Order Statistics

November 25, 2008

The **order statistics**

$$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$$

is the increasing ordered arrangement of the sample

$$X_1, X_2, \ldots X_n.$$

The **sample range** $R = X_{(n)} - X_{(1)}$. The **sample median**

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

Fix a value x and consider the Bernoulli trials

$$Y_1 = I_{(-\infty,x]}(X_1),\ Y_2 = I_{(-\infty,x]}(X_2),\ \ldots,\ Y_n = I_{(-\infty,x]}(X_n).$$

If F is the common cumulative distribution function for the X_i, then

$$p = P\{Y = 1\} = P\{X_1 \le x\} = F(x).$$

Let $S_n = Y_1 + Y_2 + \cdots + Y_n$, then $S_n \ge j$ if and only if at least j of the $X_i \le x$. In turn, this is true if and only if the j-th order statistic $X_{(j)} \le x$. Consequently,

$$P\{X_{(j)} \le x\} = P\{S_n \ge j\} = \sum_{k=j}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=j}^{n} \binom{n}{k} F(x)^k (1 - F(x))^{n-k}.$$

For X_i discrete random variables, the mass function,

$$f_{X_{(j)}}(x) = P\{X_{(j)} = x\} = P\{X_{(j)} \le x\} - P\{X_{(j)} < x\} = \sum_{k=j}^{n} \binom{n}{k} \left(F(x)^k (1 - F(x))^{n-k} - F(x-)^k (1 - F(x-))^{n-k}\right).$$

For X_i continuous random variables having common density function f, the density function,

$$f_{X_{(j)}}(x) = \frac{d}{dx} \left(\sum_{k=j}^{n} \binom{n}{k} F(x)^k (1 - F(x))^{n-k} \right).$$

Exercise 1.

$$\binom{n}{k+1}(k+1) = \binom{n}{k}(n-k).$$

Differentiating the individual terms

$$\begin{aligned}\frac{d}{dx}\left(\binom{n}{k}F(x)^k(1-F(x))^{n-k}\right) &= \binom{n}{k}\left(kF(x)^{k-1}(1-F(x))^{n-k} - (n-k)F(x)^k(1-F(x))^{n-k-1}\right)f(x)\\ &= \binom{n}{k}kF(x)^{k-1}(1-F(x))^{n-k}f(x) - \binom{n}{k+1}(k+1)F(x)^k(1-F(x))^{n-k-1}f(x)\end{aligned}$$

Summing the first term, we obtain

$$\sum_{k=j}^{n}\binom{n}{k}kF(x)^{k-1}(1-F(x))^{n-k}f(x).$$

Summing the second term, we obtain

$$\sum_{k=j}^{n}\binom{n}{k+1}(k+1)F(x)^k(1-F(x))^{n-k-1}f(x) = \sum_{k=j+1}^{n+1}\binom{n}{k}kF(x)^{k-1}(1-F(x))^{n-k}f(x).$$

The $k = n+1$ term in the second sum is zero. Thus, the difference leaves only the first term in the first sum.

$$f_{X_{(j)}}(x) = \binom{n}{j}jF(x)^{j-1}(1-F(x))^{n-j}f(x).$$

If the X_i are uniform random variables on the interval $[0,1]$, then

$$f_{X_{(j)}}(x) = \binom{n}{j}jx^{j-1}(1-x)^{n-j}, \qquad 0 \le x \le 1,$$

a beta random variable with parameters $\alpha = j$ and $\beta = n-j+1$

For continuous random variables, the probability that two random variables take on the same value $P\{X_i = X_j, \text{ for some } i \neq j\} = 0$.

To find the joint density of the order statistics, write π for an arbitrary permutation on $\{1,2,\ldots,n\}$. Each of these $n!$ permutations of an equal chance of being the order statistics for an given sample. Indeed, the mapping from the observations to the order statistics is an $n!$ to one map. In addition, the permutation matrix associated with π has determinant ± 1 depending on whether or not the permutation is even or odd. Because this is a linear map, the Jacobian J_π of the inverse transformation is equal to the inverse of the permutation matrix.

For example, assume that $X_1 = X_{(2)}, X_2 = X_{(3)}, X_3 = X_{(1)}$. The inverse transformation

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} X_{(1)} \\ X_{(2)} \\ X_{(3)} \end{bmatrix}.$$

The matrix is the Jacobian of the inverse transformation

Write $B_\pi = \{X_{\pi(1)} < X_{\pi(2)} < \cdots < X_{\pi(n)}\}$. Then,

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_{(1)},x_{(2)},\ldots,x_{(n)}) = \sum_\pi |\det(J_\pi)| f(x_{\pi(1)})f(x_{\pi(2)})\cdots f(x_{\pi(n)}) = n! f(x_1)f(x_2)\cdots f(x_n),$$

$x_{(1)} < x_{(2)} < \ldots < x_{(n)}$

Example 2. *Integrating to find the lower dimensional marginals is an arduous task. We shall shall that we can use a hierarchical model strategy to compute the density. We show this in the case of two order statistics and the uniform distribution on* $[0,1]$.

Write, for $i < j$,

$$f_{x_{(i)},X_{(j)}}(u,v) = f_{X_{(i)}|X_{(j)}}(u|v) f_{X_{(j)}}(v)$$

From the computation above, we know that

$$f_{X_{(j)}}(v) = \binom{n}{j} j v^{j-1}(1-v)^{n-j}, \qquad 0 \le v \le 1,$$

Given that $X_{(j)} = v$, *we have* $j-1$ *random variables uniformly distributed on* $[0,v]$. *Among these random variable, we are looking for the ith order statistics. Now,*

$$F_X(u|X \le v) = \frac{u}{v}, \quad \text{for } 0 \le u \le v.$$

Thus,

$$f_{X_{(i)}|X_{(j)}}(u|v) = \binom{j-1}{i} i \left(\frac{u}{v}\right)^{i-1} \left(\frac{v-u}{v}\right)^{j-i-1}.$$

Now multiply

$$\begin{aligned}
f_{x_{(i)},X_{(j)}}(u,v) &= \binom{j-1}{i} i \left(\frac{u}{v}\right)^{i-1} \left(\frac{v-u}{v}\right)^{j-i-1} \cdot \binom{n}{j} j v^{j-1}(1-v)^{n-j} \\
&= \binom{n}{j} j \binom{j-1}{i} i u^{i-1}(v-u)^{j-i-1}(1-v)^{n-j} \\
&= \frac{n!}{(j-1)!(n-j)!} \frac{(j-1)!}{(i-1)!(j-i-1)!} u^{i-1}(v-u)^{j-i-1}(1-v)^{n-j} \\
&= \frac{n!}{(n-j)!(i-1)!(j-i-1)!} u^{i-1}(v-u)^{j-i-1}(1-v)^{n-j}
\end{aligned}$$

Note that the coefficient is a trinomial coefficient. This (correctly) suggests that an approach using a trinomial distribution similar to the strategy used above with the binomial distribution will succeed.