HOMEWORK 03

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

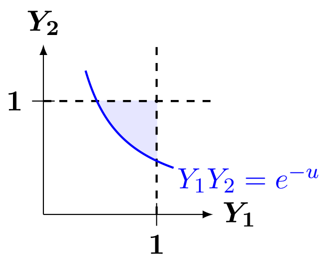
1. (**Modified from #6.95**) Let $Y_1, Y_2 \overset{\text{i.i.d.}}{\sim} \text{Unif}[0, 1]$.

   (a) Find the density of $U_1 := -\ln(Y_1 Y_2)$. Use this to recognize the distribution of U_1 by name, including any/all relevant parameter(s)!

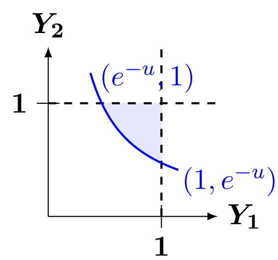
Solution: First note that since $S_{Y_1} = [0, 1]$ and $S_{Y_2} = [0, 1]$, we have that the support of $Y_1 Y_2$ is also $[0, 1]$. The negative logarithm function maps the unit interval to the interval $[0, \infty)$; hence $S_U = [0, \infty)$. Now, let us proceed using the CDF method:

$$F_{U_1}(u) = \mathbb{P}(U_1 \leq u) = \mathbb{P}(-\ln(Y_1 Y_2) \leq u) = \mathbb{P}(Y_1 Y_2 \geq e^{-u})$$

Let's sketch a picture, noting that the level curves of the function $g(y_1 y_2) = y_1 y_2$ are hyperbolas.



We should find the coordinates of the points of intersection between the hyperbola and the lines $\{Y_1 = 1\}$ and $\{Y_2 = 1\}$. The upper point of intersection satisfies $y_1 y_2 = e^{-u}$ and $y_2 = 1$; hence its y_1 coordinate is e^{-u}. Similarly, the lower point of intersection satisfies $y_1 y_2 = e^{-u}$ and $y_1 = 1$; hence its y_2 coordinate is e^{-u}. Hence:



Now we are ready to integrate! Either order of integration will be fine: let's (somewhat arbitrarily) use $\mathrm{d}y_2 \, \mathrm{d}y_1$.

$$\begin{aligned} F_U(u) &= \mathbb{P}(Y_1 Y_2 \geq e^{-u}) \\ &= \int_{e^{-u}}^{1} \int_{e^{-u}/y_1}^{1} f_{Y_1, Y_2}(y_1, y_2) \, \mathrm{d}y_2 \, \mathrm{d}y_1 \\ &= \int_{e^{-u}}^{1} \int_{e^{-u}/y_1}^{1} (1) \, \mathrm{d}y_2 \, \mathrm{d}y_1 \end{aligned}$$

$$= \int_{e^{-u}}^{1} \left(1 - \frac{1}{y_1} e^{-u}\right) \, \mathrm{d}y_1$$

$$= \left[y_1 - \ln(y_1) e^{-u}\right]_{y_1 = e^{-u}}^{y_1 = 1} = 1 - e^{-u} + u e^{-u} = 1 - (u+1) e^{-u}$$

Taking the derivative wrt. u and reincorporating the support yields

$$f_{U_1}(u_1) = \left[-e^{-u} + (u+1)e^{-u}\right] \cdot \mathbb{1}_{\{u \geq 0\}} = u e^{-u} \cdot \mathbb{1}_{\{u \geq 0\}}$$

This shows that $U_1 \sim \text{Gamma}(2, 1)$.

(b) Find the density of $U_2 := Y_1 Y_2$

Solution: There are a couple of ways to go about this problem. One is to use the CDF method directly. Instead, however, I'll demonstrate another way to solve this problem. Note that $U_2 = e^{-U_1}$, where U_1 is defined in part (a). Since we found that $U_1 \sim \text{Gamma}(2, 1)$, we can essentially reduce this problem to a *univariate* transformation: let $U_1 \sim \text{Gamma}(2, 1)$, and find the density of $U_2 := e^{-U_1}$.

Let's use the Change of Variable method: $g(u_1) = e^{-u_1} \implies g^{-1}(u_2) = -\ln(u_2)$, and so

$$\left| \frac{\mathrm{d}}{\mathrm{d}u_2} g^{-1}(u_2) \right| = \left| \frac{\mathrm{d}}{\mathrm{d}u_2} [-\ln(u_2)] \right| = \left| -\frac{1}{u_2} \right| = \frac{1}{u_2}$$

Hence, plugging into the Change of Variable formula:

$$\begin{aligned} f_{U_2}(u_2) &= f_{U_1}(g^{-1}(u_2)) \cdot \left| \frac{\mathrm{d}}{\mathrm{d}u_2} g^{-1}(u_2) \right| \\ &= [-\ln(u_2)] e^{-[-\ln(u_2)]} \cdot \frac{1}{u_2} \cdot \mathbb{1}_{\{-\ln(u_2) \geq 0\}} \\ &= [-\ln(u_2)] \cancel{u_2} \cdot \frac{1}{\cancel{u_2}} \cdot \mathbb{1}_{\{0 \leq u_2 \leq 1\}} = -\ln(u_2) \cdot \mathbb{1}_{\{0 \leq u_2 \leq 1\}} = \ln\left(\frac{1}{u_2}\right) \cdot \mathbb{1}_{\{0 \leq u_2 \leq 1\}} \end{aligned}$$

(c) Find the density of $U_3 := Y_1^2$. Is this the same as the density of U_2 from part (b) above?

Solution: We can (again) use either the CDF method or the Change of Variable method. I'll demonstrate using the Change of Variable method: $g(u) = u^2$ so $g^{-1}(u) = \sqrt{u}$ and

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

Hence, plugging into the Change of Variable formula:

$$f_{U_3}(u) = f_{Y_1}(g^{-1}(u)) \cdot \left| \frac{\mathrm{d}}{\mathrm{d}u} g^{-1}(u) \right|$$

$$= \mathbb{1}_{\{0 \leq \sqrt{u} \leq 1\}} \cdot \frac{1}{2\sqrt{u}} = \boxed{\frac{1}{2\sqrt{u}} \cdot \mathbb{1}_{\{0 \leq u \leq 1\}}}$$

Note that this is not the same as the density of U_3. This reveals a very important fact: even if Y_1 and Y_2 follow the same distribution, the distributions of $Y_1 Y_2$ and Y_1^2 will (in general) be different.

2. Let $Y_1, Y_2 \overset{\text{i.i.d.}}{\sim} \text{Exp}(\theta)$. Find the distribution of U, where

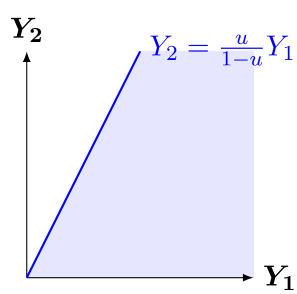
$$U := \frac{Y_2}{Y_1 + Y_2}$$

Be sure to include both the distribution's name as well as any/all relevant parameter(s)!

Solution: First note that $S_U = [0, 1]$. Hence, for a fixed $u \in [0, 1]$,

$$F_U(u) := \mathbb{P}(U \leq u) = \mathbb{P}\left(\frac{Y_2}{Y_1 + Y_2} \leq u \right)$$
$$= \mathbb{P}\left(Y_2 \leq uY_1 + uY_2\right) = \mathbb{P}(Y_2 - uY_2 \leq uY_1) = \mathbb{P}\left(Y_2 \leq \frac{u}{1-u} Y_1 \right)$$

Note $u/(1-u)$ is just a constant - hence the desired probability is a double integral over the following region:



Since the joint density $f_{Y_1,Y_2}(y_1, y_2) = (1/\theta^2)e^{-(y_1+y_2)/\theta} \cdot \mathbb{1}_{\{y_1 \geq 0\ ,\ y_2 \geq 0\}}$ has terms with negative exponents, integration will be a little easier if we can include infinities in our bounds. Hence, let's use the order of integration dy_1 dy_2:

$$F_U(u) = \mathbb{P}\left(Y_2 \leq \frac{u}{1-u} Y_1 \right)$$
$$= \int_0^\infty \int_{\left(\frac{1-u}{u}\right) y_2}^\infty \frac{1}{\theta^2} e^{-(y_1+y_2)/\theta} \, \mathrm{d}y_1 \, \mathrm{d}y_2$$
$$= \frac{1}{\theta} \int_0^\infty e^{-y_2/\theta} e^{-\left(\frac{1-u}{u}\right) y_2/\theta} \, \mathrm{d}y_2 = \frac{1}{\theta} \int_0^\infty e^{-y_2/(u\theta)} \, \mathrm{d}y_2$$
$$= \frac{1}{\theta} \cdot u\theta \int_0^\infty \frac{1}{u\theta} e^{-y_2/(u\theta)} \, \mathrm{d}y_2 = u$$

Additionally, for $u < 0$ we clearly have $F_U(u) = 0$ and for $u \geq 1$ we have $F_U(u) = 1$; hence

$$F_U(u) = \begin{cases} 0 & \text{if } u < 0 \\ u & \text{if } 0 \leq u < 1 \\ 1 & \text{if } u \geq 1 \end{cases}$$

which allows us to conclude $\boxed{U \sim \text{Unif}[0, 1]}$.

As an Aside: this is a very useful result that we will use later in the course!

3. In this problem, we'll consider the exercise of deriving the distribution of the minimum of two *non*-i.i.d. random variables. Suppose $(X, Y) \sim f_{X,Y}$, where

$$f_{X,Y} = 2e^{-(x+y)} \cdot \mathbb{1}_{\{0 \leq x \leq y < \infty\}}$$

Define $U := \min\{X, Y\}$ to be the minimum of X and Y.

(a) Argue that $\overline{F_U}(u)$, the survival of U, is given by $\mathbb{P}(X > u,\ Y > u)$. [Yes, we've done this before but it's good practice to revisit the argument for why this fact holds!]

Solution: By definition,

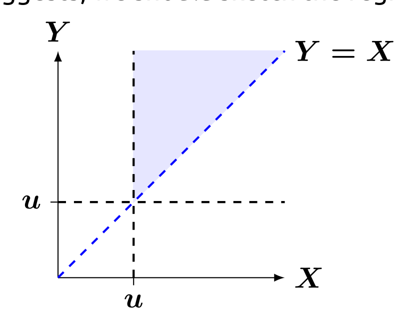
$$\overline{F_U}(u) := 1 - F_U(u = \mathbb{P}(U > u) = \mathbb{P}(\min\{X, Y\} \geq u)$$

If the smallest of two numbers is larger than u, then both numbers must be larger than u. Hence,

$$\overline{F_U}(u) = \mathbb{P}(X > u\ ,\ Y > u)$$

(b) Compute $\overline{F_U}(u)$ as a function of u. **Hint:** You should sketch the region of integration here.

Solution: As the problem suggests, we should sketch the region of integration here.



Hence:

$$\overline{F_U}(u) = \mathbb{P}(X > u,\ Y > u) = \int_u^\infty \int_x^\infty 2e^{-(x+y)}\ \mathrm{d}y\ \mathrm{d}x$$

$$= 2\int_u^\infty e^{-x} \cdot e^{-x}\, \mathrm{d}x = \int_u^\infty e^{-2x}\, \mathrm{d}x = e^{-2u}$$

Additionally, if $u < 0$ then $\overline{F_U}(u) = 1$; hence

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

(c) Use your answer to part (b) to identify the distribution of U by name, including any/all relevant parameter(s).

Solution: This is the survival function of the Exp$(1/2)$ distribution; hence $U \sim$ Exp$(1/2)$.

4. Let $Y_1, Y_2, \cdots \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.

(a) What is the distribution of $U_1 := \sum_{i=1}^6 Y_i^2$? Include both the distribution's name along with any relevant parameter(s)!

Solution: Since the Y_i's are independent, so to will the Y_i^2's be independent. Furthermore, the square of a standard normal distribution follows a χ_1^2 distribution; hence

$$Y_1^2, Y_2^2, \cdots \overset{\text{i.i.d.}}{\sim} \chi_1^2$$

We have previously seen that independent χ^2 random variables add, with their degrees of freedom adding; hence,

$$\sum_{i=1}^6 Y_i^2 \sim \chi_6^2$$

(b) What is the distribution of $U_2 := \sum_{i=1}^6 (Y_i - \overline{Y}_6)^2 + Y_7^2$, where $\overline{Y}_6 := (1/6)\sum_{i=1}^n Y_i$? Include both the distribution's name along with any relevant parameter(s)!

Solution: In general, for $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ we know that

$$\frac{n-1}{\sigma^2} S_n^2 := \frac{n-1}{\sigma^2} \cdot \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \overline{X}_n)^2 \sim \chi_{n-1}^2$$

Plugging in $n = 6$ and $\sigma^2 = 1$ tells us that

$$\sum_{i=1}^n (Y_i - \overline{Y}_n)^2 \sim \chi_{6-1}^2 \sim \chi_5^2$$

Additionally, $Y_7^2 \sim \chi_1^2$, with $Y_7 \perp \sum_{i=1}^6 (Y_i - \overline{Y}_6)^2$; hence, we can again use the closure of the χ^2 distribution under addition to conclude

$$\left(\sum_{i=1}^{6}(Y_i - \overline{Y}_6)^2 + Y_7^2\right) \sim \chi_6^2$$

(c) What is the distribution of $U_3 := 4\overline{Y}_{16}/S_{16}$, where

$$\overline{Y}_{16} := \frac{1}{16}\sum_{i=1}^{n} Y_i \qquad \text{and} \qquad S_{16} := \sqrt{\frac{1}{15}\sum_{i=1}^{16}(Y_i - \overline{Y}_{16})^2}$$

Include both the distribution's name along with any relevant parameter(s)!

Solution: By a result shown in class,

$$\sqrt{16}\left(\frac{\overline{Y}_{16} - 0}{S_{16}}\right) = 4 \cdot \frac{\overline{Y}_{16}}{S_{16}} \sim t_{16-1} \sim t_{15}$$