CS 70 Discrete Mathematics and Probability Theory
Spring 2021

HW 10

Due: Friday, 04/09 at 10:00 PM
Grace period until Sunday, 04/11 at 11:59 PM

Sundry

Before you start writing your final homework submission, state briefly how you worked on it. Who else did you work with? List names and email addresses. (In case of homework party, you can just describe the group.)

1 Random Variables Warm-Up

Let X and Y be random variables, each taking values in the set $\{0,1,2\}$, with joint distribution

$$\begin{array}{lll} \mathbb{P}[X=0,Y=0]=1/3 & \mathbb{P}[X=0,Y=1]=0 & \mathbb{P}[X=0,Y=2]=1/3 \\ \mathbb{P}[X=1,Y=0]=0 & \mathbb{P}[X=1,Y=1]=1/9 & \mathbb{P}[X=1,Y=2]=0 \\ \mathbb{P}[X=2,Y=0]=1/9 & \mathbb{P}[X=2,Y=1]=1/9 & \mathbb{P}[X=2,Y=2]=0. \end{array}$$

(a) What are the marginal distributions of X and Y?

(b) What are $\mathbb{E}[X]$ and $\mathbb{E}[Y]$?

(c) (optional) What are $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$?

(d) Let I be the indicator that $X=1$, and J be the indicator that $Y=1$. What are $\mathbb{E}[I]$, $\mathbb{E}[J]$ and $\mathbb{E}[IJ]$?

(e) In general, let I_A and I_B be the indicators for events A and B in a probability space $(\Omega,\mathbb{P})$. What is $\mathbb{E}[I_A I_B]$, in terms of the probability of some event?

2 Optimal Gambling

Jonathan has a coin that may be biased, but he doesn't think so. You disagree with him though, and he challenges you to a bet. You start off with X_0 dollars. You and Jonathan then play multiple rounds, and each round, you bet an amount of money of your choosing, and then coin is tossed. Jonathan will match your bet, no matter what, and if the coin comes up heads, you win and you take both yours and Jonathan's bet, and if it comes up tails, then you lose your bet.

(a) Now suppose you actually secretly know that the bias of the coin is $\frac{1}{2} < p < 1$! You use the following strategy: on each round, you will bet a fraction q of the money you have at the start of the round. Let's say you play n rounds. What is the probability that you win exactly k of the rounds? What is the amount of money you would have if you win exactly k rounds? [*Hint*: Does the order in which you win the games affect your profit?]

(b) Let X_n denote the amount of money you have on round n. X_0 represents your initial assets and is a constant value. Show that $\mathbb{E}[X_n] = ((1-p)(1-q)+p(1+q))^n X_0$.

You may use the binomial theorem in your answer:

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

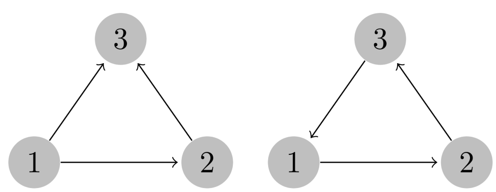
[*Hint*: Try computing a sum over the number of rounds you win out of the n rounds you play - use your answers from the previous part.]

(c) What value of q will maximize $\mathbb{E}[X_n]$? For this value of q, what is the distribution of X_n? Can you predict what will happen as $n \to \infty$? [*Hint*: Under this betting strategy, what happens if you ever lose a round?]

(d) The problem with the previous approach is that we were too concerned about expected value, so our gambling strategy was too extreme. Let's start over: again we will use a gambling strategy in which we bet a fraction q of our money at each round. Express X_n in terms of n, q, X_0, and W_n, where W_n is the number of rounds you have won up until round n. [*Hint*: Does the order in which you win the games affect your profit?]

3 Random Tournaments

A *tournament* is a directed graph in which every pair of vertices has exactly one directed edge between them—for example, here are two tournaments on the vertices $\{1,2,3\}$:



In the first tournament above, $(1,2,3)$ is a *Hamiltonian path*, since it visits all the vertices exactly once, without repeating any edges, but $(1,2,3,1)$ is not a valid *Hamiltonian cycle*, because the tournament contains the directed edge $1 \to 3$ and not $3 \to 1$. In the second tournament, $(1,2,3,1)$ is a *Hamiltonian cycle*, as are $(2,3,1,2)$ and $(3,1,2,3)$; for this problem we'll say that these are all different Hamiltonian cycles, since their start/end points are different.

Consider the following way of choosing a random tournament T on n vertices: independently for each (unordered) pair of vertices $\{i,j\} \subset \{1,...,n\}$, flip a coin and include the edge $i \to j$ in the graph if the outcome is heads, and the edge $j \to i$ if tails. What is the expected number of Hamiltonian paths in T? What is the expected number of Hamiltonian cycles?

4 Class Enrollment

Lydia has just started her CalCentral enrollment appointment. She needs to register for a marine science class and CS 70. There are no waitlists, and she can attempt to enroll once per day in either class or both. The CalCentral enrollment system is strange and picky, so the probability of enrolling successfully in the marine science class on each attempt is μ and the probability of enrolling successfully in CS 70 on each attempt is λ. Also, these events are independent.

(a) Suppose Lydia begins by attempting to enroll in the marine science class everyday and gets enrolled in it on day M. What is the distribution of M?

(b) Suppose she is not enrolled in the marine science class after attempting each day for the first 5 days. What is the conditional distribution of M given $M > 5$?

(c) Once she is enrolled in the marine science class, she starts attempting to enroll in CS 70 from day $M+1$ and gets enrolled in it on day C. Find the expected number of days it takes Lydia to enroll in both the classes, i.e. $\mathbb{E}[C]$.

(d) Suppose instead of attempting one by one, Lydia decides to attempt enrolling in both the classes from day 1. Let M be the number of days it takes to enroll in the marine science class, and C be the number of days it takes to enroll in CS 70. What is the distribution of M and C now? Are they independent?

(e) Let X denote the day she gets enrolled in her first class and let Y denote the day she gets enrolled in both the classes. What is the distribution of X?

(f) What is the expected number of days it takes Lydia to enroll in both classes now, i.e. $\mathbb{E}[Y]$.

(g) What is the expected number of classes she will be enrolled in by the end of 14 days?

5 Poisson Coupling

(a) Let X, Y be discrete random variables taking values in $\mathbb{N}$. A common way to measure the "distance" between two probability distributions is known as the total variation norm, and it is given by

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

Show that

$$d(X,Y) \le \mathbb{P}(X \neq Y). \tag{1}$$

[*Hint*: Use the Law of Total Probability to split up the events according to $\{X=Y\}$ and $\{X \neq Y\}$.]

(b) Show that if X_i, Y_i, $i \in \mathbb{Z}_+$ are discrete random variables taking values in $\mathbb{N}$, then $\mathbb{P}(\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i) \le \sum_{i=1}^n \mathbb{P}(X_i \neq Y_i)$. [*Hint*: Maybe try the Union Bound.]

Notice that the LHS of (1) only depends on the *marginal* distributions of X and Y, whereas the RHS depends on the *joint* distribution of X and Y. This leads us to the idea that we can find a good bound for $d(X,Y)$ by choosing a special joint distribution for (X,Y) which makes $\mathbb{P}(X \neq Y)$ small.

We will now introduce a coupling argument which shows that the distribution of the sum of independent Bernoulli random variables with parameters p_i, $i = 1,\dots,n$, is close to a Poisson distribution with parameter $\lambda = p_1 + \cdots + p_n$.

(c) Let (X_i,Y_i) and (X_i,Y_j) be independent for $i \neq j$, but for each i, X_i and Y_i are *coupled*, meaning that they have the following discrete distribution:

$$\begin{aligned}
&\mathbb{P}(X_i = 0, Y_i = 0) = 1 - p_i, \\
&\mathbb{P}(X_i = 1, Y_i = y) = \frac{e^{-p_i} p_i^y}{y!}, && y = 1,2,\dots, \\
&\mathbb{P}(X_i = 1, Y_i = 0) = e^{-p_i} - (1 - p_i), \\
&\mathbb{P}(X_i = x, Y_i = y) = 0, && \text{otherwise.}
\end{aligned}$$

Recall that all valid distributions satisfy two important properties. Argue that this distribution is a valid joint distribution.

(d) Show that X_i has the Bernoulli distribution with probability p_i.

(e) Show that Y_i has the Poisson distribution with parameter $\lambda = p_i$.

(f) Show that $\mathbb{P}(X_i \neq Y_i) \leq p_i^2$.

(g) Finally, show that $d(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i) \leq \sum_{i=1}^n p_i^2$.

6 Combining Distributions

Let $X \sim Pois(\lambda), Y \sim Pois(\mu)$ be independent. 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: Recall that we can prove $X|X+Y$ is binomial if it's PMF is of the same form