CS 70 Fall 2021 — Discrete Mathematics and Probability Theory — HW 12

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?

Solution:

(a) The number of heads you get is binomially distributed with parameters i and $\frac{1}{2}$. Thus, the expected number of heads you see is $\frac{i}{2}$.

(b) Let D be the outcome of the die roll and H be the number of heads you get. We have that

$$\begin{aligned}\mathbb{E}[H] &= \sum_{i=1}^{6} \mathbb{E}[H|D=i]\cdot\mathbb{P}[D=i] \\ &= \sum_{i=1}^{6} \frac{i}{2}\cdot\frac{1}{6} \\ &= \frac{1}{12}\sum_{i=1}^{6} i\end{aligned}$$

We know that $\sum_{i=1}^{n} i$ comes out to $\frac{n(n+1)}{2}$, so $\mathbb{E}[H] = \frac{1}{12}\cdot\frac{6\cdot 7}{2} = \frac{7}{4}$.

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

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

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

Solution:

(a) One has

$$\begin{aligned}
d(X,Y) &= \frac{1}{2} \sum_{k=0}^{\infty} |\mathbb{P}(X = k) - \mathbb{P}(Y = k)| \\
&= \frac{1}{2} \sum_{k=0}^{\infty} |\mathbb{P}(X = k, X = Y) + \mathbb{P}(X = k, X \neq Y) - \mathbb{P}(Y = k, X = Y) \\
&\qquad - \mathbb{P}(Y = k, X \neq Y)|.
\end{aligned}$$

Note that the event $\{X = k, X = Y\}$ is the same as $\{Y = k, X = Y\}$ (they both equal the event $\{X = Y = k\}$). Hence, these terms cancel and we have

$$\begin{aligned}
d(X,Y) &= \frac{1}{2} \sum_{k=0}^{\infty} |\mathbb{P}(X = k, X \neq Y) - \mathbb{P}(Y = k, X \neq Y)| \\
&\leq \frac{1}{2} \left(\sum_{k=0}^{\infty} \mathbb{P}(X = k, X \neq Y) + \sum_{k=0}^{\infty} \mathbb{P}(Y = k, X \neq Y) \right) = \frac{1}{2} (\mathbb{P}(X \neq Y) + \mathbb{P}(X \neq Y)).
\end{aligned}$$

We see that the factor of $1/2$ disappears and we are left with

$$d(X,Y) \le \mathbb{P}(X \neq Y). \tag{2}$$

(b) Note that the event $\{\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i\} \subseteq \{\exists i\ X_i \neq Y_i\}$, since if the two summations $\sum_{i=1}^n X_i$ and $\sum_{i=1}^n Y_i$ are different, then there must be at least one term which is different between the summations. Now, we can write

$$\mathbb{P}\left(\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i\right) \le \mathbb{P}(X_i \neq Y_i \text{ for some } i) = \mathbb{P}\left(\bigcup_{i=1}^n \{X_i \neq Y_i\}\right).$$

Now, we apply the Union Bound to the term on the right to obtain

$$\mathbb{P}\left(\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i\right) \le \sum_{i=1}^n \mathbb{P}(X_i \neq Y_i). \tag{3}$$

(c) We need to verify that the probabilities sum to 1. Indeed,

$$\begin{aligned}\mathbb{P}(X_i = 0, Y_i = 0) + \mathbb{P}(X_i = 1, Y_i = 0) + \sum_{y=1}^{\infty} \mathbb{P}(X_i = 1, Y_i = y) &= e^{-p_i} + \sum_{y=1}^{\infty} \frac{e^{-p_i} p_i^y}{y!}. \\ &= \cancel{e^{-p_i}} + 1 - \cancel{e^{-p_i}} = 1,\end{aligned}$$

Notice that

$$\begin{aligned}\sum_{y=1}^{\infty} \frac{e^{-p_i} p_i^y}{y!} &= e^{-p_i} \sum_{y=1}^{\infty} \frac{p_i^y}{y!} \\ &= e^{-p_i} \left(\left(\sum_{y=0}^{\infty} \frac{p_i^y}{y!} \right) - \frac{p_i^0}{0!} \right) \\ &= e^{-p_i} \left(e^{p_i} - 1 \right) \\ &= 1 - e^{-p_i},\end{aligned}$$

where the third inequality is true because of the Taylor series expansion $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$. An alternative derivation of the above equality is by considering a Poisson(p_i) random variable Z. Then the left hand side is $\sum_{y=1}^{\infty} \mathbb{P}(Z = y) = 1 - \mathbb{P}(Z = 0)$ by the Law of Total Probability, and we achieve the same equality.

In either case, we conclude that

$$\mathbb{P}(X_i = 0, Y_i = 0) + \mathbb{P}(X_i = 1, Y_i = 0) + \sum_{y=1}^{\infty} \mathbb{P}(X_i = 1, Y_i = y) = \cancel{e^{-p_i}} + 1 - \cancel{e^{-p_i}} = 1.$$

Also, the probabilities are non-negative, since $e^{-p_i} \ge 1 - p_i$ always.

(d) We know that $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 0, Y_i = 0) = 1 - p_i$, and that $\mathbb{P}(X_i = x) = 0$ for any $x \notin \{0, 1\}$. Then $\mathbb{P}(X_i = 0) + \mathbb{P}(X_i = 1) = 1$, so $\mathbb{P}(X_i = 1) = p_i$. This is a sufficient approach, but to be fully explicit, we can verify through direct calculation that $\mathbb{P}(X_i = 1) = p_i$:

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = 1, Y_i = 0) + \sum_{y=1}^{\infty} \mathbb{P}(X_i = 1, Y_i = y) = e^{-p_i} - (1 - p_i) + \sum_{y=1}^{\infty} \frac{e^{-p_i} p_i^y}{y!}$$
$$= \cancel{e^{-p_i} - 1} + p_i + \cancel{1 - e^{-p_i}} = p_i.$$

Hence, X_i has the Bernoulli distribution with probability of success p_i.

(e) We see that $\mathbb{P}(Y_i = 0) = \mathbb{P}(X_i = 0, Y_i = 0) + \mathbb{P}(X_i = 1, Y_i = 0) = e^{-p_i}$, and for $y = 1, 2, \ldots$ we have

$$\mathbb{P}(Y_i = y) = \mathbb{P}(X_i = 1, Y_i = y) = \frac{e^{-p_i} p_i^y}{y!}.$$

This is indeed the Poisson distribution with rate $\lambda = p_i$.

(f) We can recognize that $\mathbb{P}(X_i \neq Y_i) = 1 - \mathbb{P}(X_i = Y_i)$:

$$\begin{aligned}
\mathbb{P}(X_i \neq Y_i) &= 1 - \mathbb{P}(X_i = Y_i) = 1 - \mathbb{P}(X_i = 0, Y_i = 0) - \mathbb{P}(X_i = 1, Y_i = 1) \\
&= 1 - (1 - p_i) - \frac{e^{-p_i} p_i^1}{1!} \\
&= p_i - e^{-p_i} p_i \\
&= p_i (1 - e^{-p_i}) \leq p_i^2.
\end{aligned}$$

In the last line, we are using $1 - e^{-p_i} \leq p_i$. Note that this follows from $e^{-p_i} \geq 1 - p_i$ by rearranging the inequality.

Alternatively, we can compute $\mathbb{P}(X_i \neq Y_i)$ directly:

$$\mathbb{P}(X_i \neq Y_i) = \mathbb{P}(X_i = 1, Y_i = 0) + \mathbb{P}(X_i = 1, Y_i \geq 2) = e^{-p_i} - (1 - p_i) + \sum_{y=2}^{\infty} \frac{e^{-p_i} p_i^y}{y!}$$
$$= e^{-p_i} - (1 - p_i) + 1 - e^{-p_i} - p_i e^{-p_i} = p_i (1 - e^{-p_i}) \leq p_i^2.$$

(g) Thanks to the inequalities we have proved, we can write down

$$d\left(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i\right) \underbrace{\leq}_{(2)} \mathbb{P}\left(\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i\right) \underbrace{\leq}_{(3)} \sum_{i=1}^n \mathbb{P}(X_i \neq Y_i) \leq \sum_{i=1}^n p_i^2.$$

The last inequality is from Part (f).

This is known as Le Cam's Theorem. It provides precise bounds on how far the sum of independent Bernoulli random variables is from a Poisson distribution.

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

Solution:

$$\begin{aligned}
P(X=k|X+Y=n) &= \frac{P(X=k \cap X+Y=n)}{P(X+Y=n)} = \frac{P(X=k \cap Y=n-k)}{P(X+Y=n)} \\
&= \frac{\frac{\lambda^k e^{-\lambda}}{k!} \times \frac{\mu^{n-k} e^{-\mu}}{(n-k)!}}{\frac{(\lambda+\mu)^n e^{-(\lambda+\mu)}}{n!}} \\
&= \frac{n!}{k!(n-k)!} \times \frac{e^{-\lambda} e^{-\mu}}{e^{-(\lambda+\mu)}} \times \frac{\lambda^k \mu^{n-k}}{(\lambda+\mu)^n} \\
&= \binom{n}{k} (\frac{\lambda}{\lambda+\mu})^k (\frac{\mu}{\lambda+\mu})^{n-k} \\
&= \binom{n}{k} (\frac{\lambda}{\lambda+\mu})^k (1-\frac{\lambda}{\lambda+\mu})^{n-k}
\end{aligned}$$

Hence, it is a binomial distribution with $Z \sim \text{Bin}(n, \frac{\lambda}{\lambda+\mu})$.

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

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

Solution:

(a) (i) $\text{cov}(X+Y, X-Y) = \text{cov}(X,X) + \text{cov}(X,Y) - \text{cov}(Y,X) - \text{cov}(Y,Y) = \text{cov}(X,X) - \text{cov}(Y,Y) = 0$

(ii) Observe that $\mathbb{P}(X+Y=7, X-Y=0) = 0$ because if $X-Y=0$, then the sum of our two dice rolls must be even. However both $\mathbb{P}(X+Y=7)$ and $\mathbb{P}(X-Y=0)$ are nonzero so $\mathbb{P}(X+Y=7, X-Y=0) \neq \mathbb{P}(X+Y=7) \cdot \mathbb{P}(X-Y=0)$

(b) Yes. If we write $\mu = \mathbb{E}[X]$, then $0 = \text{Var}(X) = \mathbb{E}[(X-\mu)^2]$ so $(X-\mu)^2$ must be identically 0 since perfect squares are non-negative. Thus $X = \mu$.

(c) No. We have $\text{Var}(cX) = \mathbb{E}[(cX - \mathbb{E}[cX])^2] = c^2\,\mathbb{E}[(X-\mathbb{E}[X])^2] = c^2\,\text{Var}(X)$ so if $\text{Var}(X) \neq 0$ and $c \neq 0$ or $c \neq 1$ then $\text{Var}(cX) \neq c\,\text{Var}(X)$. This does prove that $\sigma(cX) = c\sigma(X)$ though.

(d) No. Let $A = X+Y$ and $B = X-Y$ from part (a). Since A and B are not constants then part (b) says they must have nonzero variances which means they also have nonzero standard deviations. Part (a) says that their covariance is 0 which means they are uncorrelated, and that they are not independent.

Recall from lecture that the converse is true though.

(e) Yes. If $\text{Corr}(X,Y) = 0$, then $\text{cov}(X,Y) = 0$. We have $\text{Var}(X+Y) = \text{cov}(X+Y, X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{cov}(X,Y) = \text{Var}(X) + \text{Var}(Y)$.

(f) Yes. For any values x, y we have $\max(x,y)\min(x,y) = xy$. Thus, $\mathbb{E}[\max(X,Y)\min(X,Y)] = \mathbb{E}[XY]$.

(g) No. You may be tempted to think that because $(\max(x,y), \min(x,y))$ is either (x,y) or (y,x), then $\text{Corr}(\max(X,Y), \min(X,Y)) = \text{Corr}(X,Y)$ because $\text{Corr}(X,Y) = \text{Corr}(Y,X)$. That reasoning is flawed because $(\max(X,Y), \min(X,Y))$ is not always equal to (X,Y) or always equal to (Y,X) and the inconsistency affects the correlation. It is possible for X and Y to be independent while $\max(X,Y)$ and $\min(X,Y)$ are not.

For a concrete example, suppose X is either 0 or 1 with probability $1/2$ each and Y is independently drawn from the same distribution. Then $\text{Corr}(X,Y) = 0$ because X and Y are independent. Even though X never gives information about Y, if you know $\max(X,Y) = 0$ then you know for sure $\min(X,Y) = 0$.

More formally, $\max(X,Y) = 1$ with probability $3/4$ and 0 with probability $1/4$, and $\min(X,Y) = 1$ with probability $1/4$ and 0 with probability $3/4$. This means

$$\mathbb{E}[\max(X,Y)] = 1 * 3/4 + 0 * 1/4 = 3/4$$

and

$$\mathbb{E}[\min(X,Y)] = 1 * 1/4 + 0 * 3/4 = 1/4.$$

Thus,

$$\text{cov}(\max(X,Y), \min(X,Y)) = \mathbb{E}[\max(X,Y)\min(X,Y)] - 3/16 = 1/4 - 3/16 = 1/16 \neq 0.$$

We conclude that $\text{Corr}(\max(X,Y), \min(X,Y)) \neq 0 = \text{Corr}(X,Y)$.

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, $\mathbb{P}(X \geq \mathbb{E}[X] + \alpha) > \frac{\text{Var}(X)}{2\alpha^2}$ or $\mathbb{P}(X \leq \mathbb{E}[X] - \alpha) > \frac{\text{Var}(X)}{2\alpha^2}$.

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 $\text{Var}(X) = 2$.

   (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.)

Solution: See next week's solutions.

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 $\sigma^2 = \text{Var}(X)$.

Solution: See next week's solutions.