STAT 345/445 Lecture 19

Order Statistics – Section 5.4

Order Statistics

Let $X_1, X_2, \ldots, X_n$ be a random sample. The statistics

- $X_{(1)} = \min\{X_1, X_3, \ldots, X_n\}$
- $X_{(2)} =$ second smallest$\{X_1, X_3, \ldots, X_n\}$
- $\ldots$
- $X_{(n-1)} =$ second largest$\{X_1, X_3, \ldots, X_n\}$
- $X_{(n)} = \max\{X_1, X_3, \ldots, X_n\}$

are called **order statistics**.

Get used to the $X_{(j)}$ notation and $X_j \neq X_{(j)}$

Could have used
$U_1 = \min(X_1, \ldots, X_n)$
$\vdots$
$U_n = \max(X_1, \ldots, X_n)$

Examples:

- Weight of smallest kitten in a litter
- Highest score on an exam

We want to find the *(sampling) distributions* of order statistics

Order Statistics - Example

- Random sample $X_1, X_2, \ldots, X_{10}$
- Order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(10)}$

Observed random sample:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1.9	3.5	9.0	2.8	3.3	2.1	6.1	1.1	0.7	7.9

Corresponding observed order statistics:

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	$x_{(5)}$	$x_{(6)}$	$x_{(7)}$	$x_{(8)}$	$x_{(9)}$	$x_{(10)}$
0.7	1.1	1.9	2.1	2.8	3.3	3.5	6.1	7.9	9.0

Distributions of order statistics

Theorem: distribution of order statistics

Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution, with pdf $f(x)$ and cdf $F(x)$. Then the cdf and pdf of the jth order statistic $X_{(j)}$ are

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

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

proof... Lets take $j = 1$ and $j = n$ first

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

Order Statistics as transformations

- Transformations from a random sample to order statistics

$$(X_1, X_2, \ldots, X_n) \longmapsto X_{(j)}$$
$$\text{or } (X_1, X_2, \ldots, X_n) \longmapsto (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$$

  are not one-to-one transformations

- **Example:** $n = 3$ and $(X_1, X_2, X_3) \longmapsto (X_{(1)}, X_{(2)}, X_{(3)})$
  - The following observations of (X_1, X_2, X_3)

$$(3,5,8),\ (3,8,5),\ (5,3,8),\ (5,8,3),\ (8,3,5),\ (8,5,3)$$

    are all mapped to $(X_{(1)}, X_{(2)}, X_{(3)}) = (3,5,8)$

- So we can't use out transformation formulas to obtain $f_{(j)}(x)$
  - So use the "cdf-method" to find $F_{(j)}(x)$ $= P(X_{(j)} < x)$

Untangling expressions in the Theorem

- Recall that by the Binomial formula

$$\sum_{k=0}^{n} \binom{n}{k} F(x)^k \, (1 - F(x))^{n-k} = (F(x) + 1 - F(x))^{n} = 1$$

- So the cdf $F_{(1)}(x)$ can be written as

$$\begin{aligned} F_{(1)}(x) &= \sum_{k=1}^{n} \binom{n}{k} F(x)^k (1 - F(x))^{n-k} \qquad \text{n-1 terms} \\ &= \sum_{k=0}^{n} \binom{n}{k} F(x)^k (1 - F(x))^{n-k} - \binom{n}{0} F(x)^0 (1 - F(x))^{n-0} \\ &= 1 - (1 - F(x))^n \end{aligned}$$

add and subtract for term k=0 →

and

$$\begin{aligned} f_{(1)}(x) &= \frac{n!}{(1-1)!(n-1)!} f(x) \, F(x)^{1-1} \, (1 - F(x))^{n-1} \\ &= n \, f(x) \, (1 - F(x))^{n-1} = \frac{d}{dx} F_{(1)}(x) \end{aligned}$$

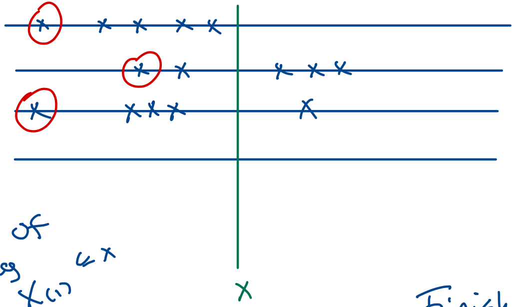
Distribution of $X_{(1)}$

- Have the joint pdf of $X_1, X_2, \dots, X_n$
- How can we relate $X_{(1)}$ to $(X_1, X_2, \dots, X_n)$?
- "Smallest value is bigger than x" is the same event as "all values are bigger than x"

$X_{(1)} \le x$

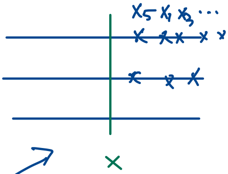
$X_{(1)} > x$?

$n = 5$



etc.

many ways of getting $X_{(1)} < x$



all X's have to be $> x$

Finished on the whiteboard

Distribution of $X_{(n)}$

- Theorem for $j = n$:

$$\frac{d}{dx} F_{(n)}(x) = \frac{d}{dx} F(x)^n = n\, F(x)^{n-1} \cdot f(x)$$

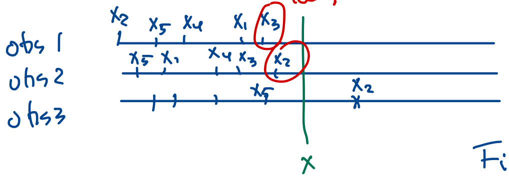
chain rule.

$$F_{(n)}(x) = \binom{n}{n} F(x)^n (1 - F(x))^{n-n} = F(x)^n$$

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

- How can we relate $X_{(n)}$ to $(X_1, X_2, \ldots, X_n)$?
- "Largest value is less than x" is the same event as "all values are less than x"

Event $X_{(n)} \le x$ $\quad n = 5$



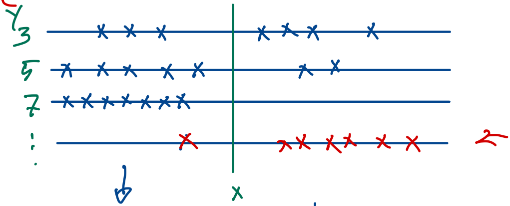
← here $X_{(n)} > x$

Finished on the whiteboard

- How can we relate $X_{(j)}$ to $(X_1, X_2, \ldots, X_n)$?
- "$X_{(j)} \leq x$" is the same event as
  "all least j of the random variables are less than x"

j or more

$n = 7 \quad j = 3 \qquad X_{(3)} \leq x$



$X_{(3)} > x$

have at least 3 obs. $\leq x$

$F_{(j)}(x)$ on the whiteboard

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

showed last time.

$(uv)' = u'v + uv'$

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

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

$$= \sum_{k=j}^{n} \Big[\binom{n}{k} k\, F(x)^{k-1} f(x) (1-F(x))^{n-k} - \binom{n}{k} F(x)^k (n-k\cancel{-1}) (1-F(x))^{n-k-1} f(x) \Big]$$

support of $f_{(j)}(x)$?

same as $X_1, \dots, X_n$

$= \dots$ some algebra, term cancel...

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

Example: Order stats for Uniform

- Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be order statistics from a random sample $X_1, X_2, \ldots, X_n$ from the $\mathrm{Uniform}(0, 1)$. What are the (marginal) pdfs of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$?

On the whiteboard...

Example: Order stats for Uniform

- Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be order statistics from a random sample $X_1, X_2, \ldots, X_n$ from the $\mathrm{Uniform}(0, 1)$.
- Then

$$f_{(j)}(x) = \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)!} x^{j-1}(1-x)^{(n-j+1)-1}$$

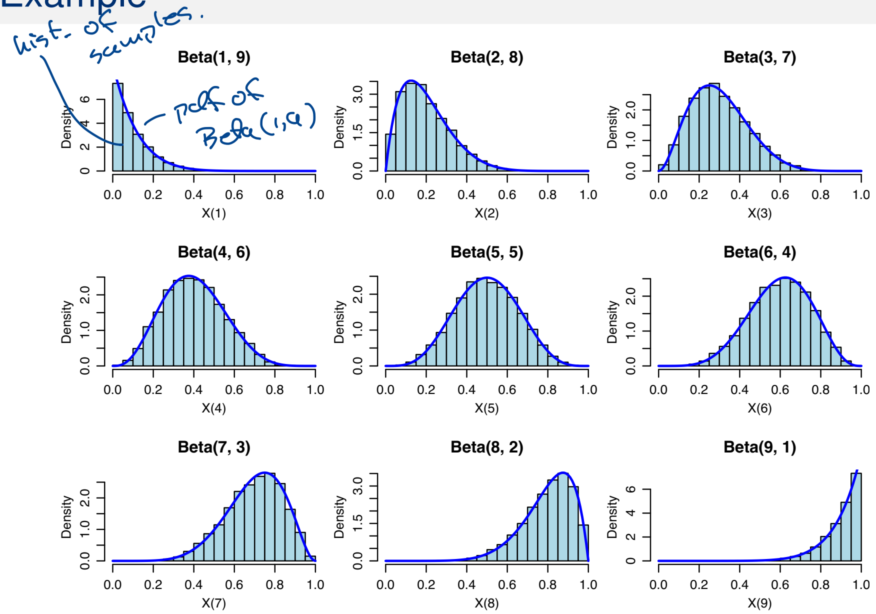
So $X_{(j)} \sim \mathrm{Beta}(j, n-j+1)$

- R simulation:

```
x.order <- c()        ← makes an empty vector
for(i in 1:10000){
sample → x <- runif(n=9)        n=9
  x.order <- rbind(x.order, x[order(x)])
}                                      ordering elements of x
```

1st col. of x.order are samples of $X_{(1)}$ etc.

Example



hist. of samples.
pdf of Beta(1,9)
Beta(1, 9)
Beta(2, 8)
Beta(3, 7)
Beta(4, 6)
Beta(5, 5)
Beta(6, 4)
Beta(7, 3)
Beta(8, 2)
Beta(9, 1)
Density
X(1)
X(2)
X(3)
X(4)
X(5)
X(6)
X(7)
X(8)
X(9)
0.0 0.2 0.4 0.6 0.8 1.0

Joint distribution of order statistics

e.g. range: $X_{(n)} - X_{(1)}$

Joint pdf of two order statistics

Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be order statistics from a random sample $X_1, X_2, \ldots, X_n$ with pdf $f(x)$ and cdf $F(x)$.
The joint pdf of $\left(X_{(i)}, X_{(j)}\right)$, $1 \le i < j \le n$ is

$$f_{(i,j)}(u, v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f(u)\, f(v) \times F(u)^{i-1} \left(F(v) - F(u)\right)^{j-1-i} \left(1 - F(v)\right)^{n-j}$$

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

Think: 3-dimensional Multinomial distribution

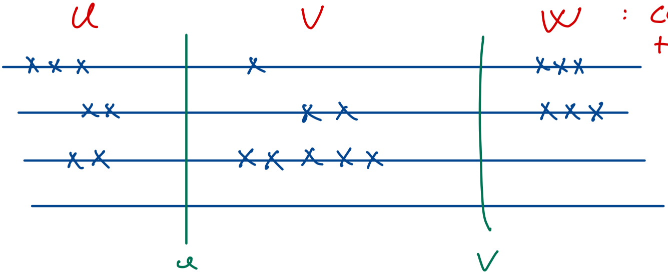
very similar argument but with multinomial instead of Binomial

"cdf method"
Joint cdf of $X_{(i)}, X_{(j)}$ $i<j$:

$$F_{(ij)}(u,v) = P(X_{(i)} \le u, X_{(j)} \le v) \quad u < v$$

$$= P(U \ge i, U+V \ge j)$$

e.g. $n=7$ $i=2$, $j=4$



: count in the three ranges

see Ex. 5.26 for details.

$X_{(i)} \le u$ and $X_{(j)} \le v$: same as at least i of $X_1, \dots, X_n$ are $\le u$ and at least j of $X_1, \dots, X_n$ are $\le v$

$$(U, V, W) \sim \text{Multinomial}(m=n, p=(F(u), F(v)-F(u), 1-F(v)))$$

Joint distribution of order statistics

Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution, with pdf $f(x)$ and cdf $F(x)$.

Joint pdf for all n order statistics $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$:

$$f_{(1),\ldots,(n)}(u_1, \ldots, u_n) = n! f(u_1) f(u_2) \cdots f(u_n)$$

for $-\infty < u_1 < u_2 < \cdots < u_n < \infty$ proof in book...

Can get any marginal and joint marginal pdfs by integrating out other variables.

see book for discrete order stats.