# 556: MATHEMATICAL STATISTICS I

## SAMPLE QUANTILES AND ORDER STATISTICS

For $n$ random variables $X_1, ...., X_n$, the **order statistics**, $Y_1, ...., Y_n$, are defined by

$$Y_i = X_{(i)} - \text{“the i'th smallest value in } X_1, ...., X_n\text{”}$$

for $i = 1, ..., n$. For example

- $Y_1 = X_{(1)} = \min\{X_1, ..., X_n\}$,
- $Y_n = X_{(n)} = \max\{X_1, ..., X_n\}$.

For $n$ independent, identically distributed random variables $X_1, ..., X_n$, with marginal density function $f_X$, the following theorem characterizes the key distributional results.

**THEOREM**
For random sample $X_1, ..., X_n$ from population with pmf/pdf $f_X$ and cdf $F_X$,

(a) $Y_1 = X_{(1)}$ has cdf

$$F_{Y_1}(y) = 1 - \{1 - F_X(y)\}^n$$

(b) $Y_n = X_{(n)}$ has cdf

$$F_{Y_n}(y) = \{F_X(y)\}^n$$

*Proof.* (a) For the marginal cdf for $Y_1$,

$$\begin{aligned}
F_{Y_1}(y_1) &= P[Y_1 \le y_1] = 1 - P[Y_1 > y_1] = 1 - P[\min\{X_1, ..., X_n\} > y_1] \\
&= 1 - P[X_1 > y_1, X_2 > y_1, ..., X_n > y_1] \\
&= 1 - \prod_{i=1}^{n} P[X_i > y_1] = 1 - \prod_{i=1}^{n} \{1 - F_X(y_1)\} \\
&= 1 - \{1 - F_X(y_1)\}^n
\end{aligned}$$

(b) For $Y_n$,

$$\begin{aligned}
F_{Y_n}(y_n) &= P[Y_n \le y_n] = P[\max\{X_1, ..., X_n\} \le y_n] = P[X_1 \le y_n, X_2 \le y_n, ..., X_n \le y_n] \\
&= \prod_{i=1}^{n} P[X_i \le y_n] = \prod_{i=1}^{n} \{F_X(y_n)\} \\
&= \{F_X(y_n)\}^n
\end{aligned}$$

The pmf/pdf can be computed from the cdf. ∎

**THEOREM (MARGINAL PMF/PDF)**

For random sample $X_1, ..., X_n$ from population with pmf/pdf $f_X$ and cdf $F_X$,

(a) In the **discrete** case, suppose that $\mathbb{X} \equiv \{x_1, x_2, \ldots\}$, where $x_1 < x_2 < \cdots$, and suppose that

$$f_X(x_i) = p_i \qquad i = 1, 2, \ldots.$$

Then the marginal cdf of $Y_j = X_{(j)}$ is defined by

$$F_{Y_j}(x_i) = \sum_{k=j}^{n} \binom{n}{k} P_i^k (1 - P_i)^{n-k} \qquad x_i \in \mathbb{X}$$

with the usual cdf behaviour at other values of $x$. The marginal pmf of $Y_j = X_{(j)}$ is

$$f_{Y_j}(x) = \sum_{k=j}^{n} \binom{n}{k} [P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k}] \qquad x_i \in \mathbb{X}$$

and zero otherwise, where

$$P_i = \sum_{k=1}^{i} p_i.$$

(b) In the **continuous** case, the marginal cdf of $Y_j = X_{(j)}$ is

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

and the marginal pdf is

$$f_{Y_j}(x) = \frac{n!}{(j-1)!(n-j)!} \{F_X(x)\}^{j-1} \{1 - F_X(x)\}^{n-j} f_X(x)$$

**THEOREM (JOINT PDF: CONTINUOUS CASE)**

For random sample $X_1, ..., X_n$ from population with pdf $f_X$, the joint pdf of order statistics $Y_1, ...., Y_n$

$$f_{Y_1,...,Y_n}(y_1, ..., y_n) = n! f_X(y_1) ... f_X(y_n) \qquad y_1 < ... < y_n$$

**NOTE:** In general, these distributions are difficult to compute for large $n$.