CS 70 Discrete Mathematics and Probability Theory
Fall 2021

HW 12

Due: Saturday 11/20, 4:00 PM
Grace period until Saturday 11/20, 5:59 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 Dice Games

Suppose you roll a fair six-sided die. You read off the number showing on the die, then flip that many fair coins.

(a) If the result of your die roll is i, what is the expected number of heads you see?

(b) What is the expected number of heads you see?

2 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

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

[*Hint*: Use the Law of Total Probability to split up the events according to $\{X = Y\}$ and $\{X \neq Y\}$. Also, the inequality $|a-b| \leq a+b$ 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 $\mathbb{P}(\sum_{i=1}^{n} X_i \neq \sum_{i=1}^{n} Y_i) \leq \sum_{i=1}^{n} \mathbb{P}(X_i \neq Y_i)$. [*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, $i = 1,\ldots,n$, 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 $d(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i) \leq \sum_{i=1}^n p_i^2$.

3 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

4 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

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

5 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 $\mathbb{E}[X] = 0$, $\text{Var}(X) = \sigma^2 < \infty$, 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).

We can plug in the minimizing value of c you found in part (b) to prove the following bound:

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

This bound is also known as Cantelli's inequality.

(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.

(i) Why is this not always correct in general?

(ii) 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 $\mathrm{Var}(X) = 2$.

(i) What bound would Markov's inequality give for $\mathbb{P}[X \geq 5]$?

(ii) What bound would Chebyshev's inequality give for $\mathbb{P}[X \geq 5]$?

(iii) What bound would Cantelli's Inequality give for $\mathbb{P}[X \geq 5]$? (*Note*: Recall that Cantelli's Inequality only applies for zero-mean random variables.)

6 Tightness of Inequalities

(a) Show by example that Markov's inequality is tight; that is, show that given some fixed $k > 0$, there exists a discrete non-negative random variable X such that $\mathbb{P}(X \geq k) = \mathbb{E}[X]/k$.

(b) Show by example that Chebyshev's inequality is tight; that is, show that given some fixed $k \geq 1$, there exists a random variable X such that $\mathbb{P}(|X-\mathbb{E}[X]| \geq k\sigma) = 1/k^2$, where $\sigma^2 = \mathrm{Var}(X)$.

7 Exploring Concepts

The following snippet illustrates a concept about polynomials.

Polynomial?
Interpolation gives one!
Few roots says; No more!

Yet another concept.

Lo! Delta sub i.
Behold! One at x sub i.

Sub j? Such empty.

And finally.

Polynomial in x.
A zero at r.
Cleft by x minus r.
And nothing remains.

1. You should write a poem or snippet to illustrate a concept from this weeks content. We are certain you can do better, but no pressure. It's all good.

2. In terms of the staff using your content for fun and "profit". Do you wish to (1) allow us to share to the class without atttribution (2) allow us to share to the class with attribution or (3) please, please do not share!

This problem will be sampled but only adds to the numerator not the "out of" (or denominator). Also, it will be generously graded and worth 10 points. If we feel this is useful to students (based on the connection to concepts in the course), we may continue it in future problem sets.