) = 9 + 9 + 9 = 27$.
- (b) $\text{Var}(\bar{X}) = \frac{\text{Var}(S)}{3^2} = \frac{27}{9} = 3$ (también $= \sigma^2/n = 9/3$). Promediar reduce la varianza: de 9 a 3.

Ejercicio 1.22 (Aproximación Normal a un conteo Binomial). En una ciudad, el 40% de los hogares recicla. Se toma una muestra de $n = 100$ hogares independientes; $X \sim \text{Bin}(100, 0.4)$ es el número que recicla. Para n grande, $X \approx \mathcal{N}(np, np(1-p))$.

- (a) Calcula $\mathbb{E}[X]$ y $\text{SD}(X)$.
- (b) Usando la aproximación Normal y la regla empírica, halla un intervalo que contenga X con probabilidad ≈ 0.95 .

Solución

- (a) $\mathbb{E}[X] = np = 100 \cdot 0.4 = 40$; $\text{Var}(X) = np(1-p) = 100 \cdot 0.4 \cdot 0.6 = 24$, así $\text{SD}(X) = \sqrt{24} \approx 4.9$.
- (b) $\approx 95\%$ cae en $\mu \pm 2\sigma = 40 \pm 2(4.9) = 40 \pm 9.8$, es decir $[30.2, 49.8]$, aprox. entre 30 y 50 hogares.

Ejercicio 1.23 (Aproximación de Poisson a la Binomial (eventos raros)). En un censo de $n = 1\,000$ personas, la probabilidad de una condición rara es $p = 0.002$. Sea $X \sim \text{Bin}(1\,000, 0.002)$ el número de casos.

- (a) Justifica que $X \approx \text{Pois}(\lambda)$ con $\lambda = np$ y da λ .
- (b) Aproxima $\mathbb{P}(X = 0)$ y $\mathbb{P}(X = 2)$ con la Poisson, y compara $\mathbb{P}(X = 0)$ con el valor Binomial exacto $(1-p)^n$. (Usa $e^{-2} \approx 0.1353$.)

Solución

- (1) Como n es grande, p chico y np moderado, la Binomial se aproxima por una **Poisson** con $\lambda = np = 1\,000 \cdot 0.002 = 2$.
- (2) Poisson: $\mathbb{P}(X = 0) = e^{-2} \approx 0.1353$; $\mathbb{P}(X = 2) = \frac{2^2}{2!}e^{-2} = 2e^{-2} \approx 0.2707$.
- Binomial exacta: $\mathbb{P}(X = 0) = (0.998)^{1000}$. Como $(0.998)^{1000} = e^{1000 \ln(0.998)} \approx e^{1000(-0.0020020)} = e^{-2.002} \approx 0.1351$. La aproximación Poisson (0.1353) es excelente.

Ejercicio 1.24 (TLC: probabilidad sobre la media muestral). El tiempo de viaje al trabajo en una ciudad tiene $\mu = 45$ min y $\sigma = 15$ min (no necesariamente normal). Se toma una muestra aleatoria de $n = 225$ personas. Por el TLC, $\bar{X} \approx \mathcal{N}(\mu, \sigma^2/n)$.

- (a) Calcula el error estándar de $\bar{X}$.
- (b) Calcula $\mathbb{P}(\bar{X} > 47)$. (Usa $\mathbb{P}(Z < 2) \approx 0.975$.)

Solución

- (a) $\text{SD}(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{225}} = \frac{15}{15} = 1$ min.
- (b) $z = \frac{47 - 45}{1} = 2$, así $\mathbb{P}(\bar{X} > 47) = \mathbb{P}(Z > 2) = 1 - \mathbb{P}(Z < 2) \approx 1 - 0.975 = 0.025$. La distribución de X podía ser cualquiera: el TLC nos permite usar la Normal porque n es grande.

Ejercicio 1.25 (TLC: probabilidad sobre una suma). Una ONG reparte $n = 100$ bonos independientes; el monto X_i de cada bono tiene $\mu = 50$ y $\sigma = 20$ soles. Sea $S = \sum_{i=1}^{100} X_i$ el gasto total. Por el TLC, $S \approx \mathcal{N}(n\mu, n\sigma^2)$.

- (a) Calcula $\mathbb{E}[S]$ y $\text{SD}(S)$.
- (b) Calcula $\mathbb{P}(S > 5\,200)$. (Usa $\mathbb{P}(Z < 1) \approx 0.84$.)

Solución

(a) $\mathbb{E}[S] = 100 \cdot 50 = 5\,000$; $\text{Var}(S) = 100 \cdot 20^2 = 40\,000$, así $\text{SD}(S) = \sqrt{40\,000} = 200$ soles.

(b) $z = \frac{5\,200 - 5\,000}{200} = 1$, así $\mathbb{P}(S > 5\,200) = \mathbb{P}(Z > 1) = 1 - 0.84 = 0.16$.

Ejercicio 1.26 (Suma de Normales independientes). Dos regiones tienen ingresos medios mensuales independientes $X \sim \mathcal{N}(1\,500, 200^2)$ e $Y \sim \mathcal{N}(1\,200, 150^2)$ (soles). Recuerda que toda combinación lineal de normales independientes es normal.

(a) Da la distribución de la suma $T = X + Y$ (media y varianza).

(b) Da la distribución de la diferencia $D = X - Y$ y calcula $\mathbb{P}(D > 550)$. (Usa $\mathbb{P}(Z < 1) \approx 0.84$.)

Solución

(a) $\mathbb{E}[T] = 1\,500 + 1\,200 = 2\,700$; por independencia $\text{Var}(T) = 200^2 + 150^2 = 40\,000 + 22\,500 = 62\,500$. Así $T \sim \mathcal{N}(2\,700, 62\,500)$ con $\text{SD}(T) = 250$.

(b) $\mathbb{E}[D] = 1\,500 - 1\,200 = 300$; $\text{Var}(D) = 200^2 + 150^2 = 62\,500$ (la varianza de la diferencia de independientes suma), así $\text{SD}(D) = 250$ y $D \sim \mathcal{N}(300, 62\,500)$. Entonces $z = \frac{550 - 300}{250} = 1$ y $\mathbb{P}(D > 550) = \mathbb{P}(Z > 1) = 1 - 0.84 = 0.16$.

Nivel 4 – Reto

Difícil: integra varios temas en un mismo problema.

Ejercicio 1.27 (Tamaño de muestra para un error estándar dado). El ingreso tiene $\sigma = 600$ soles. Quieres estimar la media con un error estándar de a lo sumo 20 soles.

(1) Plantea la condición $\sigma/\sqrt{n} \leq 20$ y despeja n .

(2) Si además quieres que el “margen de error” al 95 % (aprox. $2 \text{SD}(\bar{X})$) sea ≤ 20 soles, ¿qué n necesitas?

Solución

(1) $\frac{600}{\sqrt{n}} \leq 20 \iff \sqrt{n} \geq \frac{600}{20} = 30 \iff n \geq 900$. Se necesita $n = 900$.

(2) Ahora $2 \cdot \frac{600}{\sqrt{n}} \leq 20 \iff \frac{600}{\sqrt{n}} \leq 10 \iff \sqrt{n} \geq 60 \iff n \geq 3\,600$. Reducir el margen a la mitad cuadruplica el n requerido.

Ejercicio 1.28 (Aproximación Normal a la Binomial con corrección por continuidad). Se lanza una moneda justa $n = 100$ veces; $X \sim \text{Bin}(100, 0.5)$ es el número de caras.

- (1) Da $\mu = \mathbb{E}[X]$ y $\sigma = \text{SD}(X)$.
- (2) Aproxima $\mathbb{P}(X \leq 45)$ con la Normal *usando corrección por continuidad* (usa 45.5). (Usa $\mathbb{P}(Z < 0.9) \approx 0.816$.)

Solución

- (1) $\mu = np = 50$; $\text{Var}(X) = np(1-p) = 100 \cdot 0.5 \cdot 0.5 = 25$, así $\sigma = 5$.
- (2) Con corrección por continuidad, $\mathbb{P}(X \leq 45) \approx \mathbb{P}(Y \leq 45.5)$ donde $Y \sim \mathcal{N}(50, 25)$. Entonces $z = \frac{45.5 - 50}{5} = \frac{-4.5}{5} = -0.9$, así $\mathbb{P}(X \leq 45) \approx \mathbb{P}(Z \leq -0.9) = 1 - \mathbb{P}(Z < 0.9) = 1 - 0.816 = 0.184$. La corrección (45.5 en vez de 45) mejora la aproximación porque la Binomial es discreta y la Normal continua.

Ejercicio 1.29 (Insigadez de la media muestral y de un estimador escalado). Sea $X_1, \dots, X_n$ i.i.d. $\text{Unif}[0, \theta]$ (este caso aparece en clase). El parámetro desconocido es θ . Considera el estimador $\hat{\theta} = 2\bar{X}$.

- (1) Recordando que $\mathbb{E}[X_i] = \theta/2$, demuestra que $\hat{\theta} = 2\bar{X}$ es *insesgado* para θ (es decir $\mathbb{E}[\hat{\theta}] = \theta$).
- (2) Calcula $\text{Var}(\hat{\theta})$ usando $\text{Var}(X_i) = \theta^2/12$. ¿Qué pasa cuando $n \rightarrow \infty$?

Solución

- (1) $\mathbb{E}[\bar{X}] = \mathbb{E}[X_i] = \theta/2$, así $\mathbb{E}[\hat{\theta}] = \mathbb{E}[2\bar{X}] = 2 \cdot \frac{\theta}{2} = \theta$. Por tanto el sesgo es $\mathbb{E}[\hat{\theta}] - \theta = 0$: **insesgado**.
- (2) $\text{Var}(\hat{\theta}) = \text{Var}(2\bar{X}) = 4 \text{Var}(\bar{X}) = 4 \cdot \frac{\sigma^2}{n} = 4 \cdot \frac{\theta^2/12}{n} = \frac{\theta^2}{3n}$. Cuando $n \rightarrow \infty$, $\text{Var}(\hat{\theta}) \rightarrow 0$: el estimador se concentra en θ (es consistente).

Ejercicio 1.30 (Aproximaciones encadenadas: Binomial $\rightarrow$ Poisson y Binomial $\rightarrow$ Normal). Dos escenarios con la *misma* familia Binomial, pero que piden aproximaciones distintas.

- (1) **Eventos raros (Poisson)**. En un padrón de $n = 2000$ votantes, la probabilidad de que una cédula tenga un error de registro raro es $p = 0.0015$, independiente entre cédulas. Sea $X \sim \text{Bin}(2000, 0.0015)$. Justifica la aproximación $X \approx \text{Pois}(\lambda)$, da λ , y aproxima $\mathbb{P}(X \leq 2)$. Compara $\mathbb{P}(X = 0)$ con el valor Binomial exacto $(1-p)^n$.
- (2) **Muchos ensayos balanceados (Normal con corrección por continuidad)**. En una elección reñida, de $n = 400$ votos independientes cada uno favorece a la candidata A con probabilidad $p = 0.5$. Sea $X \sim \text{Bin}(400, 0.5)$ el número de votos para A. Aproxima $\mathbb{P}(X \leq 190)$ con la Normal *usando corrección por continuidad* (190.5), y compárala con la aproximación *sin* corrección.

(3) Discute: ¿por qué en (1) NO conviene la Normal y en (2) NO conviene la Poisson?

(Usa $e^{-3} \approx 0.0498$; $\mathbb{P}(Z < 0.95) \approx 0.8289$; $\mathbb{P}(Z < 1) \approx 0.8413$. Valor exacto de referencia: $\mathbb{P}(X \leq 2) \approx 0.4230$ en (1); $\mathbb{P}(X \leq 190) \approx 0.1711$ en (2).)

Solución

(1) Binomial $\rightarrow$ Poisson. Con n grande, p pequeño y np moderado, la regla de eventos raros da $X \approx \text{Pois}(\lambda)$ con $\lambda = np = 2000 \cdot 0.0015 = 3$. Entonces

$$\mathbb{P}(X = 0) \approx e^{-3} = 0.0498, \quad \mathbb{P}(X = 1) \approx 3e^{-3} = 0.1494,$$

$$\mathbb{P}(X = 2) \approx \frac{3^2}{2!} e^{-3} = 4.5 e^{-3} = 0.2240,$$

así $\mathbb{P}(X \leq 2) \approx 0.0498 + 0.1494 + 0.2240 = 0.4232$ (exacto Binomial 0.4230: casi idéntico). Para $\mathbb{P}(X = 0)$ el valor Binomial exacto es $(1 - p)^n = (0.9985)^{2000} = e^{2000 \ln 0.9985} \approx e^{-3.002} \approx 0.0496$, que coincide con $e^{-3} = 0.0498$. La aproximación Poisson es **excelente**.

(2) Binomial $\rightarrow$ Normal con corrección. Aquí $p = 0.5$ no es chico y $\lambda = np = 200$ es enorme: la Poisson no aplica. La Normal sí: $\mu = np = 200$, $\sigma = \sqrt{np(1 - p)} = \sqrt{400 \cdot 0.5 \cdot 0.5} = \sqrt{100} = 10$. Con corrección por continuidad (la Binomial es discreta, ≤ 190 incluye al 190):

$$\mathbb{P}(X \leq 190) \approx \mathbb{P}\left(Z \leq \frac{190.5 - 200}{10}\right) = \mathbb{P}(Z \leq -0.95) = 1 - 0.8289 = 0.1711,$$

que coincide con el exacto 0.1711. Sin corrección daría $z = \frac{190 - 200}{10} = -1.0$ y $\mathbb{P}(Z \leq -1) = 1 - 0.8413 = 0.1587$: peor por casi 0.012. La corrección (sumar 0.5 al borde) mejora claramente la aproximación.

(3) Cuándo cada una. La *Poisson* sirve para eventos raros ($p \rightarrow 0$, np moderado): la masa se concentra en valores chicos y la campana normal quedaría mal por su asimetría. La *Normal* sirve cuando np y $n(1 - p)$ son ambos grandes (regla práctica ≥ 5 o ≥ 10): la Binomial es casi simétrica y se parece a una campana. En (1), $np = 3$ es chico $\Rightarrow$ Poisson; en (2), $np = n(1 - p) = 200 \Rightarrow$ Normal. (Adaptado de MIT 18.05.)

Ejercicio 1.31 (Suma de variables independientes: Normal, Poisson y Exponencial $\rightarrow$ Erlang). Tres familias cierran bajo la suma de forma distinta. En todos los casos las variables son independientes.

- (1) **Normales.** El gasto mensual de dos programas sociales es $X \sim \mathcal{N}(1200, 150^2)$ e $Y \sim \mathcal{N}(900, 120^2)$ (soles). Da la distribución exacta de $T = X + Y$ (familia, media y varianza) y su SD.
- (2) **Poisson.** A una central de ayuda llegan llamadas de la línea 1 con $X \sim \text{Pois}(3)$ por hora y de la línea 2 con $Y \sim \text{Pois}(5)$ por hora. Da la distribución de $X + Y$ y calcula $\mathbb{E}[X + Y]$ y $\text{Var}(X + Y)$.
- (3) **Exponenciales $\rightarrow$ Erlang/Gamma.** Un trámite tiene 4 etapas independientes; cada etapa dura $T_i \sim \text{Exp}(\lambda = \frac{1}{4})$ horas (media 4 h). El tiempo total es $S = T_1 + T_2 + T_3 + T_4$. Identifica la familia de S (Gamma/Erlang con parámetros) y da $\mathbb{E}[S]$ y $\text{Var}(S)$.

Solución

(1) Toda combinación lineal de normales independientes es normal. La suma: media $\mathbb{E}[T] = 1\,200 + 900 = 2\,100$; varianza (suma de varianzas por independencia) $\text{Var}(T) = 150^2 + 120^2 = 22\,500 + 14\,400 = 36\,900$. Así $T \sim \mathcal{N}(2\,100, 36\,900)$ con $\text{SD}(T) = \sqrt{36\,900} \approx 192.1$ soles.

(2) La suma de Poissons independientes es Poisson con parámetro la suma de tasas: $X + Y \sim \text{Pois}(3 + 5) = \text{Pois}(8)$. Luego $\mathbb{E}[X + Y] = 8$ y $\text{Var}(X + Y) = 8$ (en la Poisson media y varianza coinciden).

(3) La suma de k exponenciales i.i.d. de tasa λ es una **Gamma** (Erlang) de forma k y tasa λ : $S \sim \text{Gamma}(k = 4, \lambda = \frac{1}{4})$. Su media y varianza:

$$\mathbb{E}[S] = \frac{k}{\lambda} = \frac{4}{1/4} = 16 \text{ h}, \quad \text{Var}(S) = \frac{k}{\lambda^2} = \frac{4}{(1/4)^2} = 4 \cdot 16 = 64, \quad \text{SD}(S) = 8 \text{ h}.$$

(También sale de $\mathbb{E}[T_i] = 1/\lambda = 4$, $\text{Var}(T_i) = 1/\lambda^2 = 16$, sumando 4 etapas.) (*Adaptado de MIT 18.05.*)

Ejercicio 1.32 (Método de momentos vs. MLE para la exponencial). Se observa el tiempo entre quejas ciudadanas: una muestra i.i.d. $X_1, \dots, X_n \sim \text{Exp}(\lambda)$ con densidad $f(x | \lambda) = \lambda e^{-\lambda x}$, $x \geq 0$. El parámetro desconocido es la tasa λ .

- (1) **Momentos.** Iguala el primer momento teórico $\mathbb{E}[X] = 1/\lambda$ al momento muestral $\bar{X}$ y despeja el estimador $\hat{\lambda}_{\text{MoM}}$.
- (2) **MLE.** Escribe la log-verosimilitud $\ell(\lambda) = \sum_i [\ln \lambda - \lambda x_i]$, deriva, iguala a cero y obtén $\hat{\lambda}_{\text{MLE}}$.
- (3) Compara ambos estimadores. Para la muestra $\{2.1, 3.4, 1.0, 4.8, 2.9\}$ (en días), calcula el valor numérico.

Solución

(1) **Momentos.** $\mathbb{E}[X] = 1/\lambda$. Igualando a $\bar{X}$: $\frac{1}{\lambda} = \bar{X} \Rightarrow \hat{\lambda}_{\text{MoM}} = \frac{1}{\bar{X}}$.

(2) **MLE.** Verosimilitud $L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i}$. Log-verosimilitud $\ell(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^n x_i$. Derivando e igualando a cero:

$$\ell'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0 \implies \hat{\lambda}_{\text{MLE}} = \frac{n}{\sum x_i} = \frac{1}{\bar{X}}.$$

($\ell'' = -n/\lambda^2 < 0$, es máximo.)

(3) **Comparación.** Para la exponencial, **ambos coinciden**: $\hat{\lambda}_{\text{MoM}} = \hat{\lambda}_{\text{MLE}} = 1/\bar{X}$. Con la muestra, $\sum x_i = 2.1 + 3.4 + 1.0 + 4.8 + 2.9 = 14.2$, $\bar{X} = 14.2/5 = 2.84$ días, así $\hat{\lambda} = 1/2.84 \approx 0.352$ quejas/día. (Que coincidan es especial de la exponencial; en otras familias difieren, ver el siguiente nivel.) (*Adaptado de MIT 18.05.*)

Nivel 5 – Reto+

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

Ejercicio 1.33 (MLE de la Uniforme $[0, \theta]$ y su sesgo). La duración (en años) de mandatos en un cargo se modela $X_1, \dots, X_n$ i.i.d. $\text{Unif}[0, \theta]$, con densidad $f(x | \theta) = 1/\theta$ para $0 \leq x \leq \theta$. El parámetro desconocido es θ (la duración máxima posible).

- (1) Escribe la verosimilitud $L(\theta)$ y argumenta por qué el MLE NO se obtiene derivando, sino que es $\hat{\theta}_{\text{MLE}} = \max(X_1, \dots, X_n)$.
- (2) La densidad del máximo $M = \max_i X_i$ es $f_M(m) = n m^{n-1} / \theta^n$ para $0 \leq m \leq \theta$. Demuestra que $\mathbb{E}[M] = \frac{n}{n+1} \theta$ y deduce el sesgo de $\hat{\theta}_{\text{MLE}}$.
- (3) Propone la corrección insesgada $\hat{\theta}_c = c M$ (da c) y compara cualitativamente con el estimador de momentos $\hat{\theta}_{\text{MoM}} = 2\bar{X}$.

Solución

(1) $L(\theta) = \prod_{i=1}^n \frac{1}{\theta} \mathbf{1}\{0 \leq x_i \leq \theta\} = \frac{1}{\theta^n} \mathbf{1}\{\theta \geq \max_i x_i\}$. Para $\theta < \max_i x_i$ la verosimilitud es 0 (algún dato caería fuera del soporte). Para $\theta \geq \max_i x_i$, $L = \theta^{-n}$ es *decreciente* en θ , así que se maximiza tomando el menor θ admisible: $\hat{\theta}_{\text{MLE}} = \max(X_1, \dots, X_n) = M$. Derivar no sirve porque el máximo está en el *borde* del rango admisible, no en un punto interior con derivada cero.

(2) $\mathbb{E}[M] = \int_0^\theta m \cdot \frac{n m^{n-1}}{\theta^n} dm = \frac{n}{\theta^n} \int_0^\theta m^n dm = \frac{n}{\theta^n} \cdot \frac{\theta^{n+1}}{n+1} = \frac{n}{n+1} \theta$. Sesgo: $\mathbb{E}[\hat{\theta}_{\text{MLE}}] - \theta = \frac{n}{n+1} \theta - \theta = -\frac{\theta}{n+1} < 0$. El MLE **subestima** sistemáticamente θ (es lógico: el máximo de la muestra nunca supera a θ). El sesgo $\rightarrow 0$ cuando $n \rightarrow \infty$ (es asintóticamente insesgado).

(3) Buscamos c con $\mathbb{E}[c M] = \theta$: $c \cdot \frac{n}{n+1} \theta = \theta \Rightarrow c = \frac{n+1}{n}$. Así $\hat{\theta}_c = \frac{n+1}{n} \max(X_i)$ es **insesgado**. Frente a $\hat{\theta}_{\text{MoM}} = 2\bar{X}$ (también insesgado, pues $\mathbb{E}[2\bar{X}] = 2 \cdot \theta/2 = \theta$): el corregido del máximo usa solo el dato extremo pero tiene varianza mucho menor ($\propto \theta^2/n^2$) que el de momentos ($\propto \theta^2/n$); para n grande el estimador basado en el máximo es preferible. (*Adaptado de MIT 18.05.*)

Ejercicio 1.34 (Simulación por transformada inversa). Para simular en computadora una variable con distribución dada, se parte de $U \sim \text{Unif}[0, 1]$ (que todo lenguaje genera) y se aplica la inversa de la cdf. El *método de la transformada inversa* dice: si F es la cdf y $U \sim \text{Unif}[0, 1]$, entonces $X = F^{-1}(U)$ tiene cdf F .

- (1) **Exponencial**. Para $\text{Exp}(\lambda)$, $F(x) = 1 - e^{-\lambda x}$. Despeja $x = F^{-1}(u)$ y da la fórmula generadora.

- (2) Con $\lambda = \frac{1}{2}$ y el número aleatorio $u = 0.3$, calcula el valor simulado x . (Usa $\ln(0.7) \approx -0.3567$.)
- (3) **Discreta (Bernoulli)**. Explica cómo generar $X \sim \text{Bernoulli}(p)$ a partir de U , y aplica con $p = 0.4$, $u = 0.3$.

Solución

(1) Despejamos $u = 1 - e^{-\lambda x} \Rightarrow e^{-\lambda x} = 1 - u \Rightarrow -\lambda x = \ln(1 - u) \Rightarrow$

$$x = F^{-1}(u) = -\frac{\ln(1 - u)}{\lambda}.$$

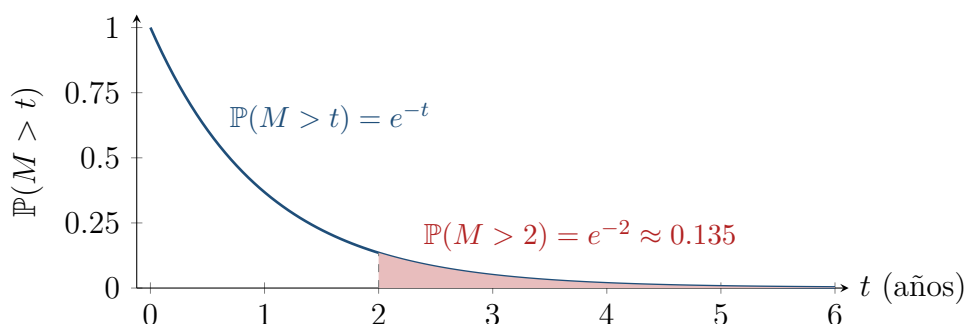
(Como $1 - U$ también es $\text{Unif}[0, 1]$, suele usarse la forma equivalente $x = -\ln(u)/\lambda$.)

(2) Con $\lambda = \frac{1}{2}$, $u = 0.3$: $x = -\frac{\ln(1 - 0.3)}{1/2} = -2\ln(0.7) = -2(-0.3567) = 0.7134$. El valor simulado es ≈ 0.71 (años/minutos según la unidad).

(3) La cdf de $\text{Bernoulli}(p)$ vale 0 hasta 0, salta a $1 - p$ en $x = 0$ y a 1 en $x = 1$. La inversa generalizada da la regla: $X = 0$ si $U \leq 1 - p$, y $X = 1$ si $U > 1 - p$. (Equivalente y más común: $X = 1$ si $U < p$, $X = 0$ si no.) Con $p = 0.4$, $u = 0.3$: como $0.3 < 0.4$, se obtiene $X = 1$ (“éxito”). (Adaptado de MIT 18.05.)

Ejercicio 1.35 (Estadísticos de orden: el mínimo de exponenciales). Un sistema social tiene $n = 5$ componentes independientes; el tiempo de falla de cada uno es $T_i \sim \text{Exp}(\lambda = 0.2)$ por año (media 5 años). El sistema falla cuando falla el *primero*, así que su vida es $M = \min(T_1, \dots, T_5)$.

- (1) Demuestra que $\mathbb{P}(M > t) = \prod_i \mathbb{P}(T_i > t)$ y deduce que $M \sim \text{Exp}(n\lambda)$. Da la tasa y la media de M .
- (2) Calcula $\mathbb{P}(M > 2)$. (Usa $e^{-2} \approx 0.1353$.)
- (3) La figura muestra la cola $\mathbb{P}(M > t) = e^{-t}$ del mínimo, con el área bajo la curva a la derecha de $t = 2$ sombreada. Interpreta por qué el mínimo de muchos componentes “vive mucho menos” que uno solo.



Solución

(1) $M > t$ ocurre sólo si *todos* sobreviven a t . Por independencia,

$$\mathbb{P}(M > t) = \mathbb{P}(T_1 > t, \dots, T_5 > t) = \prod_{i=1}^5 \mathbb{P}(T_i > t) = \prod_{i=1}^5 e^{-\lambda t} = e^{-5\lambda t} = e^{-(n\lambda)t}.$$

Esa es la función de supervivencia de una exponencial de tasa $n\lambda$, así que $M \sim \text{Exp}(n\lambda) = \text{Exp}(5 \cdot 0.2) = \text{Exp}(1)$. Media $\mathbb{E}[M] = 1/(n\lambda) = 1$ año (¡frente a 5 años de un solo componente!).

(2) $\mathbb{P}(M > 2) = e^{-(n\lambda) \cdot 2} = e^{-1 \cdot 2} = e^{-2} \approx 0.1353$, que es el área roja de la figura.

(3) El sistema vive lo que dura su componente más débil. Cuantos más componentes en serie, más “oportunidades” de fallar temprano: la tasa total $n\lambda$ crece con n y la vida media $1/(n\lambda)$ cae. Por eso el mínimo de 5 exponenciales de media 5 tiene media 1: la cola decae cinco veces más rápido. (*Adaptado de MIT 18.05.*)

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