CS 70 Spring 2020 — Discrete Mathematics and Probability Theory Course Notes — HW 11

Due: Friday, April 17, 2020 at 11:59 PM
Grace period until Sunday, April 19, 2020 at 11:59 PM

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

2 Binomial Variance

Throw n balls into m bins uniformly at random. For a specific ball i, what is the variance of the number of roommates it has (i.e. the number of other balls that it shares its bin with)?

3 Portfolio Optimization

Suppose that there are n assets, where n is a positive integer. For each unit dollar invested in asset i, for $i = 1, \ldots, n$, with probability p_i the value of the asset will grow by α_i to $1 + \alpha_i$, and with probability $1 - p_i$ the value of the asset will shrink by α_i to $1 - \alpha_i$. Let the proportion of money invested in asset i be w_i (so that $\sum_{i=1}^n w_i = 1$), and let X_i be a random variable denoting the final value of the ith asset per unit dollar. Then $X = w_1 X_1 + \cdots + w_n X_n$ is the total value. For simplicity, assume that the outcomes of the different assets are independent.

(a) Compute the expectation $\mathbb{E}[X]$. What values of w_i maximize this quantity?

(b) Compute the variance $\operatorname{Var} X$. What values of w_i minimize this quantity?

4 Will I Get My Package?

A delivery guy in some company is out delivering n packages to n customers, where $n \in \mathbb{N}$, $n > 1$. Not only does he hand a random package to each customer, 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 $\operatorname{Var}(X)$.

5 Family Planning

Mr. and Mrs. Brown decide to continue having children until they either have their first girl or until they have three children. Assume that each child is equally likely to be a boy or a girl, independent of all other children, and that there are no multiple births. Let G denote the numbers of girls that the Browns have. Let C be the total number of children they have.

(a) Determine the sample space, along with the probability of each sample point.

(b) Compute the joint distribution of G and C. Fill in the table below.

	$C=1$	$C=2$	$C=3$
$G=0$			
$G=1$			

(c) Use the joint distribution to compute the marginal distributions of G and C and confirm that the values are as you'd expect. Fill in the tables below.

$\mathbb{P}(G=0)$	
$\mathbb{P}(G=1)$	

$\mathbb{P}(C=1)$	$\mathbb{P}(C=2)$	$\mathbb{P}(C=3)$

(d) Are G and C independent?

(e) What is the expected number of girls the Browns will have? What is the expected number of children that the Browns will have?

6 More Family Planning

(a) Suppose we have a random variable $N \sim Geom(1/3)$ representing the number of children of a randomly chosen family. Assume that within the family, children are equally likely to be boys and girls. Let B be the number of boys and G the number of girls in the family. What is the joint probability distribution of B, G?

(b) Given that we know there are 0 girls in the family, what is the most likely number of boys in the family?

(c) Now let X and Y be independent random variables representing the number of children in two independently, randomly chosen families. Suppose that $X \sim Geom(p)$ and $Y \sim Geom(q)$. Find $\mathbb{P}(X < Y)$, the probability that the number of children in the first family (X) is less than the number of children in the second family (Y). (You may use the convergence formula for a Geometric Series: $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$ for $|r| < 1$)

(d) Show how you could obtain your answer from the previous part using an interpretation of the geometric distribution.

7 Marginals

(a) Can there exist three random variables X_1, X_2, X_3, each taking values in the set $\{+1, -1\}$, with the property that for every $i \neq j$, the joint distribution of X_i and X_j is given by

$$\mathbb{P}[X_i = 1, X_j = -1] = \frac{1}{2} \qquad \mathbb{P}[X_i = -1, X_j = 1] = \frac{1}{2} \qquad \mathbb{P}[X_i = X_j] = 0? \tag{2}$$

If so, specify the joint distribution of X_1, X_2, X_3; if not, prove it.

(b) For which natural numbers $n \geq 3$ can there exist random variables $X_1, X_2, \ldots, X_n$, each taking values in the set $\{+1, -1\}$, with the property that for every i and j satisfying $i - j = 1 \pmod{n}$, the joint distribution of X_i and X_j is given by (1)? For any n that work, specify the joint distribution; for those that do not, prove it.

8 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

9 Make Your Own Question

Make your own question on this week's material and solve it.

10 Homework Process and Study Group

Citing sources and collaborators are an important part of life, including being a student! We also want to understand what resources you find helpful and how much time homework is taking, so we can change things in the future if possible.

1. **What sources (if any) did you use as you worked through the homework?**
2. **If you worked with someone on this homework, who did you work with?** List names and student ID's. (In case of homework party, you can also just describe the group.)
3. **How did you work on this homework?** (For example, *I first worked by myself for 2 hours, but got stuck on problem 3, so I went to office hours. Then I went to homework party for a few hours, where I finished the homework.*)
4. **Roughly how many total hours did you work on this homework?**