0.1. Order Statistics

Random sampling assumes that the sample is taken in such a way that the random variables for each trial are independent and follow the common population density function. In this case, the joint density function is the product of the common marginal densities.

DEFINITION 1. The set random variables $X_1, X_2, ..., X_n$ is said to be a random sample of size n from a population with density function $f(x)$ if the joint pdf has the form

$$f(x_1, x_2, ..., x_n) = f(x_1) f(x_2) ... f(x_n)$$

For example, 100 months were required before all five light bulbs failed, but the first four failed in 17 months. In some cases one may desire to stop after the r smallest ordered observations out of n have been observed, because this could result in a great saving of time. The joint distribution of the ordered variables is not the same as the joint density of the unordered variables.

Consider a transformation that orders the values of $x_1, x_2, ..., x_n$. Let

$$\begin{array}{rcc} y_1 = u_1(x_1, x_2, ..., x_n) & = & \min(x_1, x_2, ..., x_n) = x_{1:n} \\ y_2 = u_2(x_1, x_2, ..., x_n) & = & x_{2:n} \\ \vdots & \vdots & \vdots \\ y_n = u_n(x_1, x_2, ..., x_n) & = & \max(x_1, x_2, ..., x_n) = x_{n:n} \end{array}$$

When this transformation is applied to a random sample $X_1, X_2, ..., X_n$, a set of ordered random variables will be obtained, called the **order statistics**, and denoted by either $X_{1:n}, X_{2:n}, ..., X_{n:n}$ or $Y_1, Y_2, ..., Y_n$.

THEOREM 2. *If $X_1, X_2, ..., X_n$ is a random sample from a population with continuous pdf $f(x)$, then the joint pdf of the order statistics $Y_1, Y_2, ..., Y_n$ is*

$$g(y_1, y_2, ..., y_n) = n! f(y_1) f(y_2) ... f(y_n)$$

if $y_1 < y_2 < ... < y_n$, and zero otherwise.

EXAMPLE 3. Suppose X_1, X_2 and X_3 that represent a random sample of size 3 from a population with pdf $f(x) = 2x, 0 < x < 1$. Find the joint pdf of the order statistics Y_1, Y_2 and Y_3.

Solution. It follows that the joint pdf of the order statistics Y_1, Y_2 and Y_3 is

$$g(y_1, y_2, y_3) = 3! (2y_1)(2y_2)(2y_3) = 48 y_1 y_2 y_3, 0 < y_1 < y_2 < y_3 < 1$$

and zero otherwise.

And marginal pdf of Y_1 is

$$g_{Y_1}(y) = \int_{...}^{...} \int_{...}^{...} 48 y_1 y_2 y_3 d...d... = ...$$

It is possible to derive an explicit generale formula for the distribution of the kth order statistic in terms of pdf, $f(x)$, CDF, $F(x)$, of the population random variable.

THEOREM 4. *Suppose $X_1, X_2, ..., X_n$ denotes a random sample of size n from a continuous pdf $f(x)$, where $f(x) > 0$ for $a < x < b$. Then the pdf of the kth order statistic Y_k is*

$$g_k(y_k) = \frac{n!}{(k-1)!\,(n-k)!} (F(y_k))^{k-1} f(y_k)(1-F(y_k))^{n-k}$$

if $a < y_k < b$, and zero otherwise.

PROOF. Having $Y_k = y_k$, one must have $k-1$ observations less than y_k, one at y_k, and $n-k$ observations greater than y_k, where $P(X \leq y_k) = F(y_k)$, the likelihood of an observation at y_k is $f(y_k)$, and $P(X \geq y_k) = 1 - F(y_k)$. There are $\frac{n!}{(k-1)!(n-k)!}$ possible orderings of the n independent observations, therefore

$$g_k(y_k) = \frac{n!}{(k-1)!\,(n-k)!} (F(y_k))^{k-1} f(y_k)(1-F(y_k))^{n-k}$$

□

A similar argument can be used to easily give the joint pdf of any set of order statistics. For example, consider a pair of order statistics Y_i and Y_j where $i < j$. To have $Y_i = y_i$ and $Y_j = y_j$, one must have $i-1$ observation less than y_i, one at y_i, $j-i-1$ between y_i and y_j, one at y_j, $n-j$ greater that y_j. Applying the multinomial form gives the joint pdf for Y_i and Y_j as

(0.1.1)

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

And for special cases, the CDF of $X_{1:n}$ and $X_{n:n}$.

(1) The CDF of $X_{n:n} = X_{maks}$ is

$$\begin{aligned} F_{X_{maks}}(x) &= P(X_{maks} \leq x) \\ &= P(X_1 \leq x, X_2 \leq x, ..., X_n \leq x) \\ &= P(X_1 \leq x) P(X_2 \leq x) ... P(X_n \leq x) \\ &= F_X(x) F_X(x) ... F_X(x) \\ &= (F_X(x))^n \end{aligned} \tag{0.1.2}$$

Thus the pdf of $X_{n:n} = X_{maks}$ is

$$f_{X_{maks}}(x) = \frac{dF_{X_{maks}}(x)}{dx} = \frac{d((F_X(x))^n)}{dx} = n(F_X(x))^{n-1} f_X(x)$$

(2) The CDF of $X_{1:n} = X_{min}$ is

$$\begin{aligned} F_{X_{min}}(x) &= P(X_{min} \leq x) \\ &= 1 - P(X_{min} > x) \\ &= 1 - P(X_1 > x, X_2 > x, ..., X_n > x) \\ &= 1 - P(X_1 > x) P(X_2 > x) ... P(X_n > x) \\ &= 1 - (P(X_1 > x))^n \\ &= 1 - (1 - P(X_1 \leq x))^n \\ &= 1 - (1 - F_X(x))^n \end{aligned} \tag{0.1.3}$$

Thus the pdf of $X_{1:n} = X_{min}$ is

$$f_{X_{min}}(x) = \frac{dF_{X_{min}}(x)}{dx} = \frac{d(1-(1-F_X(x))^n)}{dx} = n(1-F_X(x))^{n-1} f_X(x)$$

EXAMPLE 5. From Example 3, find the pdf of $X_{1:3}$,$X_{2:3}$ and $X_{3:3}$.

Solution. $X_{1:3}$,$X_{2:3}$ and $X_{3:3}$ a random sample of size 3 from a population with CDF $F_X(y) = \int_0^y f(y)\,dx = \int_0^y 2x dx = y^2$. From Theorem 4,

$$f_{Y_1}(y) = \begin{cases} \frac{3!}{(1-1)!(3-1)!}\left(y^2\right)^{1-1} 2y\left(1-y^2\right)^{3-1} = 6y\left(1-y^2\right)^2 & ,0<y<1 \\ 0 & ,otherwise \end{cases}$$

$$f_{Y_2}(y) = \begin{cases} \frac{3!}{(2-1)!(3-2)!}\left(y^2\right)^{2-1} 2y\left(1-y^2\right)^{3-2} = 12y^3\left(1-y^2\right) & ,0<y<1 \\ 0 & ,otherwise \end{cases}$$

$$f_{Y_3}(y) = \begin{cases} \frac{3!}{(3-1)!(3-3)!}\left(y^2\right)^{3-1} 2y\left(1-y^2\right)^{3-3} = 6y^5 & ,0<y<1 \\ 0 & ,otherwise \end{cases}$$

THEOREM 6. *For a random sample of size n from a discrete or continuous CDF, the marginal CDF of the kth order statistic is given by*

$$G_k(y_k) = \sum_{j=k}^{n} \binom{n}{j} \left(F(y_k)\right)^j \left(1-F(y_k)\right)^{n-j}$$

EXAMPLE 7. Consider a random sample of size n from a distribution with pdf and CDF given by $f(x) = 2x$ and $F(x) = x^2; 0 < x < 1$. Find pdf and CDF of $X_{1:n}$ and $X_{n:n}$.

Solution.

EXAMPLE 8. From Example 7, what is the density of the range of the sample, $R = Y_n - Y_1$.

Solution. From equation 0.1.1,

$$g_{in}(y_1, y_n) = \frac{n!}{(n-2)!} 2y_1 \left(y_n^2 - y_1^2\right)^{n-2} 2y_n, 0 < y_1 < y_n < 1$$

Making the transformation $R = Y_n - Y_1$ and $S = Y_1$, yields the inverse transformation y_1 =s, $y_n = r + s$ and $|J| = 1$. Thus, the joint pdf of R and S is

$$h(r,s) = g_{in}(s, r+s)\,|J| = \frac{n!}{(n-2)!} 2s\left((r+s)^2 - s^2\right)^{n-2}(r+s), 0 < s < 1-r, 0 < r < 1$$