SOME PSTAT 120A-STYLE REVIEW PROBLEMS

PSTAT 120B: Mathematical Statistics, I
Summer Session A, 2024 with Instructor: Ethan P. Marzban

1. Wait times (in minutes) at *Cajé* are uniformly distributed between 5 minutes and 15 minutes. Suppose a customer is selected at random from *Cajé*, and their wait time is recorded.

   (a) What is the probability that the selected customer waits fewer than 7 minutes?

   Solution: Let X denote the wait time of the customer; then $X \sim \text{Unif}[5, 15]$, meaning X has density

   $$f_X(x) = \frac{1}{15 - 5} \cdot \mathbb{1}_{\{x \in [5,15]\}} = \frac{1}{10} \cdot \mathbb{1}_{\{x \in [5,15]\}}$$

   Therefore,

   $$\mathbb{P}(X < 7) = \int_{-\infty}^{7} f_X(x) \, \mathrm{d}x = \int_{5}^{7} \frac{1}{10} \, \mathrm{d}x = \frac{2}{10} = \boxed{\frac{1}{5} = 20\%}$$

   (b) What is the expected amount of time this customer should expect to wait in line?

   Solution: Recall that the expectation of the Unif$[a, b]$ distribution is just $(a + b)/2$. Hence, plugging in $a = 5$ and $b = 15$ we find

   $$\mathbb{E}[X] = \frac{5 + 15}{2} = \boxed{10 \text{ mins}}$$

   (c) Given that the customer has already waited for 7 minutes but hasn't been served yet, what is the probability that they end up waiting for more than 10 minutes?

   Solution: We seek $\mathbb{P}(X > 10 \mid X > 7)$. By the definition of conditional probability,

   $$\mathbb{P}(X > 10 \mid X > 7) = \frac{\mathbb{P}(X > 10 \, , \, X > 7)}{\mathbb{P}(X > 7)}$$

   Note that $\{X > 10\} \subseteq \{X > 7\}$; that is, if the customer has waited for more than 10 minutes they most certainly have waited for more than 7 minutes. Hence,

   $$\mathbb{P}(X > 10 \, , \, X > 7) = \mathbb{P}(X > 10)$$

   and so

   $$\mathbb{P}(X > 10 \mid X > 7) = \frac{\mathbb{P}(X > 10)}{\mathbb{P}(X > 7)} = \frac{\int_{10}^{15} \frac{1}{10} \, \mathrm{d}x}{\int_{7}^{15} \frac{1}{10} \, \mathrm{d}x} = \frac{\left(\frac{5}{10}\right)}{\left(\frac{8}{10}\right)} = \boxed{\frac{5}{8} = 62.5\%}$$

2. Luna the Golden Retriever has buried a bone somewhere in the backyard. Unfortunately, she can't quite remember where she buried it! As such, she keepds digging holes in the hopes of finding her

bone - once she finds her bone, she stops digging holes. Suppose each hole Luna digs has a 25% chance of containing her bone, independently of all other holes.

(a) If X denotes the total number of holes Luna digs (including the final successful hole) before stopping, what distribution does X follow? Include both a distribution name as well as any/all relevant parameter(s).

> **Solution:** Since X counts the total number of trials before the first success (where "success" in this problem is digging a hole that contains the bone), we know X will follow a Geometric distribution. Specifically, because X is counting the *total* number of trials we know that X will gollow a Geometric distribution on $\{1, 2, \cdots\}$. The parameter of the Geometric distribution is the probability of success, which in this problem is stated to be $p = 0.25$. Hence,
>
> $$X \sim \text{Geom}(0.25) \text{ on } \{1, 2, \cdots\}$$

(b) What is the probability that Luna has to dig 20 or more holes before finding her bone?

> **Solution:** By our work to part (a), we have that the PMF of X is given by
>
> $$p_X(x) = (1 - 0.25)^{x-1}(0.25) = (0.25) \cdot (0.75)^{x-1}$$
>
> Hence,
>
> $$\begin{aligned} \mathbb{P}(X \geq 20) &= \sum_{x=20}^{\infty} p_X(x) = \sum_{x=20}^{\infty} (0.25) \cdot (0.75)^{x-1} = \frac{0.25}{0.75} \cdot \sum_{x=20}^{\infty} (0.75)^x \\ &= \frac{1}{3} \cdot \frac{\left(\frac{3}{4}\right)^{20}}{\left(\frac{1}{4}\right)} = \frac{1}{3} \cdot \frac{3^{20}}{4^{20}} \cdot 4 = \left(\frac{3}{4}\right)^{19} \approx 0.4228\% \end{aligned}$$
>
> where we have utilized the formula for an **infinite geometric series**:
>
> $$\sum_{k=a}^{\infty} r^k = \frac{r^a}{1 - r}, \qquad \text{provided } |r| < 1$$

(c) What is the probability that Luna has to dig 19 or fewer holes before finding her bone?

> **Solution:** Note that since X only admits positive integers in its support,
>
> $$\mathbb{P}(X \leq 19) = 1 - \mathbb{P}(X \geq 20) = 1 - \left(\frac{3}{4}\right)^{19} \approx 99.5772\%$$

3. Let X be a random variable, and let $a, b \in \mathbb{R}$ be deterministic constants. Use first principles to prove that expectations are **linear**; that is, $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$. As a hint: consider the discrete and continuous cases separately, and start each case by applying the LOTUS and leveraging linearity of sums and expectations. (You may assume that X is not mixed.)

Solution: First we consider the discrete case: let X have probability mass function (p.m.f.) given by $p_X(k)$, so that we have

$$\begin{aligned}\mathbb{E}[aX+b] &= \sum_k (ak+b)\cdot p_X(k) \\ &= a\underbrace{\left(\sum_k k\cdot p_X(k)\right)}_{:=\mathbb{E}[X]} + b\underbrace{\left(\sum_k p_X(k)\right)}_{=1} = a\mathbb{E}[X]+b\end{aligned}$$

Similarly, if X is continuous with probability density function (p.d.f.) given by $f_X(x)$, then

$$\begin{aligned}\mathbb{E}[aX+b] &= \int_{-\infty}^{\infty} (ax+b)\cdot f_X(x)\ \mathrm{d}x \\ &= a\underbrace{\left(\int_{-\infty}^{\infty} x f_X(x)\ \mathrm{d}x\right)}_{:=\mathbb{E}[X]} + b\underbrace{\left(\int_{-\infty}^{\infty} f_X(x)\ \mathrm{d}x\right)}_{=1} = a\mathbb{E}[X]+b\end{aligned}$$

4. The Celestial Toymaker[1] has decided to play a game with me. On a table, he lines up an infinite number of boxes (he *is* the god of games, after all). With probability $(1/2)^i$ he selects box number i [where $i = 1, 2, 3, \cdots$]. Inside box number i there are 3^i marbles, one of which is red and the remainder of which are blue. So, for example, box 1 is selected with probability (1/2), and contains 1 red marble and 2 blue marbles; box 2 is selected with probability (1/4), and contains 1 red marble and 8 blue marbles, etc. The Toymaker selects a box, and then draws a marble.

   (a) What is the probability that the Toymaker selects a red marble?

Solution: Let B_i denote the event that box i was chosen, and let R denote the event that a red marble was chosen. From the problem statement, we have that

$$\mathbb{P}(B_i) = \left(\frac{1}{2}\right)^i; \qquad \mathbb{P}(R \mid B_i) = \frac{1}{3^i}$$

We seek $\mathbb{P}(R)$; using the Law of Total Probability, we compute this as

$$\begin{aligned}\mathbb{P}(R) &= \sum_{i=1}^{\infty} \mathbb{P}(R \mid B_i)\cdot\mathbb{P}(B_i) \\ &= \sum_{i=1}^{\infty}\left(\frac{1}{3^i}\right)\cdot\left(\frac{1}{2}\right)^i = \sum_{i=1}^{\infty}\left(\frac{1}{6}\right)^i \\ &= \frac{\left(\frac{1}{6}\right)}{1-\frac{1}{6}} = \frac{1}{6}\cdot\frac{6}{5} = \boxed{\frac{1}{5}}\end{aligned}$$

[1] If you're curious, this is a character from the British Sci-Fi teleision show *Doctor Who*.

(b) Given that the Toymaker selected a red marble, what is the problem that he drew from box 4?

Solution: Using our notation from part (a), we seek $\mathbb{P}(B_4 \mid R)$. By Bayes' Rule, we compute this as

$$\begin{aligned}\mathbb{P}(B_4 \mid R) &= \frac{\mathbb{P}(R \mid B_4) \cdot \mathbb{P}(B_4)}{\mathbb{P}(R)} \\ &= \frac{\left(\frac{1}{3^4}\right) \cdot \left(\frac{1}{2^4}\right)}{\left(\frac{1}{5}\right)} = \frac{5}{1296} \approx 3.856 \times 10^{-3}\end{aligned}$$

5. Consider a sequence $\{X_i\}_{i=1}^n$ of i.i.d. random variables with common mean μ and common variance σ^2. Define

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$$

to be the sample mean. Compute Corr$(X_1, \overline{X}_n)$, the correlation between X_1 and the sample mean. **Hint:** Bilinearity.

Solution: By the definition of correlation,

$$\text{Corr}(X_1, \overline{X}_n) = \frac{\text{Cov}(X_1, \overline{X}_n)}{\text{SD}(X_1)\text{SD}(\overline{X}_n)}$$

The denominator is relatively simple to compute: we know SD$(X_1) = \sigma$, and from a previously-derived result (from PSTAT 120A) we know SD$(\overline{X}_n) = \sigma/\sqrt{n}$. As such, let's focus on the numerator:

$$\text{Cov}(X_1, \overline{X}_n) = \text{Cov}\left(X_1 \, , \, \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \frac{1}{n} \text{Cov}(X_1, X_i)$$

Since we are assuming the X_i's are i.i.d., we know that Cov$(X_1, X_i) = 0$ whenever $i \neq 1$. Furthermore, Cov$(X_1, X_1) =$ Var$(X_1) = \sigma^2$; hence,

$$\text{Cov}(X_1, \overline{X}_n) = \frac{1}{n} \sum_{i=1}^{n} \sigma^2 \mathbb{1}_{\{i=1\}} = \frac{\sigma^2}{n}$$

Therefore, putting everything together,

$$\text{Corr}(X_1, \overline{X}_n) = \frac{\text{Cov}(X_1, \overline{X}_n)}{\text{SD}(X_1)\text{SD}(\overline{X}_n)} = \frac{\left(\frac{\not{\sigma}^2}{n}\right)}{(\not{\sigma}) \cdot \left(\frac{\not{\sigma}}{\sqrt{n}}\right)} = \frac{\sqrt{n}}{n} = \sqrt{\frac{1}{n}}$$

6. Let $(X, Y) \sim f_{X,Y}$. Prove that

$$\mathbb{E}[X] = \iint_{\mathbb{R}^2} x f_{X,Y}(x, y)\, \mathsf{d}A$$

Hint: Iterate the integral on the RHS, leverage the relationship between marginal densities and joint densities, and finally recognize the definition of $\mathbb{E}[X]$.

Solution:

$$\begin{aligned}\iint_{\mathbb{R}^2} x f_{X,Y}(x, y)\, \mathsf{d}A &= \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X,Y}(x, y)\, \mathsf{d}y\, \mathsf{d}x \\ &= \int_{\mathbb{R}} x \underbrace{\left(\int_{\mathbb{R}} f_{X,Y}(x, y)\, \mathsf{d}y \right)}_{= f_X(x)} \mathsf{d}x = \int_{\mathbb{R}} x f_X(x)\, \mathsf{d}x =: \mathbb{E}[X]\end{aligned}$$

7. Let $(X, Y) \sim f_{X,Y}$ where

$$f_{X,Y}(x, y) = k(1 - y) \cdot \mathbb{1}_{\{0 \leq x \leq y \leq 1\}}$$

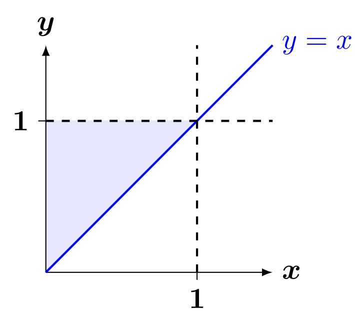
(a) Find the value of k that ensures this is a valid joint density function.

Solution: Nonnegativity is fairly trivial; for every $(x, y) \in \{(x, y) \in \mathbb{R} : 0 \leq x \leq y \leq 1\}$ we know that $y \in [0, 1]$ and so $(1 - y) \geq 0$. So, all we need for nonnegativity to hold is for k to be positive.

To find the specific value of k, we recall that $f_{X,Y}(x, y)$ must integrate to unity. Therefore, we start off by computing

$$\iint_{\mathbb{R}^2} (1 - y) \cdot \mathbb{1}_{\{0 \leq x \leq y \leq 1\}}\, \mathsf{d}A$$

Let's sketch the support:



This allows us to compute

$$\begin{aligned}\iint_{\mathbb{R}^2} k(1 - y) \cdot \mathbb{1}_{\{0 \leq x \leq y \leq 1\}}\, \mathsf{d}A &= \int_0^1 \int_0^y (1 - y)\, \mathsf{d}x\, \mathsf{d}y \\ &= \int_0^1 y(1 - y)\, \mathsf{d}y = \int_0^1 (y - y^2)\, \mathsf{d}y = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}\end{aligned}$$

Hence, in order for the density to integrate to unity, we should take $k = 6$.

(b) Compute $\mathbb{P}(X \leq {}^3\!/\!_4\,,\ Y \geq {}^1\!/\!_2)$.

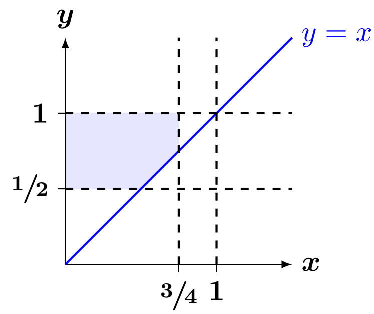
Solution: We find the desired probability by computing

$$\mathbb{P}(X \leq {}^3\!/\!_4\,,\ Y \geq {}^1\!/\!_2) = \iint_{\mathcal{R}} 6(1-y)\ \mathrm{d}A$$

where

$$\mathcal{R} := \{(x,y) \in \mathbb{R}^2 : x \leq {}^3\!/\!_4\,,\ y \geq {}^1\!/\!_2\,,\ 0 \leq x \leq y \leq 1\}$$

Sketching this region yields:



Whichever order of integration we pick, we need to split the integral into two. As such, let's (somewhat arbitrarily) use the order dx dy:

$$\mathbb{P}(X \leq {}^3\!/\!_4\,,\ Y \geq {}^1\!/\!_2) = \int_{1/2}^{3/4} \int_0^y 6(1-y)\ \mathrm{d}x\ \mathrm{d}y + \int_{3/4}^{1} \int_0^{3/4} 6(1-y)\ \mathrm{d}x\ \mathrm{d}y$$

$$\int_{1/2}^{3/4} \int_0^y 6(1-y)\ \mathrm{d}x\ \mathrm{d}y = 6\int_{1/2}^{3/4} (y-y^2)\ \mathrm{d}y = 6\left[\frac{1}{2}\left(\frac{9}{16}-\frac{1}{4}\right) - \frac{1}{3}\left(\frac{27}{64}-\frac{1}{8}\right)\right] = \frac{11}{32}$$

$$\int_{3/4}^{1} \int_0^{3/4} 6(1-y)\ \mathrm{d}x\ \mathrm{d}y = 6 \cdot \frac{3}{4} \cdot \int_{3/4}^{1} (1-y)\ \mathrm{d}y = \frac{9}{2} \cdot \left[\frac{1}{4} - \frac{1}{2}\left(1 - \frac{9}{16}\right)\right] = \frac{9}{64}$$

$$\mathbb{P}(X \leq {}^3\!/\!_4\,,\ Y \geq {}^1\!/\!_2) = \frac{11}{32} + \frac{9}{64} = \boxed{\frac{31}{64}}$$

8. Let $X \sim \text{Unif}[a,b]$.

(a) Show that X has MGF given by

$$M_X(t) = \begin{cases} \frac{e^{tb}-e^{ta}}{t(b-a)} & \text{if } t \neq 0 \\ 1 & \text{if } t = 0 \end{cases}$$

Be careful with the cases you consider!

Solution: We know that X has density given by

$$f_X(x) = \frac{1}{b-a} \cdot \mathbb{1}_{\{x \in [a,b]\}}$$

Thus, by the definition of MGFs (coupled with the LOTUS),

$$M_X(t) := \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\ \mathrm{d}x$$

First note: if $t = 0$, we have

$$M_X(0) = \int_{-\infty}^{\infty} e^0 f_X(x)\,\mathsf{d}x = \int_{-\infty}^{\infty} f_X(x)\,\mathsf{d}x = 1$$

If $t \neq 0$, then

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,\mathsf{d}x = \int_a^b e^{tx} \cdot \frac{1}{b-a}\,\mathsf{d}x = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

Hence, putting everything together,

$$M_X(t) = \begin{cases} \frac{e^{tb}-e^{ta}}{t(b-a)} & \text{if } t \neq 0 \\ 1 & \text{if } t = 0 \end{cases}$$

(b) Is the above MGF continuous at $t = 0$? (Recall that this question is important as a lot of our MGF-related results assume continuity in an interval containing $t = 0$!)

Solution: The question really boils down to whether $\lim_{t \to 0} M_X(t) = 1$ or not. First note that plugging in $t = 0$ to the formula

$$\frac{e^{tb} - e^{ta}}{t(b-a)}$$

yields an indeterminate form of $0/0$. As such, we should apply L'Hospital's rule:

$$\lim_{t \to 0} \frac{e^{tb} - e^{ta}}{t(b-a)} = \lim_{t \to 0} \frac{\frac{\mathsf{d}}{\mathsf{d}t}[e^{tb} - e^{ta}]}{\frac{\mathsf{d}}{\mathsf{d}t}[t(b-a)]} = \lim_{t \to 0} \frac{be^{tb} - ae^{ta}}{b-a} = \frac{b-a}{b-a} = 1$$

Hence, we have

$$\lim_{t \to 0} M_X(t) = 1 = M_X(1)$$

meaning the MGF *is* continuous at $t = 0$.

(c) Derive a simple closed-form expression for

$$\left.\frac{\mathsf{d}}{\mathsf{d}t^n}\right|_{t=0} \left[\frac{e^{tb} - e^{ta}}{t}\right]$$

where the notation $\frac{\mathsf{d}}{\mathsf{d}t^n}|_{t=0}$ means "the n^{th} derivative, evaluated at $t = 0$."

Solution: Though we could try and "brute-force" this by taking derivatives and hoping to see a pattern, let's see if we can be a bit more clever. Since we have established continuity of the MGF at $t = 0$, we can now invoke the fact that

$$M_X^{(n)}(0) = \mathbb{E}[X^n]$$

This means

$$\left.\frac{\mathsf{d}}{\mathsf{d}t^n}\right|_{t=0} \left[\frac{e^{tb} - e^{ta}}{t(b-a)}\right] = \mathbb{E}[X^n]$$

By the LOTUS, we can obtain another formula for $\mathbb{E}[X^n]$:

$$\mathbb{E}[X^n] = \int_a^b x^n \frac{1}{b-a} \, \mathrm{d}x = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}$$

Therefore, we have

$$\left. \frac{\mathrm{d}}{\mathrm{d}t^n} \right|_{t=0} \left[\frac{e^{tb} - e^{ta}}{t(b-a)} \right] = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}$$

Multiplying both sides by $(b-a)$ yields

$$\left. \frac{\mathrm{d}}{\mathrm{d}t^n} \right|_{t=0} \left[\frac{e^{tb} - e^{ta}}{t} \right] = \frac{b^{n+1} - a^{n+1}}{n+1}$$