STAT 400 Fall 2011

Examples for 10/14/2011
(Chapter 5.6)

Population: mean μ, standard deviation σ.

Random Sample: $X_1, X_2, \ldots, X_n$.

or $X_1, X_2, \ldots, X_n$ are i.i.d.

$E(X_1 + X_2 + \ldots + X_n) = n \cdot \mu$, $SD(X_1 + X_2 + \ldots + X_n) = \sqrt{n} \cdot \sigma$.

If the sampling is done *without* replacement from a finite population of size N, then $SD(\Sigma X) = \sqrt{n} \cdot \sigma \cdot \sqrt{\frac{N-n}{N-1}}$.

The **sample mean** $\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$.

$E(\overline{X}) = \mu$, $SD(\overline{X}) = \frac{\sigma}{\sqrt{n}}$.

If the sampling is done *without* replacement from a finite population of size N, then $SD(\overline{X}) = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$.

$\overline{X} = \mu$ + chance error.

LAW OF LARGE NUMBERS (LAW OF AVERAGES):

As the sample size, n, increases, the sample mean, $\overline{X}$, "tends to gets closer and closer" to the population mean μ.
For Bernoulli trials, as # of trials n increases, the sample proportion of "successes", X/n, "tends to gets closer and closer" to the probability of "success" p.

Central Limit Theorem:

If the sample size, n, is large, the sampling distribution of the sample total is approximately **normal** with mean $n \cdot \mu$ and standard deviation $\sqrt{n} \cdot \sigma$.

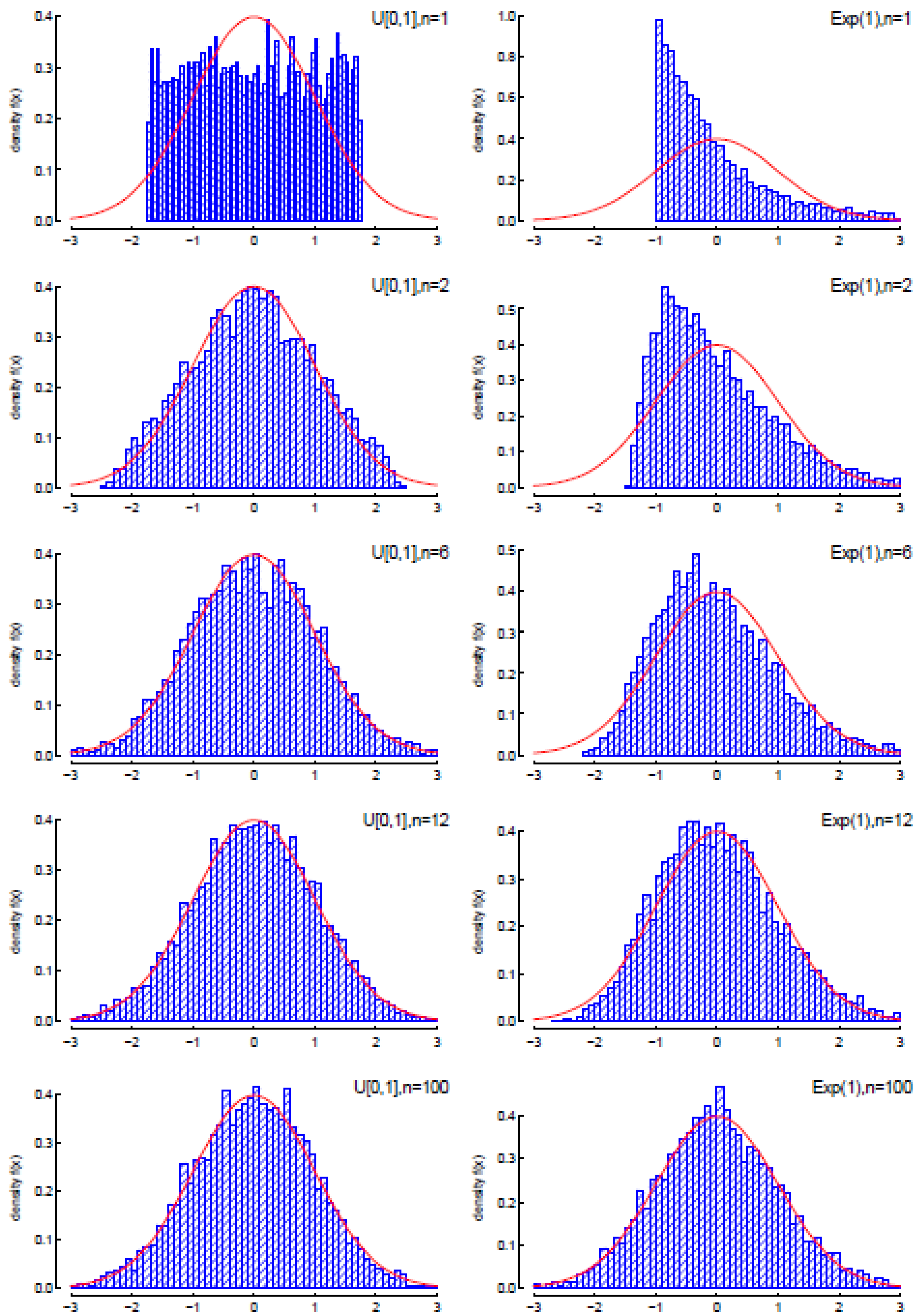
Therefore,

$$\frac{\Sigma X - n \cdot \mu}{\sqrt{n} \cdot \sigma} \approx Z.$$

If the population itself is normally distributed, the sampling distribution of the sample total is normal for ***any*** sample size n.

What happens when the population distribution we are sampling from is not "Normal"? For example, Xi ~ U[0,1] or Exp(1)?

"Sampling distribution" of $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$.



If the sample size, n, is large, the sampling distribution of the sample mean, $\overline{X}$, is approximately **normal** with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

Therefore,

$$\frac{\overline{X}-\mu}{\sigma / \sqrt{n}} \approx Z.$$

If the population itself is normally distributed, the sampling distribution of the sample mean, $\overline{X}$, is normal for ***any*** sample size n.

Remarks:

- How fast approximation becomes good depends on distribution of X_i's:
  - If it is symmetric and has tails that die off rapidly, n can be relatively small.
    Example: If $X_i \sim U[0,1]$ iid, the approximation is good for $n = 12$.
  - If it is very skewed or if its tails die down very slowly, a larger value of n is needed.
    Example: Exponential distribution.

Case 1. Any population
n – large ☺

Case 2. Normal population
Any n ☺

Case 3. Population NOT Normal
n – small ☹

1. A student commission wants to know the mean amount of money spent by college students for textbooks in one semester. Suppose the population mean is $450 and the population standard deviation is $40. A random sample of 625 students is taken.

a) What is the probability that the sample mean will be less than $452?

b) What is the probability that the sample mean will be within $2 of $450? That is, what is the probability that the sample mean will be between $448 and $452?

c) What is the probability that the sample mean will be within $10 of $450? That is, what is the probability that the sample mean will be between $440 and $460?

2. The amount of sulfur in the daily emissions from a power plant has a normal distribution with mean of 134 pounds and a standard deviation of 22 pounds. For a random sample of 5 days, find the probability that the total amount of sulfur emissions will exceed 700 pounds.

3. An economist wishes to estimate the average family income in a certain population. The population standard deviation is known to be $\$4{,}500$, and the economist uses a random sample of 225 families. What is the probability that the sample mean will fall within $\$600$ of the population mean?

4. Forty-eight measurements are recorded to several decimal places. Each of these 48 numbers is rounded off to the nearest integer. The sum of the original 48 numbers is approximated by the sum of these integers. If we assume that the errors made by rounding off are i.i.d. and have uniform distribution over the interval $\left(-\tfrac{1}{2}, \tfrac{1}{2}\right)$, compute approximately the probability that the sum of the integers is within 2 units of the true sum.