Lesson 1.1

The Cumulative Distribution Function Technique

Introduction

The objective of statistics is to make inferences about a population based on information contained in a sample taken from that population. The quantities used to estimate population parameters or to make decisions about a population are functions of the n random observations that appear in a sample.

To illustrate, consider the problem of estimating a population mean, μ. Intuitively, we draw a random sample of n observations, $y_1, y_2, \dots, y_n$, from the population and use the sample mean

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

to estimate μ.

If the probability distributuion of the $Y_i's$ is known what would be the distribution of the sample mean $\bar{Y}$. If $Y_i's$ are random variables then $\bar{Y}$ is likewise a random variable.

In this module, we shall study methods in finding the distribution of functions of random variables. These methods include:

1. Cumulative Distribution Function (CDF) Technique
2. Transformation Technique
3. Moment Generation Function (MGF) Technique

The Cumulative Distribution Function (CDF) Technique

Suppose Y is a continuous random variable with cumulative distribution function $F_Y(y) = P(Y \leq y)$. The CDF technique is useful if $F_Y(y)$ has a closed form. The steps in the CDF method are:

1. Find a closed form of the CDF of Y.
2. Find the support of $U = g(Y)$.
3. Write $F_U(u) = P(U \leq U)$ in terms of $F)Y(y)$.
4. Differentiate $F_U(u)$ to obtain the PDF of U.

Example 1.1.1

Suppose $Y \sim U(0,1)$. Find the distribution of $U = g(Y) = -lnY$.

SOLUTION

Verify that the CDF of Y is given by

$$F_Y(y) = \begin{cases} 0,\ y \leq 0 \\ y,\ 0 < y < 1 \\ 1,\ y \geq 1 \end{cases}$$

Since the support for $Y \sim U(0,1)$ is $R_Y = \{y : 0 < y < 1\}$, thus, it follows that the support for $U = -lnY$ is $R_U = \{u : u > 0\}$.

Now, we are going to find the CDF of U.

$$\begin{aligned} F_U(u) &= P(U \leq u) \\ &= P(-lnY \leq u) \\ &= P(lnY > -u) \\ &= P(Y > e^{-u}) \\ &= 1 - P(Y \leq e^{-u}) \\ &= 1 - F_Y(e^{-u}) \end{aligned}$$

Since $F_Y(y) = y$ for $0 < y < 1$, hence, for $u > 0$ we have

$$\begin{aligned} F_U(u) &= 1 - F_Y(e^{-u}) \\ &= 1 - e^{-u} \end{aligned}$$

To get the PDF of U, we have

$$\begin{aligned} f_U(u) &= \frac{d}{du} F_U(u) \\ &= \frac{d}{du}\left[1 - e^{-u}\right] \\ &= e^{-u} \end{aligned}$$

Therefore,

$$f_U(u) = \begin{cases} e^{-u}, \; u > 0 \\ 0, \text{ elsewhere} \end{cases}$$

which shows that $U \sim Exp(1)$ which is an exponential distribution with mean $\beta = 1$.

Example 1.1.2

A process for refining sugar yields up to 1 ton of pure sugar per day, but the actual amount produced, Y, is a random variable because of machine breakdowns and other slowdowns. Suppose that Y has density function given by

$$f_Y(y) = \begin{cases} 2y, \; 0 \le y \le 1 \\ 0, \text{ elsewhere} \end{cases}$$

The company is paid at the rate of \$300 per ton for the refined sugar, but it also has a fixed overhead cost of \$100 per day. Thus the daily profit, in hundreds of dollars, is $U = 3Y - 1$. Find the probability density function for U.

SOLUTION

Verify that the CDF of Y is given by

$$F_Y(y) = \begin{cases} 0, \; y < 0 \\ y^2, \; 0 \le y < 1 \\ 1, \; y \ge 1 \end{cases}$$

Note that the support of $U = 3Y - 1$ is $R_U = \{u : -1 \le u \le 2\}$.

The CDF of U is obtained as follows:

$$
\begin{aligned}
F_U(u) &= P(U \leq u) \\
&= P(3Y - 1 \leq u) \\
&= P(Y \leq \frac{u+1}{3}) \\
&= F_Y\left(\frac{u+1}{3}\right)
\end{aligned}
$$

That is, for $-1 \leq u \leq 2$,

$$
\begin{aligned}
F_U(u) &= F_Y\left(\frac{u+1}{3}\right) \\
&= \left(\frac{u+1}{3}\right)^2
\end{aligned}
$$

And the density function for U is

$$
\begin{aligned}
f_U(u) &= \frac{d}{du} F_U(u) \\
&= \frac{d}{du}\left[\left(\frac{u+1}{3}\right)^2\right] \\
&= \frac{2}{9}(u+1)
\end{aligned}
$$

Therefore,

$$
f_U(u) = \begin{cases} \frac{2}{9}(u+1), \ -1 \leq u \leq 2 \\ 0, \text{ elsewhere} \end{cases}
$$

Example 1.1.3

Let Y be a random variable with PDF

$$
f_Y(y) = \begin{cases} \frac{3}{2}y^2, \ -1 \leq y \leq 1 \\ 0, \text{ elsewhere} \end{cases}
$$

SOLUTION [Left as a classroom exercise!]