CS 70 Discrete Mathematics and Probability Theory
Summer 2018 Sinho Chewi and Vrettos Moulos

HW 6

Due: Friday, 7/27, 10pm

Sundry

Before you start your homework, write down your team. Who else did you work with on this homework? List names and email addresses. (In case of homework party, you can also just describe the group.) How did you work on this homework? Working in groups of 3-5 will earn credit for your "Sundry" grade.

Please copy the following statement and sign next to it:

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

1 Dice Distributions

A fair die with k faces ($k \geq 2$), numbered $1, \ldots, k$, is rolled n times, with each roll being independent of all other rolls. Let X_i be a random variable for the number of times the ith face shows up. For each of the following parts, your answers should be in terms of n and k.

(a) What is the size of the sample space? (How many possible outcomes are there?)

(b) What is the distribution of X_i, for $1 \leq i \leq k$?

(c) What is the joint distribution of $X_1, X_2, \ldots, X_k$?

(d) Are X_1 and X_2 independent random variables?

2 Large Three-Way Cuts

In class, we have seen that for every graph $G = (V, E)$ with $|V|$ vertices and $|E|$ edges, the vertex set V can be partitioned in two sets A, B such that the number of edges between A and B is at least $\frac{|E|}{2}$. Here, we'll prove a similar result for the case where we're allowed to partition V into

three sets: A, B, and C. (Recall that the sets A, B, C form a partition of V iff $A \cup B \cup C = V$ and $A \cap B = B \cap C = C \cap A = \emptyset$.)

(a) Suppose we sample a random partition of V by choosing which set each vertex goes into uniformly at random from $\{A, B, C\}$ independently. What is our sample space? How big is it?

(b) For two distinct vertices $u, v \in V$ calculate the probability that u and v lie in different sets.

(c) Compute the expected number of edges that cross between the sets in our random partition. (An edge "crosses between the sets" if its two endpoints are in different sets.)

(d) Prove that there exists some partition (A, B, C) of V such that the number of edges that cross between the sets is at least $\frac{2|E|}{3}$.

3 Who Has More Sisters?

Out of all families in the world with $n > 0$ children, you sample one at random and observe X, the total number of sisters that the male children in this family have, and Y, the total number of sisters that the female children in this family have (for example, if $n = 3$ and there are two males and one female, then $X = 2$ and $Y = 0$).

Assuming that each child born into the world has an equal chance of being male or female, find expressions for $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ in terms of n (these expressions should not involve summations). Based on your expressions, do males have more sisters or females have more sisters, on average?

Hint: Define a random variable B to denote the number of boys, find an expression for X as a function of B, and apply linearity of expectation. Use a similar approach for girls.

4 Combining Distributions

(a) Let $X \sim Pois(\lambda), Y \sim Pois(\mu)$ be independent. Prove that $X + Y \sin Pois(\lambda + \mu)$.

Hint: Recall the binomial theorem, which states that

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}.$$

(b) Let X and Y be defined as in the previous part. Prove that the distribution of X conditional on $X + Y$ is a binomial distribution, e.g. that $X|X+Y$ is binomial. What are the parameters of the binomial distribution?

Hint: Your result from the previous part will be helpful.

5 Unbiased Variance Estimation

We have a random variable X and want to estimate its variance, σ^2 and mean, μ, by sampling from it. In this problem, we will derive an "unbiased estimator" for the variance.

(a) We define a random variable Y that corresponds to drawing n values from the distribution for X and averaging, or $Y = (X_1 + \ldots + X_n)/n$. What is $\mathbb{E}(Y)$? Note that if $\mathbb{E}(Y) = \mathbb{E}(X)$ then Y is an unbiased estimator of $\mu = \mathbb{E}(X)$.

*H*int: This should not be difficult.

(b) Now let's assume the actual mean is 0 as variance doesn't change when one shifts the mean.

Before attempting to define an estimator for variance, show that $\mathbb{E}(Y^2) = \sigma^2/n$.

(c) In practice, we don't know the mean of X so following part (a), we estimate it as Y. With this in mind, we consider the random variable $Z = \sum_{i=1}^{n}(X_i - Y)^2$. What is $\mathbb{E}(Z)$?

(d) What is a good unbiased estimator for the $\text{var}(X)$?

(e) How does this differ from what you might expect? Why? (Just tell us your intuition here, it is all good!)

6 Variance Proofs

(a) Let X be a random variable. Prove that:

$$\text{var}(X) \geq 0$$

(b) Let $X_1, \ldots, X_n$ be random variables. Prove that:

$$\text{var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \text{var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j)$$

Hint: Without loss of generality we can assume that $\mathbb{E}[X_1] = \cdots = \mathbb{E}[X_n] = 0$. Why?

(c) Let $a_1, \ldots, a_n \in \mathbb{R}$, and $X_1, \ldots, X_n$ be random variables. Prove that:

$$\sum_{i=1}^{n} a_i^2 \cdot \text{var}(X_i) + 2 \sum_{1 \leq i < j \leq n} a_i \cdot a_j \cdot \text{cov}(X_i, X_j) \geq 0$$

7 Number Game

Sinho and Vrettos are playing a game where they each choose an integer uniformly at random from $[0, 100]$, then whoever has the larger number wins (in the event of a tie, they replay). However, Vrettos doesn't like losing, so he's rigged his random number generator such that it instead picks randomly from the integers between Sinho's number and 100. Let S be Sinho's number and V be Vrettos' number.

(a) What is $\mathbb{E}[S]$?

(b) What is $\mathbb{E}[V|S=s]$, where s is any constant such that $0 \leq s \leq 100$?

(c) What is $\mathbb{E}[V]$?

8 Student Request Collector

After a long night of debugging, Alvin has just perfected the new homework party/office hour queue system. CS 70 students sign themselves up for the queue, and TAs go through the queue, resolving requests one by one. Unfortunately, our newest TA (let's call him TA Bob) does not understand how to use the new queue: instead of resolving the requests in order, he always uses the Random Student button, which (as the name suggests) chooses a random student in the queue for him. To make matters worse, after helping the student, Bob forgets to click the Resolve button, so the student still remains in the queue! For this problem, assume that there are n total students in the queue.

(a) Suppose that Bob has already helped k students. What is the probability that the Random Student button will take him to a student who has not already been helped?

(b) Let X_i^r be the event that TA Bob has not helped student i after pressing the Random Student button a total of r times. What is $\mathbb{P}[X_i^r]$? Assume that the results of the Random Student button are independent of each other. Now use the inequality $1-x \leq e^{-x}$ to upper bound this probability.

(c) Let T_r represent the event that TA Bob presses the Random Student button r times, but still has not been able to help all n students. (In other words, it takes TA Bob longer than r Random Student button presses before he manages to help every student). What is T_r in terms of the events X_i^r?

Hint: Events are subsets of the probability space Ω, so you should be thinking of set operations.

(d) Using your answer for the previous part, what is an upper bound for $\mathbb{P}[T_r]$?

(e) Now let $r = \alpha n \ln n$. What is an upper bound for $\mathbb{P}[X_i^r]$?

(f) Calculate an upper bound for $\mathbb{P}[T_r]$ using the same value of r as before. (This is more formally known as a bound on the tail probability of the distribution of button presses required to help every student.)

(g) What value of r do you need to bound the tail probability by $1/n^2$? In other words, how many button presses are needed so that the probability that TA Bob has not helped every student is at most $1/n^2$?