Lecture 2: Order Statistics

Remark 2.1 Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$. Common measures of center are the mean and median. The *sample median* is an example of an order statistic. Order statistics are commonly denoted $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ arranged in increasing order.

Theorem 2.1 *Let the population distribution be continuous and its pdf is denoted by $f(x)$ on the support $\mathcal{S} = (a, b)$ for $-\infty < a < b < \infty$. For a random sample $X_1, X_2, \ldots, X_n$, let $Y_1 < Y_2 < \ldots < Y_n$ denote the n order statistics arranged in increasing order. Then, the joint pdf of $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ is given by*

$$pdf_{\boldsymbol{Y}}(y_1, \ldots, y_n) = \begin{cases} n! f(y_1) \cdots f(y_n), & a < y_1 < \ldots < y_n < b \\ 0, & \textit{elsewhere.} \end{cases}$$

Proof. Let $u : \mathbb{X} \to \mathbb{Y}$ be defined as $u(X_1, \ldots, X_n) = (Y_1, \ldots, Y_n)$ denoting the transformation of a random sample into the n order statistics where

$$\begin{aligned} \mathbb{X} &= \{(x_1, \ldots, x_n) : f(x_i) > 0 \ \forall i, \text{ and } x_1, \ldots, x_n \text{ are distinct real numbers}\} \\ \mathbb{Y} &= \{(y_1, \ldots, y_n) : f(y_i) > 0 \ \forall i, \text{ and } a < y_1 < \cdots < y_n < b\}. \end{aligned}$$

Note that u is an $n!$-to-1 mapping. If Π denotes the group of all permutations of $\{1, \ldots, n\}$, its order is $n!$ and for any $\boldsymbol{\pi} \in \Pi$, we know $(x_{\pi_1}, \ldots, x_{\pi_n}) \in \mathbb{X}$. Notably, $\bigsqcup_{\boldsymbol{\pi} \in \Pi} (x_{\pi_1}, \ldots, x_{\pi_n}) = \mathbb{X}$ (recall $\sqcup$ means "disjoint union"). Then, we can apply the approach learned in Example 1.6 to find the $\text{pdf}_{\boldsymbol{Y}}$. First, it is easy to see that the Jacobian for u for each $\boldsymbol{\pi} \in \Pi$ is ± 1. The easiest way to see this is to consider the special case $\boldsymbol{\pi}^*$ such that $x_{\pi_1^*} = y_1, \ldots, x_{\pi_n^*} = y_n$; here $J_{\boldsymbol{\pi}^*} = 1$. Then, we know for $a < y_1 < \ldots < y_n < b$

$$\begin{aligned} \text{pdf}_{\boldsymbol{Y}}(y_1, \ldots, y_n) &= \sum_{\boldsymbol{\pi} \in \Pi} \text{pdf}_{X_1, \ldots, X_n}((u^{\pi})^{-1}(y)) \cdot |J_{\boldsymbol{\pi}}| \\ &= \sum_{\boldsymbol{\pi} \in \Pi} \text{pdf}_{X_1, \ldots, X_n}(y_{\pi_1^{-1}}, \ldots, y_{\pi_n^{-1}}) \cdot |\pm 1| \\ &= \sum_{\boldsymbol{\pi} \in \Pi} f(y_{\pi_1^{-1}}) \cdots f(y_{\pi_n^{-1}}) && \text{(independence)} \\ &= \sum_{\boldsymbol{\pi} \in \Pi} f(y_1) \cdots f(y_n) && \text{(reordering)} \\ &= n! f(y_1) \cdots f(y_n) \end{aligned}$$

and $\text{pdf}_{\boldsymbol{Y}}(y_1, \ldots, y_n) = 0$ elsewhere. $\square$

Example 2.1 Let $X_1, X_2, X_3 \overset{iid}{\sim} \text{Exp}(1)$ where $f(x) = \exp(-x)$, $x \geq 0$. Define order statistics $Y_1 < Y_2 < Y_3$. Find the joint pdf of $\boldsymbol{Y} = (Y_1, Y_2, Y_3)$.

Solution. Let $u : \mathbb{X} \to \mathbb{Y}$ be defined as $u(X_1, X_2, X_3) = (Y_1, Y_2, Y_3)$ where

$$\begin{aligned}
\mathbb{X} &= \{(x_1, x_2, x_3) : x_i \geq 0 \ \forall i, \text{ and } x_1 \neq x_2,\ x_2 \neq x_3,\ x_3 \neq x_1\} \\
\mathbb{Y} &= \{(y_1, y_2, y_3) : y_i \geq 0 \ \forall i, \text{ and } 0 \leq y_1 < \cdots < y_n\}.
\end{aligned}$$

Then, by Theorem 2.1, we know

$$\begin{aligned}
\text{pdf}_{Y_1,Y_2,Y_3}(y_1, y_2, y_3) &= 3! \cdot \exp(-y_1)\exp(-y_2)\exp(-y_3) \cdot \mathbf{1}_{(0<y_1<y_2<y_3)} \\
&= 6 \cdot \exp[-(y_1 + y_2 + y_3)] \cdot \mathbf{1}_{(0<y_1<y_2<y_3)}.
\end{aligned}$$

■

Theorem 2.2 *Suppose the population distribution is continuous with its probability density function $f(x)$ and its cumulative distribution function $F(x)$. The marginal distribution of order statistics $X_{(1)} < \cdots < X_{(n)}$ of a random sample $X_1, \ldots, X_n$ are given as follows:*

(a) The pdf of $X_{(r)}$ $(1 \leq r \leq n)$:

$$\text{pdf}_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!}[F(x)]^{r-1} f(x)[1 - F(x)]^{n-r}.$$

(b) The joint pdf of $(X_{(r)}, X_{(s)})$ $(1 \leq r < s \leq n)$:

$$\text{pdf}_{(X_{(r)},X_{(s)})}(x, y) = \frac{n!}{(r-1)!(s-1-r)!(n-s)!}[F(x)]^{r-1} f(x)[F(y) - F(x)]^{s-1-r} f(y)[1 - F(y)]^{n-s}.$$

Proof. First, let $f_{\boldsymbol{Y}}(y_1, \ldots, y_n) = n! f(y_1) \cdots f(y_n) \mathbf{1}_{(y_1<\cdots<y_n)}$ denote the joint pdf of $\boldsymbol{Y} = (X_{(1)}, \ldots, X_{(n)})$ derived in Theorem 2.1. Next, let

$$f_r(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{Y}}(y_1, \ldots, x, \ldots, y_n)\, dy_1 \cdots dy_{r-1}\, dy_{r+1} \cdots dy_n, \tag{1}$$

$$f_{r,s}(x, y) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{Y}}(y_1, \ldots, x, \ldots, y, \ldots, y_n)\, dy_1 \cdots dy_{r-1}\, dy_{r+1} \cdots dy_{s-1}\, dy_{s+1} \cdots dy_n \tag{2}$$

denote the marginal pdfs of $X_{(r)}$ and $(X_{(r)}, X_{(s)})$, respectively.

(a) We know that within the integrand $f_{\boldsymbol{Y}}(y_1, \ldots, y_n)$, the products $[f(y_1) \cdots f(y_{r-1})]$ and $[f(y_{r+1}) \cdots f(y_n)]$ do not vary with the permutation of $y_1, \ldots, y_{r-1}$ and $y_{r+1}, \ldots, y_n$, respectively. From (1)

$$\begin{aligned}
f_r(x) &= n! \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_1), \ldots, f(y_{r-1}) f(x) f(y_{r+1}), \ldots, f(y_n) \mathbf{1}_{(y_1<\cdots<y_{r-1}<x<y_{r+1}<\cdots<y_n)}\, dy_1 \cdots dy_{r-1} dy_{r+1} \cdots dy_n, \\
&= n! f(x) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_1), \ldots, f(y_{r-1}) \mathbf{1}_{(y_1<\cdots<y_{r-1}<x)} dy_1 \cdots dy_{r-1} \right\} \\
&\quad \times f(y_{r+1}), \ldots, f(y_n) \mathbf{1}_{(x<y_{r+1}<\cdots<y_n)} dy_{r+1} \cdots dy_n.
\end{aligned}$$

First, recall that $f(\cdot)$ is the density function corresponding to the *population distribution*, not an order statistic. Next, we focus on simplifying the expression in the {brackets}. There exists two different approaches:

Approach (1) If we relax the ordering and just assume $y_i < x,\ i = 1, 2, \ldots, r-1$, then the integral becomes

$$\begin{aligned}
&\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_1), \ldots, f(y_{r-1}) \mathbf{1}_{(y_i < x,\ i=1,2,\ldots,r-1)} dy_1 \cdots dy_{r-1} \\
&\quad = \int_{-\infty}^{x} \cdots \int_{-\infty}^{x} f(y_1) \cdots f(y_{r-1})\, dy_1 \cdots dy_{r-1} \\
&\quad = [F(x)]^{r-1}
\end{aligned}$$

However, we are only interested in the particular ordering $y_1 < \cdots < y_{r-1} < x$. Therefore, among the $(r-1)!$ ways of permuting $y_1, \ldots, y_{r-1}$, we are only interested in one. Hence,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_1), \ldots, f(y_{r-1}) \mathbf{1}_{(y_1 < \cdots < y_{r-1} < x)} dy_1 \cdots dy_{r-1} \\
&\quad = \frac{1}{(r-1)!} [F(x)]^{r-1}.
\end{aligned}$$

Approach (2) Directly solve the integral:

$$\begin{aligned}
&\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_1), \ldots, f(y_{r-1}) \mathbf{1}_{(y_1 < \cdots < y_{r-1} < x)} dy_1 \cdots dy_{r-1} \\
&\quad = \int_{-\infty}^{\infty} f(y_{r-1}) \cdots \int_{-\infty}^{\infty} f(y_2) \int_{-\infty}^{\infty} f(y_1) \mathbf{1}_{(y_1 < y_2 < \cdots < y_{r-1} < x)} dy_1 \cdots dy_{r-1} \\
&\quad = \int_{-\infty}^{\infty} f(y_{r-1}) \cdots \int_{-\infty}^{\infty} f(y_2) \mathbf{1}_{(y_2 < \cdots < y_{r-1} < x)} \int_{-\infty}^{y_2} f(y_1) dy_1 \cdots dy_{r-1} \\
&\quad = \int_{-\infty}^{x} f(y_{r-1}) \cdots \int_{-\infty}^{y_3} f(y_2) \int_{-\infty}^{y_2} f(y_1) dy_1 dy_2 \cdots dy_{r-1} \\
&\quad = \int_{-\infty}^{x} f(y_{r-1}) \cdots \int_{-\infty}^{y_3} f(y_2) F(y_2) dy_2 dy_3 \cdots dy_{r-1} \\
&\quad = \int_{-\infty}^{x} f(y_{r-1}) \cdots \int_{-\infty}^{F(y_3)} u\, du\, dy_3 \cdots dy_{r-1} \quad [u = F(y_2),\ du = f(y_2) dy_2] \\
&\quad = \int_{-\infty}^{x} f(y_{r-1}) \cdots \int_{-\infty}^{y_4} \frac{[F(y_3)]^2}{2} dy_3 \cdots dy_{r-1} \\
&\quad\ \ \vdots \\
&\quad = \frac{1}{(r-1)!} [F(x)]^{r-1}.
\end{aligned}$$

Plugging this information back into the expression for the pdf of $X_{(r)}$, we have

$$f_r(x) = n! f(x) \left\{ \frac{1}{(r-1)!} [F(x)]^{r-1} \right\} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_{r+1}), \ldots, f(y_n) \mathbf{1}_{(x < y_{r+1} < \cdots < y_n)} dy_{r+1} \cdots dy_n.$$

Then, similar to the logic applied before, it is easy to show

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_{r+1}), \ldots, f(y_n) \mathbf{1}_{(x < y_{r+1} < \cdots < y_n)} dy_{r+1} \cdots dy_n = \frac{1}{(n-r)!} [1 - F(x)]^{n-r}.$$

Therefore, we know

$$f_r(x) = n!f(x)\left\{\frac{1}{(r-1)!}[F(x)]^{r-1}\right\}\frac{1}{(n-r)!}[1-F(x)]^{n-r}.$$

(For an alternative proof, see the proof of Theorem 5.4.4. of Casella and Burger 2nd edition [here])

(b) Following the same steps as in (a) starting with (2), it can be shown

$$f_{r,s}(x,y) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}[F(x)]^{r-1}f(x)[F(y)-F(x)]^{s-r-1}f(y)[1-F(y)]^{n-s}.$$

□

Example 2.2 Suppose a random sample $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Unif}(0,1)$. Find the joint distribution of the order statistics $X_{(1)} < \cdots < X_{(n)}$, and show $X_{(r)} \sim \text{Beta}(r, n-r+1)$.

Proof. The probability density function and cumulative distribution function of Unif$(0,1)$ are:

$$f(x) = \mathbf{1}_{x\in(0,1)}, \quad F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x \le 1 \\ 1, & x > 1 \end{cases}.$$

From Theorem 2.1, the joint probability density function of $Y = (X_{(1)}, \cdots, X_{(n)})^{\mathrm{T}}$ is

$$\text{pdf}_Y(y_1, \cdots, y_n) = n!\mathbf{1}_{(0<y_1<\cdots<y_n<1)},$$

and by part (a) of Theorem 2.2, the marginal distribution of $X_{(r)}$ is given by

$$\text{pdf}_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!}x^{r-1}(1-x)^{n-r}\mathbf{1}_{(0<x<1)}.$$

Therefore, $X_{(r)} \sim \text{Beta}(r, n-r+1)$ □

Definition 2.1

(1) Range:

$$X_{(n)} - X_{(1)}$$

(2) Midrange:

$$\frac{X_{(1)} + X_{(n)}}{2}$$