# STAT 302: Assignment 4

**Summer 2026. Due June 17, 11:59pm on Gradescope.**

### Question 1 [14 points]

Consider the following joint distribution between random variables $X$ and $Y$.

| (X , Y) | 1 | 2 | 3 |
|---|---|---|---|
| **0** | 0.10 | 0.15 | 0.05 |
| **1** | 0.20 | 0.30 | 0.10 |
| **2** | 0.05 | 0.03 | 0.02 |

(and of course, $\mathbb{P}(X, Y) = 0$ otherwise.)

(a) **[3 points]** What is $\mathbb{E}[X \mid Y]$?

We first need to find $\mathbb{P}(X \mid Y) = \mathbb{P}X, Y(x, y)/\mathbb{P}_Y(y)$

We can sum over the columns to get $\mathbb{P}(Y = y)$:

$$P(Y = y) = \begin{cases} 0.35 & y = 1 \\ 0.48 & y = 2 \\ 0.17 & y = 3 \end{cases}$$

(notice these probabilities add to one).

Then, we can solve for $\mathbb{P}(X = x \mid Y = y)$:

| $\mathbb{P}(X\|Y)$ | Y = 1 | Y= 2 | Y= 3 |
|---|---|---|---|
| **X = 0** | 0.10/0.35 = 2/7 | 0.15/0.48 = 5/16 | 0.05/0.17 = 5/17 |
| **X = 1** | 0.20/0.35 = 4/7 | 0.30/0.48 = 5/8 | 0.10/0.17 = 10/17 |
| **X = 2** | 0.05/0.35 = 1/7 | 0.03/0.48 = 1/16 | 0.02/0.17 = 2/17 |

$\mathbb{E}[X \mid Y=1] = \frac{2}{7}(0) + \frac{4}{7}[1] + \frac{1}{7}(2) = \frac{6}{7} = 0.857$

$\mathbb{E}[X \mid Y=2] = \frac{5}{16}(0) + \frac{5}{8}[1] + \frac{1}{16}(2) = \frac{3}{4} = 0.750$

$\mathbb{E}[X \mid Y=3] = \frac{5}{17}(0) + \frac{10}{17}[1] + \frac{2}{17}(2) = \frac{14}{17} = 0.824$

Therefore,

$$\mathbb{E}[X \mid Y=y] = \begin{cases} 0.857 & \text{when y} = 1 \\ 0.750 & \text{when y} = 2 \\ 0.824 & \text{when y} = 3 \\ \text{undefined} & \text{otherwise} \end{cases}$$

(b) **[4 points]** Calculate the correlation between $X$ and $Y$.

The correlation between $X$ and $Y$ is $\rho_{XY} = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sigma_X \sigma_Y}$

$$\begin{aligned} \mathbb{E}[X] &= \sum_x \sum_y x p_{XY}(x,y) \\ &= 0p_X(0) + 1p_X[1] + 2p_X(2) \\ &= 0 + 1(0.20 + 0.30 + 0.10) + 2(0.05 + 0.03 + 0.02) \\ &= 0.80 \end{aligned}$$

$$\begin{aligned} \mathbb{E}[Y] &= \sum_x \sum_y y p_{XY}(x,y) \\ &= 1p_Y[1] + 2p_Y(2) + 3p_Y(3) \\ &= 1(0.10 + 0.20 + 0.05) + 2(0.15 + 0.30 + 0.03) + 3(0.05 + 0.10 + 0.02) \\ &= 1.82 \end{aligned}$$

$$\begin{aligned} \mathbb{E}[XY] &= \sum_x \sum_y xy p_{XY}(x,y) \\ &= (0)[1](0.10) + (0)(2)(0.15) + (0)(3)(0.05) + [1][1](0.20) + [1](2)(0.30) + [1](3)(0.10) \\ &+ (2)[1](0.05) + (2)(2)(0.03) + (2)(3)(0.02) \\ &= 1.44 \end{aligned}$$

To calculate the standard deviation/variance, we also need:

$$
\begin{aligned}
\mathbb{E}[X^2] &= \sum_x \sum_y x^2 p_{XY}(x,y) \\
&= 0^2 p_X(0) + 1^2 p_X[1] + 2^2 p_X(2) \\
&= (0.20 + 0.30 + 0.10) + 4(0.05 + 0.03 + 0.02) \\
&= 1
\end{aligned}
$$

Thus, $\sigma_X = \sqrt{\mathbb{E}[X^2] - \mathbb{E}[X]^2} = \sqrt{1 - 0.8^2} = 0.60$.

$$
\begin{aligned}
\mathbb{E}[Y^2] &= \sum_x \sum_y y^2 p_{XY}(x,y) \\
&= 1^2 p_Y[1] + 2^2 p_Y(2) + 3^2 p_Y(3) \\
&= 1(0.10 + 0.20 + 0.05) + 4(0.15 + 0.30 + 0.03) + 9(0.05 + 0.10 + 0.02) \\
&= 3.80
\end{aligned}
$$

Thus, $\sigma_Y = \sqrt{\mathbb{E}[Y^2] - \mathbb{E}[Y]^2} = \sqrt{3.80 - 1.82^2} = 0.6983$.

Then,

$\rho_{XY} = \frac{1.44 - (0.8)(1.82)}{0.60 \times 0.6983} = -0.0382$

(c) **[3 points]** Let $Z = X + Y$. Find the MGF of $Z$. Simplify your answer as much as possible.

$$
\begin{aligned}
m_Z(t) &= \mathbb{E}[e^{tZ}] \\
&= \mathbb{E}[e^{t(X+Y)}] \\
&= \sum_x \sum_y e^{t(x+y)} p_{XY}(x,y) \\
&= e^{t(0+1)}(0.10) + e^{t(0+2)}(0.15) + e^{t(0+3)}(0.05) + e^{t(1+1)}(0.20) + \\
&\quad e^{t(1+2)}(0.30) + e^{t(1+3)}(0.10) + e^{t(2+1)}(0.05) + e^{t(2+2)}(0.03) + e^{t(2+3)}(0.02) \\
&= 0.10e^{t} + 0.15e^{2t} + 0.05e^{3t} + 0.20e^{2t} + 0.30e^{3t} + 0.10e^{4t} + 0.05e^{3t} + \\
&\quad 0.03e^{4t} + 0.02e^{5t} \\
&= 0.10e^{t} + 0.35e^{2t} + 0.40e^{3t} + 0.13e^{4t} + 0.02e^{5t}
\end{aligned}
$$

(d) **[4 points]** Use your MGF from part (c) to find the variance of $Z$. *Hint: start by finding* $\mathbb{E}(Z)$ *and* $\mathbb{E}(Z^2)$.

$$
\begin{aligned}
\mathbb{E}(Z) &= m'_Z(0) \\
&= \frac{d}{dt} 0.10e^t + 0.35e^{2t} + 0.40e^{3t} + 0.13e^{4t} + 0.02e^{5t}|_{t=0} \\
&= 0.10e^t + 0.35(2)e^{2t} + 0.40(3)e^{3t} + 0.13(4)e^{4t} + 0.02(5)e^{5t}|_{t=0} \\
&= 0.10e^(0) + 0.35(2)e^{2(0)} + 0.40(3)e^{3(0)} + 0.13(4)e^{4(0)} + 0.02(5)e^{5(0)} \\
&= 0.10 + 2(0.35) + 0.40(3) + 0.13(4) + 0.02(5) \\
&= 2.62
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}(Z^2) &= m''_Z(0) \\
&= \frac{d^2}{dt^2} 0.10e^t + 0.35e^{2t} + 0.40e^{3t} + 0.13e^{4t} + 0.02e^{5t}|_{t=0} \\
&= \frac{d}{dt} 0.10e^t + 0.35(2)e^{2t} + 0.40(3)e^{3t} + 0.13(4)e^{4t} + 0.02(5)e^{5t}|_{t=0} \\
&= 0.10e^t + 0.35(2)(2)e^{2t} + 0.40(3)(3)e^{3t} + 0.13(4)(4)e^{4t} + 0.02(5)(5)e^{5t}|_{t=0} \\
&= 0.10e^0 + 0.35(2)(2)e^{2(0)} + 0.40(3)(3)e^{3(0)} + 0.13(4)(4)e^{4(0)} + 0.02(5)(5)e^{5(0)} \\
&= 0.10 + 0.35(2)(2) + 0.40(3)(3) + 0.13(4)(4) + 0.02(5)(5) \\
&= 7.68
\end{aligned}
$$

$$
\begin{aligned}
\text{Var}[Z] &= \mathbb{E}[Z^2] - \mathbb{E}[Z] \\
&= 7.68 - 2.62^2 \\
&= 0.8156
\end{aligned}
$$

## Question 2 [8 points]

Let $X$ and $Y$ be continuous random variables with joint pdf $f_{X,Y}(x,y) = 3xI_{[0,1]}(x)I_{[0,x]}(y)$.

(a) **[4 points]** Compute the covariance between $X$ and $Y$.

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^1 \int_0^x (x)3xdydx \\
&= \int_0^1 \int_0^x 3x^2dydx \\
&= \int_0^1 \left[3x^2y\Big|_{y=0}^{y=x}\right]dx \\
&= \int_0^1 3x^2xdx \\
&= \int_0^1 3x^3dx \\
&= \frac{3}{4}x^4\Big|_{x=0}^{x=1} \\
&= \frac{3}{4}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Y] &= \int_0^1 \int_0^x (y)3xdydx \\
&= \int_0^1 \int_0^x 3xydydx \\
&= \int_0^1 \left[\frac{3}{2}xy^2\right]\Big|_{y=0}^{y=x}dx \\
&= \int_0^1 \frac{3}{2}x^3dx \\
&= \frac{3}{2\times 4}x^4\Big|_{x=0}^{x=1} \\
&= \frac{3}{8}
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}[XY] &= \int_0^1 \int_0^x (xy)3x dy dx \\
&= \int_0^1 \int_0^x 3x^2 y dy dx \\
&= \int_0^1 \left[\frac{3}{2}x^2y^2\right]\Big|_{y=0}^{y=x} dx \\
&= \int_0^1 \frac{3}{2}x^4 dx \\
&= \frac{3}{2\times 5}x^5\Big|_{x=0}^{x=1} \\
&= \frac{3}{10}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\text{Cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= \frac{3}{10} - \frac{3}{8}\frac{3}{4} \\
&= \frac{3}{160} \text{ or } 0.01875
\end{aligned}
$$

(b) **[4 points]** Find the marginal MGF of $X$, defined as $m_X(t) = \mathbb{E}[e^{tx}]$. Show all of your work.

$$
\begin{aligned}
m_X(t) &= \mathbb{E}[e^{tX}] \\
&= \int_0^1 \int_0^x (e^{tx})3x dy dx \\
&= \int_0^1 \left[3xe^{tx}y\Big|_{y=0}^{y=x}\right] dx \\
&= \int_0^1 3x^2 e^{tx} dx \\
&\dots \\
&= 3(t^2x^2 - 2tx + 2)\frac{e^{tx}}{t^3} - \frac{6}{t^3}\Big|_{x=0}^{x=1} \text{ (by integration by parts)} \\
&= \frac{3(t^2 - 2t + 2)e^t - 6}{t^3}
\end{aligned}
$$

or in other forms:

$\frac{3(t^2-2t+2)e^t-6}{t^3}$,

or

$\frac{3t^2e^t-6te^t+6e^t-6}{t^3}$

or

$\frac{3e^t}{t} - \frac{6e^t}{t^2} + \frac{6(e^t-1)}{t^3}$.

## Question 3 [13 points]

Let $\Theta \sim \text{Unif}(0, 1)$ and let $X \mid \Theta \sim \text{Binom}(4, \Theta)$.

(a) **[3 points]** What is $\mathbb{E}[X]$? You may refer to the general result of the expected values of any distribution that is found at the end of this assignment.

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[\mathbb{E}[X \mid \Theta]] \\
&= \mathbb{E}[4\Theta] \\
&= 4\mathbb{E}[\Theta] \\
&= 4 * 0.5 \\
&= 2
\end{aligned}$$

(b) **[4 points]** What is $\text{Var}[X]$? You may refer to the general result of the expected values of any distribution that is found at the end of this assignment.

$$\begin{aligned}
\text{Var}[X] &= \mathbb{E}[\text{Var}[X \mid Y]] + \text{Var}[\mathbb{E}[X \mid Y]] \\
&= \mathbb{E}[4(\Theta)(1-\Theta)] + \text{Var}[4\Theta] \\
&= 4\mathbb{E}[\Theta - \Theta^2] + 4^2\,\text{Var}[\Theta] \\
&= 4\mathbb{E}[\Theta] - 4\mathbb{E}[\Theta^2] + 4^2\,\text{Var}[\Theta] \\
&= 4[0.5] - 4\mathbb{E}[\Theta^2] + 4^2(1-0)^2/12 \\
&= 2 - 4\mathbb{E}[\Theta^2] + 4/3
\end{aligned}$$

Now, we can actually solve for $\mathbb{E}[\Theta^2]$ using $\text{Var}[\Theta] = \mathbb{E}[\Theta^2] - \mathbb{E}[\Theta]^2$. This implies $\mathbb{E}[\Theta^2] = \text{Var}[\Theta] + \mathbb{E}[\Theta]^2 = 1/12 + (1/2)^2 = 1/3$

Thus,

$$\begin{aligned}\text{Var}[X] &= \mathbb{E}[\text{Var}[X \mid Y]] + \text{Var}[\mathbb{E}[X \mid Y]] \\ &= 2 - 4(1/3) + 4/3 \\ &= 2\end{aligned}$$

(c) **[3 points]** Let $g(\Theta) = \mathbb{E}[X \mid \Theta]$ and consider the transformed variable $Y = e^{g(\Theta)}$. Establish a lower bound on $\mathbb{E}[Y]$. Show all of your work and the inequality you use. *Hint: $h(x) = e^x$ is a convex function.*

Recall Jensen's inequality that states $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$. Therefore:

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}[e^{g(\Theta)}] \\ &\geq e^{\mathbb{E}[g(\Theta)]} \\ &= e^{\mathbb{E}[\mathbb{E}[X \mid \Theta]]} \\ &= e^{\mathbb{E}[X]} \\ &= e^2 \\ &= 7.389\end{aligned}$$

Therefore, $\mathbb{E}[Y] \geq 7.389$.

(d) **[3 marks]** Find an upper bound on $P(|X - 2| \geq 3)$. Show your work and name the inequality you use.

By Chebyshev's inequality, since $\mu = E[X] = 2$ and $\text{Var}(X) = 2$ we have

$$\begin{aligned}P(|X - 2| \geq 3) &\leq \frac{2}{3^2} \\ &= 2/9 \\ &= 0.2222\end{aligned}$$