Lectures 17

Order Statistics. We end the discussion on transformations of random variables with order statistics. Let $X_1, X_2, \cdots, X_n$ be n random variables, the order statistics denoted by $X^{(1)}, X^{(2)}, \cdots, X^{(n)}$ of these random variables are defined as follows.

$$X^{(k)} = \min_{J \subseteq [n],\ \#J = k} \ \max_{j \in J} X_j,$$

where $[n]$ denote $\{1, 2, \cdots, n\}$. Observe that

$$X^{(1)} = \min\{X_1, \cdots, X_n\},\ X^{(n)} = \max\{X_1, \cdots, X_n\}.$$

For the rest of the section, we assume that $X_1, \cdots, X_n$ are i.i.d. random variables with distribution function of X_1 denoted by F.

Distribution function of marginals: First we will compute the marginals of $X^{(k)}$. Let $F_{X^{(k)}}$ denote the distribution function of $X^{(k)}$. For $x \in \mathbb{R}$,

$$\begin{aligned}
P\big(X^{(k)} \leq x\big) &= P\big(X_i \text{ are less than or equal to } x \text{ for atleast } k\ i's\big) \\
&= P\Big(\bigcup_{l=k}^{n} \bigcup_{\{i_1, i_2, \cdots, i_l\} \subseteq [n]} \{X_{i_1} \leq x, X_{i_2} \leq x, \cdots X_{i_l} \leq x, X_i > x, i \notin \{i_1, \cdots, i_l\}\}\Big) \\
&= \sum_{l=k}^{n} \sum_{\{i_1, i_2, \cdots, i_l\} \subseteq [n]} P\Big(X_{i_1} \leq x, \cdots X_{i_l} \leq x,\ X_i > x \,\forall\, i \notin \{i_1, \cdots i_l\}\Big) \\
&= \sum_{l=k}^{n} \binom{n}{l} P\Big(X_{i_1} \leq x, \cdots X_{i_l} \leq x,\ X_i > x \,\forall\, i \notin \{i_1, \cdots i_l\}\Big) \\
&= \sum_{l=k}^{n} \binom{n}{l} F(x)^l (1 - F(x))^{n-l}.
\end{aligned}$$

Hence

$$\boxed{F_{X^{(k)}}(x) = \sum_{l=k}^{n} \binom{n}{l} F(x)^l (1 - F(x))^{n-l}.} \tag{0.1}$$

For the rest of the discussion, we assume that the pdf of $(X_1, \cdots, X_n)$ exists and denote it by f. Also the distribution function F is assumed to be differentiable on $\{x : 0 < F(x) < 1\}$.

Differentiate (0.1) we get

$$
\begin{aligned}
\frac{dF_{X^{(k)}}(x)}{dx} &= \sum_{l=k}^{n} \binom{n}{l} lF'(x)F(x)^{l-1}(1-F(x))^{n-l} \\
&\quad -\sum_{l=k}^{n-1} \binom{n}{l} (n-j)F'(x)F(x)^{l}(1-F(x))^{n-l-1} \\
&= \sum_{l=k}^{n} \frac{n!}{(l-1)!(n-l)!} f(x)F(x)^{l-1}(1-F(x))^{n-l} \\
&\quad -\sum_{l=k}^{n-1} \frac{n!}{l!(n-l-1)!} f(x)F(x)^{l}(1-F(x))^{n-l-1} \\
&= \sum_{l=k}^{n} \frac{n!}{(l-1)!(n-l)!} f(x)F(x)^{l-1}(1-F(x))^{n-l} \\
&\quad -\sum_{l=k+1}^{n} \frac{n!}{(l-1)!(n-l)!} f(x)F(x)^{l-1}(1-F(x))^{n-l} \\
&= \frac{n!}{(k-1)!(n-k)!} f(x)F(x)^{k-1}(1-F(x))^{n-k}.
\end{aligned}
$$

Hence

$$
f_{X^{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} f(x)F(x)^{k-1}(1-F(x))^{n-k}, x \in \mathbb{R}.
$$

Distribution of $(X^{(1)}, \cdots, X^{(n)})$ **:**

Let $\mathbb{R}^n_< = \{(x_1, x_2, \cdots, x_n) | x_1 < x_2 < \cdots < x_n\}$. Then

$$
P\big((X^{(1)}, \cdots, X^{(n)}) \in \mathbb{R}^n_<\big) = 1.
$$

Hence if $f_{12\cdots n}$ denote the pdf of $(X^{(1)}, \cdots, X^{(n)})$, then

$$
f_{12\cdots n}(x_1, x_2, \cdots, x_n) = 0 \text{ outside } \mathbb{R}^n_<.
$$

For $(x_1, x_2, \cdots, x_n) \in \mathbb{R}^n_<$,

$$
\begin{aligned}
&\{X^{(1)} \le x_1, X^{(2)} \in (x_1, x_2], \cdots, X^{(n)} \in (x_{n-1}, x_n]\} \\
&= \big\{X_{\pi(1)} \le x_1,\ X_{\pi(2)} \in (x_1, x_2], \cdots, X_{\pi(n)} \in (x_{n-1}, x_n] \text{ for some permutation } \pi\big\} \\
&= \bigcup_{\pi \in S_n} \big\{X_{\pi(1)} \le x_1,\ X_{\pi(2)} \in (x_1, x_2], \cdots, X_{\pi(n)} \in (x_{n-1}, x_n]\big\},
\end{aligned}
$$

where S_n denote the set of permuatations of $(1, 2, \cdots, n)$. Therefore

$$\begin{aligned} P\big(X^{(1)} \leq x_1, \cdots, X^{(n)} \in (x_{n-1},\ x_n]\big) &= \sum_{\pi \in S_n} P\big(X_{\pi(1)} \leq x_1,\ X_{\pi(2)} \in (x_1,\ x_2], \cdots, X_{\pi(n)} \in (x_{n-1},\ x_n]\big) \\ &= n! F(x_1)(F(x_2) - F(x_1)) \cdots (F(x_n) - F(x_{n-1}). \end{aligned}$$

Since (exercise[1])

$$\frac{\partial^n}{\partial x_1 \cdots \partial x_n} \Big(P\big(X^{(1)} \leq x_1, \cdots, X^{(n)} \in (x_{n-1},\ x_n]\big) \Big) = f_{12\cdots n}(x_1, \cdots x_n),$$

we get

$$f_{12\cdots n}(x_1, \cdots, x_n) = \begin{cases} n! f(x_1) f(x_2) \cdots f(x_n) & \text{if} \quad (x_1, \cdots, x_n) \in \mathbb{R}^n_<, \\ 0 & \text{otherwise.} \end{cases}$$

Distribution of $(X^{(k)}, X^{(k+1)}), k = 1, 2, \cdots, n-1$ **:**

We can use $f_{12\cdots n}$ to compute the joint pdf f_{kk+1} of $(X^{(k)}, X^{(k+1)})$.

For $(x_2, \cdots, x_n) \in \mathbb{R}^{n-1}_<$, i.e. $x_2 < x_3 < \cdots < x_n$, the pdf $f_{23\cdots n}$ of $(X^{(2)}, \cdots, X^{(n)})$ is given by

$$\begin{aligned} f_{23\cdots n}(x_2, \cdots, x_n) &= \int_{-\infty}^{\infty} f_{12\cdots n}(y, x_2, \cdots, x_n) dy \\ &= n! f(x_2) \cdots f(x_n) \int_{-\infty}^{x_2} f(y) dy \\ &= n! f(x_2) \cdots f(x_n) F(x_2). \end{aligned}$$

Therefore

$$f_{2\cdots n}(x_2, \cdots, x_n) = \begin{cases} n! f(x_2) \cdots f(x_n) F(x_2) & \text{if} \quad (x_2, \cdots, x_n) \in \mathbb{R}^{n-1}_<, \\ 0 & \text{otherwise.} \end{cases}$$

For $(x_3, \cdots, x_n) \in \mathbb{R}^{n-2}_<$, i.e., $x_3 < x_4 < \cdots < x_n$, the pdf $f_{34\cdots n}$ of

[1] Hint:

$$\begin{aligned} P\big(X^{(1)} \leq x_1, \cdots, X^{(n)} \in (x_{n-1},\ x_n]\big) &= P\big((X^{(1)}, \cdots X^{(n)}) \in (-\infty,\ x_1] \times \cdots \times (x_{n-1},\ x_n]\big) \\ &= \int_{\infty}^{x_1} \int_{x_1}^{x_2} \cdots \int_{x_{n-1}}^{x_n} f_{12\cdots n}(y_1, y_2, \cdots, y_n) dy_n \cdots dy_1. \end{aligned}$$

Now take the nth order partial derivative of the above to conclude.

$(X^{(3)}, \cdots, X^{(n)})$ is given by

$$\begin{aligned} f_{3\cdots n}(x_3, \cdots, x_n) &= \int_{-\infty}^{\infty} f_{2\cdots n}(y, x_3, \cdots, x_n) dy \\ &= n! f(x_3) \cdots f(x_n) \int_{-\infty}^{x_3} f(y) F(y) dy \\ \big(\text{use the substitution } u = F(y)\big) &= n! f(x_3) \cdots f(x_n) \frac{1}{2} F(x_3)^2. \end{aligned}$$

Therefore

$$f_{3\cdots n}(x_3, \cdots, x_n) = \begin{cases} \frac{n!}{2!} f(x_3) \cdots f(x_n) F(x_3)^2 & \text{if} \quad (x_3, \cdots, x_n) \in \mathbb{R}_<^{n-2}, \\ 0 & \quad \text{otherwise.} \end{cases}$$

Repeating the above procedure upto the pdf $f_{kk+1\cdots n}$ of $(X^{(k)}, \cdots, X^{(n)})$ we get

$$f_{k\cdots n}(x_k, \cdots, x_n) = \begin{cases} \frac{n!}{(k-1)!} f(x_k) \cdots f(x_n) F(x_k)^{k-1} & \text{if} \quad (x_k, \cdots, x_n) \in \mathbb{R}_<^{n-k+1}, \\ 0 & \quad \text{otherwise.} \end{cases}$$

Now for $x_k < x_{k+1} < \cdots < x_{n-1}$, the pdf $f_{k\cdots(n-1)}$ is given by

$$\begin{aligned} f_{k\cdots(n-1)}(x_k, \cdots, x_{n-1}) &= \int_{-\infty}^{\infty} f_{k\cdots n}(x_k, \cdots, x_{n-1}, y) dy \\ &= \frac{n!}{(k-1)!} f(x_k) \cdots f(x_{n-1}) F(x_k)^{k-1} \int_{x_{n-1}}^{\infty} f(y) dy \\ &= \frac{n!}{(k-1)!} f(x_k) \cdots f(x_{n-1}) F(x_k)^{k-1} (1 - F(x_{n-1}). \end{aligned}$$

Hence

$$f_{k\cdots(n-1)}(x_k, \cdots, x_{n-1}) = \begin{cases} \frac{n!}{(k-1)!} f(x_k) \cdots f(x_{n-1}) F(x_k)^{k-1} (1 - F(x_{n-1})) \\ \qquad \text{if } (x_k, \cdots, x_{n-1}) \in \mathbb{R}_<^{n-k}, \\ 0 \qquad \text{otherwise.} \end{cases}$$

Again for $x_k < x_{k+1} < \cdots < x_{n-2}$, the pdf $f_{k\cdots(n-2)}$ is

$$\begin{aligned} f_{k\cdots(n-2)}(x_k, \cdots, x_{n-2}) &= \int_{-\infty}^{\infty} f_{k\cdots(n-1)}(x_k, \cdots, x_{n-2}, y) dy \\ &= \frac{n!}{(k-1)!} f(x_k) \cdots f(x_{n-2}) F(x_k)^{k-1} \int_{x_{n-1}}^{\infty} f(y)(1 - F(y)) dy \\ \big(u = 1 - F(y)\big) &= \frac{n!}{(k-1)!} f(x_k) \cdots f(x_{n-1}) F(x_k)^{k-1} \int_0^{1-F(x_{n-2})} u du \\ &= \frac{n!}{(k-1)!2!} f(x_k) \cdots f(x_{n-2}) F(x_k)^{k-1} (1 - F(x_{n-2}))^2. \end{aligned}$$

Hence the pdf $f_{k\cdots(n-2)}$ of $(X^{(k)},\cdots,X^{(n-2)})$ is given by

$$f_{k\cdots(n-2)}(x_k,\cdots,x_{n-2}) \;=\; \begin{cases} \frac{n!}{(k-1)!2!}f(x_k)\cdots f(x_{n-2})F(x_k)^{k-1}(1-F(x_{n-2}))^2 \\ \qquad \text{if } (x_k,\cdots,x_{n-2})\in\mathbb{R}^{n-k-1}_<, \\ 0 \qquad \text{otherwise.} \end{cases}$$

Repeat this, we get

$$f_{k(k+1)}(x_k,x_{k+1}) \;=\; \begin{cases} \frac{n!}{(k-1)!(n-k-1)!}f(x_k)f(x_{k+1})F(x_k)^{k-1}(1-F(x_{k+1}))^{n-k-1} \\ \qquad \text{if } (x_k,x_{k+1})\in\mathbb{R}^2_<, \\ 0 \qquad \text{otherwise.} \end{cases}$$

Distribution of $(X^{(1)},X^{(n)})$:

Since $P\big((X^{(1)},X^{(n)})\in\mathbb{R}^2_<\big)=1$, we compute $P\big(X^{(1)}\le x, X^{(n)}\le y\big)$ for $(x,y)\in\mathbb{R}^2_<$. Observe that

$$P\big(X^{(1)}\le x,X^{(n)}\le y\big) \;=\; P(X^{(n)}\le y)-P\big(X^{(1)}>x,X^{(n)}\le y\big).$$

$$\begin{aligned} P\big(X^{(1)}>x,X^{(n)}\le y\big) &= P\big(x<X_i\le y,\ \text{for all } i=1,\cdots,n\big) \\ &= (F(y)-F(x))^n. \end{aligned}$$

Therefore

$$P\big(X^{(1)}\le x,X^{(n)}\le y\big) \;=\; F_{X^{(n)}}(y)-(F(y)-F(x))^n \;=\; F(y)^n-(F(y)-F(x))^n.$$

Observing that

$$f_{1n}(x_1,x_n)=\frac{\partial^2}{\partial x_1\partial x_n}P\big(X^{(1)}\le x_1,X^{(n)}\le x_n\big)$$

we get

$$f_{1n}(x_1,x_n) \;=\; \begin{cases} n(n-1)f(x_1)f(x_n)(F(x_n)-F(x_1))^{n-2} & \text{if } (x_1,x_n)\in\mathbb{R}^2_<, \\ 0 & \text{otherwise.} \end{cases}$$

Example 0.1 *Let $X_1,\cdots,X_n$ be i.i.d. Uniform $(0,1)$ random variables. Compute the marginal distributions of the 2nd order statistics. Also find the joint distribution of* $\min\{X_1,\cdots,X_n\}$ *and* $\max\{X_1,\cdots,X_n\}$. *Hence find* $P(\min\{X_1,\cdots,X_n\}+\frac{1}{2}>\max\{X_1,\cdots,X_n\})$

Note the pdf f_k of the kth order statistics $X^{(k)}$ is give by

$$\begin{aligned} f_k(x) &= \frac{n!}{(k-1)!(n-k)!} f(x)F(x)^{k-1}(1-F(x))^{n-k}, x \in \mathbb{R} \\ &= \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k},\ 0 < x < 1,\ =\ 0, \text{otherwise}. \end{aligned}$$

i.e. $X^{(k)}$ *has Beta* $(k, n-k+1)$ *distribution*[2] *The joint distribution of* $X := \min\{X_1, \cdots, X_n\},\ Y := \max\{X_1, \cdots, X_n\}$ *is given by*

$$\begin{aligned} f_{1n}(x,y) &= \frac{n!}{(n-2)!} f(x)f(y)(F(y)-F(x))^{n-2},\ x < y, \\ & \quad = 0,\ \text{otherwise} \\ &= n(n-1)(y-x)^{n-2},\ 0 < x < y < 1,\ =\ 0,\ \text{otherwise}. \end{aligned}$$

Hence the required probability is given by

$$\begin{aligned} P(Y < X + \frac{1}{2}) &= n(n-1)\int_0^{\frac{1}{2}} \int_{x_1}^{x+\frac{1}{2}} (y-x)^{n-2} dy dx \\ &\quad + n(n-1)\int_{\frac{1}{2}}^1 \int_x^1 (y-x)^{n-2} dy dx \\ &= n(n-1)[\frac{1}{2^n(n-1)} + \frac{1}{2^n n(n-1)}] = \frac{n+1}{2^n}. \end{aligned}$$

[2] Beta (n, m) random variable is defined through the pdf

$$f(x) = \begin{cases} \frac{(n+m-1)!}{(m-1)!(n-1)!} x^{m-1}(1-x)^{n-1}, & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$