Total Points: 25

PSTAT 120B / **MIDTERM 1** / Summer 2024 Instructor: **Ethan P. Marzban**

Name: ______________________________ **NetID:** ____________________
(*First and Last*) (*NOT Perm Number*)

Your Section: **2pm (Hyuk-Jean)** **3pm (Hyuk-Jean)** **4pm (Minwoo)** **5pm (Minwoo)**
(*Circle One*)

Instructions:

- You will have **55 minutes** to complete this exam.
  - Nobody will be permitted to leave the exam room during the last 10 minutes of the exam.
- Please remember to write your name and NetID (not perm number) at the top of each sheet of this exam.
- You are allowed the use of a single **8.5 $\times$ 11-inch** sheet, front and back, of handwritten notes. You are also permitted the use of **calculators**; the use of any and all other electronic devices (laptops, cell phones, etc.) is prohibited.
  - You will be asked to turn in your note sheet with your exam.
- Unless otherwise specified, simplification is not needed; however, all integrals and infinite sums (unless otherwise specified) must be evaluated.
  - One exception is that, whenever applicable, answers may be left in terms of Φ, the standard normal c.d.f..
- **Good Luck!!!**

Honor Code: In signing my name below, I certify that all work appearing on this exam is entirely my own and not copied from any external source. I further certify that I have not received any unauthorized aid while taking this exam.

$\times$ ______________________________

1. Tasha and Tony have both gotten in line at the *Gaucho Boutique*, and are waiting to be served. Let Y_1 denote the amount of time (in minutes) Tasha waits in line, and let Y_2 denote the amount of time (in minutes) Tony waits in line. It is found that the joint density of (Y_1, Y_2) is given by

$$f_{Y_1,Y_2}(y_1, y_2) = e^{-y_1} \cdot \mathbb{1}_{\{0 \leq y_2 \leq y_1 < \infty\}}$$

(a) (4 points) Find $f_{Y_2}(y_2)$, the marginal density of Y_2, and use this to identify the distribution of Y_2 by name. Be sure to also include any/all relevant parameter(s)!

Solution: We must first find $f_{Y_2}(y_2)$, the marginal density of Y_2. To do so, we integrate the joint wrt. y_1:

$$\begin{aligned} f_{Y_2}(y_2) &= \int_{\mathbb{R}} f_{Y_1,Y_2}(y_1, y_2) \, \mathrm{d}y_1 \\ &= \int_{\mathbb{R}} e^{-y_1} \cdot \mathbb{1}_{\{y_1 \geq y_2\}} \cdot \mathbb{1}_{\{y_2 \geq 0\}} \, \mathrm{d}y_1 \\ &= \mathbb{1}_{\{y_2 \geq 0\}} \cdot \int_{y_2}^{\infty} e^{-y_1} \, \mathrm{d}y_1 = e^{-y_2} \cdot \mathbb{1}_{\{y_2 \geq 0\}} \end{aligned}$$

which allows us to conclude $\boxed{Y_2 \sim \text{Exp}(1)}$.

(b) (6 points) Given that Tasha ends up waiting for *exactly* 3 minutes, what is the probability that Tony ends up waiting for more than 2 minutes? You may use, without proof, the fact that $Y_1 \sim \text{Gamma}(2, 1)$.

Solution: The quantity we seek is $\mathbb{P}(Y_2 > 2 \mid Y_1 = 3)$. Since the event we are conditioning on has zero probability, we cannot use the definition of conditional probability but instead must integrate the conditional density. Hence, we should first find $f_{Y_2|Y_1}(y_2 \mid y_1)$, which we can do by appealing to the definition of conditional densities:

$$f_{Y_2|Y_1}(y_2 \mid y_1) := \frac{f_{Y_1,Y_2}(y_1, y_2)}{f_{Y_1}(y_1)}$$

We're told that $Y_1 \sim \text{Gamma}(2, 1)$ meaning

$$f_{Y_1}(y_1) = \frac{1}{\Gamma(2) \cdot 1^2} \cdot y_1^{2-1} \cdot e^{-y_1/1} \cdot \mathbb{1}_{\{y_1 \geq 0\}} = y_1 e^{-y_1} \cdot \mathbb{1}_{\{y_1 \geq 0\}}$$

and so

$$\begin{aligned} f_{Y_2|Y_1}(y_2 \mid y_1) &:= \frac{f_{Y_1,Y_2}(y_1, y_2)}{f_{Y_2}(y_2)} \\ &= \frac{\cancel{e^{-y_1}} \cdot \mathbb{1}_{\{0 \leq y_2 \leq y_1\}} \cdot \cancel{\mathbb{1}_{\{y_1 \geq 0\}}}}{y_1 \cancel{e^{-y_1}} \cdot \cancel{\mathbb{1}_{\{y_1 \geq 0\}}}} = \frac{1}{y_1} \cdot \mathbb{1}_{\{0 \leq y_2 \leq y_1\}} \end{aligned}$$

Hence, plugging in $y_2 = 3$ we have

$$f_{Y_2|Y_1}(y_2 \mid 3) = \frac{1}{3} \cdot \mathbb{1}_{\{0 \leq y_1 \leq 3\}}$$

and so

$$\begin{aligned}\mathbb{P}(Y_2 > 2 \mid Y_1 = 3) &= \int_2^\infty f_{Y_2 \mid Y_1}(y_2 \mid 3) \, \mathrm{d}y_1 \\ &= \int_2^\infty \frac{1}{3} \cdot \mathbb{1}_{\{0 \leq y_1 \leq 3\}} \, \mathrm{d}y_1 = \frac{1}{3} \cdot \int_2^3 \, \mathrm{d}y_1 = \boxed{\frac{1}{3}}\end{aligned}$$

2. A particular pesticide is composed of two compounds, called Compound A and Compound B. Suppose that the proportions Y_1 and Y_2 of compounds A and B, respectively, in a particular random sample of pesticide is given by

$$2 \cdot \mathbb{1}_{\{0 \leq y_1 \leq 1 \ , \ 0 \leq y_2 \leq 1 \ , \ 0 \leq y_1 + y_2 \leq 1\}}$$

(a) (1 point) One can show (and you do <u>not</u> need to prove this) that

$$f_{Y_1 | Y_2}(y_1 \mid y_2) = \frac{1}{1 - y_2} \cdot \mathbb{1}_{\{0 \leq y_1 \leq 1 - y_2\}}$$

Crucially, though, this density is missing a specification of what values of y_2 it is defined over. For what values of y_2 is the conditional density defined? Justify your answer.

Solution: We know that $f_{Y_1|Y_2}(y_1 \mid y_2)$ is defined only over values of y_2 for which $f_{Y_2}(y_2) \neq 0$. Upon inspection, we see that the support of Y_2 [i.e. the only y_2 values over which the density is nonzero] is $S_{Y_2} = [0, 1]$ meaning the density above is defined only when $y_2 \in [0, 1]$.

(b) (2 points) Given that a sample of pesticide contains 70% Compound B, what is the expected percentage of Compound A contained in the sample? (You may still use the conditional density provided in the statement of part (a) without proof.)

Solution: We are asked to compute $\mathbb{E}[Y_1 \mid Y_2 = 0.7]$ which can be computed using the definition of conditional expectation:

$$\begin{aligned}
\mathbb{E}[Y_1 \mid Y_2 = 0.7] &= \int_{\mathbb{R}} y_1 f_{Y_1|Y_2}(y_1 \mid 0.7) \, \mathrm{d}y_1 \\
&= \int_{\mathbb{R}} y_1 \cdot \frac{1}{1 - 0.7} \cdot \mathbb{1}_{\{0 \leq y_1 \leq 1 - 0.7\}} \, \mathrm{d}y_1 \\
&= \frac{1}{0.3} \cdot \int_0^{0.3} y_1 \, \mathrm{d}y_1 = \frac{1}{0.3} \cdot \frac{1}{2}(0.3)^2 = \frac{0.3}{2} = 0.15 = 15\%
\end{aligned}$$

Alternatively, we could have found a more general expression for $\mathbb{E}[Y_1 \mid Y_2 = y_2]$ and *then* plugged in $y_2 = 0.7$; this will also give you the correct answer.

(c) (4 points) Use the Law of Iterated Expectations to compute $\mathbb{E}[Y_1]$. You may use (without proof) the fact that $\mathbb{E}[Y_2] = 1/3$, along with the conditional density provided in part (a). Please note that if you simply double-integrate the joint density, you will not receive full points.

Solution: The Law of Iterated Expectations tells us

$$\mathbb{E}[Y_1] = \mathbb{E}[\mathbb{E}[Y_1 \mid Y_2]]$$

To compute $\mathbb{E}[Y_1 \mid Y_2]$, we utilize our two-step procedure: first find $\mathbb{E}[Y_1 \mid Y_2 = y_2]$, and then plug in Y_2 in place of y_2. To find $\mathbb{E}[Y_1 \mid Y_2 = y_2]$, we compute

$$\begin{aligned}\mathbb{E}[Y_1 \mid Y_2 = y_2] &:= \int_{\mathbb{R}} y_1 f_{Y_1|Y_2}(y_1 \mid y_2)\,\mathrm{d}y_1 \\ &= \int_{\mathbb{R}} y_1 \cdot \frac{1}{1-y_2} \cdot \mathbb{1}_{\{0 \leq y_1 \leq 1-y_2\}}\,\mathrm{d}y_1 \\ &= \frac{1}{1-y_2} \cdot \int_0^{1-y_2} y_1\,\mathrm{d}y_1 = \frac{1}{2} \cdot \frac{1}{1-y_2} \cdot (1-y_2)^2 = \frac{1-y_2}{2}\end{aligned}$$

This implies

$$\mathbb{E}[Y_1 \mid Y_2] = \frac{1-Y_2}{2}$$

and so, by the Law of Iterated Expectations,

$$\mathbb{E}[Y_1] = \mathbb{E}[\mathbb{E}[Y_1 \mid Y_2]] = \mathbb{E}\left[\frac{1-Y_2}{2}\right] = \frac{1}{2} - \frac{1}{2} \cdot \mathbb{E}[Y_2] = \frac{1}{2} - \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{3}$$

3. A random variable Y is said to follow the **Pareto Distribution**, notated $Y \sim \text{Pareto}(\theta, \alpha)$ for parameters $\theta > 0$ and $\alpha > 0$, if Y has density given by

$$f_Y(y) = \frac{\alpha\theta^\alpha}{y^{\alpha+1}} \cdot \mathbb{1}_{\{y \geq \theta\}}$$

Let $Y \sim \text{Pareto}(\theta, \alpha)$.

(a) (3 points) Define $U_1 := cY$ for a positive constant c. Derive the density $f_{U_1}(u)$ of U_1 using the Change of Variable Formula (aka the method of transformations). Be sure to include the support of U_1 as well.

Solution: Our transformation is $g(y) = cy$, meaning our inverse transformation is $g^{-1}(u) = u/c$. Hence,

$$\left|\frac{\mathrm{d}}{\mathrm{d}u} g^{-1}(u)\right| = \left|\frac{\mathrm{d}}{\mathrm{d}u}\left[\frac{u}{c}\right]\right| = \left|\frac{1}{c}\right| = \frac{1}{c}$$

where we have dropped the absolute values in the last step since c is assumed to be positive. Therefore, plugging into the Change of Variable formula:

$$f_U(u) = f_Y[g^{-1}(u)] \cdot \left|\frac{\mathrm{d}}{\mathrm{d}u} g^{-1}(u)\right|$$

$$
\begin{aligned}
&= \frac{\alpha\theta^{\alpha}}{(u/c)^{\alpha+1}} \cdot \mathbb{1}_{\{(u/c) \geq \theta\}} \cdot \frac{1}{c} \\
&= \frac{\alpha\theta^{\alpha}}{u^{\alpha+1}} \cdot \frac{c^{\alpha \cancel{+1}}}{\cancel{c}} \cdot \mathbb{1}_{\{u \geq c\theta\}} \\
&= \boxed{\frac{\alpha(c\theta)^{\alpha}}{u^{\alpha+1}} \cdot \mathbb{1}_{\{u \geq c\theta\}}}
\end{aligned}
$$

The simplification wasn't strictly necessary for this part; I mainly simplified in order to make the answer to part (b) more obvious.

(b) (1 point) Does U_1 (as defined in part (a) above) follow the Pareto distribution? If so, identify the parameters.

Solution: Since we derived the density of U_1 in part (a), we simply need to see if it of the form of a Pareto density function (which is provided in the problem statement). Matching terms, we can see that U_1 *does* follow a Pareto distribution; specifically, $U_1 \sim \text{Pareto}(c\theta, \alpha)$.

(c) (4 points) Define $U_2 := \sqrt{Y}$. Derive the density $f_{U_2}(u)$ of U_2 using any of the methods discussed in lecture

Solution: We can use either the CDF method or the Change of Variable formula; for this particular problem, the Change of variable formula will be a bit more direct (since we would need to derive an expression for the CDF of the Pareto distribution in order to use the CDF method). Hence, set $g(y) = \sqrt{y}$ so that $g^{-1}(u) = u^2$, and

$$\left| \frac{\mathrm{d}}{\mathrm{d}u} g^{-1}(u) \right| = \left| \frac{\mathrm{d}}{\mathrm{d}u} \left[u^2 \right] \right| = |2u| = 2u$$

where we have dropped the absolute values in the last step since U_2 admits only positive values in its support. Therefore, plugging into the Change of Variable formula:

$$\begin{aligned} f_U(u) &= f_Y[g^{-1}(u)] \cdot \left| \frac{\mathrm{d}}{\mathrm{d}u} g^{-1}(u) \right| \\ &= \frac{\alpha\theta^\alpha}{(u^2)^{\alpha+1}} \cdot \mathbb{1}_{\{(u^2) \geq \theta\}} \cdot 2u \\ &= \frac{\alpha\theta^\alpha}{u^{2\alpha+2}} \cdot 2u \cdot \mathbb{1}_{\{u^2 \geq \theta\}} \\ &= \frac{2\alpha\theta^\alpha}{u^{2\alpha+1}} \cdot \mathbb{1}_{\{u \geq \sqrt{\theta}\}} \end{aligned}$$

By the way, though this wasn't required for this problem, we can see that $U_2 \sim \text{Pareto}(\sqrt{\theta}, 2\alpha)$.

You may use the remainder of this page for scratch work; please note that nothing written on this page will be graded.