Statistics 351 (Fall 2015) October 23, 2015
Prof. Michael Kozdron

Lecture #16: Order Statistics (Chapter 4)

Reference. §4.1 pages 101–103

Consider a random sample $X_1, \dots, X_n$. We define the *order statistics* (or order variables) $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ to be the rearrangement (or permutation) of $X_1, \dots, X_n$ such that

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}.$$

Formally, we define

$$X_{(1)} = \min\{X_1, \dots, X_n\}, \quad X_{(n)} = \max\{X_1, \dots, X_n\},$$

and

$$X_{(k)} = \min\left\{\{X_1, \dots, X_n\} \setminus \{X_{(1)}, \dots, X_{(k-1)}\}\right\}$$

for $k = 2, \dots, n-1$.

Our goal in the next few lectures will be to determine various distributions of (combinations of) the order statistics. We begin with the easiest two, namely $X_{(1)}$ and $X_{(n)}$. For ease, we will assume throughout that $X_1, \dots, X_n$ are independent and identically distributed continuous random variables with common distribution $F_X(x)$ and common density $f_X(x)$.

The distribution of $X_{(n)}$

Our first goal is to find the distribution function and density function of $X_{(n)} = \max\{X_1, \dots, X_n\}$. By definition,

$$F_{X_{(n)}}(y) = P\{X_{(n)} \le y\} = P\{\max\{X_1, \dots, X_n\} \le y\} = P\{X_1 \le y, \dots, X_n \le y\}.$$

However, the independence of $X_1, \dots, X_n$ implies that

$$P\{X_1 \le y, \dots, X_n \le y\} = P\{X_1 \le y\} \cdots P\{X_n \le y\} = F_{X_1}(y) \cdots F_{X_n}(y) = [F_X(y)]^n$$

where the last equality follows from the fact that $X_1, \dots, X_n$ are identically distributed. In summary,

$$F_{X_{(n)}}(y) = [F_X(y)]^n.$$

In order to find the density of $X_{(n)}$, we simply differentiate. That is,

$$f_{X_{(n)}}(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_{X_{(n)}}(y) = \frac{\mathrm{d}}{\mathrm{d}y} [F_X(y)]^n = n\, [F_X(y)]^{n-1} \frac{\mathrm{d}}{\mathrm{d}y} F_X(y) = n f_X(y)\, [F_X(y)]^{n-1}.$$

The distribution of $X_{(1)}$

In order to find the distribution function and density function of $X_{(1)} = \min\{X_1, \ldots, X_n\}$ we note that by definition

$$F_{X_{(1)}}(y) = P\{X_{(1)} \le y\} = P\{\min\{X_1, \ldots, X_n\} \le y\}.$$

In contrast to the calculation of the maximum above, we note that

$$P\{\min\{X_1, \ldots, X_n\} \le y\} = 1 - P\{\min\{X_1, \ldots, X_n\} > y\}$$

and

$$P\{\min\{X_1, \ldots, X_n\} > y\} = P\{X_1 > y, \ldots, X_n > y\}.$$

It now follows from the independence of $X_1, \ldots, X_n$ that

$$\begin{aligned} P\{X_1 > y, \ldots, X_n > y\} = P\{X_1 > y\} \cdots P\{X_n > y\} &= [1 - F_{X_1}(y)] \cdots [1 - F_{X_n}(y)] \\ &= [1 - F_X(y)]^n \end{aligned}$$

where the last equality follows from the fact that $X_1, \ldots, X_n$ are identically distributed. In summary,

$$F_{X_{(1)}}(y) = 1 - [1 - F_X(y)]^n .$$

In order to find the density of $X_{(1)}$, we simply differentiate. That is,

$$\begin{aligned} f_{X_{(1)}}(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_{X_{(1)}}(y) = \frac{\mathrm{d}}{\mathrm{d}y} 1 - [1 - F_X(y)]^n &= -n\,[1 - F_X(y)]^{n-1} \frac{\mathrm{d}}{\mathrm{d}y}[1 - F_X(y)] \\ &= n f_X(y)\,[1 - F_X(y)]^{n-1} . \end{aligned}$$

Example. Suppose that X_1, X_2, X_3 are iid $\mathrm{Exp}(\lambda)$ random variables. Determine the distribution of $Y = \min\{X_1, X_2, X_3\}$.

Solution. Note that X_1, X_2, X_3 have common density

$$f_X(x) = \frac{1}{\lambda} e^{-x/\lambda}, \quad x > 0.$$

Therefore,

$$P\{Y > y\} = P\{\min\{X_1, X_2, X_3\} > y\} = P\{X_1 > y, X_2 > y, X_3 > y\} = [P\{X_1 > y\}]^3 .$$

Since

$$P\{X_1 > y\} = \int_y^\infty \frac{1}{\lambda} e^{-x/\lambda}\,\mathrm{d}x = \left(-e^{-x/\lambda}\right)\Big|_y^\infty = e^{-y/\lambda}$$

we find

$$P\{Y > y\} = \left[e^{-y/\lambda}\right]^3 = e^{-3y/\lambda}$$

and so

$$F_Y(y) = P\{Y \le y\} = 1 - P\{Y > y\} = 1 - e^{-3y/\lambda}.$$

Furthermore,

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{3}{\lambda} e^{-3y/\lambda} = \frac{1}{(\lambda/3)} e^{-y/(\lambda/3)}, \quad y > 0.$$

Hence, we conclude that $Y \in \mathrm{Exp}(\lambda/3)$.