1 Order Statistics

Definition

The order statistics of a random sample $X_1, \ldots, X_n$ are the sample values placed in ascending order. They are denoted by $X_{(1)}, \ldots, X_{(n)}$.

The order statistics are random variables that satisfy $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. The following are some statistics that are easily defined in terms of the order statistics.

The *sample range*, $R = X_{(n)} - X_{(1)}$, is the distance between the smallest and largest observations. It is a measure of the dispersion in the sample and should reflect the dispersion in the population.

The *sample median*, which we will denote by M, is a number such that approximately one-half of the observations are less than M and one-half are greater. In terms of order statistics, M is defined by

$$M = \begin{cases} X_{((n+1)/2)} & \text{if } n \text{ is odd} \\ (X_{(n/2)} + X_{(n/2+1)})/2 & \text{if } n \text{ is even.} \end{cases}$$

The median is a measure of location that might be considered an alternative to the sample mean. One advantage of the sample median over the sample mean is that it is less affected by extreme observations.

For any number p between 0 and 1, the $(100p)$th sample percentile is the observation such that approximately np of the observations are less than this observation and $n(1-p)$ of the observations are greater. The 50th percentile is the sample median, the 25th percentile is called the lower quartile, and the 75th percentile is called the upper quartile. A measure of dispersion that is sometimes used is the interquartile range, the distance between the lower and upper quartiles.

Theorem 5.3.4

Let $X_1, \ldots, X_n$ be a random sample from a discrete distribution with pmf $f_X(x_i) = p_i$, where

$x_1 < x_2 < \cdots$ are the possible values of X in ascending order. Define

$$\begin{aligned} P_0 &= 0 \\ P_1 &= p_1 \\ P_2 &= p_1 + p_2 \\ &\vdots \\ P_i &= p_1 + p_2 + \cdots + p_i \\ &\vdots \end{aligned}$$

Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics from the sample. Then

$$P(X_{(j)} \leq x_i) = \sum_{k=j}^{n} \binom{n}{k} P_i^k (1 - P_i)^{n-k}$$

and

$$P(X_{(j)} = x_i) = \sum_{k=j}^{n} \binom{n}{k} [P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k}].$$

PROOF: Fix i, and let Y be a random variable that counts the number of $X_1, \ldots, X_n$ that are less than or equal to x_i. For each of $X_1, \ldots, X_n$, call the event $\{X_j \leq x_i\}$ a success and $\{X_j > x_i\}$ a "failure". Then Y is the number of success in n trials. Thus, $Y \sim$binomial(n, P_i).

The event $\{X_{(j)} \leq x_i\}$ is equivalent $\{Y \geq j\}$; that is, at least j of the sample values are less than or equal to x_i. The two equations are then established. □

Theorem 5.4.4

Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of a random sample, $X_1, \ldots, X_n$, from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the pdf of $X_{(j)}$ is

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) [F_X(x)]^{j-1} [1 - F_X(x)]^{n-j}.$$

Example (Uniform order statistics pdf)

Let $X_1, \ldots, X_n$ be iid uniform(0,1), so $f_X(x) = 1$ for $x \in (0,1)$ and $F_X(x) = x$ for $x \in (0,1)$. Thus, the pdf of the jth order statistics is

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} x^{j-1} (1-x)^{n-j},$$

for $x \in (0,1)$. Hence, $X_{(j)} \sim Beta(j, n-j+1)$. From this we can deduce that

$$EX_{(j)} = \frac{j}{n+1},$$

and

$$\text{Var}X_{(j)} = \frac{j(n-j+1)}{(n+1)^2(n+2)}.$$

Theorem 5.4.6

Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of a random sample, $X_1, \ldots, X_n$, from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the joint pdf of $X_{(i)}$ and $X_{(j)}$, $1 \leq i < j \leq n$, is

$$f_{X_{(i)},X_{(j)}}(u,v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f_X(u) f_X(v) [F_X(u)]^{i-1} [F_X(v) - F_X(u)]^{j-1-i} [1 - F_X(v)]^{n-j}$$

for $-\infty < u < v < \infty$.

The joint pdf of three or more order statistics could be derived using similar but even more involved arguments. Perhaps the other most useful pdf is $f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n)$, the joint pdf of all the order statistics, which is given by

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = \begin{cases} n! f_X(x_1) \ldots f_X(x_n) & -\infty < x_1 < \cdots < x_n < \infty. \\ 0 & \text{otherwise} \end{cases}$$