Notes on Order statistics

[Based on Krishna, 2002, Appendix C.]

Consider n independent random draws from the same distribution on $[\underline{v}, \overline{v}]$ represented by F . We order the valuations drawn such that

$$V_{[n]} \geq V_{[n-1]} \geq \ldots\ldots \geq V_{[1]}$$

$V_{[k]}$ is referred to as the kth-order statistic. We shall occasionally refer to this as $V_{[k,n]}$ to capture that it is the kth-order statistic in a sample of n. Below, it will be useful to know the distributions of $V_{[n]}$ (highest) and $V_{[n-1]}$ (second-highest) the relationship between their distributions.

Distribution of the highest-order statistic $V_{[n]}$

Let $F_{[n,n]}(v)$ denote the distribution function of $V_{[n]}$ (the highest of n draws from F (v)). Now, $F_{[n,n]}(v)$ is just the probability that all the n draws are less than or equal to v. Hence,

$$F_{[n,n]}(v) = (F(v))^n = F^n(v) \tag{1}$$

It follows that the density, $f_{[n,n]}(v)$, of $V_{[n]}$ is

$$\frac{dF_{[n,n]}(v)}{dv} = \frac{dF^n(v)}{dv} = nf(v)F(v)^{n-1} \tag{2}$$

Combinatorially, this may be thought of as the n ways that one value could be v, while the other are lower. The expected value of $V_{[n]}$ is:

$$E(V_{[n]}) = \int_{\underline{v}}^{\overline{v}} vf_{n,n}(v)dv = \int_{\underline{v}}^{\overline{v}} nvf(v)F(v)^{n-1}dv \tag{3}$$

Distribution of the second-highest order statistic $V_{[n-1]}$

Let $F_{n-1,n}(v)$ denote the distribution function of $V_{[n-1]}$ (the second-highest of n draws from F (v)).

Now, $F_{n-1,n}(v)$ is the probability that $V_{[n-1]}$ is less than or equal to v. This, however, is the union of two disjoint events: a) all n values are less than or equal to v, and b) $(n-1)$ values are less than or equal to v and one value is greater than v. There are n ways in which ii) can occur, hence

$$F_{n-1,n}(v) = \underbrace{F(v)^n}_{i} + \underbrace{nF(v)^{n-1}(1 - F(v))}_{ii} \tag{4}$$

We can rewrite this as

$$F_{n-1,n}(v) = nF(v)^{n-1} - (n-1)F(v)^n \tag{5}$$

Therefore, the density $f_{n-1,n}(v)$ of $V_{[n-1]}$ is

$$f_{n-1,n}(v) = n(n-1)f(v)(1-F(v))F(v)^{n-2} \tag{6}$$

If we write this density as $nf(v) \times (n-1)(1-F(v)) \times F^{n-2}(v)$ it may be thought of as the n ways that one value could be v combined with the n - 1 ways that one could be higher and the remaining n - 2 lower.

The expected value of $V_{[n-1]}$ is

$$\begin{aligned} E(V_{[n-1]}) &= \int_{\underline{v}}^{\overline{v}} v f_{n-1,n}(v)dv \\ &= \int_{\underline{v}}^{\overline{v}} n(n-1)vf(v)(1-F(v))F(v)^{n-2}\, dv \end{aligned} \tag{7}$$

Relationships

We can rewrite $F_{n-1,n}(v)$ as

$$F_{n-1,n}(v) = n\left[F(v)\right]^{n-1} - (n-1)\left[F(v)\right]^n = nF_{n-1:n-1}(v) - (n-1)F_{n:n}(v)$$

from which it follows immediately that

$$f_{n-1,n}(v) = nf_{n-1,n-1}(v) - (n-1)f_{n,n}(v)$$

and

$$E(V_{[n-1,n]}) = nE(V_{[n-1,n-1]}) - (n-1)E(V_{[n,n]})$$

We also note that

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

Hence,

$$f_{n-1,n}(v) = n(1-F(v))f_{n-1,n-1}(v)$$

since

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

Joint density of $V_{[n,n]}$ and $V_{[n-1,n]}$

Let $V_n \geq V_{n-1} \geq \ldots\ldots \geq V_1$. Even if the draws are independent, the order statistics are not independent! Then the joint density of V_n and V_{n-1} (the highest and the second-highest among n) is given by

$$f_{n,n-1:n}(v_n, v_{n-1}) = n(n-1)f(v_n)f(v_{n-1})F(v_{n-1})^{n-2}$$

If $v_n > v_{n-1}$, equals zero otherwise.

Density of $V_{[n-1,n]}$ conditional on $V_{[n,n]} = y$ is (for y>z)

$$\begin{aligned} f_{n-1,n}(z \mid V_{[n,n]} = y) = \frac{f_{n,n-1:n}(y,z)}{f_{n,n}(y)} \\ = \frac{n(n-1)f(y)f(z)F^{n-2}(z)}{nf(y)F^{n-1}(y)} \\ = \frac{(n-1)f(z)F^{n-2}(z)}{F^{n-1}(y)} \end{aligned} \quad (8)$$

Also, the density of $V_{[n-1,n-1]}$ conditional on $V_{[n-1,n-1]} < y$ is given by

$$\begin{aligned} f_{n-1,n-1}\left(z \middle| V_{[n-1,n-1]} < y\right) = \frac{f_{n-1,n-1}(z)}{F_{n-1,n-1}(y)} \\ = \frac{(n-1)f(z)F(z)^{n-2}}{F(y)^{n-1}} \end{aligned}$$

and we conclude that

$$f_{n-1,n}(\cdot \mid V_{[n,n]} = y) = f_{n-1,n-1}(\cdot \mid V_{[n-1,n-1]} < y)$$

In words: The distribution of the second-highest order statistic in a sample of n conditioned on the highest-order statistic (in the same sample) being y is equivalent to the distribution of the highest-order statistic in a sample of $(n-1)$ conditioned on it being less than y.

From the above it follows that:

$$E\left[V_{n-1,n} \middle| V_{n,n} = y\right] = E\left[V_{n-1,n-1} \middle| V_{n-1,n-1} < y\right]$$