Law of Large Numbers - pg. 294 Moore's Text

When tossing a fair coin the chances of tails and heads are the same: 50% and 50%. So, if the coin is tossed a large number of times, the number of heads and the number of tails should be approximately, equal.

This is the law of large numbers.

The number of heads will be off half the number of tosses by some amount. That amount is called chance error. So we have

no. of heads = half the no. of tosses + chance error

John Kerrich's coin-tossing Experiment

While imprisoned by the Germans during WWII, John Kerrich tossed a coin 10,000 times with heads coming up 5067 or 50.67 percent of the time. In 1946 he published his finding in a monograph, "An Experimental Introduction to the Theory of Probability". Here is what Nature said when the monograph came out: "When Denmark was overrun by the Germans various British subjects were caught, Mr Kerrich among them. He was interned in a camp under Danish control and spent part of his enforced leisure in coin-tossing experiments"

His data are shown in the table below:

Table 1. John Kerrich's coin-tossing experiment. The first column shows the number of tosses. The second shows the number of heads. The third shows the difference

number of heads − half the number of tosses.

Number of tosses	Number of heads	Difference	Number of tosses	Number of heads	Difference
10	4	−1	600	312	12
20	10	0	700	368	18
30	17	2	800	413	13
40	21	1	900	458	8
50	25	0	1,000	502	2
60	29	−1	2,000	1,013	13
70	32	−3	3,000	1,510	10
80	35	−5	4,000	2,029	29
90	40	−5	5,000	2,533	33
100	44	−6	6,000	3,009	9
200	98	−2	7,000	3,516	16
300	146	−4	8,000	4,034	34
400	199	−1	9,000	4,538	38
500	255	5	10,000	5,067	67

Figure 1. Kerrich's coin-tossing experiment. The "chance error" is

number of heads − half the number of tosses.

This difference is plotted against the number of tosses. As the number of tosses goes up, the size of the chance error tends to go up. The horizontal axis is not to scale and the curve is drawn by linear interpolation.
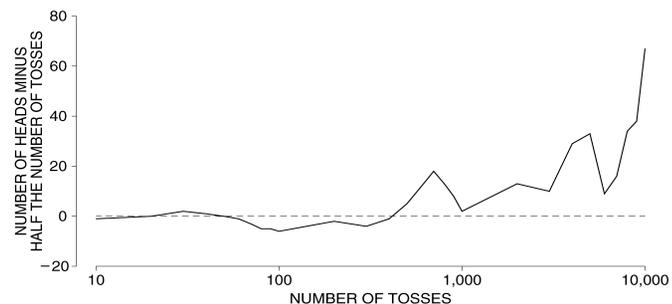
Figure 2. The chance error expressed as a percentage of the number of tosses. When the number of tosses goes up, this percentage goes down: the chance error gets smaller relative to the number of tosses. The horizontal axis is not to scale and the curve is drawn by linear interpolation.
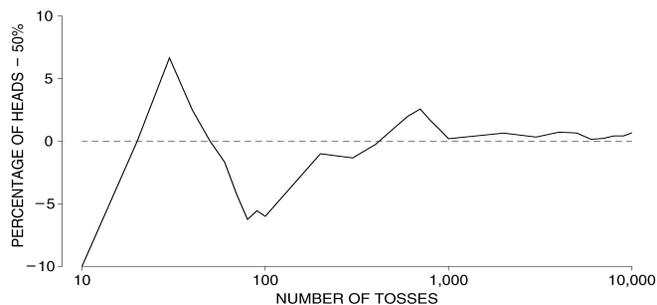
John Kerrich's coin-tossing Experiment

The **chance error increases** with the number of tosses in absolute terms, but it **decreases in relative terms**.

Q: A coin is tossed and you win $1 if there are more than 60% heads. Which is better: 10 tosses or 100?

John Kerrich's coin-tossing Experiment

The **chance error increases** with the number of tosses in absolute terms, but it **decreases in relative terms**.

Q: A coin is tossed and you win $1 if there are more than 60% heads. Which is better: 10 tosses or 100?

A: 10 tosses is better. As the number of tosses increase you are more likely to be closer to 50%, according to the law of averages.

Q: A coin is tossed and you win $1 if there are more than 48% heads. Which is better: 10 tosses or 100?

John Kerrich's coin-tossing Experiment

The **chance error increases** with the number of tosses in absolute terms, but it **decreases in relative terms**.

Q: A coin is tossed and you win $1 if there are more than 48% heads. Which is better: 10 tosses or 100?

A: 100 tosses is better, because the law of averages is working for you – with more tosses, you are more likely to be close to 50%.

Random Variables

Random variables have an element of uncertainty or variability within them. Two quantities that describe the behavior of a random variable are the Expected Value and the Standard Error.

Roulette Example

A roulette wheel has 38 pockets. 1 through 36 are alternatively colored red and black, plus 0 and 00 which are colored green. So, there are 18 red pockets, 18 black and 2 green ones.

Let's take a look at a roulette table layout.

Roulette Example

A roulette wheel has 38 pockets. 1 through 36 are alternatively colored red and black, plus 0 and 00 which are colored green. So, there are 18 red pockets, 18 black and 2 green ones.

Suppose you bet on any number. If it comes up, you win $35, otherwise you loose $1.

Your chance of winning is 1 in 38 and your chance of loosing is 37 in 38. Find the expected Gain or Loss in one spin of roulette.

$$E[Gain] = \frac{1}{38} \times (+\$35) + \frac{37}{38} \times (-\$1) \sim -\$0.053$$

Expected Value

Count the number of wins in 1000 plays of a roulette. The law of large numbers tells you that the expected value of the number of wins is $E[Wins] = \frac{1}{38} \times 1000 = 26$. You actually play the games and the results are

- 31 wins, you are off by +5

- 18 wins, you are off by -8

- 24 wins, you are off by -2

The *"amounts off"* are similar in size to the **standard error**, which we will define in a couple of slides. The expected value and the standard error depend on the random process that generates the numbers.

Expected Gain/Loss Playing Roulette

Q: Suppose you play 1,000 games of roulette betting **+$1 on #7** at each play. If you win, you get your dollar plus 35 more dollars, if you lose, the casino keeps your dollar. What are the chances that you will come out ahead from these 1,000 plays?

> We will build a **Probability Model** and identify **the Population, Sample and Imaginary Data sets** to address this question. Then we'll come back to the slides to address the **long run mean of Sums**.

Expected Gain/Loss Playing Roulette

- **What's the chance we come out ahead if we play 1000 games of roulette?**

Well, after 1 play you expect to be behind by

$$E[Gain] = \frac{1}{38} \times (+\$35) + \frac{37}{38} \times (-\$1) \sim -\$0.053$$

So, after 1000 plays, we expect to be behind by: $\underbrace{}_{?}$

And, give or take how much? Maybe the give or take is big enough that you still have a pretty good chance of coming out ahead.

The Standard Error for Sums

Sum of money after 1000 plays   = Observed Value

                           = Expected Value + chance error

The **standard error** gives a measure of how large the **chance error** is likely to be.

We can calculate the standard error for the sum of the 1000 plays as

$$\sqrt{\text{sample size}} \times (\textbf{SD of pop})$$

where "SD of pop" stands for the standard deviation of the population.

Back to our original Question

Now we'll use our **Probability Model** to address the question **What's the chance you come out ahead if you play 1000 games of roulette?**

Symbolically, that's $P(S > \$0)$, which is equivalent to wondering what percentage of the sums in the imaginary dataset are positive. If we knew what the histogram of the sums looked like, we could answer this question by working out the area under the histogram to the right of $0.

The **Central Limit Theorem** will tell us about the shape of the histogram of the sums.

Central Limit Theorem (pg. 302)

Let's discuss the central limit theorem for the population:
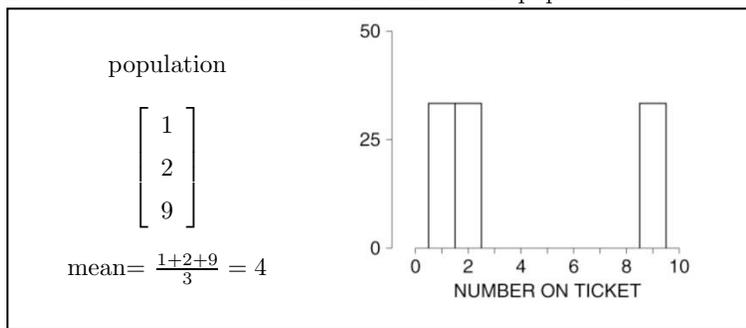
population

$$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$$

mean= $\frac{1+2+9}{3} = 4$

50

25

0

0   2   4   6   8   10
NUMBER ON TICKET

Central Limit Theorem

Consider taking an IID random sample of 25 draws from the population and sum the draws. The sum ought to be around

$$25 \times \text{mean} = 25 \times 4 = 100.$$

Now, imagine repeating this sampling story a lot – that is, take 25 IID draws from the population, work out their sum; take 25 more, work out their sum, and so on, many times. Make a histogram of all the sums you would get. Theory says it would look like the wavy histogram below:
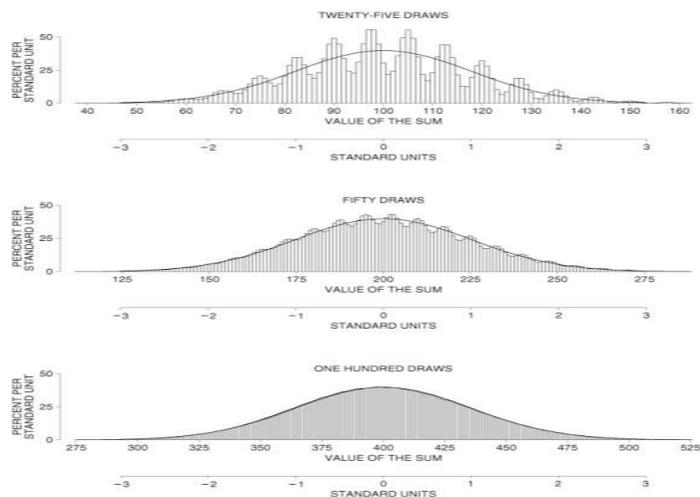
Central Limit Theorem

So, when you take 100 IID draws from the population, we see that the long run histogram of *sums* follows the normal curve. This is the Central Limit Theorem in action: as the number (n) of draws going into a sum goes up, the long-run histogram for the sum looks more and more like the normal curve.

The number of draws needed to get a good normal approximation depends on how close the shape of the *population* histogram is to normal. For example:

Central Limit Theorem

Let's consider taking an IID random sample of 100 draws from the population:
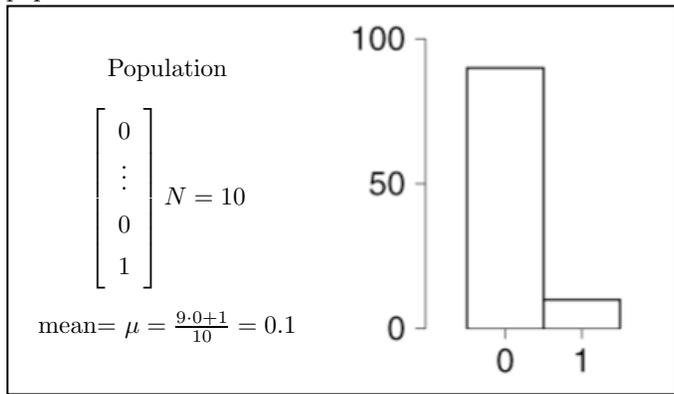
Population

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} N = 10$$

mean= $\mu = \frac{9 \cdot 0 + 1}{10} = 0.1$

Central Limit Theorem

Consider taking an IID random sample of 25 draws from the population and sum the draws. The sum ought to be around

$$25 \times \text{mean} = (25) \cdot (0.1) = 2.5$$

Now, do the same thought experiment of repeating this sampling story a lot – that is, take 100 IID draws from the population, work out their sum; take 25 more, work out their sum, and so on, many times. Make a long run histogram of all the sums you would get. We'd see that the long run histograms approximate the normal curve with fewer draws from the sample than before.

Central Limit Theorem
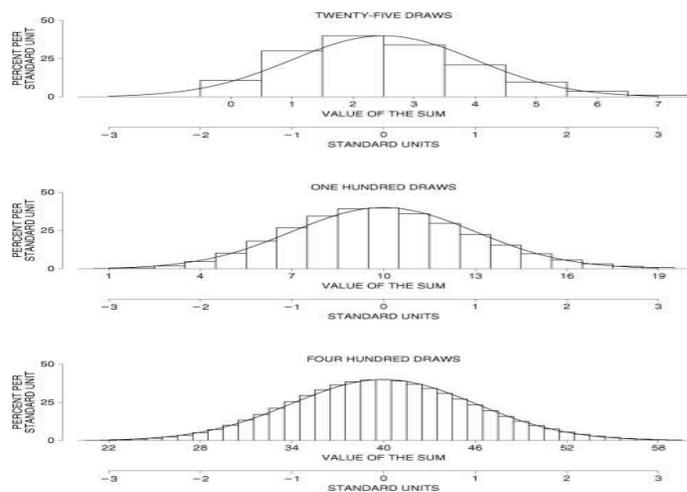
→Back to our original Question...
...Roulette, in case you've forgotten←

Now we'll use our **Probability Model** to address the question **What's the chance you come out ahead if you play 1000 games of roulette?**

Symbolically, that's $P(S > \$0)$, which is equivalent to wondering what percentage of the sums in the imaginary dataset are positive.

Now that we know that the long run histogram of the sums looks like the normal curve, we could answer this question by working out the area under the histogram to the right of $0.

→Let's work that out←

Review of Concepts

Let's review what we've learned so far. When given information about a population of data, building a *population model*, enables us to:

- Calculate long run means to find expected values

- Calculate long run *sd's*, (standard error), to find give-or-take of long run mean estimates

- Answer questions about the chances of getting a value other than the long run mean... like, the chances of coming out ahead in 1000 spins of a roulette wheel.

NOTE: The Expected Value and SE describe the **sampling distribution**.

Q: What is the difference between the SD and the SE?

Answer:

- The SD says that the amount of loss on a single roulette wheel spin when betting on #7 is about 5 cents.

- The SE says that the average loss of 1000 plays of roulette is accurate up to $182.

SD is related to the precision of **single measurements**.

SE is related to the precision of the **average**.

In the previous example, any specific loss is only accurate to about 5 cents. The estimated loss after playing 1000 games of roulette, based on the Imaginary Dataset is accurate to about $182.

Expected Gain/Loss Playing Roulette - Sums

Which has better odds? Compare the chances of coming out ahead after 1,000 plays of roulette by:

- **placing 1000 $1 bets on #7**

- **placing 1000 $1 bets on a split**

If you put $1 on a split, say 11/12, and either 11 or 12 comes up, you get back your dollar plus $17 in winnings.

> **Build a probability model describing the population, sample and imaginary data sets to address this question.**

Measurement Error

If the Potassium levels in your blood are measured you might find measurements such as:

$$\overline{Y} = \begin{bmatrix} 3.8 \\ 3.8 \\ 3.8 \\ \vdots \end{bmatrix} \qquad \text{At finer level} \rightarrow \qquad \overline{Y} = \begin{bmatrix} 3.82 \\ 3.79 \\ 3.82 \\ \vdots \end{bmatrix}$$

Not Random - Deterministic

Stochastic - Random - Probabilistic

What explains the difference in the random measures?

Basic Measurement Error Model

$$obs_1 \quad = \quad \text{true value} \quad + \quad \text{bias} \quad + \quad \text{random error}_1$$
$$obs_2 \quad = \quad \text{true value} \quad + \quad \text{bias} \quad + \quad \text{random error}_2$$
$$\vdots$$
$$obs_n \quad = \quad \text{true value} \quad + \quad \text{bias} \quad + \quad \text{random error}_n$$

Bias is a systematic tendency to over or underestimate the true value.

Bias can't be detected from the data – we need an external standard.

Measurement Error - SE$[\overline{x}]$

Alright, let's practice our skills at building a probability model for another example.

Any measurement is subject to chance error. To reduce the size of the chance error, the best thing to do is to repeat the measurements several times. This is the Law of Large Numbers at work. As the number of observation drawn increases, the mean of the observed values, $\overline{x}$, gets closer to the mean of the population, μ.

Case Study: Hypokalemia

Set up a Probability Model, describing the population, sample and imaginary data sets.

Measurement Error - Example Problem 2

100 measurements of the NB 10, a weight owned by the National Bureau of Standards, are taken to determine the **true weight**. The nominal weight of the NB 10 is 10 grams. The units, in micrograms below 10 grams, are recorded.

Given that the sample mean is 404.6 μgms with a SD of 6 μgms, find the best estimate for the true weight of the NB 10. Find the give-or-take for the estimate.

Set up a Statistical Model, describing the population, sample and imaginary data sets.

Q: What is the difference between the SD and the SE?

Answer:
- The SD says that the a single measurement is accurate up to 6 μgms or so.
- The SE says that the mean of all 100 measurements is accurate up to 0.6 μgms or so.

SD is related to the precision of **single measurements**.
SE is related to the precision of the **mean**.

In the previous example, any specific measurement is only accurate by about 6 μgms. The estimated weight of the NB 10, based on the mean of 100 measurements, is accurate by about 0.6 μgms.

Sampling with or without replacement

> Suppose we are polling New Mexico and Texas to estimate the voting intentions in a presidential election. NM has about 1.2 million voters and TX has about 12.5. Will we need to poll more people in TX to achieve the same accuracy on both state polls?

When taking a sample from a finite population it is important to bear in mind two important issues:

- **The accuracy is not determined by the size of the sample relative to the population.**

- **Sampling with or without replacement produces almost the same results when the population size is large.**

The first statement is somewhat counterintuitive. Intuitively, we think that to achieve the same accuracy we need a larger sample in Texas than New Mexico, however, **this is not true!**

Sampling with or without replacement

What counts is the absolute size of the sample. The formula for the SE of a percentage or mean **does not contain any** information of the population size.

Remember,

$$SE(\text{mean}) = \frac{\sigma}{\sqrt{n}}$$

Consider the chemical composition of a liquid. If the liquid is well mixed, then a drop should accurately tell us about the composition regardless of whether it is taken from a small test tube or from a large jug (*e.g.* whether the population size is relatively large or otherwise).

Probability vs. Statistical Inference

Probability -- DEDUCTION

Population   Sample
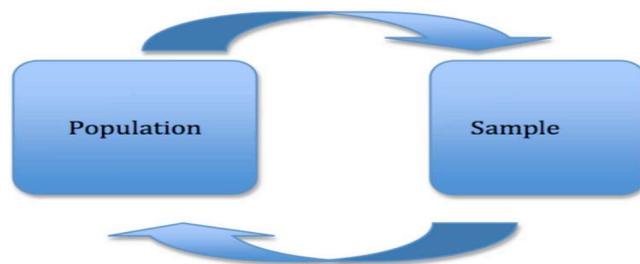
Statistical Inference -- INDUCTION

SAMPLING SCHEMES:

Independent and Identically Distributed (IID) - sampling with replacement – Math is simpler

Simple Random Sampling (SRS) - sampling without replacement – usually used in practice

---

Case Study: The Chesapeake and Ohio Freight Study

> We'll build a statistical model describing the population, sample and imaginary data sets to address this business auditing application.