CSE 312: Foundations of Computing II

Section 6: Joint Distributions

1. Random Stick

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 $f_X(x)$, mean $\mathbb{E}[X]$ **using LTE** and variance Var(X) **using LTE as well**.

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:

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

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

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

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

$$(X|Y = y) \sim Unif(0, y)$$

which tells us that

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

Thus:

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

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

$$\mathbb{E}\left[X^2\right] = Var(X) + \mathbb{E}[X]^2$$

from the definition of variance. If we add conditioning, we get:

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

Which allows us to calculate that:

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

So we finally have:

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

2. Another Urn Question

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.

$Var(X_1) = Var(X_2) = Var(X_3) = p(1-p) = \frac{5}{12}(1 - \frac{5}{12}) = \frac{35}{144}$

So:

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

(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 ith ball is red, after we choose 3 balls from the urn. There are a total of $\mathbb{P}(12, 3)$ ways to order the 3 balls we picked from the urn. There are 5 ways to fix the red ball at the ith position. There are $\mathbb{P}(12-1, 3-1)$ ways to order the remaining 11 balls for the other 2 positions. This leaves us with:

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

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, we have $Cov(X_1, X_2) = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\,\mathbb{E}[X_2]$.

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:

$$\mathbb{E}[X_1X_2] = 1 \cdot \frac{5}{12} \cdot \frac{4}{11} = \frac{5}{33}$$

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:

$$\begin{aligned}Cov(X_1, X_2) &= \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\,\mathbb{E}[X_2] \\ &= -\frac{35}{1584}\end{aligned}$$

In fact, you will find that for any i, $\mathbb{E}[X_i] = \mathbb{P}(X_1 = 1) = \frac{5}{12}$, when we consider each of these variables marginally. Further, for any $i \neq j, \mathbb{E}[X_iX_j] = \frac{5}{33}$, we can solve for these manually considering all cases, or consider the similarity to the hat check or cat and mitten problem. We have:

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

So, for any $i \neq j, Cov(X_i, X_j) = -\frac{35}{1584}$.

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:

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

3. Continuous Joint Density

The joint probability density function of X and Y is given by

$$f_{X,Y}(x, y) = \begin{cases} \frac{6}{7}\left(x^2 + \frac{xy}{2}\right) & 0 < x < 1,\ 0 < y < 2 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Verify that this is indeed a joint density function.

(b) Compute the marginal density function of X.

(c) Find $P(Y > \frac{1}{2} | X < \frac{1}{2})$.

(d) Find $E(X)$.

(e) Find $E(Y)$

Solution:

(a) A joint density function will integrate to 1 over all possible values. Thus, we integrate over the joint range range using Wolfram Alpha, and see that:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = \int_0^2 \int_0^1 \frac{6}{7}(x^2 + \frac{xy}{2}) dx dy = 1$$

We also need to check that the density is nonnegative, but that is easily seen to be true.

(b) We apply the definition of the marginal density function of X, using the fact that we only need to integrate over the values where the joint density is positive:

$$f_X(x) = \begin{cases} \int_0^2 \frac{6}{7}(x^2 + \frac{xy}{2}) dy = \frac{6}{7}x(2x + 1) & 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

(c) By the definition of conditional probability:

$$\mathbb{P}\left(Y > \frac{1}{2} | X < \frac{1}{2}\right) = \frac{\mathbb{P}(Y > \frac{1}{2},\ X < \frac{1}{2})}{\mathbb{P}(X < \frac{1}{2})}$$

For the numerator, we have

$$\mathbb{P}(Y > \frac{1}{2}, X < \frac{1}{2}) = \int_{1/2}^{\infty} \int_{-\infty}^{1/2} f_{X,Y}(x,y)dxdy$$

$$= \int_{1/2}^{2} \int_{0}^{1/2} \frac{6}{7}\left(x^2 + \frac{xy}{2}\right) dxdy = \frac{69}{448}$$

For the denominator, we can integrate using the marginal distribution that we found before:

$$\int_{0}^{1/2} \frac{6}{7} x(2x+1)dx = \frac{5}{28}$$

Putting these together, we get:

$$\mathbb{P}(Y > \frac{1}{2} | X < \frac{1}{2}) = \frac{\frac{69}{448}}{\frac{5}{28}} = 0.8625$$

(d) By definition, and using $\Omega_X = (0, 1)$:

$$\mathbb{E}[X] = \int_0^1 f_X(x)xdx = \int_0^1 \frac{6}{7}x(2x+1)xdx = \frac{5}{7}$$

(e) By definition, and using $\Omega_Y = (0, 2)$:

$$\mathbb{E}[Y] = \int_0^2 f_Y(y)ydy = \int_0^2 \frac{1}{14}(3y+4)ydy = \frac{8}{7}$$