1 Random Variables Warm-Up

Note 15

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) 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 $\mathbb{E}[IJ]$?

(d) 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 probability

$$\begin{aligned} \mathbb{P}[X=0] &= \mathbb{P}[X=0,Y=0]+\mathbb{P}[X=0,Y=1]+\mathbb{P}[X=0,Y=2] \\ &= \frac{1}{3}+0+\frac{1}{3}=\frac{2}{3} \end{aligned}$$

and similarly

$$\begin{aligned} \mathbb{P}[X=1] &= 0+\frac{1}{9}+0=\frac{1}{9} \\ \mathbb{P}[X=2] &= \frac{1}{9}+\frac{1}{9}+0=\frac{2}{9} \end{aligned}$$

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

$$\begin{aligned} \mathbb{P}[Y=0] &= \frac{1}{3}+0+\frac{1}{9}=\frac{4}{9} \\ \mathbb{P}[Y=1] &= 0+\frac{1}{9}+\frac{1}{9}=\frac{2}{9} \\ \mathbb{P}[Y=2] &= \frac{1}{3} \end{aligned}$$

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

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

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

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

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

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

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

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

2 Testing Model Planes

Note 15

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 all 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 $\mathbb{P}[X_2 = k]$? Recognize the distribution of X_2 as one of the famous ones and provide its name and parameters.

(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 $X_1 \sim \text{Binomial}(n, p)$, 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 $X_2 \sim \text{Binomial}(n, p^2)$, 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 $X_t \sim \text{Binomial}(n, p^t)$, or $\mathbb{P}[X = k] = \binom{n}{k} (p^t)^k (1-p^t)^{n-k}$, 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

$$\mathbb{P}[X_t = 0] = \binom{n}{0} (p^t)^0 (1-p^t)^{n-0} = (1-p^t)^n.$$

This means that 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) $\mathbb{P}[A_2 \cap B_2]$ is the probability that only the first model plane and second model plane crash, which is $(1-p)^2 p^{n-2}$. $\mathbb{P}[A_2]$ is the probability that the first model plane crashes, and exactly one of the remaining $n-1$ model planes crashes, so

$$\mathbb{P}[A_2] = (1-p) \cdot \binom{n-1}{1} (1-p) p^{n-1-1} = (n-1)(1-p)^2 p^{n-2}.$$

We also have $\mathbb{P}[B_2] = 1-p$, so we want to solve for n in

$$\begin{aligned}
\mathbb{P}[A_2 \cap B_2] &= \mathbb{P}[A_2]\mathbb{P}[B_2] \\
(1-p)^2 p^{n-2} &= \underbrace{(n-1)(1-p)^2 p^{n-2}}_{\mathbb{P}[A_2]} \underbrace{(1-p)}_{\mathbb{P}[B_2]} \\
(1-p)^2 p^{n-2} &= (n-1)(1-p)^3 p^{n-2} \\
1 &= (n-1)(1-p) \\
n &= 1 + \frac{1}{1-p}
\end{aligned}$$

(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. However, $\mathbb{P}[X_i = k_1] > 0$ 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.

3 Class Enrollment

Note 15
Note 19

Lydia has just started her CalCentral enrollment appointment. She needs to register for a geography class and a history class. There are no waitlists, and she can attempt to enroll once per day in either class or both. The CalCentral enrollment system is strange and picky, so the probability of enrolling successfully in the geography class on each attempt is p_g and the probability of enrolling successfully in the history class on each attempt is p_h. Also, these events are independent.

(a) Suppose Lydia begins by attempting to enroll in the geography class everyday and gets enrolled in it on day G. What is the distribution of G?

(b) Suppose she is not enrolled in the geography class after attempting each day for the first 7 days. What is $\mathbb{P}[G = i \mid G > 7]$, the conditional distribution of G given $G > 7$?

(c) Once she is enrolled in the geography class, she starts attempting to enroll in the history class from day $G+1$ and gets enrolled in it on day H. Find the expected number of days it takes Lydia to enroll in both the classes, i.e. $\mathbb{E}[H]$.

Suppose instead of attempting one by one, Lydia decides to attempt enrolling in both the classes from day 1. Let G be the number of days it takes to enroll in the geography class, and H be the number of days it takes to enroll in the history class.

(d) What is the distribution of G and H now? Are they independent?

(e) Let A denote the day she gets enrolled in her first class and let B denote the day she gets enrolled in both the classes. What is the distribution of A?

(f) What is the expected number of days it takes Lydia to enroll in both classes now, i.e. $\mathbb{E}[B]$?

(g) What is the expected number of classes she will be enrolled in by the end of 30 days?

Solution:

(a) $G \sim \text{Geometric}(p_g)$.

(b) Given that $G > 7$, the random variable G takes values in $\{8, 9, \ldots\}$. For $i = 8, 9, \ldots$,

$$\mathbb{P}[G = i \mid G > 7] = \frac{\mathbb{P}[G = i \wedge G > 7]}{\mathbb{P}[G > 7]} = \frac{\mathbb{P}[G = i]}{\mathbb{P}[G > 7]}$$
$$= \frac{p_g(1-p_g)^{i-1}}{(1-p_g)^7} = p_g(1-p_g)^{i-8}$$

If K denotes the additional number of days it takes to get enrolled in the geography class after day 7, i.e. $K = G - 7$, then conditioned on $G > 7$, the random variable K has the geometric distribution with parameter p_g. Note that this is the same as the distribution of G. This is known as the memoryless property of geometric distribution.

(c) We have $H - G \sim \text{Geometric}(p_h)$. This means that $\mathbb{E}[G] = \frac{1}{p_g}$ and $\mathbb{E}[H - G] = \frac{1}{p_h}$, and as such

$$\mathbb{E}[H] = \mathbb{E}[G] + \mathbb{E}[H - G] = \frac{1}{p_g} + \frac{1}{p_h}.$$

(d) $G \sim \text{Geometric}(p_g)$, $H \sim \text{Geometric}(p_h)$. Yes they are independent.

(e) We have $A = \min\{G, H\}$ and $B = \max\{G, H\}$. We also use the following definition of the minimum:

$$\min(g,h) = \begin{cases} g & \text{if } g \le h; \\ h & \text{if } g > h. \end{cases}$$

Now, for all $k \in \{1, 2, \ldots\}$, $\min(G, H) = k$ is equivalent to $(G = k) \cap (H \ge k)$ or $(H = k) \cap (G > k)$. Hence,

$$\begin{aligned}
\mathbb{P}[A = k] &= \mathbb{P}[\min(G,H) = k] \\
&= \mathbb{P}[(G = k) \cap (H \ge k)] + \mathbb{P}[(H = k) \cap (G > k)] \\
&= \mathbb{P}[G = k] \cdot \mathbb{P}[H \ge k] + \mathbb{P}[H = k] \cdot \mathbb{P}[G > k] && (G,\ H \text{ are independent}) \\
&= [(1-p_g)^{k-1} p_g](1-p_h)^{k-1} + [(1-p_h)^{k-1} p_h](1-p_g)^k && (G,\ H \text{ are geometric}) \\
&= ((1-p_g)(1-p_h))^{k-1} (p_g + p_h(1-p_g)) \\
&= (1 - p_g - p_h + p_h p_g)^{k-1} (p_g + p_h - p_g p_h).
\end{aligned}$$

But this final expression is precisely the probability that a geometric RV with parameter $p_g + p_h - p_g p_h$ takes the value k. Hence $A \sim \text{Geom}(p_g + p_h - p_g p_h)$.

An alternative, slightly cleaner approach is to work with the *tail probabilities* of the geometric distribution, rather than with the usual point probabilities as above. In other words, we can work with $\mathbb{P}[A \geq k]$ rather than with $\mathbb{P}[A = k]$; clearly the values $\mathbb{P}[A \geq k]$ specify the values $\mathbb{P}[A = k]$ since $\mathbb{P}[A = k] = \mathbb{P}[A \geq k] - \mathbb{P}[A \geq (k+1)]$, so it suffices to calculate them instead. We then get the following argument:

$$\begin{aligned}
\mathbb{P}[A \geq k] &= \mathbb{P}[\min(G,H) \geq k] \\
&= \mathbb{P}[(G \geq k) \cap (H \geq k)] \\
&= \mathbb{P}[G \geq k] \cdot \mathbb{P}[H \geq k] && (G, H \text{ are independent}) \\
&= (1-p_g)^{k-1}(1-p_h)^{k-1} && (G, H \text{ are geometric}) \\
&= ((1-p_g)(1-p_h))^{k-1} \\
&= (1-p_g-p_h+p_g p_h)^{k-1}.
\end{aligned}$$

This is the tail probability of a geometric distribution with parameter $p_g + p_h - p_g p_h$, so we are done.

(f) From part (e) we get $\mathbb{E}[A] = \frac{1}{p_g+p_h-p_g p_h}$. From part (d) we have $\mathbb{E}[G] = \frac{1}{p_g}$ and $\mathbb{E}[H] = \frac{1}{p_h}$. We now observe that $\min\{g,h\} + \max\{g,h\} = g+h$; using linearity of expectation, this means that $\mathbb{E}[A] + \mathbb{E}[B] = \mathbb{E}[G] + \mathbb{E}[H]$. As such, we have

$$\mathbb{E}[B] = \frac{1}{p_g} + \frac{1}{p_h} - \frac{1}{p_g + p_h - p_g p_h}.$$

(g) Let I_G and I_H be the indicator random variables of the events "$G \leq 30$" and "$H \leq 30$" respectively. Then $I_G + I_H$ is the number of classes she will be enrolled in within 30 days. Hence the answer is

$$\mathbb{E}[I_G] + \mathbb{E}[I_H] = \mathbb{P}[G \leq 30] + \mathbb{P}[H \leq 30] = 1 - (1-p_g)^{30} + 1 - (1-p_h)^{30}.$$

4 Geometric and Poisson

Note 19

Let $X \sim \text{Geometric}(p)$ and $Y \sim \text{Poisson}(\lambda)$ be independent random variables. Compute $\mathbb{P}[X > Y]$. Your final answer should not have summations.

Hint: Use the total probability rule.

Solution: We condition on Y so we can use the nice property of geometric random variables that

$\mathbb{P}[X > k] = (1-p)^k$. This gives

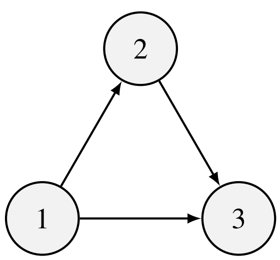
$$\begin{aligned}\mathbb{P}[X > Y] &= \sum_{y=0}^{\infty} \mathbb{P}[X > Y \mid Y = y] \cdot \mathbb{P}[Y = y] \\ &= \sum_{y=0}^{\infty} (1-p)^y \cdot \frac{e^{-\lambda}\lambda^y}{y!} \\ &= e^{-\lambda p} e^{\lambda p} \sum_{y=0}^{\infty} \frac{e^{-\lambda}(\lambda(1-p))^y}{y!} \\ &= e^{-\lambda p} \sum_{y=0}^{\infty} \frac{e^{-\lambda(1-p)}(\lambda(1-p))^y}{y!} \\ &= e^{-\lambda p}\end{aligned}$$

To simplify the last summation, we observe that the sum could be interpreted as the sum of the probabilities for a Poisson($\lambda(1-p)$) random variable, which is equal to 1. Alternatively, you can use the Taylor series $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ to simplify the sum.

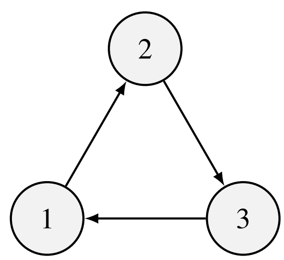
5 Random Tournaments

Note 15

A *tournament* is a directed graph in which every pair of vertices has exactly one directed edge between them—for example, Fig. 1 has examples of two tournaments on the vertices $\{1,2,3\}$.



(a)



(b)

Figure 1: Examples of tournament graphs

In the first tournament above (Fig. 1a), $(1,2,3)$ is a *Hamiltonian path*, since it visits all the vertices exactly once, without repeating any edges, but $(1,2,3,1)$ is not a valid *Hamiltonian cycle*, because the tournament contains the directed edge $1 \to 3$ and not $3 \to 1$. In the second tournament (Fig. 1b), $(1,2,3,1)$ is a *Hamiltonian cycle*, as are $(2,3,1,2)$ and $(3,1,2,3)$; for this problem we'll say that these are all different Hamiltonian cycles, since their start/end points are different.

Consider the following way of choosing a random tournament T on n vertices: independently for each (unordered) pair of distinct vertices $\{i, j\} \subset \{1,\ldots,n\}$, flip a coin and include the edge $i \to j$ in the graph if the outcome is heads, and the edge $j \to i$ if tails.

(a) What is the expected number of Hamiltonian paths in T?

(b) What is the expected number of Hamiltonian cycles in T?

Solution:

(a) Each possible Hamiltonian path in the graph corresponds to a permutation σ of the numbers $1,\ldots,n$, where $\sigma(1)$ is the starting vertex, $\sigma(2)$ is the second vertex visited, etc. If we write I_σ for the indicator random variable that σ corresponds to an actual Hamiltonian cycle in T, then

$$\mathbb{E}\left[\#\text{ Hamiltonian Paths}\right] = \mathbb{E}\left[\sum_\sigma I_\sigma\right] = \sum_\sigma \mathbb{P}[\sigma \text{ is a Hamiltonian path in } T]$$

In order for each σ to correspond to an actual Hamiltonian path in T, the edges $\sigma(i) \to \sigma(i+1)$, for $i = 1,\ldots,n-1$ must all be included in the graph. Since the orientations of the edges in T are independent, with $\sigma(i) \to \sigma(i+1)$ occurring with probability $1/2$, the probability that they are all included is $2^{-(n-1)}$. There are $n!$ possible permutations, so we have

$$\mathbb{E}\left[\#\text{ Hamiltonian Paths}\right] = \frac{n!}{2^{n-1}}.$$

(b) The situation for Hamiltonian cycles is similar. Each possible Hamiltonian cycle each possible cycle corresponds to a permutation σ, but this time in order for σ to be a valid Hamiltonian cycle, T must include the edges $\sigma(i) \to \sigma(i+1)$ for all $i = 1,\ldots,n-1$, as well as the edge $\sigma(n) \to \sigma(1)$. As above, these n edges are oriented independently of one another, so

$$\mathbb{E}\left[\#\text{ Hamiltonian Cycles}\right] = \frac{n!}{2^n}.$$

6 Swaps and Cycles

Note 15

We'll say that a permutation $\pi = (\pi(1),\ldots,\pi(n))$ contains a *swap* if there exist $i,j \in \{1,\ldots,n\}$ so that $\pi(i) = j$ and $\pi(j) = i$, where $i \neq j$.

(a) What is the expected number of swaps in a random permutation?

(b) In the same spirit as above, we'll say that π contains a *k-cycle* if there exist $i_1,\ldots,i_k \in \{1,\ldots,n\}$ with $\pi(i_1) = i_2, \pi(i_2) = i_3, \ldots, \pi(i_k) = i_1$. Compute the expectation of the number of k-cycles.

Solution:

(a) As a warm-up, let's compute the probability that 1 and 2 are swapped. There are $n!$ possible permutations, and $(n-2)!$ of them have $\pi(1) = 2$ and $\pi(2) = 1$. This means

$$\mathbb{P}[(1,2) \text{ are a swap}] = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}.$$

There was nothing special about 1 and 2 in this calculation, so for any $\{i,j\} \subset \{1,\ldots,n\}$, the probability that i and j are swapped is the same as above. Let's write $I_{i,j}$ for the indicator that i and j are swapped, and N for the total number of swaps, so that

$$\mathbb{E}[N] = \mathbb{E}\left[\sum_{\{i,j\}\subset\{1,\ldots,n\}} I_{i,j}\right] = \sum_{\{i,j\}\subset\{1,\ldots n\}} \mathbb{P}[(i,j) \text{ are swapped}] = \frac{1}{n(n-1)}\binom{n}{2} = \frac{1}{2}.$$

(b) The idea here is quite similar to the above, so we'll be a little less verbose in the exposition. However, as a first aside we need the notion of a *cyclic ordering* of k elements from a set $\{1,\ldots n\}$. We mean by this a labelling of the k beads of a necklace with elements of the set, where we say that labellings of the beads are the same if we can move them along the string to turn one into the other. For example, $(1,2,3,4)$ and $(1,2,4,3)$ are different cyclic orderings, but $(1,2,3,4)$ and $(2,3,4,1)$ are the same. There are

$$\binom{n}{k}\frac{k!}{k} = \frac{n!}{(n-k)!}\frac{1}{k}$$

possible cyclic orderings of length k from a set with n elements, since if we first count all subsets of size k, and then all permutations of each of those subsets, we have overcounted by a factor of k.

Now, let N be a random variable counting the number of k-cycles, and for each cyclic ordering $(i_1,\ldots,i_k)$ of k elements of $\{1,\ldots,n\}$, let $I_{(i_1,\ldots,i_k)}$ be the indicator that $\pi(i_1) = i_2, \pi(i_2) = i_3, \ldots, \pi(i_k) = i_1$. There are $(n-k)!$ permutations in which $(i_1,\ldots,i_k)$ form an k-cycle (since we are free to do whatever we want to the remaining $(n-k)$ elements of $\{1,\ldots,n\}$), so the probability that $(i_1,\ldots,i_k)$ are such a cycle is $\frac{(n-k)!}{n!}$, and

$$\mathbb{E}[N] = \mathbb{E}\left[\sum_{(i_1,\ldots,i_k)\text{ cyclic ordering}} I_{(i_1,\ldots,i_k)}\right] = \frac{n!}{(n-k)!}\frac{1}{k}\frac{(n-k)!}{n!} = \frac{1}{k}.$$