Order statistics - derivations

Let $X_1, X_2, \cdots, X_n$ denote independent continuous random variables with cdf $F(x)$ and pdf $f(x)$. We will denote the ordered random variables with $X_{(1)}, X_{(2)} , \cdots, X_{(n)}$, where $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$.

Probability density function of the $j_{th}$ order statistic:

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

Proof:

We will find the cdf of the $j_{th}$ order statistic and then the pdf by taking the derivative of the cdf. The cdf is denoted by $F_{X_{(j)}}(x) = P(X_{(j)} \leq x)$. Now let’s introduce a discrete random variable $Y$ that counts the number of variables less than or equal to $x$. The statement $P(X_{(j)} \leq x)$ is the same as $P(Y \geq j)$. Why? If we call “success” the event $X_i \leq x$ then $Y \sim b(n, p)$ or $Y \sim b(n, F_X(x))$.

$$F_{X_{(j)}}(x) = P(X_{(j)} \leq x) = P(Y \geq j) = \sum_{k=j}^{n} \binom{n}{k} p^k (1 - p)^{n-k}$$

Now the pdf:

$$g_{X_{(j)}}(x) = \frac{dF_{X_{(j)}}(x)}{dx} = \sum_{k=j}^{n} \binom{n}{k} k F_X(x)^{k-1} f_X(x) [1 - F_X(x)]^{n-k}$$

$$= \sum_{k=j}^{n} \binom{n}{k} (n-k) F_X(x)^k [1 - F_X(x)]^{n-k-1} f_X(x)$$

$$= \left( \binom{n}{j} j f_X(x) F_X(x)^{j-1} [1 - F_X(x)]^{n-j} \right) \quad \text{(when } k = j \text{)}$$

$$+ \sum_{k=j+1}^{n} \binom{n}{k} k F_X(x)^{k-1} f_X(x) [1 - F_X(x)]^{n-k}$$

$$- \sum_{k=j}^{n-1} \binom{n}{k} (n-k) F_X(x)^k [1 - F_X(x)]^{n-k-1} f_X(x)$$

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

$$+ \sum_{k=j}^{n-1} \binom{n}{k+1} (k + 1) F_X(x)^k f_X(x) [1 - F_X(x)]^{n-k-1}$$

$$- \sum_{k=j}^{n-1} \binom{n}{k} (n-k) F_X(x)^k [1 - F_X(x)]^{n-k-1} f_X(x)$$

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

Note: $\binom{n}{k+1}(k+1) = \binom{n}{k}(n-k)$, so the last 2 terms before the last line cancel!
An intuitive derivation of the density function of the \(j_{th}\) order statistic. This intuitive derivation is based on this result \(P(y \leq Y \leq y + dy) \approx f(y)dy\).

Consider the \(j_{th}\) order statistic \(X_{(j)}\). If \(X_{(j)}\) is in the neighborhood of \(x\) then there are \(j - 1\) random variables less than \(x\), each one with probability \(p_1 = P(X \leq x) = F_X(x)\), 1 random variable near \(x\), with probability \(p_2 = P(x \leq X \leq x + dx) \approx f_X(x)dx\), and \(n - j\) random variables larger than \(x\), with probability \(p_3 = P(X > x) = 1 - P(X \leq x) = 1 - F_X(x)\).

Therefore,
\[
P(x \leq X_{(j)} \leq x + dx) \approx g_{X_{(j)}}(x)dx
= \left(\begin{array}{c} n \\
 j - 1, 1, n - j
\end{array}\right) p_1^{j-1} p_2 p_3^{n-j} \quad \text{(multinomial distribution)}
= \frac{n!}{(j-1)!(n-j)!} F_X(x)^{j-1} f_X(x)dx [1 - F_X(x)]^{n-j}.
\]

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

Using this intuitive derivation we can now find the joint probability density function of \(X_{(i)}, X_{(j)}\). Using the same approximation as above, \(P(u \leq X_{(i)} \leq u + du, v \leq X_{(j)} \leq v + dv) \approx g_{X_{(i)}, X_{(j)}}(u, v)du dv\). For \(u < v\) we need to have the following arrangement:

\[
\begin{array}{c|c|c}
   & i - 1 & 1 \\
\hline
   j - 1 & \text{Random variables less than } u, \text{ each one with probability } p_1 = P(X \leq u) = F_X(u) & \text{Random variables near } u \text{ with probability } p_2 = P(u \leq X \leq u + du) \approx f_X(u)du \\
   j - i & \text{Random variables between } u \text{ and } v \text{ with probability } p_3 = P(u \leq X \leq v) = F_X(v) - F_X(u) & \text{Random variables near } v \text{ with probability } p_4 = P(v \leq X \leq v + dv) \approx f_X(v)dv \\
   n - j & \text{Random variables larger than } v, \text{ each one with probability } p_5 = P(X > v) = 1 - F_X(v) & \\
\end{array}
\]

Using the multinomial distribution we have:
\[
P(u \leq X_{(i)} \leq u + du, v \leq X_{(j)} \leq v + dv) \approx g_{X_{(i)}, X_{(j)}}(u, v)du dv
\]
\[
= \left(\begin{array}{c} n \\
 i - 1, j - 1 - i, 1, n - j
\end{array}\right) p_1^{i-1} p_2^{j-1-i} p_3^{1} p_4^{1} p_5^{n-j}
\]

Therefore,
\[
g_{X_{(i)}, X_{(j)}}(u, v)du dv = \left(\begin{array}{c} n \\
 i - 1, j - 1 - i, 1, n - j
\end{array}\right) F_X(u)^{i-1} f_X(u)du [F_X(v) - F_X(u)]^{j-1-i} f_X(v)dv [1 - F_X(v)]^{n-j}
\]
or
\[
g_{X_{(i)}, X_{(j)}}(u, v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} F_X(u)^{i-1} f_X(u) [F_X(v) - F_X(u)]^{j-1-i} f_X(v)[1 - F_X(v)]^{n-j}.
\]