

1 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.

2 Cliques in Random Graphs

Consider a graph $G(V, E)$ on n vertices which is generated by the following random process: for each pair of vertices u and v , we flip a fair coin and place an (undirected) edge between u and v if and only if the coin comes up heads. So for example if $n = 2$, then with probability $1/2$, $G(V, E)$ is the graph consisting of two vertices connected by an edge, and with probability $1/2$ it is the graph consisting of two isolated vertices.

- (a) What is the size of the sample space?
- (b) A k -clique in graph is a set of k vertices which are pairwise adjacent (every pair of vertices is connected by an edge). For example a 3-clique is a triangle. What is the probability that a particular set of k vertices forms a k -clique?
- (c) Prove that the probability that the graph contains a k -clique for $k = 4\lceil \log n \rceil + 1$ is at most $1/n$.

3 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 $\Pr[X_i^r]$? Assume that the results of the Random Student button are independent of each other. Now approximate the answer using the inequality $1 - x \leq e^{-x}$.
- (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 $\Pr[T_r]$? (You may leave your answer in terms of $\Pr[X_i^r]$. Use the inequality $1 - x \leq e^{-x}$ from before.)
- (e) Now let $r = \alpha n \ln n$. What is $\Pr[X_i^r]$?
- (f) Calculate an upper bound for $\Pr[T_r]$ 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. This distribution will be explored in more detail later, in the context of random variables.)
- (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 $1/n^2$?

4 Combinatorial Coins

Allen and Alvin are flipping coins for fun. Allen flips a fair coin k times and Alvin flips $n - k$ times. In total there are n coin flips.

- (a) Use a combinatorial proof to show that

$$\sum_{i=0}^k \binom{k}{k-i} \binom{n-k}{i} = \binom{n}{k}.$$

- (b) Prove that the probability that Allen and Alvin flip the same number of heads is equal to the probability that there are a total of k heads.

5 Geometric Distribution

Two faulty machines, M_1 and M_2 , are repeatedly run synchronously in parallel (i.e., both machines execute one run, then both execute a second run, and so on). On each run, M_1 fails with probability p_1 and M_2 fails with probability p_2 , all failure events being independent. Let the random variables X_1, X_2 denote the number of runs until the first failure of M_1, M_2 respectively; thus X_1, X_2 have geometric distributions with parameters p_1, p_2 respectively. Let X denote the number of runs until the first failure of *either* machine. Show that X also has a geometric distribution, with parameter $p_1 + p_2 - p_1p_2$.

6 Poisson Distribution

- (a) It is fairly reasonable to model the number of customers entering a shop during a particular hour as a Poisson random variable. Assume that this Poisson random variable X has mean λ . Suppose that whenever a customer enters the shop they leave the shop without buying anything with probability p . Assume that customers act independently, i.e. you can assume that they each simply flip a biased coin to decide whether to buy anything at all. Let us denote the number of customers that buy something as Y and the number of them that do not buy anything as Z (so $X = Y + Z$). What is the probability that $Y = k$ for a given k ? How about $\Pr[Z = k]$? Prove that Y and Z are Poisson random variables themselves.

Hint: You can use the identity

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

- (b) Prove that Y and Z are independent.
- (c) Assume that you were given two independent Poisson random variables X_1, X_2 . Assume that the first has mean λ_1 and the second has mean λ_2 . Prove that $X_1 + X_2$ is a Poisson random variable with mean $\lambda_1 + \lambda_2$.

Hint: Recall the binomial theorem.

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$$

7 Poisson Coupling

Consider the following discrete joint distribution for $p \in [0, 1]$.

$$\begin{aligned} \Pr(X = 0, Y = 0) &= 1 - p, \\ \Pr(X = 1, Y = y) &= \frac{e^{-p} p^y}{y!}, \quad y = 1, 2, \dots, \\ \Pr(X = 1, Y = 0) &= e^{-p} - (1 - p), \\ \Pr(X = x, Y = y) &= 0, \quad \text{otherwise.} \end{aligned}$$

- (a) Recall that all valid distributions satisfy two important properties. Argue that this distribution is a valid joint distribution.
- (b) Show that X has the Bernoulli distribution with probability p .
- (c) Show that Y has the Poisson distribution with parameter $\lambda = p$.
- (d) Show that $\Pr(X \neq Y) \leq p^2$.

Now, let $X_i, i = 1, 2, \dots$ be a sequence of random variables with probabilities $p_i, i = 1, 2, \dots$. Similarly, let Y_i be a Poisson random variable with parameter $\lambda = p_i, i = 1, 2, \dots$. The X_i and Y_i are coupled, so that they have the joint distribution described above (with $p = p_i$), but for $i \neq j$, (X_i, Y_i) and (X_j, Y_j) are independent.

We will now introduce a coupling argument which shows that the distribution of $\sum_{i=1}^n X_i$ approaches a Poisson distribution with parameter $\lambda = p_1 + \dots + p_n$.

- (e) A common way to measure the “distance” between two probability distributions is known as the total variation norm, and it is given by

$$d(X, Y) = \frac{1}{2} \sum_{k=0}^{\infty} |\Pr(X = k) - \Pr(Y = k)|.$$

Show that $d(X, Y) \leq \Pr(X \neq Y)$. [*Hint*: Use the Law of Total Probability to split up the events according to $\{X = Y\}$ and $\{X \neq Y\}$.]

- (f) Show that $\Pr(\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i) \leq \sum_{i=1}^n \Pr(X_i \neq Y_i)$. [*Hint*: Maybe try the Union Bound.]
- (g) Finally, for the X_i and Y_i defined above, show that $d(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i) \leq \sum_{i=1}^n p_i^2$.