CS 70 Discrete Mathematics and Probability Theory
Fall 2020 HW 11

1 Random Variables Warm-Up

Let X and Y be random variables, each taking values in the set $\{0,1,2\}$, with joint distribution

$$\begin{array}{lll} \mathbb{P}[X=0,Y=0]=1/3 & \mathbb{P}[X=0,Y=1]=0 & \mathbb{P}[X=0,Y=2]=1/3 \\ \mathbb{P}[X=1,Y=0]=0 & \mathbb{P}[X=1,Y=1]=1/9 & \mathbb{P}[X=1,Y=2]=0 \\ \mathbb{P}[X=2,Y=0]=1/9 & \mathbb{P}[X=2,Y=1]=1/9 & \mathbb{P}[X=2,Y=2]=0. \end{array}$$

(a) What are the marginal distributions of X and Y?

(b) What are $\mathbb{E}[X]$ and $\mathbb{E}[Y]$?

(c) (optional) What are $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$?

(d) Let I be the indicator that $X=1$, and J be the indicator that $Y=1$. What are $\mathbb{E}[I]$, $\mathbb{E}[J]$ and $\mathbb{E}[IJ]$?

(e) In general, let I_A and I_B be the indicators for events A and B in a probability space $(\Omega,\mathbb{P})$. What is $\mathbb{E}[I_A I_B]$, in terms of the probability of some event?

Solution:

(a) By the law of total probabilty

$$\mathbb{P}[X=0]=\mathbb{P}[X=0,Y=0]+\mathbb{P}[X=0,Y=1]+\mathbb{P}[X=0,Y=2]=1/3+0+1/3=2/3$$

and similarly

$$\mathbb{P}[X=1]=0+1/9+0=1/9$$
$$\mathbb{P}[X=2]=1/9+1/9+0=2/9.$$

As a sanity check, these three numbers are all positive and they add up to $2/3+1/9+2/9=1$ as they should. The same kind of calculation gives

$$\mathbb{P}[Y=0]=1/3+0+1/9=4/9$$
$$\mathbb{P}[Y=1]=0+1/9+1/9=2/9$$
$$\mathbb{P}[Y=2]=1/3.$$

(b) From the above marginal distributions, we can compute

$$\mathbb{E}[X]=0\mathbb{P}[X=0]+1\mathbb{P}[X=1]+2\mathbb{P}[X=2]=5/9$$
$$\mathbb{E}[Y]=0\mathbb{P}[Y=0]+1\mathbb{P}[Y=1]+2\mathbb{P}[Y=2]=8/9$$

(c) Again using our marginal distributions,

$$\mathbb{E}[X^2] = 0\mathbb{P}[X=0] + 1\mathbb{P}[X=1] + 4\mathbb{P}[X=2] = 1$$
$$\mathbb{E}[Y^2] = 0\mathbb{P}[Y=0] + 1\mathbb{P}[Y=1] + 4\mathbb{P}[Y=2] = 14/9$$

and thus

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 56/81$$
$$\text{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = 62/81.$$

We didn't ask you to do compute the covariance on the homework, but it is

$$\begin{aligned}
\text{cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] && (1)\\
&= 0\mathbb{P}[XY=0] + 1\mathbb{P}[XY=1] + 2\mathbb{P}[XY=2] + 4\mathbb{P}[XY=4] - 40/81 && (2)\\
&= 0\cdot 1/3 + 1\cdot 1/9 + 2\cdot 1/9 + 4\cdot 0 - 40/81 && (3)\\
&= -13/81. && (4)
\end{aligned}$$

(d) We know that taking the expectation of an indicator for some event gives the probability of that event, so

$$\mathbb{E}[I] = \mathbb{P}[X=1] = 1/9$$
$$\mathbb{E}[J] = \mathbb{P}[Y=1] = 2/9.$$

The random variable IJ is equal to one if $I=1$ and $J=1$, and is zero otherwise. In other words, it is the indicator for the event that $I=1$ *and* $J=1$:

$$\mathbb{E}[IJ] = \mathbb{P}[I=1, J=1] = 1/9.$$

(e) By what we said in the previous part of the solution, $I_A I_B$ is the indicator for the event $A \cap B$, so

$$\mathbb{E}[I_A I_B] = \mathbb{P}[A \cap B].$$

2 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

$$\mathbb{P}[X_i = 1, X_j = -1] = \frac{1}{2} \qquad \mathbb{P}[X_i = -1, X_j = 1] = \frac{1}{2} \qquad \mathbb{P}[X_i = X_j] = 0? \qquad (5)$$

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.

Solution:

(a) No such random variables can exist; let's prove it by contradiction. From the desired joint distribution of X_1 and X_2, we claim that $X_1 = -X_2$ (by which we mean that for every ω in the sample space $X_1(\omega) = -X_2(\omega)$). Similarly, we would need to have $X_2 = -X_3$ and $X_3 = -X_1$. But now

$$X_1 = -X_2 = X_3 = -X_1,$$

a contradiction since $X_1 \in \{+1, -1\}$.

(b) This is only possible if n is even. When $n = 2k+1$, the same argument as above gives us

$$X_1 = -X_2 = X_3 = \cdots = -X_{2k} = X_{2k+1} = -X_1,$$

a contradiction for the same reason as before. However, when $n = 2k$, we can set $X_1, \ldots, X_{2k}$ to have the joint distribution

$$\mathbb{P}[X_1 = 1, X_2 = -1, \cdots X_{2k} = -1] = 1/2$$
$$\mathbb{P}[X_1 = -1, X_2 = 1, \cdots X_{2k} = 1] = 1/2.$$

3 Testing Model Planes

Amin is testing model airplanes. He starts with n model planes which each independently have probability p of flying successfully each time they are flown, where $0 < p < 1$. Each day, he flies every single plane and keeps the ones that fly successfully (i.e. don't crash), throwing away all other models. He repeats this process for many days, where each "day" consists of Amin flying any remaining model planes and throwing away any that crash. Let X_i be the random variable representing how many model planes remain after i days. Note that $X_0 = n$. Justify your answers for each part.

(a) What is the distribution of X_1? That is, what is $\mathbb{P}[X_1 = k]$?

(b) What is the distribution of X_2? That is, what is $\mathbb{P}[X_2 = k]$? Name the distribution of X_2 and what its parameters are.

(c) Repeat the previous part for X_t for arbitrary $t \geq 1$.

(d) What is the probability that at least one model plane still remains (has not crashed yet) after t days? Do not have any summations in your answer.

(e) Considering only the first day of flights, is the event A_1 that the first and second model planes crash independent from the event B_1 that the second and third model planes crash? Recall that two events A and B are independent if $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$. Prove your answer using this definition.

(f) Considering only the first day of flights, let A_2 be the event that the first model plane crashes *and* exactly two model planes crash in total. Let B_2 be the event that the second plane crashes on the first day. What must n be equal to in terms of p such that A_2 is independent from B_2? Prove your answer using the definition of independence stated in the previous part.

(g) Are the random variables X_i and X_j, where $i < j$, independent? Recall that two random variables X and Y are independent if $\mathbb{P}[X = k_1 \cap Y = k_2] = \mathbb{P}[X = k_1]\mathbb{P}[Y = k_2]$ for all k_1 and k_2. Prove your answer using this definition.

Solution:

(a) Since Amin is performing n trials (flying a plane), each with an independent probability of "success" (not crashing), we have $X_1 \sim \text{Binom}(n, p)$, or $\mathbb{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$, for $0 \leq k \leq n$.

(b) Each model plane independently has probability p^2 of surviving both days. Whether a model plane survives both days is still independent from whether any other model plane survives both days, so we can say $X_2 \sim \text{Binom}(n, p^2)$, or $\mathbb{P}[X = k] = \binom{n}{k} p^{2k} (1-p^2)^{n-k}$, for $0 \leq k \leq n$.

(c) By extending the previous part we see each model plane has probability p^t of surviving t days, so $X_t \sim \text{Binom}(n, p^t)$, or $\mathbb{P}[X = k] = \binom{n}{k} p^{tk} (1-p^t)^{n-k}$, for $0 \leq k \leq n$.

(d) We consider the complement, the probability that no model planes remain after t days. By the previous part we know this to be $\mathbb{P}[X_t = 0] = \binom{n}{0} p^{t(0)} (1-p^t)^{n-0} = (1-p^t)^n$. So the probability of at least model plane remaining after t days is $1-(1-p^t)^n$.

(e) No. $\mathbb{P}[A_1 \cap B_1]$ is the probability that the first three model planes crash, which is $(1-p)^3$. But $\mathbb{P}[A_1]\mathbb{P}[B_1] = (1-p)^2(1-p)^2 = (1-p)^4$. So $\mathbb{P}[A_1 \cap B_1] \neq \mathbb{P}[A_1]\mathbb{P}[B_1]$ and A_1 and B_1 are not independent.

(f) $\mathbb{P}[A_1 \cap B_1]$ is the probability that only the first model plane and second model plane crash, which is $(1-p)^2 p^{n-2}$. $\mathbb{P}[A_1]$ is the probability that the first model plane crashes, and exactly one of the remaining $n-1$ model planes crashes, so $\mathbb{P}[A_2] = (1-p) \cdot \binom{n-1}{1}(1-p)p^{n-1-1} = (n-1)(1-p)^2 p^{n-2}$. Trivially, we have $\mathbb{P}[B_2] = 1-p$, so $\mathbb{P}[A_2]\mathbb{P}[B_2] = (n-1)(1-p)^3 p^{n-2}$ which is equal to $\mathbb{P}[A_2 \cap B_2] = (1-p)^2 p^{n-2}$ only when $(n-1)(1-p) = 1$, or when $n = \frac{1}{1-p} + 1$.

(g) No. Let $k_1 = 0$ and $k_2 = 1$. Then $\mathbb{P}[X_i = k_1 \cap X_j = k_2] = 0$ because you can't have 1 plane at the end of day 2 if there are no planes left at the end of day 1. But $\mathbb{P}[X_i = k_1] > 0$ and $\mathbb{P}[X_j = k_2] > 0$ so $\mathbb{P}[X_i = k_1]\mathbb{P}[X_j = k_2] > 0$. Since $\mathbb{P}[X_i = k_1]\mathbb{P}[X_j = k_2] \neq \mathbb{P}[X_i = k_1 \cap X_j = k_2]$, they are not independent.

4 Graph

Consider a random graph (undirected, no multi-edges, no self-loops) on n nodes, where each possible edge exists independently with probability p. Let X be the number of isolated nodes (nodes with degree 0).

(a) What is $\mathrm{E}(X)$? *Consider X to be the sum of the indicators X_i that vertex i is isolated. Why isn't X a binomial random variable?*

(b) (optional) What is $\mathrm{Var}(X)$?

Solution:

(a) Let's first pause and ask ourselves why X is not binomial. If we consider a trial as adding an edge, which happens with probability p, we will have $\frac{n(n-1)}{2}$ trials. If we were interested in the number of edges that the resulting graph has, then it would be binomial. But unfortunately, that is not the random variable we're looking for.

Since we are interested in the number of isolated nodes, we must instead consider a trial creating an isolated node, which happens with probability $(1-p)^{n-1}$. However, now our trials are not independent. For example, given that a node is not isolated, the conditional probability of all nodes connected to that node being isolated becomes 0.

So how can we solve this problem? Let's introduce some indicator variables $X_1, X_2, \ldots, X_n$, where $X_i = 1$ if node i is isolated.
Note that $\mathbb{P}[X_i = 1] = (1-p)^{n-1}$, and thus $\mathbb{E}(X_i) = (1-p)^{n-1}$.
Now, we can rewrite X as

$$X = X_1 + X_2 + \cdots + X_n$$

Using the Linearity of Expectation, we know

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}(X_1 + X_2 + \cdots + X_n) \\
&= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) \\
&= (1-p)^{n-1} + (1-p)^{n-1} + \cdots + (1-p)^{n-1} \\
&= n(1-p)^{n-1}
\end{aligned}$$

What happened here? We ended up with the same expectation as a binomial distribution, even though it wasn't binomial. In general, many different distributions can have the same expectation, but can vary greatly. This is one such example. Another is the two following distributions: one that is always $\frac{1}{2}$, and another is a fair coin toss. Both have the same expectation, but are very different.

5 Triangles in Random Graphs

Let's say we make a simple and undirected graph G on n vertices by randomly adding m edges, without replacement. In other words, we choose the first edge uniformly from all $\binom{n}{2}$ possible edges, then the second one uniformly from among the remaining $\binom{n}{2} - 1$ edges, etc. What is the expected number of triangles in G? (A triangle is a triplet of distinct vertices with all three edges present between them.)

Solution:

Let's label our vertices $1,\ldots n$, and first check the probability that vertices $1,2,3$ form a triangle. This event is described by a hypergeometric distribution with parameters $\binom{n}{2}$, 3, m: when we make the graph we are drawing the m edges from a bucket of $\binom{n}{2}$ possible edges, 3 of which are the ones connecting vertices 1, 2, and 3. Thus the probability that all three of these edges exist is

$$\mathbb{P}[1,\ 2,\ \text{and } 3 \text{ form a triangle}] = \frac{\binom{3}{3}\binom{\binom{n}{2}-3}{m-3}}{\binom{\binom{n}{2}}{m}}$$

In fact, there was nothing special about vertices $1,2,3$ in this calculation. The probability that the three edges connecting some triplet of distinct vertices i,j,k is equal to the quantity above. Now, for each subset $\{i,j,k\} \subset \{1,\ldots,n\}$, let $I_{i,j,k}$ be the indicator that these three vertices form a triangle. We then have

$$\begin{aligned}
\mathbb{E}[\#\text{ of triangles}] &= \mathbb{E} \sum_{\{i,j,k\}\subset\{1,\ldots,n\}} I_{i,j,k} \\
&= \sum_{\{i,j,k\}\subset\{1,\ldots,n\}} \mathbb{E}[I_{i,j,k}] && \text{linearity of expectation} \\
&= \sum_{\{i,j,k\}\subset\{1,\ldots,n\}} \mathbb{P}[i,\ j,\ \text{and } k \text{ form a triangle}] \\
&= \sum_{\{i,j,k\}\subset\{1,\ldots,n\}} \frac{\binom{3}{3}\binom{\binom{n}{2}-3}{m-3}}{\binom{\binom{n}{2}}{m}} \\
&= \binom{n}{3} \frac{\binom{3}{3}\binom{\binom{n}{2}-3}{m-3}}{\binom{\binom{n}{2}}{m}}
\end{aligned}$$