6.434/16.391 Statistics for Engineers and Scientists

Lecture 4   09/17/2012

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

Lecture 4 09/17/2012

DEFINITION OF ORDER STATISTICS

Definition of order statistics

- If $X_1,\, X_2,\, \ldots\, X_n$ are arranged in ascending order of magnitudes as $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$, where $X_{(r)}$ is the rth order statistics, $r = 1, 2, \ldots, n$.

- Notation
  - $X_1,\, X_2,\, \ldots\, X_n$: Unordered variables
  - $x_1,\, x_2,\, \ldots\, x_n$: Unordered observation
  - $X_{(1)},\, X_{(2)},\, \ldots,\, X_{(n)}$: Ordered variables
  - $x_{(1)},\, x_{(2)},\, \ldots,\, x_{(n)}$: Ordered observation
  - $X_{[1]},\, X_{[2]},\, \ldots,\, X_{[n]}$: Ordered variables in descending order
  - $x_{[1]},\, x_{[2]},\, \ldots,\, x_{[n]}$: Ordered observation in descending order
  - $X_{1:n},\, X_{2:n},\, \ldots,\, X_{n:n}$: Extensive forms

Applications of order statistics

- Antenna selection in wireless links
- Satellite relays
- Route selection for transportation

Lecture 4 09/17/2012

DISTRIBUTION OF LARGEST AND SMALLEST ORDER STATISTIC

Distribution of a single order statistic

- Let $X_1, X_2, \ldots X_n$ be independent variables with the same cumulative distribution function (CDF) $F_X(x)$.

- Let $F_r(x)$, $r = 1, 2, \ldots, n$, denote the CDF of the rth order statistics

Distribution of the largest order statistic

- The CDF of the largest order statistic

$$\begin{aligned} F_n(x) &= \mathbb{P}\left\{X_{(n)} \leq x\right\} \\ &= \mathbb{P}\left\{X_i \leq x,\, \forall i\right\} \\ &= \prod_{i=1}^{n} \mathbb{P}\left\{X_i \leq x\right\} \quad \text{(independent)} \\ &= \left[\mathbb{P}\left(X_i \leq x\right)\right]^n \quad \text{(identically distributed)} \\ &= \left[F_X(x)\right]^n \end{aligned}$$

- The pdf of the largest order statistic

$$f_n(x) = \frac{d}{dx} F_X(x) = n\left[F_X(x)\right]^{n-1} f_X(x)$$

Lecture 4 09/17/2012

DISTRIBUTION OF ANY SINGLE ORDER STATISTIC

Distribution of the smallest ordered statistic

- The CDF of the smallest order statistic

$$\begin{aligned} F_1(x) &= P(X_{(1)} \leq x) \\ &= 1 - P(X_{(1)} > x) \\ &= 1 - P(X_i > x, \forall i) \\ &= 1 - \left[1 - \underbrace{P(X_i < x)}_{F_X(x)}\right]^n \end{aligned}$$

- The pdf of the smallest order statistic

$$f_1(x) = n \left[1 - F_X(x)\right]^{n-1} f_X(x)$$

Distribution of the rth ordered statistic

- The CDF of the rth ordered statistic

$$F_r(x) = P\left(X_{(r)} \leq x\right)$$

- Approach #1: Tail probability of binomial distribution
- Approach #2: Differential approach
- Approach #3: Marginalization approach

Tail probability (1 of 3)

- First, we have

$$\begin{aligned}
F_r(x) &= P\left(X_{(r)} \leq x\right) \\
&= P(\text{At least } r \text{ of the } X_i\text{'s} \leq x) \\
&= \sum_{j=r}^{n} P(\text{Exactly } j \text{ of } X_i\text{'s are} \leq x) \\
&= \sum_{j=r}^{n} \binom{n}{j} \left[F_X(x)\right]^j \left[1 - F_X(x)\right]^{n-j}
\end{aligned}$$

which is the tail probability of binomial distribution with $F(x)$ as success probability and n the number of trials, where

$$\binom{n}{j} \left[F_X(x)\right]^j \left[1 - F_X(x)\right]^{n-j}$$

is the probability of j successes

Tail probability (2 of 3)

- Using the identity

$$\sum_{j=r}^{n} \binom{n}{j} p^j (1-p)^{n-j} = \int_0^p \frac{n!}{(r-1)!(n-r)!} t^{r-1}(1-t)^{n-r} dt$$

we have

$$F_r(x) = r\binom{n}{r} \int_0^{F_X(x)} t^{r-1}(1-t)^{n-r} dt$$

$$= \frac{1}{B(r, n-r+1)} \int_0^{F_X(x)} t^{r-1}(1-t)^{n-r} dt$$

Recall $B(\alpha_1, \alpha_2) = \dfrac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}$ and $\Gamma(n) = (n-1)!$

Tail probability (3 of 3)

- The pdf of r th order statistic is

$$f_r(x) = \frac{d}{dx} F_r(x)$$

- Using Leibniz rule

$$\frac{d}{dx} \int_{a(x)}^{b(x)} g(t,x)dt = \int_{a(x)}^{b(x)} \frac{d}{dx} g(t,x)dt + g(b,x)\frac{db}{dx} - g(a,x)\frac{da}{dx}$$

we have

$$f_r(x) = \frac{n!}{(r-1)!(n-r)!} \left[F_X(x)\right]^{r-1} \left[1 - F_X(x)\right]^{n-r} f_X(x)$$

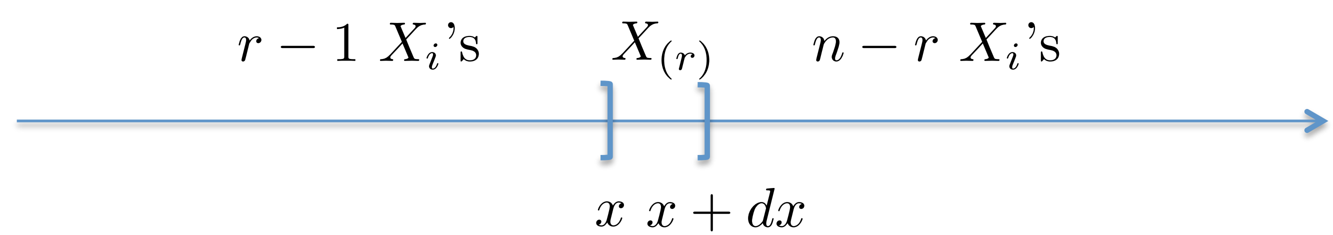
- Note: The results on CDF hold for both continuous and discrete cases. For pdf we use $f_r(x) = \frac{d}{dx} F_r(x)$, which assumes that X_i ‘s are continuous random variables.

Differential approach (1 of 3)

- In view of the importance of the order statistics, we will give alternative derivation for continuous random variables

- Consider the event $\{x \le X_{(r)} \le x + dx\}$. It can be written as

$$\begin{aligned}\{x \le X_{(r)} \le x + dx\} = \{&X_i < x,\ \text{for } r-1 \text{ of } X_i\text{'s};\\ &x \le X_i \le x + dx \text{ for exactly one of } X_i\text{'s};\\ &X_i > x + dx,\ \text{for remaining } n-r\ X_i\text{'s}\} \cup\\ &\{\text{events with probability of } o((dx)^2)\}\end{aligned}$$



Differential approach (2 of 3)

- The number of ways that n observations can be divided in the following three groups
  - one group with $r-1$ elements
  - one group with 1 elements
  - one group with $n-r$ elements

  is given by

$$\frac{n!}{(r-1)!(n-r)!} = \frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} = \frac{1}{B(r,n-r+1)}$$

- Each way has probability

$$[F_X(x)]^{r-1} f_X(x)dx\,[1-F_X(x)]^{n-r}$$

Differential approach (3 of 3)

- Thus, for a small dx

$$P(x \leq X_{(r)} \leq x + dx)$$
$$= \frac{n!}{(r-1)!(n-r)!} \left[F_X(x)\right]^{r-1} f_X(x) dx \left[1 - F_X(x)\right]^{n-r} + o((dx)^2)$$

where $o((dx)^2)$ counts for the probability of the realization of the event $\{x \leq X_{(r)} \leq x + dx\}$ in which more than one of X_i's are in $[x, x + dx]$

- Dividing both side by dx and let $dx \to 0$

$$F_r(x) = \lim_{dx \to 0} \frac{\mathbb{P}\{x \leq X_r \leq x + dx\}}{dx}$$
$$= \frac{n!}{(r-1)!(n-r)!} \left[F_X(x)\right]^{r-1} \left[1 - F_X(x)\right]^{n-r} f_X(x)$$

Joint distribution of order statistic

- Let $X_1,\, X_2,\, \ldots\, X_n$ be i.i.d. random sample. Define $Y_i = X_{(i)}$, $i = 1, 2, \ldots, n$. Then the joint distribution of $Y_1,\, Y_2,\, \ldots\, Y_n$ is

$$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = \begin{cases} n! \prod_{i=1}^{n} f_X(y_i), & -\infty < y_1 < y_2 < \ldots < y_n < \infty \\ 0, & \text{otherwise} \end{cases}$$

  – Proof idea: for $y_1 < \ldots < y_n$, we have

$$P(y_i \leq Y_i \leq y_i + dy_i,\ \forall i) = n! \prod_{i=1}^{n} f_X(y_i) dy_i$$

  Thus,

$$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = n! \prod_{i=1}^{n} f_X(y_i)$$

  for $y_1 < \ldots < y_n$. Otherwise, $f_{Y_1,Y_2,\ldots,Y_n}(y_1, \ldots, y_n) = 0$.

Example 1

- Uniform distribution: If the parent distribution

$$f_U(u) = \begin{cases} 1, & 0 \leq u \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

then

$$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = \begin{cases} n!, & 0 < y_1 < y_2 < \cdots < y_n < 1 \\ 0, & \text{otherwise} \end{cases}$$

Example 2

- Exponential distribution: If the parent distribution

$$f_X(x) = \begin{cases} e^{-x}, & 0 \leq x < \infty \\ 0, & \text{otherwise} \end{cases}$$

then

$$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = \begin{cases} n! e^{-\sum_{r=1}^{n} y_r}, & 0 < y_1 < y_2 < \cdots < y_n < \infty \\ 0, & \text{otherwise} \end{cases}$$

Marginalization approach (1 of 4)

- The pdf of Y_r is obtained by marginalizing the joint distribution (integrating out everything except y_r), i.e.,

$$f_{Y_r}(y_r) = n! f_X(y_r) \underbrace{\int_{-\infty}^{y_r} \cdots \int_{-\infty}^{y_3} \int_{-\infty}^{y_2} f_X(y_1) f_X(y_2) \ldots f_X(y_{r-1}) dy_1 dy_2 \ldots dy_{r-1}}_{\triangleq L \quad (r-1)-\text{fold integration}}$$

$$\times \underbrace{\int_{y_r}^{\infty} \cdots \int_{y_r}^{y_{r+3}} \int_{y_r}^{y_{r+2}} f_X(y_{r+1}) f_X(y_{r+2}) \ldots f_X(y_n) dy_{r+1} dy_{r+2} \ldots dy_n}_{\triangleq H \quad (n-r)-\text{fold integration}}$$

$$-\infty \leq y_1 \leq y_2 \leq \cdots \leq y_{r-1} \leq y_r \leq y_{r+1} \leq y_{r+2} \leq y_{r+3} \leq \cdots \leq y_n \leq +\infty$$

Marginalization approach (2 of 4)

- The L part

$$L(y_r) = \int_{-\infty}^{y_r} \cdots \underbrace{\int_{-\infty}^{y_3} [F(y_2) - F(\cancelto{0}{-\infty})] \overbrace{f_X(y_2)}^{F'(y_2)}}_{\frac{[F(y_3)]^2}{2}} \cdots f_X(y_{r-1}) \underbrace{dy_2} \cdots dy_{r-1}$$

$$\vdots$$

$$= \frac{[F_X(y_r)]^{r-1}}{(r-1)!}$$

Marginalization approach (3 of 4)

- The H part

$$H(y_r) = \int_{y_r}^{\infty} \cdots \int_{y_{n-2}}^{\infty} \int_{y_{n-1}}^{\infty} f_X(y_n) f_X(y_{n-1}) \cdots f_X(y_{r+1}) dy_n dy_{n-1} \cdots dy_{r+1}$$

$$= \int_{y_r}^{\infty} \cdots \underbrace{\int_{y_{n-2}}^{\infty} [\overset{1}{F(\cancel{+\infty})} - F(y_{n-1})] \overbrace{f_X(y_{n-1})}^{F'(y_{n-1})}}_{-\frac{[1-F(y_{n-1})]^2}{2}\Big|_{y_{n-2}}^{\infty} = \frac{[1-F(y_{n-2})]^2}{2}} \cdots f_X(y_{r+1}) \underbrace{dy_{n-1}}_{} \cdots dy_{r+1}$$

$$\vdots$$

$$= \frac{[1-F(y_r)]^{n-r}}{(n-r)!}$$

Marginalization approach (4 of 4)

- Finally,

$$f_r(x) = \frac{n!}{(r-1)!(n-r)!} \left[F_X(x)\right]^{r-1} \left[1 - F_X(x)\right]^{n-r} f_X(x)$$

the same result as before !

- In particular, for $r = 1$

$$f_{Y_1}(y_1) = n \left[1 - F_X(y_1)\right]^{n-1} f_X(y_1)$$

and, for $r = n$

$$f_{Y_n}(y_n) = n \left[F_X(y_n)\right]^{n-1} f_X(y_n)$$