Math Camp

Justin Grimmer

Professor
Department of Political Science
Stanford University

September 17th, 2018

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

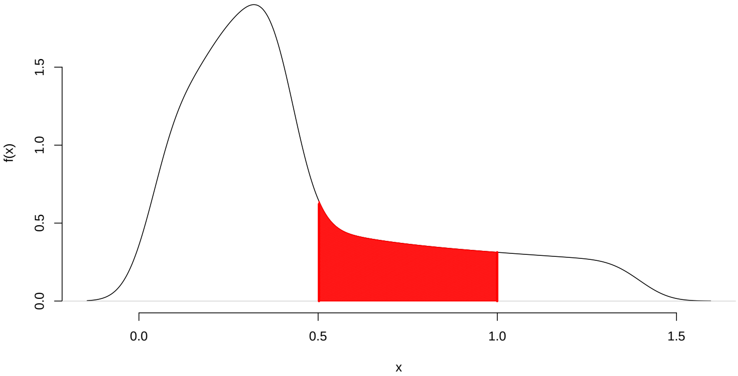
Continuous Random Variables:

- 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 $f(x)$



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 $x \in \Re$ 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

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

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 distribution of continuous random variable.

Equivalently, Cumulative distribution function characterizes continuous random variables.

Definition

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

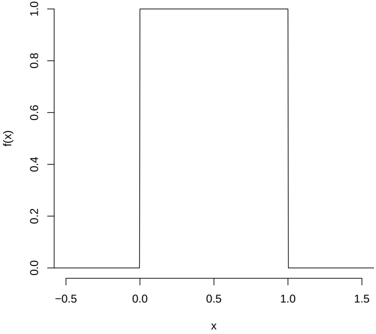
$$F(t) \quad = \quad P(X \leq t) = \int_{-\infty}^{t} f(x)dx$$

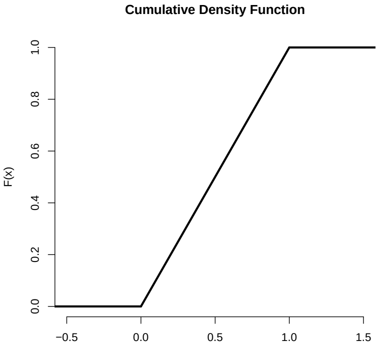
$$\begin{tikzcd}
\text{pdf} \arrow[r, bend left, "\text{Integrate}"] & \text{cdf} \arrow[l, bend left, "\text{Differentiate}"]
\end{tikzcd}$$

Uniform Random Variable

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

$$\begin{aligned} F(t) &= P(X \leq t) \\ &= 0, \text{ if } t < 0 \\ &= 1, \text{ if } t > 1 \\ &= t, \text{ if } t \in [0, 1] \end{aligned}$$





Expectation With Continuous Random Variables

Definition

If X is a continuous random variable then,

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

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

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_{-\infty}^{0} x0dx + \int_{0}^{1} x1dx + \int_{1}^{\infty} x0dx \\
&= 0 + \frac{x^2}{2}|_0^1 + 0 \\
&= 0 + \frac{1}{2} + 0 \\
&= \frac{1}{2}
\end{aligned}
$$

Expectations of Functions

Proposition

Suppose X is a continuous random variable and $g : \Re \rightarrow \Re$ (that isn't crazy). Then,

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

Expectations of Functions

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

$$
\begin{aligned}
E[g(X)] &= \int_{-\infty}^{\infty} g(x)f(x)dx \\
&= \int_0^1 x^2 dx \\
&= \frac{x^3}{3}|_0^1 \\
&= \frac{1}{3}
\end{aligned}
$$

Corollary

Suppose X is a continuous random variable. Then,

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

Proof.

$$\begin{aligned}
E[aX + b] &= \int_{-\infty}^{\infty} (ax + b) f(x) dx \\
&= a \int_{-\infty}^{\infty} x f(x) dx + b \int_{-\infty}^{\infty} f(x) dx \\
&= aE[X] + b \times 1
\end{aligned}$$

□

Definition

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

$$\begin{aligned} Var(X) &= E[(X - E[X])^2] \\ &= \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx \\ &= E[X^2] - E[X]^2 \end{aligned}$$

Variance: Random Variable

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

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

$$
\begin{aligned}
Var(X) &= E[X^2] - E[X]^2 \\
&= \frac{1}{3} - \frac{1}{4} = \frac{1}{12}
\end{aligned}
$$

Famous Continuous Distributions

- Normal Distribution
- Gamma distribution
- χ^2 Distribution
- t Distribution
- Beta, Dirichlet distributions (not today!)
- F-distribution (not today!)

Definition

Suppose X *is a random variable with* $X \in \Re$ *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* μ *and* σ^2.

Equivalently, we'll write

$$X \sim Normal(\mu, \sigma^2)$$

Support for President Obama

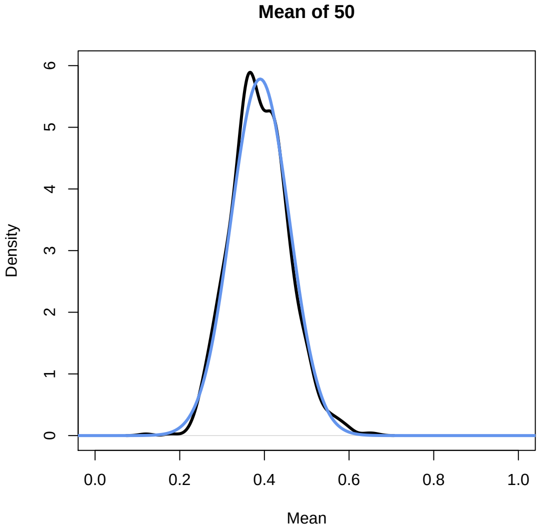
Suppose we are interested in modeling presidential approval

- 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 ($N \rightarrow \infty$)
- Then (by Central Limit Theorm) Y is Normally distributed, or

$$\begin{aligned} Y &\sim \text{Normal}(\mu, \sigma^2) \\ f(y) &= \frac{\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}} \end{aligned}$$

Central Limit Theorem

We'll prove it on Thursday.



Simulation:

Expected Value/Variance of Normal Distribution

Z is a standard normal distribution if

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

We'll call the cumulative distribution function of Z,

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

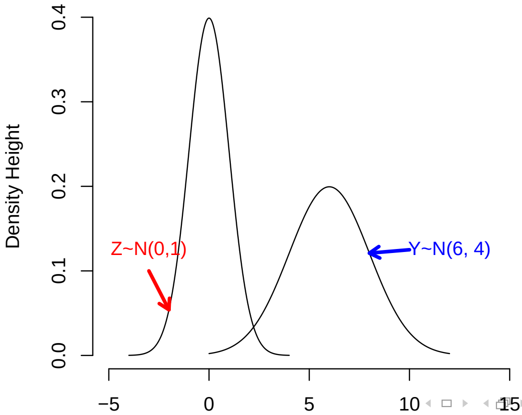
Proposition

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

$$X \sim \textit{Normal}(b, a^2)$$

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 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.

$$
\begin{aligned}
F_Y(x) &= P(Y \leq x) \\
&= P(aZ + b \leq x) \\
&= P(Z \leq \left[\frac{x-b}{a}\right]) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-b}{a}} \exp(-\frac{z^2}{2}) dz \\
&= F_Z(\frac{x-b}{a})
\end{aligned}
$$

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

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

$$
\begin{aligned}
\frac{\partial F_Y(x)}{\partial x} &= \frac{\partial F_Z(\frac{x-b}{a})}{\partial x} \\
&= f_Z(\frac{x-b}{a})\frac{1}{a} \text{ By the chain rule} \\
&= \frac{1}{\sqrt{2\pi}a}\exp\left[-\frac{\left(\frac{x-b}{a}\right)^2}{2}\right] \text{ By definition of } f_Z(x) \text{ or FTC} \\
&= \frac{1}{\sqrt{2\pi}a}\exp\left[-\frac{(x-b)^2}{2a^2}\right] \\
&= \text{Normal}(b, a^2)
\end{aligned}
$$

Expectation and Variance

Assume we know:

$$\begin{aligned} E[Z] &= 0 \\ Var(Z) &= 1 \end{aligned}$$

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

$$\begin{aligned} E[Y] &= E[\sigma Z + \mu] \\ &= \sigma E[Z] + \mu \\ &= \mu \\ Var(Y) &= Var(\sigma Z + \mu) \\ &= \sigma^2 Var(Z) + Var(\mu) \\ &= \sigma^2 + 0 \\ &= \sigma^2 \end{aligned}$$

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?) ?

$$\begin{aligned}
P(Y \geq 0.45) &= 1 - P(Y \leq 0.45) \\
&= 1 - P(0.05Z + 0.39 \leq 0.45) \\
&= 1 - P(Z \leq \frac{0.45 - 0.39}{0.05}) \\
&= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{6/5} \exp(-z^2/2) dz \\
&= 1 - F_Z(\frac{6}{5}) \\
&= 0.1150697
\end{aligned}$$

Back To Obama

Via simulation:

< code >

```
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
```

< / code >

The Gamma Function

Definition

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

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

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

Gamma Distribution

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

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

Set $X = Y/\beta$

$$\begin{aligned}
F(x) = P(X \leq x) &= P(Y/\beta \leq x) \\
&= P(Y \leq x\beta) \\
&= F_Y(x\beta) \\
\frac{\partial F_Y(x\beta)}{\partial x} &= f_Y(x\beta)\beta
\end{aligned}$$

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.

$$X \sim Gamma(\alpha, \beta)$$

Gamma Distribution

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

$$
\begin{aligned}
E[X] &= \frac{\alpha}{\beta} \\
\text{var}(X) &= \frac{\alpha}{\beta^2}
\end{aligned}
$$

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

$$
\begin{aligned}
X &\sim \text{Gamma}(1, \lambda) \\
f(x|1, \lambda) &= \lambda e^{-x\lambda}
\end{aligned}
$$

We will say

$$
X \sim \text{Exponential}(\lambda)
$$

Properties of Gamma Distributions

Proposition

Suppose we have a sequence of independent random variables, with

$$X_i \quad \sim \quad Gamma(\alpha_i, \beta)$$

Then

$$Y = \sum_{i=1}^{N} X_i$$

$Y \sim Gamma(\sum_{i=1}^{N} \alpha_i, \beta)$

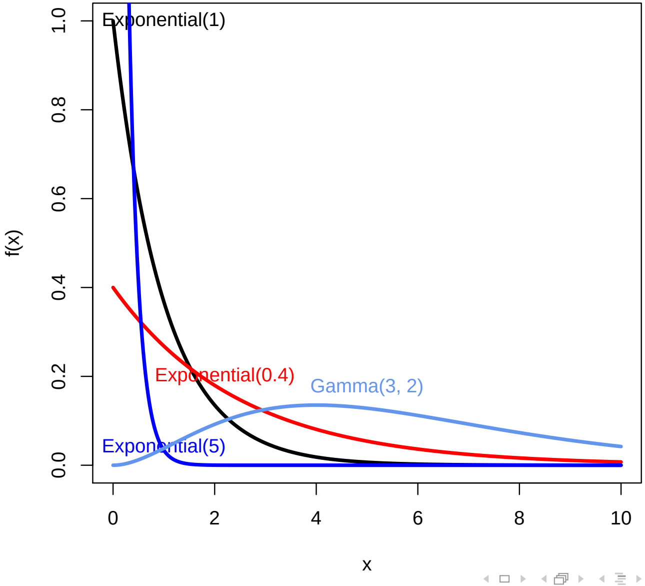
We can evaluate in `R` with `dgamma` and simulate with `rgamma`

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

```
dgamma(3, shape= 3, rate = 5)
```

and we can simulate with

```
rgamma(1000, shape = 3, rate = 5)
```


Exponential(1)
Exponential(0.4)
Gamma(3, 2)
Exponential(5)
f(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10

χ^2 Distribution

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

$$\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P(Z^2 \leq x) \\
&= P(-\sqrt{x} \leq Z \leq x) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-\frac{z^2}{2}} dz \\
&= F_Z(\sqrt{x}) - F_Z(-\sqrt{x})
\end{aligned}$$

The pdf then is

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

χ^2 Distribution

$$\begin{aligned}
\frac{\partial F_X(x)}{\partial x} &= f_Z(\sqrt{x})\frac{1}{2\sqrt{x}} + f_Z(-\sqrt{x})\frac{1}{2\sqrt{x}} \\
&= \frac{1}{\sqrt{x}}\frac{1}{2\sqrt{2\pi}}(2e^{-\frac{x}{2}}) \\
&= \frac{1}{\sqrt{x}}\frac{1}{\sqrt{2\pi}}(e^{-\frac{x}{2}}) \\
&= \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})}\left(x^{1/2-1}e^{-\frac{x}{2}}\right)
\end{aligned}$$

$X \sim \text{Gamma}(1/2, 1/2)$
Then if $X = \sum_{i=1}^{N} Z^2$
$X \sim \text{Gamma}(n/2, 1/2)$

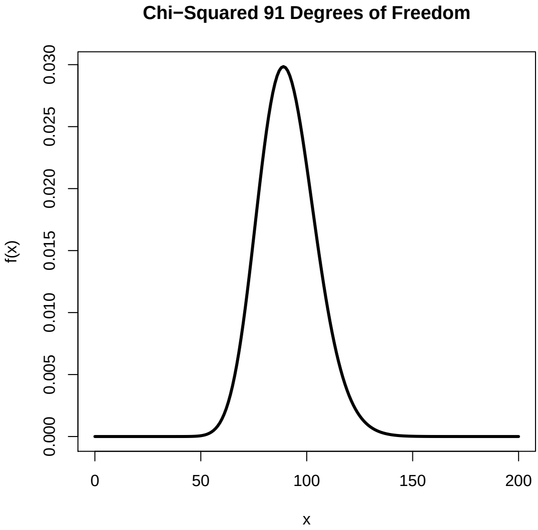
Definition

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 χ^2 distribution with n degrees of freedom. Equivalently,

$$X \;\sim\; \chi^2(n)$$


Chi−Squared 91 Degrees of Freedom
f(x)
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0
50
100
150
200
x

χ^2 Properties

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

$$
\begin{aligned}
E[X] &= E\left[\sum_{i=1}^{N} Z_i^2\right] \\
&= \sum_{i=1}^{N} E[Z_i^2] \\
\text{var}(Z_i) &= E[Z_i^2] - E[Z_i]^2 \\
1 &= E[Z_i^2] - 0 \\
E[X] &= n
\end{aligned}
$$

χ^2 Properties

$$
\begin{aligned}
\text{var}(X) &= \sum_{i=1}^{N} \text{var}(Z_i^2) \\
&= \sum_{i=1}^{N} \left(E[Z_i^4] - E[Z_i]^2 \right) \\
&= \sum_{i=1}^{N} (3 - 1) = 2n
\end{aligned}
$$

We will use the χ^2 across statistics.

Student's t-Distribution

Definition

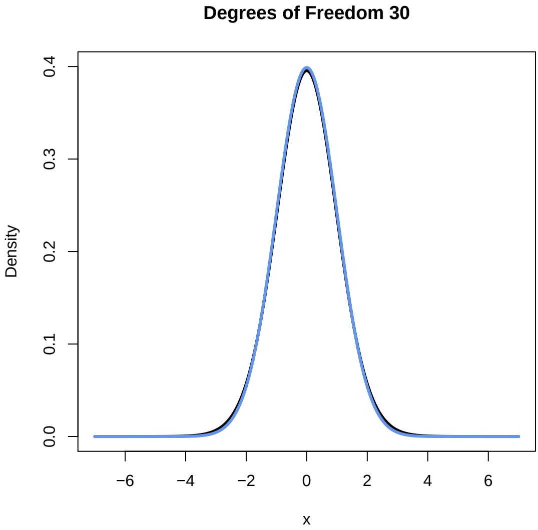
Suppose $Z \sim$ *Normal*$(0,1)$ *and* $U \sim \chi^2(n)$. *Define the random variable* Y *as,*

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

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

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

We will use the t-distribution extensively for test-statistics


Degrees of Freedom 30
Density
0.0
0.1
0.2
0.3
0.4
−6
−4
−2
0
2
4
6
x

Student's t-Distribution, Properties

Suppose $n = 1$, Cauchy distribution

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

$E[X] =$ undefined

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

If $X \sim t(2)$

$\text{E[X]} = 0$

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

Student's t-Distribution, Properties

Suppose $n > 2$, then
$var(X) = \frac{n}{n-2}$
As $n \to \infty$ var$(X) \to 1$.

Tomorrow: Joint Distributions and Multivariate Normal Distribution