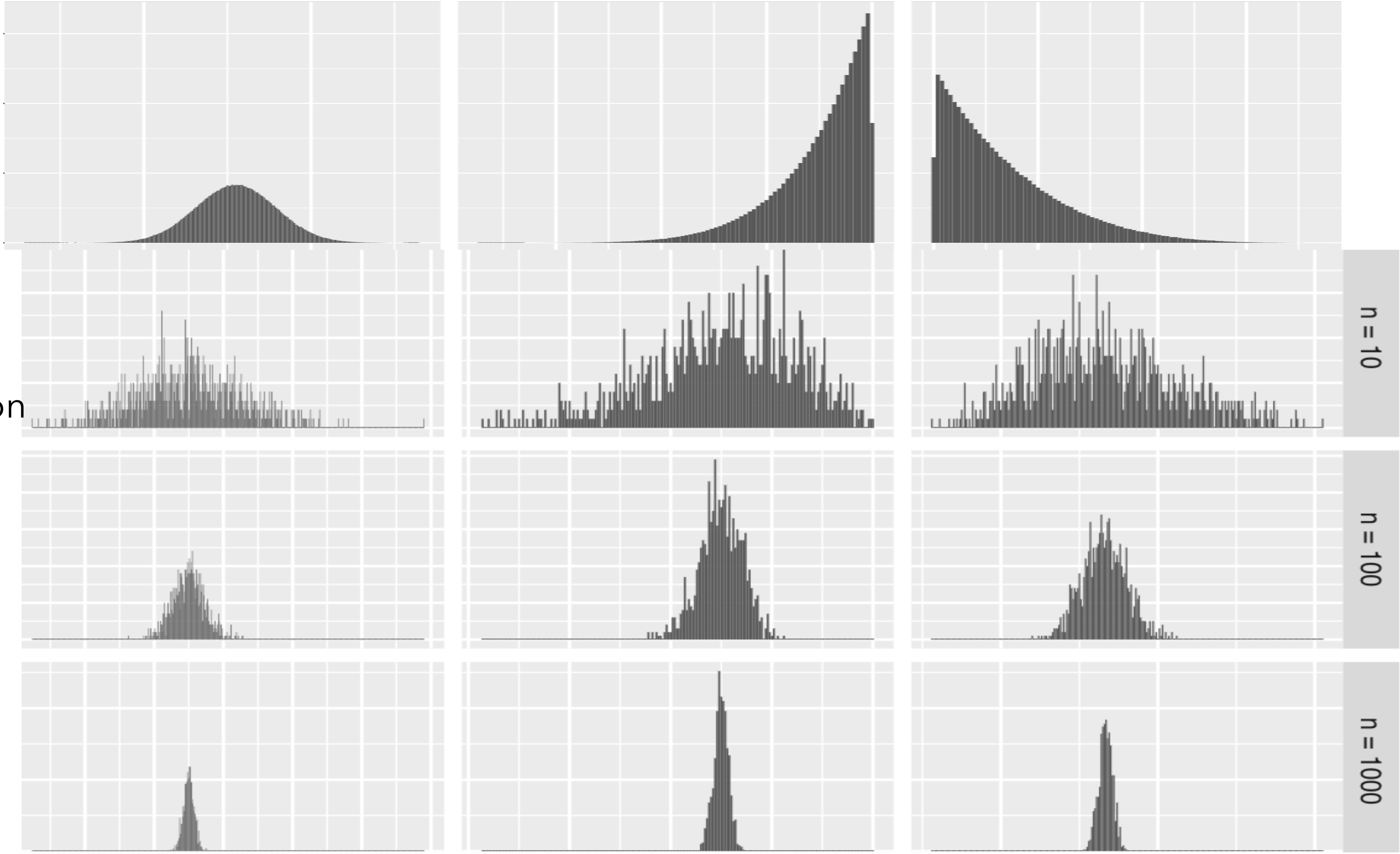
Stat 88: Prob. & Mathematical Statistics in Data Science

Prob dsn of X_k

Distribution of $\frac{S_n}{n}$



Lecture 20 PART 1: 4/5/2022

Finishing up chapter 8 and the Central Limit Theorem

The Central Limit Theorem

- Suppose that $X_1, X_2, \ldots, X_n$ are iid with mean μ and SD σ
- Let $S_n = X_1 + \cdots + X_n$ be the sample sum, and $A_n = \frac{S_n}{n}$ be the sample mean

- Then the distribution of S_n (and A_n) is *approximately normal* for large enough n.

- For S_n, the distribution is approximately normal (bell-shaped), centered at $\boldsymbol{E(S_n) = n\mu}$ and with spread given by by $\boldsymbol{SD(S_n) = \sqrt{n}\,\sigma.}$

- For A_n, the distribution is centered at $\boldsymbol{E(A_n) = \mu}$ with spread $\boldsymbol{SD(A_n) = \sigma/\sqrt{n}}$

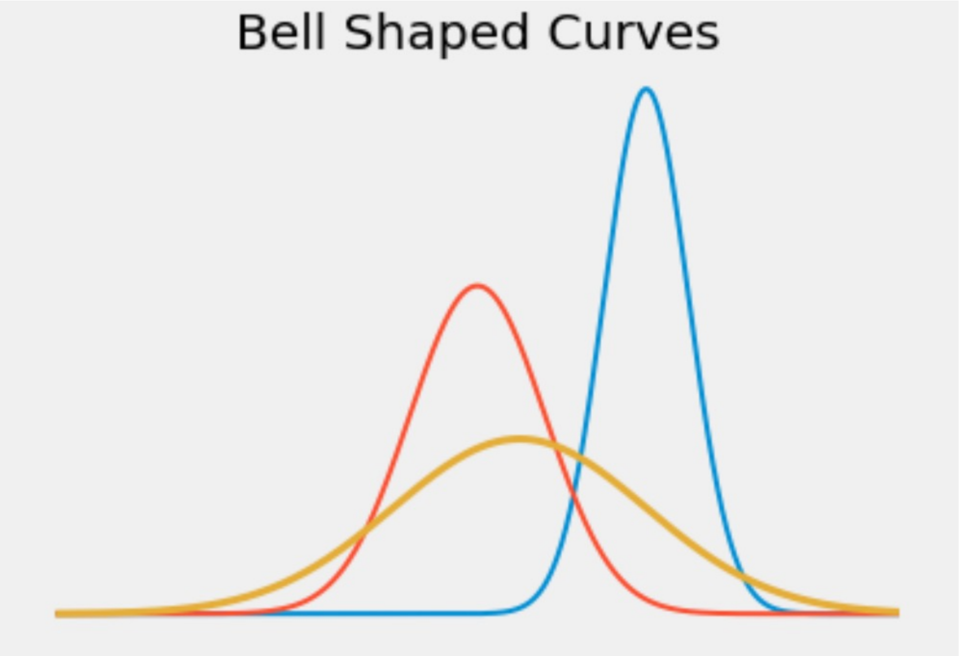
Standard normal cdf

- $\Phi(z) = \int_{-\infty}^{z} \phi(x)dx$, where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}, -\infty < x < \infty$

Normal approximations : standard units

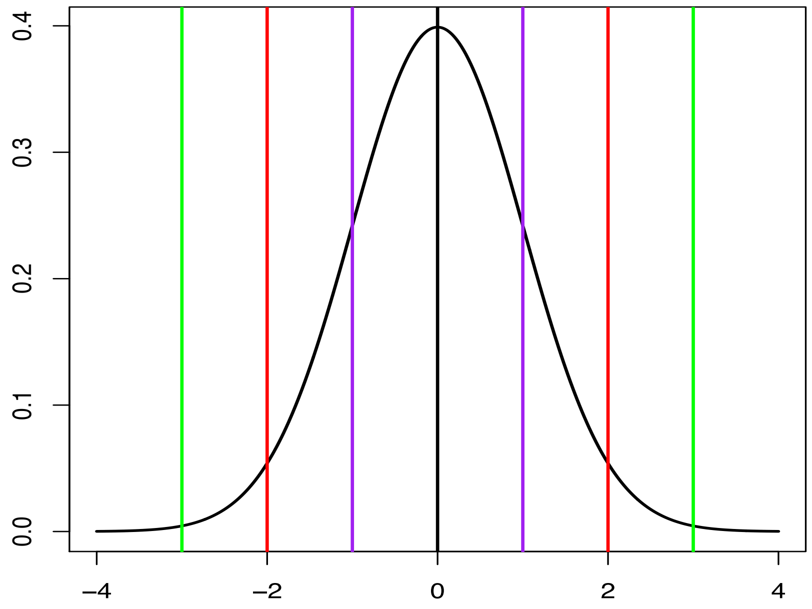
- Let X be any random variable, with expectation μ and SD σ, consider a new random variable that is a linear function of X, created by shifting X to be centered at 0, and dividing by the SD. If we call this new rv X^*, then X^* has expectation ______ and SD ______.
- $E(X^*) =$
- $SD(X^*) =$
- This new rv does not have units since it measures how far above or below the average a value is, in SD's. Now we can compare things that we may not have been able to compare.
- Because we can convert anything to standard units, ***every normal curve is the same.***

The many normal curves → the *standard normal* curve



- Just one normal curve, standard normal, centered at 0. All the rest can be derived from this one.

How do we approximate the area of a range in a histogram?



Total area under the curve = 100%
Curve is symmetric about 0
The areas between 1, 2 and 3 SDs away:
Between -1 and 1 the area is 68.27%
Between -2 and 2 the area is 95.45%
Between -3 and 3 the area is 99.73%

- Many histograms bell-shaped, but not on the same scale, and not centered at 0
- Need to convert a value to ***standard units*** – see how many SDs it is above or below the average
- Then we can **approximate** the area of the histogram using the area under the standard normal curve (using for example, stats.norm.cdf for the actual numerical computation)
- 68%-95%-99.7% rule

  (**Empirical rule**)

How to decide if a distribution could be normal

- Need enough SDs on both sides of the mean.

- In 2005, the murder rates (per 100,000 residents) for 50 states and D.C. had a mean of 8.7 and an SD of 10.7. Do these data follow a normal curve?

- If you have indicators, then you are approximating binomial probabilities. In this case, if *n* is very large, but *p* is small, so that *np* is close to 0, then you can't have many sds on the left of the mean. So need to increase *n,* stretching out the distribution and the n the normal curve begins to appear.

- If you are not dealing with indicators, then might bootstrap the distribution of the sample mean and see if it looks approximately normal.

Stat 88: Prob. & Mathematical Statistics in Data Science

I CAN'T BELIEVE SCHOOLS ARE STILL TEACHING KIDS ABOUT THE NULL HYPOTHESIS.
I REMEMBER READING A BIG STUDY THAT CONCLUSIVELY DISPROVED IT *YEARS* AGO.

https://xkcd.com/892/

Lecture 20 PART 2: 4/5/2022

Section 9.1, 9.2

Hypothesis tests & A/B testing

Hypothesis tests

- Hypothesis tests or *tests of significance* are tests in which we use data to draw conclusions, or *make inferences* about the process that generated the data, or the population from which we drew the sample.
- Underlying idea: Observed values can't be too far from the expected value, *if our assumptions are correct.*
- What if they are not correct? ***How far is too far***?

- Toss a coin 100 times. See 54 heads. Do you have reason to believe that the coin is not fair?
- What about 60 heads? 66 heads?
- How would we decide?

- Let's look at a couple of examples, and review the ideas of hypothesis testing

Example: Gender discrimination?

A large supermarket chain occasionally selects employees to receive management training. A group of women there claimed that female employees were passed over for this training in favor of their male colleagues. The company denied this claim.

Suppose that the large employee pool of the Florida chain that can be tapped for management training is half male and half female.

Since this program began, *none* of the 10 employees chosen have been female. What would be the probability of 0 out of 10 selections being female, if there truly was no gender bias?

Example: Woburn

In the early 1990s, a leukemia cluster was identified in the Massachusetts town of Woburn. Many more cases of leukemia, a malignant cancer that originates in a bone marrow cell, appeared in this small town than would be predicted. Was this evidence of a problem in the town or just chance?

Hypotheses tests: Review of steps

1. State the ***null*** hypothesis (H_0) – that is, what is the assumption we are going to make. This will determine how we compute probabilities

2. State an ***alternative*** hypothesis (H_A) Note that this should not overlap with the null hypothesis, and it may or may not define probabilities (example: there was gender bias etc)

3. Decide on a test statistic to use that will help you decide which of the two hypotheses is supported by the evidence (data). Usually there is a natural choice. Use the null hypothesis to specify probabilities for the test statistic.

4. Find the observed value of the test statistic, and see if it is ***consistent*** with the null hypothesis. That is, compute the chance that we would see such an observed value, or more extreme values of the statistic (*p-value*).

5. State your conclusion: whether you reject the null hypothesis or not. This is based on your chosen cutoff ("level of the test"). Reject if the p-value is less than the cutoff.

Observed significance levels (a.k.a p-values)

- The p*-value* decides if observed values are *consistent* with the null hypothesis. It is a *tail* probability (also called *observed significance level)*, and is the chance, **assuming that the null hypothesis is true**, of getting a test statistic equal to the one that was observed or even more in the direction of the alternative.
- If this probability is too small, then something is wrong, perhaps with your assumption (null hypothesis). That is, the data are unlikely if the null is true and therefore, your data are ***inconsistent*** with the null hypothesis.
- p-value is **not** the chance of null being true. The null is either true or not.
- The p-value is a *conditional probability* since it is computed *assuming* that the null hypothesis is true.
- The smaller the p-value, the stronger the evidence ***against*** the null and ***towards*** the alternative (in the direction of the alternative).
- Traditionally, below 5% ("result is statistically significant") and 1% ("result is highly significant") are what have been used. Significant means the p-value is small, not that the result is important.

Ex. 9.5.1

- All the patients at a doctor's office come in annually for a check-up when they are not ill. The temperatures of the patients at these check-ups are independent and identically distributed with unknown mean μ.
- The temperatures recorded in 100 check-ups have an average of 98.2 degrees and an SD of 1.5 degrees. Do these data support the hypothesis that the unknown mean μ is 98.6 degrees, commonly known as "normal" body temperature? Or do they indicate that μ is less than 98.6 degrees?

A/B testing: comparing *two* distributions

- Data 8, section 12.3, randomized controlled trial to see if botulinum toxin could help manage chronic pain.
- 31 patients → 15 in treatment group, 16 in control group. 2 patients in the control group reported pain relief and 9 in the treatment group.

- A/B testing is a term used to describe hypothesis tests which involve comparing the distributions of *two* random samples. (Earlier we had *one* sample and made a hypothesis about its distribution.)
- In particular, we can conduct an A/B test for hypothesis tests involving results of randomized controlled trials, A is the control group and B the treatment group.

Fisher’s exact test

- Control group: 16 patients, 2 reported relief
- Treatment group: 15 patients, 9 reported relief

- H_0: The treatment has no effect (there would have been 11 patients reporting pain relief no matter what, and it just so happens that 9 of them were in the treatment group)
- H_A: The treatment has an effect

Example:The Lady Tasting Tea

- The first person to describe this sort of hypothesis test was the famous British statistician Ronald Fisher. In his book *The Design of Experiments,* he describes a tea party in which a lady of his acquaintance claimed that she could tell from tasting a cup of tea if the milk had been poured first or the tea.
- Fisher immediately set up an experiment in which she was given multiple cups of tea and asked to identify which of them had had the tea poured first. She tasted 8 cups of tea, of which 4 had the tea poured first, and identified 3 of them correctly. Does this data support her claim?

Example: Gender bias ?

- Rosen and Jerdee conducted several experiments using male bank supervisors (this was in 1974) who were given a personnel file and asked to decide whether to promote or hold the file. 24 were randomly assigned to a file labeled as that of a male employee and 24 to a female.
- 21 of the 24 males were promoted, and 14 of the females. Is there evidence of gender bias?