ISyE 6739 – Statistical Methods
Sampling Distributions

Instructor: Kamran Paynabar
H. Milton Stewart School of
Industrial and Systems Engineering
Georgia Tech

Kamran.paynabar@isye.gatech.edu
Office: Groseclose 436

Review

Standard Normal

Chi-square

$X_i \sim N(\mu, \sigma^2)$ → $Z_i = \dfrac{X_i - \mu}{\sigma}$ → $V = Z_1^2 + Z_1^2 + \ldots + Z_v^2$

Student's *t*

$$\frac{Z}{\sqrt{\frac{V}{v}}}$$

$Z_i \sim N(0,1)$, $W_i \sim N(0,1)$ → $U_1 = Z_1^2 + Z_2^2 + \ldots + Z_{d_1}^2 \sim$, $U_2 = W_1^2 + W_2^2 + \ldots + W_{d_2}^2$, → $\dfrac{U_1/d_1}{U_2/d_2} \sim$

***F*-distribution**

List of Topics

- Random Sample
- Central Limit Theorem (CLT)
- Sampling Distributions
  - Sample mean of one distribution
  - Difference of sample means of two distributions
  - Sample variance of one distribution
  - Ratio of sample variances of two distributions

Sampling Distributions

Statistical inference is concerned with making decisions about a population based on the information contained in a *random sample* from that population.

The random variables $X_1, X_2, \ldots, X_n$ are a **random sample** of size n if (a) the X_i's are independent random variables, and (b) every X_i has the same probability distribution.

Observations in a random sample are also known as independent and identically distributed ***(i.i.d.)*** random variables

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

$$\text{e.g., } X_1, X_2, \cdots, X_n \rightarrow \overline{X}, S^2$$

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

Central Limit Theorem (CLT)

If we are sampling from a population that has an unknown probability distribution, the sampling distribution of the sample mean will still be approximately normal with mean μ and variance σ^2/n, if the sample size n is large. This is one of the most useful theorems in statistics, called the **central limit theorem.** The statement as follows:

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2, and if $\overline{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{7-1}$$

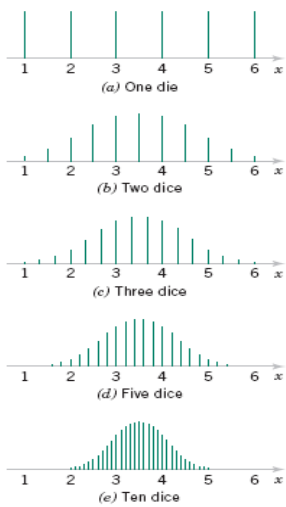
as $n \rightarrow \infty$, is the standard normal distribution.

Central Limit Theorem (CLT)

Figure 7-1 Distributions of average scores from throwing dice. [Adapted with permission from Box, Hunter, and Hunter (1978).]

Demo:

http://onlinestatbook.com/stat_sim/sampling_dist/index.html



CLT (Example)

An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is **Non-normal** Find the probability that a random sample of $n = 25$ resistors will have an average resistance less than 95 ohms.

Note that the sampling distribution of $\overline{X}$ is normal, with mean $\mu_{\overline{X}} = 100$ ohms and a standard deviation of

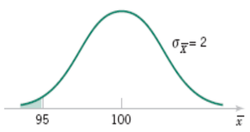
$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

Therefore, the desired probability corresponds to the shaded area in Fig. 7-1. Standardizing the point $\overline{X} = 95$ in Fig. 7-2, we find that

$$z = \frac{95 - 100}{2} = -2.5$$

and therefore,

$$P(\overline{X} < 95) = P(Z < -2.5) = 0.0062$$



Sampling Distributions

Sampling Distribution: The probability distribution of a statistic based on a random sample

Sampling distributions for sample mean and variance:

- Normal
- Chi-square distribution
- t-distribution
- F-distribution

Key issues:

- When should each distribution be used?
- How to use each distribution?

Main Assumption: samples are collected from a population following Normal distribution

Sampling Distribution of Sample Mean With Known Variance

One Population:

Xi's are normally independently distributed (a random sample from a Normal distribution with the known variance)

$$X_1, X_2, \cdots, X_n \sim NID(\mu, \sigma^2) \rightarrow \overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \rightarrow \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Two Populations:

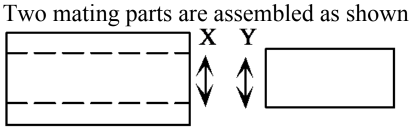
Two independent random samples from two Normal distributions with the known variances

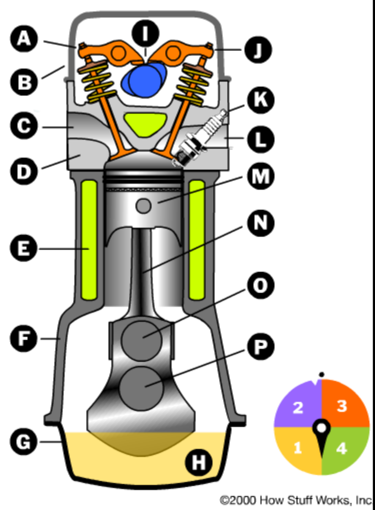
$$\left.\begin{array}{l} X_1, X_2, \cdots, X_{n_1} \sim NID(\mu_1, \sigma_1^2) \\ \\ Y_1, Y_2, \cdots, Y_{n_2} \sim NID(\mu_2, \sigma_2^2) \end{array}\right\} \rightarrow \overline{X} - \overline{Y} \sim NID(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

$$\rightarrow \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim NID(0,1)$$

Example

Two mating parts are assembled as shown below:





Assume $\mu_X = 2.010$ cm., $\sigma^2_X = .002$ cm.
$\mu_Y = 2.004$ cm., $\sigma^2_Y = .001$ cm.

X and Y are independently normally distributed

If $X_1, X_2, \cdots, X_5$ and $Y_1, Y_2, \cdots, Y_{10}$ are two independent random samples from these distributions, calculate the following probability:

$\Pr(\overline{X} - \overline{Y} > 0)$

Sampling Distribution of Sample Variance

One Population:

Xi's are normally independently distributed (a random sample from a Normal distribution)

Chi -Square

$$X_1, X_2, \cdots, X_n \sim NID(\mu, \sigma^2) \rightarrow \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Example

Patient's Waiting Time:

Suppose the patient's waiting time in an ER is normally distributed with σ = 0.5 min. If the waiting time of 3 randomly selected patients are recorded, calculate the probability that the sample standard deviation is greater than 0.5.

Sampling Distribution of Sample Mean With Unknown Variance

One Population:

Xi's are normally independently distributed (a random sample from a Normal distribution with an unknown variance)

$$X_1, X_2, \cdots, X_n \sim NID(\mu, \sigma^2) \rightarrow \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

Example

PVC pipe diameter:

The diameter of the manufactured PVC pipe follows a normal distribution with mean 1.00 and variance σ^2. Find limit *U* such that the probability that the mean diameter of a random sample of *n*=9 is greater than *U* is 0.05. The sample standard deviation is 0.1

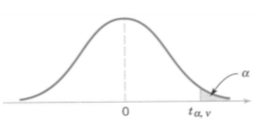


Table V Percentage Points $t_{\alpha,\nu}$ of the t Distribution

ν \ α	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587

Sampling Distribution of Sample Mean With Unknown Variance

Two Populations:

Two independent random samples from two Normal distributions with unknown (but equal) variances $\sigma_1^2 = \sigma_2^2 = ?$

$$\left.\begin{array}{l} X_1, X_2, \cdots, X_{n_1} \sim NID(\mu_1, \sigma_1^2) \\ \\ Y_1, Y_2, \cdots, Y_{n_2} \sim NID(\mu_2, \sigma_2^2) \end{array}\right\} \rightarrow \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}$$

Sampling Distribution of Sample Variance

Two Populations:

Two independent random samples from two Normal distributions

$$\left.\begin{array}{l} X_1, X_2, \cdots, X_{n_1} \sim NID(\mu_1, \sigma_1^2) \\ \\ Y_1, Y_2, \cdots, Y_{n_2} \sim NID(\mu_2, \sigma_2^2) \end{array}\right\} \rightarrow \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

Example

Ratio of variances:

If X's and Y's are independent, calculate the following probability.

$$\left.\begin{array}{l} X_1, X_2, X_3 \sim NID(\mu_1, \sigma_1^2 = 4) \\ \\ Y_1, Y_2, Y_3 \sim NID(\mu_2, \sigma_2^2 = 16) \end{array}\right\} \rightarrow \Pr\left(\frac{S_1^2}{S_2^2} \geq 2\right) = ?$$

Summary of Sampling Distributions

		One population	Two populations
Sample Mean	Known Variance	$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim NID(0,1)$
	Unknown Variance	$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$	$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$
Sample Variance		$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$

Example

If X's and Y's are independent, what are the sampling distributions of a and b.

$$X_1, X_2, \cdots, X_n \sim NID(\mu, \sigma^2)$$

$$Y_1, Y_2, \cdots, Y_n \sim NID(\mu, \sigma^2)$$

$$a)\ \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(Y_i - \bar{Y})^2}{\sigma^2}$$

$$b)\ \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$