Order Statistics (Section 5.4)

Let $X_1, X_2, \ldots, X_n$ be any values.
The order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are the values arranged in ascending order.

Theorem

If $X_1, \ldots, X_n$ are iid with cdf F,

then $$F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} (F(x))^k (1-F(x))^{n-k}$$

Proof:

$$P(X_{(j)} \leq x) = P(\#\text{ of } X_i \leq x \text{ is at least } j)$$

$$= P(Y \geq j)$$

where $$Y = \sum_{i=1}^{n} I(X_i \leq x)$$

$$\sim \text{Binomial}(n, p = F(x))$$

Note: Special Case $F_{X_{(n)}}(x) = (F(x))^n$

we already knew.

Corollary

If $X_1, \dots, X_n$ are iid with pdf f,

then $f_{X_{(j)}}(x) = n\binom{n-1}{j-1} f(x)(F(x))^{j-1}(1-F(x))^{n-j}$.

"Proof":

$$f_{X_{(j)}}(x)\,dx = (n\, f(x)\,dx)\cdot\binom{n-1}{j-1}(F(x))^{j-1}\cdot(1-F(x))^{n-j}$$

Choose one X_i to be $X_{(j)}$

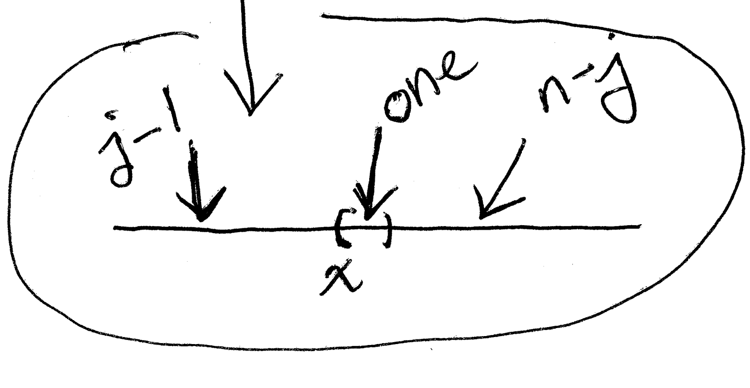
place it in $[x, x+dx)$

Choose $j-1$ of the remaining $n-1$ X_i to be $< x$

place them $< x$

Place remaining $n-j$ points $> x$

$P(X_{(j)} \in [x, x+dx))$



The Real Proof:

$$f_{X_{(j)}}(x) = \frac{d}{dx} F_{X_{(j)}}(x)$$

$$= \sum_{k=j}^{n} \binom{n}{k} \frac{d}{dx} \underbrace{\left\{ F(x)^k (1-F(x))^{n-k} \right\}}$$

$$\left\{ k\, F(x)^{k-1} (1-F(x))^{n-k} f(x) \right.$$

$$\left. - (n-k)\, F(x)^k (1-F(x))^{n-k-1} f(x) \right\}$$

$$\binom{n}{k} k = \binom{n-1}{k-1} n \quad \text{for } k = 1, \ldots, n \quad (k > 0)$$

$$\binom{n}{k}(n-k) = \binom{n-1}{k} n \quad \text{for } k = 0, \ldots, n-1 \quad (k < n)$$

(True always if $\binom{a}{b} \equiv 0$ for $b < 0$ or $b > a$.)

$$= n f(x) \sum_{k=j}^{n} \binom{n-1}{k-1} F(x)^{k-1} (1-F(x))^{n-k}$$

$$- n f(x) \sum_{k=j}^{n} \binom{n-1}{k} F(x)^k (1-F(x))^{n-k-1}$$

$$= n f(x) \sum_{l=j-1}^{n-1} \binom{n-1}{l} F(x)^{l} (1-F(x))^{n-1-l}$$

$$- n f(x) \sum_{k=j}^{n} \binom{n-1}{k} F(x)^{k} (1-F(x))^{n-k-1}$$

$$= \begin{pmatrix} \text{Term for} \\ l = j-1 \end{pmatrix} - \underbrace{\begin{pmatrix} \text{Term for} \\ k = n \end{pmatrix}}_{= 0 \text{ since } \binom{n-1}{n} = 0}$$

$$= n f(x) \binom{n-1}{j-1} F(x)^{j-1} (1-F(x))^{n-j}$$

as desired

Special Case:

When F is Uniform$(0,1)$

$F(x) = x$ for $0 < x < 1$

pdf $f(x) = 1$

$$f_{X_{(j)}}(x) = n \binom{n-1}{j-1} 1 \cdot x^{j-1} (1-x)^{n-j}$$

$$= \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} x^{j-1} (1-x)^{(n-j+1)-1}$$

Beta$(j, n-j+1)$

Note: (Connection with HW exercise 2.40)

$$F_{X_{(j)}}(x) = \int_0^x f_{X_{(j)}}(u)\,du$$

$$\sum_{k=j}^{n} \binom{n}{k} F(x)^k (1-F(x))^{n-k}$$

$$= \int_0^x n \binom{n-1}{j-1} f(u) F(u)^{j-1} (1-F(u))^{n-j} du$$

For Uniform $(0,1)$ this becomes

$$\sum_{k=j}^{n} \binom{n}{k} x^k (1-x)^{n-k}$$

$$= \int_0^x n \binom{n-1}{j-1} u^{j-1} (1-u)^{n-j} du$$

which is a close relative of the result in Exercise 2.40.

Connection with [2.40]

$$P(X_{(j)} > x) = 1 - P(X_{(j)} < x)$$

$$\sum_{k=0}^{j-1} \binom{n}{k} x^k (1-x)^{n-k}$$

$$= \int_x^1 n \binom{n-1}{j-1} u^{j-1} (1-u)^{n-j} \, du$$

$$= \int_0^{1-x} \underbrace{n \binom{n-1}{j-1}}_{= (n-(j-1)) \binom{n}{j-1}} (1-z)^{j-1} z^{n-j} \, dz$$

Multivariate Transformations (p. 185)

$X = (X_1, X_2, \ldots, X_p)$, pdf $f_X(x)$

$U = (U_1, U_2, \ldots, U_p)$

$U = g(X)$ where $g: \mathbb{R}^p \to \mathbb{R}^p$

$\mathcal{A} = \{x : f_X(x) > 0\}$

$\mathcal{B} = \{u : g(x) = u \text{ for some } x \in \mathcal{A}\}$

If $g: \mathcal{A} \to \mathcal{B}$ is 1-1 and "smooth"
with $h = g^{-1}$, then

$$f_U(u) = f_X(h(u))\,|J_h(u)| \text{ for } u \in \mathcal{B}.$$

If $\mathcal{A} = A_0 \cup A_1 \cup \cdots \cup A_K$ disjoint
with $P(X \in A_0) = 0$
and $g_i: A_i \to \mathcal{B}$ is 1-1, onto, "smooth"
(where g_i is g restricted to A_i)
with inverse $h_i = g_i^{-1}$ where $h_i: \mathcal{B} \to A_i$,

then

$$f_U(u) = \sum_{i=1}^{K} f_X(h_i(u))\,|J_{h_i}(u)| \text{ for } u \in \mathcal{B}.$$

Order statistics

Sample size $n=2$

$X = (X_1, X_2)$; X_1, X_2 iid with pdf f

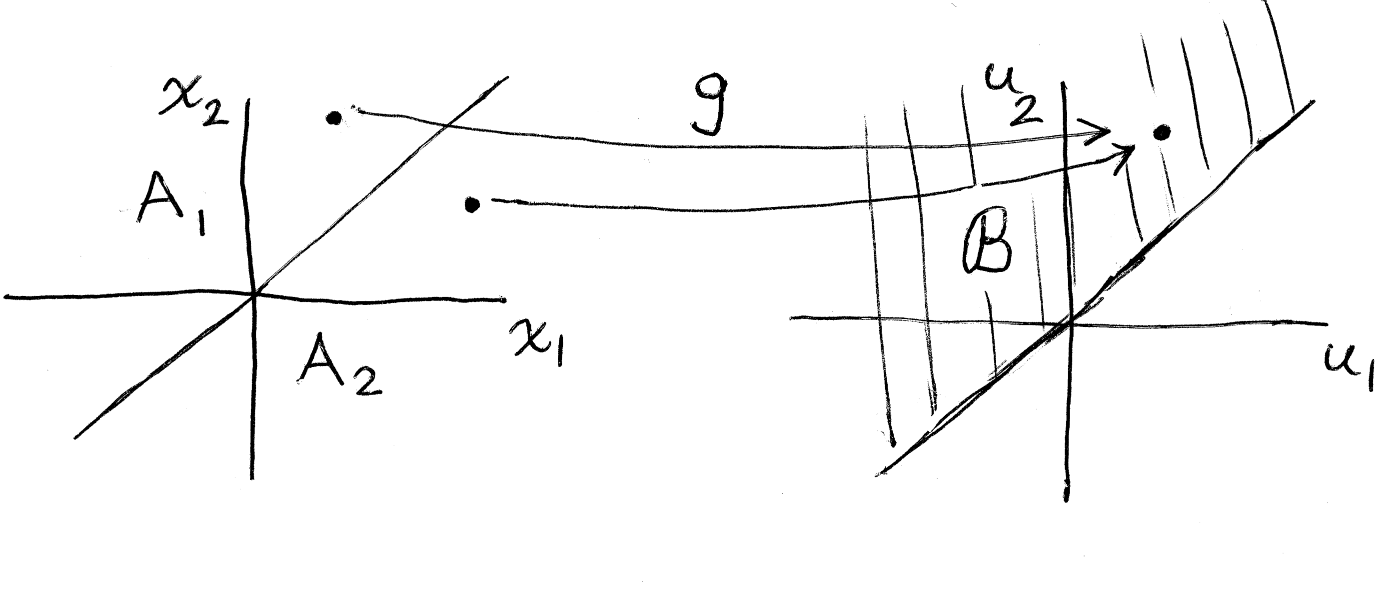
so that
$$f_X(x) = f_{X_1,X_2}(x_1,x_2) = f(x_1)f(x_2)$$

$$U = (U_1, U_2) = (X_{(1)}, X_{(2)}) = g(X)$$

where

$$g(x) = g(x_1,x_2) = \begin{cases} \overbrace{(x_1,x_2)}^{g_1} & \text{if } \overbrace{x_1 < x_2}^{A_1} \\ \underbrace{(x_2,x_1)}_{g_2} & \text{if } \underbrace{x_2 < x_1}_{A_2} \end{cases}$$

and $A_0 = \{(x_1,x_2) : x_1 = x_2\}$.



Thus $\mathcal{B} = \{(u_1, u_2) : u_1 < u_2\}$

$$h_1 : \mathcal{B} \to A_1 \; , \; h_1 = g_1^{-1}$$

$$h_1(u_1, u_2) = (u_1, u_2)$$

$$h_2 : \mathcal{B} \to A_2 \; , \; h_2 = g_2^{-1}$$

$$h_2(u_1, u_2) = (u_2, u_1) .$$

It is easily seen that

$$|J_{h_1}(u)| = |J_{h_2}(u)| = 1 .$$

For example

$$x = h_2(u) \Rightarrow \begin{array}{l} x_1 = u_2 \\ x_2 = u_1 \end{array} \Rightarrow J_{h_2}(u) =$$

$$\begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1 .$$

Also $f_X(h_1(u)) = f_X(h_2(u)) = f(u_1) f(u_2)$.

For example, $f_X(h_2(u)) = f_{X_1, X_2}(u_2, u_1)$

$$= f(u_2) f(u_1) = f(u_1) f(u_2) .$$

Thus

$$f_U(u) = \sum_{i=1}^{2} f_X(h_i(u)) \, |J_{h_i}(u)| \quad \text{for } u \in \mathcal{B}$$

$$= 2f(u_1)f(u_2) \quad \text{for } u_1 < u_2$$

(zero otherwise)

$$= 2f(u_1)f(u_2) I(u_1 < u_2)$$

<u>Sample size n=3</u>

$X = (X_1, X_2, X_3)$ where X_1, X_2, X_3 are iid with pdf f so that

$$f_X(x) = f_{X_1,X_2,X_3}(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3).$$

$$U = (U_1, U_2, U_3) = (X_{(1)}, X_{(2)}, X_{(3)})$$

$$= g(X) \text{ where } g(x) = g(x_1, x_2, x_3)$$

$$= \begin{cases} (x_1, x_2, x_3) & \text{if } x_1 < x_2 < x_3 \\ (x_2, x_1, x_3) & \text{if } x_2 < x_1 < x_3 \\ (x_1, x_3, x_2) & \text{if } x_1 < x_3 < x_2 \\ (x_3, x_2, x_1) & \text{if } x_3 < x_2 < x_1 \\ (x_2, x_3, x_1) & \text{if } x_2 < x_3 < x_1 \\ (x_3, x_1, x_2) & \text{if } x_3 < x_1 < x_2 \end{cases}$$