M378K: March 3rd, 2025.

Statistical Set-Up.

Population. e.g., all the people in this class

Sample. e.g., a committee of 4 students

We assume it's a representative sample.

Use the same word for the results of measuring/polling from our population w/ an unknown but common dist'n.

Def'n. A random sample of size n from a distribution D is a random vector

$$(Y_1, Y_2, \dots, Y_n)$$

such that:

1. $Y_1, Y_2, \dots, Y_n$ are independent
2. every Y_i has the distribution D.

Example. Consider 10 measurements $Y_1, \dots, Y_{10}$. Care was taken so that the measurements are independent.

It's a standard model to assume that

Y_i are normally distributed w/ an unknown mean μ.

Scenario #1. We know the standard deviation 0.1.

Then, $Y_i \sim N(\mu, \sigma = 0.1)$, $i = 1..10$

Scenario #2. We don't know the standard deviation σ

Then, $Y_i \sim N(\mu, \sigma)$, $i = 1..10$

Def'n. Any function of the random sample is called a STATISTIC.

A POINT ESTIMATOR is any function (rule, procedure) of the sample $(Y_1, \dots, Y_n)$ including only known constants (w/ the purpose of estimating a model parameter).

$\boxed{\text{IT MUSTN'T CONTAIN THE UNKNOWN PARAMETER WE'RE TRYING TO ESTIMATE.}}$

An interval estimator is a pair of point estimator.

e.g., • $\dfrac{\text{\# of people in the "committee" who like ice cream}}{4}$, i.e.,

a sample proportion

• in the normal example, we look @ the sample mean

$$\boxed{\bar{Y} = \frac{Y_1 + Y_2 + \dots + Y_n}{n}}$$

M378K Introduction to Mathematical Statistics
Problem Set #13
Order Statistics.

Problem 13.1. *An insurance company is handling claims from two categories of drivers: the good drivers and the bad drivers. The waiting time for the first claim from a* **good** *driver is modeled by an exponential random variable T_g with mean 6 (in years). The waiting time for the first claim from a* **bad** *driver is modeled by an exponential random variable T_b with mean 3 (in years). We assume that the random variables T_g and T_b are independent.*

What is the distribution of the waiting time T until the first claim occurs (regardless of the type of driver this claim was filed by)?

→: $T = \min(T_g, T_b)$ $\quad S_T = [0, \infty)$

$t > 0$: $F_T(t) = \mathbb{P}[T \leq t] = \mathbb{P}[\min(T_g, T_b) \leq t] = 1 - \mathbb{P}[\min(T_g, T_b) > t]$

$= 1 - \mathbb{P}[T_g > t, T_b > t] = 1 - \mathbb{P}[T_g > t] \cdot \mathbb{P}[T_b > t] = 1 - e^{-t/\tau_g} \cdot e^{-t/\tau_b}$

$= 1 - e^{-t\left(\frac{1}{\tau_g} + \frac{1}{\tau_b}\right)} \Rightarrow T \sim E(\tau)$ w/ $\tau = \dfrac{1}{\frac{1}{\tau_g} + \frac{1}{\tau_b}} = \dfrac{1}{\frac{1}{6} + \frac{1}{3}} = \dfrac{1}{1/2} = 2$ $\square$

Definition 13.1. *Let $Y_1, \ldots, Y_n$ be a* **random sample**. *The random sample ordered in an increasing order is called an* order statistic *and denoted by*

$$Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}.$$

Question Write $Y_{(1)}$ as a function of $Y_1, Y_2, \ldots, Y_n$.

Question Write $Y_{(n)}$ as a function of $Y_1, Y_2, \ldots, Y_n$.