Week 4: Sampling Distributions

Jessie Yeung

STA 220

Winter 2024

Overview

- This week will cover Module 4
  - Module 4: Sampling Distributions
- Topics include:
  - Distribution of a Sample Mean
  - Distribution of a Sample Proportion
  - Mean and Variance of a Proportion
  - Sampling Distributions

Recall from last week: Random Variables

- Last week we introduced random variables, which are variables that take on specific realized values, at specific probabilities
- Random variables are denoted using capital letters
- Some random variables follow known distributions:
  - $X \sim Bern(p)$
  - $X \sim Bin(n, p)$
  - $X \sim N(\mu, \sigma^2)$

Recall from last week: Repeating Experiments

- Each time I can run an experiment, I can define my random variable X_i to represent the possible outcomes of the experiment
- If I decide I will run my experiment n times, then I have the random variables $X_1, X_2, \dots, X_n$ representing the possible outcomes for each time I run the experiment
- Each X_i has the same probability distribution representing the possible outcomes
- Oftentimes, we can use the random variables $X_1, X_2, \dots, X_n$ to represent our data

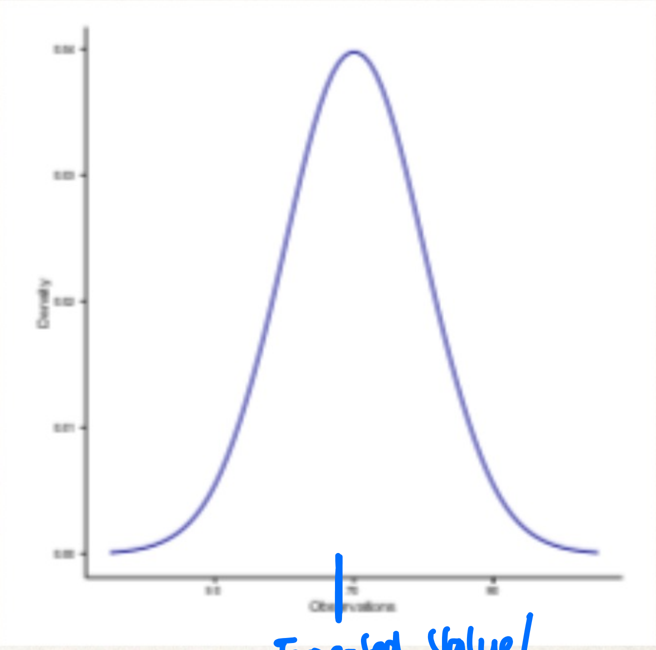
Goal for this week

- Functions of random variables are also random variables
- Our data from our experiment, $X_1, X_2, \dots, X_n$, are all random variables. This means that **sample statistics** (summary measures of the data such as sample mean) are also random variables
- Question: Can we find the distribution of the sample statistics?

Distribution of Sample Mean

Probability Distributions

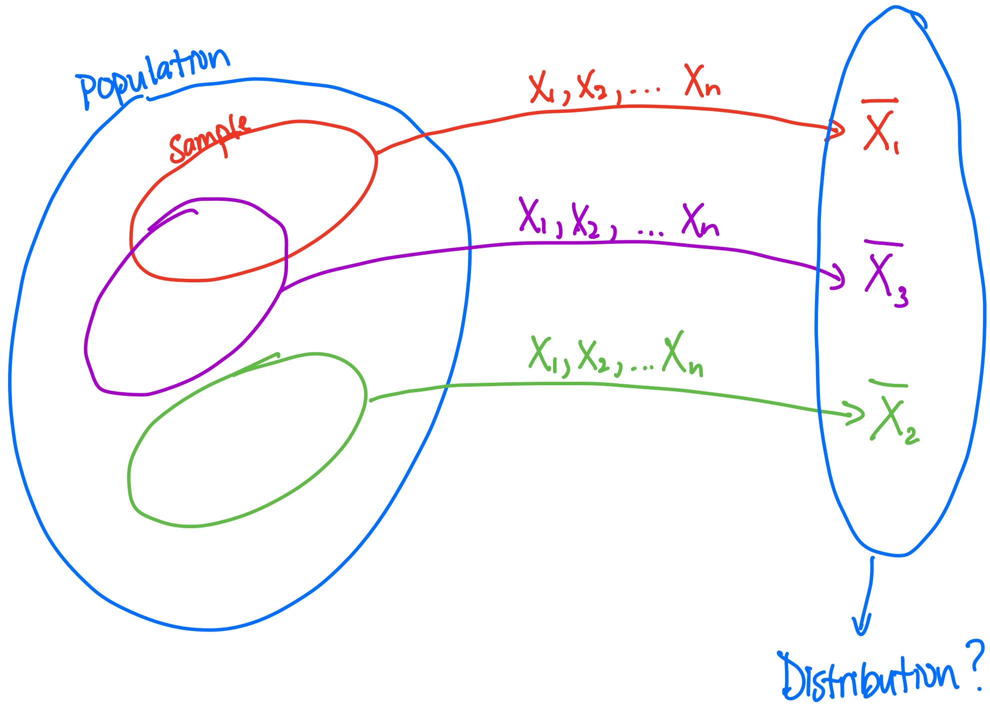
- When we have a **probability distribution**, we are representing the chances of seeing each possible outcome of the experiment
- So the graph shows us the probability for each possible outcome
  - possible outcomes that are more likely to be observed are closer to the peak
  - as we move away from the middle/peak of the distribution, the chances of seeing those possible outcomes decreases
  - the **expected value** roughly corresponds to the peak of the distribution - it has the highest chance of being observed



Expected Value/
Mean

Distribution of Sample Means

- When we run an experiment or collect some data, we are not guaranteed to see all possible outcomes but we could see any of them.
- But because each run is taking samples from the same probability distribution, we have <u>higher chances of seeing values in the centre of the distribution</u> than at the ends
- Every time we run an experiment, we will get a different collection of outcomes each time
  - one collection of outcomes represents our **sample data** from one run of the experiment
  - once we have a sample, we often like to summarize it using a **sample mean**
  - each time we run the experiment, we can then compute a sample mean for each sample we get.
  - I can treat my collection of sample means from each run as my new data.



Population
Sample
$X_1, X_2, \ldots X_n$
$X_1, X_2, \ldots X_n$
$X_1, X_2, \ldots X_n$
$\bar{X}_1$
$\bar{X}_3$
$\bar{X}_2$
Distribution?

Distribution of Sample Means

data set of $\bar{X}$'s

- Now that I have new data, my collection of sample means, I want to know their distribution
- But what does the *distribution of sample means* even mean?
  - The data I used to get the sample means were sampled from a distribution of possible values
  - If I were to keep sampling data forever, I would be able to eventually collect all the sample means corresponding to all possible combinations of sampled data
    - Then I would know how often I would see any particular sample mean
    - This is the distribution of sample means.
  - The reason I have to talk about a distribution of sample means is because I don't ever see all of the possible sample means - just a sample of them
    - just like I never see all possible values of my data - only a sample of them

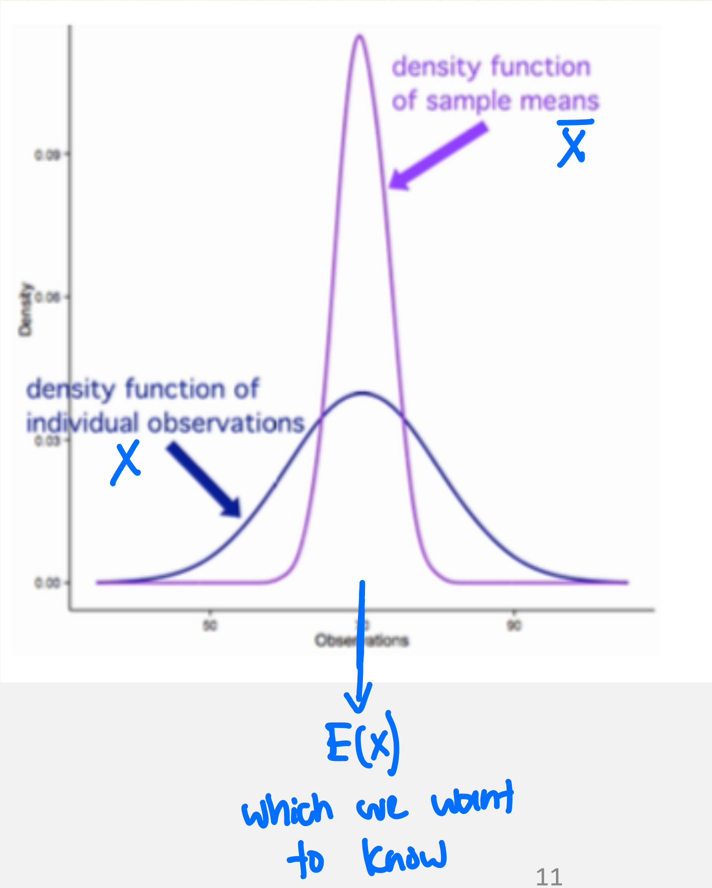
$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

Distribution of Sample Means

- Example: Suppose we are interested in the average arm span of 5 year old females. We collect a sample of n measurements.
  - A measurement X (the arm span of a single 5yo female) is a random variable and follows some probability distribution
  - A sample mean $\bar{X}$ (the sample mean of a sample of size n) is also a random variable and follows some probability distribution
  - However, as we will see, they do not follow the same probability distribution
  - Key for now is to understand that these are 2 separate random variables and both have their own distributions

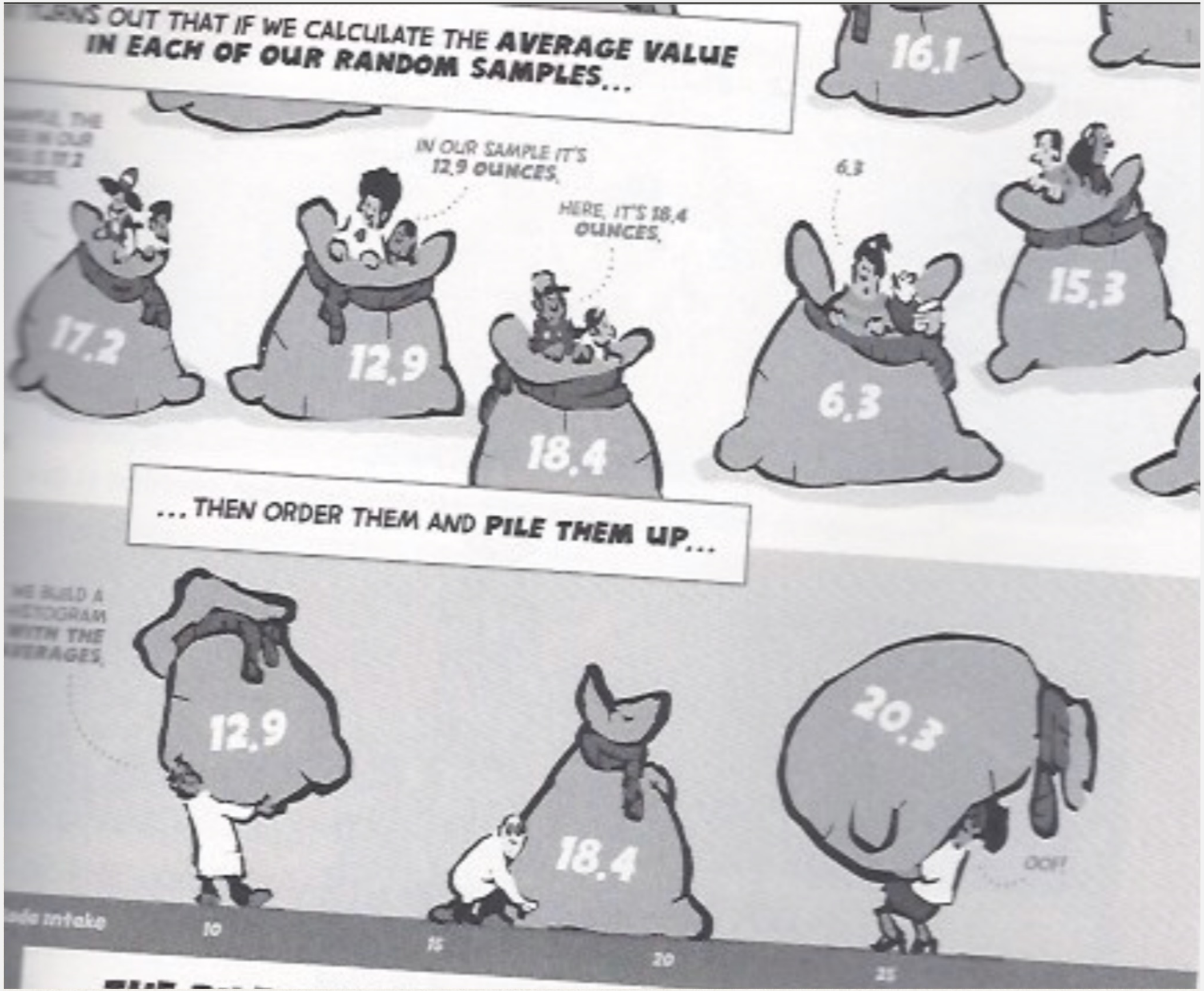
Why do I need this? $E(\bar{X}) = E(X)$

- We saw last week that the expected value of the sample mean $E(\bar{X})$, i.e. value I am most likely to get for $\bar{X}$, is the same as $E(X)$
  - centre of data distribution is same as centre of distribution of means
- But if I try to estimate $E(X)$ from one measurement of X, the variability/precision of my estimate is $Var(X)$
- If instead I collect n measurements, calculate $\bar{X}$ and instead use $\bar{X}$ to estimate $E(X)$:
  - I will still get $E(\bar{X}) = E(X)$
  - But $Var(\bar{X}) = \frac{Var(X)}{n}$, so my estimate is much more precise/less variable

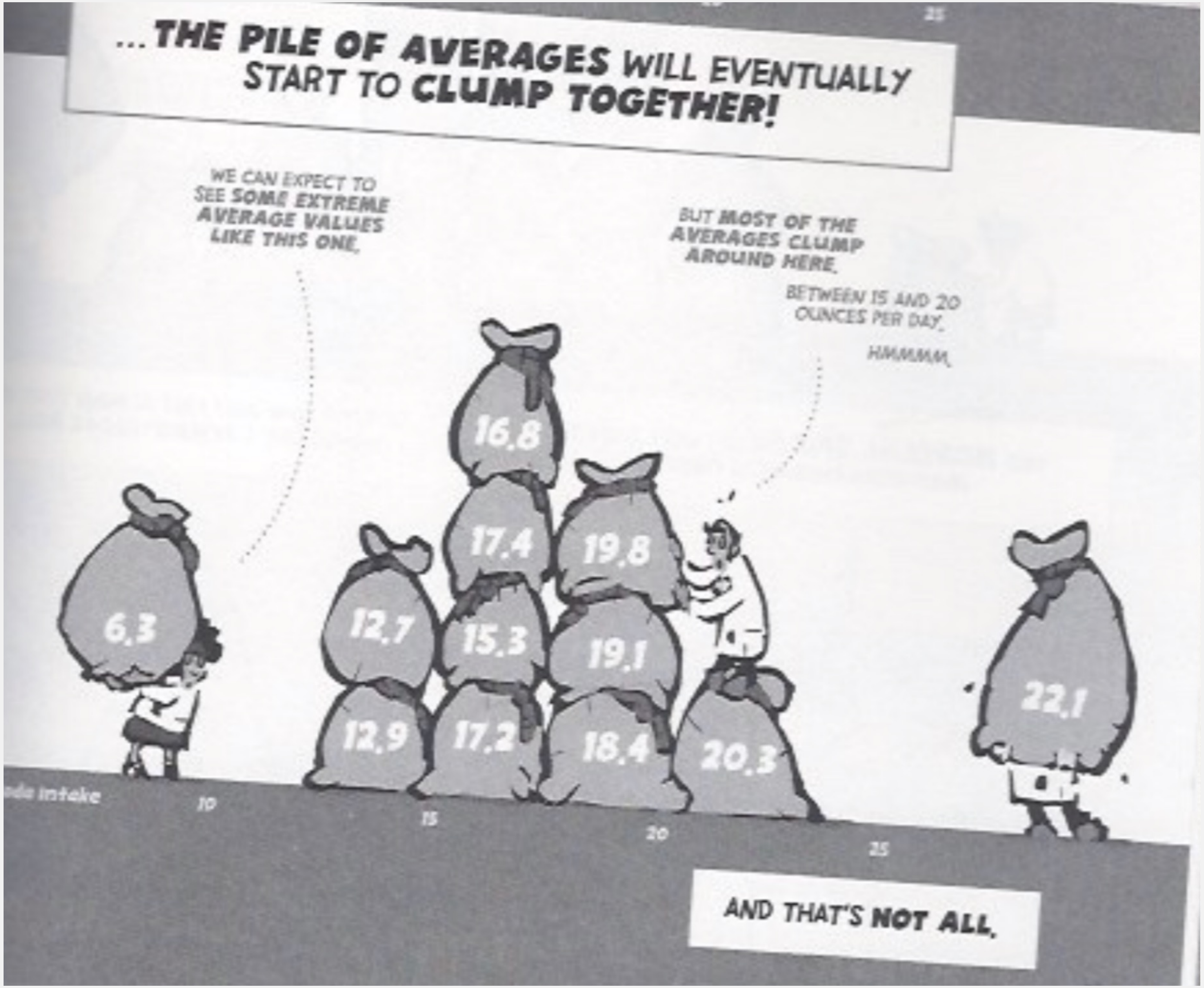


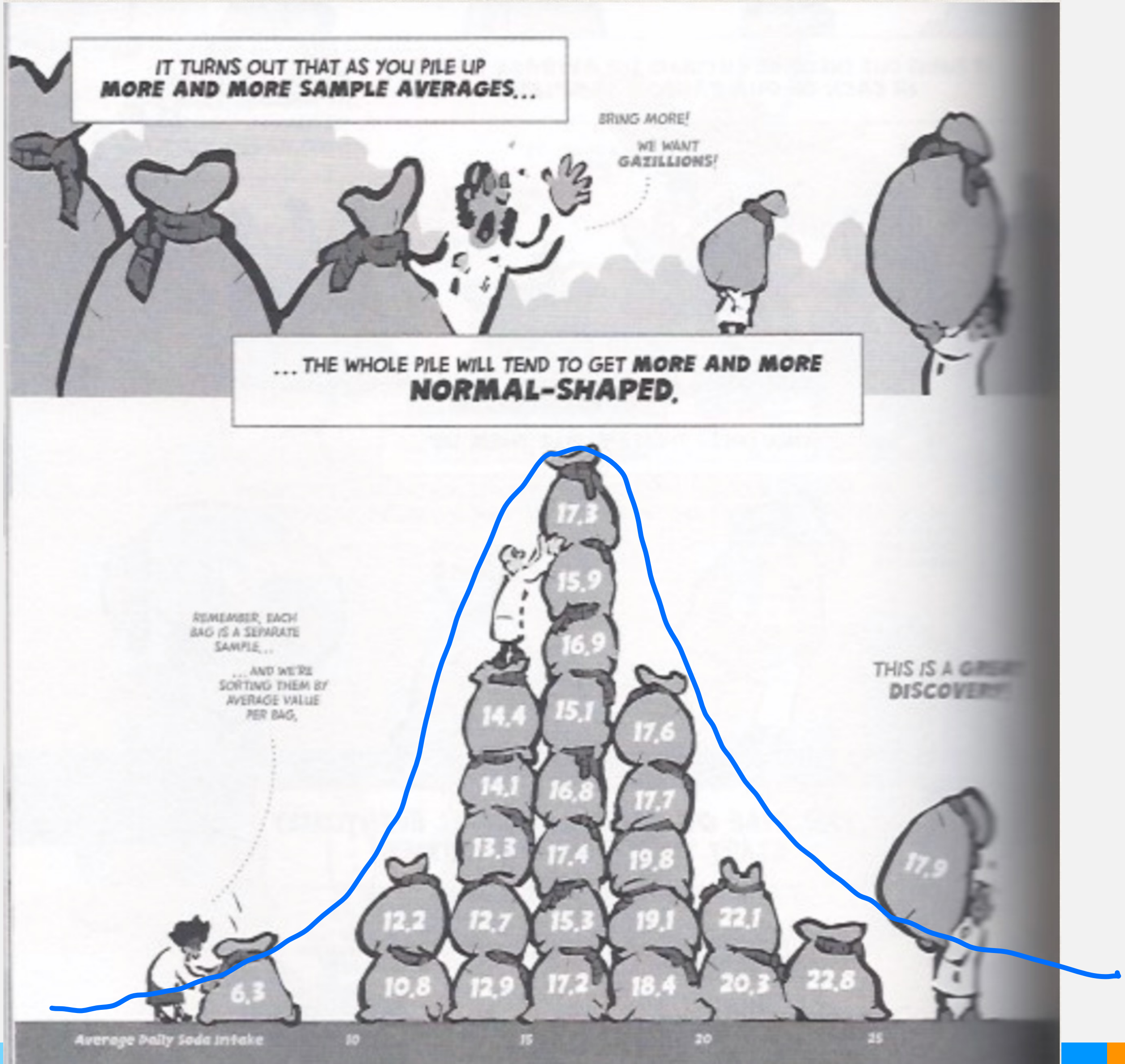

LET'S IMAGINE THAT WE WANT TO KNOW THE AVERAGE VALUE IN A CERTAIN POPULATION.
HOW MUCH SODA DOES EACH AMERICAN DRINK PER DAY?
SLURP.


THEN LET'S IMAGINE WE GO OUT AND GATHER A WHOLE BUNCH OF SEPARATE RANDOM SAMPLES FROM THAT POPULATION.
HERE ARE 50 RANDOM AMERICANS.
HERE ARE 50 OTHER RANDOM AMERICANS.
HERE ARE 50 OTHER RANDOM AMERICANS.
EACH SAMPLE HAS 50 RANDOM AMERICANS IN IT.
WE PUT EACH SAMPLE IN A BAG TO HELP US KEEP TRACK OF THEM.



OUT THAT IF WE CALCULATE THE AVERAGE VALUE IN EACH OF OUR RANDOM SAMPLES...
16.1
IN OUR SAMPLE IT'S 12.9 OUNCES.
HERE, IT'S 18.4 OUNCES.
6.3
17.2
12.9
18.4
6.3
15.3
...THEN ORDER THEM AND PILE THEM UP...
12.9
18.4
20.3
OOF!
10
15
20
25


...THE PILE OF AVERAGES WILL EVENTUALLY START TO CLUMP TOGETHER!
WE CAN EXPECT TO SEE SOME EXTREME AVERAGE VALUES LIKE THIS ONE.
BUT MOST OF THE AVERAGES CLUMP AROUND HERE.
BETWEEN 15 AND 20 OUNCES PER DAY.
HMMMM.
6.3
16.8
17.4
19.8
12.7
15.3
19.1
12.9
17.2
18.4
20.3
22.1
10
15
20
25
AND THAT'S NOT ALL.


IT TURNS OUT THAT AS YOU PILE UP MORE AND MORE SAMPLE AVERAGES...
BRING MORE!
WE WANT GAZILLIONS!
... THE WHOLE PILE WILL TEND TO GET MORE AND MORE NORMAL-SHAPED.
REMEMBER, EACH BAG IS A SEPARATE SAMPLE...
... AND WE'RE SORTING THEM BY AVERAGE VALUE PER BAG.
17.3
15.9
16.9
14.4
15.1
17.6
14.1
16.8
17.7
13.3
17.4
19.8
12.2
12.7
15.3
19.1
22.1
17.9
6.3
10.8
12.9
17.2
18.4
20.3
22.8
Average Daily Soda Intake
10
15
20
25

Description of the Illustrations

- Each bag has 50 Americans in it (random sample of 50 people)
- For each bag, I can compute the average soda consumption for those Americans
- Now each bag has a value attached to it (i.e. data) - it just happens to be a sample mean
- I can treat each value now as a single data point
- Which means I can take the expected value/average of these new data points (sample means for each bag)
- But because we know that when we collect data, we can write it out as a distribution. What I find when I get the average of sample means, is just the mean of the distribution of soda consumption of all Americans
  - This is a **distribution of sample means**!

Sample Mean of Normal Observations

Sample Mean when $X_i \sim N(\mu, \sigma^2)$

Key findings when $X_i \sim N(\mu, \sigma^2)$

1. When measurements are random values that follow a normal distribution, the probability distribution of sample means is also a normal distribution.
2. The mean of the normal probability distribution of the sample means is the same as the mean of the probability distribution of the individual measurements.
3. The standard deviation of the probability distribution of the sample means is smaller than the standard deviation of the probability distribution of the individual observations. If there are n values in the random sample and σ is the standard deviation of the probability distribution of the individual observations, the standard deviation of the probability distribution of the sample means is $\frac{\sigma}{\sqrt{n}}$.

Let’s take this apart step-by-step

1. When measurements are random values that follow a normal distribution, the probability distribution of sample means is also a normal distribution.

- Let $X_1, X_2, \ldots X_n$ be our random sample data, and $\bar{X}$ is the sample mean
- If $X_i \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N(\quad , \quad)$
- This is due to the fact that if you average independent normal variables, it is still normal

Let's take this apart step-by-step

2. The mean of the normal probability distribution of the sample means is the same as the mean of the probability distribution of the individual measurements.

- Let $X_1, X_2, \ldots X_n$ be our random sample data, and $\bar{X}$ is the sample mean
- If $X_i \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N(\boldsymbol{\mu},\quad)$
- This relates to the result that $E(\bar{X}) = E(X) = \mu$

Let’s take this apart step-by-step

3. The standard deviation of the probability distribution of the sample means is smaller than the standard deviation of the probability distribution of the individual observations. If there are n values in the random sample and σ is the standard deviation of the probability distribution of the individual observations, the standard deviation of the probability distribution of the sample means is $\frac{\sigma}{\sqrt{n}}$

- Let $X_1, X_2, \ldots X_n$ be our random sample data, and $\bar{X}$ is the sample mean
- If $X_i \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
- This relates to the result that $Var(\bar{X}) = \frac{Var(X)}{n}$

$Var(x) = \sigma^2$

IQ Example

$n = 5$

IQ measurements are designed to follow a normal distribution with a mean of 100 and a standard deviation of 15. Suppose you collect a random sample 5 individuals and measure their IQ scores. State the probability distribution of the sample mean of these 5 measurements.

$$X_1, X_2, X_3, X_4, X_5 \sim N(100, 15^2)$$

$$\bar{X} = (X_1 + \ldots + X_5)/5$$

$$\bar{X} \sim N\left(100, \frac{15^2}{5}\right)$$

Central Limit Theorem

Not normally distributed

- In the previous slides, we discussed a case when the individual observations were normally distributed
- But what if they are not normally distributed?
- This leads us into a different but similar set of results about the sample mean

Sample Mean (in General)

Key findings:

large enough n

1. **Central Limit Theorem**: *For large enough sample size*, the probability distribution of sample means is approximately a Normal distribution, regardless of the probability distribution of the individual measurements.

$E(\bar{X}) = E(X)$

2. The mean of the normal probability distribution of the sample means is the same as the mean of the probability distribution of the individual measurements.

$Var(\bar{X}) = \frac{Var(X)}{n}$

3. The standard deviation of the probability distribution of the sample means is smaller than the standard deviation of the probability distribution of the individual observations. If there are n values in the random sample and σ is the standard deviation of the probability distribution of the individual observations, the standard deviation of the probability distribution of the sample means is $\frac{\sigma}{\sqrt{n}}$.

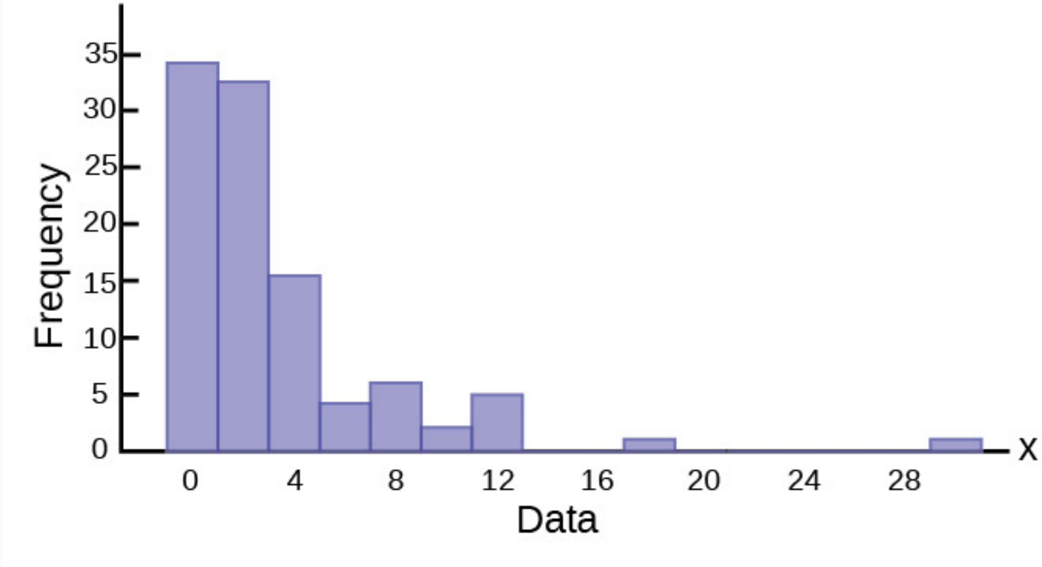
Note: The 2nd and 3rd points remain the same

Central Limit Theorem

- The central limit theorem asserts that: As more and more measurements are taken, the probability distribution for the possible averages of the measurements will converge to a Normal distribution
- Let $X_1, X_2, \ldots X_n$ be our random sample data, and $\bar{X}$ is the sample mean
- CLT states that $\bar{X}$ is approximately $N(\mu, \frac{\sigma^2}{n})$ when n is sufficiently large
- The key is that the sample size n, must be large enough for the distribution of the sample mean to resemble a Normal distribution
- How large is large enough? It depends but $n \geq 30$ is a safe bet

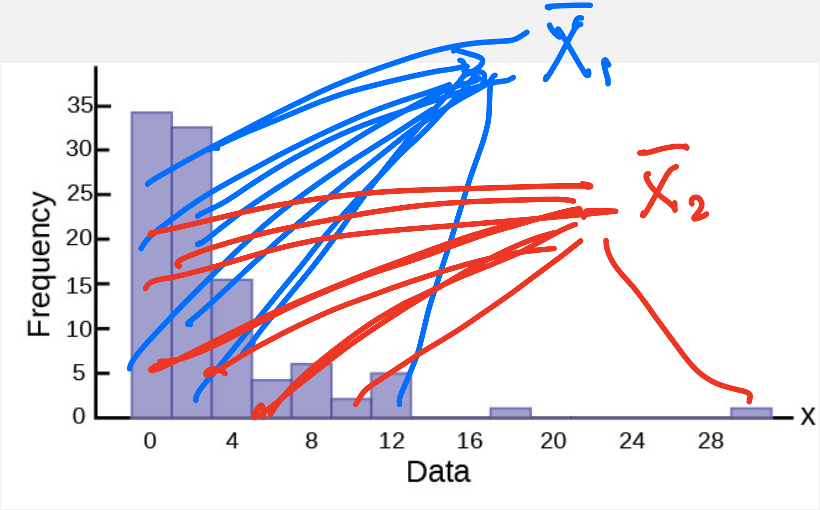
CLT Example

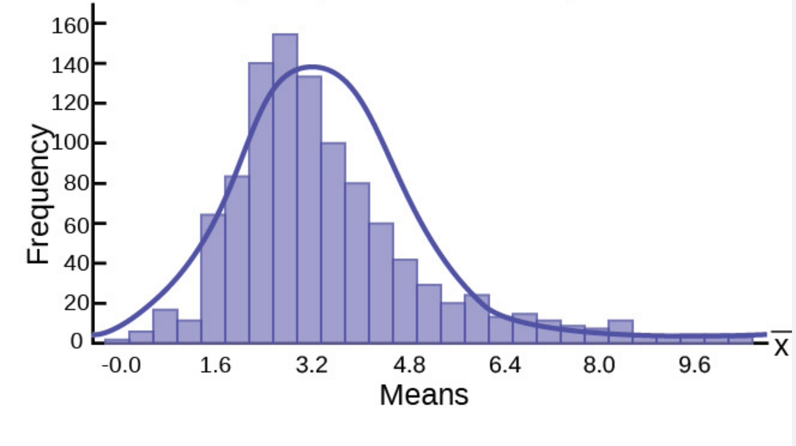
- Consider data on the time between subway arrivals.
- The data is not normal. Most come between 0-4 mins, but some are very delayed



CLT Example

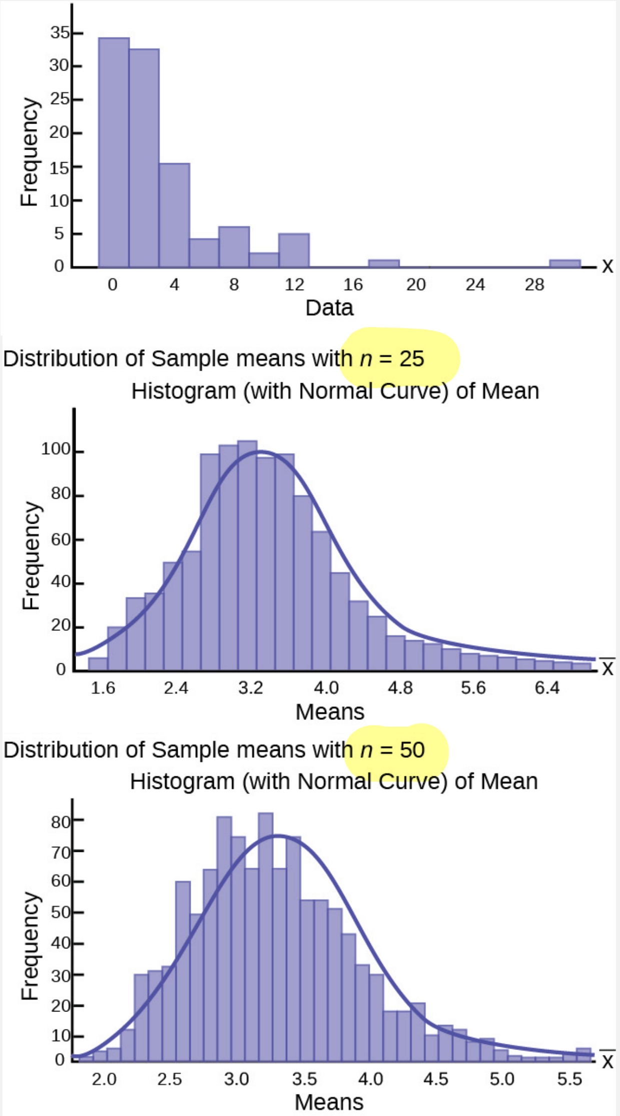
- Suppose we take 10 measurements and we have a sample of size 10. Then we calculate the sample mean.
- Let's repeat this 1000s of times… get a sample of size 10 and calculate the sample mean
- Now, let's examine the distribution of the sample mean by plotting all those sample means in a histogram





CLT Example

- We can do the same thing for a sample size of 25 and a sample size of 50
- The larger the sample size n, the more the distribution of $\bar{X}$ resembles a normal distribution
- Notice that this result holds true, even when the original measurements were not even close to a normal distribution



Example: Laptop lifespans

The lifespan of a laptop is normally distributed with a mean 3.4 years with a standard deviation of 0.8 years. Suppose we collect information on a sample of 25 laptops.

a) What is the distribution for the average lifespan for this sample of 25 laptops?

Let $X_1, \ldots, X_{25}$ be the lifespan of the 25 laptops.
Then, $\bar{X}$ is the average lifespan of the sample of 25

$$X_i \sim N(3.4, 0.8^2) \qquad \bar{X} \sim N\left(3.4, \frac{0.8^2}{25}\right)$$

b) What is the probability that the average lifespan in the sample is more than 4 years?

$$P(\bar{X} > 4) = P\left(\frac{\bar{X} - 3.4}{0.8/5} > \frac{4 - 3.4}{0.8/5}\right) = P(Z > 3.75), \quad Z \sim N(0,1)$$

$$\approx 0$$

Proportions

Proportions are Just Means

- Often when we are running experiments/conducting studies, we may not be interested in the sample average of continuous data.
- Sometimes we may be interested in a summary of an experiment that uses a discrete random variable to represent the possible outcomes, such as a Binomial.
- So what would be the sample average if we are dealing with a Binomial random variable?
  - Binomials are just a collection of Bernoulli's, which have values 1 or 0
  - If we are considering 5 independent Bernoulli trials and want the sample average of the number of 1's, what we get is

$$\frac{0+1+1+0+0}{5}=\frac{2}{5}$$

  - So my sample mean is that 2/5 of my trials were 1's, which is just the proportion of 1's in 5 trials
  - Sample proportion = sample mean of Bernoulli random variables

Distributions of Sample Proportions

- We know that Bernoulli and Binomial random variables are discrete variables
- It turns out though that if we take larger and larger samples of Bernoulli random variables (i.e. look at a Binomial with bigger and bigger n) the distribution of the Binomial starts to look different
    - it starts to look less discrete in nature (smoother curve)
    - and it starts to look symmetric, bell-shaped and unimodal -> Normal
- So even though we have a discrete variable, the Central Limit Theorem still applies, as long as we are taking a large enough number of trials.
- This is because we saw that proportions are just sample averages.

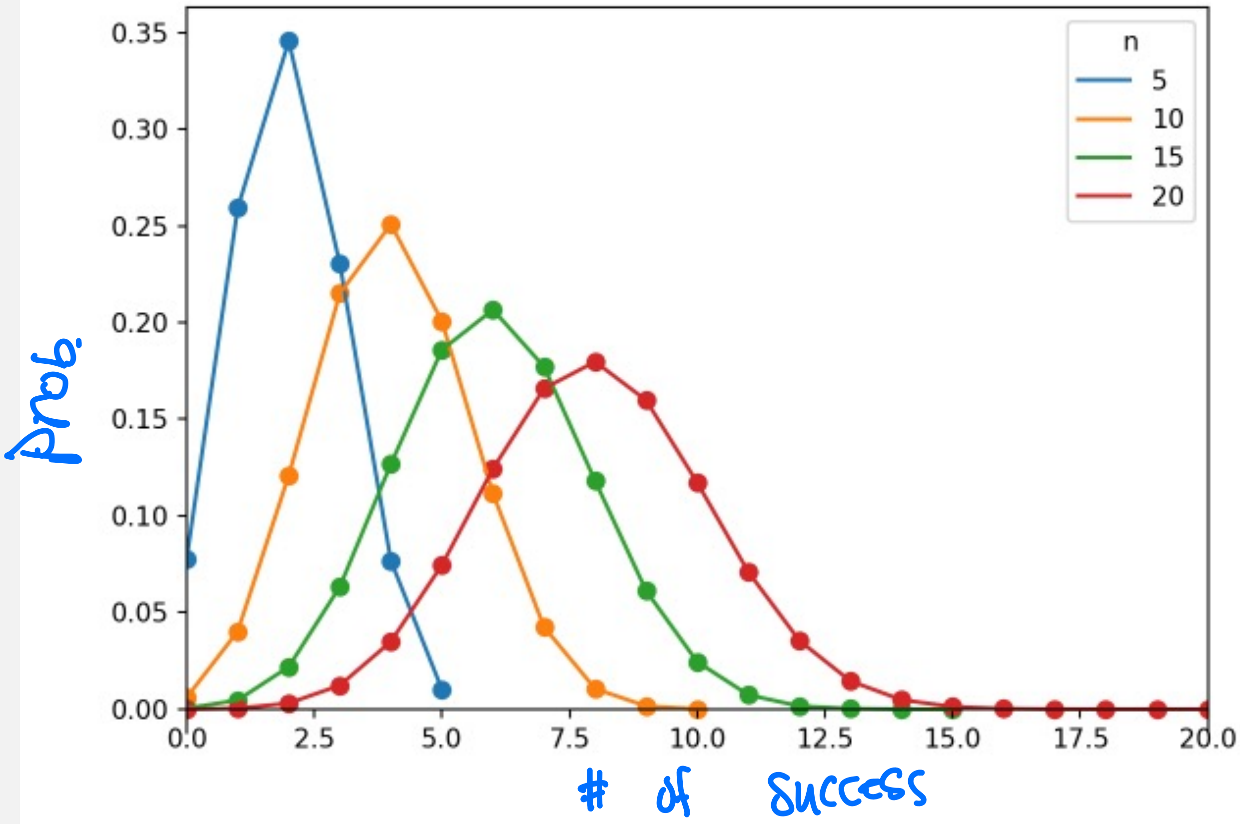


Figure 6.1: Probability mass functions for Binomial(n, 0.4) distributions for $n = 5, 10, 15, 20$.

Binomial Mean and Variance

- In order to figure out what the *distribution of sample proportions* is, we need to know how to refer to the mean and variance of a Binomial random variable
- Recall that we can find the mean/expected value of a discrete variable by

$$E(X) = \sum_{i=1}^{k} x_i P(X = x_i)$$

- When we have a Binomial, we just have a bunch of 0's and 1's, and probabilities P(X = 1) and P(X = 0) (i.e. Bernoulli random variables).
- So since we only have two probabilities, let's call $P(X = 1) = p$ and $P(X = 0) = 1 - p$. We can get the expected value for one Bernoulli by

$$E(X) = 0 \times (1 - p) + 1 \times p = p$$

Binomial Mean and Variance

- Now it's important to remember that a Binomial random variable is just the sum of independent Bernoulli random variables
  - think of one coin flip = Bernoulli; three coin flips = sum of the result on each coin
- So if X is Bernoulli, we found that $E(X) = p$, and now we use one of the helpful properties of expected values,

$$E(X + Y) = E(X) + E(Y)$$

- Since a Binomial with n trials is just summing n Bernoulli's, we can find the mean of the Binomial by

$$E(\underbrace{X_1 + X_2 + \cdots + X_n}_{\text{Binomial}}) = \underbrace{E(X_1)}_{p} + \underbrace{E(X_2)}_{p} + \cdots + \underbrace{E(X_n)}_{p} = p + \cdots + p = np$$

Since $X_i \sim Bern(p)$

Binomial Mean and Variance

- We can basically do the same thing to get the variance of a Binomial distribution
- Start from the Bernoulli and use the definition of Var(X):

$$Var(X) = \sum_i \left(x_i - E(X)\right)^2 P(X = x_i) = (0-p)^2(1-p) + (1-p)^2 p = p(1-p)$$

- So since our Binomial is made up of the sum of independent Bernoulli's, we can use another helpful property of variances: $Var(X + Y) = Var(X) + Var(Y)$
- We now get

$$Var(X_1 + X_2 + \cdots + X_n) = \underbrace{Var(X_1)}_{p(1-p)} + \cdots + Var(X_n) = \overbrace{p(1-p) + \cdots + p(1-p)}^{n \text{ times}} = np(1-p)$$

- So the variance of a Binomial distribution is $np(1-p)$

Binomial Mean and Variance

- In summary, let $X \sim Bin(n, p)$. Then,

$$E(X) = np$$

$$Var(X) = np(1-p)$$

Example: Kittens

In a litter of seven kittens, three are female. You pick two kittens at random with replacement.

a) What is the probability model for the number of male kittens you get?

b) What is the expected number of male kittens?

c) What is the standard deviation for the number of male kittens?

a) Let X be the number of male kittens chosen

$X \sim Bin(2, 4/7)$

b) $E(X) = np = 2 \times \frac{4}{7} = \frac{8}{7}$

c) $Var(X) = np(1-p) = 2 \times \frac{4}{7} \times \frac{3}{7} = 0.49$

$SD(X) = \sqrt{Var(X)} = 0.7$

Back to Distribution of Proportions

- We’ve already discussed how the larger the number of Bernoulli trials I have, the more my probability distribution for the Binomial starts looking a lot like a Normal.
- This is because the Central Limit Theorem can also apply here.
- So we know that
  - for large enough samples, the probability distribution of the Binomial is approximately Normal distributed
- But we need to find out what the mean and standard deviation of this sampling distribution will be, in the case of proportions.

Mean of Sampling Distribution

- The key to this is to realize that a proportion of successes is just a sample average.
- Like before, when we are taking a sample, the data we are collecting are just a bunch of 0's and 1's because we are sampling from a Bernoulli distribution.
- So we know that the number of successes (i.e. the Binomial we want) is just the sum of these Bernoulli random variables: $X_1 + X_2 + \ldots + X_n$
- If we wanted to take the sample average of these, we would divide by how many we have, so $\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \hat{p}$, the proportion of successes
  - The sample proportion $\hat{p}$ is an average of Bernoulli random variables

Mean of Sampling Distribution

- Now we just use the same helpful property as finding the Binomial expected value

sample proportion

- We know that $\hat{p} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$
- If we want to find the expected value, we can just do

$\sim Bern(p)$

$$E(\hat{p}) = \frac{E(X_1+X_2+\cdots+X_n)}{n} = \frac{E(X_1)+E(X_2)+\cdots+E(X_n)}{n} = \frac{np}{n} = p$$

- But we also know what each of those expected values are, since each X_i is a Bernoulli random variable: $E(X_i) = p$
- So we end up adding up n things on the top and dividing by n on the bottom, so we get that $E(\hat{p}) = \text{p}$

Variance of Sampling Distributions

$Var(bX) = b^2 Var(X)$

- Again we basically use the same logic as with the mean.
- We are still trying to find $Var(\hat{p})$ which is an average of independent Bernoulli variables
- We also have the useful result that when we have independent variables, $Var(X+Y) = Var(X) + Var(Y)$
- The variance of a Bernoulli is $p(1-p)$ so we can use this to find the variance of $\hat{p}$

$$Var(\hat{p}) = Var\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{\overbrace{Var(X_1)}^{p(1-p)} + \cdots + Var(X_n)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

- This is analogous to the sample average case, because $p(1-p)$ is the variance of one measurement, and we divide by the sample size to get the variance of the sampling distribution

$Var(\bar{X}) = \frac{Var(X)}{n}$

CLT for Proportions

- So putting all this together, we have that the Central Limit Theorem applies to proportions, as they are just sample averages of Bernoulli variables.
- Therefore, as long as we are taking a large enough sample of Bernoulli measurements, the probabilities corresponding to the average result, the proportion of successes, will converge to a Normal distribution
  - mean of the Normal distribution is the same as that of the individual Bernoulli distribution
  - standard deviation of the Normal distribution is the standard deviation of the individual Bernoulli distribution divided by the square root of the sample size.
  - We can write $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$ is the sampling distribution of a sample proportion

Example: Apples

When a truckload of thousands of apples arrives at a packing plant, a random sample of 150 apples are selected and examined for defects. Suppose that 8% of the apples on the truck have defects.

Find the distribution for the **number** of apples in the sample with defects.

Let X be the number of apples with defects in the sample

$$X \sim Bin(n = 150, \ p = 0.08)$$

Find the distribution for the **proportion** of apples in the sample with defects.

$$\hat{p} \sim N\left(0.08, \ \frac{0.08 \times 0.92}{150}\right), \text{ by CLT since } n \geq 30$$

Example: Apples

The whole truckload of apples will be rejected if more than 5% of the sample have defects. What is the probability that the truckload will be accepted?

$$P(\hat{p} > 0.05) = P\left(\frac{\hat{p} - 0.08}{c} > \frac{0.05 - 0.08}{c}\right)$$

$$= P(Z > -1.35), \quad Z \sim N(0,1)$$

$$= 0.91$$

$$P(\hat{p} < 0.05) = 0.09$$

$$c = \sqrt{\frac{0.08 \times 0.92}{150}}$$

Sampling Distributions

- A **sample statistic** is any function of your data $X_1, X_2, \ldots, X_n$. The sample mean is only one example of a sample statistic
- **Sampling distributions** refer to the probability distribution of a sample statistic (such as sample mean, sample proportion, etc.).
  - The probability distribution accounts for the variability from one sample to another
- The sample distributions will be in terms of **parameters**. Parameters describe the population (such as the mean in the whole population, μ, or the proportion in the whole population, p)

Note: n is not a parameter

Sampling Distributions

- In summary, we have seen two sampling distributions
  - Sample averages: $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
  - Sample proportions: $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$
- These sampling distributions are very important for making inference about certain aspects of a population where it is not possible to measure everyone.
  - Inference is just a fancy word for guessing the value of something, but it also incorporates the idea that there is error and randomness associated with that guess

Sampling Distributions

- Let's take our defective apple example again.
- For the purpose of actually being able to calculate things on a test, you are given the true proportion of apples in the truckload that have defects (p = 0.08).
  - In real life, you don't know this value.
  - In fact, you would be using your sample of 150 apples to estimate the value of p that you don't know
  - So we say that $\hat{p}$ is an estimate for p and we use the sampling distribution to talk about how close this estimate really is to the unknown true proportion.
  - The sampling distribution is therefore the probability distribution of all possible values of the estimator $\hat{p}$
- The variance of the sampling distribution represents the variability of my estimator due to the fact that I took a sample.

How to Work with Sampling Distributions?

- There's really nothing special involved with working with these distributions
- The trick is just realizing from a question that you are dealing with one in the first place.
- Look out for:
  - you are given a mean μ and standard deviation σ, or a proportion p
  - you are told that there is some sort of selection or sampling happening
- These are good indicators that you should use the sampling distribution.
  - Once you have figured out the mean and standard deviation of the sampling distribution, use these to solve whatever probability the question is asking.

Exercise

The weight of potato chips in a medium-size bag is stated to be 300g. The amount that the packaging machine puts in the bags is believed to have a Normal model with mean 306g and standard deviation 3.6g.

a) We want to determine how often the machine is overfilling the bags, so we take a sample of 100 bags. What is the probability that the average of the sampled bags has weight more than 305g?

$n = 100$

Let X be the mass of a bag of chips

$X \sim N(306, 3.6^2)$

$\bar{X} \sim N\left(306, \frac{3.6^2}{100}\right)$

$P(\bar{X} > 305) = P\left(\frac{\bar{X} - 306}{3.6/10} > \frac{305 - 306}{3.6/10}\right) = P(Z > -2.78)$

$= 0.9973$

Exercise Continued

b) Based on the previous exercise: Some of the bags of chips are sold in packs of 5. What is the expected number of bags in a pack that have been filled with more than 305g of chips? (Assume the bags chosen to be in a pack are selected at random)

What is the probability that single bag of chips has more than 305 g?

$$P(X > 305) = P\left(\frac{X-306}{3.6} > \frac{305-306}{3.6}\right) = P(Z > -0.278)$$

$$= 0.6103$$

Let Y be the number of bags within a pack of 5 that has more than 305g.

$Y \sim Bin(n=5,\ p=0.6103)$

$E(Y) = np = 5 \times 0.6103 = 3.0515$

Exercise Continued

c) What is the distribution for the proportion of bags with weight more than 305g in a sample of 100 bags?

$\hat{p}$ is the proportion of bags (out of 100) that have more than 305g.

$\hat{p} \sim N\left(0.6103, \dfrac{0.6103(1-0.6103)}{100}\right)$ by CLT since $n \geq 30$

Videos and Practice Problems

- In this lecture, we covered all of Module 4: Sampling Distributions
  - Note: The video lectures use the notation $N(\mu, \sigma)$ where the second parameter is the standard deviation σ. In this course, we use the notation $N(\mu, \sigma^2)$ where the second parameter is the variance σ^2.

Next Week

- In-person term test is in two weeks.
    - Official announcement will come soon
    - Go over lecture notes and examples
    - Do practice problems
    - Ask questions