# 6.434/16.391 Statistics for Engineers and Scientists

Lecture 5 09/19/2012

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

Lecture 5 09/19/2012

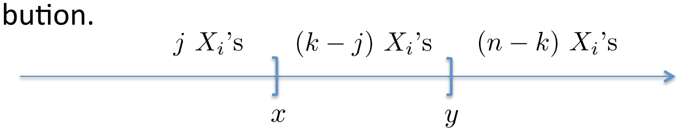
# JOINT DISTRIBUTION OF TWO ORDER STATISTIC

# Tail probability (1 of 3)

- The joint CDF of $X_{(r)}$ and $X_{(s)}$ for $x \le y$ is given by

$$
\begin{aligned}
F_{X_r, X_{(s)}}(x,y) &= \mathbb{P}\{X_{(r)} \le x,\, X_{(s)} \le y\} \\
&= \mathbb{P}\{\text{at least } r \text{ of the } X_i\text{'s} \le x \text{ and at least } s \text{ of the } X_i\text{'s} \le y\} \\
&= \sum_{k=s}^{n}\sum_{j=r}^{k} \mathbb{P}\{\text{exactly } j \text{ of the } X_i\text{'s} \le x \text{ and exactly } k \text{ of the } X_i\text{'s} \le y\} \\
&= \sum_{k=s}^{n}\sum_{j=r}^{k} \frac{n!}{j!(k-j)!(n-k)!}\,[F_X(x)]^j\,[F_X(y)-F_X(x)]^{k-j}\,[1-F_X(y)]^{n-k}
\end{aligned}
$$

which is the tail probability (over the rectangle region consisting of the points $(s,r),\ (s,r+1),\ \ldots,\ (n,n)$ ) of a bivariate binomial distribution.




# Tail probability (2 of 3)

- Using the identity (for $0 < p_1 \le p_2 < 1$ )

$$
\begin{aligned}
&\sum_{k=s}^{n}\sum_{j=r}^{k} \frac{n!}{j!(k-j)!(n-k)!} p_1^j (p_2-p_1)^{k-j}(1-p_2)^{n-k} \\
&= \int_0^{p_1}\int_{t_1}^{p_2} \frac{n!}{(r-1)!(s-r-1)!(n-s)!} t_1^{r-1}(t_2-t_1)^{s-r-1}(1-t_2)^{n-s}\,dt_2\,dt_1
\end{aligned}
$$

we have (for $-\infty < x \le y < \infty$ )

$$
\begin{aligned}
&F_{X_{(r)},X_{(s)}}(x,y) \\
&= \int_0^{F(x)}\int_{t_1}^{F(y)} \frac{n!}{(r-1)!(s-r-1)!(n-s)!} t_1^{r-1}(t_2-t_1)^{s-r-1}(1-t_2)^{n-s}\,dt_2\,dt_1
\end{aligned}
$$

for $-\infty < x \le y < \infty$ , and $F_{X_{(r)},X_{(s)}}(x,y) = F_{X_{(s)}}(y)$ if $y < x$

# Tail probability (3 of 3)

- Finally, the pdf is given by

$$f_{X_{(r)},X_{(s)}}(x,y)$$
$$= \frac{n!}{(r-1)!(s-r-1)!(n-s)!}$$
$$[F_X(x)]^{r-1}\,[F_X(y)-F_X(x)]^{s-r-1}\,[1-F_X(y)]^{n-s}\,f_X(x)f_X(y)$$

for $-\infty < x \le y < \infty$

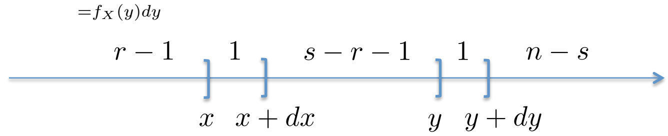
# Differential approach (1 of 2)

- Consider the event

$$\{x \le X_{(r)} \le x+dx,\ y \le X_{(s)} \le y+dy\}$$

- Using similar approach as a single order statistic, we have

$$\mathbb{P}\{x \le X_{(r)} \le x+dx,\ y \le X_{(s)} \le y+dy\}$$
$$= \frac{n!}{(r-1)!(s-r-1)!(n-s)!}$$
$$[F_X(x)]^{r-1}\,\underbrace{[F_X(x+dx)-F_X(x)]}_{=f_X(x)dx}\,[F_X(y)-F_X(x+dx)]^{s-r-1}$$
$$\underbrace{[F_X(y+dy)-F_X(y)]}_{=f_X(y)dy}\,[1-F_X(y+dy)]^{n-s}+O(dx^2dy)+O(dxdy^2)$$



# Differential approach (2 of 2)

- Dividing both sides by $dxdy$ and letting $dx \to 0$ and $dy \to 0$,

$$f_{X_{(r)},X_{(s)}}(x,y)$$
$$= \frac{n!}{(r-1)!(s-r-1)!(n-s)!}$$
$$[F_X(x)]^{r-1}[F_X(y)-F_X(x)]^{s-r-1}[1-F_X(y)]^{n-s}f_X(x)f_X(y)$$

for $-\infty < x \leq y < \infty$

# Marginalization approach (1 of 4)

- The joint pdf of $X_{(r)}$ and $X_{(s)}$ can be obtained by integrating out all other variables except $Y_r$ and $Y_s$

- Recall that the joint distribution is given by

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

# Marginalization approach (2 of 4)

- The joint pdf of $Y_r$ and $Y_s$ is

$$f_{Y_r,Y_s}(y_r, y_s) = n! f_X(y_r) f_X(y_s)$$
$$\times \underbrace{\int_{-\infty}^{y_r} \cdots \int_{-\infty}^{y_3} \int_{-\infty}^{y_2} f_X(y_1) f_X(y_2) \cdots f_X(y_{r-1}) dy_1 dy_2 \cdots dy_{r-1}}_{\triangleq L \quad (r-1)\text{-fold integer}}$$
$$\times \underbrace{\int_{y_r}^{y_s} \cdots \int_{y_r}^{y_{r+3}} \int_{y_r}^{y_{r+2}} f_X(y_{r+1}) f_X(y_{r+2}) \cdots f_X(y_{s-1}) dy_{r+1} dy_{r+2} \cdots dy_{s-1}}_{\triangleq M \quad (s-r-1)\text{-fold integer}}$$
$$\times \underbrace{\int_{y_s}^{\infty} \cdots \int_{y_s}^{y_{s+3}} \int_{y_s}^{y_{s+2}} f_X(y_{s+1}) f_X(y_{s+2}) \cdots f_X(y_n) dy_{s+1} dy_{s+2} \cdots dy_n}_{\triangleq H \quad (n-s)\text{-fold integer}}$$

$$-\infty < y_1 < y_2 < \cdots < y_{r-1} < y_r < y_{r+1} < \cdots < y_{s-1} < y_s < y_{s+1} < \cdots < y_{n-1} < y_n < +\infty$$

# Marginalization approach (3 of 4)

- After some algebra

$$L = \frac{[F_X(y_r)]^{r-1}}{(r-1)!}$$
$$H = \frac{[1 - F_X(y_s)]^{n-s}}{(n-s)!}$$

$$\begin{aligned}
M &= \int_{y_r}^{y_s} \cdots \int_{y_r}^{y_{r+3}} [F_X(y_{r+2}) - F_X(y_r)]\, f_X(y_{r+2}) \cdots f_X(y_{s-1}) dy_{r+2} \cdots dy_{s-1} \\
&= \int_{y_r}^{y_s} \cdots \int_{y_r}^{y_{r+4}} \frac{[F_X(y_{r+3}) - F_X(y_r)]^2}{2} f_X(y_{r+3}) \cdots f_X(y_{s-1}) dy_{r+3} \cdots dy_{s-1} \\
&\vdots \\
&= \frac{[F_X(y) - F_X(x)]^{s-r-1}}{(s-r-1)!}
\end{aligned}$$

## Marginalization approach (4 of 4)

- Finally, we have

$$f_{X_{(r)},X_{(s)}}(x,y)$$
$$= \frac{n!}{(r-1)!(s-r-1)!(n-s)!}$$
$$[F_X(x)]^{r-1}\,[F_X(y)-F_X(x)]^{s-r-1}\,[1-F_X(y)]^{n-s}\,f_X(x)f_X(y)$$

Lecture 5 09/19/2012

## PRINCIPLES OF DATA REDUCTION

## Motivation (1 of 3)

- We clearly would like to separate out any aspects of data
  - that are irrelevant in the context of model (in the context of the parameter of interest), and
  - that may obscure our understanding of the situation
- We begin by formalizing what we mean by "a reduction of data".

## Motivation (2 of 3)

- Definition: Regular model
  - All of $P_\theta$ are continuous with pdf $f(\mathbf{x}|\theta)$
  - All of $P_\theta$ are discrete with pmf $p(\mathbf{x}|\theta)$ and there exists $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ that is independent of $\theta$ such that $\sum_{i=1}^{\infty} p(\mathbf{x}_i|\theta) = 1$

## Motivation (3 of 3)

- Suppose we want to estimate $\theta \in \Theta$ by observing
$$\mathbf{X} = [X_1\ X_2\ \ldots\ X_n]^{\mathrm{T}}$$
- Often $\mathbf{X}$ can be reduced without losing *any* information about $\theta$
- We can simply base our inference about $\theta$ on $T(\mathbf{X})$ which can be considerably simpler than $\mathbf{X}$
  - $T(\mathbf{X})$ is an example for reduction of data

## Example 1

- Consider $X \sim N(\mu, 1)$
- We observe $\mathbf{X} = [X_1\ X_2\ \ldots\ X_n]^{\mathrm{T}}$
- One statistic for estimating $\mu$ is given by
$$T(\mathbf{X}) = \sum_{i=1}^{n} X_i$$
i.e., instead of storing $n$ values $\mathbf{X} = [X_1\ X_2\ \ldots\ X_n]^{\mathrm{T}}$ , we store only the value $T(\mathbf{X})$

# Example 2 (1 of 4)

- A factory produces $n$ blades. Each blade is defected with probability $p$ and is bad with probability $1-p$ . Suppose that there is no dependency between the quality of the items produced. Consider Bernoulli variable

$$X_i = \begin{cases} 1, & \text{if the } i\text{th item is defected} \\ 0, & \text{if the } i\text{th item is good} \end{cases}$$

whose pmf is given by

$$f_{X_i}(x_i|\theta) = \begin{cases} p^{x_i}(1-p)^{1-x_i}, & x_i = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

  – Note: $X_i$ is Bernoulli r.v

# Example 2 (2 of 4)

- When $x_i \in \{0, 1\}$,

$$\begin{aligned} \mathbb{P}\{\mathbf{X} = \mathbf{x}\} &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \\ &= p^{T(\mathbf{x})}(1-p)^{n-T(\mathbf{x})} \end{aligned}$$

where $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is a statistic

- Note that $T(\mathbf{X})$ is binomial ( $n, p$ )

# Example 2 (3 of 4)

- Let $A$ and $B$ be the events that $\mathbf{X} = \mathbf{x}$, and $T(\mathbf{X}) = t$. Then

$$\mathbb{P}\left\{\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t\right\} = \mathbb{P}\left\{A | B\right\} = \frac{P\{A, B\}}{P\{B\}}$$

  – note: $\mathbb{P}\{A, B\} = 0$ if $\sum_{i=1}^{n} x_i \neq t$

$$P\left(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t\right) = \begin{cases} \text{Something}, & \text{if } x_i \in \{0,\, 1\} \text{ and } \sum_{i=1}^n x_i = t \\ 0, & \text{otherwise} \end{cases}$$

# Example 2 (4 of 4)

- Non-zero term: if $x_i \in \{0,\, 1\}$ and $\sum_{i=1}^n x_i = t$,

$$\begin{aligned} \mathbb{P}\left\{\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t\right\} &= \frac{\mathbb{P}\left\{\mathbf{X} = \mathbf{x},\, T(\mathbf{X}) = t\right\}}{\mathbb{P}\left\{T(\mathbf{X}) = t\right\}} \\ &= \frac{p^t(1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}} \end{aligned}$$

- Finally, we obtain the conditional probability

$$P\left(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t\right) = \begin{cases} \frac{1}{\binom{n}{t}}, & \text{if } x_i \in \{0,\, 1\} \text{ and } \sum_{i=1}^n x_i = t \\ 0, & \text{otherwise} \end{cases}$$

- Key observation: the conditional prob. of $\mathbf{X} = \mathbf{x}$, given the statistic $T(\mathbf{X}) = t$, does not depend on $p$.