Lecture #18: Distribution of $X_{(k)}$, $k = 1, 2, \ldots, n$

Let $X_{(k)}$ denote the kth order variable so that $X_{(k)}$ is the kth smallest among $\{X_1, \ldots, X_n\}$. Thus, if $X_{(k)}$ is to be less than or equal to y_k, it must be the case that at least k of $\{X_1, \ldots, X_n\}$ are less than or equal to y_k. That is,

$$P\{X_{(k)} \le y_k\} = P\{\text{at least } k \text{ of } \{X_1, \ldots, X_n\} \le y_k\}$$

and so

$$P\{X_{(k)} \le y_k\} = \sum_{j=k}^{n} P\{\text{exactly } j \text{ of } \{X_1, \ldots, X_n\} \le y_k\}.$$

Since $X_1, \ldots, X_n$ are independent and identically distributed, it follows that the event that $\{\text{exactly } j \text{ of } \{X_1, \ldots, X_n\} \le y_k\}$ has the same probability as the event that there are exactly j successes in n trials of an experiment with success probability $p = P\{X \le y_k\} = F(y_k)$ on each trial. Thus,

$$P\{\text{exactly } j \text{ of } \{X_1, \ldots, X_n\} \le y_k\} = \binom{n}{j} [F(y_k)]^j [1 - F(y_k)]^{n-j}.$$

In other words,

$$P\{X_{(k)} \le y_k\} = \sum_{j=k}^{n} \binom{n}{j} [F(y_k)]^j [1 - F(y_k)]^{n-j}$$

is an expression for the distribution function of $X_{(k)}$.

There is, however, a useful identity for binomial distributions in terms of beta distributions, namely

$$\sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j} = k \binom{n}{k} \int_0^p x^{k-1} (1-x)^{n-k} \, \mathrm{d}x \qquad (*)$$

where $k \in \{0, 1, \ldots, n\}$ and $0 < p < 1$.

Remark. Problem #1 on Assignment #6 outlines the proof of this identity.

Since

$$k \binom{n}{k} = \frac{n!}{(k-1)!(n-k)!} = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)}$$

we see that $(*)$ is equivalent to

$$\sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j} = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} \int_0^p x^{k-1} (1-x)^{n-k} \, \mathrm{d}x.$$

Thus, if $X \in \text{Bin}(n, p)$ and $Y \in \beta(k, n-k+1)$, then this says that

$$P(X \ge k) = P(Y \le p), \quad \text{or, equivalently,} \quad 1 - F_X(k-1) = F_Y(p).$$

In our case of the distribution of $X_{(k)}$, we see that

$$P\{X_{(k)} \le y_k\} = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} \int_0^{F(y_k)} x^{k-1}(1-x)^{n-k}\,\mathrm{d}x$$

and so

$$f_{X_{(k)}}(y_k) = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)}[F(y_k)]^{k-1}[1-F(y_k)]^{n-k}\, f(y_k).$$

In summary, we have proved the following theorem.

Theorem. *If $X_1, \ldots, X_n$ are i.i.d. with common distribution function $F(x)$ and common density function $f(x)$, then the density of $X_{(k)}$, $k = 1, \ldots, n$, is*

$$f_{X_{(k)}}(y_k) = \frac{\Gamma(n+1)}{\Gamma(k)\,\Gamma(n-k+1)}[F(y_k)]^{k-1}\,[1-F(y_k)]^{n-k}\, f(y_k)$$

and the distribution function of $X_{(k)}$, $k = 1, \ldots, n$, is

$$F_{X_{(k)}}(y_k) = \frac{\Gamma(n+1)}{\Gamma(k)\,\Gamma(n-k+1)} \int_0^{F(y_k)} x^{k-1}(1-x)^{n-k}\,\mathrm{d}x.$$

Exercise. Check that for $k = 1$ and $k = n$, we recover the formulæ for $X_{(1)}$ and $X_{(n)}$ that we had earlier.

Remark. An alternate derivation of this result is given on pages 103–105 of [1].

Example. Suppose that X_1, X_2, X_3, X_4 are i.i.d. $U(0, 1)$ random variables so that

$$F(x) = \begin{cases} 0, & \text{if } x \le 0, \\ x, & \text{if } 0 < x < 1, \\ 1, & \text{if } x \ge 1, \end{cases} \quad \text{and} \quad f(x) = \begin{cases} 1, & \text{if } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

We find

$$F_{X_{(4)}}(y_4) = [F(y_4)]^4 = \begin{cases} 0, & \text{if } y_4 \le 0, \\ y_4^4, & \text{if } 0 < y_4 < 1, \\ 1, & \text{if } y_4 \ge 1, \end{cases}$$

and so

$$f_{X_{(4)}}(y_4) = \frac{d}{dy_4} F_{X_{(4)}}(y_4) = \begin{cases} 4y_4^3, & \text{if } 0 < y_4 < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Furthermore, we find

$$F_{X_{(1)}}(y_1) = 1 - [1 - F(y_1)]^4 = \begin{cases} 1-(1-0)^4 = 0, & \text{if } y_1 \le 0, \\ 1-(1-y_1)^4, & \text{if } 0 < y_1 < 1, \\ 1-(1-1)^4 = 1, & \text{if } y_1 \ge 1, \end{cases}$$

and so

$$f_{X_{(1)}}(y_1) = \frac{d}{dy_1} F_{X_{(1)}}(y_1) = \begin{cases} 4(1-y_1)^3, & \text{if } 0 < y_1 < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Using our earlier theorem, we find in general that the density of $X_{(k)}$, $k = 1, 2, 3, 4$, is given by

$$f_{X_{(k)}}(y_k) = \frac{\Gamma(5)}{\Gamma(k)\Gamma(5-k)} y_k^{k-1}(1-y_k)^{4-k}, \quad 0 < y_k < 1.$$

That is, $X_{(k)} \in \beta(k, 5-k)$. In particular, we note that $\mathbb{E}(X_{(k)}) = k/5$.

Example. If X_1, X_2, X_3, X_4 are i.i.d. $U(0,1)$ random variables, determine $\mathbb{E}(R)$, the expected value of the range $R = X_{(4)} - X_{(1)}$.

Solution. Using the results of Lecture #17, we know that

$$f_R(r) = 4(4-1) \int_{-\infty}^{\infty} f(v) f(r+v) [F(r+v) - F(v)]^{4-2} \, \mathrm{d}v, \quad r > 0.$$

However, we need to be careful with our limits in this example. Since X_1, X_2, X_3, X_4 are i.i.d. $U(0,1)$, it is clear that $f_R(r) = 0$ for $r \leq 0$ or $r \geq 1$. However, in our notation, given $R = r$, it must be the case that the allowable values of $X_{(1)} = v$ range from $0 \leq v \leq 1 - r$. This gives

$$f_R(r) = 12 \int_0^{1-r} [(r+v) - v]^2 \, \mathrm{d}u = 12r^2(1-r)$$

for $0 < r < 1$. We recognize this as the density of a $\beta(3,2)$ random variable, and so

$$\mathbb{E}(R) = \frac{3}{5}.$$

Note that this answer is "obvious" since we expect 4 points uniformly distributed to be evenly spread out: $\mathbb{E}(X_{(1)}) = 1/5$, $\mathbb{E}(X_{(2)}) = 2/5$, $\mathbb{E}(X_{(3)}) = 3/5$, $\mathbb{E}(X_{(4)}) = 4/5$. Thus,

$$\mathbb{E}(R) = \mathbb{E}(X_{(4)}) - \mathbb{E}(X_{(1)}) = \frac{3}{5}.$$