[6.5] Order Statistics

Consider a random sample of data $X_1, \ldots, X_n$.

Often it is useful to consider the "ordered" random sample, denoted by:

$$X_{1:n}, X_{2:n}, \cdots, X_{n:n}$$

where $$X_{1:n} \leqslant X_{2:n} \leqslant \cdots \leqslant X_{n:n}$$

The joint dist of $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ is not the same as the joint density of the unordered variables

We will also use Y_{i} as a label for the order statistics as well. The next couple of items illustrate that. (It's only for ease of writing)

We will consider transformations that transform from $x_1, x_2, \cdots, x_n$ to $y_1, y_2, \cdots, y_n$ where

$$y_1 \le y_2 \le \cdots \le y_n$$

For example,

$y_1 = u_1(x_1, \cdots, x_n) = \min(x_1, \cdots, x_n)$

$y_2 = u_2(x_1, \cdot, x_n) = \text{second smallest}(x_1, x_n)$

$\vdots$

$y_i = u_i(x_1, , x_n) = i^{th} \text{ smallest}(x_1, x_n)$

$\vdots$

$y_n = u_n(x_1, x_n) = \max(x_1, x_n)$

$\boxed{\text{Thm}}$ If $X_1, X_2, \ldots, X_n$ is a RS from a population with continuous pdf $f(x)$, then the joint pdf of the order statistics, $Y_1, Y_2, \ldots, Y_n$ is

$$g(y_1, y_2, \ldots, y_n) = n! f(y_1) f(y_2) \cdots f(y_n)$$

$\boxed{\text{EX:}}$ Suppose X_1, X_2, X_3 represents a RS from a population with pdf

$$f(x) = 2x I_{(0,1)}(x)$$

$$g(y_1, y_2, y_3) = 3!(2y_1)(2y_2)(2y_3), 0 < y_1 < y_2 < y_3 < 1$$

$$= 48 y_1 y_2 y_3, 0 < y_1 < y_2 < y_3 < 1$$

Marginal of Y_1?

$$g_1(y_1) = \int_{y_1}^{1} \int_{y_2}^{1} 48 y_1 y_2 y_3 \, dy_3 \, dy_2$$

= and then a miracle occurs.

$$= 6 y_1 (1 - y_1^2)^2 I_{(0,1)}(y_1)$$

= finish this problem on your own!

<u>EX:</u> X is cont $f(x) > 0$ on $a < x < b$
(a can be $-\infty$, and b may be ∞).

Let $n = 3$.

$$g_1 = \int_{y_1}^{b} \int_{y_2}^{b} 3!\, f(y_1) f(y_2) f(y_3)\, dy_3\, dy_2$$

$$= \int_{y_1}^{b} 3!\, f(y_1) f(y_2) \left[\int_{y_2}^{b} f(y_3)\, dy_3 \right] dy_2$$

$$= \int_{y_1}^{b} 3!\, f(y_1) f(y_2) \Big[F(y_3) \Big]_{y_2}^{b} dy_2$$

$$= \int_{y_1}^{b} 3!\, f(y_1) f(y_2) \Big[\underbrace{F(b)}_{1} - F(y_2) \Big] dy_2$$

$$= 3!\, f(y_1) \int_{y_1}^{b} f(y_2) \big[1 - F(y_2) \big] dy_2$$

$u = 1 - F(y_2)$

$du = -f(y_2)\, dy_2$

$$= -3!\, f(y_1) \int_{1-F(y_1)}^{1-F(b)=0} u\, du = 3!\, f(y_1) \int_{0}^{1-F(y_1)} u\, du$$

$$= 3!\, f(y_1) \left. \frac{u^2}{2} \right]_0^{1-F(y_1)}$$

$$= \frac{3!\, f(y_1) \big[1 - F(y_1)\big]^2}{2}$$

$$= 3\, f(y_1) \big(1 - F(y_1)\big)^2 I_{(a,b)}(y_1)$$

$$= 3\, (2y_1)\big(1 - y_1^2\big)^2 I_{(0,1)}(y_1)$$

Thm: Suppose $X_1, \dots, X_n$ denotes a RS of size n from a cont. pdf $f(x)$, $f(x) > 0$ for $a < x < b$. Then the pdf of the k^{th} order stat Y_k is given by

$$g_k(y_k) = \frac{n!}{(k-1)!(n-k)!}\left[F(y_k)\right]^{k-1}\left[1-F(y_k)\right]^{n-k} f(y_k)\, I_{(a,b)}(y_k)$$

A mnemonic to memorize:

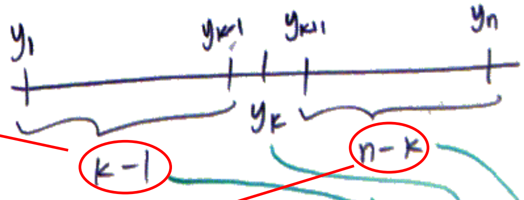
To have $Y_k = y_k$, one must have

$k-1$ observations less than y_k

one observation at y_k

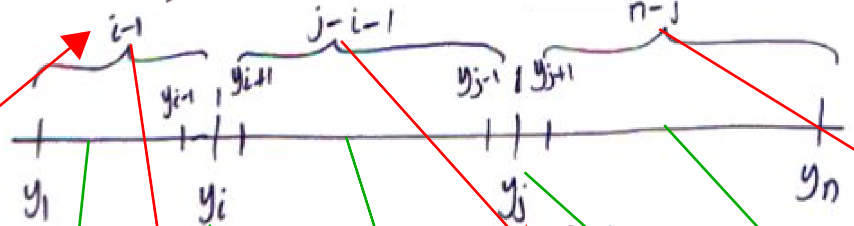
$n-k$ observations larger than y_k

There are $\dfrac{n!}{(k-1)!\,1!\,(n-k)!}$ possible orderings. Thus,



$$g_k(x_k) = \frac{n!}{(k-1)!\,1!\,(n-k)!}\left[F(y_k)\right]^{k-1} f(y_k)\left[1-F(y_k)\right]^{n-k}$$

Similarly, to find the joint pdf of X_i & X_j, $i \neq j$.



$$g_{ij}(y_i, y_j) = \frac{n!}{(i-1)!\,1!\,(j-i-1)!\,1!\,(n-j)!}\left[F(y_i)\right]^{i-1} f(y_i)\left[F(y_j)-F(y_i)\right]^{j-i-1} f(y_j)\left[1-F(y_j)\right]^{n-j}, \text{ if } a<y_i<y_j<b$$

Special order stats:

$$\text{(smallest order statistic) } Y_1$$

$$\text{(largest order statistic) } Y_n$$

$$\text{(sample range) } R = Y_n - Y_1$$

$$\text{(sample median) } Y_k, \text{ where } k = \frac{n+1}{2} (n \text{ odd})$$

Similarly, to find the joint pdf of X_i and X_j, where $i \neq j$

$$\underbrace{y_1 \quad \cdots \quad y_{i-1}}_{i-1} \quad y_i \quad \underbrace{y_{i+1} \quad \cdots \quad y_{j-1}}_{j-i-1} \quad y_j \quad \underbrace{y_{j+1} \quad \cdots \quad y_n}_{n-j}$$

$$c \cdot [F(y_i)]^{i-1} \; f(y_i) \; [F(y_j) - F(y_i)]^{j-i-1} \; f(y_j) \; [1 - F(y_j)]^{n-j}$$

$$g_{ij}(y_i, y_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}$$

(smallest order statistic) Y_1

(largest order statistic) Y_n (sample median) Y_k, where $k = \dfrac{n+1}{2}$ (n odd)

(sample range) $R = Y_n - Y_1$

pdfs and CDFs of Special Order Stats:

Smallest order statistic

Y_1

$$g_1(y) = n[1 - F(y)]^{n-1} f(y), \quad a < y < b$$

$$G_1(y) = \begin{cases} 0, & y \leqslant a \\ 1 - [1 - F(y)]^n, & a < y < b \\ 1, & y \geqslant b \end{cases}$$

Largest order statistic

Y_n

$$g_n(y) = n[F(y)]^{n-1} f(y), \quad a < y < b$$

$$G_n(y) = \begin{cases} 0, & y \leqslant a \\ [F(y)]^n, & a < y < b \\ 1, & y \geqslant b \end{cases}$$