HW5 Solutions

1. (50 pts.) Random homeworks again

(a) (8 pts.) Show that if two random variables X and Y are independent, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Answer: Applying the definition of expectation we have

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_x \sum_y xy \; p_{X,Y}(x,y) \\
&= \sum_x \sum_y xy \; p_X(x) p_Y(y) \\
&= \left(\sum_x x \; p_X(x) \right) \left(\sum_y y \; p_Y(y) \right) \\
&= \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

(b) Define the covariance of two random variables X and Y as:

$$\mathrm{Cov}(X,Y) := \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right]$$

(i) (7 pts.) What is the covariance between two independent random variables?
Answer: Note that $X - \mu_X$ is independent of $Y - \mu_Y$ (since they are shifted versions of X and Y respectively). Hence using part(a), we write expectation of the product as the product of the expectations.

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right] \\
&= \mathbb{E}[(X - \mu_X)]\mathbb{E}[(Y - \mu_Y)] \\
&= (\mathbb{E}[X] - \mu_X)(\mathbb{E}[Y] - \mu_Y) \\
&= 0.
\end{aligned}$$

(ii) (7 pts.) For two general random variables X and Y (not necessarily independent), express the variance of $X + Y$ in terms of the variance of X, the variance of Y, and the covariance of X and Y.

Answer: Using the definition of variance we have

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}\left[(X + Y - \mathbb{E}[X+Y])^2\right] \\
&= \mathbb{E}\left[(X - \mu_X + Y - \mu_Y)^2\right] \\
&= \mathbb{E}\left[(X - \mu_X)^2 + (Y - \mu_Y)^2 + 2(X - \mu_X)(Y - \mu_Y)\right]
\end{aligned}$$

Now applying the linearity of expectation we have

$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathbb{E}[(X-\mathbb{E}[X])^2] + \mathbb{E}[(Y-\mathbb{E}[Y])^2] + 2\mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\ \mathrm{Cov}(X,Y)\end{aligned}$$

(iii) (7 pts.) Show that if X and Y are two independent random variances, then

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Give an example of a dependent X and Y for which it is not true.

Answer: Combining our results from the previous parts we can write

$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\ \mathrm{Cov}(X,Y) \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y).\end{aligned}$$

As an example of dependent X and Y choose $Y = X$ and $\mathrm{Var}(X) > 0$. then $\mathrm{Var}(X+Y) = \mathrm{Var}(2X) = 4\mathrm{Var}(X) \neq \mathrm{Var}(X) + \mathrm{Var}(X) = 2\mathrm{Var}(X)$

(c) (7 pts.) Generalize part (b)(ii) to compute the variance of the sum of $X+Y+Z$ of three random variables in terms of the variances of X, Y, and Z as well as the covariances between each of the random variables. What happens when X, Y, Z are mutually independent?

Answer: Using our answer from part (b) we can write

$$\begin{aligned}\mathrm{Var}(X+Y+Z) &= \mathrm{Var}(X+Y) + \mathrm{Var}(Z) + 2\cdot\mathrm{Cov}(X+Y,Z) \\ &= [\mathrm{Var}(X) + \mathrm{Var}(Y) + 2\cdot\mathrm{Cov}(X,Y)] + \mathrm{Var}(Z) + 2\cdot\mathbb{E}[(X+Y-\mu_x-\mu_Y)(Z-\mu_Z)]\end{aligned}$$

We can expand out the last term in the previous line as

$$\begin{aligned}\mathbb{E}[(X+Y-\mu_x-\mu_Y)(Z-\mu_Z)] &= \mathbb{E}[(X-\mu_x)(Z-\mu_Z) + (Y-\mu_Y)(Z-\mu_Z)] \\ &= \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)\end{aligned}$$

Plugging back into our previous equation we have

$$\mathrm{Var}(X+Y+Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Var}(Z) + 2\mathrm{Cov}(X,Z) + 2\mathrm{Cov}(Y,Z) + 2\mathrm{Cov}(X,Z)$$

If X, Y, Z are mutually independent, then $\mathrm{Cov}(X,Z) = \mathrm{Cov}(Y,Z) = \mathrm{Cov}(X,Z) = 0$, hence

$$\mathrm{Var}(X+Y+Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Var}(Z)$$

(d) In the problem where $n = 3$ homeworks are randomly returned to $n = 3$ students, compute the variance of the number of students that get their own homeworks back in two ways.

(i) (7 pts.) Directly by using the distribution of the number of students that get their homework back which we calculated in class.

Answer: Recall from class that the pmf for the $n = 3$ case is given by

$$p_X(x) = \begin{cases} 1/3 & x = 0 \\ 1/2 & x = 1 \\ 1/6 & x = 3 \end{cases}$$

Also recall that the mean is 1. Using this information we calculate the variance as

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_{x\in\{0,1,3\}} (x-1)^2 p_X(x) \\
&= (0-1)^2 \cdot 1/3 + (1-1)^2 \cdot 1/2 + (3-1)^2 \cdot 1/6 \\
&= 1
\end{aligned}$$

(ii) (7 pts.) Using part(d) by an appropriate choice of X,Y, and Z.
Answer: Let X, Y, and Z be indicator random variables such that they are 1 when student 1,2, or 3 gets their homework back respectively and 0 otherwise. We also note that the mean of these indicator random variables is 1/3 (in general the mean of an indicator random variable is the probability that it is 1). By symmetry

$$\mathrm{Var}(X) = \mathrm{Var}(Y) = \mathrm{Var}(Z) = (0-1/3)^2 \cdot 2/3 + (1-1/3)^2 \cdot (1/3) = 2/9$$

Now we compute the covariances as

$$\begin{aligned}
\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,Z) = \mathrm{Cov}(X,Z) &= \mathbb{E}[(X-\mu_X)(Z-\mu_Z)] \\
&= \mathbb{E}[XZ] - \mu_X\mu_Z
\end{aligned}$$

Since there is only one outcome in which both student 1 and student 2 get their correct homework back (namely when all 3 get them back), it has $\mathbf{P}(XZ=1) = 1/6$ so the expectation is

$$\mathbb{E}[XZ] = 1 \cdot 1/6 + 0 \cdot 5/6$$

Plugging into our expression for the covariance we find

$$\mathrm{Cov}(X,Z) = 1/6 - 1/9 = 1/18$$

Using part(d) we have

$$\mathrm{Var}(X+Y+Z) = 3 \cdot 2/9 + 2 \cdot 3 \cdot 1/18 = 1$$

2. (50 pts.) Bit Torrent

Consider the Bit Torrent problem where m chunks of a movie are randomly distributed across infinite number of servers. Each server has a uniformly and independently selected chunk of the movie. I query the servers one-by-one and download whatever chunk I find on each server to my local hard disk. Suppose each server query takes 1 second and each chunk of the movie plays for t seconds (t integer). Answer each of these questions. None of the final answers should involve summations.

(a) (10 pts.) What is the expected time to get the first chunk of the movie so that I can start watching the movie? What is the distribution of this random variable?

Answer: Let T_1 be the random variable denoting the time to get the first chunk of the movie then $T_1 \sim Geom(1/m)$ which we know has mean given by $\mathbb{E}[T_1] = m$

(b) (20 pts.) What is the expected time to get the first two chunks of the movie? (Hint: consider the first time to see either the first chunk or the second chunk of the movie.)

Answer: Let $\tilde{T}_1 \sim Geom(2/m)$ be the time to get either the first or second chunk of the movie and $\tilde{T}_2 \sim Geom(1/m)$ be the time to get the remaining chunk. The second random variable is geometric by the memoryless property of the geometric distribution. Our total time to collect the first two movies is $\tilde{T} = \tilde{T}_1 + \tilde{T}_2$ so by linearity of expectation and the mean of a geometric random variable,

$$\begin{aligned}
\mathbb{E}[\tilde{T}] &= \mathbb{E}[\tilde{T}_1] + \mathbb{E}[\tilde{T}_2] \\
&= m/2 + m \\
&= 3/2 \cdot m
\end{aligned}$$

(c) (20 pts.) I start playing the movie once I get the first chunk of the movie. What is the probability that I have the second chunk of the movie on my hard disk *before* I finish playing the first chunk so that I can immediately continue playing the movie?

Answer: Let T_1 and T_2 be the time until we collect the first and second chunks respectively then we want to calculate the following probability $\mathbf{P}(T_2 - T_1 \leq t)$ where $t \in \mathbb{Z}_{++}$ (this notation means t is a *positive* integer. With just one + means including 0). Using total probability we can write

$$\begin{aligned}
\mathbf{P}(T_2 - T_1 \leq t) &= \mathbf{P}(T_2 - T_1 \leq t, T_2 > T_1) + \mathbf{P}(T_2 - T_1 \leq t, T_2 < T_1) \\
&= \mathbf{P}(T_2 - T_1 \leq t | T_2 > T_1)\mathbf{P}(T_2 > T_1) + \mathbf{P}(T_2 - T_1 \leq t | T_2 < T_1)\mathbf{P}(T_2 < T_1)
\end{aligned}$$

By the symmetry of the problem we know $\mathbf{P}(T_2 < T_1) = \mathbf{P}(T_2 > T_1) = 1/2$. Furthermore, the quantity $\mathbf{P}(T_2 - T_1 \leq t | T_2 < T_1) = 1$ since t is defined to be a positive quantity. Plugging in we can write,

$$\begin{aligned}
\mathbf{P}(T_2 - T_1 \leq t) &= 1/2 + \mathbf{P}(T_2 - T_1 \leq t | T_2 > T_1)1/2 \\
&= 1/2 + 1/2 \cdot \mathbf{P}(T_2 \leq t)
\end{aligned}$$

Now using the PMF of the geometric distribution and the sum of a geometric series we can simplify

$$\mathbf{P}(T_2 - T_1 \leq t) = 1/2 + 1/2 \sum_{i=1}^{t} (1 - 1/m)^{i-1} \frac{1}{m}$$

$$\mathbf{P}(T_2 - T_1 \leq t) = 1/2 + 1/2 \cdot \frac{1}{m} \sum_{i=0}^{t-1} (1 - 1/m)^{i}$$

$$\begin{aligned} \mathbf{P}(T_2 - T_1 \leq t) &= 1/2 + 1/2 \cdot \frac{1}{m} \frac{1 - (1 - 1/m)^t}{1 - (1 - 1/m)} \\ &= 1 - 1/2 \cdot (1 - 1/m)^t \end{aligned}$$

So our final answer is

$$\mathbf{P}(T_2 - T_1 \leq t) = 1 - \frac{1}{2}(1 - 1/m)^t$$