Lesson 1.5

Probability Distribution of Order Statistics

Introduction

Many functions of random variables of interest in practice depend on the relative magnitudes of the observed variables. For instance, we may be interested in the fastest time in an automobile race or the heaviest mouse among those fed on a certain diet. Thus, we often order observed random variables according to their magnitudes. The resulting ordered variables are called **order statistics**.

Formally, let $Y_1, Y_2, \cdots, Y_n$ denote independent continuous random variables with distribution function $F_Y(y)$ and density function $f_Y(y)$. We denote the ordered random variables Y_i by $Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}$, where $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$. (Because the random variables are continuous, the equality signs can be ignored.) Using this notation,

$$
\begin{aligned}
Y_{(1)} &\equiv \text{smallest (minimum) of } Y_1, Y_2, \cdots, Y_n \\
Y_{(2)} &\equiv \text{second smallest of } Y_1, Y_2, \cdots, Y_n \\
&\vdots \\
Y_{(n)} &\equiv \text{largest (maximum) of } Y_1, Y_2, \cdots, Y_n
\end{aligned}
$$

Our goal is to determine the probability distribution of these order statistics, particularly the minimum and the maximum. We shall do this using the cumulative distribution function technique.

Probability Density Function of the Minimum

Suppose let $Y_1, Y_2, \cdots, Y_n$ is a random sample from the distribution $f_Y(y)$. Using the CDF technique we have,

$$
\begin{aligned}
F_{Y_{(1)}}(y) &= P(Y_{(1)} \leq y) \\
&= 1 - P(Y_{(1)} > y) \\
&= 1 - P[\{Y_1 > y\} \cap \{Y_2 > y\} \cap \cdots \cap \{Y_n > y\}] \\
&= 1 - P(Y_1 > y) \times P(Y_2 > y) \times \cdots \times P(Y_n > y) \\
&= 1 - [P(Y > y)]^n \\
&= 1 - [1 - P(Y \leq y)]^n \\
&= 1 - [1 - F_Y(y)]^n
\end{aligned}
$$

Thus, for values of y in the support of $Y_{(1)}$, the PDF of $Y_{(1)}$ is given by

$$
\begin{aligned}
f_{Y_{(1)}}(y) &= \frac{d}{dy}\left[F_{Y_{(1)}}(y)\right] \\
&= \frac{d}{dy}\left[1 - [1 - F_Y(y)]^n\right] \\
&= -n[1 - F_Y(y)]^{n-1} \times [-f_Y(y)] \\
&= n f_Y(y)[1 - F_Y(y)]^{n-1}
\end{aligned}
$$

and 0, otherwise.

Example 1.5.1

Let Y_1 and Y_2 be independent and uniformly distributed over the interval (0, 1). Find

a. the probability density function of $Y_{(1)} = min(Y_1, Y_2)$.

b. the mean and variance of $Y_{(1)}$.

SOLUTION

a. The common density function of each Y_1 and Y_2 is

$$
f_Y(y) = \begin{cases} 1,\ 0 < y < 1 \\ 0,\ \text{otherwise} \end{cases}
$$

and the CDF is given by

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

Therefore, the PDF of $Y_{(1)}$ is

$$\begin{aligned} f_{Y_{(1)}}(y) &= n f_Y(y)[1 - F_Y(y)]^{n-1} \\ &= 2(1)[1-y]^{2-1} \\ &= 2(1-y), \ \ 0 < y < 1 \end{aligned}$$

and 0, otherwise.

b. The above PDF resembles a Beta density with $\alpha = 1$ and $\beta = 2$, thus,

$$\begin{aligned} E[Y_{(1)}] &= \frac{\alpha}{\alpha + \beta} \\ &= \frac{1}{1+2} \\ &= \frac{1}{3} \end{aligned}$$

and

$$\begin{aligned} V[Y_{(1)}] &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\ &= \frac{1(2)}{(1+2)^2(1+2+1)} \\ &= \frac{2}{36} \\ &= \frac{1}{18} \end{aligned}$$

Probability Density Function of the Maximum

Suppose let $Y_1, Y_2, \cdots, Y_n$ is a random sample from the distribution $f_Y(y)$. Using the CDF technique we have,

$$\begin{aligned} F_{Y_{(n)}}(y) &= P(Y_{(n)} \leq y) \\ &= P[\{Y_1 < y\} \cap \{Y_2 < y\} \cap \cdots \cap \{Y_n < y\}] \\ &= P(Y_1 < y) \times P(Y_2 < y) \times \cdots \times P(Y_n < y) \\ &= [P(Y < y)]^n \\ &= [F_Y(y)]^n \end{aligned}$$

Thus, for values of y in the support of $Y_{(n)}$, the PDF of $Y_{(n)}$ is given by

$$\begin{aligned} f_{Y_{(n)}}(y) &= \frac{d}{dy}\left[F_{Y_{(n)}}(y)\right] \\ &= \frac{d}{dy}\left[[F_Y(y)]^n\right] \\ &= n[F_Y(y)]^{n-1} \times [f_Y(y)] \\ &= n f_Y(y)[F_Y(y)]^{n-1} \end{aligned}$$

Example 1.5.2

Let Y_1 and Y_2 be independent and uniformly distributed over the interval (0, 1). Find

a. the probability density function of $Y_{(n)} = max(Y_1, Y_2)$.

b. the mean and variance of $Y_{(n)}$.

SOLUTION

a. From the solution in *Example 1.5.1*, we have the PDF and CDF of Y, respectively,

$$f_Y(y) = \begin{cases} 1, \ 0 < y < 1 \\ 0, \ \text{otherwise} \end{cases}$$

and

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

Thus, for $0 < y < 1$, the PDF of $Y_{(n)}$ is

$$f_{Y_{(n)}}(y) = 2(1)y^{2-1} = 2y$$

and 0, otherwise.

b. The above PDF resembles a Beta density with $\alpha = 2$ and $\beta = 1$, hence

$$\begin{aligned} E[Y_{(n)}] &= \frac{\alpha}{\alpha + \beta} \\ &= \frac{2}{1+2} \\ &= \frac{2}{3} \end{aligned}$$

and

$$\begin{aligned} V[Y_{(1)}] &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\ &= \frac{2(1)}{(2+1)^2(2+1+1)} \\ &= \frac{2}{36} \\ &= \frac{1}{18} \end{aligned}$$

Example 1.5.3 [Left as a classroom exercise!]

An engineering system consists of five components placed in a series, that is, the system fails if the first conponent fails. Suppose the lifetimes of the five components denoted as Y_1, Y_2, Y_3, Y_4, Y_5 are a random sample from an exponential distribution with mean $\beta = 1$ year. Since the system fails if the first component fails, system failures can be determined probabilistically by deriving the PDF of the $Y_{(1)}$.

a. Find the PDF of $Y_{(1)}$.

b. What is the chance that the system lasts more than 6 months?