**Total Points:** 25

PSTAT 120B / **MIDTERM 2** / 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 $\Phi$, 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. (3 points) Let $Y_1, \cdots Y_{10} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 9)$ [i.e. Var$(Y_i) = 9$ for all $i = 1, 2, \cdots, 10$]. Define $U$ as

$$U := \sum_{i=1}^{10} Y_i^2$$

Identify the distribution of $U$ by name, taking care to include any/all relevant parameter(s).

**Solution:** We know that

$$V := \sum_{i=1}^{10} \left( \frac{Y_i}{3} \right)^2 = \frac{1}{9} \sum_{i=1}^{10} Y_i^2 \sim \chi^2_{10} \sim \text{Gamma}\left( \frac{10}{2}\, ,\, 2 \right) \sim \text{Gamma}(5, 2)$$

Additionally, $U := 9V$. Hence, since if $Y \sim$ Gamma$(\alpha, \beta)$ we have $(cY) \sim$ Gamma$(\alpha, c\beta)$ for any positive constant $c$, we can conclude that

$$U \sim \text{Gamma}(5, 2 \cdot 9) \sim \text{Gamma}(5, 18)$$

2. (4 points) The amount (in miles) Carl drives in a given day varies from day to day. Assume the amount he drives in one day is independent of the amount he drives in any other day. From prior experience, Carl knows that the true average amount he drives daily is 10 miles, with a standard deviation of 4 miles.

   Suppose Carl selects a random 25-day period, over which he has enough gas to travel 260 miles. **What is the probability that Carl doesn't run out of gas during this period (i.e. that the total amount of miles he drives in this randomly-selected 25-day period does not exceed the 260 miles worth of gas he has)?**

   Leave your answer in terms of $\Phi(\cdot)$, the standard normal CDF. If you use a theorem (e.g. Central Limit Theorem, Weak Law of Large Numbers, etc.), clearly state where in your work you use it.

**Solution:** Let $Y_i$ denote the amount Carl drives in any given day. We don't know what distribution $Y_i$ follows, but we do know $\mathbb{E}[Y_i] = 10$ and SD$(Y) = 4$.

Define $S_{25} = (\sum_{i=1}^{25} Y_i)$; the probability we seek is $\mathbb{P}(S_{25} \leq 260)$. By the CLT,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \rightsquigarrow \mathcal{N}(0,1)$$

Hence,

$$\frac{S_{25} - (25)(10)}{4\sqrt{25}} = \frac{S_{36} - 250}{20} \overset{\text{d}}{\approx} \mathcal{N}(0,1)$$

and so

$$\begin{aligned}\mathbb{P}(S_{25} \leq 260) &= \mathbb{P}\left(\frac{S_{25} - 250}{20} \leq \frac{260 - 250}{20}\right) \\ &= \mathbb{P}\left(\frac{S_{25} - 250}{20} \leq \frac{10}{20}\right) = \mathbb{P}\left(\frac{S_{25} - 250}{20} \leq \frac{1}{2}\right) \\ &\approx \Phi\left(\frac{1}{2}\right)\end{aligned}$$

Though this was not required for the exam (since computers were not allowed), you could use a computer software to find this probability to be around 69.15%.

3. Let $Y_1, Y_2 \overset{\text{i.i.d.}}{\sim} \text{Unif}[0,1]$, and define $U := \ln(Y_2/Y_1)$.

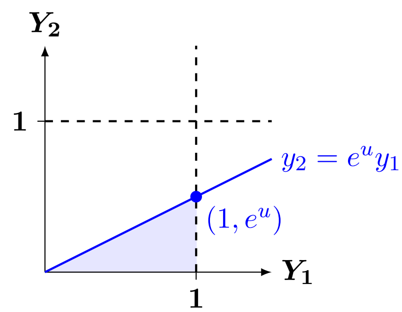
(a) (3 points) For $u < 0$, show that $F_U(U)$, the CDF (cumulative distribution function) of $U$ at $u$, is given by

$$F_U(u) = \frac{1}{2}e^u$$

**Solution:** By definition,

$$F_U(u) = \mathbb{P}(U \leq u) = \mathbb{P}\left(\ln\left(\frac{Y_2}{Y_1}\right) \leq u\right) = \mathbb{P}(Y_2 < e^u Y_1)$$

Since we are assuming $u < 0$, the curve prescribed by $y_2 = e^u y_1$ will be a line with positive slope less than 1: hence, the desired probability will be the double integral of the joint over




Since the joint density $f_{Y_1,Y_2}(y_1, y_2)$ is simply 1 over the unit square, the desired probability is simply the area of the blue shaded region, meaning

$$F_U(u) = \frac{1}{2}(1)(e^u) = \boxed{\frac{1}{2}e^u}$$

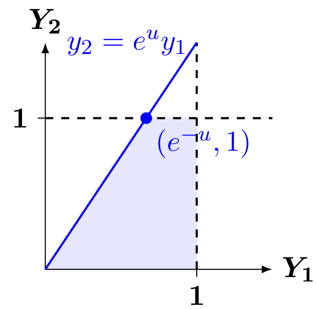
(b) (3 points) For $u \geq 0$, show that $F_U(u)$, the CDF (cumulative distribution function) of $U$ at $u$, is given by

$$F_U(u) = 1 - \frac{1}{2}e^{-u}$$

**Solution:** Similar to part (a),

$$F_U(u) = \mathbb{P}(U \leq u) = \mathbb{P}\left(\ln\left(\frac{Y_2}{Y_1}\right) \leq u\right) = \mathbb{P}(Y_2 < e^u Y_1)$$

Now we are assuming $u < 0$, meaning the curve prescribed by $y_2 = e^u y_1$ will be a line with positive slope *greater than* 1: hence, the desired probability will be the double integral of the joint over




Since the joint density $f_{Y_1,Y_2}(y_1, y_2)$ is simply 1 over the unit square, the desired probability is simply the area of the blue shaded region. This is easier to compute as 1 minus the area of the unshaded triangle, meaning

$$F_U(u) = 1 - \frac{1}{2}(1)(e^{-u}) = 1 - \frac{1}{2}e^{-u}$$

(c) (1 point) Is it true that the density of $U$ can be expressed as:

$$f_U(u) = \frac{1}{2} e^{-|u|}$$

Justify your answer!

**Solution:** Combining the results of parts (a) and (b), we have

$$F_U(u) = \begin{cases} \frac{1}{2} e^u & \text{if } u < 0 \\ 1 - \frac{1}{2} e^{-u} & \text{if } u \geq 0 \end{cases}$$

which, when differentiated, yields

$$\boxed{f_U(u) = \begin{cases} \frac{1}{2} e^u & \text{if } u < 0 \\ \frac{1}{2} e^{-u} & \text{if } u \geq 0 \end{cases} = \frac{1}{2} e^{-|u|}}$$

Hence, the answer is yes; this is the correct density for $U$.

4. Let $Y_1, \cdots, Y_n \sim$ Ray$(\theta)$. Recall that this is the notation used to denote the **Rayleigh distribution**, with density function given by

$$f_Y(y) = \left( \frac{2y}{\theta} \right) e^{-y^2/\theta} \cdot \mathbb{1}_{\{y \geq 0\}}$$

You may use, without proof, the fact that the expectation of the Ray$(\theta)$ distribution is given by

$$\mathbb{E}[Y_i] = \frac{\sqrt{\pi \theta}}{2}$$

(a) (6 points) Find $f_{Y_{(1)}}(y)$, the density of the first order statistic (i.e. the sample minimum). **Note:** This should show that $Y_{(1)} \sim$ Ray$(\theta/n)$, a fact you may use in all future parts of this problem.

**Solution:** We know that

$$f_{Y_{(1)}}(y) = n[\overline{F}(y)]^{n-1} f_Y(y)$$

Hence, we should find an expression for the survival function of $Y$. A simple $u-$substitution reveals that, for a $y > 0$,

$$\overline{F}(y) = \int_y^\infty \left( \frac{2y}{\theta} \right) e^{-y^2/\theta} \, \mathrm{d}t = e^{-y^2/\theta}$$

Therefore,

$$\begin{aligned} f_{Y_{(1)}}(y) &= n[\overline{F}(y)]^{n-1} f_Y(y) \\ &= n \left( e^{-y^2/\theta} \right)^{n-1} \cdot \left( \frac{2y}{\theta} \right) e^{-y^2/\theta} \cdot \mathbb{1}_{\{y \geq 0\}} \\ &= \boxed{\left( \frac{2ny}{\theta} \right) e^{-ny^2/\theta} \cdot \mathbb{1}_{\{y \geq 0\}}} \end{aligned}$$

Matching terms with the density provided in the problem statement, we immediately see that

$$\boxed{Y_{(1)} \sim \text{Ray}(\theta/n)}$$

(b) (3 points) Is $Y_{(1)}$ an unbiased estimator for $\theta$? Justify your answer.

**Solution:** Since $Y_{(1)} \sim \text{Ray}(\theta/n)$, we can use the formula for the expectation of a Rayleigh-distributed random variable (provided in the problem statement) to compute

$$\mathbb{E}[Y_{(1)}] = \frac{\sqrt{\pi(\theta/n)}}{2} = \frac{1}{2}\sqrt{\frac{\pi\theta}{n}}$$

This is not equal to $\theta$, and therefore $Y_{(1)}$ is a biased estimator for $\theta$.

(c) (2 points) Consider the following estimator for $\theta$:

$$\widehat{\theta}_n = \frac{4}{\pi}(\overline{Y}_n)^2$$

Is this a consistent estimator for $\theta$? Justify your answer carefully.

**Solution:**

- By the WLLN,

$$\overline{Y}_n \xrightarrow{p} \mathbb{E}[Y_1] = \frac{\sqrt{\pi\theta}}{2}$$

- By the CMT (Continuous Mapping Theorem) with $g(x) = (4/\pi)x^2$, we have

$$\frac{4}{\pi}(\overline{Y}_n)^2 \xrightarrow{p} \frac{4}{\pi}\left(\frac{\sqrt{\pi\theta}}{2}\right)^2 = \theta$$

Therefore, since $\widehat{\theta}_n \xrightarrow{p} \theta$, we have that $\widehat{\theta}_n$ is a consistent estimator for $\theta$.

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

## DO NOT REMOVE THIS PAGE FROM YOUR EXAM

**Name:**________________________________________ **NetID:**____________________

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

## DO NOT REMOVE THIS PAGE FROM YOUR EXAM