Math 180A Homework 8

Fall 2021

Due date: **11:59pm** (Pacific Time) on **Mon., Nov. 29** (via Gradescope)

Section 0 (CAPE evaluations)

This is a reminder to fill out your CAPE evaluations! (`https://cape.ucsd.edu/students`)
If you do it now, you'll be done...

Section 1 (input directly in Gradescope)

Submit the answers to these problems directly through the Gradescope interface. You do not need to write up or explain your work.

Problem 1. True or false: For all random variables X, Y, and Z,

$$\operatorname{Corr}(X+Y,Z) = \operatorname{Corr}(X,Z) + \operatorname{Corr}(Y,Z).$$

Problem 2 (numerical answers)**.** Let X and Y be random variables with

$$E[X] = 1, \quad E[Y] = -2, \quad E[X^2] = 10, \quad E[Y^2] = 8, \quad E[XY] = 2.$$

(a) Compute $\operatorname{Var}(Y)$.

(b) Compute $\operatorname{Corr}(X,Y)$.

(c) Compute $\operatorname{Cov}(X-1, 2Y)$.

Section 2 (upload files)

For each problem, write your solution on a page by itself, and upload it as a separate file to Gradescope (either typed or scanned from handwritten work). You should write your solutions to these problems neatly and carefully and provide full justification for your answers.

Problem 3. Let (X, Y) be a uniformly random point in the square with corners $(1, 0)$, $(0, 1)$, $(-1, 0)$, and $(0, -1)$.

(a) What is the joint density function $f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$?

(b) What is the correlation $\mathrm{Corr}(X, Y)$?

(c) Are X and Y independent? Why or why not?

Problem 4. Suppose that Y is a random variable with moment generating function

$$M_Y(t) = \frac{1}{3} + \frac{1}{9}e^{-5t} + \frac{1}{18}e^{t} + \frac{1}{2}e^{2t}.$$

Compute the mean of Y in two ways:

(a) Compute $E[Y]$ from the derivative of $M_Y(t)$.

(b) Compute the distribution of Y as in Example 5.15, and then compute the mean of Y directly.

Problem 5. (a) Let $Y = aX + b$ where a and b are real numbers and X and Y are random variables. Express $M_Y(t)$ in terms of $M_X(t)$. **Hint:** Use the definition of the moment generating function and the properties of expectation.

(b) Let X be a random variable with distribution $\mathrm{Exp}(1/5)$, and let $Y = 2X + 1$. Compute $M_Y(t)$.

Problem 6 (ASV Exercise 5.18)**.** Let $X \sim \mathrm{Geom}(p)$.

(a) Compute the moment generating function of X. Be careful about the possibility that $M_X(t)$ may be infinite for some t.

(b) Compute the mean and variance of X from the moment generating function. (Note we computed the mean and variance earlier by a different method.)

Problem 7. You're out surfing and waiting for a big wave to appear. Because you're extremely bored, you decide to estimate how long you'll have to wait, by modeling the wait time with an exponential random variable X with mean 10 (minutes).

(a) Use Markov's inequality to estimate the probability that you have to wait at least 30 minutes.

(b) Use Chebyshev's inequality to estimate the probability that $X \geq 30$.

(c) Compute $P(X \geq 30)$ exactly.

Problem 8. For any random variable X and real number a, show that

$$P(X \geq a) \leq e^{-a} M_X(1).$$

Section 3 (practice on later topics – NOT TO BE TURNED IN)

These problems are intended to provide practice with the final topics covered in Math 180A – the Law of Large Numbers and the Central Limit Theorem. These topics are not covered in the homework, but will be included on the final exam.

Problem 1 (ASV Exercise 9.4)**.** The European style roulette wheel has the following probabilities: a red number appears with probability $\frac{18}{37}$, a black number appears with probability $\frac{18}{37}$, and a green number appears with probability $\frac{1}{37}$. Ben bets exactly \$1 on black each round. Explain why this is not a good long-term strategy.

Problem 2 (ASV Exercise 9.16)**.** Every morning I take either bus number 5 or bus number 8 to work. Every morning the waiting time for the number 5 is exponential with mean 10 minutes, while the waiting time for the number 8 is exponential with mean 20 minutes. Assume all waiting times independent of each other. Let S_n be the total amount of bus-waiting (in minutes) that I've done during n mornings, and let T_n be the number of times I've taken the number 5 bus during n mornings.

(a) Find the limit $\lim_{n\to\infty} P(S_n \leq 7n)$.

(b) Find the limit $\lim_{n\to\infty} P(T_n \geq 0.6n)$.

Problem 3 (ASV Exercise 4.20)**.** You flip a fair coin 10,000 times. Approximate the probability that the difference between the number of heads and number of tails is at most 100.

Problem 4 (ASV Exercise 9.7)**.** A car insurance company has 2,500 policy holders. The expected claim paid to a policy holder during a year is \$1,000 with a standard deviation of \$900. What premium should the company charge each policy holder to assure that with probability 0.999, the premium income will cover the cost of the claims?

(a) Answer the question using Chebyshev's inequality.

(b) Answer the question using the Central Limit Theorem.

Problem 5 (ASV Exercise 9.20)**.** Let $X_1, X_2, X_3, \ldots$ be i.i.d. random variables with mean zero and finite variance σ^2. Let $S_n = X_1 + \cdots + X_n$. Determine the limits below, with precise justifications.

(a) $\lim_{n\to\infty} P(S_n \geq 0.01n)$

(b) $\lim_{n\to\infty} P(S_n \geq 0)$

(c) $\lim_{n\to\infty} P(S_n \geq -0.01n)$

Problem 6. Harper and Heloise are real estate agents for a corporate firm. Once a week, each of them is assigned to close an important deal. It is known that one of the two associates closes her deals successfully 60 percent of the time (model these as independent coin tosses) and the other 50 percent (also independent coin tosses) but you are not sure which is which. You formulate a plan: you will wait N weeks, so that each associate gets to attempt N different deals, and then you will offer a permanent job to the associate who is ahead in number of closings. The **main question** we'd like to answer is this: roughly how large does N have to be to ensure that there is a 95 percent chance that the more capable closer (i.e., the one with closing probability .6) is ahead after N steps? We'll approximately solve this using the Central Limit Theorem in three steps:

(a) Let X_N and Y_N be the number of deals closed by (respectively) the more and less capable agents agent after N steps. So X_N and Y_N represent the number of heads in N tosses of a p-coin with (respectively) $p = .6$ and $p = .5$. Compute (in terms of N) the mean and variance of the random variable $S_N = X_N - Y_N$.

(b) For the random variable S_N, compute (in terms of N) how many standard deviations 0 is below the mean. That is, find $E(S_N)/SD(S_N)$ where SD denotes standard deviation.

(c) The Central Limit Theorem says that if N is large, both X_N and Y_N are approximately normal variables. Since X_N and Y_N are independent (and since the difference between two independent normal random variables is itself normal) one can argue that $S_N = X_N - Y_N$ is also roughly Gaussian. Use this fact to approximate $P(S_N > 0)$ when $N = 143$. We can interpret this as an approximation for the probability that S_N is positive (so the better closer wins). Conclude that 143 is roughly the answer to the main question.

Remark: Even though there is a *huge* difference between the two agents, it actually takes *years* to determine with confidence which is better. If you as the manager *think* you can tell based on just a few outcomes, you are deluding yourself — the noise to signal ratio is too high. This problem appeared (without the real estate agent story) in the 538 Riddler (which often has great probability puzzles) and also in an academic paper which surveyed financial experts to see how many flips they thought were necessary. Feel free to look up these references for more detailed calculations. The paper states:

"The median guess was 40 flips. While lower than the full-credit answer of 143, it does show that the respondents in general appreciate it takes a long time to identify a phenomenon with this kind of risk/reward ratio simply by history. We include in Appendix 1 the calculation used to arrive at 143.3. Our respondents are a pretty mathematical bunch, and we suspect that if they took their time to calculate an answer, rather than giving a quick guess as we requested, most would have arrived at the correct answer. But the point of the exercise was to illustrate how when we are thinking fast, we tend to overweight the value of small samples: a full 30% of respondents, the single largest bucket, thought 10 flips or less was sufficient. This built-in bias to over-weight small samples results in a tendency to ignore the investing dictum 'past performance is not indicative of future results' when we clearly should not."