# Appendix | C

# Order Statistics

Let $X_1, X_2, \ldots, X_n$ be $n$ independent draws from a distribution $F$ with associated density $f$. Let $Y_1^{(n)}, Y_2^{(n)}, \ldots, Y_n^{(n)}$ be a rearrangement of these so that $Y_1^{(n)} \geq Y_2^{(n)} \geq \cdots \geq Y_n^{(n)}$. The random variables $Y_k^{(n)}$, $k = 1, 2, \ldots, n$ are referred to as *order statistics*.

Let $F_k^{(n)}$ denote the distribution of $Y_k^{(n)}$, with corresponding probability density function $f_k^{(n)}$.

When the "sample size" $n$ is fixed and there is no ambiguity, we will economize on notation and write $Y_k$ instead of $Y_k^{(n)}$, $F_k$ instead of $F_k^{(n)}$ and $f_k$ instead of $f_k^{(n)}$.

We will typically be interested in properties of the highest and second-highest order statistics, $Y_1$ and $Y_2$.

## HIGHEST-ORDER STATISTIC

The distribution of the highest order statistic $Y_1$ is easy to derive. The event that $Y_1 \leq y$ is the same as the event: for all $k$, $X_k \leq y$. Since each $X_k$ is an independent draw from the same distribution $F$, we have[1]

$$F_1(y) = F(y)^n \tag{C.1}$$

The associated probability density function is

$$f_1(y) = nF(y)^{n-1} f(y)$$

Observe that if $F$ stochastically dominates $G$ (for all $x$, $F(x) \leq G(x)$) and $F_1$ and $G_1$ are the distributions of the highest-order statistics of $n$ draws from $F$ and $G$, respectively, then $F_1$ stochastically dominates $G_1$.

[1]We write $F(y)^n$ to denote $(F(y))^n$.

## SECOND-HIGHEST-ORDER STATISTIC

The distribution the second-highest-order statistic $Y_2$ can also be easily derived. The event that $Y_2$ is less than or equal to $y$ is the union of the following disjoint events: (i) all $X_k$'s are less than or equal to $y$, and (ii) $n-1$ of the $X_k$'s are less than or equal to $y$ and one is greater than $y$. There are $n$ different ways in which (ii) can occur, so we have

$$\begin{aligned} F_2(y) &= \underbrace{F(y)^n}_{\text{(i)}} + \underbrace{nF(y)^{n-1}(1-F(y))}_{\text{(ii)}} \\ &= nF(y)^{n-1} - (n-1)F(y)^n \end{aligned} \tag{C.2}$$

The associated probability density function is

$$f_2(y) = n(n-1)(1-F(y))F(y)^{n-2}f(y) \tag{C.3}$$

Again, it can be verified that if $F$ stochastically dominates $G$ and $F_2$ and $G_2$ are the distributions of the second-highest-order statistics of $n$ draws from $F$ and $G$, respectively, then $F_2$ stochastically dominates $G_2$.

## RELATIONSHIPS

Observe that

$$\begin{aligned} F_2^{(n)}(y) &= nF(y)^{n-1} - (n-1)F(y)^n \\ &= nF_1^{(n-1)}(y) - (n-1)F_1^{(n)}(y) \end{aligned}$$

and so

$$f_2^{(n)}(y) = nf_1^{(n-1)}(y) - (n-1)f_1^{(n)}(y) \tag{C.4}$$

This immediately implies that

$$E\left[Y_2^{(n)}\right] = nE\left[Y_1^{(n-1)}\right] - (n-1)E\left[Y_1^{(n)}\right]$$

Also note that

$$\begin{aligned} f_2^{(n)}(y) &= n(n-1)(1-F(y))F(y)^{n-2}f(y) \\ &= n(1-F(y))f_1^{(n-1)}(y) \end{aligned} \tag{C.5}$$

## JOINT AND CONDITIONAL DISTRIBUTIONS OF ORDER STATISTICS

Even though $X_1, X_2, \ldots, X_n$ are independently drawn, the order statistics $Y_1^{(n)}, Y_2^{(n)}, \ldots, Y_n^{(n)}$ are not independent. The joint density of $\mathbf{Y} = \left(Y_1^{(n)}, Y_2^{(n)}, \ldots, Y_n^{(n)}\right)$ is

$$f_{\mathbf{Y}}^{(n)}(y_1, y_2, \ldots, y_n) = n!f(y_1)f(y_2)\ldots f(y_n)$$

if $y_1 \geq y_2 \geq \ldots \geq y_n$ and 0 otherwise. From this it is routine to deduce that the joint density of the first- and second-order statistics is

$$f_{1,2}^{(n)}(y_1, y_2) = n(n-1) f(y_1) f(y_2) F(y_2)^{n-2}$$

if $y_1 \geq y_2$ and 0 otherwise.

The density of $Y_2^{(n)}$ conditional on $Y_1^{(n)} = y$ is

$$\begin{aligned} f_2^{(n)}\left(z \mid Y_1^{(n)} = y\right) &= \frac{f_{1,2}^{(n)}(y,z)}{f_1^{(n)}(y)} \\ &= \frac{n(n-1) f(y) f(z) F(z)^{n-2}}{n f(y) F(y)^{n-1}} \\ &= \frac{(n-1) f(z) F(z)^{n-2}}{F(y)^{n-1}} \end{aligned}$$

if $y \geq z$ and 0 otherwise. On the other hand, the density of $Y_1^{(n-1)}$ conditional on the event $Y_1^{(n-1)} < y$ is

$$\begin{aligned} f_1^{(n-1)}\left(z \mid Y_1^{(n-1)} < y\right) &= \frac{f_1^{(n-1)}(z)}{F_1^{(n-1)}(y)} \\ &= \frac{(n-1) f(z) F(z)^{n-2}}{F(y)^{n-1}} \end{aligned}$$

Thus,

$$f_2^{(n)}\left(\cdot \mid Y_1^{(n)} = y\right) = f_1^{(n-1)}\left(\cdot \mid Y_1^{(n-1)} < y\right) \tag{C.6}$$

## ORDER STATISTICS FOR SYMMETRIC NONINDEPENDENT RANDOM VARIABLES

Suppose that the random variables $X_1, X_2, \ldots, X_n$ are distributed on $[0,\omega]^n$ according to the joint density function $f$ and that $f$ is a symmetric function of its arguments. Let $Y_1^{(n)}, Y_2^{(n)}, \ldots, Y_n^{(n)}$ be the random variables obtained by rearranging $X_1, X_2, \ldots, X_n$ in decreasing order so that $Y_1^{(n)} \geq Y_2^{(n)} \geq \ldots \geq Y_n^{(n)}$.

To determine the joint density of $\mathbf{Y} \equiv (Y_1^{(n)}, Y_2^{(n)}, \ldots, Y_n^{(n)})$, first notice that the support of $\mathbf{Y}$ is the set $\{\mathbf{y} \in [0,\omega]^n : y_1 \geq y_2 \geq \cdots \geq y_n\}$. Since $f$ is symmetric, if $g_{\mathbf{Y}}$ is the joint density of $\mathbf{Y}$, then

$$g_{\mathbf{Y}}(y_1, y_2, \ldots, y_n) = \begin{cases} n! f(y_1, y_2, \ldots, y_n) & \text{if } y_1 \geq y_2 \geq \ldots \geq y_n \\ 0 & \text{otherwise} \end{cases}$$

To see why this is correct, fix a $\mathbf{y} \in [0,\omega]^n$ such that $y_1 \geq y_2 \geq \cdots \geq y_n$. The "event" $\mathbf{Y} = \mathbf{y}$ occurs as long as some permutation of the variables $\mathbf{X}$ equals $\mathbf{y}$. There are $n$! such permutations and since the density of $\mathbf{X}$ is symmetric, the preceding formula results.

## NOTES ON APPENDIX C

A basic discussion of order statistics can be found in any standard text on probability theory. The books by David (1969) and Arnold, Balakrishnan, and Nagaraja (1992) contain more specialized treatments.

We caution the reader that the terminology and notation in these and other specialized texts are different from that employed in auction theory. We have denoted the *highest* of $X_1, X_2, \ldots, X_n$ by $Y_1^{(n)}$, the *second-highest* by $Y_2^{(n)}$, and so on. In statistics it is conventional to call the *smallest* of $X_1, X_2, \ldots, X_n$ as the "first" order statistic, the *second-smallest* as the "second" order statistic and so on, and these are denoted, respectively, by $Y_1^{(n)}, Y_2^{(n)}$, and so on (or some similar notation); in other words, the order is reversed. The definitions and results in statistics texts like those mentioned here need to be read with some care to account for these differences in terminology and notation.