Math Camp 2025 – Problem Set 10

Read the following problems carefully and justify all your work. Avoid using calculators or computers. (This was adapted almost verbatim from a problem set by Christopher Lucas.)

1. Suppose that X and Y are independently distributed random variables with $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. Note that μ_X represents the expectation of X, σ_X represents the standard deviation, and σ_X^2 the variance.

Calculate the following quantities:

1. $\mathbb{E}(5X - 2Y + 8)$
2. $\text{Var}(X + 3Y)$
3. $\mathbb{E}(4X^2 - 10XY + 25Y^2)$

Answer.

1. $\mathbb{E}(5X - 2Y + 8) = 5\mathbb{E}(X) - 2\mathbb{E}(Y) - 8 = 5\mu_X - 2\mu_Y - 8$.
2. Using that X, Y are independent, $X, 3Y$ are also independent [not entirely obvious, but intuitive, and you can verify it using the definition]. Therefore $\text{Var}(X + 3Y) = \text{Var}(X) + \text{Var}(3Y) = \text{Var}(X) + 9\,\text{Var}(Y) = \sigma_X^2 + 9\sigma_Y^2$.
3. We have $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, so $\sigma_X^2 = \mathbb{E}(X^2) - \mu_X^2$, and $\mathbb{E}(X^2) = \mu_X^2 + \sigma_X^2$. Similarly, $\mathbb{E}(Y^2) = \mu_Y^2 + \sigma_Y^2$. We have $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) = \mu_X\mu_Y$ because X, Y are independent. Therefore,

$$\begin{aligned}\mathbb{E}(4X^2 - 10XY + 25Y^2) &= 4\mathbb{E}(X^2) - 10\mathbb{E}(XY) + 25\mathbb{E}(Y^2)\\ &= 4(\mu_X^2 + \sigma_X^2) - 10\mu_X\mu_Y + 25(\mu_Y^2 + \sigma_Y^2).\end{aligned}$$

2. X and Y are discrete random variables with the following joint distribution:

	X			
Y	1	2	3	4
1	.10	.07	.03	.01
2	.08	.13	.04	.02
3	.03	.04	.11	.09
4	.02	.03	.12	.08

Answer the following questions:

(a) Calculate $\mathbb{E}[X]$, $\mathbb{E}[Y]$, and $\mathbb{E}[XY]$.

(b) Calculate $\text{Var}[X]$ and $\text{Var}[Y]$.

(c) Prove that the following two expressions of the covariance between X and Y are equivalent:

$$\text{cov}(X,Y) = \mathbb{E}\big[\big(X - \mathbb{E}[X]\big)\big(Y - \mathbb{E}[Y]\big)\big]$$
$$\text{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

(d) Calculate $\text{cov}(X,Y)$.

(e) Calculate the correlation between X and Y: ρ_{XY} with the following formula:

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{\text{Var}[X]\,\text{Var}[Y]}}$$

(f) Write out the probability mass function (PMF) of the random variable $Z = \mathbb{E}[Y \mid X]$.

Comment. This means: find the possible values that $Z = \mathbb{E}[Y \mid X]$ can take, and calculate the probability that it takes each of those values.

Answer.

(a) Recall that the expectation of a discrete random variable X is:

$$\mathbb{E}[X] = \sum_x \Pr(X = x) \cdot x$$

Then we have:

$$\begin{aligned}
\mathbb{E}[X] &= \Pr(X=1)\cdot 1 + \Pr(X=2)\cdot 2 + \Pr(X=3)\cdot 3 + \Pr(X=4)\cdot 4 \\
&= 1\cdot(0.10+0.08+0.03+0.02) + 2\cdot(0.07+0.13+0.04+0.03) \\
&\quad + 3\cdot(0.03+0.04+0.11+0.12) + 4\cdot(0.01+0.02+0.09+0.08) \\
&= 1\cdot 0.23 + 2\cdot 0.27 + 3\cdot 0.30 + 4\cdot 0.20 \\
&= 2.47
\end{aligned}$$

Similarly, we have:

$$\mathbb{E}[Y] = 2.56$$
$$\mathbb{E}[XY] = 6.89$$

(b) Recall the variance of a random variable X is

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

And the expectation of the function of a discrete random variable X is

$$\mathbb{E}[f(X)] = \sum_x f(x) \cdot \Pr(X = x).$$

Then we have

$$\begin{aligned}
\mathbb{E}[X^2] &= \Pr(X = 1) \cdot 1^2 + \Pr(X = 2) \cdot 2^2 + \Pr(X = 3) \cdot 3^2 + \Pr(X = 4) \cdot 4^2 \\
&= 1 \cdot (0.10 + 0.08 + 0.03 + 0.02) + 4 \cdot (0.07 + 0.13 + 0.04 + 0.03) \\
&\quad + 6 \cdot (0.03 + 0.04 + 0.11 + 0.12) + 16 \cdot (0.01 + 0.02 + 0.09 + 0.08) \\
&= 7.21.
\end{aligned}$$

Therefore, $\mathrm{Var}[X] = 7.21 - 2.47^2 = 1.1091$. Similarly, we have

$$\mathrm{Var}[Y] = 1.1664.$$

(c) Using linearity of expectations, we have:

$$\begin{aligned}
\mathrm{cov}(X, Y) &= \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big] \\
&= \mathbb{E}\big[XY - \mathbb{E}[X]Y - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]\big] \\
&= \mathbb{E}[XY] - \mathbb{E}\big[\mathbb{E}[X]Y\big] - \mathbb{E}\big[X\mathbb{E}[Y]\big] + \mathbb{E}\big[\mathbb{E}[X]\mathbb{E}[Y]\big] \\
&\qquad (\mathbb{E}[X],\ \mathbb{E}[Y] \text{ are constants}) \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

(d) $\mathrm{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 6.89 - 2.47 \cdot 2.56 = 0.5668$.

(e) $$\begin{aligned}
\rho_{XY} &= \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} \\
&= \frac{0.5668}{\sqrt{1.1091 \cdot 1.1664}} \\
&\approx 0.4983
\end{aligned}$$

(f) Recall that the PMF of a discrete random variable is simply $f_X(x) \equiv \Pr(X = x)$.

Note that the random variable $Z = \mathbb{E}[Y \mid X]$ takes different values based on the value of x. For example:

$$\mathbb{E}[Y \mid X = 1] = \frac{1}{0.10 + 0.08 + 0.03 + 0.02} \cdot \left(1 \cdot 0.10 + 2 + 0.08 + 3 \cdot 0.03 + 4 \cdot 0.02\right) = \frac{43}{23}$$

Similarly:

$$\mathbb{E}[Y \mid X = 2] = \frac{57}{27}$$
$$\mathbb{E}[Y \mid X = 3] = \frac{46}{15}$$
$$\mathbb{E}[Y \mid X = 4] = \frac{16}{5}$$

Therefore, the PMF of Z is:

$$f_Z(z) = \Pr(Z = z) = \begin{cases} 0.23 & \text{if } z = \frac{43}{23} \\ 0.27 & \text{if } z = \frac{57}{27} \\ 0.30 & \text{if } z = \frac{46}{15} \\ 0.20 & \text{if } z = \frac{16}{5} \\ 0 & \text{otherwise.} \end{cases}$$

3. Imagine that every person in the United States has a fixed preference for redistribution, which is defined as a continuous variable, where smaller values mean that the individual favors less redistribution.

Let N equals the size of the entire population of the US. Let Y_i be the preference for redistribution of person $i \in \{1, \ldots, N\}$, and Y be preference for redistribution of a person chosen uniformly at random. Suppose that the average preference for redistribution of the entire population is μ and the variance is σ^2.

As political scientists, we would like to make inferences about the aggregated preference for redistribution of the entire population, but we can't go out and measure every single person's view. So instead, we are going to sample n people from the entire population at random (but not necessarily with the same probability), and measure their preferences for each person in our sample (imagine for now that every person who is sampled responds, and that we measure their views correctly each time).

Let S_i be an indicator variable for person i being sampled, i.e., $S_i = 1$ if i is in the sample and $S_i = 0$ if not. Let S be an indicator variable for a person chosen uniformly at random being sampled.

Let $\hat{\mu}$ be the average preference for redistribution in our sample. The "hat" notation indicates that $\hat{\mu}$ is intended to estimate μ using only our sample.

Now answer the following questions:

(a) Consider Y_i, S_i, μ, and $\hat{\mu}$. Which of these are random variables? Which, if any, are not? Please make sure to explain your answer.

(b) Conceptually, what does $\mathbb{E}[Y]$ refer to?

(c) Conceptually, what do $\mathbb{E}[\hat{\mu}]$ and $\text{Var}[\hat{\mu}]$ refer to?

(d) Calculate $\mathbb{E}[S]$ and $\text{Var}[S]$.

(e) Write $\hat{\mu}$ as an expression of n, N, S_i, and Y_i.

(f) Write $\mathbb{E}[\hat{\mu}]$ as an expression of n, N, S and Y. What is $\mathbb{E}[\hat{\mu}]$ if S and Y are independent?

Hint. If I is selected uniformly at random from $\{1, \ldots, N\}$ and $X_1, \ldots, X_N$ are numbers then $\mathbb{E}[X_I] = \frac{1}{N}\sum_{i=1}^{N} X_i$. Notice that $Y = Y_I$ and $S = S_I$.

(g) If S and Y were not independent, would we necessarily get the same result as in (f)? Practically, what does this mean for surveys? Could you provide an example where S and Y might not be independent?

Answer.

(a) There are two sources of randomness: the sampling (i.e., who we are including in the sample), and which individual we are selecting when we calculate S and Y. The sampling is the choice of $S_1, \ldots, S_N$. Which individual we are selecting is a variable I that results from choosing an individual uniformly at random. Given I, we define $S = S_I$ and $Y = Y_I$. Notice that we are implicitly taking these two sources of randomness, $(S_1, \ldots, S_N)$ and I, to be independent.

So, Y_i for each $i \in \{1, \ldots, N\}$ are just numbers. Not random. S_i for each i are random variables, because we said that we are randomly sampling. The number μ is just the average of Y_i, so it's not random. Finally, $\hat{\mu}$ is random, because it is the average within the sample, and who is in the sample is random.

(b) Recall from the previous answer that $Y = Y_I$. $\mathbb{E}(Y)$ is the expected value of that. The only random part of Y is which individual we are choosing. Each one is selected with the same probability. So, $\mathbb{E}(Y)$ is the average of Y_i for $i \in \{1, \ldots, N\}$. It's just the population average μ.

(c) As we said, $\hat{\mu}$ is random because it depends on who is sampled. So, $\mathbb{E}(\hat{\mu})$ is the expected value of the sample average, and $\text{Var}(\hat{\mu})$ is the variance.

(d) We have to use the so-called "law of iterated expectations." First, if we have two discrete random variables X and Y and a value $y \in \mathbb{R}$ that Y can take, the *conditional expectation* of X given $Y = y$, or $\mathbb{E}(X|Y = y)$, is the expected value of X taking conditional probabilities with respect to the event $Y = y$:

$$\mathbb{E}(X|Y = y) = \sum_{x \in \text{supp}(X)} x \cdot \Pr(X = x \mid Y = y).$$

Then, the *conditional expectation* of X given Y is the random variable $f(Y)$ where $f(y) = \mathbb{E}(X|Y = y)$. The *law of iterated expectations* states that $\mathbb{E}[\mathbb{E}(X|Y)] = \mathbb{E}(X)$. (You can verify that it is true.)

The idea is that when we are calculating the expectation of X we can first calculate its expected value assuming that Y is fixed (so we get $\mathbb{E}(X|Y)$, which depends on Y), and then we calculate the expected value of that, taking into account that Y is random.

This is useful to calculate the expected value of $S = S_I$ because it has two sources of randomness: the random sample, i.e., the values of $S_1, \ldots, S_N$, and the particular individual we are choosing, $I \in \{1, \ldots, N\}$. These two sources of randomness are independent, so conditioning on one will not change the probabilities of the other one.

Noting this, we can use the law of iterated expectations:

$$\begin{aligned}
\mathbb{E}(S) = \mathbb{E}(S_I) &= \mathbb{E}[\mathbb{E}(S_I \mid S_1, \ldots, S_N)] && \text{(law of iterated expectations)} \\
&= \mathbb{E}\left[\sum_{i=1}^{N} S_i \Pr(I = i \mid S_1, \ldots, S_N)\right] && \text{(using that, given } S_1, \ldots, S_N, S_I \text{ is a function of } I) \\
&= \mathbb{E}\left[\sum_{i=1}^{N} S_i \Pr(I = i)\right] && \text{(using that } I \text{ is independent of } S_1, \ldots, S_N) \\
&= \mathbb{E}\left[\sum_{i=1}^{N} S_i \frac{1}{N}\right] && \text{(using that } I \text{ is uniform on } \{1, \ldots, N\}) \\
&= \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^{N} S_i\right] \\
&= \mathbb{E}\left[\frac{1}{N} n\right] && \text{(using that } n \text{ people are in the sample)} \\
&= \frac{n}{N} && (\mathbb{E}(c) = c \text{ for } c \text{ constant}).
\end{aligned}$$

We can now calculate $\mathrm{Var}(S)$ with the formula

$$\begin{aligned}\mathrm{Var}(S) &= \mathbb{E}(S^2) - \mathbb{E}(S)^2 \\ &= \mathbb{E}(S) - \mathbb{E}(S)^2 && \text{(using that } S \text{ is 0 or 1, and so } S^2 = S) \\ &= \mathbb{E}(S)(1 - \mathbb{E}(S)) = \frac{n}{N}\left(1 - \frac{n}{N}\right).\end{aligned}$$

(e) Recall that $\hat{\mu}$ is the average of Y_i within the sample, which has n elements. We can take the sum over the whole population but only including the individuals who are in the sample; multiplying by S_i does precisely that. So,

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{N} S_i Y_i.$$

(f) The idea is that when we take the expectation of $SY = S_I Y_I$ taking $S_1, \ldots, S_N$ as fixed (not random), we get the average of $S_i Y_i$ over the population: $\frac{1}{N}\sum_{i=1}^{N} S_i Y_i$. Formally,

$$\mathbb{E}(SY \mid S_1, \ldots, S_N) = \frac{1}{N}\sum_{i=1}^{N} S_i Y_i.$$

So

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{N} S_i Y_i = \frac{N}{n}\frac{1}{N}\sum_{i=1}^{N} S_i Y_i = \frac{N}{n}\mathbb{E}(SY \mid S_1, \ldots, S_N).$$

Therefore,

$$\mathbb{E}(\hat{\mu}) = \mathbb{E}\left[\frac{N}{n}\mathbb{E}(SY \mid S_1, \ldots, S_N)\right] = \frac{N}{n}\mathbb{E}[\mathbb{E}(SY \mid S_1, \ldots, S_N)] = \frac{N}{n}\mathbb{E}(SY)$$

by the law of iterated expectations.

If S, Y are independent then $\mathbb{E}(SY) = \mathbb{E}(S)\mathbb{E}(Y)$. Now, $\mathbb{E}(S) = \frac{n}{N}$ and $\mathbb{E}(Y)$ is the population average μ. So, in this case, $\mathbb{E}(SY) = \frac{n}{N}\mu$, and

$$\mathbb{E}(\hat{\mu}) = \frac{N}{n}\mathbb{E}(SY) = \frac{N}{n}\frac{n}{N}\mu = \mu.$$

(g) In general, $\mathbb{E}(SY) = \mathbb{E}(S)\mathbb{E}(Y) + \mathrm{Cov}(S, Y)$, so

$$\mathbb{E}(\hat{\mu}) = \frac{N}{n}\mathbb{E}(SY) = \frac{N}{n}(\mathbb{E}(S)\mathbb{E}(Y) + \mathrm{Cov}(S, Y)) = \mu + \frac{N}{n}\mathrm{Cov}(S, Y).$$

If S, Y are independent then their covariance $\text{Cov}(S, Y)$ is 0, because

$$\text{Cov}(S, Y) = \mathbb{E}[(S - \mathbb{E}(S))(Y - \mathbb{E}(Y))] = \mathbb{E}[S - \mathbb{E}(S)]\mathbb{E}[Y - \mathbb{E}(Y)] = 0$$

using independence. But if they are not, they could, for example, be positively associated (meaning that S is more likely to be 1 if Y is larger, or, in other words, people with greater preference for redistribution are more likely to be sampled), which would create a positive bias ("bias" is the difference between the expectation of the estimator and the estimand). This is intuitive: if the probability of being sampled is not correlated with preferences for redistribution then the sample average should be unbiased. But if, say, we oversample rich people, who are likely to prefer less redistribution, then the sample average will be biased.

4. You are preparing to run a field experiment on the effectiveness of oversight in curbing corruption, as in Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy*, vol. 115, no.2.

As part of a major infrastructure project, the Indonesian government has allocated funds for roads in an "infinite" (we assume for the purpose of this question) number of villages. Your experimental intervention is whether you inform the village that: "after funds had been awarded but before construction began, the project would subsequently be audited by the central government audit agency."

After the road is constructed, you will conduct an extensive (and expensive) data-collection effort to estimate the amount actually spent on each village's roads, based on the quality of materials found in excavated core samples, estimated wages based on interviews, and so on. The discrepancy between allocated funds and estimated expenditures will be your outcome variable.

Let T_i be an indicator variable for the treatment, so $T_i = 1$ refers to village i has been "treated." Let Y_i be the outcome you measured as described above, then the observed outcome is a function of the treatment variable, i.e. $Y_i = Y_i(T_i)$. Each village i has two potential outcomes:

- $Y_i(0)$: the outcome that would be observed if village i is assigned to the "control," or non-informed, group.
- $Y_i(1)$: the outcome that would be observed if village i is assigned to the "treatment," or informed, group.

Because the field experiment is such a large undertaking, you want to carefully design your experiment to ensure success. At the same time, you can only afford to measure 100 villages. Suppose that you randomly sample N_T villages to assign to treatment and N_C to control (so $N_T + N_C = 100$). Now answer the following questions:

(a) Interpret the meaning of $Y_i(1) - Y_i(0)$ for a single village i.

(b) Interpret the meaning of $\mathbb{E}[Y_i(1) - Y_i(0)]$, where i now is a random village.

(c) Interpret the meaning of $\mathbb{E}[Y_i(0)]$ and $\mathbb{E}[Y_i(1)]$, where i denotes a random village. As an expert on local bureaucracy and corruption, what kind of relationship do you expect between these two quantities?

(d) Interpret the meaning of $\frac{1}{100}\sum_{i\in S}\big(Y_i(1) - Y_i(0)\big)$, where S denotes the set of villages in your sample of 100 villages. Compare it to the quantity in (b).

(e) Interpret the meaning of $\text{Var}[Y_i(0)]$ and $\text{Var}[Y_i(1)]$, where again i denotes a random village. Again, as an expert, what kind of relationship do you expect between these two quantities?

(f) Assume that the treatment assignment is random, that is, $\Pr(T_i = 1) = \frac{N_T}{100}$ for each $i \in S$. For simplicity, assume that villages $i = 1$ through $i = 100$ are in our sample, or equivalently, $S = \{1, 2, \ldots, 100\}$. Calculate the expectation of the following expression and interpret your result:

$$\hat{\tau} = \frac{1}{N_T}\sum_{i=1}^{100} T_i Y_i - \frac{1}{N_C}\sum_{i=1}^{100}(1 - T_i)Y_i.$$

Answer.

(a) $Y_i(1) - Y_i(0)$ is the treatment effect for the unit i. It's the difference in outcomes between receiving and not receiving the treatment.

(b) $\mathbb{E}[Y_i(1) - Y_i(0)]$ is the average treatment effect over the population of villages.

(c) $\mathbb{E}[Y_i(0)]$ is the average baseline outcome, without the intervention. $\mathbb{E}[Y_i(0)]$ is the average outcome if every village receives the treatment. The idea of the experiment is that we expect that auditing will reduce corruption. $Y_i(t)$ is actual expenditure on the road, so what we expect is that if we reduce corruption then village authorities will spend more money on the roads (rather than stealing it).

(d) $\frac{1}{100}\sum_{i\in S}(Y_i(1) - Y_i(0))$ is the average treatment effect within the sample. If the sample is uniformly random (i.e., each village has the same chance of being sampled) then it will be an unbiased estimator of the average treatment effect. (But notice we cannot calculate this estimator, because we can't observe both potential outcomes for any given village.)

(e) $\text{Var}(Y_i(0))$ and $\text{Var}(Y_i(1))$ are the variance of the outcomes without and with the treatment, respectively. We may expect that the variance in expenditures should be smaller with the treatment, if the effect is to completely reduce corruption, and we believe that baseline levels of corruption are heterogeneous.

(f) Recall that $Y_i = Y_i(T_i)$. We have

$$\begin{aligned}
\mathbb{E}(\hat{\tau}) &= \mathbb{E}\left[\frac{1}{N_T}\sum_{i=1}^{100} T_i Y_i(T_i) - \frac{1}{N_C}\sum_{i=1}^{100}(1-T_i)Y_i(T_i)\right] \\
&= \frac{1}{N_T}\sum_{i=1}^{100}\mathbb{E}[T_i Y_i(T_i)] - \frac{1}{N_C}\sum_{i=1}^{100}\mathbb{E}[(1-T_i)Y_i(T_i)] \\
&\qquad \text{(linearity of expectation)} \\
&= \frac{1}{N_T}\sum_{i=1}^{100}[1\cdot Y_i(1)\cdot \Pr(T_i=1) + 0\cdot Y_i(0)\cdot \Pr(T_i=0)] \\
&\quad - \frac{1}{N_C}\sum_{i=1}^{100}[0\cdot Y_i(1)\cdot \Pr(T_i=1) + 1\cdot Y_i(0)\cdot \Pr(T_i=0)] \\
&= \frac{1}{N_T}\sum_{i=1}^{100} Y_i(1)\cdot \Pr(T_i=1) - \frac{1}{N_C}\sum_{i=1}^{100} Y_i(0)\cdot \Pr(T_i=0) \\
&= \frac{1}{N_T}\sum_{i=1}^{100} Y_i(1)\cdot \frac{N_T}{100} - \frac{1}{N_C}\sum_{i=1}^{100} Y_i(0)\cdot \frac{N_C}{100} \\
&= \frac{1}{100}\sum_{i=1}^{100} Y_i(1) - \frac{1}{100}\sum_{i=1}^{100} Y_i(0) \\
&= \frac{1}{100}\sum_{i=1}^{100}(Y_i(1) - Y_i(0)),
\end{aligned}$$

which is the average treatment effect in the sample. If the sample itself is random, this will be an unbiased estimator of the average treatment effect.