Alex Tsun
CSE 312: Foundations of Computing II Section #6
July 30th, 2020

Section #6 Solutions

Luxi Wang, Shreya Jayaraman, Alex Tsun, Pemi Nguyen

1. Let X and Y be independent geometric random variables with the same success parameter $p < 1$. Find the distribution of $X + Y$, using n as the value of $X + Y$. That is solve for: $\mathbb{P}(X + Y = n)$ in terms of p.

Solution

Since X and Y are both geometric RVs, both of their PMFs are as follows:

$$\mathbb{P}(X = k) = \mathbb{P}(Y = k) = p(1-p)^{k-1}$$

So, by convolution, we have the following:

$$\mathbb{P}(X + Y = n) = \sum_{k \in \Omega_X} \mathbb{P}(X = k)\mathbb{P}(Y = n - k)$$

We need to consider which values in this sum are nonzero. That is, for what values of k are $\mathbb{P}(X = k)$ and $\mathbb{P}(Y = n - k)$ both nonzero. Well, k must be at least 1, since X is a geometric random variable. Similarly, $n - k$ must be at least 1, which means $1 \leq k \leq n - 1$.

So:

$$\begin{aligned}
\mathbb{P}(X + Y = n) &= \sum_{k \in \Omega_X} \mathbb{P}(X = k)\mathbb{P}(Y = n - k) \\
&= \sum_{k=1}^{n-1} \mathbb{P}(X = k)\mathbb{P}(Y = n - k) \\
&= \sum_{k=1}^{n-1} p(1-p)^{k-1} p(1-p)^{n-k-1} \\
&= \sum_{k=1}^{n-1} p^2(1-p)^{n-2} \\
&= (n-1)p^2(1-p)^{n-2}
\end{aligned}$$

2. A new diner specializing in waffles opens on our street. It will be open 24 hours a day, seven days a week. It is assumed that the inter-arrival times between customers will be i.i.d. Exponential random variables with mean 10 minutes. Approximate the probability that the 120th customer will arrive after the first 21 hours of operation.

Solution

We will let X_k denote the waiting time between the $(k-1)^{\text{th}}$ customer and the k^{th} customer. Note that each $X_k \sim Exp(6)$ since we have a rate of 6 customers per hour to get a mean of ten minutes.

This means that $\mathbb{E}[X_k] = \frac{1}{6}$ and $Var(X_k) = \frac{1}{36}$.

Then, we will use S_{120} to be the waiting time until the 120^{th} customer. Note that:

$$S_{120} = \sum_{k=1}^{120} X_k$$

By the central limit theorem we have:

$$S_{120} \approx \mathcal{N}(120 \cdot \frac{1}{6}, 120 \cdot \frac{1}{36}) = \mathcal{N}(20, (\frac{\sqrt{120}}{6})^2)$$

We do not need to use the continuity correction, because the random variable we are approximating is continuous. So we have:

$$\begin{aligned}
\mathbb{P}(S_{120} > 21) &= \mathbb{P}(\frac{S_{120} - 20}{\frac{\sqrt{120}}{6}} > \frac{21 - 20}{\frac{\sqrt{120}}{6}}) \\
&= \mathbb{P}(Z > 0.55) \\
&= 1 - \Phi(0.55) \\
&\approx 1 - 0.71 \\
&= 0.29
\end{aligned}$$

What if we asked for an exact answer instead though? Notice that for $C \sim Gamma(120, 6)$, which describes the actual probability:

$$\mathbb{P}(C > 21) = \int_{21}^{\infty} \frac{6^{120} \cdot x^{119}}{119!} e^{-6x} dx \approx 0.285$$

3. You hold a stick of unit length (1). Someone comes along and breaks off a random piece at some point $Y \sim Unif(0,1)$. Now you hold a stick of length Y. Another person comes along and breaks off another piece from the remaining part of the stick that you hold at point $X \sim Unif(0,Y)$. You are left with a stick of length X. Find the PDF f_X , mean $\mathbb{E}[X]$ and variance $Var(X)$.

Solution

a. First, let's solve for the PDF $f_X(x)$. First notice that:

$$f_Y(y) = \begin{cases} 1, & y \in (0,1) \\ 0, & \text{otherwise} \end{cases}$$

Further:

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{1}{y}, & 0 < x < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

This means that by the law of total probability and definition of marginal distributions we have:

$$\begin{aligned} f_X(x) &= \int_{\infty}^{\infty} f_{X|Y}(x \mid y) f_Y(y) dy \\ &= \int_x^1 \frac{1}{y} dy \\ &= -\ln(x), \text{ for } x \in (0,1) \end{aligned}$$

b. To solve for the expected value of X we will use conditional expectation. First note that:

$$\mathbb{E}[X \mid Y = y] = \frac{1}{2}(0 + y) = \frac{1}{2}y$$

So:

$$\begin{aligned} \mathbb{E}[X] &= \int_0^1 \mathbb{E}[X \mid Y = y] f_Y(y) dy \\ &= \int_0^1 \frac{1}{2} y \cdot 1 dy \\ &= \frac{1}{4} \end{aligned}$$

c. We can similarly solve for the variance of X. First we note that:

$$\begin{aligned} \mathbb{E}[X^2 \mid Y = y] &= Var(X \mid Y = y) + \mathbb{E}[X \mid Y = y]^2 \\ &= \frac{1}{12}(y - 0)^2 + (\frac{1}{2}y)^2 \\ &= \frac{1}{3}y^2 \end{aligned}$$

Which allows us to calculate that:

$$\begin{aligned}\mathbb{E}[X^2] &= \int_0^1 \mathbb{E}[X^2 \mid Y = y] f_Y(y) dy \\ &= \int_0^1 \frac{1}{3} y^2 \cdot 1 dy \\ &= \frac{1}{9}\end{aligned}$$

So we finally have:

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{9} - (\frac{1}{4})^2 = \frac{7}{144}$$

4. An urn has 12 balls, 5 red ones and 7 green ones. Draw 3 balls. Let X denote the number of red balls in the sample. Compute $Var(X)$ when sampling is done:

   (a) With replacement

   (b) Without replacement

Solution

a. We start by introducing the indicator variables:

$$X_1 = \begin{cases} 1, & \text{first ball is red} \\ 0, & \text{first ball is green} \end{cases}$$

$$X_2 = \begin{cases} 1, & \text{second ball is red} \\ 0, & \text{second ball is green} \end{cases}$$

$$X_3 = \begin{cases} 1, & \text{third ball is red} \\ 0, & \text{third ball is green} \end{cases}$$

Then, note that X_1, X_2, X_3 are all $Ber(\frac{5}{12})$ and are independent.
$Var(X_1) = Var(X_2) = Var(X_3) = p(1-p) = \frac{5}{12}(1-\frac{5}{12}) = \frac{5}{12} \cdot \frac{7}{12} = \frac{35}{144}$
So:

$$\begin{aligned} Var(X) &= Var(X_1 + X_2 + X_3) \\ &= Var(X_1) + Var(X_2) + Var(X_1) \quad [\text{since } X_1, X_2, X_3 \text{ are all independent}] \\ &= \frac{35}{144} + \frac{35}{144} + \frac{35}{144} \\ &= \frac{35}{48} \end{aligned}$$

b. We can consider this as taking out 3 balls. We will define X_1 for the first ball being red, X_2 for the second ball being red, and X_3 for the third ball being red. The independence of X_1, X_2, and X_3 is no longer true. So, we will want to solve for the covariance matrix and find the sum of the entries. However, the marginal distributions of X_1, X_2, and X_3 are $Ber(\frac{5}{12})$, since each has a probability $\frac{5}{12}$ of being red.
We can prove this because there are 5 red balls and 12 total balls. We want to calculate the probability that the i^{th} ball is red, after we choose 3 balls from the urn. There are a total of $P(12,3)$ ways to order the 3 balls we picked from the urn. There are 5 ways to fix the red ball at the i^{th} position. There are $P(12-1, 3-1)$ ways to order the remaining

11 balls for the other 2 positions. This leaves us with:

$$\begin{aligned}\mathbb{P}(X_i = 1) &= \frac{5 \cdot P(11,2)}{P(12,3)} \\ &= \frac{5 \cdot \frac{11!}{9!}}{\frac{12!}{9!}} \\ &= \frac{5}{12}\end{aligned}$$

Which means that $X_i \sim Ber(\frac{5}{12})$.
We calculated the variance for this above, and we have:

$$Cov(X_i, X_i) = Var(X_i) = \frac{35}{144}$$

Now, $Cov(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]$.
Note that $X_1 \cdot X_2$ is only 1 when both are 1. We start with 5 red balls of the 12 in the first choice, and then 4 of the remaining 11 in the second choice. So:

$$\mathbb{E}[X_1 X_2] = 1 \cdot \frac{5}{12} \cdot \frac{4}{11} = \frac{5}{33}$$

Then, since X_1 is the first choice, with 5 red balls of the 12:

$$\mathbb{E}[X_1] = \frac{5}{12}$$

For the second choice we have:

$$\begin{aligned}\mathbb{E}[X_2] &= \mathbb{E}[X_2 \mid X_1 = 1]\mathbb{P}(X_1 = 1) + \mathbb{E}[X_2 \mid X_1 = 0]\mathbb{P}(X_1 = 0) \\ &= \frac{4}{11} \cdot \frac{5}{12} + \frac{5}{11} \cdot \frac{7}{12} \\ &= \frac{5}{12}\end{aligned}$$

All together this means:

$$\begin{aligned}Cov(X_1, X_2) &= \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] \\ &= \frac{5}{33} - \frac{5}{12} \cdot \frac{5}{12} \\ &= -\frac{35}{1584}\end{aligned}$$

In fact, you will find that for any i, $\mathbb{E}[X_i] = \mathbb{P}(X_i = 1) = \frac{5}{12}$, when we consider each of these variables marginally. Further, for any $i \neq j$, $\mathbb{E}[X_i X_j] = \frac{5}{33}$, we can solve for these manually considering all cases, or consider the similarity to the hat check or cat and mitten problem. We have:

$$\begin{aligned}\mathbb{E}[X_iX_j] &= \mathbb{P}(X_i = 1) \cdot \mathbb{P}(X_j = 1 \mid X_i = 1)\\ &= \frac{5}{12} \cdot \frac{4}{11}\\ &= \frac{5}{33}\end{aligned}$$

So, for any $i \neq j$, we have $Cov(X_i, X_j) = -\frac{35}{1584}$.
This gives us the following covariance matrix:

	X_1	X_2	X_3
X_1	$-\frac{35}{144}$	$-\frac{35}{1584}$	$-\frac{35}{1584}$
X_2	$-\frac{35}{1584}$	$-\frac{35}{144}$	$-\frac{35}{1584}$
X_3	$-\frac{35}{1584}$	$-\frac{35}{1584}$	$-\frac{35}{144}$

So, we have the following:

$$\begin{aligned}Cov(X, X) &= Cov(\sum_{i=1}^{3} X_i, \sum_{j=1}^{3} X_j)\\ &= \sum_{i=1}^{3}\sum_{j=1}^{3} Cov(X_i, X_j)\\ &= 3 \cdot \frac{35}{144} + 2\binom{3}{2} \cdot (-\frac{35}{1584})\\ &= \frac{105}{176}\end{aligned}$$

5. Now let's consider the general case. Suppose we have N balls in an urn, K red balls, (N-K) green balls, and we draw n times. Denote the total number of red balls to be X. What is Var(X)?

Solution
Again, we will let X_i be the indicator random variable for the i^{th} ball being a red. Note that, $X_i \sim Ber(\frac{K}{N})$. This is because we have K red balls, N total balls. We want to calculate the probability that $\mathbb{P}(X_i = 1)$. We can turn this into: Select n balls from the urn. What's the probability that the ball at position i is red? There are $P(N, n)$ ways to draw the n balls for n positions There are K ways to fix the red ball at a position i. There are $P(N-1, n-1)$ ways to draw the remaining $N-1$ balls for the remaining $n-1$ position. So we have:

$$\begin{aligned}\mathbb{P}(X_i = 1) &= \frac{K \cdot P(N-1, n-1)}{P(N, n)} \\ &= \frac{K\frac{(N-1)!}{(N-n)!}}{\frac{N!}{(N-n)!}} \\ &= \frac{K}{N}\end{aligned}$$

So we have:

$$Var(X_i) = p(1-p) = \frac{K}{N}(1 - \frac{K}{N})$$

Then for any i, we know that $\mathbb{E}[X_i] = \frac{K}{N}$, and for $i \neq j$, $\mathbb{E}[X_i X_j] = \frac{K}{N} \cdot \frac{K-1}{N-1}$ since we have K successes of N when drawing for X_i and then one less for each when drawing X_j after. So, for $i \neq j$, we have:

$$Cov(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] = \frac{K}{N} \cdot \frac{K-1}{N-1} - \frac{K^2}{N^2}$$

Overall this gives us:

$$\begin{aligned} Var(X) &= Var(\sum_{i=1}^{n} X_i) \\ &= Cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} Cov(X_i, X_j) \\ &= \sum_{i=1}^{n} Var(X_i) + 2 \sum_{i<j} Cov(X_i, X_j) \\ &= n\frac{K}{N}(1 - \frac{K}{N}) + 2\binom{n}{2}(\frac{K}{N} \cdot \frac{K-1}{N-1} - \frac{K^2}{N^2}) \\ &= n\frac{K\,(N-K)\,(N-n)}{N^2(N-1)} \end{aligned}$$