Homework 6

CS 70, Summer 2024

Due by Wednesday, July 31st at 11:59 PM

This content is protected and may not be shared, uploaded, or distributed.

Instructions. Start each problem on a separate page. The subparts of each problem can be on the same page. Every answer should contain a calculation or reasoning. Your answers should be clear, organized, and legible—your final submission should not include scratch work or failed attempts. You must always commit to a final answer; if multiple answers are provided, the most incorrect one will be graded. You may leave all algebraic expressions unsimplified, but you must simplify any integrals or infinite sums unless otherwise stated.

If you are completing the homework using LaTeX, you may use the templates. Homeworks must be submitted through Gradescope. See the end of the homework for submission instructions.

Sundry. Before you start writing your final homework submission, state briefly how you worked on it (e.g., if you went to office hours, how frequently you worked on it, etc.). If you worked on the assignment in a group with other students, list their names and email addresses.

1 Variances Practice

Find the requested variance or covariance in each of the following settings.

(a) A vendor sells 5 varieties of cookies. The vendor starts the day with 20 boxes of each variety. Over the course of the day, each of 25 customers buys one box at random from all the boxes which the vendor has at the time the customer visits. No other customer buys cookies that day.

Let V be the number of varieties which still have all 20 boxes left at the end of the day. Find $\text{Var}[V]$.

(b) During her turn in a game, Charlotte must roll a fair nine-sided die with faces numbered 1 through 9. Charlotte's score is computed as $|X - 5|$, where X is the result of the roll.

Find the variance of Charlotte's score.

(c) During his turn in a game, Casey must roll a fair six-sided die $n \geq 4$ times. Casey's score is the number of faces which appear exactly twice. For example, if $n = 8$ and Casey rolls $1, 3, 1, 1, 4, 2, 3, 4$, then his score is 2.

Find the variance of Casey's score.

(d) Arjun throws k balls independently at random into n bins. Let X be the number of balls which land in the first bin and Y be the number of balls which land in the last bin. Find $\text{Cov}[X, Y]$.

(e) Mosi rolls a six-sided die repeatedly. Let W be the number of rolls until Mosi has seen each of the six faces. Find $\text{Var}[W]$.

2 Gap Distributions

In the class and in Homework 5, we have examined the distributions of waiting times when we are sampling with replacement, e.g. the number of rolls of a die until we see the face with six spots or the number of calls until a professor has called on a data science major three times. In this question, we will examine the distribution of such waiting times when we instead sample without replacement.

Our motivating example will be the waiting time until the first ace is seen when cards are drawn from a standard deck at random without replacement.

Suppose cards are drawn at random without replacement from a standard deck of cards. Let

- G_1 be the number of cards before the first ace,
- G_2 be the number of cards after the first ace but before the second,
- G_3 be the number of cards after the second ace but before the third,
- G_4 be the number of cards after the third ace but before the fourth,
- G_5 be the number of cards after the fourth ace.

For example, if the four aces appear on the 12^{th}, 23^{rd}, 24^{th}, and 52^{nd} draws, then $G_1 = 11$, $G_2 = 10$, $G_3 = 0$, $G_4 = 27$, and $G_5 = 0$.

Note. This question has many parts. Parts **(a)**, **(e)**, and **(h)** are the important ones. The other parts should not take too long, and are only there to build up your answers to the important parts.

(a) Find the distribution of G_1.

(b) For each $k \in \{1, \ldots, 52\}$, find the chance that the card drawn on the k^{th} draw is an ace.

(c) Argue that G_1, G_2, G_3, G_4, and G_5 are identically distributed.

(*Hint.* Suppose without loss of generality that G_i is more likely to take on large values than G_j for some $i \neq j$. Argue that this contradicts part **(b)**.)

(d) Find the distribution of $G_1 + G_2 + G_3 + G_4 + G_5$ and briefly explain your answer.

(e) Let W_1 be the number of cards drawn until the first ace is seen. Use parts **(c)** and **(d)** to find $\mathrm{E}[G_1]$ and hence find $\mathrm{E}[W_1]$.

(f) To find $\mathrm{Var}[W_1]$, we'll need to consider another approach. Express G_1 as the sum of indicator random variables.

(g) Find the chance that any two specified cards which are not aces (e.g. the king of diamonds and the queen of hearts) are both before the first ace.

(h) Use parts **(f)** and **(g)** to find $\mathrm{Var}[W_1]$.

3 Mean Concentration

Let $X_1, \ldots, X_{25}$ be a sample drawn at random with replacement from a population of men with an average height of 68 inches and a standard deviation in height of 3 inches.

(a) If it possible, find the chance that the heights of the first two men in the sample differ by fewer than 12 inches. Otherwise, provide the tightest possible bounds on the chance.

(b) Let

$$\bar{X} = \frac{1}{25} \sum_{i=1}^{25} X_i$$

be the sample mean. Find $\mathrm{E}[\bar{X}]$ and $\mathrm{Var}[\bar{X}]$.

(c) Find the tightest possible bounds on the chance that the average height of the men in the sample is between 5 feet and 6 feet. That is, find the tightest possible bounds on $\mathrm{P}(60 < \bar{X} < 72)$.

(d) Find the tighest possible bounds on $\mathrm{P}(\bar{X}^2 \geq 5625)$.

4 Doubly Stochasticity

A *stochastic matrix* is a matrix in which each row sums to 1. A matrix is *doubly stochastic* if each column also sums to 1. The transition matrix of any Markov chain is stochastic by definition, since each row of the transition matrix is a distribution.

(a) Prove or disprove that the transition matrix of any Markov chain is doubly stochastic.

(b) Suppose an irreducible, aperiodic Markov chain on a finite state space has a doubly stochastic transition matrix. Prove that the stationary distribution is uniform on the state space.

(c) For $n \in \mathbb{N}$, let S_n be the total of n rolls of a six-sided die with faces numbered 1 through 6. Show that $X_n = S_n \bmod 7$ is a Markov chain.

(d) Find the probability that S_n is a multiple of 7 as $n \to \infty$.

Submission. Homeworks must be submitted through Gradescope. If you are completing your homeworks on paper, please scan the pages of your homework into a PDF using any scanner or phone application such as CamScanner. **It is your responsibility to ensure that all the work on the scanned pages is legible.**

Once you upload your submission to the Gradescope assignment, you will be prompted to select pages. **It is your responsibility to correctly select the pages of your homework corresponding to each question.** If you are having difficulties scanning, uploading, or submitting your homework, post a follow-up on the main thread corresponding to this homework on Ed.