

STA 130A – Homework 5

Submission due: Thu, November 5 at 11:59 PM PT

Instructor: Dogyoon Song

Instructions: Upload a single PDF file to Gradescope via Canvas (“Homework 5” under “Assignments”). Name the file using the prefix of your UC Davis email ID and the homework number (e.g., `dgsong_hw5.pdf`). Include “STA 130A,” your name, and the last four digits of your student ID on the front page. No late submissions will be accepted; any submission received after the deadline will receive 0 points. For full information about submission requirements and the late submission policy, see the syllabus.

Problem 1 (20 points in total). Covariance and variance: core calculations

For parts (a)–(b), suppose

$$\text{Var}(X) = 4, \quad \text{Var}(Y) = 9, \quad \text{Cov}(X, Y) = -3.$$

- (a) **(5 points)** Compute the correlation $\rho(X, Y)$. What does its sign indicate about how X and Y vary together?
- (b) **(5 points)** Compute $\text{Var}(X + Y)$ and $\text{Var}(2X - Y)$.
- (c) **(5 points)** In a separate model, let $G \sim \text{Bernoulli}(1/2)$, with

$$Z \mid G = 0 \sim \mathcal{N}(0, 1), \quad Z \mid G = 1 \sim \mathcal{N}(4, 9),$$

where the second normal parameter is the variance. Write $\mathbb{E}[Z \mid G]$ and $\text{Var}(Z \mid G)$ as functions of G . Then compute $\mathbb{E}[Z]$.

- (d) **(5 points)** For the model in part (c), compute

$$\mathbb{E}[\text{Var}(Z \mid G)] \quad \text{and} \quad \text{Var}(\mathbb{E}[Z \mid G]),$$

and use them to obtain $\text{Var}(Z)$. Which term represents variability within groups, and which represents variability between group means?

Problem 2 (20 points in total). Dependence and conditional variability

The joint PMF of X and Y is

$p_{X,Y}(x, y)$	$y = 0$	$y = 1$
$x = -1$	1/6	1/6
$x = 0$	1/3	0
$x = 1$	1/6	1/6

and is zero for all other pairs.

- (a) (5 points) Find the marginal PMFs of X and Y .
- (b) (5 points) Compute $\text{Cov}(X, Y)$ and $\rho(X, Y)$. Are X and Y independent? Justify directly from the table.
- (c) (5 points) For each $y \in \{0, 1\}$, compute $\mathbb{E}[X \mid Y = y]$ and $\text{Var}(X \mid Y = y)$. Use these results to verify the law of total variance for X .
- (d) (5 points) A student claims that observing Y cannot change the distribution of X , because their correlation is zero. Evaluate this claim by comparing $\text{Var}(X \mid Y = 1)$ with $\text{Var}(X)$. Explain why your comparison does not contradict the law of total variance.

Problem 3 (20 points in total). Building and using MGFs

For parts (a)–(c), let X have PMF

$$P(X = 0) = \frac{1}{4}, \quad P(X = 1) = \frac{1}{2}, \quad P(X = 2) = \frac{1}{4}.$$

- (a) (5 points) Find $M_X(t) = \mathbb{E}[e^{tX}]$ directly from the PMF.
- (b) (5 points) Use derivatives of M_X at zero to compute $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, and $\text{Var}(X)$.
- (c) (5 points) Let X_1, X_2 be independent copies of X . Find the MGFs of $X_1 + X_2$ and $2X$. Do these two random variables have the same distribution? Justify using the MGFs.
- (d) (5 points) Separately, let T have PDF $f_T(x) = 2e^{-2x}$ for $x \geq 0$, and zero otherwise. Derive $M_T(t)$ from its defining integral, including its range of finiteness.

Problem 4 (20 points in total). Markov and Chebyshev inequalities

Suppose $X \geq 0$, $\mathbb{E}[X] = 20$, and $\text{Var}(X) = 64$. No particular distribution is assumed.

- (a) (5 points) Use Markov's inequality to give an upper bound for $P(X \geq 50)$.
- (b) (5 points) Use Chebyshev's inequality to give another upper bound for $P(X \geq 50)$. State the event inclusion that justifies applying a two-sided deviation bound to this one-sided event. Which of your two bounds is smaller?
- (c) (5 points) Apply Chebyshev's inequality to $P(|X - 20| \geq 4)$. Combine the result with the fact that a probability is at most one. Does this give any information beyond that fact? Explain briefly.
- (d) (5 points) Find the largest possible value of $P(|X - 20| \geq 16)$ under the assumptions above. Justify your answer by giving an upper bound and constructing a PMF that attains it. Verify that your PMF has the required mean and variance.

Hint: Try three possible values, symmetric about 20.

Problem 5 (20 points in total). A random number of repair requests

A service center receives $N \sim \text{Poisson}(4)$ requests during an hour. Each request independently requires a replacement part with probability $1/4$. More precisely, let $B_1, B_2, \dots$ be independent Bernoulli($1/4$) variables, independent of N , and define

$$K = \sum_{i=1}^N B_i,$$

with $K = 0$ when $N = 0$. Thus K is the number of requests requiring a part.

- (a) **(5 points)** Identify the distribution of $K \mid N = n$, and give $\mathbb{E}[K \mid N = n]$ and $\text{Var}(K \mid N = n)$, including the case $n = 0$.
- (b) **(5 points)** Use total expectation and total variance to compute $\mathbb{E}[K]$ and $\text{Var}(K)$. Show both terms in the variance decomposition.
- (c) **(5 points)** Derive $M_K(t)$ by conditioning on N , and use it to identify the distribution of K .
Hint: First find $\mathbb{E}[e^{tK} \mid N = n]$, then average over N . You may use $M_N(u) = \exp\{4(e^u - 1)\}$.
- (d) **(5 points)** The center has three replacement parts available. Find the exact probability $P(K \geq 4)$ that this supply is insufficient. Also give the Markov and Chebyshev upper bounds for this probability. Report the exact probability and both bounds numerically to three decimal places.

Problem 6* (up to 10 bonus points). Two further connections

Each part is optional and worth up to 5 additional points.

- (a) **(5 bonus points)** Adapted from [BT08, Chapter 4, Problem 19]. Suppose

$$\mathbb{E}[X] = 0, \quad \mathbb{E}[X^2] = 1, \quad \mathbb{E}[X^3] = 0, \quad \mathbb{E}[X^4] = 3,$$

and let $Y = a + bX + cX^2$, where $b^2 + c^2 > 0$. Find $\rho(X, Y)$. Does Y being a function of X force $|\rho(X, Y)| = 1$? Use your formula to explain.

- (b) **(5 bonus points)** Return to K in Problem 5. For $t > 0$, apply Markov's inequality to e^{tK} to show that

$$P(K \geq 4) \leq e^{-4t} M_K(t).$$

Evaluate this bound at $t = \log 4$ and compare it with the bounds in Problem 5(d). No optimization over t is required.

References

- [BT08] Dimitri P. Bertsekas and John N. Tsitsiklis. *Introduction to Probability*, 2nd edition. Athena Scientific, 2008.