Math Camp

Justin Grimmer

Associate Professor
Department of Political Science
Stanford University

September 14th, 2015
Questions?

1) What is a random variable? Where does the randomness in the random variable come from?

2) What is the pmf? How would we derive it?

3) What does iid mean?

4) Define $E[X]$, var($X$)

5) What does it mean for a random variable, $Y \sim \text{Poisson}(\lambda)$?
Questions?

1) What is a random variable? Where does the randomness in the random variable come from?
Questions?

1) What is a random variable? Where does the randomness in the random variable come from?

2) What is the pmf? How would we derive it?
Questions?

1) What is a random variable? Where does the randomness in the random variable come from?
2) What is the pmf? How would we derive it?
3) What does iid mean?
Questions?

1) What is a random variable? Where does the randomness in the random variable come from?
2) What is the pmf? How would we derive it?
3) What does iid mean?
4) Define $E[X]$, $\text{var}(X)$
Questions?

1) What is a random variable? Where does the randomness in the random variable come from?
2) What is the pmf? How would we derive it?
3) What does iid mean?
4) Define $E[X], \text{var}(X)$
5) What does it mean for a random variable, $Y \sim \text{Poisson}(\lambda)$?
Where We’ve Been, Where We’re Going

Continuous Random Variables:
- Random variables that are not discrete
- Widely used:
  - Approval ratings
  - Vote Share
  - GDP
  - ...
- Many analogues to distributions used yesterday
Continuous Random Variables

- Wait time between wars: $X(t) = t$ for all $t$
- Proportion of vote received: $X(v) = v$ for all $v$
- Stock price: $X(p) = p$ for all $p$
- Stock price, squared: $Y(p) = p^2$ for all $p$

We'll need calculus to answer questions about probability.
Continuous Random Variables

Continuous Random Variables:

- Wait time between wars: \( X(t) = t \) for all \( t \)
- Proportion of vote received: \( X(v) = v \) for all \( v \)
- Stock price: \( X(p) = p \) for all \( p \)
- Stock price, squared: \( Y(p) = p^2 \) for all \( p \)

We'll need calculus to answer questions about probability.
Continuous Random Variables

Continuous Random Variables:

- Wait time between wars: \( X(t) = t \) for all \( t \)

We'll need calculus to answer questions about probability.
Continuous Random Variables

Continuous Random Variables:
- Wait time between wars: $X(t) = t$ for all $t$
- Proportion of vote received: $X(v) = v$ for all $v$
Continuous Random Variables

Continuous Random Variables:

- Wait time between wars: \( X(t) = t \) for all \( t \)
- Proportion of vote received: \( X(v) = v \) for all \( v \)
- Stock price \( X(p) = p \) for all \( p \)
Continuous Random Variables

- Wait time between wars: \( X(t) = t \) for all \( t \)
- Proportion of vote received: \( X(v) = v \) for all \( v \)
- Stock price \( X(p) = p \) for all \( p \)
- Stock price, squared \( Y(p) = p^2 \) for all \( p \)

We'll need calculus to answer questions about probability.
Continuous Random Variables

Continuous Random Variables:
- Wait time between wars: $X(t) = t$ for all $t$
- Proportion of vote received: $X(v) = v$ for all $v$
- Stock price $X(p) = p$ for all $p$
- Stock price, squared $Y(p) = p^2$ for all $p$

We’ll need calculus to answer questions about probability.
Integration

Suppose we have some function $f(x)$.
Integration

Suppose we have some function $f(x)$

What is the area under $f(x)$ between $\frac{1}{2}$ and 1?
Integration

Suppose we have some function $f(x)$

What is the area under $f(x)$ between $\frac{1}{2}$ and 1?

Area under curve $= \int_{1/2}^{1} f(x) \, dx = F(1) - F(1/2)$
Continuous Random Variable

Definition

$X$ is a continuous random variable if there exists a nonnegative function defined for all $x \in \mathbb{R}$ having the property for any (measurable) set of real numbers $B$,

$$P(X \in B) = \int_B f(x) \, dx$$

We’ll call $f(\cdot)$ the probability density function for $X$. 
Example: Uniform Random Variable

$X \sim \text{Uniform}(0, 1)$ if

$$f(x) = \begin{cases} 1 & \text{if } x \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$
Example: Uniform Random Variable

\[ X \sim \text{Uniform}(0, 1) \text{ if } \]

\[ f(x) = 1 \text{ if } x \in [0, 1] \]
Example: Uniform Random Variable

$X \sim \text{Uniform}(0, 1)$ if

\[
f(x) = \begin{cases} 
1 & \text{if } x \in [0, 1] \\
0 & \text{otherwise}
\end{cases}
\]
Example: Uniform Random Variable

\[ X \sim \text{Uniform}(0, 1) \text{ if} \]

\[ f(x) = 1 \text{ if } x \in [0, 1] \]

\[ f(x) = 0 \text{ otherwise} \]

\[
P(X \in [0.2, 0.5]) = \int_{0.2}^{0.5} 1\,dx
= X|_{0.2}^{0.5}
= 0.5 - 0.2
= 0.3
\]
Example: Uniform Random Variable

\[ X \sim \text{Uniform}(0, 1) \text{ if} \]

\[
f(x) = \begin{cases} 
1 & \text{if } x \in [0, 1] \\
0 & \text{otherwise}
\end{cases}
\]

\[
P(X \in [0, 1]) = \int_{0}^{1} 1 \, dx = X|_{0}^{1} = 1 - 0 = 1
\]
Example: Uniform Random Variable

$X \sim \text{Uniform}(0, 1)$ if

\[
f(x) = \begin{cases} 
1 & \text{if } x \in [0, 1] \\
0 & \text{otherwise}
\end{cases}
\]

\[
P(X \in [0.5, 0.5]) = \int_{0.5}^{0.5} 1 \, dx = X\big|_{0.5}^{0.5} = 0.5 - 0.5 = 0
\]
Example: Uniform Random Variable

$X \sim \text{Uniform}(0, 1)$ if

$$f(x) = \begin{cases} 1 & \text{if } x \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

$$P(X \in \{[0, 0.2] \cup [0.5, 1]\}) = \int_0^{0.2} 1 \, dx + \int_{0.5}^{1} 1 \, dx$$

$$= X_{0.2}^0 + X_{0.5}^1$$

$$= 0.2 - 0 + 1 - 0.5$$

$$= 0.7$$
Example: Uniform Random Variable

\[ X \sim \text{Uniform}(0, 1) \text{ if } \]

\[
  f(x) = 1 \text{ if } x \in [0, 1] \\
  f(x) = 0 \text{ otherwise}
\]

To summarize

- \( P(X = a) = 0 \)
- \( P(X \in (-\infty, \infty)) = 1 \)
- If \( F \) is antiderivative of \( f \), then \( P(X \in [c, d]) = F(d) - F(c) \) (Fundamental theorem of calculus)
Cumulative Mass Function

Probability density function \((f)\) characterizes \textit{distribution} of continuous random variable.
Cumulative Mass Function

Probability density function ($f$) characterizes *distribution* of continuous random variable.

Equivalently, Cumulative density (distribution) function characterizes continuous random variables.
Cumulative Mass Function

Probability density function \( f \) characterizes distribution of continuous random variable.

Equivalently, Cumulative density (distribution) function characterizes continuous random variables.

Definition

*Cumulative Distribution function.* For a continuous random variable \( X \) define its cumulative density function \( F(x) \) as,

\[
F(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)dx
\]
Cumulative Mass Function

Probability density function \((f)\) characterizes distribution of continuous random variable.

Equivalently, Cumulative density (distribution) function characterizes continuous random variables.

Definition

*Cumulative Distribution function.* For a continuous random variable \(X\) define its cumulative density function \(F(x)\) as,

\[
F(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)\,dx
\]
Cumulative Mass Function

Probability density function ($f$) characterizes distribution of continuous random variable.
Equivalently, Cumulative density (distribution) function characterizes continuous random variables.

Definition

*Cumulative Distribution function. For a continuous random variable $X$ define its cumulative density function $F(x)$ as,*

$$F(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)dx$$
Cumulative Mass Function

Probability density function \( f \) characterizes distribution of continuous random variable.

Equivalently, Cumulative density (distribution) function characterizes continuous random variables.

**Definition**

*Cumulative Distribution function.* For a continuous random variable \( X \) define its cumulative density function \( F(x) \) as,

\[
F(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)dx
\]
Cumulative Mass Function

Probability density function \((f)\) characterizes distribution of continuous random variable.

Equivalently, Cumulative density (distribution) function characterizes continuous random variables.

Definition

Cumulative Distribution function. For a continuous random variable \(X\) define its cumulative density function \(F(x)\) as,

\[
F(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)dx
\]
Cumulative Mass Function
Probability density function ($f$) characterizes distribution of continuous random variable.
Equivalently, Cumulative density (distribution) function characterizes continuous random variables.

Definition
Cumulative Distribution function. For a continuous random variable $X$ define its cumulative density function $F(x)$ as,

$$F(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)dx$$
Uniform Random Variable

Suppose $X \sim \text{Uniform}(0, 1)$, then

$$F(t) = \begin{cases} 
0, & \text{if } t < 0 \\
1, & \text{if } t > 1 \\
t, & \text{if } t \in [0, 1] 
\end{cases}$$
Uniform Random Variable

Suppose $X \sim \text{Uniform}(0, 1)$, then

$$F(t) = P(X \leq t)$$
Uniform Random Variable

Suppose $X \sim Uniform(0, 1)$, then

$$F(t) = P(X \leq t)$$

$$= 0, \text{ if } t < 0$$
Uniform Random Variable

Suppose $X \sim \text{Uniform}(0, 1)$, then

$$F(t) = P(X \leq t)$$

$$= 0, \text{ if } t < 0$$

$$= 1, \text{ if } t > 1$$
Uniform Random Variable

Suppose $X \sim \text{Uniform}(0, 1)$, then

$$F(t) = P(X \leq t)$$

$$= 0, \text{ if } t < 0$$

$$= 1, \text{ if } t > 1$$

$$= t, \text{ if } t \in [0, 1]$$

Cumulative Density Function
Definition

*If $X$ is a continuous random variable then,*

$$E[X] = \int_{-\infty}^{\infty} xf(x)\,dx$$
Suppose $X \sim \text{Uniform}(0, 1)$. What is $E[X]$?
Suppose $X \sim Uniform(0, 1)$. What is $E[X]$?

$$E[X]$$
Suppose $X \sim \text{Uniform}(0,1)$. What is $E[X]$?

\[ E[X] = \int_{-\infty}^{\infty} xf(x) \, dx \]
Suppose $X \sim \text{Uniform}(0, 1)$. What is $E[X]$?

$$
E[X] = \int_{-\infty}^{\infty} x f(x) \, dx
$$

$$
= \int_{-\infty}^{0} x \cdot 0 \, dx + \int_{0}^{1} x \cdot 1 \, dx + \int_{1}^{\infty} x \cdot 0 \, dx
$$
Suppose $X \sim Uniform(0, 1)$. What is $E[X]$?

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

$$= \int_{-\infty}^{0} x0dx + \int_{0}^{1} x1dx + \int_{1}^{\infty} x0dx$$

$$= 0 + \frac{x^2}{2} \bigg|_{0}^{1} + 0$$

$$= \frac{1}{2}$$
Suppose $X \sim \text{Uniform}(0, 1)$. What is $E[X]$?

$$E[X] = \int_{-\infty}^{\infty} xf(x) \, dx$$

$$= \int_{-\infty}^{0} x \cdot 0 \, dx + \int_{0}^{1} x \cdot 1 \, dx + \int_{1}^{\infty} x \cdot 0 \, dx$$

$$= 0 + \frac{x^2}{2} \bigg|_{0}^{1} + 0$$

$$= 0 + \frac{1}{2} + 0$$
Suppose $X \sim \text{Uniform}(0, 1)$. What is $E[X]$?

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

$$= \int_{-\infty}^{0} x0dx + \int_{0}^{1} x1dx + \int_{1}^{\infty} x0dx$$

$$= 0 + \frac{x^2}{2} \bigg|_{0}^{1} + 0$$

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

$$= \frac{1}{2}$$
Expectations of Functions

Proposition

Suppose $X$ is a continuous random variable and $g: \mathbb{R} \to \mathbb{R}$ (that isn’t crazy). Then,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$
Suppose $g(X) = X^2$ and $X \sim \text{Uniform}(0, 1)$. What is $E[g(X)]$?
Expectations of Functions

Suppose \( g(X) = X^2 \) and \( X \sim \text{Uniform}(0, 1) \). What is \( E[g(X)] \)?

\[
E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) \, dx = \int_{0}^{1} x^2 \, dx = \frac{x^3}{3} \bigg|_{0}^{1} = \frac{1}{3}
\]
Expectations of Functions

Suppose $g(X) = X^2$ and $X \sim \text{Uniform}(0, 1)$. What is $E[g(X)]$?

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$
Expectations of Functions

Suppose $g(X) = X^2$ and $X \sim \text{Uniform}(0, 1)$. What is $E[g(X)]$?

\[
E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx
\]

\[
= \int_{0}^{1} x^2 dx
\]
Expectations of Functions

Suppose $g(X) = X^2$ and $X \sim \text{Uniform}(0, 1)$. What is $E[g(X)]$?

\[
E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx
\]
\[
= \int_{0}^{1} x^2 dx
\]
\[
= \left. \frac{x^3}{3} \right|_{0}^{1} = \frac{1}{3}
\]
Expectations of Functions

Suppose $g(X) = X^2$ and $X \sim \text{Uniform}(0, 1)$. What is $E[g(X)]$?

\[
E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx
\]

\[
= \int_{0}^{1} x^2 \,dx
\]

\[
= \left. \frac{x^3}{3} \right|_{0}^{1}
\]

\[
= \frac{1}{3}
\]
Corollary

Suppose $X$ is a continuous random variable. Then,

$$E[aX + b] = aE[X] + b$$

Proof.
Corollary

Suppose \( X \) is a continuous random variable. Then,

\[
E[aX + b] = aE[X] + b
\]

Proof.

\[
E[aX + b] = \int_{-\infty}^{\infty} (ax + b)f(x)dx
\]
Corollary

Suppose $X$ is a continuous random variable. Then,

$$E[aX + b] = aE[X] + b$$

Proof.

$$E[aX + b] = \int_{-\infty}^{\infty} (ax + b)f(x)\,dx$$

$$= a \int_{-\infty}^{\infty} xf(x)\,dx + b \int_{-\infty}^{\infty} f(x)\,dx$$
Corollary

Suppose $X$ is a continuous random variable. Then,

$$E[aX + b] = aE[X] + b$$

Proof.

$$E[aX + b] = \int_{-\infty}^{\infty} (ax + b)f(x)dx$$

$$= a \int_{-\infty}^{\infty} xf(x)dx + b \int_{-\infty}^{\infty} f(x)dx$$

$$= aE[X] + b \times 1$$
Definition

Variance. If $X$ is a continuous random variable, define its variance, $\text{Var}(X)$,

$$\text{Var}(X) = E[(X - E[X])^2]$$

$$= \int_{-\infty}^{\infty} (x - E[X])^2 f(x) \, dx$$

$$= E[X^2] - E[X]^2$$
Variance: Random Variable

$X \sim \text{Uniform}(0, 1)$. What is $\text{Var}(X)$?
Variance: Random Variable

$X \sim \text{Uniform}(0, 1)$. What is $\text{Var}(X)$?

$$E[X^2] = \frac{1}{3}$$
Variance: Random Variable

$X \sim \text{Uniform}(0, 1)$. What is $\text{Var}(X)$?

\[
E[X^2] = \frac{1}{3} \\
E[X]^2 = \left(\frac{1}{2}\right)^2
\]
Variance: Random Variable

\( X \sim \text{Uniform}(0, 1). \) What is \( \text{Var}(X) \)?

\[
E[X^2] = \frac{1}{3}
\]

\[
E[X]^2 = \left( \frac{1}{2} \right)^2 = \frac{1}{4}
\]

\[
\text{Var}(X) = E[X^2] - E[X]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}
\]
Variance: Random Variable

\( X \sim \text{Uniform}(0, 1) \). What is \( \text{Var}(X) \)?

\[
\begin{align*}
E[X^2] & = \frac{1}{3} \\
E[X]^2 & = \left( \frac{1}{2} \right)^2 \\
& = \frac{1}{4}
\end{align*}
\]

\[
\text{Var}(X) = E[X^2] - E[X]^2
\]
Variance: Random Variable

$X \sim \text{Uniform}(0, 1)$. What is $\text{Var}(X)$?

\[
E[X^2] = \frac{1}{3}
\]
\[
E[X]^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4}
\]

\[
\text{Var}(X) = E[X^2] - E[X]^2
\]
\[
= \frac{1}{3} - \frac{1}{4} = \frac{1}{12}
\]
Famous Continuous Distributions

- Normal Distribution
- Gamma distribution
- $\chi^2$ Distribution
- $t$ Distribution
- Beta, Dirichlet distributions (not today!)
- $F$-distribution (not today!)
Definition

Suppose $X$ is a random variable with $X \in \mathbb{R}$ and density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Then $X$ is a normally distributed random variable with parameters $\mu$ and $\sigma^2$.

Equivalently, we’ll write

$$X \sim \text{Normal}(\mu, \sigma^2)$$
Support for President Obama

Suppose we are interested in modeling \textit{presidential approval}
Support for President Obama

Suppose we are interested in modeling presidential approval

- Let $Y$ represent random variable: proportion of population who “approves job president is doing”
Support for President Obama

Suppose we are interested in modeling presidential approval

- Let $Y$ represent random variable: proportion of population who “approves job president is doing”

- Individual responses (that constitute proportion) are independent and identically distributed (sufficient, not necessary) and we take the average of those individual responses
Suppose we are interested in modeling \textit{presidential approval}

- Let \( Y \) represent random variable: proportion of population who
  “approves job president is doing”

- Individual responses (that constitute proportion) are \textit{independent and
  identically distributed} (sufficient, not necessary) and we take the
  average of those individual responses

- Observe \textit{many} responses (\( N \to \infty \))
Support for President Obama

Suppose we are interested in modeling presidential approval

- Let $Y$ represent random variable: proportion of population who “approves job president is doing”
- Individual responses (that constitute proportion) are independent and identically distributed (sufficient, not necessary) and we take the average of those individual responses
- Observe many responses ($N \to \infty$)
- Then (by Central Limit Theorem) $Y$ is Normally distributed, or
Support for President Obama

Suppose we are interested in modeling presidential approval

- Let $Y$ represent random variable: proportion of population who “approves job president is doing”

- Individual responses (that constitute proportion) are independent and identically distributed (sufficient, not necessary) and we take the average of those individual responses

- Observe many responses ($N \to \infty$)

- Then (by Central Limit Theorem) $Y$ is Normally distributed, or

$$Y \sim \text{Normal}(\mu, \sigma^2)$$
Support for President Obama

Suppose we are interested in modeling presidential approval
- Let $Y$ represent random variable: proportion of population who “approves job president is doing”
- Individual responses (that constitute proportion) are independent and identically distributed (sufficient, not necessary) and we take the average of those individual responses
- Observe many responses ($N \to \infty$)
- Then (by Central Limit Theorem) $Y$ is Normally distributed, or

$$Y \sim \text{Normal}(\mu, \sigma^2)$$

$$f(y) = \frac{\exp \left( - \frac{(y-\mu)^2}{2\sigma^2} \right)}{\sqrt{2\pi \sigma^2}}$$
Central Limit Theorem
We’ll prove it on Thursday.
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem

We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:

Mean of 5
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:

Mean of 8

Density

Justin Grimmer (Stanford University)
Methodology I

September 14th, 2015 20 / 45
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.
Central Limit Theorem

We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:

Mean of 12
Central Limit Theorem

We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:

Mean of 15

Justin Grimmer (Stanford University)
Central Limit Theorem
We’ll prove it on Thursday.
Central Limit Theorem

We’ll prove it on Thursday.

Simulation:

Mean of 17
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:

Mean of 18

Justin Grimmer (Stanford University)
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem

We’ll prove it on Thursday.

Simulation:

Mean of 21
Central Limit Theorem
We’ll prove it on Thursday.
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:

Mean of 23
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:

Mean of 25
Central Limit Theorem

We’ll prove it on Thursday.
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:

Mean of 29

Density

Justin Grimmer (Stanford University)
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:

Mean of 31

Density

Mean

Justin Grimmer (Stanford University)
Methodology I
September 14th, 2015 20 / 45
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:

Mean of 32
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem

We’ll prove it on Thursday.
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:

Mean of 36
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem

We’ll prove it on Thursday.

Simulation:

Mean of 38

Justin Grimmer (Stanford University)
Central Limit Theorem
We’ll prove it on Thursday.
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:

Mean of 41
Central Limit Theorem

We’ll prove it on Thursday.

Simulation:

Mean of 42

Density

Mean

Justin Grimmer (Stanford University)
Central Limit Theorem
We’ll prove it on Thursday.
Central Limit Theorem
We’ll prove it on Thursday.

Mean of 44

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.
Central Limit Theorem

We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:

Mean of 48

Justin Grimmer (Stanford University)
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Central Limit Theorem
We’ll prove it on Thursday.

Simulation:
Expected Value/Variance of Normal Distribution

$Z$ is a standard normal distribution if

$$Z \sim \text{Normal}(0, 1)$$

The cumulative density function of $Z$, $F_Z(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{z^2}{2}\right) dz$.

Proposition Scale/Location. If $Z \sim \text{N}(0, 1)$, then $X = aZ + b$ is, $X \sim \text{Normal}(b, a^2)$. 
Expected Value/Variance of Normal Distribution

$Z$ is a standard normal distribution if

$$Z \sim \text{Normal}(0, 1)$$
Expected Value/Variance of Normal Distribution

$Z$ is a standard normal distribution if

$$Z \sim \text{Normal}(0, 1)$$

We’ll call the cumulative density function of $Z$,
Expected Value/Variance of Normal Distribution

$Z$ is a standard normal distribution if

$$Z \sim \text{Normal}(0, 1)$$

We’ll call the cumulative density function of $Z$,

$$F_Z(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-z^2/2)dz$$
Expected Value/Variance of Normal Distribution

\( Z \) is a standard normal distribution if

\[
Z \sim \text{Normal}(0, 1)
\]

We’ll call the cumulative density function of \( Z \),

\[
F_Z(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-z^2/2) \, dz
\]

Proposition

Scale/Location. If \( Z \sim N(0, 1) \), then \( X = aZ + b \) is,

\[
X \sim \text{Normal}(b, a^2)
\]
Intuition

Suppose $Z \sim \text{Normal}(0, 1)$. 

![Normal Distribution Graph]

$Z \sim \text{N}(0, 1)$
Intuition

Suppose $Z \sim \text{Normal}(0, 1)$.

$Y = 2Z + 6$
Intuition

Suppose $Z \sim \text{Normal}(0, 1)$.

$Y = 2Z + 6$

$Y \sim \text{Normal}(6, 4)$
Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove
Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for $Y$ is a normal distribution.
Proof: \( Z \sim N(0, 1) \) and \( Y = aZ + b \), then \( Y \sim N(b, a^2) \)

To prove we need to show that density for \( Y \) is a normal distribution. That is, we’ll show \( F_Y(x) \) is Normal cdf.

Call \( F_Z(x) \) cdf for standardized normal. \( F_Y(x) = P(Y \leq x) = P(aZ + b \leq x) = P(Z \leq \frac{x - b}{a}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x-b/a} \exp(-z^2/2) \, dz = F_Z(x-b/a) \)
Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for $Y$ is a normal distribution. That is, we’ll show $F_Y(x)$ is Normal cdf. Call $F_Z(x)$ cdf for standardized normal.
Proof: $Z \sim N(0,1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for $Y$ is a normal distribution. That is, we’ll show $F_Y(x)$ is Normal cdf. Call $F_Z(x)$ cdf for standardized normal.

$$F_Y(x) = P(Y \leq x)$$
Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for $Y$ is a normal distribution. That is, we’ll show $F_Y(x)$ is Normal cdf. Call $F_Z(x)$ cdf for standardized normal.

\[
F_Y(x) = P(Y \leq x) = P(aZ + b \leq x)
\]
Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for $Y$ is a normal distribution. That is, we’ll show $F_Y(x)$ is Normal cdf. Call $F_Z(x)$ cdf for standardized normal.

$$F_Y(x) = P(Y \leq x) = P(aZ + b \leq x) = P(Z \leq \frac{x - b}{a})$$
Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for $Y$ is a normal distribution. That is, we’ll show $F_Y(x)$ is Normal cdf.
Call $F_Z(x)$ cdf for standardized normal.

\[
F_Y(x) = P(Y \leq x)
= P(aZ + b \leq x)
= P(Z \leq \frac{x - b}{a})
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-b}{a}} \exp\left(-\frac{z^2}{2}\right)dz
\]
Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for $Y$ is a normal distribution. That is, we’ll show $F_Y(x)$ is Normal cdf.

Call $F_Z(x)$ cdf for standardized normal.

\[
F_Y(x) = P(Y \leq x) \\
= P(aZ + b \leq x) \\
= P(Z \leq \left[ \frac{x - b}{a} \right]) \\
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-b}{a}} \exp\left(-\frac{z^2}{2}\right) dz \\
= F_Z\left(\frac{x - b}{a}\right)
\]
Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

So, we can work with $F_Z(\frac{x-b}{a})$. 
Proof: \( Z \sim N(0, 1) \) and \( Y = aZ + b \), then \( Y \sim N(b, a^2) \)

So, we can work with \( F_Z\left(\frac{x-b}{a}\right) \).

\[
\frac{\partial F_Y(x)}{\partial x} = \frac{\partial F_Z\left(\frac{x-b}{a}\right)}{\partial x}
\]
Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

So, we can work with $F_Z\left(\frac{x-b}{a}\right)$.

$$\frac{\partial F_Y(x)}{\partial x} = \frac{\partial F_Z\left(\frac{x-b}{a}\right)}{\partial x}$$

$$= f_Z\left(\frac{x-b}{a}\right) \frac{1}{a} \text{ By the chain rule}$$
Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

So, we can work with $F_Z\left(\frac{x-b}{a}\right)$.

$$\frac{\partial F_Y(x)}{\partial x} = \frac{\partial F_Z\left(\frac{x-b}{a}\right)}{\partial x}$$

$$= f_Z\left(\frac{x-b}{a}\right) \frac{1}{a} \text{ By the chain rule}$$

$$= \frac{1}{\sqrt{2\pi}a} \exp \left[ -\frac{(x-b)^2}{2a^2} \right] \text{ By definition of } f_Z(x) \text{ or FTC}$$
Proof: $Z \sim N(0,1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

So, we can work with $F_Z\left(\frac{x-b}{a}\right)$.

\[
\frac{\partial F_Y(x)}{\partial x} = \frac{\partial F_Z\left(\frac{x-b}{a}\right)}{\partial x} \\
= f_Z\left(\frac{x-b}{a}\right) \frac{1}{a} \quad \text{By the chain rule} \\
= \frac{1}{\sqrt{2\pi a}} \exp \left[ -\frac{(x-b)^2}{2a^2} \right] \quad \text{By definition of } f_Z(x) \text{ or FTC} \\
= \frac{1}{\sqrt{2\pi a}} \exp \left[ -\frac{(x-b)^2}{2a^2} \right]
\]
Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

So, we can work with $F_Z\left(\frac{x-b}{a}\right)$.

$$\frac{\partial F_Y(x)}{\partial x} = \frac{\partial F_Z\left(\frac{x-b}{a}\right)}{\partial x}$$

$$= f_Z\left(\frac{x-b}{a}\right) \frac{1}{a} \text{ By the chain rule}$$

$$= \frac{1}{\sqrt{2\pi a}} \exp\left[- \left(\frac{x-b}{a}\right)^2 \right] \text{ By definition of } f_Z(x) \text{ or FTC}$$

$$= \frac{1}{\sqrt{2\pi a}} \exp\left[- \frac{(x-b)^2}{2a^2} \right]$$

$$= \text{Normal}(b, a^2)$$
Expectation and Variance

Assume we know:

\[
E[Z] = 0 \\
Var(Z) = 1
\]
Expectation and Variance

Assume we know:

\[ E[Z] = 0 \]
\[ Var(Z) = 1 \]

This implies that, for \( Y \sim \text{Normal}(\mu, \sigma^2) \)
Expectation and Variance

Assume we know:

\[ E[Z] = 0 \]
\[ Var(Z) = 1 \]

This implies that, for \( Y \sim \text{Normal}(\mu, \sigma^2) \)

\[ E[Y] = E[\sigma Z + \mu] \]
Expectation and Variance

Assume we know:

\[ E[Z] = 0 \]
\[ Var(Z) = 1 \]

This implies that, for \( Y \sim \text{Normal}(\mu, \sigma^2) \)

\[
E[Y] = E[\sigma Z + \mu] \\
= \sigma E[Z] + \mu
\]
Expectation and Variance

Assume we know:

\[ E[Z] = 0 \]
\[ Var(Z) = 1 \]

This implies that, for \( Y \sim \text{Normal}(\mu, \sigma^2) \)

\[
E[Y] = E[\sigma Z + \mu] \\
= \sigma E[Z] + \mu \\
= \mu
\]
Expectation and Variance

Assume we know:

\[ E[Z] = 0 \]
\[ \text{Var}(Z) = 1 \]

This implies that, for \( Y \sim \text{Normal}(\mu, \sigma^2) \)

\[
E[Y] = E[\sigma Z + \mu] \\
= \sigma E[Z] + \mu \\
= \mu \\
\text{Var}(Y) = \text{Var}(\sigma Z + \mu)
\]
Expectation and Variance

Assume we know:

\[
E[Z] = 0 \\
\text{Var}(Z) = 1
\]

This implies that, for \( Y \sim \text{Normal}(\mu, \sigma^2) \)

\[
E[Y] = E[\sigma Z + \mu] \\
= \sigma E[Z] + \mu \\
= \mu \\
\text{Var}(Y) = \text{Var}(\sigma Z + \mu) \\
= \sigma^2 \text{Var}(Z) + \text{Var}(\mu)
\]
Expectation and Variance

Assume we know:

\[ E[Z] = 0 \]
\[ Var(Z) = 1 \]

This implies that, for \( Y \sim \text{Normal}(\mu, \sigma^2) \)

\[ E[Y] = E[\sigma Z + \mu] \]
\[ = \sigma E[Z] + \mu \]
\[ = \mu \]

\[ Var(Y) = Var(\sigma Z + \mu) \]
\[ = \sigma^2 Var(Z) + Var(\mu) \]
\[ = \sigma^2 + 0 \]
Expectation and Variance

Assume we know:

\[ E[Z] = 0 \]
\[ Var(Z) = 1 \]

This implies that, for \( Y \sim \text{Normal}(\mu, \sigma^2) \)

\[
E[Y] = E[\sigma Z + \mu] \\
= \sigma E[Z] + \mu \\
= \mu
\]

\[
Var(Y) = Var(\sigma Z + \mu) \\
= \sigma^2 Var(Z) + Var(\mu) \\
= \sigma^2 + 0 \\
= \sigma^2
\]
Back To Obama

Suppose $\mu = 0.39$ and $\sigma^2 = 0.0025$
Back To Obama

Suppose $\mu = 0.39$ and $\sigma^2 = 0.0025$

$P(Y \geq 0.45)$ (What is the probability it isn’t that bad?) ?
Suppose $\mu = 0.39$ and $\sigma^2 = 0.0025$

$P(Y \geq 0.45)$ (What is the probability it isn’t that bad?)

$$P(Y \geq 0.45) = 1 - P(Y \leq 0.45)$$
Back To Obama

Suppose $\mu = 0.39$ and $\sigma^2 = 0.0025$
$P(Y \geq 0.45)$ (What is the probability it isn’t that bad?) ?

$$P(Y \geq 0.45) = 1 - P(Y \leq 0.45)$$
$$= 1 - P(0.05Z + 0.39 \leq 0.45)$$
Suppose $\mu = 0.39$ and $\sigma^2 = 0.0025$

$P(Y \geq 0.45)$ (What is the probability it isn’t that bad?)

\[
P(Y \geq 0.45) = 1 - P(Y \leq 0.45)
\]

\[
= 1 - P(0.05Z + 0.39 \leq 0.45)
\]

\[
= 1 - P(Z \leq \frac{0.45 - 0.39}{0.05})
\]

\[
= 1 - P(Z \leq 0.12)
\]

\[
= 1 - F_Z(0.12)
\]

\[
= 0.1150697
\]
Back To Obama

Suppose $\mu = 0.39$ and $\sigma^2 = 0.0025$

$P(Y \geq 0.45)$ (What is the probability it isn’t that bad?)

$$P(Y \geq 0.45) = 1 - P(Y \leq 0.45)$$

$$= 1 - P(0.05Z + 0.39 \leq 0.45)$$

$$= 1 - P(Z \leq \frac{0.45 - 0.39}{0.05})$$

$$= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{6/5} \exp(-z^2/2)dz$$
Back To Obama

Suppose $\mu = 0.39$ and $\sigma^2 = 0.0025$

$P(Y \geq 0.45)$ (What is the probability it isn’t that bad?)

$$P(Y \geq 0.45) = 1 - P(Y \leq 0.45)$$
$$= 1 - P(0.05Z + 0.39 \leq 0.45)$$
$$= 1 - P(Z \leq \frac{0.45 - 0.39}{0.05})$$
$$= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{6/5} \exp(-z^2/2)dz$$
$$= 1 - F_Z\left(\frac{6}{5}\right)$$
Suppose $\mu = 0.39$ and $\sigma^2 = 0.0025$

$P(Y \geq 0.45)$ (What is the probability it isn’t that bad?)

\[
P(Y \geq 0.45) = 1 - P(Y \leq 0.45)
= 1 - P(0.05Z + 0.39 \leq 0.45)
= 1 - P(Z \leq \frac{0.45 - 0.39}{0.05})
= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{6/5} \exp(-z^2/2)dz
= 1 - F_Z\left(\frac{6}{5}\right)
= 0.1150697
\]
Via simulation:

```r
draws <- rnorm(1e7, mean = 0.39, sd = sqrt(0.0025))
greater <- which(draws > 0.45)
p.45 <- length(greater) / 1e7
print(p.45)
[1] 0.1149824
```

Justin Grimmer (Stanford University)
The Gamma Function

Definition

Suppose $\alpha > 0$. Then define $\Gamma(\alpha)$ as

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} \, dy$$

- For $\alpha \in \{1, 2, 3, \ldots\}$
  $$\Gamma(\alpha) = (\alpha - 1)!$$
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
Gamma Distribution

Suppose we have $\Gamma(\alpha)$,
Gamma Distribution

Suppose we have $\Gamma(\alpha)$,

$$\frac{\Gamma(\alpha)}{\Gamma(\alpha)} = \int_0^\infty y^{\alpha-1}e^{-y} dy$$

$$1 = \int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1}e^{-y} dy$$
Gamma Distribution

Suppose we have $\Gamma(\alpha)$,

\[
\frac{\Gamma(\alpha)}{\Gamma(\alpha)} = \int_0^\infty y^{\alpha - 1} e^{-y} dy
\]

\[
1 = \int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha - 1} e^{-y} dy
\]

Set $X = \frac{Y}{\beta}$
Gamma Distribution

Suppose we have \( \Gamma(\alpha) \),

\[
\frac{\Gamma(\alpha)}{\Gamma(\alpha)} = \frac{\int_0^\infty y^{\alpha-1}e^{-y} \, dy}{\Gamma(\alpha)}
\]

\[
1 = \int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1}e^{-y} \, dy
\]

Set \( X = Y/\beta \)

\[
F(x) = P(X \leq x) = P(Y/\beta \leq x)
\]
Gamma Distribution

Suppose we have $\Gamma(\alpha)$,

$$\frac{\Gamma(\alpha)}{\Gamma(\alpha)} = \frac{\int_0^\infty y^{\alpha-1}e^{-y} \, dy}{\Gamma(\alpha)}$$

$$1 = \int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1}e^{-y} \, dy$$

Set $X = Y/\beta$

$$F(x) = P(X \leq x) = P(Y/\beta \leq x) = P(Y \leq x\beta)$$
Gamma Distribution

Suppose we have $\Gamma(\alpha)$,

$$\frac{\Gamma(\alpha)}{\Gamma(\alpha)} = \frac{\int_0^\infty y^{\alpha-1}e^{-y} dy}{\Gamma(\alpha)}$$

$$1 = \int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1}e^{-y} dy$$

Set $X = Y/\beta$

$$F(x) = P(X \leq x) = P(Y/\beta \leq x)$$

$$= P(Y \leq x\beta)$$

$$= F_Y(x\beta)$$
Gamma Distribution

Suppose we have $\Gamma(\alpha)$,

$$\frac{\Gamma(\alpha)}{\Gamma(\alpha)} = \int_0^\infty y^{\alpha-1}e^{-y} \, dy$$

$$1 = \int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1}e^{-y} \, dy$$

Set $X = Y/\beta$

$$F(x) = P(X \leq x) = P(Y/\beta \leq x) = P(Y \leq x\beta) = F_Y(x\beta)$$

$$\frac{\partial F_Y(x\beta)}{\partial x} = f_Y(x\beta)\beta$$
Gamma Distribution

Suppose we have $\Gamma(\alpha)$,

$$\frac{\Gamma(\alpha)}{\Gamma(\alpha)} = \frac{\int_0^\infty y^{\alpha-1}e^{-y} dy}{\Gamma(\alpha)}$$

$$1 = \int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1}e^{-y} dy$$

Set $X = Y/\beta$

$$F(x) = P(X \leq x) = P(Y/\beta \leq x)$$

$$= P(Y \leq x\beta)$$

$$= F_Y(x\beta)$$

$$\frac{\partial F_Y(x\beta)}{\partial x} = f_Y(x\beta) \beta$$

The result is:
Gamma Distribution

Suppose we have $\Gamma(\alpha)$,

\[
\frac{\Gamma(\alpha)}{\Gamma(\alpha)} = \frac{\int_0^\infty y^{\alpha-1} e^{-y} \, dy}{\Gamma(\alpha)}
\]

\[
1 = \int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y} \, dy
\]

Set $X = Y/\beta$

\[F(x) = P(X \leq x) = P(Y/\beta \leq x)\]
\[= P(Y \leq x\beta) = F_Y(x\beta)\]

\[
\frac{\partial F_Y(x\beta)}{\partial x} = f_Y(x\beta)\beta
\]

The result is:

\[f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-x\beta}\]
Definition

Suppose $X$ is a continuous random variable, with $X \geq 0$. Then if the pdf of $X$ is

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-x\beta}$$

if $x \geq 0$ and 0 otherwise, we will say $X$ is a Gamma distribution.

$$X \sim \text{Gamma}(\alpha, \beta)$$
Gamma Distribution

Suppose $X \sim \text{Gamma}(\alpha, \beta)$
Gamma Distribution

Suppose $X \sim \text{Gamma}(\alpha, \beta)$

$$E[X] = \frac{\alpha}{\beta}$$
Gamma Distribution

Suppose $X \sim \text{Gamma}(\alpha, \beta)$

\[ E[X] = \frac{\alpha}{\beta} \]
\[ \text{var}(X) = \frac{\alpha}{\beta^2} \]

Suppose $\alpha = 1$ and $\beta = \lambda$. If $X \sim \text{Gamma}(1, \lambda)$

\[ f(x | 1, \lambda) = \lambda e^{-x \lambda} \]

We will say $X \sim \text{Exponential}(\lambda)$. 
Gamma Distribution

Suppose $X \sim \text{Gamma}(\alpha, \beta)$

$$E[X] = \frac{\alpha}{\beta}$$

$$\text{var}(X) = \frac{\alpha}{\beta^2}$$

Suppose $\alpha = 1$ and $\beta = \lambda$. If
Gamma Distribution

Suppose $X \sim \text{Gamma}(\alpha, \beta)$

$$E[X] = \frac{\alpha}{\beta}$$

$$\text{var}(X) = \frac{\alpha}{\beta^2}$$

Suppose $\alpha = 1$ and $\beta = \lambda$. If

$$X \sim \text{Gamma}(1, \lambda)$$
Gamma Distribution

Suppose $X \sim \text{Gamma}(\alpha, \beta)$

\[
E[X] = \frac{\alpha}{\beta} \\
\text{var}(X) = \frac{\alpha}{\beta^2}
\]

Suppose $\alpha = 1$ and $\beta = \lambda$. If

\[
X \sim \text{Gamma}(1, \lambda) \\
f(x|1, \lambda) = \lambda e^{-x\lambda}
\]
Gamma Distribution

Suppose $X \sim \text{Gamma}(\alpha, \beta)$

$$E[X] = \frac{\alpha}{\beta}$$

$$\text{var}(X) = \frac{\alpha}{\beta^2}$$

Suppose $\alpha = 1$ and $\beta = \lambda$. If

$$X \sim \text{Gamma}(1, \lambda)$$

$$f(x|1, \lambda) = \lambda e^{-x\lambda}$$

We will say
Gamma Distribution

Suppose $X \sim \text{Gamma}(\alpha, \beta)$

$$E[X] = \frac{\alpha}{\beta}$$

$$\text{var}(X) = \frac{\alpha}{\beta^2}$$

Suppose $\alpha = 1$ and $\beta = \lambda$. If

$$X \sim \text{Gamma}(1, \lambda)$$

$$f(x|1, \lambda) = \lambda e^{-x\lambda}$$

We will say

$$X \sim \text{Exponential}(\lambda)$$
Properties of Gamma Distributions

Proposition

Suppose we have a sequence of independent random variables, with

\[ X_i \sim \text{Gamma}(\alpha_i, \beta) \]

Then

\[ Y = \sum_{i=1}^{N} X_i \]

\[ Y \sim \text{Gamma}(\sum_{i=1}^{N} \alpha_i, \beta) \]
We can evaluate in R with `dgamma` and simulate with `rgamma`.

\[ X \sim \text{Gamma}(3, 5) \] and we evaluate at 3,

\[ \text{dgamma}(3, \text{shape}= 3, \text{rate} = 5) \]

and we can simulate with

\[ \text{rgamma}(1000, \text{shape} = 3, \text{rate} = 5) \]
$\chi^2$ Distribution

Suppose $Z \sim \text{Normal}(0, 1)$.
\( \chi^2 \) Distribution

Suppose \( Z \sim \text{Normal}(0, 1) \).
Consider \( X = Z^2 \)
\( \chi^2 \) Distribution

Suppose \( Z \sim \text{Normal}(0, 1) \).
Consider \( X = Z^2 \)

\[
F_X(x) = P(X \leq x)
\]
Suppose $Z \sim \text{Normal}(0, 1)$. Consider $X = Z^2$

$$F_X(x) = P(X \leq x) = P(Z^2 \leq x)$$
\( \chi^2 \) Distribution

Suppose \( Z \sim \text{Normal}(0,1) \).

Consider \( X = Z^2 \)

\[
F_X(x) = P(X \leq x) \\
= P(Z^2 \leq x) \\
= P(-\sqrt{x} \leq Z \leq \sqrt{x})
\]

The pdf then is

\[
\frac{\partial}{\partial x} F_X(x) = f_{Z}(\sqrt{x}) \frac{1}{2\sqrt{x}} + f_{Z}(-\sqrt{x}) \frac{1}{2\sqrt{x}}
\]
\( \chi^2 \) Distribution

Suppose \( Z \sim \text{Normal}(0, 1) \).
Consider \( X = Z^2 \)

\[
F_X(x) = P(X \leq x)
= P(Z^2 \leq x)
= P(-\sqrt{x} \leq Z \leq \sqrt{x})
= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-\frac{z^2}{2}} \, dz
\]
\( \chi^2 \) Distribution

Suppose \( Z \sim \text{Normal}(0, 1) \).
Consider \( X = Z^2 \)

\[
F_X(x) = P(X \leq x) \\
= P(Z^2 \leq x) \\
= P(-\sqrt{x} \leq Z \leq \sqrt{x}) \\
= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-\frac{z^2}{2}} \, dz \\
= F_Z(\sqrt{x}) - F_Z(-\sqrt{x})
\]
\( \chi^2 \) Distribution

Suppose \( Z \sim \text{Normal}(0, 1) \).
Consider \( X = Z^2 \)

\[
F_X(x) = P(X \leq x) = P(Z^2 \leq x) = P(-\sqrt{x} \leq Z \leq \sqrt{x}) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-\frac{z^2}{2}} dz = F_Z(\sqrt{x}) - F_Z(-\sqrt{x})
\]

The pdf then is
\( \chi^2 \) Distribution

Suppose \( Z \sim \text{Normal}(0, 1) \).
Consider \( X = Z^2 \)

\[
F_X(x) = P(X \leq x) = P(Z^2 \leq x) = P(-\sqrt{x} \leq Z \leq \sqrt{x}) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-\frac{z^2}{2}} \, dz = F_Z(\sqrt{x}) - F_Z(-\sqrt{x})
\]

The pdf then is

\[
\frac{\partial F_X(x)}{\partial x} = f_Z(\sqrt{x}) \frac{1}{2\sqrt{x}} + f_Z(-\sqrt{x}) \frac{1}{2\sqrt{x}}
\]
χ² Distribution

\[
\frac{\partial F_X(x)}{\partial x} = f_Z(\sqrt{x}) \frac{1}{2\sqrt{x}} + f_Z(-\sqrt{x}) \frac{1}{2\sqrt{x}}
\]
\( \chi^2 \) Distribution

\[
\frac{\partial F_X(x)}{\partial x} = f_Z(\sqrt{x}) \frac{1}{2\sqrt{x}} + f_Z(-\sqrt{x}) \frac{1}{2\sqrt{x}}
\]

\[
= \frac{1}{\sqrt{x}} \frac{1}{2\sqrt{2\pi}} \left(2e^{-\frac{x}{2}}\right)
\]
\( \chi^2 \) Distribution

\[
\frac{\partial F_X(x)}{\partial x} = f_Z(\sqrt{x}) \frac{1}{2\sqrt{x}} + f_Z(-\sqrt{x}) \frac{1}{2\sqrt{x}}
\]

\[
= \frac{1}{\sqrt{x}} \frac{1}{2\sqrt{2\pi}} (2e^{-\frac{x}{2}})
\]

\[
= \frac{1}{\sqrt{x}} \frac{1}{\sqrt{2\pi}} (e^{-\frac{x}{2}})
\]
\[ \frac{\partial F_X(x)}{\partial x} = f_Z(\sqrt{x}) \frac{1}{2\sqrt{x}} + f_Z(-\sqrt{x}) \frac{1}{2\sqrt{x}} \]

\[ = \frac{1}{\sqrt{x}} \frac{1}{2\sqrt{2\pi}} (2e^{-\frac{x}{2}}) \]

\[ = \frac{1}{\sqrt{x}} \frac{1}{\sqrt{2\pi}} (e^{-\frac{x}{2}}) \]

\[ = \frac{1}{\sqrt{\pi}} (e^{-\frac{x}{2}}) \]

\[ = \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} \left( x^{1/2-1} e^{-\frac{x}{2}} \right) \]
\( \chi^2 \) Distribution

\[
\frac{\partial F_X(x)}{\partial x} = f_Z(\sqrt{x}) \frac{1}{2\sqrt{x}} + f_Z(-\sqrt{x}) \frac{1}{2\sqrt{x}}
\]

\[
= \frac{1}{\sqrt{x}} \frac{1}{2\sqrt{2\pi}} (2e^{-\frac{x}{2}})
\]

\[
= \frac{1}{\sqrt{x}} \frac{1}{\sqrt{2\pi}} (e^{-\frac{x}{2}})
\]

\[
= \frac{1}{\sqrt{x}} \frac{1}{\sqrt{2\pi}} \left( e^{-\frac{x}{2}} \right)
\]

\[
= \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} \left( x^{1/2-1} e^{-\frac{x}{2}} \right)
\]

\( X \sim \text{Gamma}(1/2, 1/2) \)
$\chi^2$ Distribution

$$\frac{\partial F_X(x)}{\partial x} = f_Z(\sqrt{x}) \frac{1}{2\sqrt{x}} + f_Z(-\sqrt{x}) \frac{1}{2\sqrt{x}}$$

$$= \frac{1}{\sqrt{x}} \frac{1}{2\sqrt{2\pi}} (2e^{-\frac{x}{2}})$$

$$= \frac{1}{\sqrt{x}} \frac{1}{\sqrt{2\pi}} (e^{-\frac{x}{2}})$$

$$= \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} \left( x^{1/2-1} e^{-\frac{x}{2}} \right)$$

$X \sim \text{Gamma}(1/2, 1/2)$

Then if $X = \sum_{i=1}^{N} Z^2$
\[ \frac{\partial F_X(x)}{\partial x} = f_Z(\sqrt{x}) \frac{1}{2\sqrt{x}} + f_Z(-\sqrt{x}) \frac{1}{2\sqrt{x}} \]

\[ = \frac{1}{\sqrt{x}} \frac{1}{2\sqrt{2\pi}} (2e^{-\frac{x}{2}}) \]

\[ = \frac{1}{\sqrt{x}} \frac{1}{\sqrt{2\pi}} (e^{-\frac{x}{2}}) \]

\[ = \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} \left( x^{1/2-1} e^{-\frac{x}{2}} \right) \]

\[ X \sim \text{Gamma}(1/2, 1/2) \]

Then if \( X = \sum_{i=1}^{N} Z^2 \)

\[ X \sim \text{Gamma}(n/2, 1/2) \]
Definition

Suppose $X$ is a continuous random variable with $X \geq 0$, with pdf

$$f(x) = \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2}$$

Then we will say $X$ is a $\chi^2$ distribution with $n$ degrees of freedom. Equivalently,

$$X \sim \chi^2(n)$$
Chi-Squared 1 Degrees of Freedom

$\chi^2$ distribution with 1 degree of freedom.
Chi-Squared 21 Degrees of Freedom
Chi-Squared 41 Degrees of Freedom

x

f(x)

0 50 100 150 200
0.00 0.01 0.02 0.03 0.04

Chi-Squared 41 Degrees of Freedom

Justin Grimmer (Stanford University)
Chi-Squared 51 Degrees of Freedom

![Chi-Squared Distribution](image-url)
Chi-Squared 61 Degrees of Freedom
Chi-Squared 91 Degrees of Freedom

![Chi-Squared Distribution Graph](image)

Justin Grimmer (Stanford University)
Methodology I
September 14th, 2015 38 / 45
\( \chi^2 \) Properties

Suppose \( X \sim \chi^2(n) \)

\[
E[X] = E \left[ \sum_{i=1}^{N} Z_i^2 \right] = \sum_{i=1}^{N} E[Z_i^2]
\]

\[
\text{var}(Z_i) = E[Z_i^2] - E[Z_i]^2 = E[Z_i^2] - 0 = E[Z_i^2]
\]

\[
E[X] = n
\]
\[ \chi^2 \text{ Properties} \]

\[ \text{var}(X) = \sum_{i=1}^{N} \text{var}(Z_i^2) \]

\[ = \sum_{i=1}^{N} (E[Z_i^4] - E[Z_i]^2) \]

\[ = \sum_{i=1}^{N} (3 - 1) = 2n \]

We will use the \( \chi^2 \) in 350a, 350b, and across statistics.
Student’s $t$-Distribution

Definition

Suppose $Z \sim \text{Normal}(0, 1)$ and $U \sim \chi^2(n)$. Define the random variable $Y$ as,

$$Y = \frac{Z}{\sqrt{\frac{U}{n}}}$$

If $Z$ and $U$ are independent then $Y \sim t(n)$, with pdf

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

We will use the $t$-distribution extensively for test-statistics
Degrees of Freedom 1

\[ x \quad \text{Density} \]

\begin{tabular}{cccccc}
-6 & -4 & -2 & 0 & 2 & 4 & 6 \\
0.0 & 0.1 & 0.2 & 0.3 & 0.4 & \\
\end{tabular}

Justin Grimmer (Stanford University)
Degrees of Freedom 3
Degrees of Freedom 4

Density

Degrees of Freedom 4

x

Density

Justin Grimmer (Stanford University)
Degrees of Freedom 6

Density

x

Justin Grimmer (Stanford University)
Degrees of Freedom 7
Degrees of Freedom 8

x

Density

Justin Grimmer (Stanford University)
Degrees of Freedom 9

![Graph of a density function with degrees of freedom 9. The x-axis ranges from -6 to 6, and the density values range from 0 to 0.4. The graph shows a bell-shaped curve centered around 0.]
Degrees of Freedom 10
Degrees of Freedom 11

Density

Degrees of Freedom 11

x

Density

Justin Grimmer (Stanford University)
Degrees of Freedom 18

![Density plot with degrees of freedom 18](image)
Degrees of Freedom 20

Density

x

Justin Grimmer (Stanford University)
Methodology I
September 14th, 2015 42 / 45
Degrees of Freedom 21
Degrees of Freedom 22

-6 -4 -2 0 2 4 6
0.0 0.1 0.2 0.3 0.4

Justin Grimmer (Stanford University)
Methodology I
September 14th, 2015 42 / 45
Student’s $t$-Distribution, Properties

Suppose $n = 1$, Cauchy distribution
Suppose $n = 1$, Cauchy distribution

If $X \sim \text{Cauchy}(1)$, then:

$E[X] = \text{undefined}$

$\text{var}(X) = \text{undefined}$

If $X \sim t(2)$

$E[X] = 0$

$\text{var}(X) = \text{undefined}$
Suppose $n > 2$, then

$$\text{var}(X) = \frac{n}{n-2}$$

As $n \to \infty$ \text{var}(X) \to 1.$
Tomorrow: Joint Distributions and Multivariate Normal Distribution