Recitation Final Practice

Math-UA 185: Probability and Statistics / Spring 2026

Ed Chen

I've put together a few practice problems in light of the final exam coming up. I'll go over these during the final recitation on May 1st. Problems 1 through 4 are decently comprehensive and touch on most of the more challenging concepts in this course, whilst the latter 10 questions are more straightforward standard issue questions.

A fully explained solution set is also on Brightspace. Portions of the last 10 questions are pulled from the Dekking textbook, which is also a good resource for more practice. Partial inspiration for Problems 1 and 2 are taken from EECS126, at Berkeley.

Problem 1. Office Hours

A student arrives to office hours at 1 pm. The TA is particularly lazy today, and instead of being in his office continuously, checks his office according to a Poisson process with rate $\lambda = 5$ times per hour independently of the student's arrival time.

When the TA finally arrives, he gives the student bonus points linearly proportional to the time since his previous visit to the office, as he does not know how long the student has been waiting. Assume this rate is 1 point per hour.

a) Let W be the amount of time, in hours, that the student waits until the TA next appears. What is the distribution of W? What is $\mathbb{E}[W]$?

Ans. This follows from the homogeneity of Poisson process, note that the distribution of W is equivalent to the probability that $N_{1\text{pm}+t} - N_{1\text{pm}} = 0$, which follows $\text{Poi}(5t)$ at $k = 0$.

$$\text{P}(W > t) = \text{P}(N_{1\text{pm}+t} - N_{1\text{pm}} = 0) = \frac{(5t)^0}{0!}e^{-5t} = e^{-5t}$$

What we have found here, is the converse of the CDF of W.

$$P(W > t) = 1 - P(W \le t) = 1 - \int_0^t p_W(t)dt$$

Indeed, if we solve for p_W, we have $p_W(t) = d_t(1 - e^{-5t}) = 5e^{-5t}$, which is precisely the pdf for Exponential(5).

If you are not asked to *show* this, you could of course, directly claim that by memorylessness of Poisson processes, the waiting time until the next TA visit is exponential.

$$W \sim \text{Exponential}(5) \qquad \mathbb{E}[W] = 1/5$$

Converting, 12 minutes.

b) Let A be the amount of time, in hours, since the TA's previous office visit when the student arrives. What is the distribution of A? Are A and W independent?

Ans. Again, this follows from homogeneity of Poisson processes. Namely, the distribution of A is equivalent to that of $N_{1\text{pm}} - N_{1\text{pm}-t} = 0$. By the same logic, $A \sim \text{Exponential}(5)$.

Moreover, A and W are independent. We have argued that the distribution of W depends on the probability of arrivals within $[1\text{pm}, 1\text{pm} + t]$ and the distribution of A depends on the probability of arrivals within $[1\text{pm} - t, 1\text{pm}]$, which are disjoint intervals, and hence independent axiomatically.

c) Let B be the number of bonus points. Compute $\mathbb{E}[B]$, and claim the distribution of B.

Ans. The time since the TA's previous visit, measured at the moment he next arrives, is $B = A + W$. Since $A, W \sim \text{Exponential}(5)$ independently,

$$\mathbb{E}[B] = \mathbb{E}[A] + \mathbb{E}[W] = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$$

So the expected bonus is 0.4 points. Also note since B is the sum of two independent exponential RVs with the same rate 5,

$$B \sim \text{Gamma}(2, 5) \qquad f_B(b) = 25be^{-5b},\ b \ge 0.$$

d) What is the probability that the student receives more than 1 bonus point?

Ans. We want $\mathrm{P}(B > 1)$, which is the complement of the CDF at 1.

$$\mathrm{P}(B > 1) = 1 - \mathrm{P}(B \leq 1) = 1 - \int_0^1 25be^{-5b}\, db = 1 - (1 - 6e^{-5}) = 6e^{-5}.$$

e) While waiting, the student sends emails according to a separate Poisson process at rate 3 emails per hour, independently of the TA's office-checking process. Let N be the number of emails sent before the TA appears. Find $\mathrm{P}(N = 0)$.

Ans. The tricky part here is that the time the student waits for the TA is itself a RV, which we have identified previously as W. If we condition that W takes some determined value w, we can claim the distribution of N under a fixed time interval,

$$\mathrm{P}(N = 0 \mid W = w) = e^{-3w}$$

To recover the unconditioned probability, we integrate out the "randomness" of W

$$\mathrm{P}(N = 0) = \mathbb{E}[e^{-3W}] = \int_0^\infty \mathrm{P}(N = 0 \mid W = w) f_W(w)\, dw = \int_0^\infty e^{-3w} 5e^{-5w}\, dw = \frac{5}{8}$$

So with probability 5/8, the TA arrives before the student sends an email.

f) Now find the distribution of N, the number of emails sent before the TA appears.

Ans. Again conditioning on W, but now generalized to k emails,

$$\mathrm{P}(N = k \mid W = w) = e^{-3w} \frac{(3w)^k}{k!}$$

We again integrate out W, using the Gamma function integral

$$\mathrm{P}(N = k) = \int_0^\infty e^{-3w} \frac{(3w)^k}{k!} 5e^{-5w}\, dw = \frac{5(3^k)}{k!} \int_0^\infty w^k e^{-8w} dw = \frac{5(3^k)}{\cancel{k!}} \frac{\cancel{\Gamma(k+1)}}{8^{k+1}} = \frac{5}{8} \left(\frac{3}{8}\right)^k$$

So $N \sim$ Geometric(5/8) with success probability 5/8, where N counts the number of failures before the first success.

Problem 2. Concert Time

After taking the final, you decide to go to a concert. There are only two staff members letting fans into the venue. Staff 1 has service times distributed as Exponential(2), and staff 2, who is more experienced at their job, has service times distributed as Exponential(3). All service times are independent, and measured in minutes.

When you arrive, both staff are busy with one fan each, and there is a single line of $n-1$ fans in front of you waiting for the first available staff member. Let T be the total time, in minutes, from when you enter the line until you are let into the venue.

a) Let $X_1 \sim$ Exponential(2) be the remaining time until staff 1 finishes with their current fan, and let $X_2 \sim$ Exponential(3) be the remaining time until staff 2 finishes with their current fan. Find the distribution and expectation of

$$M = \min\{X_1, X_2\}$$

Ans. For $t \geq 0$ the events $\{M > t\}$ and $\{X_1 > t \cap X_2 > t\}$ are equivalent. Since $X_1 \perp X_2$,

$$\mathrm{P}(M > t) = \mathrm{P}(X_1 > t, X_2 > t) = (1 - F_{X_1}(t))(1 - F_{X_2}(t)) = e^{-2t}e^{-3t} = e^{-5t}$$

This is the complement of the CDF of M. As with Q1a, we can identify that

$$M \sim \text{Exponential}(5) \qquad \mathbb{E}[M] = \frac{1}{5}$$

In words, a staff member becomes available at rate $2 + 3 = 5$ per minute.

b) Find the CDF of

$$Z = \max\{X_1, X_2\}$$

Interpret what Z represents in this scenario.

Ans. For $t \geq 0$ the events $\{Z \leq t\}$ and $\{X_1 \leq t \cap X_2 \leq t\}$ are equivalent.

$$F_Z(t) = \mathrm{P}(Z \leq t) = \mathrm{P}(X_1 \leq t, X_2 \leq t) = (1 - e^{-2t})(1 - e^{-3t}) = 1 - e^{-2t} - e^{-3t} + e^{-5t}$$

Here Z is the amount of time until both of the fans who were already being served when you arrived have been let into the venue.

c) After waiting for a while, you are now first in line. Yay! You notice staff 1 is giving out free merch. What is the probability that staff 1 is the first worker to become available? Which staff is more likely to be available first?

Ans. This is the event that $\{X_1 < X_2\}$. Note the distinction between $<$ and $\leq$ does not matter in continuous RVs as the probability of any exact value is 0. This is similar to the situation in Q1e, the tricky part being that X_2 is a RV, so we condition on it having a determined value, $\mathrm{P}(X_1 < x \mid X_2 = x)$, and compute the total probability by integrating out the "randomness" of X_2, ergo over all possible values of x.

$$\mathrm{P}(X_1 < X_2) = \int_0^\infty \mathrm{P}(X_1 < x) f_{X_2}(x)\, dx$$

We use the CDF for X_1 and the PDF for X_2

$$\mathrm{P}(X_1 < x) = 1 - e^{-2x} \qquad \text{and} \qquad f_{X_2}(x) = 3e^{-3x}$$

to get

$$\mathrm{P}(X_1 < X_2) = \int_0^\infty (1 - e^{-2x})3e^{-3x}\, dx = 3\int_0^\infty e^{-3x}\, dx - 3\int_0^\infty e^{-5x}\, dx = 1 - \frac{3}{5} = \frac{2}{5}$$

As $\{X_1 > X_2\}$ and $\{X_2 > X_1\}$ completely span the event space, we can also state

$$P(X_2 < X_1) = \frac{3}{5}$$

Staff 2 is more likely to open first because staff 2 has the faster service rate. A tricky misconception may be that since the problem statement does not state *when* the current fan has reached a given staff member, that you cannot directly appeal to the exponential distribution. However, note that due to memorylessness, it does not actually matter when the current fan has reached a staff member.

d) Find the expected amount of time from arrival until you first reach a staff member.

Ans. Each time a staff member finishes with a fan, the line moves forward by one person. While both staff are busy, the time until the next completion follows $M \sim \text{Exponential}(5)$ as we identified in a).

There are $n-1$ fans in front of you, so after $n-1$ completions those fans have reached staff members. You then need one more completion before you reach a staff member. Therefore you wait for n service completions. Let W describe the time from arrival until you reach a staff member. Then,

$$W = \sum_{i=1}^{n} E_i, \qquad E_i \perp E_j \;\forall\, i \neq j$$

where

$$E_i \sim \text{Exponential}(5)$$

independently. Therefore

$$\mathbb{E}[W] = n \cdot \frac{1}{5} = \frac{n}{5}.$$

e) Once you reach a staff member, what is your expected service time?

Ans. The next available staff member is staff 1 with probability 2/5 and staff 2 with probability 3/5 as we have determined in c). Take S to be the RV describing your expected service time, then we can use total probability to state

$$\mathbb{E}[S] = \sum_{i=1}^{2} \mathbb{E}[S \mid \text{you get staff } i] P(\text{you get staff } i)$$

$$\mathbb{E}[S] = \frac{2}{5} \cdot \frac{1}{2} + \frac{3}{5} \cdot \frac{1}{3} = \frac{2}{5}.$$

f) Find $\mathbb{E}[T]$, the expected total time from entering the line until being let into the concert.

Ans. The total time is

$$T = W + S$$

where W is the time until you reach a staff member and S is your own service time. Therefore

$$\mathbb{E}[T] = \mathbb{E}[W] + \mathbb{E}[S] = \frac{n}{5} + \frac{2}{5} = \frac{n+2}{5}$$

g) Imagine now that the line is quite busy. The concert organizers, who are getting flak for only hiring two staff members, are promising a free drink ticket if your wait time is over 60 minutes. Use Markov's inequality to give an upper bound on the probability that your total time is over 60 minutes in terms of the number of people in line, including yourself n.

Ans. Since $T \geq 0$, a direct application of Markov's inequality gives

$$P(T \geq 60) \leq \frac{\mathbb{E}[T]}{60} = \frac{(n+2)/5}{60} = \frac{n+2}{300}.$$

h) Your friend is quite excited at the prospect of a free drink and demands to know a tighter probability bound after they count a total of $n = 148$ fans in line (they haven't taken this course). Compute $\text{Var}(T)$ and use Chebyshev's inequality to bound

$$P\left(|T - \mathbb{E}[T]| \geq r\right)$$

What can you say about $P(T \geq 60)$ using the bound you have found?

Ans. We seek the variance of T. Note, that since W and S are independent, $\text{Cov}(W, S) = 0$, so

$$\text{Var}(T) = \text{Var}(W + S) = \text{Var}(W) + \text{Var}(S)$$

Recall we have

$$W = E_1 + \cdots + E_n, \qquad E_i \sim \text{Exponential}(5)$$

so, using the variance computation rule

$$\text{Var}(W) = n \cdot \frac{1}{5^2} = \frac{n}{25}$$

Now compute the variance of your own service time S. We already found $\mathbb{E}[S] = 2/5$. Also,

$$\mathbb{E}[S^2] = \frac{2}{5} \cdot \frac{2}{2^2} + \frac{3}{5} \cdot \frac{2}{3^2}$$

Hence

$$\text{Var}(S) = \mathbb{E}[S^2] - \mathbb{E}[S]^2 = \frac{1}{3} - \left(\frac{2}{5}\right)^2 = \frac{1}{3} - \frac{4}{25} = \frac{13}{75}$$

Now combining,

$$\text{Var}(T) = \text{Var}(W) + \text{Var}(S) = \frac{n}{25} + \frac{13}{75} = \frac{3n+13}{75}$$

Applying Chebyshev's inequality,

$$P\left(|T - \mathbb{E}[T]| \geq r\right) = P\left(\left|T - \frac{n+2}{5}\right| \geq r\right) \leq \frac{\text{Var}(T)}{r^2} = \frac{3n+13}{75r^2}$$

we get, using the prior part $\mathbb{E}[T] = (148 + 2)/5 = 30$

$$P\left(|T - 30| \geq r\right) \leq \frac{457}{75r^2}$$

To get a bound on $P(T \geq 60)$, we set $r = 30$ to get

$$P(|T - 30| \geq 30) \leq \frac{457}{75(30)^2} \approx 0.00677$$

Since the event $\{T \geq 60\}$ is contained in the event $\{|T - 30| \geq 30\}$, this also gives

$$P(T \geq 60) \leq 0.00677$$

Alas, the chances are slim. Indeed, this 0.67% bound is much tighter than the 50% bound from Markov!

Problem 3. Boba Price Ceiling

An eager NYU Data Science graduate has decided to capitalize on a lucrative business opportunity to open a new boba shop near NYU. She is trying to decide how aggressively she can price drinks before students remember they have rent to pay.

She wants to better understand the distribution that describes the willingness-to-pay of students, X. As an avid data scientist, she assumes it follows a bounded rescaled Beta distribution $X \sim \text{Boba}(\alpha, \theta)$, where θ describes the maximum possible price the population would be willing to pay for a drink, and where α describes how clustered others' willingness-to-pay is around that maximum. The probability density is

$$f_{\alpha,\theta}(x) = \begin{cases} \dfrac{\alpha x^{\alpha-1}}{\theta^\alpha}, & 0 < x < \theta, \\ 0, & \text{otherwise} \end{cases}$$

For student i, let X_i be the maximum price, in dollars, that student is willing to pay for one drink. Suppose a poll is taken of n students, collecting samples

$$X_1, \ldots, X_n$$

which we assume are i.i.d. with density $X \sim \text{Boba}(\alpha, \theta)$, with unknown $\alpha > 0$ and $\theta > 0$.

a) Confirm that $f_{\alpha,\theta}$ is a valid density for arbitrary α, θ, and find the CDF of X_i.

Ans. We check that the PDF normalizes to 1:

$$\int_0^\theta \frac{\alpha x^{\alpha-1}}{\theta^\alpha}\, dx = \frac{\alpha}{\theta^\alpha}\left[\frac{x^\alpha}{\alpha}\right]_0^\theta = \frac{\theta^\alpha}{\theta^\alpha} = 1$$

So $f_{\alpha,\theta}$ is a valid density. For the CDF, over $0 \le x \le \theta$,

$$F_X(x) = \mathrm{P}(X_i \le x) = \int_0^x \frac{\alpha t^{\alpha-1}}{\theta^\alpha}\, dt = \frac{x^\alpha}{\theta^\alpha}$$

Thus, altogether

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \dfrac{x^\alpha}{\theta^\alpha}, & 0 \le x \le \theta, \\ 1, & x > \theta. \end{cases}$$

b) Write the likelihood and log-likelihood functions for α, θ. Recall since θ dictates the support of the distribution, you must use an indicator function in accurately describing the PDF.

Ans. The likelihood is the product of the PDFs for each polled student,

$$L(\alpha, \theta) = \prod_{i=1}^n \frac{\alpha X_i^{\alpha-1}}{\theta^\alpha} \mathbf{1}_{\{0 < X_i < \theta\}}$$

Note now, that if $\theta < \max_i X_i$ the likelihood is 0, intuitively, the chances that a sample $X_i > \theta$ is drawn from a distribution with support $[0, \theta]$ is 0. For $\theta \ge \max_i X_i$, the log-likelihood is

$$\begin{aligned} \ell(\alpha, \theta) &= \log L(\alpha, \theta) \\ &= \log\left(\left(\frac{\alpha}{\theta^\alpha}\right)^n \prod_{i=1}^n X_i^{\alpha-1}\right) \\ &= n\log\alpha - n\alpha\log\theta + (\alpha - 1)\sum_{i=1}^n \log X_i \qquad (\theta \ge \max_i X_i) \end{aligned}$$

c) Assuming a fixed α, find the Maximum Likelihood Estimator of θ, denoted $\widehat{\theta}$.

Ans. We seek the argmax θ value which maximizes $\ell(\alpha, \theta)$. Note this is a bounded optimization problem, exactly as you have seen in Calculus. The partial derivative of the log-likelihood function, keeping in mind the constraint is

$$\partial_\theta \ell(\alpha, \theta) = -\frac{n\alpha}{\theta} \qquad (\theta \geq \max_i X_i)$$

Note now that given $\alpha, \theta > 0$ this is strictly negative. So, the maximal choice of θ must occur at the left endpoint as our function is strictly decreasing in the interval. Our smallest choice of θ is constrained by $\theta \geq \max_i X_i$, and indeed becomes

$$\widehat{\theta} = \max_i X_i$$

d) Now using $\widehat{\theta}$, find the Maximum Likelihood Estimator of α, denoted $\widehat{\alpha}$.

Ans. Plugging $\theta = \widehat{\theta} = \max_i X_i$ into the log-likelihood gives

$$\ell(\alpha, \widehat{\theta}) = n \log \alpha - n\alpha \log \widehat{\theta} + (\alpha - 1) \sum_{i=1}^{n} \log X_i$$

$$\partial_\alpha \ell(\alpha, M) = \frac{n}{\alpha} - n \log \widehat{\theta} + \sum_{i=1}^{n} \log X_i$$

We solve for $\partial_\alpha \ell(\alpha, \widehat{\theta}) = 0$ to get

$$\widehat{\alpha} = \frac{n}{n \log \widehat{\theta} - \sum_{i=1}^{n} \log X_i} = \frac{n}{\sum_{i=1}^{n} \log(\widehat{\theta}/X_i)}$$

Intuitively, you can interpret this as the estimator $\widehat{\theta}$ using the largest observed price willingness, while $\widehat{\alpha}$ estimates how close the rest of the population is to that bound.

e) Compute the bias of the Maximum Likelihood Estimator $\widehat{\theta}$ of θ. You may find it helpful to first attain the PDF/CDF of $\widehat{\theta}$.

Ans. Again, we have $\widehat{\theta} = \max_i X_i$, and we seek to find $\text{Bias}(\widehat{\theta}) = \mathbb{E}[\widehat{\theta}] - \theta$. As we are dealing with a maximum, finding $\mathbb{E}[\widehat{\theta}]$ does not admit an immediate linearization. Instead, consider that

$$\text{P}(\widehat{\theta} \leq m) = \text{P}(X_1 \leq m, \ldots, X_n \leq m)$$

and as X_i are iid variables, the CDF within $0 \leq m \leq \theta$

$$F_{\widehat{\theta}}(m) = \prod_{i=1}^{n} F_X(m) = \left[\left(\frac{m}{\theta}\right)^{\alpha}\right]^{n} = \left(\frac{m}{\theta}\right)^{\alpha n}$$

Differentiating, we have the PDF

$$f_{\widehat{\theta}}(m) = \frac{\alpha n \, m^{\alpha n - 1}}{\theta^{\alpha n}}$$

Then, we can directly calculate the expectation

$$\mathbb{E}[\widehat{\theta}] = \int_0^\theta m \frac{\alpha n \, m^{\alpha n - 1}}{\theta^{\alpha n}} \, dm = \frac{\alpha n}{\theta^{\alpha n}} \int_0^\theta m^{\alpha n} \, dm = \frac{\alpha n}{\theta^{\alpha n}} \frac{\theta^{\alpha n + 1}}{\alpha n + 1} = \frac{\alpha n}{\alpha n + 1} \theta$$

Altogether, we have

$$\text{Bias}(\widehat{\theta}) = \mathbb{E}[\widehat{\theta}] - \theta = -\frac{\theta}{\alpha n + 1}$$

So M is negatively biased as an estimator of the true price ceiling. Intuitively, the highest price willingness observed is probably still not the theoretical limit of price willingness among all students.

f) Assuming a known α, construct an unbiased estimator, denoted $\widehat{\theta}_0$ of θ using $\widehat{\theta}$.

Ans. We seek to find an estimator $\widehat{\theta}_0$ such that $\mathbb{E}[\widehat{\theta}_0] - \theta = 0$. Using our calculation from the previous part of $\mathbb{E}[\widehat{\theta}]$, we solve for a constant c such that $c\mathbb{E}[\widehat{\theta}] = \mathbb{E}[\widehat{\theta}_0]$

$$\theta = \mathbb{E}[\widehat{\theta}_0] = c\mathbb{E}[\widehat{\theta}] = c\left(\frac{\alpha n}{\alpha n + 1}\theta\right)$$

$$c = \frac{\alpha n + 1}{\alpha n}$$

As such, our estimator is

$$\widehat{\theta}_0 = \frac{\alpha n + 1}{\alpha n} \max_i X_i$$

g) Suppose now a helpful friend who also runs a boba shop lets our business owner know that their prior market research has established that the maximum possible willingness-to-pay is

$$\theta = 12$$

That is, we assume no student is willing to pay more than \$12 for one boba. Restate the MLE $\widehat{\alpha}$, compute the Fisher Information $I_n(\alpha)$, and state the Cramer–Rao lower bound for unbiased estimators of α.

Ans. We use the formula derived in part d) to get

$$\widehat{\alpha} = \frac{n}{n \log 12 - \sum_{i=1}^{n} \log X_i} = \frac{n}{\sum_{i=1}^{n} \log(12/X_i)}$$

To compute Fisher information, differentiate $\ell(\alpha, 12)$ from part d) again:

$$\ell''(\alpha, 12) = -\frac{n}{\alpha^2}$$

Therefore

$$I_n(\alpha, 12) = -\mathbb{E}[\ell''(\alpha, 12)] = \frac{n}{\alpha^2}$$

So if T is an unbiased estimator of α, the Cramér–Rao lower bound gives

$$\text{Var}(T) \geq \frac{1}{I_n(\alpha)} = \frac{\alpha^2}{n}$$

h) A related large-sample principle to Cramer–Rao is that under regularity conditions the MLE is approximately normal with variance that meets the lower bound given by Cramer–Rao, ergo

$$\widehat{\alpha} \approx \mathcal{N}\left(\alpha, I_n^{-1}(\alpha)\right)$$

Use this approximation to give an approximate 95% confidence interval for α.

Ans. Using the large-sample MLE approximation,

$$\widehat{\alpha} \approx \mathcal{N}\left(\alpha, I_n^{-1}(\alpha)\right) = \mathcal{N}\left(\alpha, \frac{\alpha^2}{n}\right)$$

We apply the transformation

$$Z = \frac{\widehat{\alpha} - \mathbb{E}[\widehat{\alpha}]}{\sqrt{\text{Var}(\widehat{\alpha})}} = \frac{\widehat{\alpha} - \alpha}{\alpha/\sqrt{n}} \approx \mathcal{N}(0, 1)$$

Then, with the notation $z_a = \Phi^{-1}(1-a)$ we have

$$\mathrm{P}\left(-z_{0.025} \le \frac{\widehat{\alpha} - \alpha}{\alpha/\sqrt{n}} \le z_{0.025}\right) \approx 0.95$$

Since α is unknown, we plug in $\widehat{\alpha}$ in the standard transformation. Thus an approximate 95% confidence interval for α is

$$\widehat{\alpha} \pm z_{0.025}\frac{\widehat{\alpha}}{\sqrt{n}} \qquad \widehat{\alpha} = \frac{n}{\sum_{i=1}^{n}\log(12/X_i)}$$

Note this is not an exact finite-sample confidence interval.

i) With this new information, suppose she sets the price of one drink to be p, where $0 \le p \le 12$. A randomly chosen student buys the drink if and only if it is below their willingness-to-pay, that is

$$X_i \ge p$$

Find the expected revenue from one student as a function of p.

Ans. The probability of a sale is

$$\mathrm{P}(X_i \ge p) = 1 - F_X(p) = 1 - \left(\frac{p}{12}\right)^{\alpha}$$

The revenue from one student is p if they buy and 0 otherwise. Thus the expected revenue is

$$r(p,\alpha) = p\mathrm{P}(X_i \ge p) = p\left[1 - \left(\frac{p}{12}\right)^{\alpha}\right]$$

Our prospective business owner has circuitously rediscovered demand curves!

j) Eager at this revelation, the polling study is conducted and the MLE $\widehat{\alpha}$ is determined to be 2. Find the price p^* that maximizes the expected revenue $r(p)$.

Ans. We maximize

$$r(p,2) = p\left[1 - \left(\frac{p}{12}\right)^{2}\right] = p - \frac{p^3}{12^2}$$

over $0 \le p \le 12$. Differentiating and setting to 0, gives

$$r'(p) = 1 - \frac{3p^2}{12^2} \qquad p^* = \sqrt{\frac{144}{3}} = 4\sqrt{3}$$

To verify this is a maximum, we check the second derivative

$$r''(p) = -\frac{6p}{144} < 0$$

for $p > 0$, indeed this is a maximum. So the optimal price given our shop owner's estimator $\widehat{\alpha}$ is $4\sqrt{3}$, or around \$6.93 (not bad!). One note, the demand curve here is prescribed *probabilistically*, which is slightly different from what you may be used to in deterministic models used in introductory economics.

At price p, the probability that a randomly chosen student buys the drink is $\mathrm{P}(X \ge p) = 1 - F_X(p)$, which is in fact the complement of the CDF of X. The quantity of boba sold is then, this probability multiplied by a number of students.

Problem 4. Not Exam Advice

A student wants to understand how caffeine consumption relates to exam performance. For student i, let

$$x_i = \text{number of caffeinated drinks consumed in the previous 24 hours} \qquad y_i = \text{exam score}$$

The student polls some friends, and fits the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad \mathbb{E}[\varepsilon_i] = 0,\ \mathrm{Var}[\varepsilon_i] = \sigma^2$$

For the full dataset of $n = 8$ students, the following summary statistics are given:

$$\overline{x} = 3, \qquad \overline{y} = 86$$

$$S_{xx} = \sum_{i=1}^{8}(x_i - \overline{x})^2 = 28, \qquad S_{xy} = \sum_{i=1}^{8}(x_i - \overline{x})(y_i - \overline{y}) = 140$$

$$TSS = \sum_{i=1}^{8}(y_i - \overline{y})^2 = 1260$$

Afterwards, the student realizes there is actually another important distinction, and separates the data points into high and low sleepers. Within each group, the summary statistics are:

	n	$\overline{x}$	$\overline{y}$	S_{xx}	S_{xy}
Low sleep	4	1.5	73.5	5	-5
High sleep	4	4.5	98.5	5	-5

a) Compute the least-squares line of the overall data

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$$

Ans. Using the $\widehat{\beta}$ estimators from the least squares objective,

$$\widehat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{140}{28} = 5$$

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} = 86 - 5(3) = 71$$

So the overall fitted line is

$$\widehat{y} = 71 + 5x$$

Intuitively, each additional caffeinated drink is associated with 5 more exam points.

b) Compute the residual sum of squares RSS and R^2 for the overall data.

Ans. We compute RSS directly, using $\widehat{y}_i = \overline{y} + \widehat{\beta}_1(x_i - \overline{x})$

$$\begin{aligned}
RSS &= \sum_{i=1}^{8}(y_i - \widehat{y}_i)^2 \\
&= \sum_{i=1}^{8}\Big[(y_i - \overline{y}) - \widehat{\beta}_1(x_i - \overline{x})\Big]^2 \\
&= \sum_{i=1}^{8}(y_i - \overline{y})^2 - 2\widehat{\beta}_1 \sum_{i=1}^{8}(x_i - \overline{x})(y_i - \overline{y}) + \widehat{\beta}_1^2 \sum_{i=1}^{8}(x_i - \overline{x})^2 \\
&= TSS - 2\widehat{\beta}_1 S_{xy} + \widehat{\beta}_1^2 S_{xx} \\
&= 1260 - 1400 + 700 \\
&= 560
\end{aligned}$$

Now for the coefficient of determination,

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{560}{1260} = 1 - \frac{4}{9} = \frac{5}{9}$$

So the overall model explains about 55.6% of the variation in exam scores.

c) Compute the least-squares slope separately within each sleep group.

Ans. Again using the $\widehat{\beta}$ estimators from the least squares objective, for the low-sleep group,

$$\widehat{\beta}_{1,L} = \frac{S_{xy,L}}{S_{xx,L}} = \frac{-5}{5} = -1$$

Similarly for the high-sleep group,

$$\widehat{\beta}_{1,H} = \frac{S_{xy,H}}{S_{xx,H}} = \frac{-5}{5} = -1$$

So conditioned on each sleep group, each additional caffeinated drink is 1 fewer exam point.

d) Compare the overall slope with the within-group slopes. Can you explain what is happening?

Ans. The overall slope is positive as we have seen, $\widehat{\beta}_1 = 5$. But conditioned on the sleep group, $\widehat{\beta}_{1,L} = \widehat{\beta}_{1,H} = -1$. Thus the direction of the relationship reverses after conditioning on sleep group. This is Simpson's paradox.

The issue is that sleep group is a confounding variable. The high-sleep group both drinks more caffeine and scores higher, so the overall regression can be misleading.

e) Suppose the model errors in the overall caffeine-only regression are normally distributed with known standard deviation

$$\sigma = \sqrt{7}$$

Argue that the estimator $\widehat{\beta}_1$ is normally distributed and utilizing the transformation to the standard normal, give an approximate 95% confidence interval for the overall slope β_1.

Ans. Under the normal-error regression model, each observation satisfies

$$Y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$$

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{S_{xy}}{S_{xx}} \\
&= \frac{\sum_{i=1}^n (x_i - \overline{x})(Y_i - \overline{Y})}{S_{xx}} \\
&= \frac{\sum_{i=1}^n (x_i - \overline{x})Y_i}{S_{xx}} \\
&= \sum_{i=1}^n \frac{x_i - \overline{x}}{S_{xx}} Y_i
\end{aligned}$$

Since $\widehat{\beta}_1$ is computed as a linear combination of the observed Y_i's, it is also normally distributed. We find the distribution of $\widehat{\beta}_1$:

$$\begin{aligned}
\widehat{\beta}_1 &\sim \mathcal{N}\left(\sum_{i=1}^{n} \frac{x_i - \overline{x}}{S_{xx}}(\beta_0 + \beta_1 x_i),\ \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{S_{xx}}\right)^2 \sigma^2\right) \\
&= \mathcal{N}\left(\frac{\beta_0 \sum_{i=1}^{n}(x_i - \overline{x}) + \beta_1 \sum_{i=1}^{n}(x_i - \overline{x})x_i}{S_{xx}},\ \frac{\sigma^2 \sum_{i=1}^{n}(x_i - \overline{x})^2}{S_{xx}^2}\right) \\
&= \mathcal{N}\left(\frac{0 + \beta_1 S_{xx}}{S_{xx}},\ \frac{\sigma^2 S_{xx}}{S_{xx}^2}\right) \\
&= \mathcal{N}\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)
\end{aligned}$$

Here $\sigma^2 = 7$ and $S_{xx} = 28$ so $\text{Var}(\widehat{\beta}_1) = 1/4$. Hence the transformation to the standard normal is

$$Z = \frac{\widehat{\beta}_1 - \beta_1}{1/2} \sim \mathcal{N}(0, 1)$$

Therefore an approximate 95% confidence interval is

$$\widehat{\beta}_1 \pm z_{0.025}\left(\frac{1}{2}\right) = 5 \pm \frac{1.96}{2}$$

Intuitively, this confidence interval describes uncertainty in the overall caffeine-only regression slope.

Problem 5. Menu Combinations

A food truck sells rice bowls. There are 4 bases, 5 proteins, 3 sauces, and 6 toppings. A regular bowl has 1 base, 1 protein, 1 sauce, and 1 topping. A deluxe bowl has 1 base, 2 different proteins, 2 different sauces, and 2 different toppings.

a) How many regular bowls are possible?

Ans. We have disjoint "tasks" corresponding to each component of the bowl. Using the product rule, this is the product of the number of choices per each task. $4 \cdot 5 \cdot 3 \cdot 6 = 360$.

b) How many deluxe bowls are possible?

Ans. Expanding on the prior part, we want the number of possible subsets of each "task", these are unordered so we use the combination $4\binom{5}{2}\binom{3}{2}\binom{6}{2} = 4 \cdot 10 \cdot 3 \cdot 15 = 1800$.

c) Suppose 7 students sit in a row, but two particular students refuse to sit next to each other. How many arrangements are possible?

Ans. We use the inclusion-exclusion principle. Ignoring the restriction, there are 7! total ordered arrangements of 7 students.

Now count the arrangements where the two students are sat adjacently. Treating them as a block, this group and the other 5 students can be arranged in 6! ways, and the two students have 2 internal orders. Hence the number with them adjacent is $2 \cdot 6!$.

The desired count is $7! - 2 \cdot 6! = 3600$.

d) A randomly chosen regular bowl is equally likely among all regular bowls. Let A be the event that base 1 is chosen and B be the event that topping 1 is chosen. Are A and B independent?

Ans. We have $\mathrm{P}(A) = 1/4$ and $\mathrm{P}(B) = 1/6$. The joint probability is $\mathrm{P}(A \cap B) = 1/(4 \cdot 6) = 1/24 = \mathrm{P}(A)\mathrm{P}(B)$, hence independent.

Problem 6. Disease Testing

A disease affects 2% of a population. A rapid test is positive with probability 0.96 for someone with the disease. For someone without the disease, the test is falsely positive with probability 0.07. A symptom checker, conditionally independent of the rapid test given disease status, flags someone with probability 0.80 if they have the disease and with probability 0.10 otherwise.

a) Find the probability that both the test and symptom checker are positive.

Ans. Let D be disease, T test positive, and S symptom checker positive. We want to find $\mathrm{P}(T \cap S)$. Since we are given *conditional* probabilities in the problem statement, we calculate this through conditioning on D

$$\begin{aligned} \mathrm{P}(T \cap S) &= \mathrm{P}((T \cap S)|D)\mathrm{P}(D) + \mathrm{P}((T \cap S)|D^c)\mathrm{P}(D^c) \\ &= 0.02(0.96)(0.80) + 0.98(0.07)(0.10) \\ &= 0.01536 + 0.00686 = 0.02222 \end{aligned}$$

b) Given that both are positive, find the probability that the person has the disease.

Ans. We use Bayes. The denominator is the total probability calculated in the previous part.

$$\mathrm{P}(D \mid T \cap S) = \frac{\mathrm{P}((T \cap S)|D)\mathrm{P}(D)}{\mathrm{P}(T \cap S)} = \frac{0.02(0.96)(0.80)}{0.02222} \approx 0.6913.$$

c) Explain why the probability of having the disease given the positive rapid test and the flagged symptom checker result is much larger than the probability from the positive rapid test alone.

Ans. The symptom checker supplies another conditionally informative signal. A false rapid-test positive can happen with probability 0.07, but a false positive on both signals has probability $0.07 \cdot 0.10 = 0.007$ among non-diseased people, making the combined evidence stronger.

Problem 7. Functions of RVs

Let X have probability mass function

x	-2	-1	0	1	2
$p_X(x)$	c	$2c$	$3c$	$2c$	c

a) Find c.

Ans. Normalization gives that $\sum_x p_X(x) = c + 2c + 3c + 2c + c = 9c = 1$, so $c = 1/9$.

b) Compute $\mathbb{E}[X]$, $\mathrm{Var}(X)$, and $\mathbb{E}[X^3]$.

Ans. Notice symmetry, $\mathbb{E}[X] = 0$. Next,

$$\mathbb{E}[X^2] = \sum_x x^2 p_X(x) = 4 \cdot \frac{1}{9} + 1 \cdot \frac{2}{9} + 0 + 1 \cdot \frac{2}{9} + 4 \cdot \frac{1}{9} = \frac{12}{9} = \frac{4}{3}$$

Therefore $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 4/3 - 0 = 4/3$.

Recall the odd moments of symmetric random variables about 0 vanish. Precisely, note that X and $-X$ have the same distribution, so for odd n, $(-X)^n = -X^n$ and

$$\mathbb{E}[X^n] = \mathbb{E}[(-X)^n] = -\mathbb{E}[X^n]$$

Ergo, $\mathbb{E}[X^n] = 0$. Of course, you can also verify this by the standard definition.

c) Let $Y = X^2$. Find the PMF of Y.

Ans. The attainable values, the support of Y, is given by $\{0, 1, 4\}$. We have

$$\mathrm{P}(Y = 0) = \frac{3}{9} = \frac{1}{3}, \quad \mathrm{P}(Y = 1) = \frac{2}{9} + \frac{2}{9} = \frac{4}{9}, \quad \mathrm{P}(Y = 4) = \frac{1}{9} + \frac{1}{9} = \frac{2}{9}.$$

d) Are X and $Y = X^2$ independent?

Ans. No. For example, $\mathrm{P}(Y = 0) = 1/3$, but $\mathrm{P}(Y = 0 \mid X = 0) = 1$. If $Y = f(X)$ where f is nontrivial (not, say a constant), then Y and X are dependent, $Y \not\perp X$.

Problem 8. Discrete RVs

A machine produces independent items, each defective with probability $p = 0.08$.

a) Let X be the number of items inspected until the first defective item, counting the defective item. Identify the distribution of X and derive $\mathbb{E}[X]$.

Ans. This is a discrete distribution until the first "success", which is here, a defective item. It follows the Geometric distribution, with probability 0.08, $X \sim \text{Geometric}(0.08)$.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{\infty} x p_X(x) \\
&= \sum_{x=1}^{\infty} x(1-p)^{x-1} p \\
&= \sum_{x=1}^{\infty} (x-1)(1-p)^{x-1} p + \underbrace{\sum_{x=1}^{\infty} (1-p)^{x-1} p}_{\text{normalization of Geo(p)}} \\
&= 1 + (1-p) \sum_{x=1}^{\infty} (x-1)(1-p)^{x-1} p \\
&= 1 + (1-p)\mathbb{E}[X]
\end{aligned}$$

Solving this recursive relation, we have $p\mathbb{E}[X] = 1$, so $\mathbb{E}[X] = 1/p = 1/0.08 = 12.5$

b) Compute $\mathrm{P}(X > 20)$.

Ans. The first 20 items must be non-defective, so $\mathrm{P}(X > 20) = (1 - 0.08)^{20} = 0.92^{20}$

c) Given that the first 20 items are not defective, compute the probability that more than 30 items are needed until the first defective item.

Ans. By memorylessness, conditioning on earlier failures is independent of the trials needed to the next success

$$\mathrm{P}(X > 30 \mid X > 20) = \mathrm{P}(X > 10) = 0.92^{10}.$$

d) Let Y be the number of defective items among the first 50 items. Identify the distribution of Y and write $\mathrm{P}(Y \leq 3)$ as a summation.

Ans. Each item is independent, and we have 50 of them, ergo $Y \sim \text{Binomial}(50, 0.08)$ and

$$\mathrm{P}(Y \leq 3) = \sum_{k=0}^{3} \binom{50}{k} (0.08)^k (0.92)^{50-k}.$$

Problem 9. Triangles

Let X have density

$$f_X(x) = \begin{cases} cx, & 0 \leq x \leq 1, \\ c(2-x), & 1 < x \leq 2, \\ 0, & \text{otherwise} \end{cases}$$

a) Find c.

Ans. Normalization requires $\int_{\mathbb{R}} f_X(x)dx = 1$

$$c\int_0^1 x\,dx + c\int_1^2 (2-x)\,dx = c\left(\frac{1}{2}+\frac{1}{2}\right) = c$$

so $c = 1$.

b) Find the CDF $F_X(x)$.

Ans. We evaluate the disjoint integral across the domain to get

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \frac{x^2}{2}, & 0 \leq x \leq 1, \\ 1 - \frac{(2-x)^2}{2}, & 1 < x \leq 2, \\ 1, & x > 2. \end{cases}$$

c) Compute $\mathbb{E}[X]$ without integrating.

Ans. The PDF is symmetric about $x = 1$, so $\mathbb{E}[X] = 1$.

d) Now consider random variables X, Y that admit a joint PDF

$$f_{XY} = \begin{cases} 2, & 0 < y < x < 1, \\ 0, & \text{otherwise} \end{cases}$$

Find the marginal densities $f_X(x)$ and $f_Y(y)$.

Ans. For $0 < x < 1$, we integrate out Y, which is upper bounded by the line $y = x$

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}\,dy = \int_0^x 2\,dy = 2x$$

For $0 < y < 1$,

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}\,dx = \int_y^1 2\,dx = 2(1-y).$$

e) Are X and Y independent?

Ans. No. $f_X(x)f_Y(y) = 4x(1-y)$ is not equal to 2 on the support.

Problem 10. RV Functions

Let X, Y, Z be independent random variables such that

$$X \sim \text{Uniform}(-1,1), \qquad Y \sim \text{Uniform}(-1,1), \qquad \mathbb{E}[Z]=0, \qquad \text{Var}(Z)=1$$

Define

$$U = X + Z, \qquad V = Y + 2Z, \qquad R = X^2$$

a) Find the CDF of R.

Ans. For $0 \le r \le 1$,

$$F_R(r) = \text{P}(R \le r) = \text{P}(X^2 \le r) = \text{P}(-\sqrt{r} \le X \le \sqrt{r})$$

Since $X \sim \text{Uniform}(-1,1)$, this is the width of the interval multiplied by the height of the distribution, $1/2$. Ergo

$$F_R(r) = \frac{2\sqrt{r}}{2} = \sqrt{r}$$

Thus

$$F_R(r) = \begin{cases} 0, & r < 0, \\ \sqrt{r}, & 0 \le r \le 1, \\ 1, & r > 1. \end{cases}$$

For continuous RVs, it is often the case that we find the CDF before finding the PDF, if it exists. The CDF is a more fundamental concept, whilst the PDF requires continuity.

b) Find the PDF of R.

Ans. For $0 < r < 1$,

$$f_R(r) = F_R'(r) = \frac{1}{2\sqrt{r}}$$

So

$$f_R(r) = \begin{cases} \frac{1}{2\sqrt{r}}, & 0 < r < 1, \\ 0, & \text{otherwise} \end{cases}$$

c) Compute $\mathbb{E}[R]$ and $\text{Var}(R)$.

Ans. Since $R = X^2$,

$$\mathbb{E}[R] = \mathbb{E}[X^2]$$

For $X \sim \text{Uniform}(-1,1)$,

$$\mathbb{E}[X^2] = \int_{-1}^{1} x^2 \frac{1}{2}\, dx = \frac{1}{3}$$

Also

$$\mathbb{E}[R^2] = \mathbb{E}[X^4] = \int_{-1}^{1} x^4 \frac{1}{2}\, dx = \frac{1}{5}$$

Therefore

$$\text{Var}(R) = \mathbb{E}[R^2] - \mathbb{E}[R]^2 = \frac{1}{5} - \frac{1}{9} = \frac{4}{45}.$$

d) Compute $\mathrm{Cov}(U, V)$.

Ans. Using bilinearity of covariance,

$$\mathrm{Cov}(U, V) = \mathrm{Cov}(X + Z, Y + 2Z)$$

Expanding,

$$\mathrm{Cov}(U, V) = \mathrm{Cov}(X, Y) + 2\,\mathrm{Cov}(X, Z) + \mathrm{Cov}(Z, Y) + 2\,\mathrm{Var}(Z)$$

Since X, Y, Z are independent, all covariance terms involving different variables are zero. Therefore

$$\mathrm{Cov}(U, V) = 2\,\mathrm{Var}(Z) = 2.$$

e) Compute $\mathrm{Var}(U)$, $\mathrm{Var}(V)$, and $\mathrm{Cor}(U, V)$.

Ans. Since X and Z are independent,

$$\mathrm{Var}(U) = \mathrm{Var}(X + Z) = \mathrm{Var}(X) + \mathrm{Var}(Z) = \frac{1}{3} + 1 = \frac{4}{3}$$

Similarly,

$$\mathrm{Var}(V) = \mathrm{Var}(Y + 2Z) = \mathrm{Var}(Y) + 2^2\,\mathrm{Var}(Z) = \frac{1}{3} + 4 = \frac{13}{3}$$

Therefore

$$\mathrm{Cor}(U, V) = \frac{\mathrm{Cov}(U, V)}{\sqrt{\mathrm{Var}(U)}\sqrt{\mathrm{Var}(V)}} = \frac{2}{\sqrt{4/3}\sqrt{13/3}} = \frac{3}{\sqrt{13}}.$$

f) Are U and V independent?

Ans. No. $\mathrm{Cov}(U, V) = 2 \neq 0$

g) Are R and V independent?

Ans. Yes. The random variable R depends only on X, whilst

$$V = Y + 2Z$$

depends only on Y and Z. Since X is independent of both Y and Z, any function of X is independent of any function of (Y, Z). Thus R and V are independent.

Problem 11. LLN vs CLT

Let $X_1, \dots, X_n$ be i.i.d. with $\mathbb{E}[X_i] = 40$ and $\mathrm{Var}(X_i) = 25$. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

a) State the weak law of large numbers for $\overline{X}_n$.

Ans. For every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}(|\overline{X}_n - 40| > \varepsilon) = 0.$$

b) Use Chebyshev's inequality to bound $\mathrm{P}(|\overline{X}_{100} - 40| \geq 1)$.

Ans. $\mathrm{Var}(\overline{X}_{100}) = 25/100 = 1/4$. Hence

$$\mathrm{P}(|\overline{X}_{100} - 40| \geq 1) \leq \frac{1/4}{1^2} = \frac{1}{4}.$$

c) Use the CLT to approximate $\mathrm{P}(39 < \overline{X}_{100} < 41)$ in terms of Φ.

Ans. The standard deviation of $\overline{X}_{100}$ is $5/10 = 1/2$. Therefore

$$\mathrm{P}(39 < \overline{X}_{100} < 41) \approx \Phi\left(\frac{41-40}{1/2}\right) - \Phi\left(\frac{39-40}{1/2}\right) = \Phi(2) - \Phi(-2).$$

d) Explain why the Chebyshev bound and the CLT approximation are not the same kind of statement.

Ans. Chebyshev gives a rigorous bound that is a *finite sample bound*, ergo it is always valid but often not very tight. The CLT gives an *asymptotic approximation* that may be accurate for large n, but it uses limiting behavior and is not a guarantee for finite samples.

Problem 12. Sample Variance Bias

Let $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance σ^2. Define

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

a) Show that $\mathbb{E}[S_n^2] = \frac{n-1}{n}\sigma^2$.

Ans. Use the identity

$$\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\overline{X}_n - \mu)^2$$

Taking expectations gives

$$\mathbb{E}\left[\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\right] = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2$$

Dividing by n gives $\mathbb{E}[S_n^2] = \frac{n-1}{n}\sigma^2$.

b) Give an unbiased estimator of σ^2.

Ans. $\widehat{\sigma}^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$.

Problem 13. Least Squares

For data (x_i, y_i), fit $y_i \approx \beta_0 + \beta_1 x_i$ by minimizing

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

a) Derive the normal equations.

Ans. Differentiate and set equal to zero:

$$\frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

Thus

$$\sum y_i = n\beta_0 + \beta_1 \sum x_i, \qquad \sum x_i y_i = \beta_0 \sum x_i + \beta_1 \sum x_i^2.$$

b) Show that the fitted line passes through $(\overline{x}, \overline{y})$.

Ans. Divide the first normal equation by n:

$$\overline{y} = \beta_0 + \beta_1 \overline{x}$$

Thus the fitted line passes through $(\overline{x}, \overline{y})$.

c) Give the formula for $\widehat{\beta}_1$.

Ans. Using $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$ in the second normal equation, modulo some algebra work, gives

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}.$$

d) If $\widehat{\beta}_1 = 0$, must $R^2 = 0$?

Ans. In simple linear regression of the form we have identified, yes. If $\widehat{\beta}_1 = 0$, then $\widehat{\beta}_0 = \overline{y}$. Directly, we have

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} = 1 - 1 = 0$$

Intuitively, you can think the regression explains no variation beyond the mean and naturally the coefficient of determination is 0. Note this is not generally true of other regression forms.

Problem 14. Cramer Rao

Suppose $X_1, \dots, X_n$ are i.i.d. Exponential(λ) with density $f_\lambda(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.

a) Find the MLE of λ.

Ans. The likelihood is $L(\lambda) = \lambda^n e^{-\lambda \sum X_i}$, so

$$\ell(\lambda) = n \log \lambda - \lambda \sum X_i$$

Setting $\ell'(\lambda) = n/\lambda - \sum X_i = 0$ gives

$$\widehat{\lambda} = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\overline{X}}.$$

b) What does the LLN imply about $\widehat{\lambda}$?

Ans. Since $\mathbb{E}[X_i] = 1/\lambda$, the LLN gives $\overline{X} \to 1/\lambda$ in probability. By continuity, $1/\overline{X} \to \lambda$ in probability. Thus $\widehat{\lambda}$ is consistent.

c) Compute the Fisher information $I_n(\lambda)$.

Ans. Because

$$\ell''(\lambda) = -\frac{n}{\lambda^2}$$

we have $I_n(\lambda) = n/\lambda^2$.

d) State the Cramer–Rao lower bound for unbiased estimators of λ.

Ans. Any unbiased estimator T of λ satisfies

$$\text{Var}(T) \geq \frac{\lambda^2}{n}.$$

Problem 15. Confidence Interval

A sample of $n = 64$ measurements has sample mean $\overline{x} = 12.4$. The distribution is known to be Gaussian and the population standard deviation is known to be $\sigma = 3.2$.

a) Give a 95% confidence interval for the population mean μ using $z_{0.025} = 1.96$.

Ans. The standard error is $\sigma/\sqrt{n} = 3.2/8 = 0.4$. Thus

$$12.4 \pm 1.96(0.4) = 12.4 \pm 0.784$$

so the interval is $(11.616, 13.184)$.

b) What changes if σ is unknown and estimated by the sample standard deviation s?

Ans. Use a t critical value with $n - 1 = 63$ degrees of freedom and apply the transformation $s/\sqrt{n}$ instead of $\sigma/\sqrt{n}$. Note, I am not sure if this is covered in this course, but if so it should be just about the last topic.

c) Interpret the 95% t-confidence level.

Ans. The procedure has 95% long-run coverage: if we repeatedly sampled and constructed intervals this way, about 95% of the intervals would contain μ. This does *not guarantee* that this specific fixed interval has a 95% chance of containing μ.