# Recitation, Week 6

Ye Wang

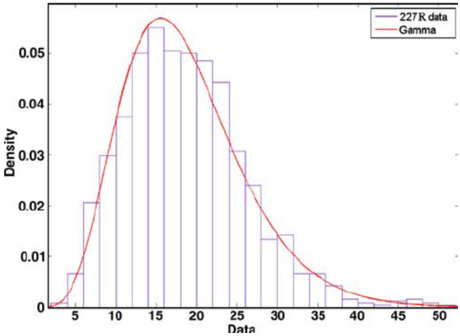
New York University

POL-850

Spring 2018

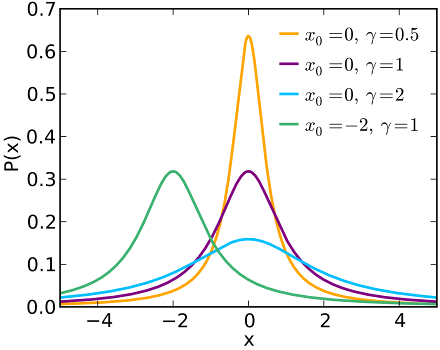
# Outline

# Probability Distribution




- Probability Distribution can be seen as the limit of a histogram
- X-axis: the index of events; Y-axis: the probability for the event to happen
- Imagine that we can measure X as precisely as possible...

# Probability Distribution




Properties of Probability Distribution

1. Y-value of the curve is always greater than or equal to 0
   $y = f(x) \geq 0$
2. The area underneath the curve is always 1
   $\int f(x)dx = 1$ (what does this integral mean?)

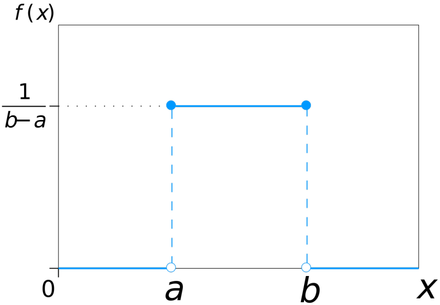
# Uniform Distribution



- Suppose you're waiting for a bus and you know that the bus comes every 5 minutes (but you don't know when the last bus passed the bust stop)
- X: the waiting time for the next bus
- X obeys uniform distribution ($X \sim U[0, 5]$)
- From now on, the probability for the bus to arrive is the same in any given time interval (e.g. one minute)

# Group exercise 1

- What is the probability that a bus comes between 0 and 5 minutes?
- Probability that a bus comes after 3 minutes?
- Probability that a bus comes between 2 minutes and 4 minutes?
- Why the value of the curve below takes always should take $\frac{1}{b-a}$?



# Standard Normal Distribution



We learned normal and standard normal distribution in class. But,

1. Why do we care about a normal distribution?
2. Why do we standardize using Z-score? (i.e. Why do we use the standard normal dist.?)

$$Z = \frac{X-\mu}{\sigma}$$

3. Why do we have the Z-table?

Because..

1. It is the most common distribution in real world and also in "statistical world" (due to the central limit theorem)
2. Normal distributions may have different means and standard deviation. By standardizing, we can convert all them into one distribution.
3. It helps us to calculate the area underneath the curve (i.e. to calculate a probability).

Simply put, we want to calculate a probability!

# Normal Distribution

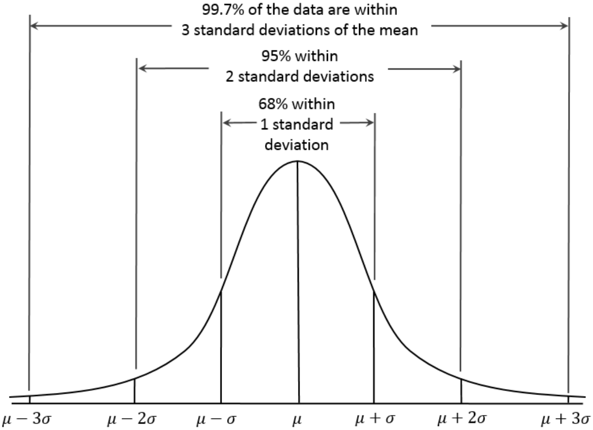
- Suppose you're studying obesity and interested in average weight of American men

# Normal Distribution

- Suppose you're studying obesity and interested in average weight of American men
- X: weight of an male adult in the US

# Normal Distribution

- Suppose you're studying obesity and interested in average weight of American men
- X: weight of an male adult in the US
- You know that X is normally distributed



# Normal Distribution

- Suppose the mean is 190 pounds and the standard deviation is 20 pounds

# Normal Distribution

- Suppose the mean is 190 pounds and the standard deviation is 20 pounds
- Then, we write $X \sim N(190, 20^2)$

# Normal Distribution

- Suppose the mean is 190 pounds and the standard deviation is 20 pounds
- Then, we write $X \sim N(190, 20^2)$

Properties of the Normal Distribution

# Normal Distribution

- Suppose the mean is 190 pounds and the standard deviation is 20 pounds
- Then, we write $X \sim N(190, 20^2)$

Properties of the Normal Distribution

- Symmetric

# Normal Distribution

- Suppose the mean is 190 pounds and the standard deviation is 20 pounds
- Then, we write $X \sim N(190, 20^2)$

Properties of the Normal Distribution

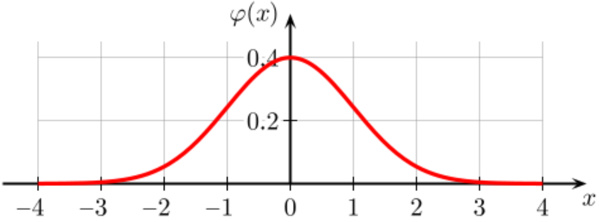
- Symmetric
- Unimodal $\rightarrow$ Mean $=$ Median $=$ Mode

# Normal Distribution

- Suppose the mean is 190 pounds and the standard deviation is 20 pounds
- Then, we write $X \sim N(190, 20^2)$

Properties of the Normal Distribution

- Symmetric
- Unimodal $\rightarrow$ Mean $=$ Median $=$ Mode



# Standard Normal Distribution

- This is a special case of the Normal Distribution

# Standard Normal Distribution

- This is a special case of the Normal Distribution
- Its mean is 0 and its standard deviation is 1

# Standard Normal Distribution

- This is a special case of the Normal Distribution
- Its mean is 0 and its standard deviation is 1
- $X \sim N(0, 1^2)$

# Standard Normal Distribution



- This is a special case of the Normal Distribution
- Its mean is 0 and its standard deviation is 1
- $X \sim N(0, 1^2)$
- When we calculate the z-score, we actually convert the weight normal distribution into the Standard Normal Distribution!

# Standard Normal Distribution



- This is a special case of the Normal Distribution
- Its mean is 0 and its standard deviation is 1
- $X \sim N(0, 1^2)$
- When we calculate the z-score, we actually convert the weight normal distribution into the Standard Normal Distribution!

$$Z = \frac{X - \mu}{\sigma}$$

# Example

- Suppose you want to get the probability that a man's weight is greater than 210 pounds (when randomly drawn)?

# Example

- Suppose you want to get the probability that a man's weight is greater than 210 pounds (when randomly drawn)?

$$Z = \frac{X-\mu}{\sigma} = \frac{210-190}{20} = 1$$

# Example

- Suppose you want to get the probability that a man's weight is greater than 210 pounds (when randomly drawn)?
$$Z = \frac{X - \mu}{\sigma} = \frac{210 - 190}{20} = 1$$
- Looking at the Z-table returns $Pr(Z \geq 1) = 0.16$

# Group exercise 2

1. Visualize the previous result ($Pr(Z \geq 1) = 0.16$) using the graph of Standard Normal Distribution.
2. How to calculate the probability that a man randomly drawn from the population weighs greater than 180 pounds?
3. How to calculate the probability that a man randomly drawn from the population weighs greater than 170 and smaller than 200 pounds?

# Why?

- Why do we study sampling distribution and central limit theorem?

# Why?

- Why do we study sampling distribution and central limit theorem?
- That's because they are the basis of statistical inference

# Why?

- Why do we study sampling distribution and central limit theorem?
- That's because they are the basis of statistical inference
- What was the goal of statistical inference?

# Why?

- Why do we study sampling distribution and central limit theorem?
- That's because they are the basis of statistical inference
- What was the goal of statistical inference?
- It's about finding a population parameter (population mean $\mu$)

# Why?

- Why do we study sampling distribution and central limit theorem?
- That's because they are the basis of statistical inference
- What was the goal of statistical inference?
- It's about finding a population parameter (population mean $\mu$)
- How do we get it without census?

# Why?

- Why do we study sampling distribution and central limit theorem?
- That's because they are the basis of statistical inference
- What was the goal of statistical inference?
- It's about finding a population parameter (population mean $\mu$)
- How do we get it without census?
- We randomly draw a sample from population

# Why?

- Why do we study sampling distribution and central limit theorem?
- That's because they are the basis of statistical inference
- What was the goal of statistical inference?
- It's about finding a population parameter (population mean $\mu$)
- How do we get it without census?
- We randomly draw a sample from population
- And then we make a guess about the population parameter based on the sample

# Why?

- But how accurate is our guess?

# Why?

- But how accurate is our guess?
- We need confidence interval, an estimate of the guess' uncertainty

# Why?



- But how accurate is our guess?
- We need confidence interval, an estimate of the guess' uncertainty
- It is an interval estimate something like this: [5 feet 4 inches, 5 feet 9 inches] for average height of US adults

# Why?

- But how accurate is our guess?
- We need confidence interval, an estimate of the guess' uncertainty
- It is an interval estimate something like this: [5 feet 4 inches, 5 feet 9 inches] for average height of US adults
- To understand the origin of confidence interval, we need to make sense of sampling distribution and central limit theorem

# Why?

- But how accurate is our guess?
- We need confidence interval, an estimate of the guess' uncertainty
- It is an interval estimate something like this: [5 feet 4 inches, 5 feet 9 inches] for average height of US adults
- To understand the origin of confidence interval, we need to make sense of sampling distribution and central limit theorem
- That's because they form the basis of confidence interval

# Animation

A cool animation about rabbits' weight and dragons' wing span
(Length: about 4 minutes)

# Sampling distribution

- What is the sampling distribution?

# Sampling distribution

- What is the sampling distribution?
- It's a kind of a probability distribution. But it's probability distribution of a sample statistic (e.g. sample mean $\overline{X}$).

# Sampling distribution

- What is the sampling distribution?
- It's a kind of a probability distribution. But it's probability distribution of a sample statistic (e.g. sample mean $\overline{X}$).
- They are so special that we call them "sampling distribution"

# Sampling distribution

- What is the sampling distribution?
- It's a kind of a probability distribution. But it's probability distribution of a sample statistic (e.g. sample mean $\overline{X}$).
- They are so special that we call them "sampling distribution"
- Therefore we can plot the sampling distribution of sample mean by:

# Sampling distribution

- What is the sampling distribution?
- It's a kind of a probability distribution. But it's probability distribution of a sample statistic (e.g. sample mean $\overline{X}$).
- They are so special that we call them "sampling distribution"
- Therefore we can plot the sampling distribution of sample mean by:
- Randomly draw a sample $\rightarrow$ Calculate the sample mean $\rightarrow$ Repeat this process many times $\rightarrow$ Plot its distribution (or the histogram)

# Central Limit Theorem

- This is arguably the most important theorem in probability theory

# Central Limit Theorem

- This is arguably the most important theorem in probability theory
- Basically, it's saying that as the size of sample increases, the sampling distribution looks more and more like a normal distribution (and narrower)

# Central Limit Theorem

- This is arguably the most important theorem in probability theory
- Basically, it's saying that as the size of sample increases, the sampling distribution looks more and more like a normal distribution (and narrower)

# Central Limit Theorem

- What's fantastic about this theorem is that this holds regardless of the population distribution!

# Central Limit Theorem

- What's fantastic about this theorem is that this holds regardless of the population distribution!
- It implies that you don't need to care about the population distribution. You just form a sampling distribution with large sample size!

# Central Limit Theorem

- What's fantastic about this theorem is that this holds regardless of the population distribution!
- It implies that you don't need to care about the population distribution. You just form a sampling distribution with large sample size!
- That is enough for guessing a population mean (by calculating a confidence interval)