STAT 512 su 2021 Lec 03 slides

Random samples, statistics, esp. order statistics

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Random sample

A collection of independent rvs with the same distribution is a *random sample*.

- Often denote by $X_1, \dots, X_n$, where n is the *sample size*.
- In random sample, $X_1, \dots, X_n$ are *iid*: independent and identically distributed.
- Common distribution of $X_1, \dots, X_n$ called the *population distribution*.

Goal is to learn from $X_1, \dots, X_n$ about the population distribution.

Sample statistic

A *statistic* is any function of the rvs in the random sample.

Setup: Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} F_X$, where F_X has support on $\mathcal{X}$, and let T_n be the rv

$$T_n := T(X_1, \dots, X_n),$$

where $T : \mathcal{X}^n \to \mathcal{Y}$, for some set $\mathcal{Y}$.

Examples. . .

Sampling distribution

The distribution of a statistic is called the *sampling distribution* of the statistic.

- We want to learn about population parameters from sample statistics.
- What we can learn from a statistic depends on its sampling distribution.

Our favorite statistics

Given a rs $X_1, \ldots, X_n$, the *sample mean* and *sample variance* are the statistics

$$\bar{X}_n = (X_1 + \cdots + X_n)/n \quad \text{and} \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

Theorem (Moment results for favorite statistics)

Let $X_1, \ldots, X_n$ *be a rs from a dist. with mean* μ *and variance* $\sigma^2 < \infty$. *Then*

- $\mathbb{E}\bar{X}_n = \mu$
- $\mathrm{Var}\,\bar{X}_n = \sigma^2/n$
- $\mathbb{E}S_n^2 = \sigma^2$

Exercise: Prove the above.

Exercises:

1. Let $X_1, \dots, X_n$ be ind. Gamma(α, β) rvs. Find the sampling dist. of $\bar{X}_n$.
2. Let $X_1, \dots, X_n$ be ind. Normal(μ, σ^2) rvs. Find the sampling dist. of $\bar{X}_n$.

Order statistics

Given a random sample $X_1, \ldots, X_n$, define

$$\begin{aligned} X_{(1)} &= \text{ the least of } X_1, \ldots, X_n \\ X_{(2)} &= \text{ the next-to-least of } X_1, \ldots, X_n \\ &\vdots \\ X_{(n)} &= \text{ the greatest of } X_1, \ldots, X_n. \end{aligned}$$

Then $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ are called the *order statistics* of the rs.

Exercise: Define range, midrange, and median with order statistics.

Theorem (pdf of kth order statistic)

Let $X_{(1)}, \ldots, X_{(n)}$ be the order statistics of a rs with cdf F_X and pdf f_X.

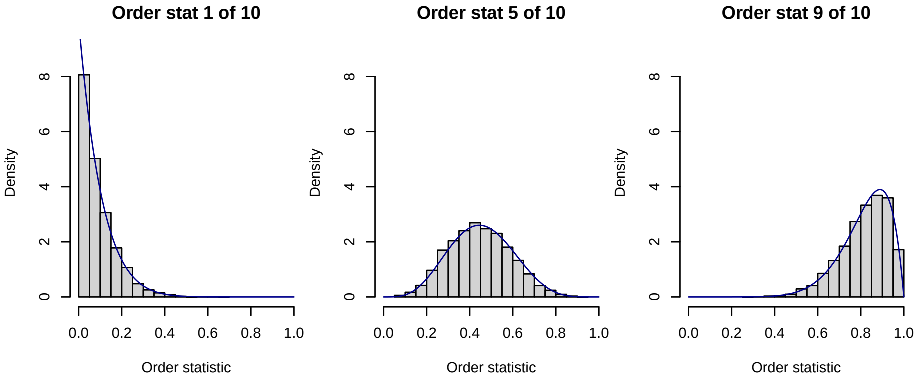
Then the pdf of $X_{(k)}$ is given by

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

for $k = 1, \ldots, n$.

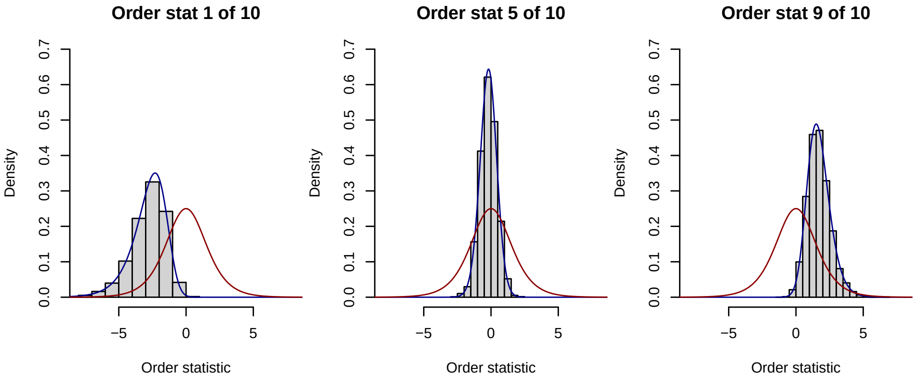
Exercises:

1. Derive the above.
2. Let $U_1, \ldots, U_n \stackrel{\text{ind}}{\sim} \text{Uniform}(0, 1)$. Find pdf of $U_{(k)}$, find $\mathbb{E}U_{(k)}$ and $\text{Var}\, U_{(k)}$.
3. Draw samples $X_1, \ldots, X_n \stackrel{\text{ind}}{\sim} \text{Uniform}(0, 1)$ and record kth order statistic. Make histograms. Use $n = 10$ and consider $k = 1, 5, 9$.


Order stat 1 of 10
Order stat 5 of 10
Order stat 9 of 10
Density
Order statistic
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0

Exercise: Let $X_1, \ldots, X_n$ be independent rvs with cdf $F_X(x) = (1 + e^{-x})^{-1}$.

1. Find the pdf of the kth order statistic.
2. Draw samples $X_1, \ldots, X_n$ from F_X and record kth order statistic. Make histograms. Use $n = 10$ and consider $k = 1, 5, 9$.



Order stat 1 of 10
Order stat 5 of 10
Order stat 9 of 10
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
−5
0
5
Order statistic

Corollary (pdf of minimum and maximum)

Let $X_{(1)}, \dots, X_{(n)}$ *be the order statistics of a rs with cdf* F_X *and pdf* f_X. *Then*

- $X_{(1)}$ *has cdf and pdf given by*

$$F_{X_{(1)}}(x) = 1 - [1 - F_X(x)]^n$$
$$f_{X_{(1)}}(x) = n[1 - F_X(x)]^{n-1} f_X(x)$$

- $X_{(n)}$ *has cdf and pdf given by*

$$F_{X_{(n)}}(x) = [F_X(x)]^n$$
$$f_{X_{(n)}}(x) = n[F_X(x)]^{n-1} f_X(x)$$

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Exponential}(\lambda)$.

1. Find the pdf of $X_{(n)}$.
2. Find the pdf of $X_{(1)}$ and identify the distribution.

Theorem (Joint pdf of two order statistics)

Let $X_{(1)}, \ldots, X_{(n)}$ *be the order statistics of a rs with cdf* F_X *and pdf* f_X.

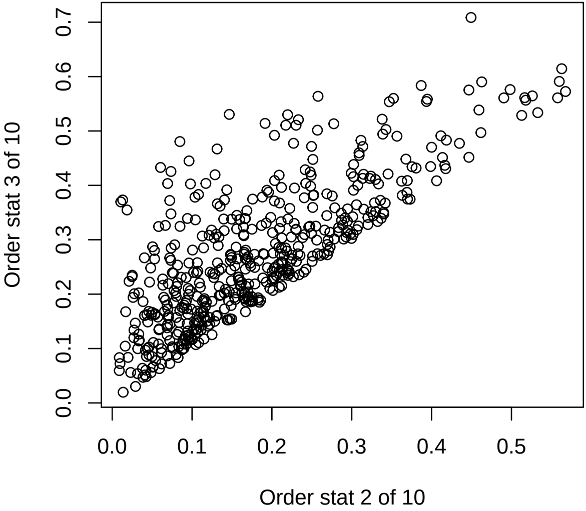
Then the joint pdf of $X_{(j)}$ *and* $X_{(k)}$, $1 \leq j < k \leq n$, *is given by*

$$\begin{aligned} f_{X_{(j)},X_{(k)}}(u, v) = & \frac{n!}{(j-1)!(k-j-1)!(n-k)!} f_X(u) f_X(v) \\ & \times [F_X(u)]^{j-1}[F_X(v) - F_X(u)]^{k-j-1}[1 - F_X(v)]^{n-k} \end{aligned}$$

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

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Uniform}(0,1)$.

1. Find the joint pdf of the order statistics $U = X_{(k)}$ and $V = X_{(k+1)}$.
2. Draw samples $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Uniform}(0,1)$ and record $(X_{(k)}, X_{(k+1)})$. Make a scatterplot of the values. Use $n = 10$, $k = 2$.


Order stat 3 of 10
Order stat 2 of 10
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

Corollary (Joint pdf of min and max)

Let $X_{(1)}, \ldots, X_{(n)}$ be the order statistics of a rs with cdf F_X and pdf f_X.

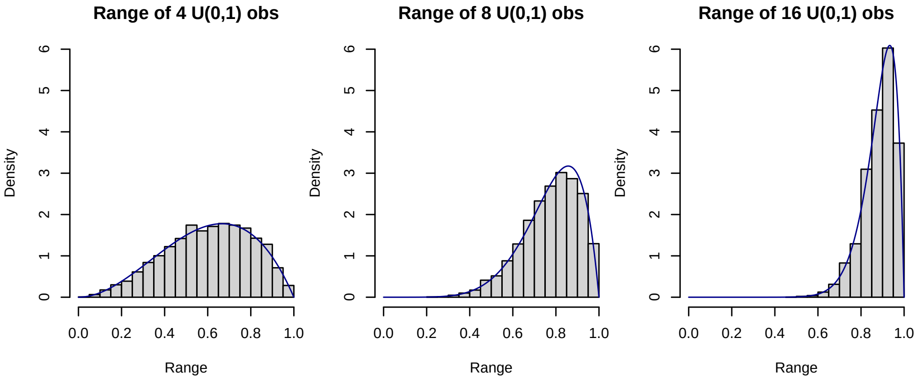
The joint pdf of $X_{(1)}$ and $X_{(n)}$ is given by

$$f_{X_{(1)},X_{(n)}}(u,v) = n(n-1)f_X(u)f_X(v)[F_X(v) - F_X(u)]^{n-2}$$

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

Exercise: Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Uniform}(0, \theta)$.

1. Find the joint pdf of $X_{(1)}$ and $X_{(n)}$.
2. Find the joint pdf of $R = X_{(n)} - X_{(1)}$ and $M = X_{(n)}$.
3. Find the marginal pdf of R.
4. Draw realizations of the range of $\text{Uniform}(0, 1)$ samples with $n = 4, 8, 16$. Make histograms and overlay densities.


Range of 4 U(0,1) obs
Range of 8 U(0,1) obs
Range of 16 U(0,1) obs
Density
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
Range