STAT 302: Assignment 3

Summer 2026. Due June 10, 11:59pm on Gradescope.

Question 1 [9 points]

In any given 15 minute block of time, the R4 shows up with the "SORRY, BUS FULL" sign 5 times. Assume the arrival times of full buses are independent. Let X be the time in minutes between full buses.

(a) **[2 points]** Write out the CDF of X. Be sure to include the support.

X is exponentially distributed, with rate parameter $5/15 = 1/3$.

$X \sim \text{Exp}(1/3)$

The CDF is $F_X(x) = 1 - e^{-\frac{1}{3}x} I_{[0,\infty)}$

(b) **[3 points]** What is the variance of the time between full buses? Show the full calculation.

First solve for $\mathbb{E}[X]$

$$
\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty x f_X(x) dx \\
&= \int_0^\infty x \frac{1}{3} e^{-1/3x} dx \\
&= \int_0^\infty \frac{1}{3} x e^{-1/3x} dx
\end{aligned}
$$

Let $u = 1/3x \rightarrow du = 1/3dx$. Let $dv = e^{1/3x}dx \rightarrow v = -3e^{-1/3x}$ Then by integration by parts....

$$
\begin{aligned}
\mathbb{E}[X] &= \frac{1}{3}x(-3)e^{-1/3x} - \int_0^\infty (-3)e^{-1/3x}\frac{1}{3}dx \\
&= -xe^{-1/3x}|_{x=0}^{x=\infty} + \int_0^\infty e^{-1/3x}dx \\
&= -xe^{-1/3x}|_{x=0}^{x=\infty} - 3e^{-1/3x}|_{x=0}^{x=\infty} \\
&= (0-0) - (0-3) \\
&= 3
\end{aligned}
$$

Now solve for $\mathbb{E}[X^2]$

$$
\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty x^2 f_X(x)dx \\
&= \int_0^\infty x^2 \frac{1}{3}e^{-1/3x}dx \\
&= \int_0^\infty \frac{1}{3}x^2 e^{-1/3x}dx
\end{aligned}
$$

Let $u = 1/3x^2 \to du = 1/3xdx$. Let $dv = e^{1/3x}dx \to v = -3e^{-1/3x}$ Then by integration by parts....

$$
\begin{aligned}
\mathbb{E}[X^2] &= \frac{1}{3}x^2(-3)e^{-1/3x} - \int_0^\infty (-3)e^{-1/3x}\frac{2}{3}xdx \\
&= -x^2e^{-1/3x}|_{x=0}^{x=\infty} + \int_0^\infty e^{-1/3x}dx \\
&= -x^2e^{-1/3x}|_{x=0}^{x=\infty} + 6\int_0^\infty \frac{1}{3}e^{-1/3x}dx \\
&= (0-0) - 6(\mathbb{E}[X]) \\
&= 6(3) = 18
\end{aligned}
$$

Then:

$$
\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 18 - 3^2 = 9
$$

Aside: here are some general results for the exponential distribution:

If $X \sim \mathrm{Exp}(\lambda)$, $\mathbb{E}[X] = \frac{1}{\lambda}$, and $\mathrm{Var}[X] = \frac{1}{\lambda^2}$

(c) [**4 points**] What is the probability that the total time for 2 consecutive full buses to arrive is less than 12 minutes?

We can find this by finding the joint PDF of X_1 and X_2, which is

$$\begin{aligned} f_(X_1, X_2)(x_1, x_2) &= \frac{1}{3}e^{-\frac{1}{3}x_1}\frac{1}{3}e^{-\frac{1}{3}x_2} I_{[0,\infty)}(x_1) I_{[0,\infty)}(x_2) \\ &= \frac{1}{9}e^{-\frac{1}{3}(x_1+x_2)} I_{[0,\infty)}(x_1) I_{[0,\infty)}(x_2) \end{aligned}$$

Now, we are interested in $\mathbb{P}(X_1 + X_2 < 12)$. We can think of this as $\mathbb{P}(X_1 < 12 - X_2)$ (or $\mathbb{P}(X_2 < 12 - X_1)$. WLOG we will show the solution using $\mathbb{P}(X_1 < 12 - X_2)$.

We can integrate this, but we do need to be careful about the bounds of the integral. If $X_2 > 12$, this then $12 - X_2 < 0$, and our support is only for $X_1 \geq 0$. Therefore, this part of the integral will just be 0. If $0 \leq X_2 \leq 12$, then $12 - X_2$ will be between 0 and 1, which is allowed for our support of X_1. So, we really only care about the case where $0 \leq X_2 \leq 12$.

$$\begin{aligned} \mathbb{P}(X_1 + X_2 < 12) &= \mathbb{P}(X_1 < 12 - X_2) \\ &= \int_0^{12} \int_0^{12-x_2} \frac{1}{9} e^{-\frac{1}{3}(x_1+x_2)} dx_1 dx_2 \\ &= \frac{1}{9} \int_0^{12} e^{-\frac{1}{3}x_2} \left[\int_0^{12-x_2} e^{-\frac{1}{3}x_1} dx_1 \right] dx_2 \\ &= \frac{1}{9} \int_0^{12} e^{-\frac{1}{3}x_2} \left[(-3) e^{-\frac{1}{3}x_1} \Big|_{x_1=(0)}^{x_1=12-x_2} \right] dx_2 \\ &= \frac{1}{3} \int_0^{12} e^{-\frac{1}{3}x_2} \left(1 - e^{-\frac{1}{3}(12-x_2)} \right) dx_2 \\ &= \frac{1}{3} \int_0^{12} \left(e^{-\frac{1}{3}x_2} - e^{-\frac{1}{3}(12)} \right) dx_2 \\ &= \frac{1}{3} \left((-3) e^{-\frac{1}{3}x_2} - e^{-\frac{1}{3}(12)} x_2 \right) \Big|_{x_2=0}^{x_2=12} \\ &= \left(-e^{-\frac{1}{3}x_2} - \frac{1}{3} e^{-\frac{1}{3}(12)} x_2 \right) \Big|_{x_2=0}^{x_2=12} \\ &= \left(-e^{-\frac{1}{3}(12)} - \frac{1}{3} e^{-\frac{1}{3}(12)}(12) \right) - \left(-e^{-\frac{1}{3}(0)} - \frac{1}{3} e^{-\frac{1}{3}(12)}(0) \right) \\ &= 1 - e^{-4} - 4e^{-4} \\ &= 1 - 5e^{-4} \\ &= \approx 0.9084 \end{aligned}$$

You could also look at this as a Poisson distribution. With $\lambda = \frac{1}{3}$ buses/min, the expected number of full buses in 12 minutes is: $\mu = \lambda \cdot t = \frac{1}{3} \times 12 = 4$. So we can define Y as the number of busses full within 12 minutes, where $Y \sim Poisson(4)$. Then,

$P(Y \geq 1) = 1 - \mathbb{P}(Y = 0) - \mathbb{P}(Y = 1) = 1 - \frac{e^{-4} \cdot 4^0}{0!} - \frac{e^{-4} \cdot 4^1}{1!} = 0.9084$.

Question 2 [15 points]

The provincial avalanche centre monitors two adjacent weather stations in the Cascade Mountains. Let X be the snow depth (cm) at the lower station and Y be the snow depth (cm) at the upper station on a randomly selected day in January. Their joint PDF is modelled as:

$$f_{X,Y}(x, y) = \frac{1}{3000}(x + 2y), \qquad 0 \leq x \leq 10, 0 \leq y \leq 15$$

(a) **[2 points]** On any given day, what is the probability that the snow depth at the lower station is greater than 5cm and the snow depth of the upper station is less than 9cm?

$$\begin{aligned}
P(X > 5,\ Y < 9) &= \frac{1}{3000} \int_5^{10} \int_0^9 \frac{1}{3000}(x + 2y)\ dy\ dx \\
&= \frac{1}{3000} \int_5^{10} \left[xy + y^2\right]_0^9\ dx \\
&= \frac{1}{3000} \int_5^{10} (9x + 81)\ dx \\
&= \frac{1}{3000} \left[\frac{9x^2}{2} + 81x\right]_5^{10} \\
&= \frac{1}{3000} \left[(450 + 810) - \left(\frac{225}{2} + 405\right)\right] \\
&= \frac{1}{3000} \cdot \frac{1485}{2} \\
&= 0.2475
\end{aligned}$$

(b) **[2 points]** Calculate the marginal pdf $f_X(x)$.

$$f_X(x) = \int_0^{15} \frac{1}{3000}(x + 2y)\ dy = \frac{1}{3000}\left[xy + y^2\right]_0^{15} = \frac{1}{3000}(15x + 225) I_{[0,10](x)}$$

(c) **[2 points]** Derive the joint CDF for $0 \leq x \leq 10, 0 \leq y \leq 15$

$$
\begin{aligned}
F_{X,Y}(x,y) &= \int_0^x \int_0^y \frac{1}{3000}(s+2t)\ dt\ ds \\
&= \frac{1}{3000}\int_0^x \left[st + t^2\right]_0^y\ ds \\
&= \frac{1}{3000}\int_0^x \left[sy + y^2\right]\ ds \\
&= \frac{1}{3000}\left[\frac{s^2y}{2} + sy^2\right]_0^x \\
&= \frac{x^2y + 2xy^2}{6000} I_{[0,10]}(x) I_{[0,15]}(y)
\end{aligned}
$$

(d) **[2 points]** On any given day, what is the probability that the snow depth at the lower station is more than 3cm greater than snow depth at the upper station?

We need the region $x > y+3$, while also considering the support $0 \leq x \leq 10$, $0 \leq y \leq 15$. Since $x \leq 10$.

We can consider $y < x - 3 \leq 7$, so y ranges from 0 to 7 and x from $y+3$ to 10.

$$
\begin{aligned}
P(X > Y+3) &= \int_0^7 \int_{y+3}^{10} \frac{1}{3000}(x+2y)\ dx\ dy \\
&= \frac{1}{3000}\int_0^7 \left[\frac{x^2}{2} + 2xy\right]_{y+3}^{10}\ dy \\
&= \frac{1}{3000}\int_0^7 \left[(50+20y) - \left(\frac{(y+3)^2}{2} + 2y(y+3)\right)\right]\ dy \\
&= \frac{1}{3000}\int_0^7 \left[\frac{91}{2} + 11y - \frac{5y^2}{2}\right]\ dy \\
&= \frac{1}{3000}\left[\frac{91y}{2} + \frac{11y^2}{2} - \frac{5y^3}{6}\right]_0^7 \\
&= \frac{1}{3000}\cdot\frac{1813}{6} \\
&= 0.1007
\end{aligned}
$$

You can also compute $P(Y < X - 3)$.

We need to be super careful about the integration region. For $y \geq 0$, we need $x - 3 > 0$, so $x > 3$. Thus x runs from 3 to 10, and for each x, y runs from 0 to $x - 3$:

$$
\begin{aligned}
P(X > Y + 3) &= \int_3^{10} \int_0^{x-3} \frac{1}{3000}(x + 2y)\ dy\ dx \\
&= \frac{1}{3000} \int_3^{10} \left[xy + y^2\right]_0^{x-3}\ dx \\
&= \frac{1}{3000} \int_3^{10} (x(x-3) + (x-3)^2)\ dx \\
&= \frac{1}{3000} \int_3^{10} (x^2 - 3x + x^2 - 6x + 9)\ dx \\
&= \frac{1}{3000} \int_3^{10} (2x^2 - 9x + 9)\ dx \\
&= \frac{1}{3000} \left[(\frac{2}{3}x^3 - \frac{9}{2}x^2 + 9x)\right]_{x=3}^{x=10} \\
&= \frac{1}{3000} \left[(\frac{2}{3}(10)^3 - \frac{9}{2}(10^2) + 9(10)) - (\frac{2}{3}(3)^3 - \frac{9}{2}(3^2) + 9(3))\right] \\
&= 0.1007
\end{aligned}
$$

(e) **[3 points]** What is the conditional distribution $f_{X|Y(x|y)}$?

First we can derive the marginal $f_Y(y)$.

$$
\begin{aligned}
f_Y(y) &= \int_0^{10} \frac{1}{3000}(x + 2y)\ dx \\
&= \frac{1}{3000} \left[\frac{x^2}{2} + 2xy\right]_0^{10} \\
&= \frac{50 + 20y}{3000} \\
&= \frac{5 + 2y}{300} I_{[0,15]}(y)
\end{aligned}
$$

Then,

$$
f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} I_{[0,10]}(x) = \frac{\frac{x+2y}{3000}}{\frac{5+2y}{300}} I_{[0,10]}(x) = \frac{x + 2y}{10(5 + 2y)} I_{[0,10]}(x)
$$

(which technically is defined for $0 \le y \le 15$, but it's okay if you don't say that explicitly here! Adding an indicator for y is fine too.)

(f) [**4 points**] Are X and Y independent? Discuss why or why not, providing mathematical reasoning.

$$f_X(x) \cdot f_Y(y) = \frac{x+15}{200} \cdot \frac{5+2y}{300} = \frac{(x+15)(2y+5)}{60000} \neq f_{X,Y}(x,y)$$

Therefore they are not independent.

(You can also show for a point (x,y) within the support that this equality fails.)

Question 3 [11 points]

The joint PMF of discrete random variables X and Y is given by:

$X \backslash Y$	$Y=0$	$Y=1$	$P(X=x)$
$X=0$	0.30	0.05	0.35
$X=1$	0.20	0.25	0.45
$X=2$	0.10	0.10	0.20
$P(Y=y)$	0.60	0.40	1.00

(a) [**2 points**] Compute the conditional distribution $\mathbb{P}_{Y|X}(y \mid X=1)$ for all $y \in \mathbb{R}$.

(a) By definition, $P_{Y \mid X}(y \mid X=x) = \dfrac{P(X=x, Y=y)}{P(X=x)}$. For $X=1$, we have $P(X=1) = 0.45$:

$$P_{Y \mid X}(0 \mid X=1) = \frac{P(X=1, Y=0)}{P(X=1)} = \frac{0.20}{0.45} = \frac{4}{9}$$

$$P_{Y \mid X}(1 \mid X=1) = \frac{P(X=1, Y=1)}{P(X=1)} = \frac{0.25}{0.45} = \frac{5}{9}$$

Check: $\dfrac{4}{9} + \dfrac{5}{9} = 1$.

(b) [**3 points**] Calculate $\mathbb{P}(Y=1|X \geq 1)$. Show your work.

(b) We need $\mathbb{P}(Y=1 \mid X \geq 1)$. By the definition of conditional probability,

$$\mathbb{P}(Y = 1 \mid X \geq 1) = \frac{\mathbb{P}(Y = 1 \text{ and } X \geq 1)}{\mathbb{P}(X \geq 1)}$$

For the numerator, use the joint table: $X \geq 1$ and $Y = 1$ means $(X = 1, Y = 1)$ or $(X = 2, Y = 1)$, so

$$\mathbb{P}(Y = 1 \text{ and } X \geq 1) = P(X = 1, Y = 1) + P(X = 2, Y = 1) = 0.25 + 0.10 = 0.35$$

For the denominator: $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - 0.35 = 0.65$. Therefore,

$$\mathbb{P}(Y = 1 \mid X \geq 1) = \frac{0.35}{0.65} = \frac{7}{13} \approx 0.5385$$

(c) **[3 points]** Compute the conditional distribution $P_{X|Y}(x \mid Y = 0)$ for all $x \in \mathbb{R}$.

By definition, $P_{X \mid Y}(x \mid Y = 0) = \dfrac{P(X = x, Y = 0)}{P(Y = 0)}$. For $Y = 0$, we have $P(Y = 0) = 0.60$:

$$P_{X \mid Y}(0 \mid Y = 0) = \frac{P(X = 0, Y = 0)}{P(Y = 0)} = \frac{0.30}{0.60} = \frac{1}{2}$$

$$P_{X \mid Y}(1 \mid Y = 0) = \frac{P(X = 1, Y = 0)}{P(Y = 0)} = \frac{0.20}{0.60} = \frac{1}{3}$$

$$P_{X \mid Y}(2 \mid Y = 0) = \frac{P(X = 2, Y = 0)}{P(Y = 0)} = \frac{0.10}{0.60} = \frac{1}{6}$$

Check: $\dfrac{1}{2} + \dfrac{1}{3} + \dfrac{1}{6} = 1$.

(d) **[2 points]** Are X and Y independent? Justify your answer mathematically ~~by checking at least two conditions from the definition of independence~~.

Two equivalent definitions of independence:

1. Checking by factorization: X and Y are independent if $P(X = x, Y = y) = P(X = x)\, P(Y = y)$ for all (x, y) in the support. We can thus check each cell:

$$P(0,0) = 0.30 \stackrel{?}{=} P(X=0)\,P(Y=0) = (0.35)(0.60) = 0.21 \quad \text{(fails)}$$
$$P(0,1) = 0.05 \stackrel{?}{=} P(X=0)\,P(Y=1) = (0.35)(0.40) = 0.14 \quad \text{(fails)}$$
$$P(1,0) = 0.20 \stackrel{?}{=} P(X=1)\,P(Y=0) = (0.45)(0.60) = 0.27 \quad \text{(fails)}$$
$$P(1,1) = 0.25 \stackrel{?}{=} P(X=1)\,P(Y=1) = (0.45)(0.40) = 0.18 \quad \text{(fails)}$$
$$P(2,0) = 0.10 \stackrel{?}{=} P(X=2)\,P(Y=0) = (0.20)(0.60) = 0.12 \quad \text{(fails)}$$
$$P(2,1) = 0.10 \stackrel{?}{=} P(X=2)\,P(Y=1) = (0.20)(0.40) = 0.08 \quad \text{(fails!}$$

Therefore, X and Y are **not independent**. (A single counterexample is sufficient to conclude X and Y are not independent).

2. Checking if conditional distribution = marginal distribution: X and Y are independent if $P_{X\mid Y}(x \mid y) = P(X=x)$ for all (x,y). From part (c):

$$P_{X\mid Y}(0 \mid Y=0) = \frac{1}{2} = 0.50 \neq P(X=0) = 0.35$$

(A single counterexample is sufficient to conclude X and Y are not independent).

(e) [**3 point**] Compute $\mathbb{E}(XY)$.

By definition of expectation for discrete random variables,

$$\mathbb{E}(XY) = \sum_x \sum_y xy\,P(X=x, Y=y).$$

Going cell by cell through the joint table:

$$\begin{aligned}
\mathbb{E}(XY) &= 0\cdot 0\cdot P(0,0) + 0\cdot 1\cdot P(0,1) + 1\cdot 0\cdot P(1,0) + 1\cdot 1\cdot P(1,1) + 2\cdot 0\cdot P(2,0) + 2\cdot 1\cdot P(2,1) \\
&= 0 + 0 + 0 + 1(0.25) + 0 + 2(0.10) \\
&= 0.25 + 0.20 \\
&= 0.45
\end{aligned}$$