BIOSTAT 704 - Homework 1

Austin Allen

January 30, 2024

Problem 1

BE Exercise 7.1: Consider a random sample of size n from a distribution with CDF $F(x) = 1 - 1/x$ if $1 \leq x < \infty$, and zero otherwise.

a) Derive the CDF of the smallest order statistic, $X_{1:n}$

For $x \geq 1$,

$$\begin{aligned}
G_n(y) &= P(Y_n \leq y) \\
&= 1 - P(Y_n > y) \\
&= 1 - P(X_{1:n} > y) \\
&= 1 - P(X_1 > y, X_2 > y, ..., X_n > y) \\
&= 1 - P(X_1 > y)P(X_2 > y)...P(X_n > y) \\
&= 1 - [P(X_i > y)]^n \\
&= 1 - [1 - P(X_i \leq y)]^n \\
&= 1 - [1 - F_X(x)]^n \\
&= 1 - [1 - (1 - 1/x)]^n \\
&= 1 - (1/x)^n \\
G_{X_{1:n}}(x) &= \begin{cases} 1 - \frac{1}{x^n}, & \text{for } x \geq 1 \\ 0, & \text{otherwise} \end{cases}
\end{aligned}$$

b) Find the limiting distribution of $X_{1:n}$

For $x \geq 1$,

$$\lim_{n \to \infty} 1 - 1/x^n = 1$$

Thus the limiting distribution of $G_{X_{1:n}}(x)$ is,

$$G(y) = \begin{cases} 1, & \text{for } x \geq 1 \\ 0, & \text{otherwise} \end{cases}$$

c) Find the limiting distribution of $X_{1:n}^n$

Let $Y_n = X_{1:n}^n$. For $y \geq 1$,

$$
\begin{aligned}
G_n(y) &= P(Y_n \leq y) \\
&= P(X_{1:n}^n \leq y) \\
&= P(X_{1:n} \leq y^{1/n}) \\
&= 1 - (\frac{1}{y^{1/n}})^n \\
&= 1 - \frac{1}{y} \\
\lim_{n\to\infty} 1 - \frac{1}{y} &= 1 - \frac{1}{y}
\end{aligned}
$$

Thus, the limiting distribution of $X_{1:n}^n$ is given as follows:

$$
G(y) = \begin{cases} 1 - \frac{1}{y}, & y \geq 1 \\ 0, & \text{otherwise} \end{cases}
$$

BE Exercise 7.18: In exercise 1 (above), find the limiting distribution of $n\ln(X_{1:n})$

Let $Y_n = n\ln(X_{1:n})$. For $y \geq 1$,

$$
\begin{aligned}
G_n(y) &= P(Y_n \leq y) \\
&= P(n\ln(X_{1:n}) \leq y) \\
&= P(\ln(X_{1:n}) \leq \frac{y}{n}) \\
&= P(X_{1:n} \leq e^{y/n}) \\
&= 1 - (\frac{1}{e^{\frac{y}{n}}})^n \\
&= 1 - e^{-y}
\end{aligned}
$$

The limiting distribution of $n\ln(X_{1:n})$ is given below:

$$
G(y) = \begin{cases} 1 - e^{-y}, & y \geq 1 \\ 0, & \text{otherwise} \end{cases}
$$

Problem 2

BE Exercise 7.2: Consider a random sample of size n from a distribution with CDF $F(x) = (1+e^{-x})^{-1}$ for all real x.

a) Does the largest order statistic, $X_{n:n}$, have a limiting distribution?

Let $Y_n = X_{n:n}$. For all real x,

$$\begin{aligned}
G_n(y) &= P(Y_n \le y) \\
&= P(X_{n:n} \le y) \\
&= P(X_1 \le y, X_2 \le y, ..., X_n \le y) \\
&= P(X_1 \le y)P(X_2 \le y)...P(X_n \le y) \\
&= [P(X_i \le y)]^n \\
&= [(1+e^{-y})^{-1}]^n \\
&= (1+e^{-y})^{-n}
\end{aligned}$$

At this point, I'm not sure how to proceed. I believe that I should take the limit, but I do not know how to do this. Entering the function into Desmos reveals that the function will not be right-continuous and non-decreasing for arbitrarily large values of n, so I would conclude that $X_{n:n}$ does not have a limiting distribution.

b) Does $X_{n:n} - \ln(n)$ have a limiting distribution?

Let $Y_n = X_{n:n} - \ln(n)$. For all real x,

$$\begin{aligned}
G_n(y) &= P(Y_n \le y) \\
&= P(X_{n:n} - \ln(n) \le y) \\
&= P(X_{n:n} \le y + \ln(n)) \\
&= P(X_1 \le y + \ln(n), X_2 \le y + \ln(n), ..., X_n \le y + \ln(n)) \\
&= P(X_1 \le y + \ln(n))P(X_2 \le y + \ln(n))...P(X_n \le y + \ln(n)) \\
&= [P(X_i \le y + \ln(n))]^n \\
&= [(1 - e^{-y-\ln(n)})^{-1}]^n \\
&= (1 - e^{-y-\ln(n)})^{-n} \\
&= (1 - e^{-y}e^{-\ln(n)})^{-n} \\
&= (1 - \frac{e^{-y}}{n})^{-n}
\end{aligned}$$

$$\lim_{n\to\infty} (1 - \frac{e^{-y}}{n})^{-n} = e^{-e^{-y}}$$

To verify if the function $e^{-e^{-y}}$ is a CDF, we need to check three requirements:

1. The limit as $y \to \infty$ is 1, and the limit as $y \to -\infty$ is 0
2. It is a non-decreasing function
3. It is continuous over all its domain, so it is right continuous

All three conditions are met, so we can conclude that it is a CDF. Thus, $X_{n:n} - \ln(n)$ does have a limiting distribution, and $G(y) = e^{-e^{-y}}$.

BE Exercise 7.19: In exercise 2 (above), find the limiting distribution of $(1/n)e^{X_{n:n}}$

Let $(1/n)e^{X_{n:n}}$. For all real x,

$$
\begin{aligned}
G_n(y) &= P(Y_n \leq y) \\
&= P(1/ne^{X_{n:n}} \leq y) \\
&= P(X_{n:n} \leq \ln(ny)) \\
&= [(1 + e^{-\ln(ny)})^{-1}]^n \\
&= (1 + \frac{1}{ny})^{-n} \\
&= e^{-1/y}
\end{aligned}
$$

Note that this function does not meet the requirements of a CDF for $y < 0$, so the limiting distribution, $G^*(y)$, is as follows:

$$
\begin{cases} e^{-1/y}, & \text{y 0} \\ 0, & \text{otherwise} \end{cases}
$$

BE Exercise 7.3: Consider a random sample of size n from a distribution with CDF $F(x) = 1 - x^{-2}$ if $x > 1$ for $x > 1$, and 0 otherwise. Determine whether each of the following sequences has a limiting distribution; if so, then give the limiting distribution.

a) $X_{1:n}$

Let $Y_n = X_{1:n}$. For $x > 1$,

$$
\begin{aligned}
G_n(y) &= P(Y_n \leq y) \\
&= P(X_{1:n} \leq y) \\
&= 1 - P(X_{1:n} > y) \\
&= 1 - P(X_1 > y, X_2 > y, ..., X_n > y) \\
&= 1 - P(X_1 > y)P(X_2 > y)...P(X_n > y) \\
&= 1 - [P(X_i > y)]^n \\
&= 1 - [1 - P(X_i \leq y)]^n \\
&= 1 - [1 - (1 - x^{-2})]^n \\
&= 1 - [y^{-2}]^n \\
&= 1 - y^{-2n} \\
\lim_{n \to \infty} 1 - y^{-2n} &= 1
\end{aligned}
$$

We have a small problem in that we need to adjust the support so that the function is right continuous at $y = 1$, such that the limiting distribution of Y_n is

$$
G^*(y) = \begin{cases} 1, & y \geq 1 \\ 0, & \text{otherwise} \end{cases}
$$

b) $X_{n:n}$

Let $Y_n = X_{n:n}$. For $x > 1$,

$$
\begin{aligned}
G_n(y) &= P(Y_n \leq y) \\
&= P(X_{n:n} \leq y) \\
&= P(X_1 \leq y, X_2 \leq y, ..., X_n \leq y) \\
&= P(X_1 \leq y)P(X_2 \leq y)...P(X_n \leq y) \\
&= [P(X_i \leq y)]^n \\
&= [1 - y^{-2}]^n \\
\lim_{n\to\infty} (1 - x^{-2})^n &= 0
\end{aligned}
$$

This sequence does not have a limiting distribution.

NOTE: I HAVE NO IDEA HOW TO CALCULATE THIS LIMIT. I GOT THIS FROM DESMOS, SO I'D LKE HELP KNOWING HOW TO CALCULATE IT

c) $n^{-1/2}X_{n:n}$

Let $Y_n = n^{-1/2}X_{n:n}$. For $x > 1$,

$$
\begin{aligned}
G_n(y) &= P(Y_n \leq y) \\
&= P(n^{-1/2}X_{n:n} \leq y) \\
&= P(X_{n:n} \leq n^{1/2}y) \\
&= P(X_1 \leq n^{1/2}y, X_2 \leq n^{1/2}y, ..., X_n \leq n^{1/2}y) \\
&= P(X_1 \leq n^{1/2}y)P(X_2 \leq n^{1/2}y)...P(X_n \leq n^{1/2}y) \\
&= [P(X_i \leq n^{1/2}y)]^n \\
&= [1 - (n^{1/2}y)^{-2}]^n \\
&= [1 - \frac{1}{ny^2}]^n \\
\lim_{n\to\infty} [1 - \frac{1}{ny^2}]^n &= e^{-1/y^2}
\end{aligned}
$$

The function is not right continuous at $y = 1$, but the other two conditions are met. That leads us to give the following limiting distribution:

$$
G^*(y) = \begin{cases} e^{-1/y^2}, & y \geq 1 \\ 0, & \text{otherwise} \end{cases}
$$

Problem 3

Consider Example 7.2.2 in the textbook, on pages 233. Let $X1, X2, ..., X_n$ be a random sample from an exponential distribution, $Xi \sim EXP(\theta)$ and let $Y_n = X_{1:n}$ be the smallest order statistic.

a) Using the definition of a CDF (Theorem 2.2.3), show that $G_n(Y)$ is a proper CDF. That is, show that the conditions (2.2.8 - 2.2.11) hold for $G_n(y)$.

I BELIEVE THIS IS ASKING TO VERIFY THAT $G_n(y)$ IS A PROPER CDF AS OPPOSED TO $G(y)$. I NEED TO CONFIRM THIS IN OFFICE HOURS ON MONDAY.

$$
\begin{aligned}
G_n(y) &= P(Y_n \leq y) \\
&= P(X_{1:n} \leq y) \\
&= 1 - P(X_{1:n} > y) \\
&= 1 - P(X_1 > y, X_2 > y, ..., X_n > y) \\
&= 1 - P(X_1 > y)P(X_2 > y)...P(X_n > y) \\
&= 1 - [P(X_i > y)]^n \\
&= 1 - [1 - P(X_i \leq y)]^n \\
&= 1 - [1 - (1 - e^{-y/\theta})]^n \\
&= 1 - e^{-ny/\theta}
\end{aligned}
$$

The following conditions must be met for $G_n(y)$ to be a proper CDF. I will evaluate each one in turn:

1. $\lim_{y\to\infty} G_n(y)$ must equal 1, and $\lim_{y\to-\infty} G_n(y)$ must equal 0

- Note that n is required to be a positive integer for this to be a valid sequence. This being the case, no matter what arbitrary value we may have for n, the limit as y goes to positive infinity is 1, and the limit as y goes to negative infinity is 0. This condition holds

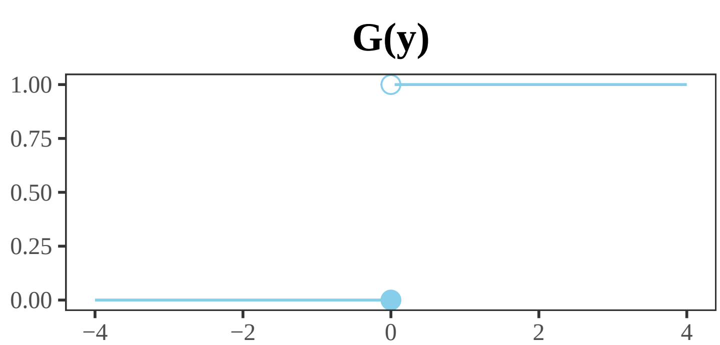
2. The function must not be decreasing

- Again, because n must be a positive integer, any value we choose for n over $G_n(y)$'s domain will not alter the fact that the function is not decreasing

3. The function must be right continuous over its domain

- The funcion is continuous for all $y > 0$.

b) Sketch the graph of the limiting function $G(Y)$.



c) Explain why the authors state that G(y) being discontinuous (and not even right-continuous) at y = 0 ”is not a problem”. How does this connect with our explanations in the lecture notes on pages 1–2?

In the definition on page 1, it states that a sequence converges in distribution as long as "$\lim_{n\to\infty} G_n(y) = G(y)$ for all values of y at which $G(y)$ is continuous." The authors can say that the fact that $G(y)$ is discontinous from the right at $y = 0$ "isn't a problem" because $G(y)$ is continous at the same points that $G_n(y)$ is continous.

Problem 4

(Example 7.2.6 in the textbook, on page 235)

Consider $X1, ..., X_n$ **a random sample where** $X_i \sim EXP(\theta)$**. Define the random sequence** $Y_n = (1/\theta)X_{n:n} - \ln(n)$**. The purpose of this exercise is to demonstrate that** Y_n **converges in distribution**

a) Prove that the CDF of Y_n **is**

$$G_n(Y) = \begin{cases} \left[1 - \frac{1}{n}e^{-y}\right]^n, & \text{when } y > -\ln(n) \\ 0, & \text{otherwise} \end{cases}$$

For all values of y in the support of $G_n(y)$,

$$\begin{aligned} G_n(Y) &= P(Y_n \leq y) \\ &= P((1/\theta)X_{n:n} - \ln(n) \leq y) \\ &= P(X_{n:n} \leq \theta y + \theta \ln(n)) \\ &= P(X_1 \leq \theta y + \theta \ln(n), ..., X_n \leq \theta y + \theta \ln(n)) \\ &= P(X_1 \leq \theta y + \theta \ln(n))...P(X_n \leq \theta y + \theta \ln(n)) \\ &= [F(\theta y + \theta \ln(n))]^n \\ &= [1 - e^{-1/\theta(\theta y + \theta \ln(n))}]^n \\ &= [1 - e^{-y - \ln(n)}]^n \\ &= [1 - e^{y} e^{\ln(n)}]^n \\ &= [1 - \frac{1}{n}e^{-y}]^n \end{aligned}$$

Note that the original support for $F_X(x)$ was $x > 0$. However, because we are subtracting $\ln(n)$ from $X_{n:n}$, and because we want x to be positive, the support for y is $y > -\ln(n)$. This leads to the following result:

$$G_n(Y) = \begin{cases} \left[1 - \frac{1}{n}e^{-y}\right]^n, & \text{when } y > \ln(n) \\ 0, & \text{otherwise} \end{cases}$$

b) Show that $\lim_{n\to\infty} G_n(Y) = G(Y)$**, with** $G(y) = e^{-e^{-y}}$**, where** $-\infty < y < \infty$**.**

$$\lim_{n\to\infty} (1 + c/n + d(n)/n)^{nb} = e^{cb}$$

$$\begin{aligned} \lim_{n\to\infty} \left(1 - \frac{1}{n}e^{-y}\right)^n &= e^{e^{-y}} \\ &= G(y) \end{aligned}$$

Note that the function is continuous over all real numbers, and that $\lim_{y\to-\infty}[1 - \frac{1}{n}e^{-y}]^n = 0$ and that $\lim_{y\to\infty}[1 - \frac{1}{n}e^{-y}]^n = 1$. Thus, the support of $G(y)$ is $-\infty < y < \infty$.

c) Show that $G(y)$ **is a CDF.**

We will show that $G(y)$ is a CDF by showing that it satisfies the three conditions in the definition of a CDF:

1. $\lim_{y\to\infty} G_n(y)$ must equal 1, and $\lim_{y\to-\infty} G_n(y)$ must equal 0

- We showed this in part b. This condition is met

2. The function must not be decreasing

- The function $e^{-e^{-y}}$ does not decrease as y increases. This condition is met

3. The function must be right continuous over its domain

- Because of the exponential nature of this function, there are no real values of $G(y)$ that are discontinuous, so it is also right continous from $-\infty$ to ∞. This condition is met, and the function $G(y)$ is a CDF

Problem 5

BE Exercise 7.5: Suppose that $Z_i \sim N(0,1)$ and that $Z_1, Z_2, \ldots$ are independent. Use moment generating functions to find the limiting distribution of $\sum_{i=1}^{n}(Z_i + 1/n)/\sqrt{n}$.

Let $Y_n = \sum_{i=1}^{n}(Z_i + 1/n)/\sqrt{n}$:

$$\begin{aligned}
M_{Y_n}(t) &= E[\exp(\sum_{i=1}^{n}(Z_i + 1/n)(t/\sqrt{n}))] \\
&= E[\exp(t/\sqrt{n}(\sum_{i=1}^{n}(Z_i + 1/n)))] \\
&= E[\exp(t/\sqrt{n}(\sum_{i=1}^{n} Z_i + \sum_{i=1}^{n} 1/n))] \\
&= E[\exp(t/\sqrt{n}(\sum_{i=1}^{n} Z_i + 1))] \\
&= E[\exp(\sum_{i=1}^{n} Z_i t/\sqrt{n} + t/\sqrt{n}))] \\
&= E[\exp(\sum_{i=1}^{n} Z_i t/\sqrt{n})\exp(t/\sqrt{n}))] \\
&= \exp(t/\sqrt{n})E[\exp(\sum_{i=1}^{n} Z_i t/\sqrt{n})] \\
&= \exp(t/\sqrt{n})E[\prod_{i=1}^{n} \exp(Z_i t/\sqrt{n})] \\
&= \exp(t/\sqrt{n})\prod_{i=1}^{n} E[\exp(Z_i t/\sqrt{n})] \\
&= \exp(t/\sqrt{n})[E[\exp(Z_i t/\sqrt{n})]]^n \\
&= \exp(t/\sqrt{n})[M_Z(t/\sqrt{n})]^n \\
&= \exp(t/\sqrt{n})[\exp(t^2/2n)]^n \\
&= \exp(t/\sqrt{n})\exp(t^2/2) \\
\lim_{n\to\infty} \exp(t/\sqrt{n})\exp(t^2/2) &= \exp(0)\exp(t^2/2) \\
&= \exp(t^2/2)
\end{aligned}$$

Thus, the limiting distribution of $\sum_{i=1}^{n}(Z_i + 1/n)/\sqrt{n}$ is the standard normal distribution, $Y \to^d N(0,1)$.