ORDER STATISTICS

Order statistics

- Definition 1: The ORDER STATISTICS of *independent and identically distributed* (iid) random variables $Y_1, Y_2, \ldots, Y_n$ are:
  - $Y_{(1)}$ is the smallest value in the sample
  - $Y_{(2)}$ is the second smallest value
  - …
  - $Y_{(n)}$ is the largest value in the sample
- And so $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$.

Order statistics

- Some common statistics are functions of the order statistics, e.g.,
  - Median is $Y_{\left(\frac{n+1}{2}\right)}$ if n is odd; $\frac{1}{2}\left(Y_{\left(\frac{n}{2}\right)} + Y_{\left(\frac{n}{2}+1\right)}\right)$ if n is even.
  - Range is $Y_{(n)} - Y_{(1)}$
- These are random variables, so we can speak about their distributions.

Order statistics: Distribution of max

$P(A \cap B) = P(A)P(B)$
IF A & B
indep

- Example 1: If $Y_1, Y_2, \ldots, Y_n$ are iid random variables with cdf F and pdf f, find the pdf for the maximum order statistic $Y_{(n)}$.

cdf $F_{Y_{(n)}}(y) = P(Y_{(n)} \le y) = P(Y_1 \le y, Y_2 \le y, \ldots, Y_n \le y)$

AND AND

indep

$= P(Y_1 \le y) \times P(Y_2 \le y) \times \ldots \times P(Y_n \le y)$

def. of cdf

$= F(y) \times F(y) \times \ldots \times F(y) = [F(y)]^n$ cdf of $Y_{(n)}$

pdf $f_{Y_{(n)}}(y) = \frac{d}{dy}[F(y)]^n = \boxed{n[F(y)]^{n-1} f(y) \quad -\infty < y < \infty}$

chain rule

Order statistics: Distribution of min

- Example 12: If $Y_1, Y_2, \ldots, Y_n$ are iid random variables with cdf F and pdf f, find the pdf for the minimum order statistic $Y_{(1)}$.

cdf $F_{Y_{(1)}}(y) = P(Y_{(1)} \le y) = 1 - P(Y_{(1)} > y)$

$= 1 - P(Y_1 > y, Y_2 > y, \ldots, Y_n > y)$

$= 1 - P(Y_1 > y) \times P(Y_2 > y) \times \ldots \times P(Y_n > y)$

$= 1 - (1 - F(y)) \times (1 - F(y)) \times \ldots \times (1 - F(y))$

$= 1 - (1 - F(y))^n$ (cdf)

pdf $f_{Y_{(1)}}(y) = \frac{d}{dy}\left[1 - (1 - F(y))^n\right] = -n(1 - F(y))^{n-1}(-f(y))$

$$\boxed{= n(1 - F(y))^{n-1} f(y)}$$

$f(y) = \frac{1}{10} e^{-\frac{1}{10}y}$ $F(y) = 1 - e^{-\frac{1}{10}y}$

(Book)

$\boxed{f(y) = \frac{1}{\beta} e^{-\frac{1}{\beta}y}}$

- Example 3: Survival times of rats in an experiment are exponentially distributed with mean 10 days. Find the pdf of the time to the first death among three rats.

$E[Y] = \beta$

$n = 3$

pdf of $Y_{(1)}$

$\beta = 10$

$$n(1 - F(y))^{n-1} f(y)$$

$$3\left(1 - \left(1 - e^{-\frac{1}{10}y}\right)\right)^2 \frac{1}{10} e^{-\frac{1}{10}y}$$

$$= 3\left(e^{-\frac{1}{10}y}\right)^2 \frac{1}{10} e^{-\frac{1}{10}y} = \boxed{\frac{3}{10} e^{-\frac{3}{10}y}}$$

$\beta = \frac{10}{3}$

$Y_{(1)} \sim exp\left(\frac{10}{3}\right)$

$E[Y_{(1)}] = \frac{10}{3}$

$\star$ Assuming Y_1, Y_2, Y_3 independent

- Example 3 revisited: Find the expected time until the first death.

$$Y_{(1)} \sim \exp\left(\frac{10}{3}\right)$$

$$E\left[Y_{(1)}\right] = \frac{10}{3} \text{ days}$$

$k=1$

$k=n$

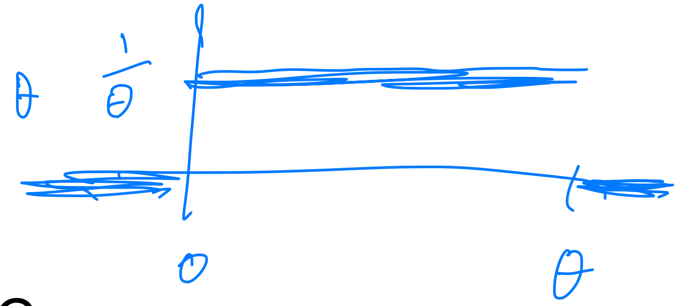
Order statistics: Distribution of general $Y_{(k)}$

- Theorem 1: If $Y_1, Y_2, \dots, Y_n$ are iid random variables with cdf F and pdf f, then the pdf of $Y_{(k)}$ is

$Y_{(n)}$ $Y_{(1)}$

$$f_{(k)}(y_k) = \frac{n!}{(k-1)!(n-k)!}[F(y_k)]^{k-1}[1-F(y_k)]^{n-k}f(y_k), \quad -\infty < y_k < \infty$$

$$f(y) = \begin{cases} \frac{1}{\theta} & 0 \le u \le \theta \\ 0 & \text{o.w.} \end{cases} \qquad F(y) = \begin{cases} 0 & y < 0 \\ 1 - \frac{y}{\theta} & 0 \le y \le \theta \\ 1 & y > \theta \end{cases}$$



Order statistics: Another example

$n = 5$

- Example 4: If U_1, U_2, U_3, U_4, U_5 are independent uniform$(0, \theta)$ random variables, find the distribution of the median.

$k = 3$ $\qquad Y_{(3)}$

$$f_{Y_{(3)}}(y) = \frac{n!}{(k-1)!\,(n-k)!} \left[F(y)\right]^{k-1} \left(1 - F(y)\right)^{n-k} f(y)$$

$$= \frac{5!}{2!\,2!} \left(1 - \frac{y}{\theta}\right)^2 \left(1 - 1 + \frac{y}{\theta}\right)^2 \frac{1}{\theta}$$

$$= \frac{5 \cdot \overset{2}{\cancel{4}} \cdot 3 \cdot \cancel{2!}}{\cancel{2!}\,\cancel{2}} \left(\frac{\theta - y}{\theta}\right)^2 \left(\frac{y}{\theta}\right)^2 \frac{1}{\theta} = \boxed{\frac{30}{\theta^5} y^2 (\theta - y)^2 \quad 0 < y < \theta}$$