# Order Statistics

Frank R. Kschischang

January 5, 2019

## 1 Definition

The $r$th order statistic $X_{(r)}$ of a sample of $n$ random variables $X_1, \ldots, X_n$ is equal to its $r$th smallest value. Thus $X_{(1)}$ denotes $\min\{X_1, \ldots, X_n\}$ (the minimum of the $X_i$'s), $X_{(2)}$ denotes $\min\left(\{X_1, \ldots, X_n\} \setminus \{X_{(1)}\}\right)$ (the second minimum), and in general,

$$X_{(r)} = \min\left(\{X_1, \ldots, X_n\} \setminus \{X_{(1)}, \ldots, X_{(r-1)}\}\right).$$

Thus $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

## 2 Distribution and Density Functions

Fix $x \in \mathbb{R}$ and for any random variable $X$, designate the event $X \leq x$ a "success." At the crux of understanding order statistics is the following observation: $X_{(r)}$ is a success if and only if at least $r$ of $X_1, \ldots, X_n$ are successes.

Suppose that the $X_1, \ldots, X_n$ are independent and identically distributed with cumulative distribution function $F(x)$. The probability of success for each $X_i$ is then given by $F(x)$, and the probability that $X_{(r)}$ is a success, i.e., the cumulative distribution function of $X_{(r)}$, is given as the binomial probability

$$F_{(r)}(x) = \sum_{k=r}^{n} \binom{n}{k} (F(x))^k (1 - F(x))^{n-k} = 1 - \sum_{k=0}^{r-1} \binom{n}{k} (F(x))^k (1 - F(x))^{n-k}.$$

Thus, for example,

$$F_{(1)}(x) = 1 - (1 - F(x))^n \text{ and } F_{(n)}(x) = (F(x))^n.$$

If $F(x)$ has derivative $f(x)$, then we can compute the probability density function for

$X_{(r)}$ by differentiating $F_{(r)}(x)$. We get

$$\begin{aligned}
f_{(r)}(x) &= \frac{\mathrm{d}}{\mathrm{d}x} F_{(r)}(x) \\
&= \frac{\mathrm{d}}{\mathrm{d}x} \sum_{k=r}^{n} \binom{n}{k} (F(x))^k (1-F(x))^{n-k} \\
&\overset{(a)}{=} \frac{\mathrm{d}}{\mathrm{d}x} \left( \sum_{k=r}^{n-1} \binom{n}{k} (F(x))^k (1-F(x))^{n-k} + (F(x))^n \right) \\
&= \sum_{k=r}^{n-1} \binom{n}{k} k (F(x))^{k-1} f(x) (1-F(x))^{n-k} + n(F(x))^{n-1} \\
&\quad + \sum_{k=r}^{n-1} \binom{n}{k} (F(x))^k (n-k)(1-F(x))^{n-k-1}(-f(x)) \\
&= \sum_{k=r}^{n} \frac{n!}{(k-1)!(n-k)!} (F(x))^{k-1} (1-F(x))^{n-k} f(x) \\
&\quad - \sum_{k=r}^{n-1} \frac{n!}{k!(n-k-1)!} (F(x))^{k} (1-F(x))^{n-k-1} f(x) \\
&\overset{(b)}{=} \sum_{k=r}^{n} \frac{n!}{(k-1)!(n-k)!} (F(x))^{k-1} (1-F(x))^{n-k} f(x) \\
&\quad - \sum_{k'=r+1}^{n} \frac{n!}{(k'-1)!(n-k')!} (F(x))^{k'-1} (1-F(x))^{n-k'} f(x) \\
&\overset{(c)}{=} \frac{n!}{(r-1)!(n-r)!} (F(x))^{r-1} (1-F(x))^{n-r} f(x) \\
&= r\binom{n}{r} (F(x))^{r-1} (1-F(x))^{n-r} f(x).
\end{aligned}$$

In (a) we isolated the last term in the sum to avoid negative exponents when differentiating, in (b) we re-indexed the second summation by substituting $k' = k+1$, and we get (c) by observing the miraculous cancellation of all but one term of the first summation by the terms of the second summation. This derivation is the centerpiece of this note. Thus, for example,

$$f_{(1)}(x) = n(1-F(x))^{n-1} f(x) \text{ and } f_{(n)}(x) = n(F(x))^{n-1} f(x).$$

in agreement with the direct differentiation of their cumulative distribution functions.

To illustrate these formulas in action, suppose that $X_1, \ldots, X_5$ are uniformly distributed on the unit interval $I = [0,1]$, so that $F(x) = x$ and $f(x) = 1$ for $x \in I$. Then, for $x \in I$, we have

$$
\begin{aligned}
f_{(1)}(x) &= 5(1-x)^4 \\
f_{(2)}(x) &= 20x(1-x)^3 \\
f_{(3)}(x) &= 30x^2(1-x)^2 \\
f_{(4)}(x) &= 20x^3(1-x) \\
f_{(5)}(x) &= 5x^4,
\end{aligned}
$$

all of them being examples of *beta distributions*. From these expressions we find (the intuitively plausible) expected values

$$
E[X_{(1)}] = \frac{1}{6},\ E[X_{(2)}] = \frac{1}{3},\ E[X_{(3)}] = \frac{1}{2},\ E[X_{(4)}] = \frac{2}{3},\ E[X_{(5)}] = \frac{5}{6}
$$

and variances

$$
\begin{aligned}
\mathrm{VAR}[X_{(1)}] &= \frac{5}{252} \approx 0.0198, \\
\mathrm{VAR}[X_{(2)}] &= \frac{2}{63} \approx 0.0317, \\
\mathrm{VAR}[X_{(3)}] &= \frac{1}{28} \approx 0.0357, \\
\mathrm{VAR}[X_{(4)}] &= \frac{2}{63} \approx 0.0317, \\
\mathrm{VAR}[X_{(5)}] &= \frac{5}{252} \approx 0.0198.
\end{aligned}
$$

Evidently, in this case, the extremal order statistics (minimum and maximum) have the least variance, as their distributions become squeezed by the finite support over which the $X_i$'s are defined.

When $X_1, \ldots, X_5$ are normally distributed with zero mean and unit variance, we get

$$
E[X_{(1)}] \approx -1.16,\ E[X_{(2)}] \approx -0.495,\ E[X_{(3)}] = 0, E[X_{(4)}] \approx 0.495,\ E[X_{(5)}] \approx 1.16;
$$

these values are called *rankits* and they are used for the abscissa of a so-called *normal plot*. The variances are

$$
\begin{aligned}
\mathrm{VAR}[X_{(1)}] &\approx 0.448, \\
\mathrm{VAR}[X_{(2)}] &\approx 0.312, \\
\mathrm{VAR}[X_{(3)}] &\approx 0.287, \\
\mathrm{VAR}[X_{(4)}] &\approx 0.312, \\
\mathrm{VAR}[X_{(5)}] &\approx 0.448.
\end{aligned}
$$

Now the extremal order statistics have the most variance.

It is interesting to consider the limit distributions of the extremal order statistics as the number of samples $n$ becomes large. As $n \to \infty$, we get

$$\lim_{n\to\infty} F_{(1)}(x) = \lim_{n\to\infty} 1 - (1 - F(x))^n = \begin{cases} 0 & \text{if } F(x) = 0, \\ 1 & \text{if } F(x) > 0; \end{cases}$$

$$\lim_{n\to\infty} F_{(n)}(x) = \lim_{n\to\infty} (F(x))^n = \begin{cases} 1 & \text{if } F(x) = 1, \\ 0 & \text{if } F(x) < 1. \end{cases}$$

In case $F(x) = 0$ is not achieved by any $x$, the minimum $X_{(1)}$ does not approach a valid distribution as $n \to \infty$ (in effect the density approaches a Dirac delta at $-\infty$). Likewise, in case $F(x) = 1$ is not achieved by any $x$, the maximum $X_{(n)}$ also does not approach a valid distribution as $n \to \infty$ (in effect the density approaches a Dirac delta at $+\infty$).

# 3 Joint Distributions

A remarkable result, from which all joint densities (including all marginal densities) of the order statistics can in principle be derived via marginalization, is the following. Again assuming that $X_1, \ldots, X_n$ are independent and identically distributed with density $f(x)$, the joint density of the order statistics is given by

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = \begin{cases} n! f(x_1) f(x_2) \ldots f(x_n) & \text{if } x_1 < x_2 < \cdots < x_n; \\ 0 & \text{otherwise.} \end{cases}$$

To see this, recall the well known result that if $g$ smoothly transforms the random vector $X = (X_1, \ldots, X_n)$ to the vector $Y = (Y_1, \ldots, Y_n) = g(X)$, then the probability density at a point $(y_1, \ldots, y_n)$ in the range of $g$ is given by

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \sum_{(x_1,\ldots,x_n) \in g^{-1}(y_1,\ldots,y_n)} \frac{f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)}{|J(x_1, \ldots, x_n)|},$$

where

$$g^{-1}(y_1, \ldots, y_n) = \{(x_1, \ldots, x_n) : g(x_1, \ldots, x_n) = (y_1, \ldots, y_n)\}$$

is the set of points in the domain of $g$ that map to $(y_1, \ldots, y_n)$ and the Jacobian $J(x_1, \ldots, x_n)$ is given as

$$\det \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \cdots & \frac{\partial y_n}{\partial x_n} \end{bmatrix},$$

where $(y_1, \ldots, y_n) = g(x_1, \ldots, x_n)$. Intuitively, a point $(x_1, \ldots, x_n)$ in the domain of $g$ contributes an inverse-Jacobian-scaled version of the density $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ to the density at $g(x_1, \ldots, x_n)$ in the range of $g$.

From now on, let $g$ be the mapping that takes a vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ to $g(x) = (x_{(1)}, \ldots, x_{(n)})$. This function is certainly not one-to-one (or smooth). However, it is piecewise one-to-one and smooth when restricted to certain regions of $\mathbb{R}^n$ defined as follows. For every fixed permutation $\pi$ on $\{1, 2, \ldots, n\}$ define the region

$$\mathcal{R}_\pi = \{x \in \mathbb{R}^n : x_{\pi(1)} < x_{\pi(2)} < \cdots < x_{\pi(n)}\}.$$

For example, if $n = 2$, there are two possible permutations on $\{1, 2\}$. If $\pi = \text{id}$ is the identity permutation, then $R_{\text{id}} = \{(x_1, x_2) : x_1 < x_2\}$, a half-plane. On the other hand, if $\pi$ is the permutation such that $\pi(1) = 2$, then $R_\pi = \{(x_1, x_2) : x_2 < x_1\}$, the complementary half-plane (neglecting the zero-measure boundary set $\{(x_1, x_2) : x_1 = x_2\}$).

In general, the different $R_\pi$ regions partition $\mathbb{R}^n$ into $n!$ disjoint regions (neglecting zero-measure boundary sets in which one or more of the coordinates are equal).

Now let $T_\pi : \mathcal{R}_\pi \to \mathcal{R}_{\text{id}}$ denote the map that takes $(x_1, \ldots, x_n) \in \mathcal{R}_\pi$ to $(x_{\pi(1)}, \ldots, x_{\pi(n)}) \in \mathcal{R}_{\text{id}}$. This map is linear and one-to-one, and the corresponding Jacobian (being the absolute value of the determinant of a permutation matrix) has unit magnitude.

Returning now to the map $g$, observe that $g$ can be written as

$$g(x) = T_\pi(x) \text{ if } x \in \mathcal{R}_\pi$$

for any $x \in \bigcup_\pi \mathcal{R}_\pi$. Every point $y = (y_1, \ldots, y_n) \in \mathcal{R}_{\text{id}}$ has exactly $n!$ inverse images under $g$, one in each region $\mathcal{R}_\pi$. The joint probability density of $X_1, \ldots, X_n$ at each such point is identical, given by $f(y_1)f(y_2)\cdots f(y_n)$. Since the Jacobian corresponding to $g$ at each such point has unit magnitude, the Jacobian transformation rule gives us

$$f(y_1, \ldots, y_n) = \sum_\pi f(y_1)f(y_2)\cdots f(y_n) = n!f(y_1)f(y_2)\cdots f(y_n), \quad y_1 < y_2 < \cdots < y_n.$$

For example, when $n = 3$, we have

$$\begin{aligned}
f_{X_{(1)},X_{(2)}}(x_1, x_2) &= \int_{x_2}^{\infty} 6f(x_1)f(x_2)f(x_3)\,\mathrm{d}x_3 = 6f(x_1)f(x_2)(1 - F(x_2)),\\
f_{X_{(1)},X_{(3)}}(x_1, x_3) &= \int_{x_1}^{x_3} 6f(x_1)f(x_2)f(x_3)\,\mathrm{d}x_2 = 6f(x_1)f(x_3)(F(x_3) - F(x_1)),\\
f_{X_{(2)},X_{(3)}}(x_2, x_3) &= \int_{-\infty}^{x_2} 6f(x_1)f(x_2)f(x_3)\,\mathrm{d}x_1 = 6f(x_2)f(x_3)F(x_2).
\end{aligned}$$

We can marginalize these further, for example recovering

$$\begin{aligned}
f_{X(1)}(x_1) &= \int_{x_1}^{\infty} 6f(x_1)f(x_2)(1 - F(x_2))\,\mathrm{d}x_2\\
&= 6f(x_1)(1 - F(x_1) - \frac{1}{2}(1 - (F(x_1))^2))\\
&= 3(1 - F(x_1))^2 f(x_1),
\end{aligned}$$

in agreement with

$$\begin{aligned}
f_{X(1)}(x_1) &= \int_{x_1}^{\infty} 6f(x_1)f(x_3)(F(x_3) - F(x_1))\,\mathrm{d}x_3\\
&= 6f(x_1)\left(\frac{1}{2}(1 - (F(x_1))^2) - F(x_1)(1 - F(x_1))\right)\\
&= 3(1 - F(x_1))^2 f(x_1).
\end{aligned}$$