14.310x – Data Analysis for Social Scientists

MicroMasters en Statistics and Data Science (SDS) – MITx

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

Clase 12

Comprehensive Probability Review

Solucionario (Read-Aloud Solutions)

Módulos oficiales del MicroMaster:
M1–M5 – Fundamentals of Probability, Conditional Probability & Bayes, Joint Distributions, Expectation & Variance, Special Distributions, Sample Mean & CLT

B.Sc. Cristian Lazo Quispe
clazoq@uni.pe
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).

11 de agosto de 2026

How to use this document

This is the **read-aloud solutions** companion to the Clase 12 question deck (`slides_clase12.tex`). Each problem below follows the same three-part structure: **theory used** (the tool being invoked), **thought process** (how to recognize which tool applies, before any computation), and a **step-by-step solution** with a boxed final answer. Difficulty tiers match the slides. **15 core problems** (Easy ×3 Intermediate ×6 Hard ×6), plus **5 bonus problems** (16–20) at the end in case there is extra time.

1. Problem 1 — Committee Counting

Problem 1 – Easy

Class 1 – Counting (combinations)

Setup. A district education office selects a 3-person committee (no distinct roles) from a pool of 7 eligible staff members. How many different committees can be formed?

Theory used

When we choose a subset of k objects from n **without** regard to order, the count is the binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

Thought process

Two signals tell us this is a *combination*, not a *permutation*: (i) the committee members have "equal standing, no distinct roles" — order does not matter — and (ii) nobody can serve twice, so it is selection **without replacement**. That combination of features is exactly what $\binom{n}{k}$ counts.

Step-by-step solution

With $n = 7$ and $k = 3$:

$$\binom{7}{3} = \frac{7!}{3!\,4!} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = \frac{210}{6} = \boxed{35}.$$

There are 35 different 3-person committees. *Sanity check:* $\binom{7}{3} = \binom{7}{4}$ (choosing who is *on* the committee is the same count as choosing who is *left out*); both give 35.

2. Problem 2 — Training and Employment

Problem 2 – Easy

Class 2 – Conditional probability

Setup. Of $n = 200$ program graduates: 120 completed training (90 employed, 30 not) and 80 did not complete it (40 employed, 40 not). Find $\mathbb{P}(\text{Employed} \mid \text{Completed})$, $\mathbb{P}(\text{Employed})$, and say whether the two events are independent.

Theory used

Conditional probability restricts the sample space to the conditioning event: $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$. Two events are **independent** iff $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ (equivalently $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$).

Thought process

This is pure "read the table": the row "Completed training" is our conditioning event B (120 people), and within that row we read off how many are employed. Then we compare that conditional probability to the *overall* (unconditional) employment rate — if restricting to "Completed" changed the employment rate, the two events are dependent.

Step-by-step solution

(a) Restrict to the 120 who completed training; 90 of them are employed:

$$\mathbb{P}(\text{Employed} \mid \text{Completed}) = \frac{90}{120} = \boxed{0.75}.$$

(b) Overall employment rate, using the full sample of 200 (130 employed in total, $90 + 40$):

$$\mathbb{P}(\text{Employed}) = \frac{130}{200} = \boxed{0.65}.$$

(c) Since $0.75 \neq 0.65$, knowing someone completed training **changes** our assessment of their employment probability, so **Employed and Completed are not independent** (completing training is associated with a higher employment rate).

3. Problem 3 — Field-Worker Pay

Problem 3 – Easy

Class 4 – Linearity of expectation

Setup. A field worker's weekly pay is $50+8X$ soles, where X is the (random) number of households visited, with $\mathbb{E}[X] = 18$. Find $\mathbb{E}[50 + 8X]$.

Theory used

Linearity of expectation: for constants a, b, $\mathbb{E}[aX + b] = a\,\mathbb{E}[X] + b$. This holds **regardless of the distribution of** X and even when X and other random variables are dependent — it never requires independence.

Thought process

We are never told the full distribution of X (how many households, with what probabilities) — only its expectation. That is a strong hint: the problem is designed so you don't need the distribution at all, only linearity.

Step-by-step solution

$$\mathbb{E}[50 + 8X] = 50 + 8\,\mathbb{E}[X] = 50 + 8(18) = 50 + 144 = \boxed{194 \text{ soles}}.$$

Expected weekly payment is S/. 194.

4. Problem 4 — Newborn Referrals

Problem 4 – Intermediate

Class 2 – Bayes' theorem

Setup. Births: 50 % urban, 30 % peri-urban, 20 % rural. Referral probabilities: 2 %, 5 %, 9 % respectively. Given a newborn was referred, find the probability it came from a rural clinic.

Theory used

Bayes' theorem with a partition $\{H_1, \ldots, H_k\}$ of the sample space:

$$\mathbb{P}(H_i \mid E) = \frac{\mathbb{P}(H_i)\,\mathbb{P}(E \mid H_i)}{\sum_j \mathbb{P}(H_j)\,\mathbb{P}(E \mid H_j)}.$$

The denominator is the **law of total probability** applied to E.

Thought process

We are asked to reverse a conditional: we know $\mathbb{P}(\text{Referral} \mid \text{clinic type})$ but we want $\mathbb{P}(\text{clinic type} \mid \text{Referral})$. That reversal is the signature of a Bayes problem. With three clinic types, the "prior" partition has three pieces, so the denominator sums three terms.

Step-by-step solution

Let U, P, R denote urban/peri-urban/rural and E = referral.

$$\mathbb{P}(E) = \underbrace{(0.50)(0.02)}_{U} + \underbrace{(0.30)(0.05)}_{P} + \underbrace{(0.20)(0.09)}_{R} = 0.010 + 0.015 + 0.018 = 0.043.$$

$$\mathbb{P}(R \mid E) = \frac{(0.20)(0.09)}{0.043} = \frac{0.018}{0.043} = \frac{18}{43} \approx \boxed{0.419}.$$

Even though rural clinics account for only 20 % of *all* births, they account for about 41.9 % of *referrals* — because their referral rate (9 %) is much higher than urban (2 %) or peri-urban (5 %). *Sanity check:* 18/43 + (urban share 10/43) + (peri-urban share 15/43) = 43/43 = 1. ✓

5. Problem 5 — Household Composition

Problem 5 – Intermediate

Class 3 – Joint distributions

Setup. Joint pmf of X (adults, 1 or 2) and Y (children, 0, 1, 2): row X=1: $0.10, 0.15, 0.05$; row X=2: $0.10, 0.25, 0.35$. Find the marginal pmf of Y and check independence.

Theory used

The **marginal** pmf of Y sums the joint pmf over all values of X: $\mathbb{P}(Y=y) = \sum_x \mathbb{P}(X=x, Y=y)$. Independence requires $\mathbb{P}(X=x, Y=y) = \mathbb{P}(X=x)\mathbb{P}(Y=y)$ for *every* pair (x, y) — a **single** mismatch is enough to rule it out.

Thought process

To get the marginal, add *down each column* of the table (summing out X). To check independence, we don't need to check all six cells: finding just *one* cell where the joint probability disagrees with the product of marginals is sufficient to conclude "not independent."

Step-by-step solution

(a) Marginal of Y (sum each column):

$$\mathbb{P}(Y{=}0) = 0.10 + 0.10 = 0.20, \quad \mathbb{P}(Y{=}1) = 0.15 + 0.25 = 0.40, \quad \mathbb{P}(Y{=}2) = 0.05 + 0.35 = 0.40.$$

Check: $0.20 + 0.40 + 0.40 = 1$. ✓

(b) Independence. Marginal of X: $\mathbb{P}(X{=}1) = 0.10 + 0.15 + 0.05 = 0.30$. Compare the cell $(X{=}1, Y{=}0)$:

$$\mathbb{P}(X{=}1, Y{=}0) = 0.10 \quad \text{vs.} \quad \mathbb{P}(X{=}1)\mathbb{P}(Y{=}0) = (0.30)(0.20) = 0.06.$$

Since $0.10 \neq 0.06$, X **and** Y **are not independent**. (Households with only 1 adult have *fewer* children than independence would predict, which matches intuition: single-adult households skew toward fewer children.)

6. Problem 6 — Revenue and Expense Growth

Problem 6 – Intermediate

Class 4 – Covariance & variance

Setup. $\mathrm{Var}(X) = 4$, $\mathrm{Var}(Y) = 9$, $\mathrm{Cov}(X, Y) = -2$. Find $\mathrm{Var}(X - Y)$ and its standard deviation.

Theory used

For any random variables X, Y: $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\,\mathrm{Cov}(X, Y)$. This follows from expanding $\mathrm{Var}(X - Y) = \mathrm{Cov}(X - Y, X - Y)$ and using bilinearity of covariance. Note

the **minus** sign on the covariance term (it would be $+2\,\mathrm{Cov}(X,Y)$ for $\mathrm{Var}(X+Y)$).

Thought process

The key trap here is the sign: many students default to the $\mathrm{Var}(X+Y)$ formula out of habit. Since the question defines net margin growth as a *difference* $(X-Y)$, we need the difference formula, and a **negative** covariance in that formula (subtracting a negative) *adds* to the variance rather than reducing it.

Step-by-step solution

$$\mathrm{Var}(X-Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\,\mathrm{Cov}(X,Y) = 4 + 9 - 2(-2) = 4 + 9 + 4 = \boxed{17}.$$

$$\mathrm{SD}(X-Y) = \sqrt{17} \approx \boxed{4.12}.$$

Intuition: because revenue and expense growth move in *opposite* directions ($\mathrm{Cov} < 0$), their *difference* is *more* variable than either alone, not less — the negative covariance reinforces the swings in $X - Y$ instead of canceling them.

7. Problem 7 — Tele-Health Call Center

Problem 7 – Intermediate

Class 5 – Poisson distribution

Setup. Requests arrive Poisson with $\lambda = 6$ per hour. Find $\mathbb{P}(X = 4)$ and $\mathbb{P}(X \geq 1)$.

Theory used

Poisson pmf: $\mathbb{P}(X = k) = \dfrac{e^{-\lambda}\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$, used to model counts of rare/independent events over a fixed interval (calls, arrivals, defects) given a constant average rate λ.

Thought process

"Rate per unit time" plus "count of events in that interval" is the classic Poisson setup. For part (b), computing $\mathbb{P}(X \geq 1)$ directly would mean summing infinitely many terms ($k = 1, 2, 3, \ldots$); it's far easier to use the complement: $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0)$.

Step-by-step solution

(a)

$$\mathbb{P}(X = 4) = \frac{e^{-6}6^4}{4!} = \frac{e^{-6}(1296)}{24} \approx \boxed{0.1339}.$$

(b)

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - \frac{e^{-6}6^0}{0!} = 1 - e^{-6} \approx 1 - 0.00248 = \boxed{0.9975}.$$

With a rate of 6 per hour, it is nearly certain (99.75 %) that the center gets at least one request in any given hour.

8. Problem 8 — Volunteers in a Row

Problem 8 – Intermediate

Class 1 – Counting with a constraint

Setup. 6 distinct volunteers line up; two specific ones (the co-founders) must stand next to each other. Count the valid arrangements.

Theory used

Grouping trick for "must be adjacent" constraints: glue the two required-adjacent people into a single block. If n items become $n-1$ units after gluing, arrange the units in $(n-1)!$ ways, then multiply by the number of internal orderings of the glued block (here $2! = 2$, since the pair can appear in either order).

Thought process

Without the adjacency requirement this would be a plain 6! permutation problem. The constraint "must stand next to each other" is what signals the block/gluing technique rather than a direct factorial or combination formula. It helps to physically picture the two co-founders as one person wearing a two-person costume.

Step-by-step solution

Treat the two co-founders as one block $\Rightarrow$ we now arrange 5 units (the block + the other 4 volunteers):

$$5! = 120 \text{ ways to order the units.}$$

Within the block, the two co-founders can stand in either order:

$$2! = 2 \text{ internal orderings.}$$

$$\text{Total} = 5! \times 2! = 120 \times 2 = \boxed{240}.$$

Sanity check: out of $6! = 720$ total arrangements of 6 distinct people, exactly $240/720 = 1/3$ have the two co-founders adjacent — which matches the general rule that the probability two particular people are adjacent in a random line of n is $2/n = 2/6 = 1/3$.

9. Problem 9 — Sample Mean of Food Expenditure

Problem 9 – Intermediate

Class 5 – CLT & sample mean

Setup. Population mean S/. 450, SD S/. 120, sample size $n = 64$. Approximate $\mathbb{P}(\bar{X} > 470)$.

Theory used

Central Limit Theorem: for large n, $\bar{X} \overset{\text{approx}}{\sim} \mathcal{N}\left(\mu,\ \dfrac{\sigma^2}{n}\right)$, so the standard error is $\text{SE} = \sigma/\sqrt{n}$, and $Z = \dfrac{\bar{X} - \mu}{\text{SE}} \overset{\text{approx}}{\sim} \mathcal{N}(0,1)$.

Thought process

We are asked about the **sample mean** $\bar{X}$, not a single household's expenditure X — that is the cue to use the CLT and divide σ by $\sqrt{n}$ (not use σ directly). With $n = 64 = 8^2$, the square root is clean, which is a hint the problem was designed to have a tidy standard error.

Step-by-step solution

$$\text{SE} = \frac{\sigma}{\sqrt{n}} = \frac{120}{\sqrt{64}} = \frac{120}{8} = 15.$$

$$Z = \frac{470 - 450}{15} = \frac{20}{15} = 1.33.$$

$$\mathbb{P}(\bar{X} > 470) = \mathbb{P}(Z > 1.33) = 1 - \Phi(1.33) \approx 1 - 0.9088 = \boxed{0.0912}.$$

About a 9.1 % chance the sample mean overshoots S/. 470, even though *individual* households often spend far more than that (recall $\sigma = 120$ for individuals, vs. SE = 15 for the mean of 64 of them) — averaging shrinks the spread.

10. Problem 10 — Splitting an NGO Budget

Problem 10 – Hard

Class 3 – Joint densities

Setup. (X, Y) supported on the triangle $\{x \geq 0, y \geq 0, x + y \leq 1\}$ with $f(x, y) = k(x + 2y)$. Find k, $f_X(x)$, $\mathbb{P}(Y > 2X)$, and $\mathbb{E}[X]$.

Theory used

For a continuous joint density: (i) it must integrate to 1 over its support, which pins down the normalizing constant k; (ii) the marginal $f_X(x) = \int f(x, y)\, dy$, integrating out y over its *valid range given* x; (iii) probabilities of regions are double integrals of f over that region; (iv) $\mathbb{E}[X] = \int x\, f_X(x)\, dx$.

Thought process

The tricky part of triangular supports is getting the **limits of integration** right: for a fixed $x \in [0, 1]$, y ranges over $[0,\ 1 - x]$ (not $[0, 1]$) because of the constraint $x + y \leq 1$. For part (c), $\mathbb{P}(Y > 2X)$ also intersects with the triangle, so we first need to find *where* the line $y = 2x$ crosses the hypotenuse $x + y = 1$: setting $2x = 1 - x$ gives $x = 1/3$ — that becomes the new upper limit for x.

Step-by-step solution

(a) Find k. Integrate over the triangle, y from 0 to $1-x$, then x from 0 to 1:

$$\int_0^{1-x} k(x+2y)\,dy = k\Big[xy+y^2\Big]_0^{1-x} = k\big[x(1-x)+(1-x)^2\big] = k(1-x)\big[x+(1-x)\big] = k(1-x).$$

$$\int_0^1 k(1-x)\,dx = k\Big[x-\tfrac{x^2}{2}\Big]_0^1 = k\left(1-\tfrac{1}{2}\right) = \frac{k}{2}.$$

Set equal to 1: $k = \boxed{2}$.

(b) Marginal $f_X(x)$. From the computation above, with $k=2$:

$$f_X(x) = k(1-x) = \boxed{2(1-x)}, \qquad 0 \le x \le 1.$$

(Check: $\int_0^1 2(1-x)\,dx = 1$. ✓)

(c) $\mathbb{P}(Y > 2X)$. Within the triangle, also require $y > 2x$; this region is nonempty only where $2x < 1-x$, i.e. $x < 1/3$. For such x, y ranges from $2x$ to $1-x$:

$$\int_{2x}^{1-x} 2(x+2y)\,dy = 2\Big[xy+y^2\Big]_{2x}^{1-x} = 2\Big[\big(x(1-x)+(1-x)^2\big) - \big(2x^2+4x^2\big)\Big] = 2\big[(1-x)-6x^2\big].$$

$$\int_0^{1/3} (2-2x-12x^2)\,dx = \Big[2x-x^2-4x^3\Big]_0^{1/3} = \frac{2}{3}-\frac{1}{9}-\frac{4}{27} = \frac{18-3-4}{27} = \frac{11}{27} \approx \boxed{0.407}.$$

(d) $\mathbb{E}[X]$.

$$\mathbb{E}[X] = \int_0^1 x\cdot 2(1-x)\,dx = \int_0^1 (2x-2x^2)\,dx = \Big[x^2-\tfrac{2}{3}x^3\Big]_0^1 = 1-\frac{2}{3} = \boxed{\frac{1}{3} \approx 0.333}.$$

Interpretation: on average, about a third of the budget goes to training, with the density $f_X(x) = 2(1-x)$ decreasing in x — large training shares are less likely than small ones, because the density $k(x+2y)$ favors higher y (materials).

11. Problem 11 — Ride-Hailing Trip Times

Problem 11 – Hard

Class 4 – Law of total variance

Setup. Mixture of two driver pools (60 %/40 %), each with its own mean and variance of trip time T. Find $\mathbb{E}[T]$ and $\mathrm{Var}(T)$ unconditionally.

Theory used

Let G be the (unobserved) pool. **Law of total expectation:** $\mathbb{E}[T] = \mathbb{E}\big[\mathbb{E}[T\,|\,G]\big]$. **Law of total variance:**

$$\mathrm{Var}(T) = \underbrace{\mathbb{E}\big[\mathrm{Var}(T\,|\,G)\big]}_{\text{within-group variance}} + \underbrace{\mathrm{Var}\left(\mathbb{E}[T\,|\,G]\right)}_{\text{between-group variance}}.$$

What *is* $\mathrm{Var}(\mathbb{E}[T \mid G])$, concretely? The expression $\mathbb{E}[T \mid G]$ looks like a single number, but it **depends on** G, and G is random — so $\mathbb{E}[T \mid G]$ is itself a random variable (it's a *function* of the random pool G, the same way T^2 or $3T+1$ would be functions of T). Give it a name, $M := \mathbb{E}[T \mid G]$. Since G only takes 2 possible values (experienced / new), M also only takes (at most) 2 possible values: whatever $\mathbb{E}[T \mid G]$ works out to be in each case. Once you see M as an ordinary random variable with a known probability for each of its 2 possible values, $\mathrm{Var}(M)$ is computed **exactly like the variance of any 2-outcome random variable**:

$$\mathrm{Var}(M) = \sum_i \mathbb{P}(M = m_i)\,(m_i - \mathbb{E}[M])^2 \quad \text{or equivalently} \quad \mathrm{Var}(M) = \mathbb{E}[M^2] - (\mathbb{E}[M])^2.$$

Nothing new is happening here beyond "compute the variance of a 2-point random variable" — the only subtlety is realizing that *the conditional mean itself* is the random variable in question.

Thought process

This is a *mixture*: the rider doesn't know the pool, so unconditional uncertainty about T has **two** sources — the randomness *within* each pool (captured by the within-group variance term) *and* the randomness of *which pool* we land in (the between-group term, since the two pools have different average trip times). Only reporting $\mathbb{E}[\mathrm{Var}(T \mid G)]$ and forgetting the second term is the classic mistake here.

Step-by-step solution

$\mathbb{E}[T]$:

$$\mathbb{E}[T] = (0.6)(18) + (0.4)(24) = 10.8 + 9.6 = \boxed{20.4 \text{ min}}.$$

Within-group term $\mathbb{E}[\mathrm{Var}(T \mid G)]$:

$$\mathbb{E}[\mathrm{Var}(T \mid G)] = (0.6)(16) + (0.4)(25) = 9.6 + 10 = 19.6.$$

Between-group term $\mathrm{Var}(\mathbb{E}[T \mid G])$. Following the theory box: name the conditional mean $M := \mathbb{E}[T \mid G]$. Plugging in each value of G, M takes exactly 2 possible values:

$$M = \mathbb{E}[T \mid G = \text{experienced}] = 18 \quad \text{with probability } 0.6,$$

$$M = \mathbb{E}[T \mid G = \text{new}] = 24 \quad \text{with probability } 0.4.$$

So M is just a 2-point random variable, and we already know its mean: $\mathbb{E}[M] = \mathbb{E}\big[\mathbb{E}[T \mid G]\big] = \mathbb{E}[T] = 20.4$ by the law of total expectation (computed just above) — *this is the same* 20.4, not a new number to find. Now apply the ordinary variance formula $\mathrm{Var}(M) = \sum_i \mathbb{P}(M{=}m_i)(m_i - \mathbb{E}[M])^2$ to this 2-point variable, i.e. weight each squared deviation *from that mean* by its probability:

$$\mathrm{Var}(M) = (0.6)\underbrace{(18-20.4)^2}_{\text{experienced's deviation}} + (0.4)\underbrace{(24-20.4)^2}_{\text{new's deviation}}$$

$$= (0.6)(-2.4)^2 + (0.4)(3.6)^2 = (0.6)(5.76) + (0.4)(12.96) = 3.456 + 5.184 = \boxed{8.64}.$$

Cross-check with the shortcut formula $\mathrm{Var}(M) = \mathbb{E}[M^2] - (\mathbb{E}[M])^2$:

$$\mathbb{E}[M^2] = (0.6)(18)^2 + (0.4)(24)^2 = (0.6)(324) + (0.4)(576) = 194.4 + 230.4 = 424.8,$$

$$\mathrm{Var}(M) = 424.8 - (20.4)^2 = 424.8 - 416.16 = 8.64. \quad \checkmark \text{ matches.}$$

Total:

$$\mathrm{Var}(T) = 19.6 + 8.64 = \boxed{28.24}, \qquad \mathrm{SD}(T) = \sqrt{28.24} \approx \boxed{5.31 \text{ min}}.$$

Note $\mathrm{Var}(T) = 28.24 > \mathbb{E}[\mathrm{Var}(T \mid G)] = 19.6$: not knowing the driver pool genuinely adds uncertainty beyond what either pool has on its own.

12. Problem 12 — Sequential Screening for Default Risk

Problem 12 – Hard

Class 2 – Sequential Bayesian updating

Setup. Prior $\mathbb{P}(\text{High}) = 0.30$. Two conditionally independent indicators, both come back "Yes." Find the posterior probability of High-risk two ways: sequentially, and in one shot.

Theory used

Bayes' theorem, and why we can work with "un-normalized weights." For a two-way partition $\{H, L\}$,

$$\mathbb{P}(H \mid E) = \frac{\mathbb{P}(H)\mathbb{P}(E \mid H)}{\mathbb{P}(E)}, \qquad \mathbb{P}(E) = \mathbb{P}(H)\mathbb{P}(E \mid H) + \mathbb{P}(L)\mathbb{P}(E \mid L)$$

(the denominator is just the law of total probability). Notice the denominator $\mathbb{P}(E)$ is *exactly* the sum of the two numerators $\mathbb{P}(H)\mathbb{P}(E \mid H)$ and $\mathbb{P}(L)\mathbb{P}(E \mid L)$ — so instead of computing $\mathbb{P}(E)$ separately, we can compute the two numerators ("un-normalized weights"), and **divide each by their sum**. That's why the solution below never computes $\mathbb{P}(E)$ as its own step.

Why does sequential updating give the same answer as one shot? This is not a coincidence to take on faith — it follows algebraically from conditional independence. After Step 1, treat $\mathbb{P}(H \mid E_1)$ as the new prior and apply Bayes again with E_2:

$$\mathbb{P}(H \mid E_1, E_2) = \frac{\mathbb{P}(H \mid E_1)\,\mathbb{P}(E_2 \mid H, E_1)}{\mathbb{P}(H \mid E_1)\,\mathbb{P}(E_2 \mid H, E_1) + \mathbb{P}(L \mid E_1)\,\mathbb{P}(E_2 \mid L, E_1)}.$$

Because $E_1 \perp\!\!\!\perp E_2 \mid H$ (and likewise given L), knowing E_1 already happened doesn't change the probability of E_2: we may replace $\mathbb{P}(E_2 \mid H, E_1)$ with the simpler $\mathbb{P}(E_2 \mid H)$ (and $\mathbb{P}(E_2 \mid L, E_1)$ with $\mathbb{P}(E_2 \mid L)$). Then substitute $\mathbb{P}(H \mid E_1) = \mathbb{P}(H)\mathbb{P}(E_1 \mid H)/\mathbb{P}(E_1)$ (and the analogous expression for L) — the common factor $1/\mathbb{P}(E_1)$ appears in *every* term, top and bottom, so it **cancels out of the ratio entirely**, leaving

$$\mathbb{P}(H \mid E_1, E_2) = \frac{\mathbb{P}(H)\mathbb{P}(E_1 \mid H)\mathbb{P}(E_2 \mid H)}{\mathbb{P}(H)\mathbb{P}(E_1 \mid H)\mathbb{P}(E_2 \mid H) + \mathbb{P}(L)\mathbb{P}(E_1 \mid L)\mathbb{P}(E_2 \mid L)},$$

which is exactly the one-shot formula with both likelihoods multiplied together. In words: **conditional independence is precisely the assumption that lets E_1's information get "absorbed" into the prior without contaminating how E_2 updates things next.**

Thought process

The problem explicitly asks for *two* routes to the same number, which is a good way to build confidence that sequential updating is not some special trick — it's the *same* Bayes' rule, just applied twice with an intermediate "today's posterior becomes tomorrow's prior" step (proven algebraically in the theory box, not just asserted). The conditional independence assumption is what lets us multiply the two likelihoods together without worrying about how the indicators interact — if E_1 and E_2 were *not* conditionally independent (e.g., missing a payment reminder call made a credit-bureau flag more or less likely, even within the same risk group), the two routes would **not** agree, and we would be forced to use a genuine joint likelihood $\mathbb{P}(E_1, E_2 \mid H)$ instead of a product.

Step-by-step solution

Step 1 — update on Indicator 1. Multiply each hypothesis' *prior* by how likely Indicator 1 (E_1="Yes") is *under that hypothesis* — these products are the "un-normalized weights" from the theory box:

$$\mathbb{P}(H)\mathbb{P}(E_1{=}Y \mid H) = (0.30)(0.60) = 0.18, \qquad \mathbb{P}(L)\mathbb{P}(E_1{=}Y \mid L) = (0.70)(0.15) = 0.105.$$

Their sum, $0.18 + 0.105 = 0.285$, *is* $\mathbb{P}(E_1{=}Y)$ (law of total probability) — we get it for free by adding the two weights, no separate computation needed. Divide each weight by that sum to get the (now properly normalized) posterior:

$$\mathbb{P}(H \mid E_1{=}Y) = \frac{0.18}{0.285} = \frac{12}{19} \approx 0.632, \qquad \mathbb{P}(L \mid E_1{=}Y) = \frac{0.105}{0.285} = \frac{7}{19} \approx 0.368.$$

(Check: $\frac{12}{19} + \frac{7}{19} = 1$. ✓)

Step 2 — update again on Indicator 2. Now repeat *the exact same move*, but starting from 12/19 and 7/19 (Step 1's posteriors) as the priors, and using Indicator 2's likelihoods ($\mathbb{P}(E_2{=}Y \mid H) = 0.50$, $\mathbb{P}(E_2{=}Y \mid L) = 0.10$) — this is allowed precisely because of the conditional-independence argument in the theory box:

$$\left(\frac{12}{19}\right)(0.50) = \frac{6}{19}, \qquad \left(\frac{7}{19}\right)(0.10) = \frac{0.7}{19}.$$

Both weights share the same denominator 19 (inherited from Step 1), so when we divide one by their sum, that 19 cancels out — we can work directly with the numerators 6 and 0.7:

$$\mathbb{P}(H \mid E_1{=}Y, E_2{=}Y) = \frac{6/19}{6/19 + 0.7/19} = \frac{6}{6 + 0.7} = \frac{6}{6.7} = \frac{60}{67} \approx \boxed{0.896}.$$

Check — one-shot Bayes with both signals at once:

$$\mathbb{P}(H)\mathbb{P}(E_1{=}Y \mid H)\mathbb{P}(E_2{=}Y \mid H) = (0.30)(0.60)(0.50) = 0.09,$$

$$\mathbb{P}(L)\mathbb{P}(E_1{=}Y \mid L)\mathbb{P}(E_2{=}Y \mid L) = (0.70)(0.15)(0.10) = 0.0105.$$

$$\mathbb{P}(H \mid \text{both}) = \frac{0.09}{0.09 + 0.0105} = \frac{0.09}{0.1005} = \frac{60}{67} \approx \boxed{0.896}. \quad \checkmark \text{ matches Step 2.}$$

Two mildly-informative signals compounded: from a 30 % prior to an ≈ 90 % posterior belief the applicant is high-risk.

13. Problem 13 — MGF of the Exponential and a Sum of Loans

Problem 13 – Hard

Classes 4–5 – MGFs & sums of random variables

Setup. $X \sim \text{Exponential}(\lambda)$, $M_X(t) = \lambda/(\lambda - t)$. Derive $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, $\text{Var}(X)$; then for $\lambda = 3$ and 5 independent copies, find the MGF, distribution family, mean and variance of the sum.

Theory used

Definition. The moment generating function of X is $M_X(t) = \mathbb{E}[e^{tX}]$, a function of a dummy variable t (*not* a random variable) — it takes the random quantity e^{tX} and averages it away, so $M_X(t)$ is just an ordinary real-valued function of t, one we can differentiate with respect to t like any other function.

Why does differentiating it produce moments? Two equivalent ways to see it:

(i) Differentiate under the expectation. Under mild regularity conditions, we're allowed to swap the order of "take $\mathbb{E}[\cdot]$" and "take d/dt":

$$M_X'(t) = \frac{d}{dt}\,\mathbb{E}[e^{tX}] = \mathbb{E}\left[\frac{d}{dt}e^{tX}\right] = \mathbb{E}[X\,e^{tX}].$$

The extra factor of X appears simply from the chain rule ($\frac{d}{dt}e^{tX} = Xe^{tX}$, treating X as a constant while differentiating in t). Evaluating at $t = 0$ makes $e^{tX} = e^0 = 1$ vanish, leaving exactly $\mathbb{E}[X]$:

$$M_X'(0) = \mathbb{E}[X\,e^0] = \mathbb{E}[X].$$

Differentiate *again* and the same chain-rule step pulls down *another* factor of X: $M_X''(t) = \mathbb{E}[X^2 e^{tX}]$, so $M_X''(0) = \mathbb{E}[X^2]$. Each extra derivative brings down one more factor of X, so in general $M_X^{(n)}(0) = \mathbb{E}[X^n]$ — the n-th derivative at 0 is the n-th moment. This is exactly why it's called a **moment generating** function.

(ii) Taylor-series intuition (same idea, no calculus needed). Expand e^{tX} in its Maclaurin series in t and take expectations term by term:

$$M_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}\left[\sum_{n=0}^{\infty}\frac{(tX)^n}{n!}\right] = \sum_{n=0}^{\infty}\frac{t^n}{n!}\,\mathbb{E}[X^n] = 1 + t\,\mathbb{E}[X] + \frac{t^2}{2!}\,\mathbb{E}[X^2] + \frac{t^3}{3!}\,\mathbb{E}[X^3] + \cdots.$$

So $M_X(t)$ *is*, term by term, a power series whose t^n coefficient is $\mathbb{E}[X^n]/n!$. Comparing this to the general Taylor expansion of any function, $M_X(t) = \sum_n \frac{M_X^{(n)}(0)}{n!}t^n$, and matching coefficients of t^n on both sides gives $M_X^{(n)}(0)/n! = \mathbb{E}[X^n]/n!$, i.e. $M_X^{(n)}(0) = \mathbb{E}[X^n]$ again — the moments are *literally sitting inside* the Taylor coefficients of M_X, which is why extracting them is just repeated differentiation at $t = 0$.

Sums of independent variables. The MGF of a sum of *independent* variables is the **product** of their MGFs, $M_{X_1+\cdots+X_n}(t) = \prod_i M_{X_i}(t)$ (because $\mathbb{E}[e^{t(X_1+\cdots+X_n)}] = \mathbb{E}[e^{tX_1}\cdots e^{tX_n}] = \prod_i \mathbb{E}[e^{tX_i}]$ by independence). Since an MGF uniquely determines its distribution, recognizing the product's *shape* tells us the sum's distribution family without ever computing a convolution integral.

Thought process

Part (a) is mechanical once you trust the fact above: differentiate twice, evaluate at $t = 0$, done. The trap to avoid is differentiating with respect to the wrong variable — λ is a fixed constant here, and t is the only variable we differentiate against; also remember to evaluate *at* $t = 0$, not leave t floating, since $M'_X(0)$ is a fixed number ($\mathbb{E}[X]$) but $M'_X(t)$ for general t is not.
Part (b) is the more MIT-flavored step: rather than deriving the distribution of a sum of 5 exponentials from scratch (which would require a convolution integral, i.e. repeatedly computing the density of $X_1 + X_2$, then of that sum plus X_3, and so on), we use the MGF *product* shortcut and pattern-match the result to a known family — the sum of n iid Exponential(λ) is exactly the definition of a Gamma(n, λ) (a.k.a. Erlang(n, λ)) random variable.

Step-by-step solution

Where does $M_X(t) = \lambda/(\lambda - t)$ come from? (Given in the problem, but worth seeing once.) Directly from the definition, using the Exponential(λ) density $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$:

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_0^\infty e^{tx}\, \lambda e^{-\lambda x}\, dx = \lambda \int_0^\infty e^{-(\lambda - t)x}\, dx = \lambda \cdot \frac{1}{\lambda - t} = \frac{\lambda}{\lambda - t},$$

valid only for $t < \lambda$ (otherwise the exponent $-(\lambda - t)x$ doesn't decay and the integral diverges to ∞ — this is why every MGF formula comes with a range of valid t).
(a) Moments from the MGF, using $M_X^{(n)}(0) = \mathbb{E}[X^n]$ from the theory box above:

$$M_X(t) = \lambda(\lambda - t)^{-1} \;\Rightarrow\; M'_X(t) = \lambda(\lambda - t)^{-2} \;\Rightarrow\; \mathbb{E}[X] = M'_X(0) = \lambda \cdot \lambda^{-2} = \boxed{\frac{1}{\lambda}}.$$

(Sanity check: this matches the well-known mean of an exponential, $\mathbb{E}[X] = 1/\lambda$, obtained here *without* integrating $x\, f(x)$ directly.)

$$M''_X(t) = 2\lambda(\lambda - t)^{-3} \;\Rightarrow\; \mathbb{E}[X^2] = M''_X(0) = 2\lambda \cdot \lambda^{-3} = \boxed{\frac{2}{\lambda^2}}.$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \boxed{\frac{1}{\lambda^2}}.$$

(b) Sum of 5 independent Exponential(3)'s. With $\lambda = 3$: $\mathbb{E}[X] = 1/3$, $\mathrm{Var}(X) = 1/9$. The MGF of the sum $T = X_1 + \cdots + X_5$ is the product of the five (identical) MGFs:

$$M_T(t) = \left(\frac{3}{3 - t}\right)^5.$$

This is exactly the MGF of a **Gamma**(k=5, λ=3) (**Erlang**) random variable — the "waiting time until the 5th event" in a Poisson process of rate 3. By linearity of expectation and additivity of variance across independent variables:

$$\mathbb{E}[T] = 5 \cdot \frac{1}{3} = \boxed{\frac{5}{3} \approx 1.67 \text{ months}}, \qquad \mathrm{Var}(T) = 5 \cdot \frac{1}{9} = \boxed{\frac{5}{9} \approx 0.556}.$$

14. Problem 14 — Estimating a Certification Rate

Problem 14 – Hard

Class 5 – CLT & confidence intervals

Setup. $n = 225$, 153 successes. Find $\hat{p}$, a 95 % CI for p, and the sample size needed to halve the margin of error.

Theory used

For a sample proportion, $\hat{p} \overset{\text{approx}}{\sim} \mathcal{N}\left(p,\ \dfrac{p(1-p)}{n}\right)$ by the CLT (treating each respondent as a Bernoulli trial and $\hat{p}$ as their sample mean). A $(1-\alpha)$ confidence interval is $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$; for 95 %, $z_{0.025} = 1.96$. The **margin of error** is $\text{ME} = z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} \propto 1/\sqrt{n}$.

Thought process

A proportion is just a sample mean of 0/1 outcomes, so this is the CLT applied to Bernoulli data — the standard error formula $\sqrt{p(1-p)/n}$ replaces $\sigma/\sqrt{n}$ because $\text{Var}(\text{Bernoulli}) = p(1-p)$. For part (c), since $\text{ME} \propto 1/\sqrt{n}$, halving the ME requires **quadrupling** n — a good number to have memorized, since it recurs constantly in sample-size planning.

Step-by-step solution

(a)

$$\hat{p} = \frac{153}{225} = \boxed{0.68}.$$

(b) Standard error:

$$\text{SE} = \sqrt{\frac{(0.68)(0.32)}{225}} = \sqrt{\frac{0.2176}{225}} \approx 0.0311.$$

$$\text{ME} = 1.96 \times 0.0311 \approx 0.0610.$$

$$95\,\%\ \text{CI} = 0.68 \pm 0.0610 = \boxed{(0.619,\ 0.741)}.$$

We are approximately 95 % confident the true certification rate lies between 61.9 % and 74.1 %.
(c) $\text{ME} = z_{0.025}\sqrt{\hat{p}(1-\hat{p})/n}$, so $\text{ME} \propto n^{-1/2}$. To cut ME in half, we need $\sqrt{n}$ to double, i.e. n to **quadruple**:

$$n' = 4 \times 225 = \boxed{900}.$$

15. Problem 15 — Capstone: Which Program Version, and What Next?

Problem 15 – Hard (Capstone)

Classes 2 & 5 – Bayes + Poisson

Setup. Version A (65 % of sites, Poisson(7)) vs. Version B (35 %, Poisson(4)) placements per cohort of 10. A random site's cohort reports 6 placements. Find $\mathbb{P}(\text{A} \mid X{=}6)$ and the posterior-predictive expected placements for a new cohort at the same site.

Theory used

This combines two tools from the course: **Bayes' theorem** with a *discrete* hypothesis (which version?) and a **Poisson likelihood** as the evidence (instead of a simple "yes/no" indicator). Once we have posterior probabilities $\mathbb{P}(A \mid X{=}6)$ and $\mathbb{P}(B \mid X{=}6)$, the **posterior-predictive mean** for a new, independent cohort at the *same* site is a probability-weighted mixture of the two conditional means:

$$\mathbb{E}[X_{\text{new}} \mid X{=}6] = \mathbb{P}(A \mid X{=}6) \cdot \mathbb{E}[X \mid A] + \mathbb{P}(B \mid X{=}6) \cdot \mathbb{E}[X \mid B].$$

Thought process

The structure is identical to Problem 4 (Bayes with a partition), except the "evidence" is no longer a simple event like "referred" — it is a specific Poisson *count*, so the likelihoods $\mathbb{P}(X{=}6 \mid A)$ and $\mathbb{P}(X{=}6 \mid B)$ come from the Poisson pmf rather than being given directly. For part (b), the key realization is that *if* we knew the version for certain, the best prediction for a new cohort would just be that version's Poisson mean (7 or 4); since we're *not* certain, we blend the two means using our updated (posterior) beliefs about which version this site uses — not the original 65 %/35 % prior.

Step-by-step solution

(a) Likelihoods (Poisson pmf at $k = 6$):

$$\mathbb{P}(X{=}6 \mid A) = \frac{e^{-7}7^6}{6!} \approx 0.1490, \qquad \mathbb{P}(X{=}6 \mid B) = \frac{e^{-4}4^6}{6!} \approx 0.1042.$$

Bayes' theorem:

$$\mathbb{P}(A \mid X{=}6) = \frac{(0.65)(0.1490)}{(0.65)(0.1490) + (0.35)(0.1042)} = \frac{0.0969}{0.0969 + 0.0365} = \frac{0.0969}{0.1333} \approx \boxed{0.726}.$$

$$\mathbb{P}(B \mid X{=}6) = 1 - 0.726 = \boxed{0.274}.$$

Six placements is *more consistent* with the higher-rate Version A (mean 7) than the lower-rate Version B (mean 4), so the posterior shifts *up* from the 65 % prior to about 72.6 %.

(b) Posterior-predictive mean. Using $\mathbb{E}[X \mid A] = 7$ and $\mathbb{E}[X \mid B] = 4$ (Poisson mean equals its rate parameter):

$$\mathbb{E}[X_{\text{new}} \mid X{=}6] = (0.726)(7) + (0.274)(4) = 5.082 + 1.096 \approx \boxed{6.18}.$$

For the next cohort at this same site, we would predict about **6.18** successful placements — noticeably higher than the population-average prediction of $(0.65)(7) + (0.35)(4) = 6.05$ we would have made *before* observing this site's data, because the observation of 6 placements pulled our belief toward Version A.

16. Problem 16 — Appointment Codes

Problem 16 – Easy (Bonus)

Class 1 – Counting (permutations)

Setup. 3 distinct letters from $\{A, B, C, D, E\}$, arranged in order, no repeats. How many codes?

Theory used

When order **matters** and repetition is **not** allowed, the count is the permutation $P(n,k) = \dfrac{n!}{(n-k)!} = n(n-1)\cdots(n-k+1)$.

Thought process

Unlike Problem 1 (committee, order irrelevant), here `ABC` and `BAC` are explicitly called out as *different* codes — that is the signal to use a permutation, not a combination.

Step-by-step solution

$$P(5,3) = 5 \times 4 \times 3 = \boxed{60}.$$

There are 60 different appointment codes.

17. Problem 17 — Transit and Car Ownership

Problem 17 – Intermediate (Bonus)

Class 2 – Independent events

Setup. $\mathbb{P}(\text{transit}) = 0.4$, $\mathbb{P}(\text{car}) = 0.5$, independent. Find $\mathbb{P}(\text{transit and car})$, $\mathbb{P}(\text{transit or car})$, $\mathbb{P}(\text{transit} \mid \text{car})$.

Theory used

Independence: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Inclusion–exclusion for "or": $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$. And by definition of independence, $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

Thought process

The word "independent" is given directly, so there's no need to check it — we just plug straight into the three formulas. Part (c) is really a conceptual check: conditioning on an independent event should not move the probability at all.

Step-by-step solution

(a) $\mathbb{P}(\text{transit and car}) = (0.4)(0.5) = \boxed{0.20}$.
(b) $\mathbb{P}(\text{transit or car}) = 0.4 + 0.5 - 0.20 = \boxed{0.70}$.

(c) $\mathbb{P}(\text{transit} \mid \text{car}) = \mathbb{P}(\text{transit}) = \boxed{0.40}$ — exactly because the two events are independent, knowing someone owns a car tells us nothing new about whether they use transit.

18. Problem 18 — Uniform on a Rectangle

Problem 18 – Intermediate (Bonus)

Class 3 – Joint uniform density

Setup. (X, Y) uniform on $[0, 2] \times [0, 3]$. Find $f(x, y)$, $\mathbb{P}(X < 1, Y < 2)$, and whether X, Y are independent.

Theory used

A uniform density on a region of area A is the constant $f(x, y) = 1/A$ on that region. When the support is a **rectangle** (a product of intervals) *and* the density is constant, the joint density **factors** as $f(x, y) = f_X(x) f_Y(y)$, which is exactly the definition of independence.

Thought process

This is a deliberate contrast with Problem 10: there, the triangular (non-rectangular) support tangled X and Y together even though the density itself was simple. Here, the support *is* a rectangle, so X and Y are independent **without any calculation** — a rectangular support with constant density always factors into a uniform-X times a uniform-Y.

Step-by-step solution

(a) Area $= 2 \times 3 = 6$, so

$$f(x, y) = \boxed{\tfrac{1}{6}}, \qquad 0 \le x \le 2,\ 0 \le y \le 3.$$

(b) The event is itself a rectangle $[0, 1) \times [0, 2)$ of area $1 \times 2 = 2$:

$$\mathbb{P}(X < 1, Y < 2) = \frac{2}{6} = \boxed{\tfrac{1}{3} \approx 0.333}.$$

(c) Yes, independent. $X \sim \text{Unif}[0, 2]$ (density $\frac{1}{3}$ per unit $\times$ length-3 range for Y integrates out) and $Y \sim \text{Unif}[0, 3]$, and $f(x, y) = \frac{1}{6} = (\frac{1}{2})(\frac{1}{3}) = f_X(x) f_Y(y)$ factors cleanly. *Check via part (b):* $\mathbb{P}(X{<}1)\mathbb{P}(Y{<}2) = (\frac{1}{2})(\frac{2}{3}) = \frac{1}{3}$, matching (b) exactly. ✓

19. Problem 19 — Best Linear Predictor

Problem 19 – Hard (Bonus)

Class 4 – Best linear predictor

Setup. $\mathrm{Var}(X) = 5$, $\mathrm{Var}(Y) = 12$, $\mathrm{Cov}(X, Y) = 6$. Find the slope b and R^2 of the best linear predictor of Y from X.

Theory used

The best linear predictor $\hat{Y} = a + bX$ (minimizing mean squared error) has slope $b = \mathrm{Cov}(X, Y)/\mathrm{Var}(X)$. Its quality of fit is measured by $R^2 = \mathrm{Corr}(X, Y)^2 = \dfrac{\mathrm{Cov}(X, Y)^2}{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$ — the share of Y's variance the linear predictor explains.

Thought process

This is exactly the "intro to regression" idea previewed in Class 4: the best linear predictor's slope is a ratio of covariance to variance (the same formula that becomes the OLS slope in Class 9), and R^2 is that same covariance, but standardized by *both* variances so it's unit-free and bounded in $[0, 1]$.

Step-by-step solution

(a)

$$b = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} = \frac{6}{5} = \boxed{1.2}.$$

(b)

$$R^2 = \frac{\mathrm{Cov}(X, Y)^2}{\mathrm{Var}(X)\,\mathrm{Var}(Y)} = \frac{6^2}{(5)(12)} = \frac{36}{60} = \boxed{0.6}.$$

The linear predictor explains 60 % of Y's variance; the slope tells us each 1-unit increase in X predicts a 1.2-unit increase in Y, on average.

20. Problem 20 — Sum of Independent Poissons

Problem 20 – Hard (Bonus)

Class 5 – Sums of Poisson variables

Setup. $X \sim \mathrm{Poisson}(3)$, $Y \sim \mathrm{Poisson}(5)$, independent. Find the distribution of $X + Y$ and $\mathbb{P}(X + Y = 4)$.

Theory used

Reproductive property of the Poisson: if $X \sim \mathrm{Pois}(\lambda_1)$ and $Y \sim \mathrm{Pois}(\lambda_2)$ are independent, then $X + Y \sim \mathrm{Pois}(\lambda_1 + \lambda_2)$. (This also falls out of the MGF technique from Problem 13: the MGF of a sum of independents is the product of MGFs, and the product of two Poisson MGFs is again a Poisson MGF, with rates adding.)

Thought process

Rather than computing the pmf of $X+Y$ from a convolution sum (summing $\mathbb{P}(X{=}j)\mathbb{P}(Y{=}4{-}j)$ over $j = 0, \ldots, 4$), recognizing the reproductive property lets us skip straight to a single named

distribution — the same "recognize the family" shortcut used with the MGF in Problem 13.

Step-by-step solution

(a) Since X and Y are independent Poissons,

$$X + Y \sim \text{Poisson}(3 + 5) = \boxed{\text{Poisson}(8)}.$$

(b)

$$\mathbb{P}(X + Y = 4) = \frac{e^{-8}8^4}{4!} \approx \boxed{0.0573}.$$

Sanity check: computing this instead as a convolution, $\sum_{j=0}^{4} \mathbb{P}(X{=}j)\mathbb{P}(Y{=}4{-}j)$, gives the same 0.0573 — just with far more arithmetic than using the reproductive property directly.

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. This review session was designed around the probability core of Classes 1–5 of this course (counting, conditional probability & Bayes, joint distributions, expectation & variance, special distributions, sample mean & the CLT). All 15 problems were authored fresh for this session; none are copied from the course's per-class problem banks or Kahoot question sets.

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`; `hanmochen/Probability-Theory-2020`; `mynameisjanus/14310xDataAnalysis`.