7 Lecture 7: Order Statistics

7.1 Motivation

Let $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2}, \dots, \mathbf{X_n})$ denote a random sample from a population with a continuous distribution function F_X. Since F_X is assumed to be *continuous*, the probability of any two of these random variables assuming the same value is zero. After reordering the n values we get $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)})$ in which, as mentioned, the $\leq$ sign could also be replaced by $<$. These values are collectively termed the *order statistic* of the random sample $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2}, \dots, \mathbf{X_n})$. The subject of order statistics generally deals with properties of $X_{(r)}$ $(r = 1, 2, \dots, n)$ which is called the r-th order statistic.

Order statistics are particularly useful in nonparametric statistics because of the following:

Theorem 7.1. (Probability-integral transformation). *If the random variable X has a continuous cdf F_X then the random variable $Y = F_X(X)$ has the uniform probability distribution over the interval (0,1). Further, given a sample $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2}, \dots, \mathbf{X_n})$ of n i.i.d. random variables with cdf F_X, the transformation $U_{(r)} = F_X(X_{(r)})$ produces a random variable $U_{(r)}$ which is the r-th order statistic from the uniform population in (0,1), regardless of what F_X is, i.e. $U_{(r)}$ is distribution-free.*

Proof: It is your textbook and was also discussed in the introductory Lecture 1.

Note: The above theorem has also an extremely important practical application in the generation **(computer simulation)** of observations from any specific continuous distribution function. There are several well-developed **uniform random number generators** that implement methods to generate sequences of uniform in (0,1) pseudo-random numbers. These numbers are **pseudo** since in fact they are generated by a deterministic algorithm (therefore are not random) **but** look as **random** (hence the word pseudo-random) in the sense that they pass usual statistical tests about randomness of the generated sequence. Every program system (Fortran, SPLUS, C, SAS, etc.) has such uniform random number generators and we will not discuss their specific implementation here. What we would like to discuss is how we could use these **uniform** random number generators to generate random numbers with **arbitrary continuous** cumulative distribution function F_X. The answer is:

1) Generate Y as uniformly distributed in (0,1) using the uniform random number generator

2) Calculate $\xi = F_X^{-1}(Y)$.

Then ξ is distributed according to $F_X(.)$ since its cumulative distribution function is:

$$P(F_X^{-1}(Y) < x) = P(F_X^{-1}(F_X(X)) < x) = P(X < x) = F_X(x).$$

Some further important obvious applications of order statistics are listed below:

- $X_{(n)}$ is of interest in studying floods, earthquakes and other extreme phenomena, sports records, financial markets etc.

- $X_{(1)}$ is useful, for example, in estimating strength of a chain that would depend on the weakest link.
- the sample median defined as $X_{[(n+1)/2]}$ for n odd and any number between $X_{(n/2)}$ and $X_{(n/2+1)}$ for n even, is a measure of location and an estimate of the population central tendency.
- the sample midrange $(X_{(n)} + X_{(1)})/2$ is also a measure of central tendency, whereas the sample range $X_{(n)} - X_{(1)}$ is a measure of dispersion.

7.2 Multinomial distribution.

At this point, we shall give the definition of the multinomial distribution. It is used to prove in an easy way many of the results related to distributions of order statistics and for that reason, it will be given first:
Suppose a single trial can result in k $(k \geq 2)$ possible outcomes numbered $1, 2, \ldots, k$ and let $w_i = P(\text{a single trial results in outcome } i)$ $(\sum_{i=1}^k w_i = 1)$. For n independent trials, let X_i denote the number of trials resulting in outcome i (then $\sum_{i=1}^k X_i = n$). Then we say that the distribution of $(X_1, X_2, \ldots, X_k) \sim \text{Multinomial}(n; w_1, w_2, \ldots, w_k)$ and it holds

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1! x_2! \ldots x_k!} w_1^{x_1} w_2^{x_2} \ldots w_k^{x_k}, 0 < w_i < 1, \sum_{i=1}^k w_i = 1$$

Note that when $k = 2$ this is just the familiar Binomial distribution, i.e. the Multinomial can be considered as its generalization. It is also easy to see that

$$E(X_i) = nw_i; Var(X_i) = nw_i(1 - w_i), i = 1, 2, \ldots, k$$

holds by noting that the **marginal** distribution of the i-th component is in fact binomial (Bin$(n; w_i)$). In a bit more complicated way (for example, using moment generating functions), one can also show that $Cov(X_i, X_j) = -nw_i w_j$ holds for $j \neq i$.

7.3 Distributions related to order statistics

A lot of material discussed in this lecture can be found in Section 5.4 of the textbook.

Let X be a random variable with a density $f_X(x)$ and a cumulative distribution function $F_X(x)$ and let there be n independent copies $X_1, X_2, \ldots, X_n$ of X.

Theorem 7.2. *The joint density* $f_{X_{(1)}, X_{(2)}, \ldots, X_{(n)}}(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ *is given by:*

$$f_{X_{(1)}, X_{(2)}, \ldots, X_{(n)}}(x_{(1)}, x_{(2)}, \ldots, x_{(n)}) = n! \prod_{i=1}^n f_X(x_{(i)}) \text{ for } x_{(1)} < x_{(2)} < \cdots < x_{(n)}$$

Proof: We consider a trial consisting of $k = 2n+1$ possible outcomes that is repeated independently n times. Each of the outcomes is a realization in one of the $2n+1$ intervals:

$$(-\infty, x_{(1)}), [x_{(1)}, x_{(1)} + \Delta x_{(1)}), [x_{(1)} + \Delta x_{(1)}, x_{(2)}), \ldots, [x_{(n-1)} + \Delta x_{(n-1)}, x_{(n)}),$$

$$[x_{(n)}, x_{(n)} + \Delta x_{(n)}), [x_{(n)} + \Delta x_{(n)}, \infty)$$

where $\Delta x_{(i)}, i = 1, 2, \dots, n$ are chosen sufficiently small so that no overlap of the intervals occurs. The trial's outcome can be interpreted as a realization of multinomial

$$\text{Multin}(n; F_X(x_{(1)}), F_X(x_{(1)}+\Delta x_{(1)}) - F_X(x_{(1)}), F_X(x_{(2)}) - F_X(x_{(1)}+\Delta x_{(1)}), \dots, (1 - F_X(x_{(n)}+\Delta x_{(n)}))$$

distribution. We are looking at the probability for one very particular outcome of the trial, namely realization in the second, fourth, $2n$th interval for this multinomial distribution. On one hand, this probability is

$$\frac{n!}{0!1!0!1!\dots0!1!0!}[F_X(x_{(1)} + \Delta x_{(1)}) - F_X(x_{(1)}] \dots [F_X(x_{(n)} + \Delta x_{(n)}) - F_X(x_{(n)}].$$

On the other hand, it is just $P(x_{(i)} \leq X_{(i)} < x_{(i)} + \Delta x_{(i)}, i = 1, 2, \dots, n)$. Having in mind the definition

$$f_{X_{(1)}, X_{(2)}, \dots, X_{(n)}}(x_{(1)}, x_{(2)}, \dots, x_{(n)}) = \lim_{\Delta x_{(i)} \to 0} \frac{P(x_{(i)} \leq X_{(i)} < x_{(i)} + \Delta x_{(i)}, i = 1, 2, \dots, n)}{\prod_{i=1}^{n} \Delta x_{(i)}}$$

we get

$$f_{X_{(1)}, X_{(2)}, \dots, X_{(n)}}(x_{(1)}, x_{(2)}, \dots, x_{(n)}) = n! \prod_{i=1}^{n} f_X(x_{(i)}) \text{ for } x_{(1)} < x_{(2)} < \dots < x_{(n)}$$

Theorem 7.3. *It holds:*

$$f_{X_{(n)}}(y_n) = n[F_X(y_n)]^{n-1} f_X(y_n)$$

$$f_{X_{(1)}}(y_1) = n[1 - F_X(y_1)]^{n-1} f_X(y_1)$$

$$f_{X_{(r)}}(y_r) = \frac{n!}{(r-1)!(n-r)!}[F_X(y_r)]^{r-1}[1 - F_X(y_r)]^{n-r} f_X(y_r)$$

Proof: Since the previous theorem gives us the joint distribution, the marginal distributions formulated in Theorem (7.2) can be obtained through integration. This method is a straightforward as an idea but the integration is tiresome. A much simpler method can be used which appeals to probability theory instead to pure mathematical integration. It will be illustrated at the lecture.

Note: From Theorem (7.3) we realize that for the particular case of the F_X being uniform distribution on (0,1), we get

$$f_{X_{(r)}}(y) = \frac{n!}{(r-1)!(n-r)!} y^{r-1}(1-y)^{n-r}, y \in (0,1) (\text{ and zero elsewhere})$$

which is the density of the *Beta distribution* with parameters r and $n-r+1$. In particular, using the properties of the beta distribution, we can now show that for the rth order statistic $X_{(r)}$ of the beta distribution we have $EX_{(r)} = \frac{r}{n+1}, Var(X_{(r)}) = \frac{r(n+1-r)}{(n+1)^2(n+2)}$.

Joint densities of couples $(X_{(i)}, X_{(j)})$ can also be derived through integration of the joint density of all the n order statistics. But it is again easier to use some probabilistic arguments instead. The idea is illustrated in the following result:

Theorem 7.4. *It holds for $1 \leq i < j \leq n$:*

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

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

Proof (idea): We obtain the cdf $F_{X_{(i)},X_{(j)}}(u,v)$ first and then find its partial derivative $f_{X_{(i)},X_{(j)}}(u,v) = \frac{\partial^2}{\partial u \partial v} F_{X_{(i)},X_{(j)}}(u,v)$. To obtain $F_{X_{(i)},X_{(j)}}(u,v)$, let U be a random variable that counts the number of $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n})$ that are less or equal to u and V be the random variable that counts the number of $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n})$ greater than u and less or equal to v. Then

$$(U, V, n-U-V) \sim \text{Multinomial}(n; F_X(u), F_X(v) - F_X(u), 1 - F_X(v)).$$

Then, obviously

$$F_{X_{(i)},X_{(j)}}(u,v) = P(U \geq i, U+V \geq j) = \sum_{k=i}^{j-1} \sum_{m=j-k}^{n-k} P(U=k, V=m) + P(U \geq j) =$$

$$\sum_{k=i}^{j-1} \sum_{m=j-k}^{n-k} \frac{n!}{k!m!(n-k-m)!} [F_X(u)]^k [F_X(v) - F_X(u)]^m [1 - F_X((v)]^{n-k-m} + P(U \geq j)$$

Looking at this expression, we realize that when calculating $\frac{\partial^2}{\partial u \partial v} F_{X_{(i)},X_{(j)}}(u,v)$, the probability $P(U \geq j)$ is irrelevant since this term only depends on u and *not* on v. Carefully calculating the mixed partial derivative of the other term, yields the formula given in the Theorem.

Of course, joint densities of three or more order statistics could also be derived using similar arguments like above but the calculations will be more exhausting. We could use mathStatica's help to alleviate the calculations.

Note: The above derivations have very important applications. For instance, the density of the sample median, or of the sample range can now easily be derived by using standard formulae for density of transformed random variables. In the case of the range $R = X_{(n)} - X_{(1)}$, for example, we know from Theorem (7.4) that

$$f_{X_{(1)},X_{(n)}}(x,y) = n(n-1)[F_X(y) - F_X(x)]^{n-2} f_X(x) f_X(y)$$

and if we make the transformation $u = y - x, v = y$ (with absolute value of Jacobian equal to one), we obtain for the range: for all $u > 0$:

$$f_R(u) = \int_{-\infty}^{\infty} n(n-1)[F_X(v) - F_X(v-u)]^{n-2} f_X(v-u) f_X(v) dv$$

For the uniform distribution, the integrand is non-zero for the intersection of the regions $0 < v - u < 1$ and $0 < v < 1$ which is just the region $0 < u < v < 1$. Therefore after integration, we get in this case $f_R(u) = n(n-1)u^{n-2}(1-u)$ for $0 < u < 1$.
The "trick" in the above derivation was to introduce, on top of the transformation $u = y - x$ of interest, one more transformation ($v = y$) so that we could tranform the original

vector $(X_{(1)}, X_{(n)})$ in a new random vector (U, V) whose first component is the statistic of interest. This allows us, by using the density transformation formula for random vectors, to get the **joint** density of (U, V) first. Then we integrated out the unwanted variable V to obtain the marginal density of the main component of interest (the component U). Such a trick is used very often when working with order statistics. Clearly, the variable V plays an intermediate role here and in its choice, we are mainly guided by convenience of the calculations. For example, we could have chosen $v = x$ instead as a component of our tranformation. You are advised to go along similar lines by using this new transform instead, derive the marginal of U again and convince yourself that you get the same result for the density of the range.