STAT 3375Q: Introduction to Mathematical Statistics I

Lecture 20: Order Statistics

Mary Lai Salvaña, Ph.D.

Department of Statistics
University of Connecticut

April 15, 2024

Outline

1. Previously...
   - Functions of Random Variables (Multivariate)
   - The Jacobian Transformation Method
   - The MGF Method
   - Probability Integral Transform
2. Order Statistics
3. The Distribution of the Maximum
4. The Distribution of the Minimum
5. The Joint Distribution of the Minimum and Maximum
6. The Distribution of the kth Order Statistic

Previously...

Functions of Random Variables (Multivariate)

- Suppose we have two random variables X_1 and X_2 with joint PDF $f_{X_1,X_2}(x_1, x_2)$.
- Define new random variables $U_1 = h_1(X_1, X_2)$ and $U_2 = h_2(X_1, X_2)$, where h_1 and h_2 are (one-to-one) monotone functions.
- What is the joint PDF of U_1 and U_2?
  - Jacobian Method (PDF-to-PDF Method or change of variable)
  - MGF Method

The Jacobian Transformation Method

Suppose $U_1 = h_1(X_1, X_2)$ and $U_2 = h_2(X_1, X_2)$ such that has X_1 and X_2 has joint PDF $f_{X_1,X_2}(x_1, x_2)$.

- The joint PDF of U_1 and U_2, $f_{U_1,U_2}(u_1, u_2)$ can be obtained as follows:

  $$f_{U_1,U_2}(u_1, u_2) = f_{X_1,X_2}\{h_1^{-1}(u_1, u_2), h_2^{-1}(u_1, u_2)\}|J|,$$

  where $|J|$ is the absolute value of the Jacobian.
- We need the following to obtain the new PDF:
  - original joint PDF: $f_{X_1,X_2}(x_1, x_2)$
  - transformations: $u_1 = h_1(x_1, x_2)$ and $u_2 = h_2(x_1, x_2)$
  - inverse of the transformation: $x_1 = h_1^{-1}(u_1, u_2)$ and $x_2 = h_2^{-1}(u_1, u_2)$
  - Jacobian: (determinant of the matrix of partial derivatives)

    $$J = \begin{vmatrix} \frac{\partial h_1^{-1}(u_1,u_2)}{\partial u_1} & \frac{\partial h_1^{-1}(u_1,u_2)}{\partial u_2} \\ \frac{\partial h_2^{-1}(u_1,u_2)}{\partial u_1} & \frac{\partial h_2^{-1}(u_1,u_2)}{\partial u_2} \end{vmatrix},$$

    where $|\cdot|$ takes the determinant of the Jacobian matrix.

The MGF Method

- MGF of a sum of independent RVs:
  If $U = Y_1 + Y_2 + \ldots + Y_n$, then

$$m_U(t) = m_{Y_1}(t) m_{Y_2}(t) \cdots m_{Y_n}(t)$$

- Distribution of a sum of independent Gaussian RVs:
  Suppose $Y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$ are independent Gaussian RVs. If $U = \sum_{i=1}^n a_i Y_i$, then $U \sim \mathcal{N}\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$.

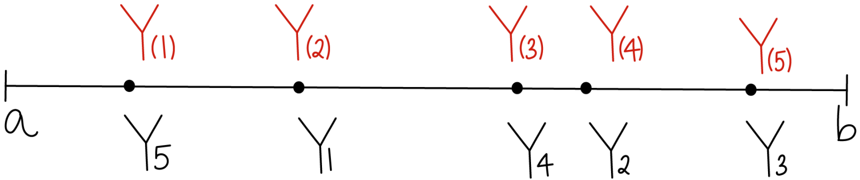
Probability Integral Transform

- Let X have a continuous and strictly increasing CDF $F_X(x)$. Define $U = F_X(X)$. Then, $U \sim \mathcal{U}(0,1)$.
- Let $U \sim \mathcal{U}(0,1)$ and let F be a continuous CDF with quantile function F^{-1}. Let $X = F^{-1}(U)$ Then, X has CDF $F(x)$.

Order Statistics

Order Statistics: Introduction

- Order statistics is concerned with distributions of random variables that follow a certain order.
- Notation for ordered RVs: $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$, where $Y_{(1)} \leq Y_{(2)} \leq \ldots \leq Y_{(n)}$.



- $Y_{(k)}$ is the kth smallest Y, also called the *kth* order statistics.
- We will be more interested in the two extreme RVs:

$$Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n) = Y_{\min}$$
$$Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n) = Y_{\max}$$

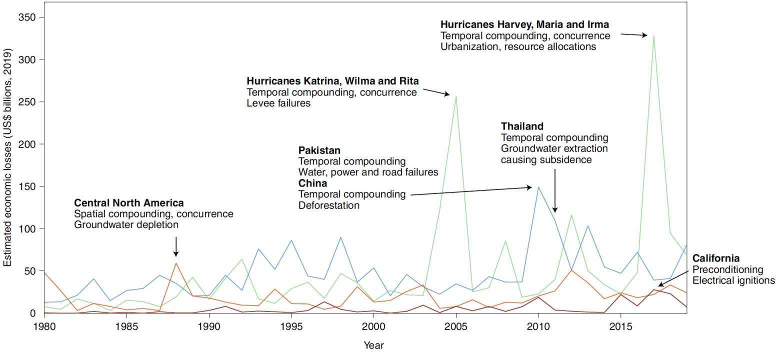
Order Statistics: Motivation

Extreme events are record-shattering...

Extraordinary rainfall that historically would have been extremely rare...

Order Statistics: Motivation

Extreme events are costly...



Economic losses per disaster: tropical cyclones (green), floods (blue), droughts (orange), and wildfires (red). Source: Raymond et al. (2020)

Order Statistics: Motivation

Modeling extreme events is important for disaster risk management...

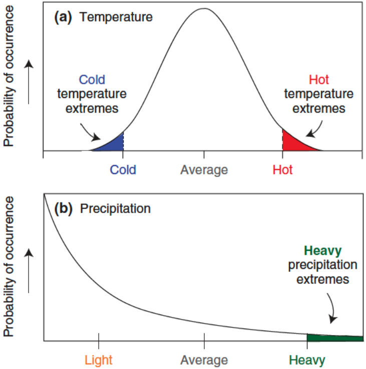
The Oosterschelde barrier / Eastern Scheldt storm surge barrier is the largest of the 13 dams under the Delta Works in Netherlands.

- Delta Works is a system of flood defense structures built as a response to the North Sea flood of 1953.
- The flood defenses were built for a failure of 1 in 10,000 years.

What can we say about the 10,000-year flood based on 100 years of data?

This is when Extreme Statistics is used...

What are extremes?



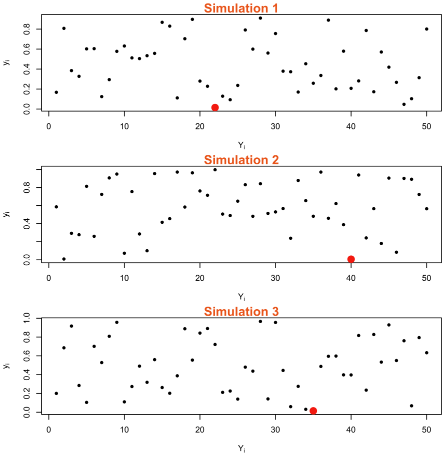
Extremes are those in the shaded region. Source: Zhang et al. (2011)

- Events near the tails, such as extreme rainfall, are given very small probability of occurrence such that the distribution might imply they would never happen.

Order Statistics: Preliminaries

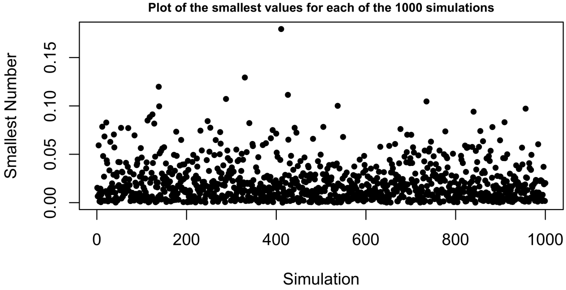
Consider the following experiment:

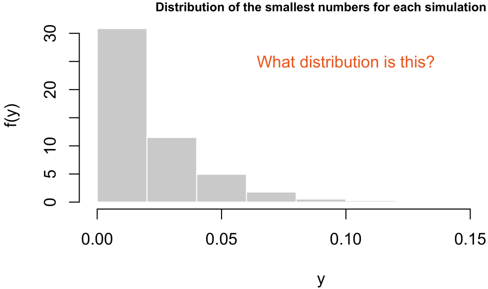
1. Simulate sample values for random variables $Y_1, Y_2, \ldots, Y_{50}$ which are independent and identically distributed uniform RVs on $[0, 1]$.
2. Record or collect the smallest number generated.
3. Repeat 1000x.



The smallest number for each simulation is shown as a red dot.

Order Statistics: Preliminaries

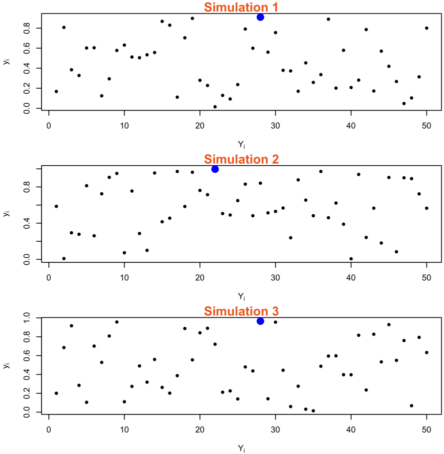




Order Statistics: Preliminaries

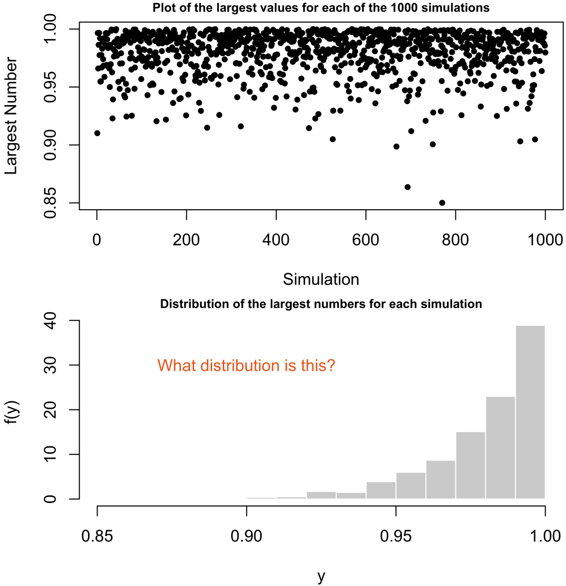
Consider this 2nd experiment:

1. Simulate sample values for random variables $Y_1, Y_2, \ldots, Y_{50}$ which are independent and identically distributed uniform RVs on $[0, 1]$.
2. Record or collect the largest number generated.
3. Repeat 1000x.



The largest number for each simulation is shown as a blue dot.

Order Statistics: Preliminaries



The Distribution of the Maximum

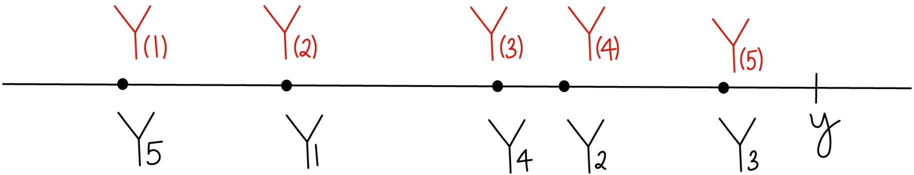
The Distribution of the Maximum $Y_{(n)}$

Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed continuous RVs with PDF $f(y)$ and CDF $F(y)$. Deriving the CDF of the maximum RV $Y_{(n)}$, we have

$$
\begin{aligned}
F_{Y_{(n)}}(y) &= P(Y_{(n)} \leq y) \quad \text{def'n of CDF} \\
&= P(Y_1 \leq y, Y_2 \leq y, \ldots, Y_n \leq y)
\end{aligned}
$$

the prob. that the maximum RV $Y_{(n)}$ will be less than or equal to y is equal to the prob. all the RVs are less than or equal to y.

$$
\begin{aligned}
&= P(Y_1 \leq y)P(Y_2 \leq y)\cdots P(Y_n \leq y) \quad \text{independence} \\
&= \{P(Y_1 \leq y)\}^n \quad \text{identically distributed} \\
&= \{F(y)\}^n. \quad \text{def'n of CDF} \quad \square
\end{aligned}
$$



The Distribution of the Maximum $Y_{(n)}$

Deriving the PDF of the maximum RV $Y_{(n)}$, we have

$$
\begin{aligned}
f_{Y_{(n)}}(y) &= \frac{d}{dy}\{F_{Y_{(n)}}(y)\} && \text{def'n of PDF} \\
&= \frac{d}{dy}\left[\{F(y)\}^n\right] && \text{derived CDF from previous slide} \\
&= n\{F(y)\}^{n-1} f(y). && \text{derivative of CDF } F(y) \text{ is PDF } f(y)\text{; chain rule}
\end{aligned}
$$

□

The Distribution of the Maximum $Y_{(n)}$

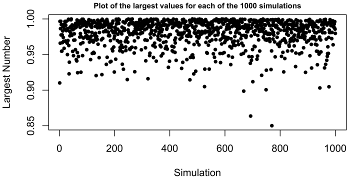
Example 1: (Maximum of Uniforms)

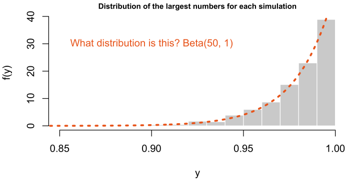
Recall the 2nd experiment...

- Given: $Y_1, Y_2, \ldots, Y_{50}$ which are independent and identically distributed uniform RVs on $[0, 1]$.
- $f(y) = 1$, since $Y_1, Y_2, \ldots, Y_{50} \sim \mathcal{U}(0, 1)$
- $F(y) = y$, $y \in [0, 1]$ CDF of $\mathcal{U}(0, 1)$
- $n = 50$

$$
\begin{aligned}
f_{Y_{(50)}}(y) &= n\{F(y)\}^{n-1} f(y) \quad \text{formula} \\
&= 50y^{50-1}(1) \\
&= 50y^{49}, \quad y \in [0, 1].
\end{aligned}
$$

This is a Beta$(50, 1)$ distribution.





Recall PDF of Beta: $f(y) = \begin{cases} \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1}, & 0 \le y \le 1, \\ 0, & \text{elsewhere}, \end{cases}$ where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

The Distribution of the Maximum $Y_{(n)}$

Example 2: (Maximum of Exponentials)

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\beta)$. Find the PDF of $X_{(n)}$.

Solution:

- CDF of exponential: $F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x/\beta}, & 0 \leq x < \infty \end{cases}$
- PDF of exponential: $f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & 0 \leq x < \infty, \\ 0, & \text{elsewhere} \end{cases}$
- Use the formula of PDF of Maximum:

$$\begin{aligned} f_{X_{(n)}}(x) &= n\{F(x)\}^{n-1} f(x) \quad \text{formula} \\ &= n(1 - e^{-x/\beta})^{n-1} \frac{1}{\beta} e^{-x/\beta}. \end{aligned}$$

□

The Distribution of the Minimum

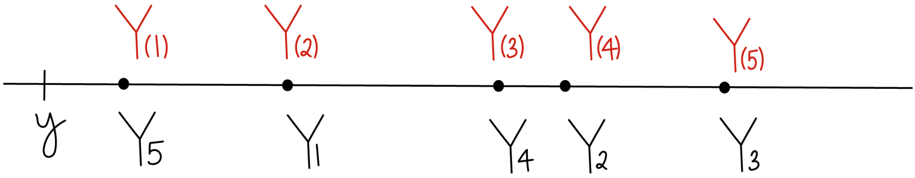
The Distribution of the Minimum $Y_{(1)}$

Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed continuous RVs with PDF $f(y)$ and CDF $F(y)$. Deriving the CDF of the minimum RV $Y_{(1)}$, we have

$$
\begin{aligned}
F_{Y_{(1)}}(y) &= P(Y_{(1)} \leq y) && \text{def'n of CDF} \\
&= 1 - P(Y_{(1)} > y) && \text{complement} \\
&= 1 - P(Y_1 > y, Y_2 > y, \ldots, Y_n > y)
\end{aligned}
$$

the prob. that the minimum RV $Y_{(n)}$ will be greater than y is equal to the prob. all the RVs are greater than y.

$$
\begin{aligned}
&= 1 - P(Y_1 > y)P(Y_2 > y)\cdots P(Y_n > y) && \text{independence} \\
&= 1 - \{P(Y_1 > y)\}^n && \text{identically distributed} \\
&= 1 - \{1 - P(Y_1 \leq y)\}^n && \text{complement} \\
&= 1 - \{1 - F(y)\}^n. && \text{def'n of CDF} \quad \square
\end{aligned}
$$



The Distribution of the Minimum $Y_{(1)}$

Deriving the PDF of the minimum RV $Y_{(1)}$, we have

$$
\begin{aligned}
f_{Y_{(1)}}(y) &= \frac{d}{dy}\{F_{Y_{(1)}}(y)\} && \text{def'n of PDF} \\
&= \frac{d}{dy}\left[1 - \{1 - F(y)\}^n\right] && \text{derived CDF from previous slide} \\
&= n\{1 - F(y)\}^{n-1} f(y). && \text{derivative of CDF } F(y) \text{ is PDF } f(y)\text{; chain rule}
\end{aligned}
$$

□

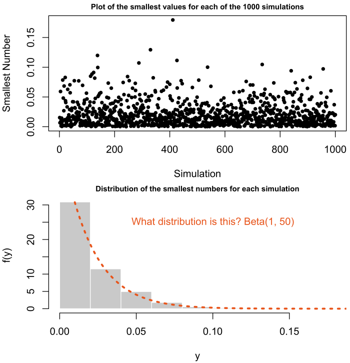
Example 3: (Minimum of Uniforms)

Recall the 1st experiment...

- Given: $Y_1, Y_2, \ldots, Y_{50}$ which are independent and identically distributed uniform RVs on $[0, 1]$.
- $f(y) = 1$, since $Y_1, Y_2, \ldots, Y_{50} \sim \mathcal{U}(0, 1)$
- $F(y) = y$, $y \in [0, 1]$ CDF of $\mathcal{U}(0, 1)$
- $n = 50$

$$\begin{aligned} f_{Y_{(1)}}(y) &= n\{1 - F(y)\}^{n-1} f(y) \quad \text{formula} \\ &= 50(1-y)^{50-1}(1) \\ &= 50(1-y)^{49}, \quad y \in [0,1]. \end{aligned}$$

This is a Beta(1, 50) distribution.



Recall PDF of Beta: $f(y) = \begin{cases} \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1}, & 0 \le y \le 1, \\ 0, & \text{elsewhere}, \end{cases}$ where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

The Distribution of the Minimum $Y_{(1)}$

Example 4: (Minimum of Exponentials)

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\beta)$. Find the PDF of $X_{(1)}$.

Solution:

- CDF of exponential: $F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x/\beta}, & 0 \leq x < \infty \end{cases}$
- PDF of exponential: $f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & 0 \leq x < \infty, \\ 0, & \text{elsewhere} \end{cases}$
- Use the formula of PDF of Minimum:

$$\begin{aligned} f_{X_{(1)}}(y) &= n\{1 - F(x)\}^{n-1} f(x) \quad \text{formula} \\ &= n\{1 - (1 - e^{-x/\beta})\}^{n-1} \frac{1}{\beta} e^{-x/\beta} \\ &= \frac{n}{\beta} e^{-x/\beta} (e^{-x/\beta})^{n-1} \\ &= \frac{n}{\beta} e^{-nx/\beta}. \end{aligned}$$

□

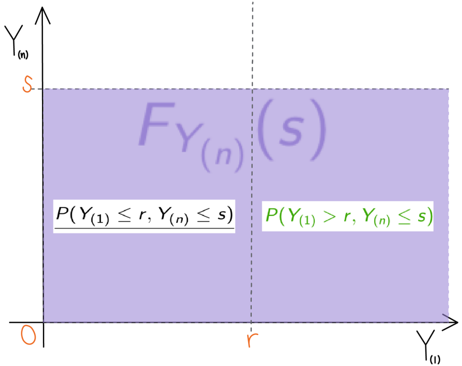
The Joint Distribution of the Minimum and Maximum

The Joint Distribution of the Minimum and Maximum

Goal: Derive the joint CDF of the minimum $Y_{(1)}$ and maximum $Y_{(n)}$:

$$F_{Y_{(1)},Y_{(n)}}(r,s) = \underline{P(Y_{(1)} \leq r, Y_{(n)} \leq s)}.$$

How do we solve for $\underline{P(Y_{(1)} \leq r, Y_{(n)} \leq s)}$?



$$\Rightarrow \underline{P(Y_{(1)} \leq r, Y_{(n)} \leq s)} \quad = \quad F_{Y_{(n)}}(s) - P(Y_{(1)} > r, Y_{(n)} \leq s).$$

(cont'd next slide...)

The Joint Distribution of the Minimum and Maximum

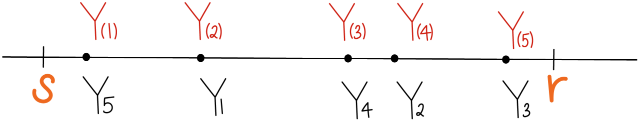
$$\underline{P(Y_{(1)} \leq r, Y_{(n)} \leq s)} = F_{Y_{(n)}}(s) - P(Y_{(1)} > r, Y_{(n)} \leq s)$$

- $F_{Y_{(n)}}(s)$: we know the formula for CDF of maximum:

$$F_{Y_{(n)}}(s) = \{F(s)\}^n$$

Slide 20

- $P(Y_{(1)} > r, Y_{(n)} \leq s)$:
  - If $r \geq s$:



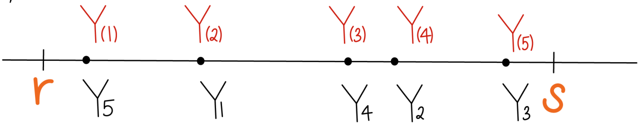
$$P(Y_{(1)} > r, Y_{(n)} \leq s) = 0.$$

This is the uninteresting case since $\underline{P(Y_{(1)} \leq r, Y_{(n)} \leq s)} = F_{Y_{(n)}}(s)$.

(cont'd next slide...)

The Joint Distribution of the Minimum and Maximum

- $P(Y_{(1)} > r, Y_{(n)} \leq s)$:
  - If $r < s$,



$$
\begin{aligned}
P(Y_{(1)} > r, Y_{(n)} \leq s) &= P(r < Y_1 \leq s, r < Y_2 \leq s, \ldots, r < Y_n \leq s) \\
&= P(r < Y_1 \leq s)P(r < Y_2 \leq s) \cdots P(r < Y_n \leq s)
\end{aligned}
$$

independent RVs

$$= \{P(r < Y_1 \leq s)\}^n$$ identical RVs

$$= \{F(s) - F(r)\}^n.$$ def'n of CDF

(cont'd next slide...)

The Joint Distribution of the Minimum and Maximum

- Putting everything together, the joint CDF of $Y_{(1)}$ and $Y_{(n)}$ is:

$$
\begin{aligned}
F_{Y_{(1)},Y_{(n)}}(r,s) &= \underline{P(Y_{(1)} \le r, Y_{(n)} \le s)} \\
&= F_{Y_{(n)}}(s) - P(Y_{(1)} > r, Y_{(n)} \le s) \\
&= \{F(s)\}^n - \{F(s) - F(r)\}^n. \quad \text{from Slides 31-32}
\end{aligned}
$$

- Consequently, the joint PDF of $Y_{(1)}$ and $Y_{(n)}$ is:

$$
\begin{aligned}
f_{Y_{(1)},Y_{(n)}}(r,s) &= \frac{d}{dr}\frac{d}{ds}\left[\{F(s)\}^n - \{F(s) - F(r)\}^n\right] \\
&= \frac{d}{dr}\left[n\{F(s)\}^{n-1}f(s) - n\{F(s) - F(r)\}^{n-1}f(s)\right] \\
&= -n(n-1)\{F(s) - F(r)\}^{n-2}f(s)\{-f(r)\} \\
&= n(n-1)\{F(s) - F(r)\}^{n-2}f(s)f(r), \quad r < s. \quad \square
\end{aligned}
$$

The Joint Distribution of the Minimum and Maximum

Example 5:

Let $X_1, X_2, \ldots, X_{15} \overset{iid}{\sim} \mathcal{U}(0,1)$. Find the joint PDF of $X_{(1)}$ and $X_{(15)}$.

Solution:

- CDF of $\mathcal{U}(0,1)$: $F(x) = x, \quad 0 \leq x \leq 1$
- PDF of $\mathcal{U}(0,1)$: $f(x) = 1, \quad 0 \leq x \leq 1$
- $n = 15$
- Use the formula of joint PDF of Minimum and Maximum:

$$\begin{aligned} f_{X_{(1)},X_{(15)}}(r,s) &= n(n-1)\{F(s)-F(r)\}^{n-2} f(s) f(r) \quad \text{formula} \\ &= 15(15-1)(s-r)^{15-2}(1)(1) \\ &= 15(14)(s-r)^{13}, \quad 0 \leq r \leq 1, \quad 0 \leq s \leq 1. \end{aligned}$$

□

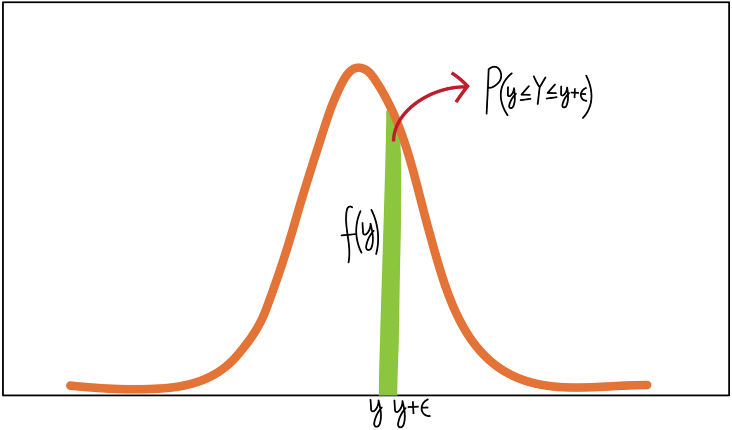
The Distribution of the kth Order Statistic

The Distribution of the kth Order Statistic: Preliminaries

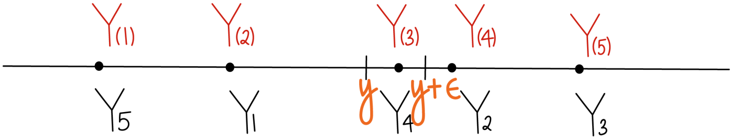
Let Y be a continuous RV.

$$f(y) = \lim_{\epsilon \to 0} \frac{P(Y \in [y, y+\epsilon])}{\epsilon}$$

$$P(Y \in [y, y+\epsilon]) \approx f(y)\epsilon$$ probability = area under the curve



The Distribution of the kth Order Statistic



Let $Y_1, Y_2, \ldots, Y_n$ be iid continuous RVs with PDF $f(y)$ and $F(y)$.

$$
\begin{aligned}
P(y \leq Y_{(k)} \leq y+\epsilon) &= P(\text{one of the Y's} \in [y, y+\epsilon] \text{ and exactly } k-1 \text{ of them} < y) \\
&= \sum_{i=1}^{n} P(Y_i \in [y, y+\epsilon] \text{ and exactly } k-1 \text{ of them are} < y) \\
&= nP(Y_1 \in [y, y+\epsilon] \text{ and exactly } k-1 \text{ of them are} < y)
\end{aligned}
$$

Y_i's have identical distributions

$$
= nP(Y_1 \in [y, y+\epsilon])P(\text{exactly } k-1 \text{ of them are} < y)
$$

Y_i's are independent

$$
= nP(Y_1 \in [y, y+\epsilon])\left\{\binom{n-1}{k-1} P(Y_1 < y)^{k-1} P(Y_1 \geq y)^{n-k}\right\}
$$

binomial probability with $n-1$ trials and prob. of success $P(Y_1 < y)$

$$
= nP(Y_1 \in [y, y+\epsilon])\binom{n-1}{k-1} F(y)^{k-1}(1-F(y))^{n-k}
$$

(cont'd next slide...)

The Distribution of the kth Order Statistic

$$
\begin{aligned}
P(y \leq Y_{(k)} \leq y + \epsilon) &= nP(Y_1 \in [y, y + \epsilon]) \binom{n-1}{k-1} F(y)^{k-1}(1 - F(y))^{n-k} \\
f_{Y_{(k)}}(y)\epsilon &= nf(y)\epsilon \binom{n-1}{k-1} F(y)^{k-1}(1 - F(y))^{n-k} \quad \text{def'n of PDF Slide 36} \\
f_{Y_{(k)}}(y) &= nf(y) \binom{n-1}{k-1} F(y)^{k-1}(1 - F(y))^{n-k}. \quad \text{cancel } \epsilon \text{ on both sides} \quad \square
\end{aligned}
$$

The Distribution of the kth Order Statistic

Example 6:

Let $X_1, X_2, \ldots, X_{15} \overset{iid}{\sim} \mathcal{U}(0,1)$. Find the PDF of $X_{(8)}$.

Solution:

- CDF of $\mathcal{U}(0,1)$: $F(x) = x, \quad 0 \leq x \leq 1$
- PDF of $\mathcal{U}(0,1)$: $f(x) = 1, \quad 0 \leq x \leq 1$
- $n = 15$
- $k = 8$
- Use the formula for the PDF of the kth order statistic:

$$\begin{aligned} f_{X_{(8)}}(x) &= nf(x)\binom{n-1}{k-1}F(x)^{k-1}(1-F(x))^{n-k} \quad \text{formula} \\ &= 15(1)\binom{15-1}{8-1}x^{8-1}(1-x)^{15-8} \\ &= 15\binom{14}{7}x^7(1-x)^7 = 15\frac{14!}{(14-7)!7!}x^7(1-x)^7 \\ &= \frac{15!}{7!7!}x^7(1-x)^7, \quad 0 \leq x \leq 1. \end{aligned}$$

□

Questions?

Homework Exercises: 6.72, 6.73, 6.74, 6.75, 6.80

Solutions will be discussed this Friday by the TA.