Fall 2018 STSCI 5080 Supplemental Material 2 (9/27)

Densities of order statistics

The purpose of this supplementary material is to the prove the following theorem, which is Theorem A in Rice p. 105.

Theorem 1. *Let $X_1, \dots, X_n$ be independent continuous random variables with common pdf f, and let $X_{(1)} < \cdots < X_{(n)}$ be order statistics of $X_1, \dots, X_n$. For any $k = 1, \dots, n$, the pdf of $X_{(k)}$ is*

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

where F denotes the cdf corresponding to f.

Proof. Fix any $x \in \mathbb{R}$, and let $p = F(x)$. For $i = 1, \dots, n$, define

$$Z_i = \begin{cases} 1 & \text{if } X_i \le x \\ 0 & \text{otherwise} \end{cases}.$$

The variables $Z_1, \dots, Z_n$ are independent Bernoulli trials with success probability p (since $P(Z_i = 1) = P(X_i \le x) = F(x) = p$). So, the variable

$$Y = \sum_{i=1}^{n} Z_i$$

follows the binomial distribution with parameters n and p. Now, since

$$\begin{aligned} X_{(k)} \le x &\Leftrightarrow \text{the } k \text{ smallest variables among } X_1, \dots, X_n \text{ are at most } x \\ &\Leftrightarrow \text{there are at least } k \text{ variables among } X_1, \dots, X_n \text{ that are at most } x \\ &\Leftrightarrow Y \ge k, \end{aligned}$$

we have

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

We will differentiate $F_{X_{(k)}}(x)$ to find the pdf of $X_{(k)}$. To this end, define a function

$$p(j, m) = \binom{m}{j} p^j (1-p)^{m-j}, \ j = 0, 1, \dots, m,$$

and $p(j, m) = 0$ for $j > m$. We note that for $j = 1, \dots, n-1$,

$$\begin{aligned} \frac{d}{dp} \binom{n}{j} p^j (1-p)^{n-j} &= \frac{n!}{(j-1)!(n-j)!} p^{j-1} (1-p)^{n-j} - \frac{n!}{j!(n-j-1)!} p^j (1-p)^{n-j-1} \\ &= n\{p(j-1, n-1) - p(j, n-1)\}. \end{aligned}$$

For $j = n$,

$$\frac{d}{dp}\binom{n}{n}p^n = np^{n-1} = np(n-1, n-1).$$

Therefore, we have

$$\begin{aligned}
f_{X_{(k)}}(x) = \frac{d}{dx}F_{X_{(k)}}(x) &= nf(x)\left\{\sum_{j=k}^{n-1}\{p(j-1, n-1) - p(j, n-1)\} + p(n-1, n-1)\right\} \\
&= nf(x)\{p(k-1, n-1) - p(n-1, n-1) + p(n-1, n-1)\} = nf(x)p(k-1, n-1) \\
&= \frac{n!}{(k-1)!(n-k)!}F(x)^{k-1}\{1 - F(x)\}^{n-k}.
\end{aligned}$$

This completes the proof. $\square$