Math 502AB - Lecture 16

Dr. Jamshidian

October 23, 2017

1 Lecture - Part 1

1.1 Section 5.4 - Order Statistics

Think given the sample $X_1, X_2, ..., X_n$, consider the ordered sample:

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

From these orders, you can consider:

1. $X_{(1)} = min(X_1, ..., X_n)$
2. $X_{(n)} = max(X_1, ..., X_n)$
3. Median:

$$\begin{cases} X_{\left(\frac{n+1}{2}\right)} & n \text{ is odd} \\ \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} & n \texttt{ is even} \end{cases}$$

We can calculate the **Distribution of** $X_{(n)}$:

$$\begin{aligned} F_{X_{(n)}} &= P\left(X_{(n)} \leq x\right) \\ &= P\left(max(X_1, ..., X_n) \leq x\right) \\ &= P\left(X_1 \leq x, X_2 \leq x, ..., X_n \leq x\right) \\ &= \prod_{i=1}^{n} P(X_i \leq x) = \left(F_{x_1}(x)\right)^n \\ f_{x_{(n)}}(x) &= n\left(F_{x_1}(x)\right)^{n-1} f_{x_1}(x) \end{aligned}$$

We can calculate the **Distribution of** $X_{(1)}$:

$$\begin{aligned}
F_{X_{(1)}} &= P\left(X_{(1)} \leq x\right) \\
&= 1 - P\left(X_{(1)} \geq x\right) \\
&= 1 - P\left(min(X_1, ..., X_n) > x\right) \\
&= 1 - P\left(X_1 > x, X_2 > x, ..., X_n > x\right) \\
&= 1 - \prod_{i=1}^{n} P(X_i > x) = 1 - \left(1 - F_{x_1}(x)\right)^n \\
f_{x_{(1)}}(x) &= n f_{x_1}(x) \left(1 - F_{x_1}(x)\right)^{n-1}
\end{aligned}$$

Example:

Suppose that n system components are connected in a series, so that if one of the components fails. Let $T_1, ..., T_n$ denote the lifetime of the components, with:

$$T_i \sim exp(\lambda) \qquad T_i \sim iid$$

Let V be the length of time that the system operates. Obtain the distribution of V.

$$\begin{aligned}
V &= min(T_1, T_2, ..., T_n) \\
f_V(x) &= n \frac{1}{\lambda} e^{-x/\lambda} \left(e^{-x/\lambda}\right)^{n-1} \qquad x > 0
\end{aligned}$$

Theorem:

The *pdf* of the k^{th} order statistic $X_{(k)}$ is:

$$f_{x_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} f(x) \left(F(x)\right)^{k-1} \left[1 - F(x)\right]^{n-k}$$

Proof:

The differential argument is based on:

$$f_{x_{(k)}}(x) dx \cong \int_x^{x+dx} f_{X_{(k)}}(t) dt$$

Essentially, given an arbitrarily small differential, we want to calculate the probability that $x < X_{(k)} < x + dx$. So:

$$P\left(x < X_{(k)} < x + dx\right) \cong \int_x^{x+dx} f_{X_{(k)}}(x) dx$$

On the other hand, in order for $X_{(k)}$ to fall in this range, we need $k-1$ observations below x, and $n-k$ observations above $x + dx$. The probability that something falls below x is $F_X(x)$, and the probability that something is above $x + dx$ is $1 - F_X(x + dx)$. This forms a *multinomial* distribution:

$$
\begin{aligned}
P(x < X_k < x + dx) &= \binom{n}{k-1, 1, n-k} \left(F_X(x)\right)^{k-1} f(x)dx \left(1 - F_X(x + dx)\right)^{n-k} \\
&\cong \binom{n}{k-1, 1, n-k} \left(F_X(x)\right)^{k-1} \left(1 - F_X(x + dx)\right)^{n-k} f(x)dx
\end{aligned}
$$

Example:

If $X_1, ..., X_n \sim unif(0, 1)$ (**iid**), then the k^{th} order statistic is:

$$
\begin{aligned}
f_{x_{(k)}}(x) &= \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k} \quad 0 \le x \le 1 \\
X_{(k)} &\sim Beta(\alpha = k, \beta = n - k + 1) \\
E\left[X_{(k)}\right] &= \frac{k}{n+1}
\end{aligned}
$$

Let $n = 11$. If we wanted to find the *median*, $X_{(6)}$, we will have:

$$
E\left[X_{(6)}\right] = \frac{6}{12}
$$

2 Lecture - Part 2

2.1 Joint Distributions of Order Statistics

Theorem:

Let $X_1, ..., X_n$ be *iid* from a distribution with *pdf* $f(x)$ and *cdf* $F(x)$. Then, the joint density of $X_{(i)}$ and $X_{(j)}$, $1 \le i < j \le n$ is:

$$
\begin{aligned}
f_{X_{(i)}, X_{(j)}}(u, v) =& \frac{n!}{(i-1)!(j-i-1)!(n-j)!} f(u)f(v) \left[F(u)\right]^{i-1} \\
& \times \left[F(v) - F(u)\right]^{j-i-1} \left[1 - F(v)\right]^{n-j}
\end{aligned}
$$

Proof:

Again, we can use a similar differential argument. We have:

$$
\begin{aligned}
f_{X_{(i)}, X_{(j)}}(u, v) \cong& \binom{n}{i-1, j-i-1, 1, n-j} \left[F(u)\right]^{i} - 1 f(u)du \left[F(v) - F(u + du)\right]^{j-i-1} \\
& \times f(v)dv \left(1 - F(v + dv)\right)^{n-j}
\end{aligned}
$$

Example:

Let $X_1, ..., X_n \sim Unif(0,1)$ (**iid**). Obtain the distribution of the range:

$$R = X_{(n)} - X_{(1)}$$

First, we need the joint distribution:

$$\begin{aligned} f_{X_{(1)},X_{(n)}}(u,v) &= n(n-1)f(u)f(v)[F(v)-F(u)]^{n-2}, \quad u \leq v \\ &= n(n-1)(v-u)^{n-2} \end{aligned}$$

Now, we need the *cdf*:

$$\begin{aligned} F_R(r) &= P(R \leq r) \\ &= P(V - U \leq r) \\ &= P(V \leq U + r) \end{aligned}$$

We now integrate this:

$$\begin{aligned} * &= \int_0^{1-r} \int_u^{r+u} n(n-1)(v-u)^{n-2} dvdu + \int_{1-r}^1 \int_u^1 n(n-1)(v-u)^{n-2} dvdu \\ &= nr^{n-1} - (n-1)r^n \end{aligned}$$

And we end up with:

$$\begin{aligned} f_R(r) &= n(n-1)r^{n-2}(1-r) \quad 0 \leq r \leq 1 \\ R &\sim Beta(n-1, 2) \end{aligned}$$

Theorem:

For $X_1 \leq X_2 \leq ... \leq X_n$:

$$f_{X_{(1)},...,X_{(n)}}(x_1, x_2, ..., x_n) = n! f(x_1) f(x_2) \cdots f(x_n)$$

Proof (Consider n=3):

Let $x_1 \leq x_2 \leq x_3$. Consider the *CDF*:

$$F_{X_{(1)},X_{(2)},X_{(3)}}(x_1, x_2, x_3) = P(X_{(1)} \leq x_1, X_{(2)} \leq x_2, X_{(3)} \leq x_3)$$

Let N_1 be the number of x_1, x_2, x_3 that fall below x_1.
Let N_2 be the number of x_1, x_2, x_3 that fall between x_1 and x_2.
Let N_3 be the number of x_1, x_2, x_3 that fall between x_2 and x_3.

We are observing three values.

1. One possibility is that all three fall below the first value:
   $(N_1 = 3, N_2 = 0, N_3 = 0)$
   $$P = (F(x_1))^3$$
2. One possibility is to get two below the first value, one between x_1 and x_2:
   $(N_1 = 2, N_2 = 1, N_3 = 0)$
   $$P = \binom{3}{2} (F(x_1))^2 (F(x_2) - F(x_1))$$
3. One possibility is to get two below the first value, one between x_2 and x_3:
   $(N_1 = 2, N_2 = 0, N_3 = 1)$
   $$P = \binom{3}{2} (F(x_1))^2 (F(x_3) - F(x_2))$$
4. One possibility is to get one below the first value, two between x_1 and x_2:
   $(N_1 = 1, N_2 = 2, N_3 = 0)$
   $$P = \binom{3}{2} (F(x_1)) (F(x_2) - F(x_1))^2$$
5. One possibility is to get one below the first value, one between x_1 and x_2, and one between x_2 and x_3:
   $(N_1 = 1, N_2 = 1, N_3 = 1)$
   $$3! \cdot F(x_1)(F(x_2) - F(x_1))(F(x_3) - F(x_2))$$

It turns out that the **CDF** we wish to find is the sum of all of these probabilities.