CS 70 Spring 2021 — Discrete Mathematics and Probability Theory — HW 11

Due: Friday, 04/16 at 10:00 PM
Grace period until Friday, 04/16 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 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

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

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

(b) Show that if X_i, Y_i, $i \in \mathbb{Z}_+$ are discrete random variables taking values in $\mathbb{N}$, then $\mathbb{P}(\sum_{i=1}^{n} X_i \ne \sum_{i=1}^{n} Y_i) \le \sum_{i=1}^{n} \mathbb{P}(X_i \ne Y_i)$. [*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 $\mathbb{P}(X \ne Y)$ 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:

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

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 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

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 $\mathrm{Corr}(X,Y) = 0$, and X and Y have nonzero standard deviations, then is $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$?

(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

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

4 Just One Tail, Please

Let X be some random variable with finite mean and variance which is not necessarily non-negative. The *extended* version of Markov's Inequality states that for a non-negative function $\phi(x)$ which is monotonically increasing for $x > 0$ and some constant $\alpha > 0$,

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[\phi(X)]}{\phi(\alpha)}$$

Suppose $\mathbb{E}[X] = 0$, $\mathrm{Var}(X) = \sigma^2 < \infty$, and $\alpha > 0$.

(a) Use the extended version of Markov's Inequality stated above with $\phi(x) = (x+c)^2$, where c is some positive constant, to show that:

$$\mathbb{P}(X \geq \alpha) \leq \frac{\sigma^2 + c^2}{(\alpha + c)^2}$$

(b) Note that the above bound applies for all positive c, so we can choose a value of c to minimize the expression, yielding the best possible bound. Find the value for c which will minimize the RHS expression (you may assume that the expression has a unique minimum). Plug in the minimizing value of c to prove the following bound:

$$\mathbb{P}(X \geq \alpha) \leq \frac{\sigma^2}{\alpha^2 + \sigma^2}.$$

(c) Recall that Chebyshev's inequality provides a two-sided bound. That is, it provides a bound on $\mathbb{P}(|X - \mathbb{E}[X]| \geq \alpha) = \mathbb{P}(X \geq \mathbb{E}[X] + \alpha) + \mathbb{P}(X \leq \mathbb{E}[X] - \alpha)$. If we only wanted to bound the probability of one of the tails, e.g. if we wanted to bound $\mathbb{P}(X \geq \mathbb{E}[X] + \alpha)$, it is tempting to just divide the bound we get from Chebyshev's by two. Why is this not always correct in general? Provide an example of a random variable X (does not have to be zero-mean) and a constant α such that using this method (dividing by two to bound one tail) is not correct, that is, $\mathbb{P}(X \geq \mathbb{E}[X] + \alpha) > \frac{\mathrm{Var}(X)}{2\alpha^2}$ or $\mathbb{P}(X \leq \mathbb{E}[X] - \alpha) > \frac{\mathrm{Var}(X)}{2\alpha^2}$.

Now we see the use of the bound proven in part (b) - it allows us to bound just one tail while still taking variance into account, and does not require us to assume any property of the random variable. Note that the bound is also always guaranteed to be less than 1 (and therefore at least somewhat useful), unlike Markov's and Chebyshev's inequality!

(d) Let's try out our new bound on a simple example. Suppose X is a positively-valued random variable with $\mathbb{E}[X] = 3$ and $\text{Var}(X) = 2$. What bound would Markov's inequality give for $\mathbb{P}[X \geq 5]$? What bound would Chebyshev's inequality give for $\mathbb{P}[X \geq 5]$? What about for the bound we proved in part (b)? (*Note*: Recall that the bound from part (b) only applies for zero-mean random variables.)

5 Law of Large Numbers

Recall that the *Law of Large Numbers* holds if, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| > \varepsilon\right) = 0.$$

In class, we saw that the Law of Large Numbers holds for $S_n = X_1 + \cdots + X_n$, where the X_i's are i.i.d. random variables. This problem explores if the Law of Large Numbers holds under other circumstances.

Packets are sent from a source to a destination node over the Internet. Each packet is sent on a certain route, and the routes are disjoint. Each route has a failure probability of $p \in (0,1)$ and different routes fail independently. If a route fails, all packets sent along that route are lost. You can assume that the routing protocol has no knowledge of which route fails.

For each of the following routing protocols, determine whether the Law of Large Numbers holds when S_n is defined as the total number of received packets out of n packets sent. Answer **Yes** if the Law of Large Number holds, or **No** if not, and give a brief justification of your answer. (Whenever convenient, you can assume that n is even.)

(a) **Yes** or **No**: Each packet is sent on a completely different route.

(b) **Yes** or **No**: The packets are split into $n/2$ pairs of packets. Each pair is sent together on its own route (i.e., different pairs are sent on different routes).

(c) **Yes** or **No**: The packets are split into 2 groups of $n/2$ packets. All the packets in each group are sent on the same route, and the two groups are sent on different routes.

(d) **Yes** or **No**: All the packets are sent on one route.

6 Practical Confidence Intervals

(a) It's New Year's Eve, and you're re-evaluating your finances for the next year. Based on previous spending patterns, you know that you spend \$1500 per month on average, with a standard deviation of \$500, and each month's expenditure is independently and identically distributed. As a college student, you also don't have any income. How much should you have in your bank account if you don't want to run out of money this year, with probability at least 95%?

(b) As a UC Berkeley CS student, you're always thinking about ways to become the next billionaire in Silicon Valley. After hours of brainstorming, you've finally cut your list of ideas down to 10, all of which you want to implement at the same time. A venture capitalist has agreed to back all 10 ideas, as long as your net return from implementing the ideas is positive with at least 95% probability.

Suppose that implementing an idea requires 50 thousand dollars, and your start-up then succeeds with probability p, generating 150 thousand dollars in revenue (for a net gain of 100 thousand dollars), or fails with probability $1-p$ (for a net loss of 50 thousand dollars). The success of each idea is independent of every other. What is the condition on p that you need to satisfy to secure the venture capitalist's funding?

(c) One of your start-ups uses error-correcting codes, which can recover the original message as long as at least 1000 packets are received (not erased). Each packet gets erased independently with probability 0.8. How many packets should you send such that you can recover the message with probability at least 99%?

7 Balls in Bins Estimation

We throw $n > 0$ balls into $m \geq 2$ bins. Let X and Y represent the number of balls that land in bin 1 and 2 respectively.

(a) Calculate $\mathbb{E}[Y \mid X]$. [*Hint*: Your intuition may be more useful than formal calculations.]

(b) What are $L[Y \mid X]$ and $Q[Y \mid X]$ (where $Q[Y \mid X]$ is the best quadratic estimator of Y given X)? [*Hint*: Your justification should be no more than two or three sentences, no calculations necessary! Think carefully about the meaning of the MMSE.]

(c) Unfortunately, your friend is not convinced by your answer to the previous part. Compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

(d) Compute $\text{Var}(X)$.

(e) Compute $\text{cov}(X,Y)$.

(f) Compute $L[Y \mid X]$ using the formula. Ensure that your answer is the same as your answer to part (b).