CS 70 Fall 2022	Discrete Mathematics and Probability Theory Satish Rao and Babak Ayazifar	HW 11

Due: Saturday, 11/12, 4:00 PM
Grace period until Saturday, 11/12, 6:00 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 Balls and Bins

Throw n balls into m bins, where m and n are positive integers. Let X be the number of bins with exactly one ball. Compute $\text{Var}(X)$. Your final answer should not contain any summations.

Solution: Let I_i be the indicator variable for the event that bin i has 1 ball. Now, we use:

$$E\left(\sum I_i\right) = \sum P(I_i)$$

to calculate the expected value. Computing the probability of any indicator variable I_i:

$$P(I_i) = \frac{1}{m}\left(\frac{m-1}{m}\right)^{n-1} n$$

We multiply by n at the end because there are n balls that we could choose from to go into bin i. Now, there are m bins, so therefore the total expected value is simply $mP(I_i)$, so therefore:

$$\sum P(I_i) = n\left(\frac{m-1}{m}\right)^{n-1}$$

2 Will I Get My Package?

A delivery guy in some company is out delivering n packages to n customers, where n is a natural number greater than 1. Not only does he hand each customer a package uniformly at random from the remaining packages, he opens the package before delivering it with probability $1/2$. Let X be the number of customers who receive their own packages unopened.

(a) Compute the expectation $\mathbb{E}[X]$.

(b) Compute the variance $\mathrm{Var}(X)$.

3 Double-Check Your Intuition Again

(a) You roll a fair six-sided die and record the result X. You roll the die again and record the result Y.

(i) What is $\text{cov}(X+Y, X-Y)$?

(ii) Prove that $X+Y$ and $X-Y$ are not independent.

For each of the problems below, if you think the answer is "yes" then provide a proof. If you think the answer is "no", then provide a counterexample.

(b) If X is a random variable and $\text{Var}(X) = 0$, then must X be a constant?

(c) If X is a random variable and c is a constant, then is $\text{Var}(cX) = c\,\text{Var}(X)$?

(d) If A and B are random variables with nonzero standard deviations and $\text{Corr}(A,B) = 0$, then are A and B independent?

(e) If X and Y are not necessarily independent random variables, but $\text{Corr}(X,Y) = 0$, and X and Y have nonzero standard deviations, then is $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$?

(f) If X and Y are random variables then is $\mathbb{E}[\max(X,Y)\min(X,Y)] = \mathbb{E}[XY]$?

(g) If X and Y are independent random variables with nonzero standard deviations, then is

$$\text{Corr}(\max(X,Y), \min(X,Y)) = \text{Corr}(X,Y)?$$

4 Fishy Computations

Assume for each part that the random variable can be modelled by a Poisson distribution.

(a) Suppose that on average, a fisherman catches 20 salmon per week. What is the probability that he will catch exactly 7 salmon this week?

(b) Suppose that on average, you go to Fisherman's Wharf twice a year. What is the probability that you will go at most once in 2018?

(c) Suppose that in March, on average, there are 5.7 boats that sail in Laguna Beach per day. What is the probability there will be *at least* 3 boats sailing throughout the *next two days* in Laguna?

5 Geometric and Poisson

Let $X \sim \text{Geo}(p)$ and $Y \sim \text{Poisson}(\lambda)$ be independent. random variables. Compute $\mathbb{P}[X > Y]$. Your final answer should not have summations.

Hint: Use the total probability rule.

6 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 $\{X \neq Y\}$. Also, the inequality $|a-b| \le a+b$ might be helpful.]

(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 $\mathbb{P}[X_i \neq Y_i] \le p_i^2$.

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