Calculus 131, section 13.X Central Limit Theorem

notes by Tim Pilachowski

If you haven't done it yet, go to the Math 131 page and download the handout The Central Limit Theorem. Today's lecture will use the material on pages 1 through 7, and tonight's homework assignment is #1 through #7 on pages 7-8 from that supplement.

Read the supplement for more extensive explanations of

– population vs. sample
– bias
– random sampling
– representative sample

including **Example 1** and **Example 2**.

The Central Limit Theorem

Probability models exist in a theoretical world where everything is known. If you constructed every possible sample of a specified size n from the population of Example 1, or were able to toss the coin of Example 2 an infinite number of times, you would create what statisticians call a **sampling distribution**. The ***Central Limit Theorem*** tells us, in short, that a sampling distribution is close to a normal distribution.

– population parameter
– sample statistic

What does this mean for random sampling? It tells us that 68% of the time, a random sample will give us a result—a statistic—within 1 standard deviation of the "true" parameter. We would expect that 95% of the time, a random sample will give a statistic within 2 standard deviations of the population parameter, and 99.7% of the time, a random sample will give a statistic within 3 standard deviations of the population parameter. In a beginning level statistics course, you would be introduced to confidence intervals and hypothesis tests, each of which make use of the percents listed above.

In the Examples below, we'll use a hypothetical population Ψ consisting of the numbers 10, 20, 30, 40 and 50. The parameter and statistic we'll consider first is the mean.

Example A-1: Calculate the mean and standard deviation of a population Ψ which consists of elements from the set {10, 20, 30, 40, 50} with probabilities given in the table below.

Value (x)	10	20	30	40	50
probability	0.4	0.2	0.2	0.0	0.2

Example A-2: Construct a histogram for population Ψ.

We would get the same histogram if we were to consider all possible samples of size $n = 1$ that could be taken from the population Ψ and calculated each sample's expected value (mean).

Example A-3: Construct all possible samples of size $n = 2$ that can be made from the elements of Ψ, designate the probability of each being picked, and calculate each sample's expected value (mean).

sample	P	$\bar{x}$	sample	P	$\bar{x}$	sample	P	$\bar{x}$	sample	P	$\bar{x}$
10, 10			20, 10			30, 10			50, 10		
10, 20			20, 20			30, 20			50, 20		
10, 30			20, 30			30, 30			50, 30		
10, 50			20, 50			30, 50			50, 50		

Example A-4: Draw the histogram for the sampling distribution from Example A-3.

$\bar{x}$	10	15	20	25	30	35	40	45	50
P									

Example A-5: Calculate mean and standard deviation of the sampling distribution of Ψ for sample size $n = 2$.

Mean of sampling distribution = $E(\bar{x})$ =

$$10(0.16)+15(0.16)+20(0.20)+25(0.08)+30(0.20)+35(0.08)+40(0.08)+45(0.0)+50(0.04)=$$

Variance of sampling distribution = $\text{Var}(\bar{x})$ =

$$(10-24)^2(0.16)+(15-24)^2(0.16)+(20-24)^2(0.20)+(25-24)^2(0.08)+(30-24)^2(0.20)+(35-24)^2(0.08)$$
$$+(40-24)^2(0.08)+(45-24)^2(0.0)+(50-24)^2(0.04)=$$

Standard deviation of sampling distribution = standard error =

Theory:

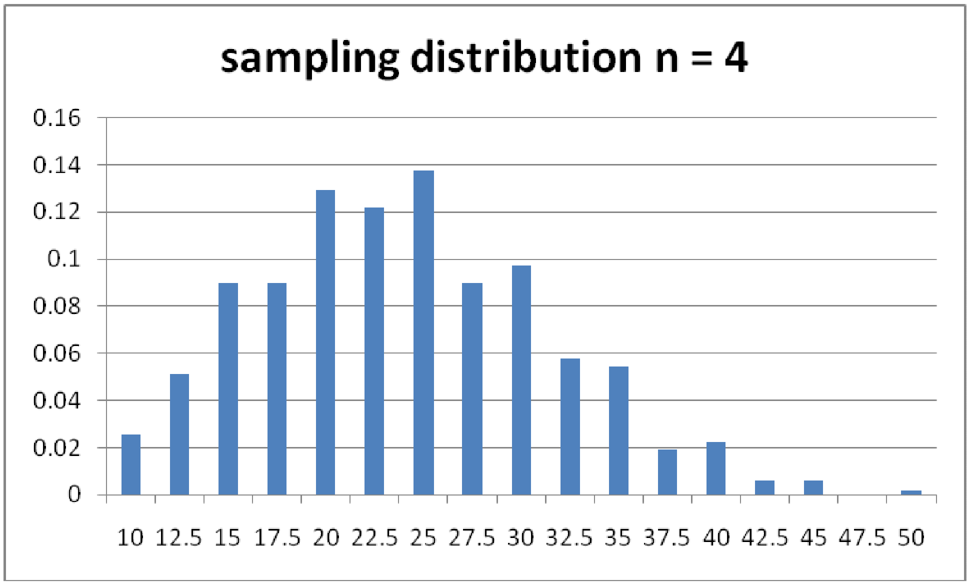
Example A-6: Draw the histogram and calculate the mean and standard deviation of the sampling distribution of Ψ for sample size $n = 4$.

Mean of sampling distribution = $E(\bar{x})$ =

Variance of sampling distribution = $\text{Var}(\bar{x})$ =

Standard deviation of sampling distribution = standard error =

The histogram for this sampling distribution (sample size $n = 4$) looks like this.



Example A-7: Explore the histogram for sampling distributions of Ψ for various sample sizes.

We'll use one of two sources:
http://www.chem.uoa.gr/applets/AppletCentralLimit/Appl_CentralLimit2.html
http://www.intuitor.com/statistics/CentralLim.html.

Notes:

Here's what the Central Limit Theorem says:

Given a population with mean μ and standard deviation σ:

1) As the sample size n increases, or as the number of trials n approaches infinite, the shape of a sampling distribution becomes increasingly like a normal distribution.

2) The mean of sampling distribution = the mean of the population.

3) The standard deviation of sampling distribution = standard error = $\dfrac{\sigma}{\sqrt{n}}$.

For statistics, a sample size of 30 is usually large enough to use the normal distribution probability table for hypothesis tests and confidence intervals. For Lecture examples, and for homework exercises from the handout, we'll use the normal distribution table to find various probabilities for sample statistics.

Example B. A population has mean $\mu = 150$ and standard deviation $\sigma = 22$. For a random sample of size 47, calculate a) the expected value of the sample mean and b) the standard error. Find the following probabilities: c) $P\left(145 < \bar{x} < 153\right)$ and d) $P\left(\bar{x} > 154\right)$.

answers: 150, $\frac{22}{\sqrt{47}}$, 0.7644, 0.1056

WARNING*:** Example B is ***not like those we did at the end of section 8.5 and in section 8.6! That is, it does not find probabilities for single values of x for a normally-distributed population, which uses $z = \dfrac{x-\mu}{\sigma}$.

Rather, Example B looks at one sample and finds probabilities involving the mean of that sample, $\bar{x}$, considered as part of all the hypothetical samples which could have been constructed (the sampling distribution). Therefore, for example B we used $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.

Also, in some homework exercises, you'll need to use the skills from previous sections to determine μ and σ.

Example C. A random variable X has probability density function $f(x) = 6x(1-x)$ on the interval $0 \le x \le 1$.

a) What are the expected value and standard deviation for a single randomly-chosen value of X?

b) You randomly select a sample of size $n = 100$. What is the expected value for the sample mean, $\bar{x}$? What is the standard error for the sampling distribution?

c) What is the probability that a single randomly-chosen subject from this population will exhibit a value of at least 0.55?

d) You randomly select a sample of size $n = 100$. What is the probability that the sample mean will be at least 0.55?

e) You randomly select a sample of size $n = 100$. There is a 25% probability that the sample mean will be below what value?

answers: $\frac{1}{2}$, $\frac{1}{2\sqrt{5}}$; $\frac{1}{2}$, $\frac{1}{20\sqrt{5}}$; 0.42525; 0.0125; 0.485

Go back to the supplement for more examples worked out. There are ones similar to each of the homework exercises.