The Central Limit Theorem

Patrick Breheny

March 5

Kerrich's experiment

- A South African mathematician named John Kerrich was visiting Copenhagen in 1940 when Germany invaded Denmark
- Kerrich spent the next five years in an interment camp
- To pass the time, he carried out a series of experiments in probability theory
- One of them involved flipping a coin 10,000 times

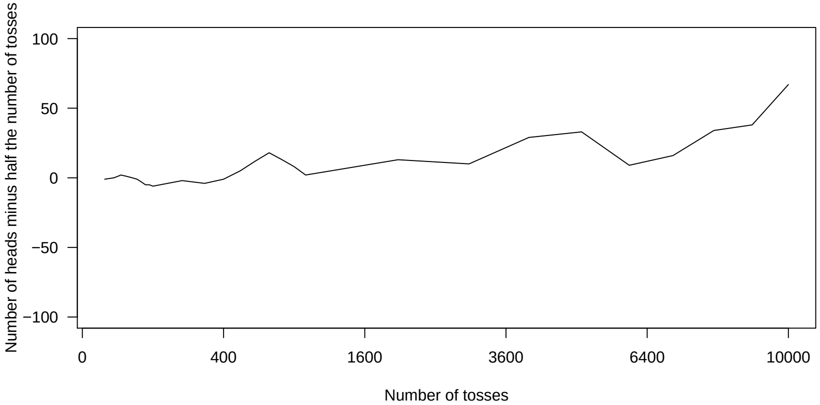
The law of averages

- We know that a coin lands heads with probability 50%
- Thus, after many tosses, the law of averages says that the number of heads should be about the same as the number of tails . . .
- . . . or does it?

Kerrich's results

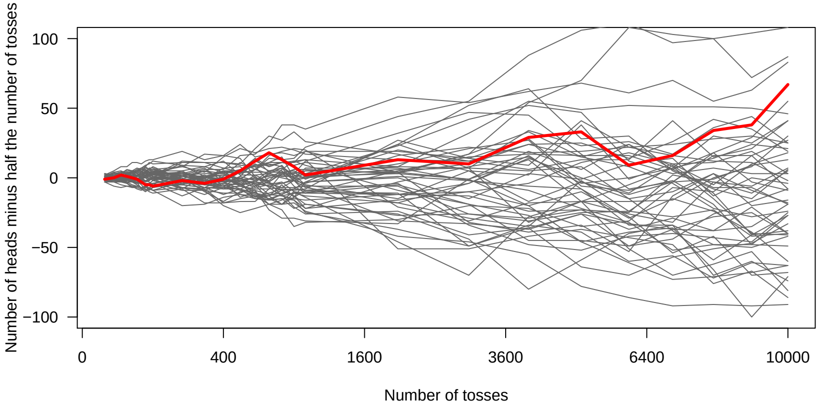
Number of tosses (n)	Number of heads	Heads - $0.5\cdot$Tosses
10	4	-1
100	44	-6
500	255	5
1,000	502	2
2,000	1,013	13
3,000	1,510	10
4,000	2,029	29
5,000	2,533	33
6,000	3,009	9
7,000	3,516	16
8,000	4,034	34
9,000	4,538	38
10,000	5,067	67

Kerrich's results plotted



Instead of getting closer, the numbers of heads and tails are getting farther apart

Repeating the experiment 50 times

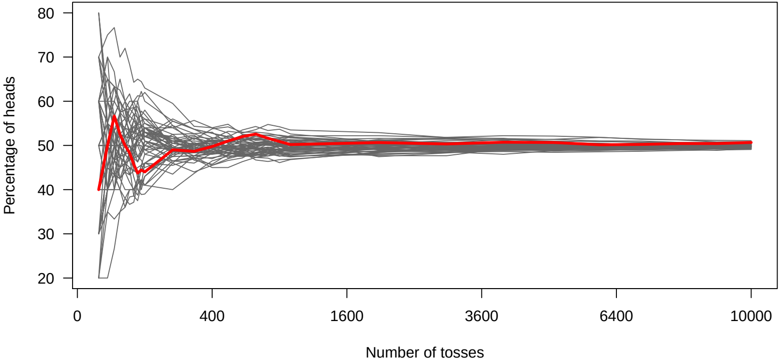


This is not a fluke – instead, it occurs systematically and consistently in repeated simulated experiments

Where's the law of averages?

- So where's the law of averages?
- Well, the law of averages does **not** say that as n increases the number of heads will be close to the number of tails
- What it says instead is that, as n increases, the average number of heads will get closer and closer to the long-run average (in this case, 0.5)
- The technical term for this is that the sample average, which is an estimate, *converges* to the population mean, which is a parameter

Repeating the experiment 50 times, Part II



Trends in Kerrich's experiment

- There are three very important trends going on in this experiment
- We'll get to those three trends in a few minutes, but first, I want to introduce two additional, important facts about the binomial distribution: its mean (expected value) and standard deviation

The expected value of the binomial distribution

- Recall that the probability of an event is the long-run percent of time it occurs
- An analogous idea exists for random variables: if we were to measure a random variable over and over again an infinite number of times, the average of those measurements would be the *expected value* of the random variable
- For example, the expected value of a random variable X following a binomial distribution with n trials and probability p is np:

$$E(X) = np$$

- This makes sense; if you flip a coin 10 times, you can expect 5 heads

The standard deviation of the binomial distribution

- Of course, you won't always get 5 heads
- Because of variability, we are also interested in the standard deviation of random variables
- For the binomial distribution, the standard deviation is

$$\mathrm{SD}(X) = \sqrt{np(1-p)}$$

- To continue our example of flipping a coin 10 times, here the SD is $\sqrt{10(0.5)(0.5)} = 1.58$, so we can expect the number of heads to be 5 ± 3 about 95% of the time (by the 95% rule of thumb)
- Note that the SD is highest when $p = 0.5$ and gets smaller as p is close to 0 or 1 – this makes sense, as if p is close to 0 or 1, the event is more predictable and less variable

Trends in Kerrich's experiment

- As I said a few minutes ago, there are three very important trends going on in this experiment
- These trends can be observed visually from the computer simulations or proven via the binomial distribution
- We'll work with both approaches so that you can get a sense of how they both work and how they reinforce each other

The expected value of the mean

- The expected value of the binomial distribution is np; what about the expected value of its *mean*?
- The mean (i.e., the sample proportion) is

$$\hat{p} = \frac{X}{n},$$

so its expected value is

$$\begin{aligned} E(\hat{p}) &= \frac{E(X)}{n} \\ &= \frac{np}{n} \\ &= p \end{aligned}$$

- In other words, for any sample size, the expected value of the sample proportion is equal to the true proportion (i.e., it is not biased)

The standard error of the mean

- Likewise, but the standard deviation of the binomial distribution is $\sqrt{np(1-p)}$, but what about the SD of the mean?
- As before,

$$\begin{aligned} \mathrm{SD}(\hat{p}) &= \frac{\mathrm{SD}(X)}{n} \\ &= \frac{\sqrt{np(1-p)}}{n} \\ &= \sqrt{\frac{p(1-p)}{n}} \end{aligned}$$

Standard errors

- Note that, as n goes up, the variability of the # of heads goes up, but the variability of the average goes down – just as we saw in our simulation
- Indeed, the variability goes to 0 as n gets larger and larger – this is the law of averages
- The standard deviation of the average is given a special name in statistics to distinguish it from the sample standard deviation of the data
- The standard deviation of the average is called the *standard error*
- The term *standard error* refers to the variability of any estimate, to distinguish it from the variability of individual tosses or people

The square root law

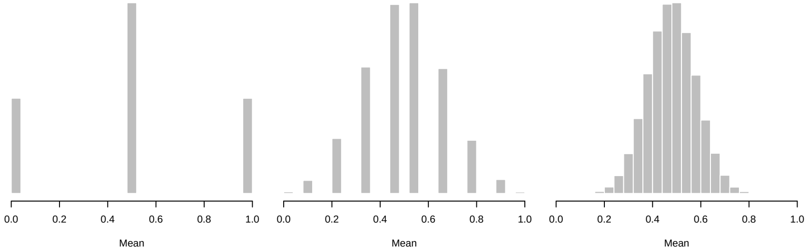
- The relationship between the variability of an individual (toss) and the variability of the average (of a large number of tosses) is a very important relationship, sometimes called the *square root law*:

$$\mathrm{SE} = \frac{\mathrm{SD}}{\sqrt{n}},$$

  where SE is the standard error of the mean and SD is the standard deviation of an individual (toss)
- We saw that this is true for tosses of a coin, but it is in fact true for all averages
- Once again, we see this phenomenon visually in our simulation results

The distribution of the mean

Finally, let's look at the distribution of the mean by creating histograms of the mean in our simulation



The central limit theorem

- In summary, there are three very important phenomena going on here concerning the sampling distribution of the sample average:
  - **#1** The expected value is always equal to the population average
  - **#2** The standard error is always equal to the population standard deviation divided by the square root of n
  - **#3** As n gets larger, the sampling distribution looks more and more like the normal distribution
- Furthermore, these three properties of the sampling distribution of the sample average hold for **any distribution** – not just the binomial

The central limit theorem (cont'd)

- This result is called the *central limit theorem*, and it is one of the most important, remarkable, and powerful results in all of statistics
- In the real world, we rarely know the distribution of our data
- But the central limit theorem says: we don't have to

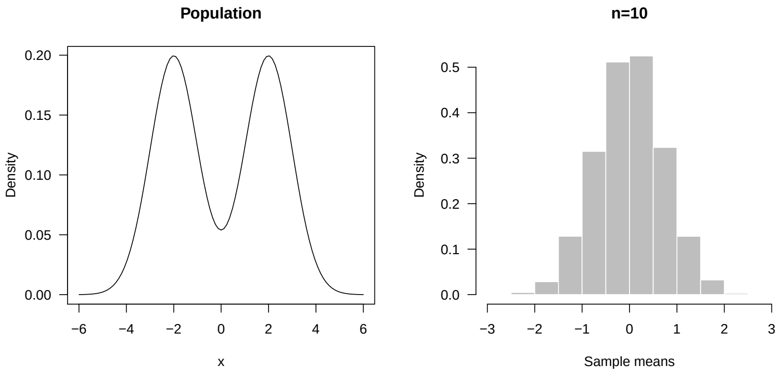
The central limit theorem (cont'd)

- Furthermore, as we have seen, knowing the mean and standard deviation of a distribution that is approximately normal allows us to calculate anything we wish to know with tremendous accuracy – and the sampling distribution of the mean is always approximately normal
- The only caveats:
  - Observations must be independently drawn from and representative of the population
  - The central limit theorem applies to the sampling distribution of the mean – not necessarily to the sampling distribution of other statistics
  - How large does n have to be before the distribution becomes close enough in shape to the normal distribution?

How large does n have to be?

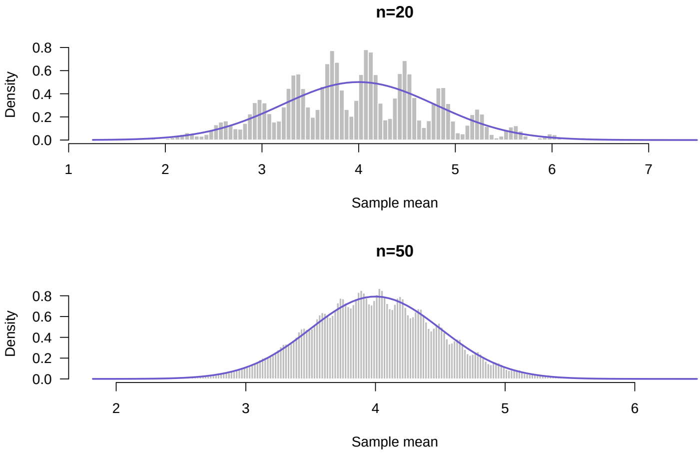
- Rules of thumb are frequently recommended that $n = 20$ or $n = 30$ is "large enough" to be sure that the central limit theorem is working
- There is some truth to such rules, but in reality, whether n is large enough for the central limit theorem to provide an accurate approximation to the true sampling distribution depends on how close to normal the population distribution is
- If the original distribution is close to normal, $n = 2$ might be enough
- If the underlying distribution is highly skewed or strange in some other way, $n = 50$ might not be enough

Example #1

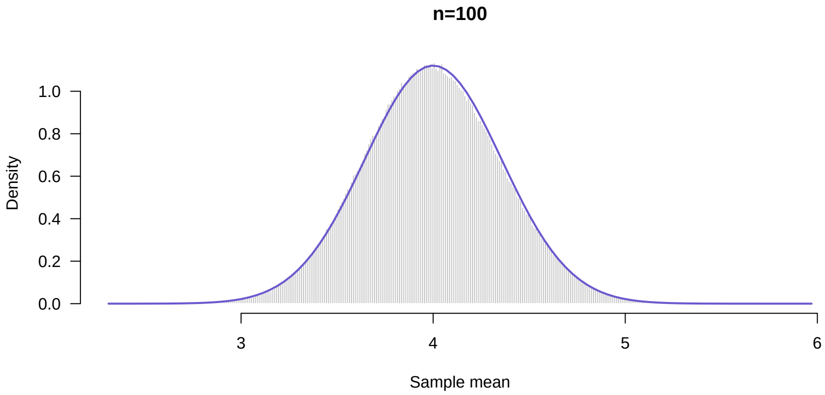

Population
Density
0.20
0.15
0.10
0.05
0.00
−6
−4
−2
0
2
4
6
x
n=10
Density
0.5
0.4
0.3
0.2
0.1
0.0
−3
−2
−1
0
1
2
3
Sample means

Example #2

Now imagine an urn containing the numbers 1, 2, and 9:



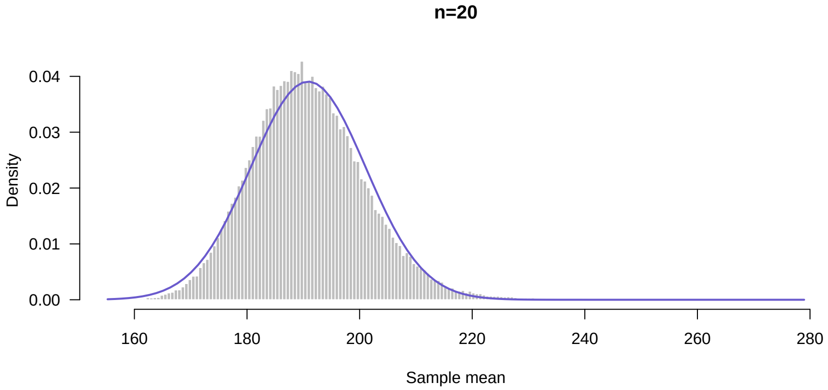
Example #2 (cont'd)



Example #3

- Weight tends to be skewed to the right (far more people are overweight than underweight)
- Let's perform an experiment in which the NHANES sample of adult men is the population
- I am going to randomly draw twenty-person samples from this population (*i.e.* I am re-sampling the original sample)

Example #3 (cont'd)



Why do so many things follow normal distributions?

- We can see now why the normal distribution comes up so often in the real world: any time a phenomenon has many contributing factors, and what we see is the average effect of all those factors, the quantity will follow a normal distribution
- For example, there is no one cause of height – thousands of genetic and environmental factors make small contributions to a person's adult height, and as a result, height is normally distributed
- On the other hand, things like eye color, cystic fibrosis, broken bones, and polio have a small number of (or a single) contributing factors, and do not follow a normal distribution

Summary

- Central limit theorem:
  - The expected value of the average is always equal to the population average
  - $\mathrm{SE} = \mathrm{SD}/\sqrt{n}$
  - As n gets larger, the sampling distribution looks more and more like the normal distribution
- Generally speaking, the sampling distribution looks pretty normal by about $n = 20$, but this could happen faster or slower depending on the population and how skewed it is