Continuous Variables

Noah Arthurs
CS109, Stanford University

Review

Poisson can be used to approximate a Binomial where n is large and p is small.

Tender (Central) Moments with Poisson

- Recall: Y ~ Bin(n, p)
  - E[Y] = np
  - Var(Y) = $np(1 - p)$
- X ~ Poi(λ) where $\lambda = np$ ($n \rightarrow \infty$ and $p \rightarrow 0$)
  - E[X] = $np = \lambda$
  - Var(X) = $np(1 - p) = \lambda(1 - 0) = \lambda$
  - Yes, expectation and variance of Poisson are same

Geometric Random Variable

- X is **Geometric** Random Variable: X ~ Geo(*p*)
  - X is number of independent trials until first success
  - *p* is probability of success on each trial
  - X takes on values 1, 2, 3, …, with probability:

$$P(X = n) = (1 - p)^{n-1} p$$

$$E[X] = \frac{1}{p}$$

$$\text{Var}(X) = \frac{1 - p}{p^2}$$

Negative Binomial Random Variable

- X is **Negative Binomial** RV: X ~ NegBin(r, p)
  - X is number of independent trials until r successes
  - p is probability of success on each trial
  - X takes on values r, r + 1, r + 2…, with probability:

$$P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \text{ where } n = r, r+1, \ldots$$

  - E[X] = r/p Var(X) = $r(1 - p)/p^2$
- Note: Geo(p) ~ NegBin(1, p)

Grid of Random Variables

	number of **successes**	**trials/time** to get successes	
One trial	$X \sim \text{Ber}(p)$	$X \sim \text{Geo}(p)$	One success
	$\boldsymbol{n} = 1$	$\boldsymbol{r} = 1$	
Several trials	$X \sim \text{Bin}(n, p)$	$X \sim \text{NegBin}(r, p)$	Several successes
Interval of time	$X \sim \text{Poi}(\lambda)$	$X \sim \text{Exp}(\lambda)$	One success

Not all values are discrete

Riding the Marguerite

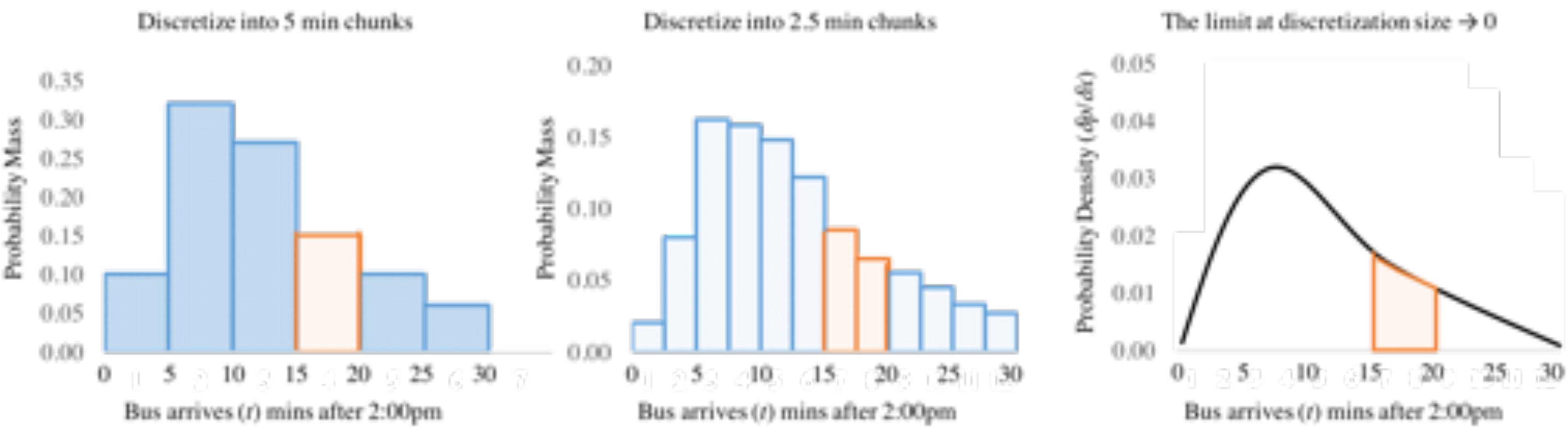
You are running to the bus stop. You don't know exactly when the bus arrives. You have a distribution of probabilities.

You show up at 2:20pm.

What is P(wait < 5 minutes)?

What is the probability that the bus arrives at: 2:17pm and 12.12333911102389234 seconds?

Discrete to Continuous



Probability Density Function

The **probability density function** (PDF) of a continuous random variable represents the relative likelihood of various values.

Units of probability *divided by units of X*.
Integrate it to get probabilities!

$$P(a < X < b) = \int_{x=a}^{b} \boxed{f(X = x)} dx$$

Probability Density Function

The **probability density function** (PDF) of a continuous random variable represents the relative likelihood of various values.

Units of probability *divided by units of X*.
Integrate it to get probabilities!

$$P(a < X < b) = \int_{x=a}^{b} \boxed{f_X(x)} \, dx$$

This is another way to write the PDF

Properties of PDFs

The integral of a PDF gives a probability. Thus:

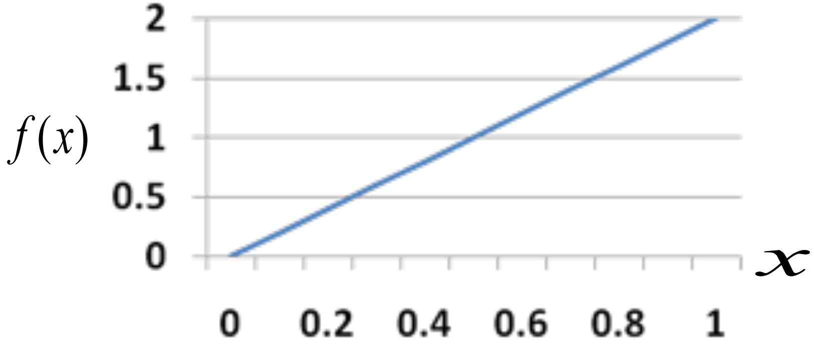
$$0 \leq \int_{x=a}^{b} f(X = x)\, dx \leq 1$$

$$\int_{x=-\infty}^{\infty} f(X = x)\, dx = 1$$

Finding Constants

- X is a continuous random variable with PDF:

$$f(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$



What about $f(x) = 3x$?

Not a valid PDF

$$\int_0^1 2x \; dx = x^2 |_0^1 = 1$$

valid PDF

$$\int_0^1 3x \; dx = \frac{3}{2} x^2 |_0^1 = \frac{3}{2}$$

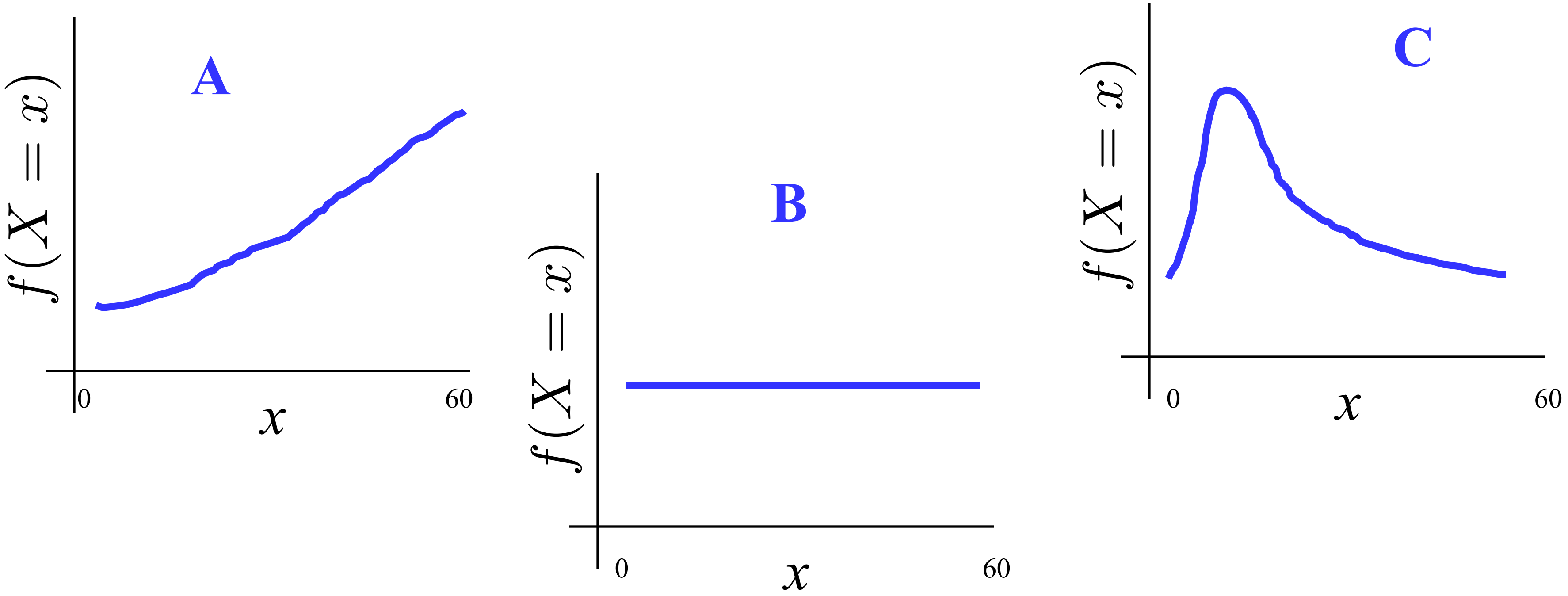
What do you get if you integrate over a probability *density* function?

A probability!

Probability Density Function

Probability density functions articulate ***relative*** belief.

Let X be a random variable which is the # of minutes after 2pm that the bus arrives at the stop:

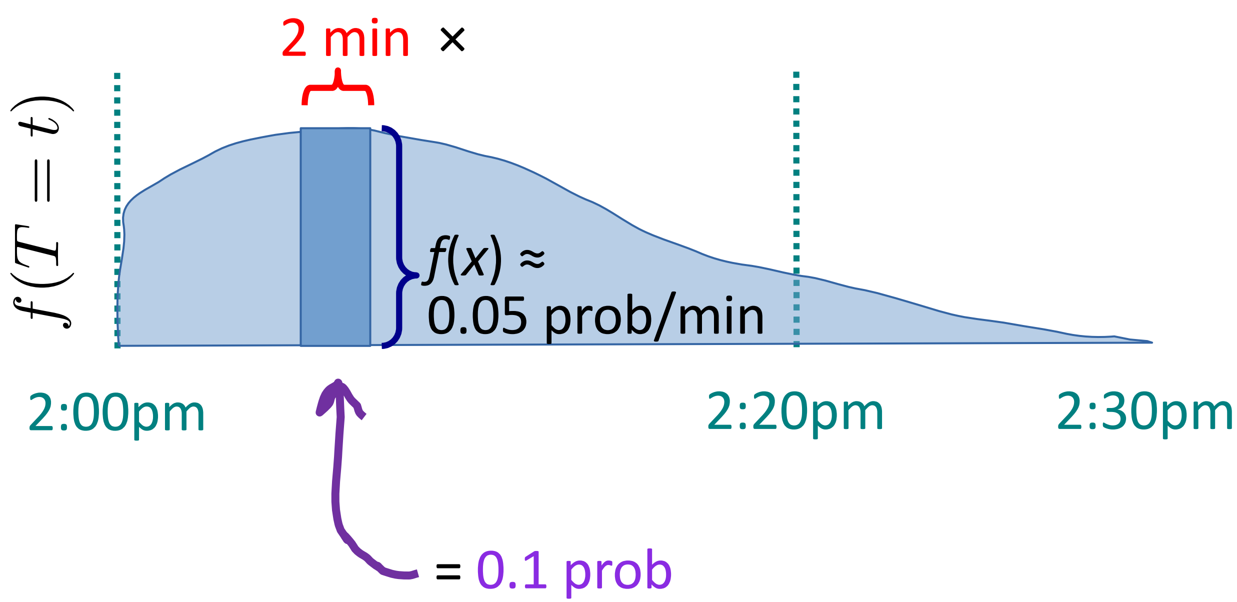


Which of these represent that you think the arrival is more likely to be close to 3:00pm

The ratio of probability densities is meaningful

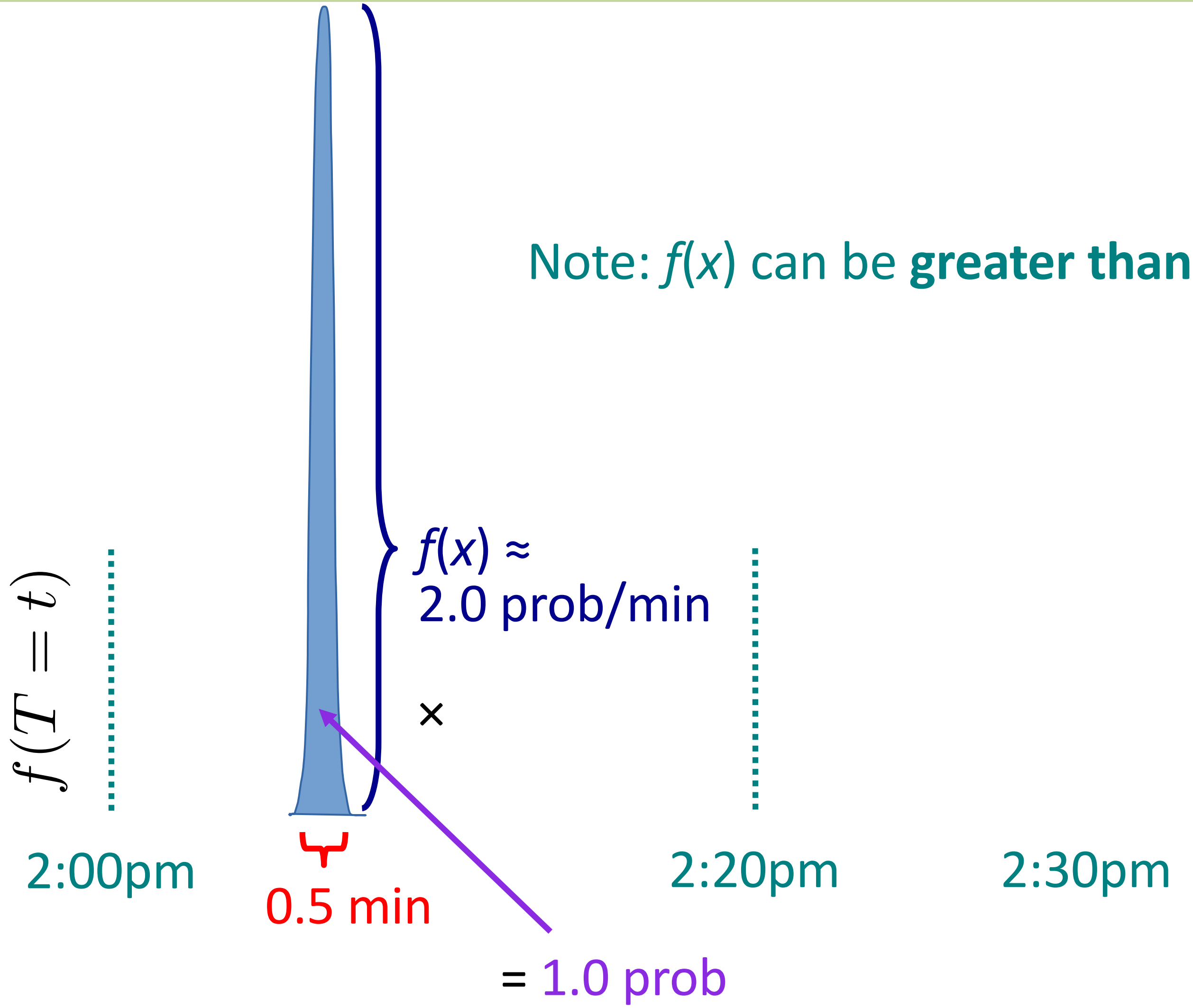
$f(X = x)$ is *Not* a Probability

Rather, it has "units" of:
probability divided by units of *X*.



$f(X = x)$ is ***Not*** a Probability

Note: $f(x)$ can be **greater than 1**!



$f(X = x)$ vs $P(X = x)$

"The probability that a **discrete** random variable X takes on the value little *x*."

$$P(X = x)$$

Aka the PMF

"The ***derivative*** of the probability that a **continuous** random variable X takes on the value little *x*."

$$f(X = x)$$

Aka the PDF

They are both measures of how ***likely*** *X is to take on the value x.*

End Review

Simple Example

Consider a random 5000x5000 matrix, where each element in the matrix is Uniform(0,1). What is the probability that a selected eigenvalue (λ) of the matrix is greater than 0?*

* We need help from Wigner

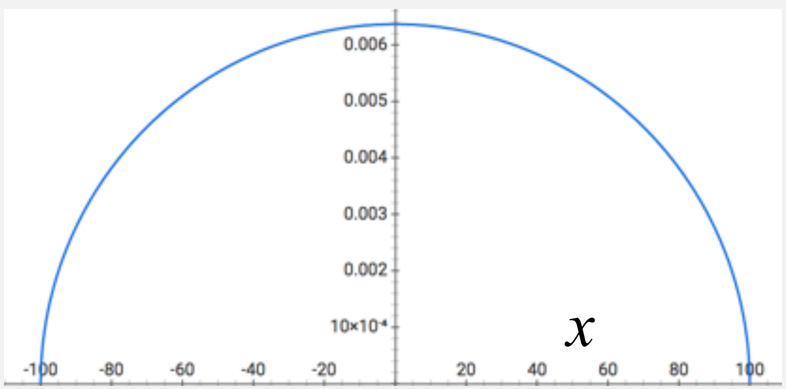
Simple Example from Quantum Physics

Let X be a continuous random variable[1]

Theory

$$f(X = x) = \frac{1}{15708}\sqrt{100^2 - x^2}$$

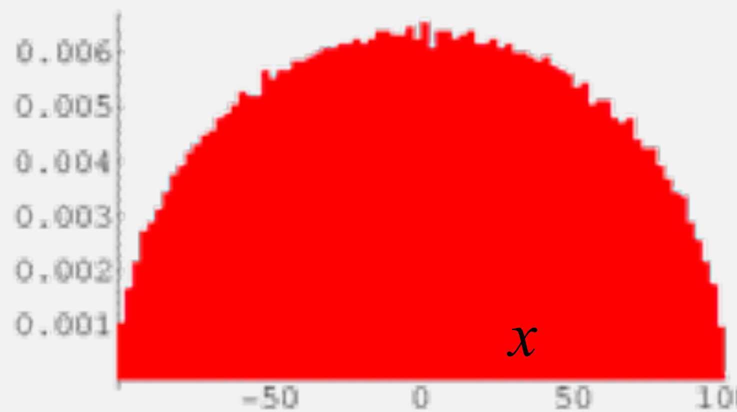
$f(X = x)$



Practice

From simulations

$P(X = x)$



$$P(X > 0) = ?$$

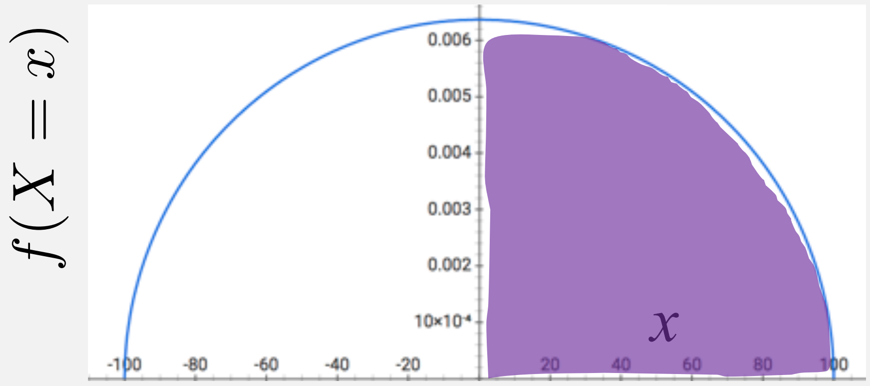
[1] X represents the eigenvalue of a 5000x5000 matrix of uniform random variables

Simple Example from Quantum Physics

Let X be a continuous random variable:

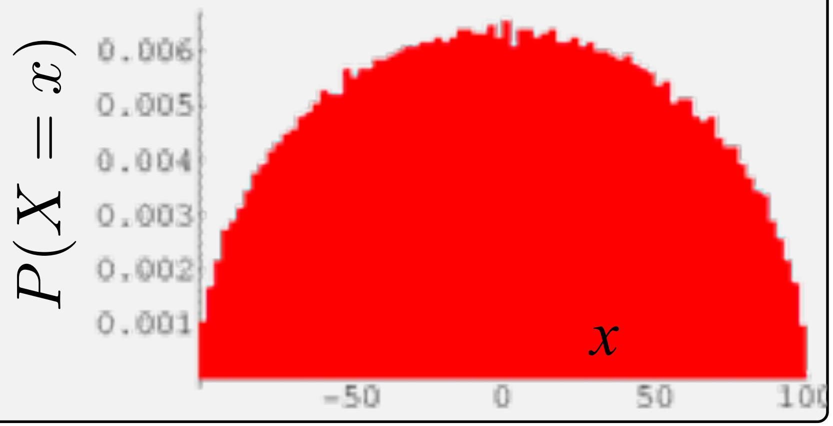
Theory

$$f(X = x) = \frac{1}{15708}\sqrt{100^2 - x^2}$$



Practice

From simulations



Approach #1: Integrate over the PDF

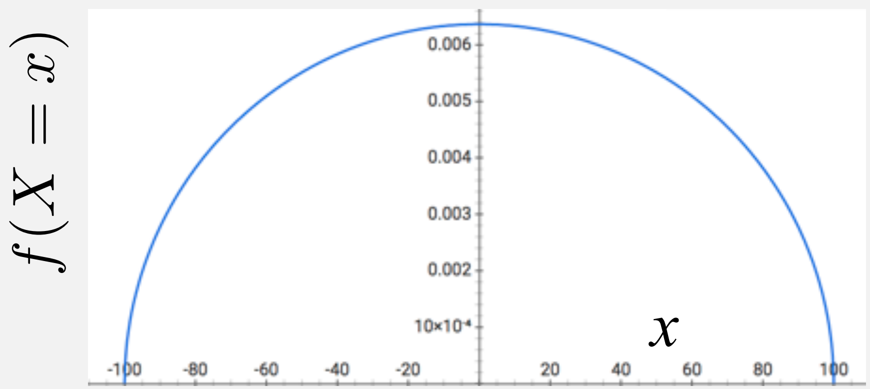
$$P(X > 0) = \int_0^{100} f(X = x)\, dx$$

Simple Example from Quantum Physics

Let X be a continuous random variable:

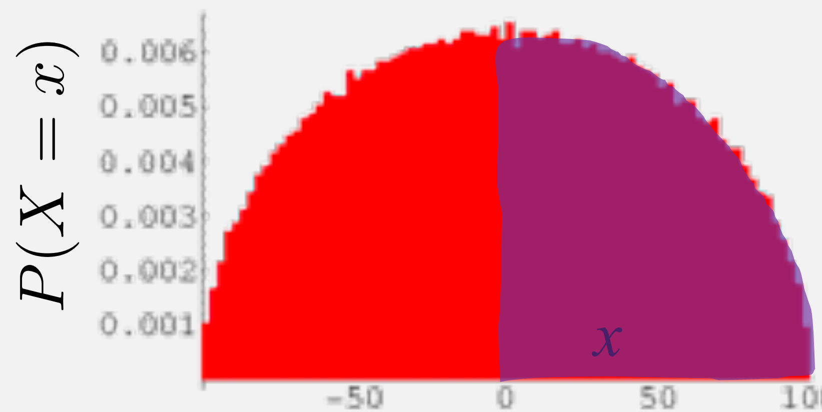
Theory

$$f(X = x) = \frac{1}{15708}\sqrt{100^2 - x^2}$$



Practice

From simulations



Approach #2: Discrete Approximation

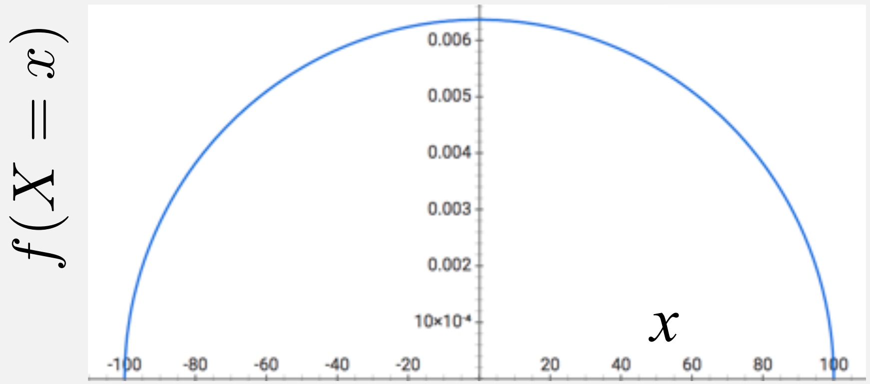
$$P(X > 0) \approx \sum_{i=0}^{100} P(X = i)$$

Simple Example from Quantum Physics

Let X be a continuous random variable:

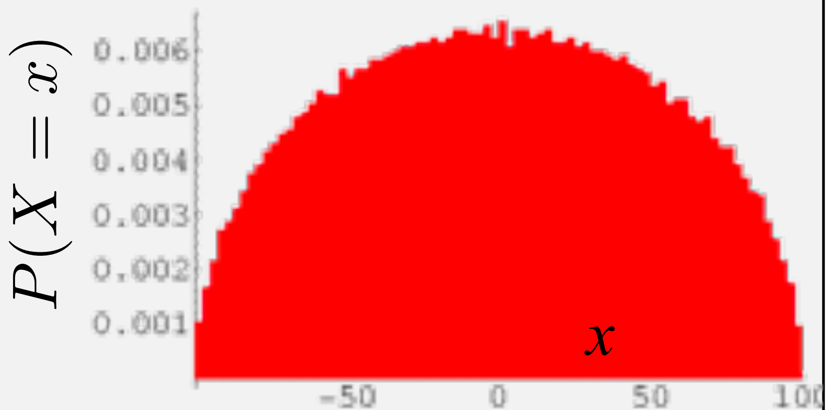
Theory

$$f(X = x) = \frac{1}{15708}\sqrt{100^2 - x^2}$$



Practice

From simulations



Approach #3: Know Semi-Circles

$$P(X > 0) = \frac{1}{2}$$

What do you get if you integrate over a probability *density* function?

A probability!

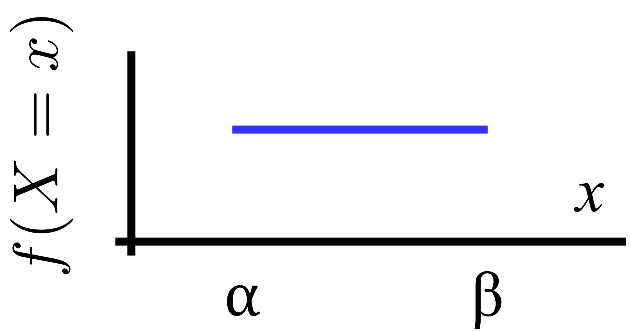
Uniform Random Variable

A **uniform** random variable is **equally likely** to be any value in an interval.

$$X \sim \mathrm{Uni}(\alpha, \beta)$$

Probability Density

$$f(X = x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha \le x \le \beta \\ 0 & \text{otherwise} \end{cases}$$



Properties

$$E[X] = \frac{\beta - \alpha}{2}$$

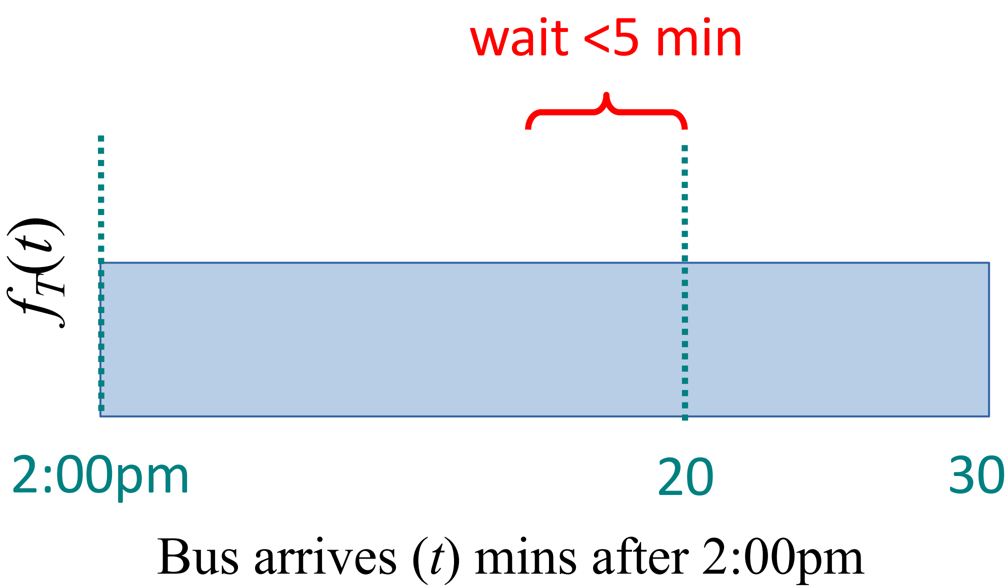
$$\mathrm{Var}(X) = \frac{(\beta - \alpha)^2}{12}$$

Uniform Bus

You are running to the bus stop. You don’t know exactly when the bus arrives. **You believe all times between 2 and 2:30 are equally likely**.

You show up at 2:15pm. What is P(wait < 5 minutes)?

$$T \sim \text{Uni}(\alpha = 0, \beta = 30)$$



$$P(\text{Wait} < 5) = \int_{15}^{20} \frac{1}{\beta - \alpha}\, dx$$

$$= \frac{x}{\beta - \alpha}\Big|_{15}^{20}$$

$$= \frac{x}{30 - 0}\Big|_{15}^{20} = \frac{5}{30}$$

Expectation and Variance

For discrete RV X:

$$E[X] = \sum_x x \cdot p(X = x)$$

$$E[g(X)] = \sum_x g(x) \cdot p(X = x)$$

$$E[X^n] = \sum_x x^n \cdot p(X = x)$$

For continuous RV X:

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

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

$$E[X^n] = \int_{-\infty}^{\infty} x^n \cdot f_X(x)$$

For both discrete and continuous RVs:

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

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

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

Expectation of Uniform

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

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

$$= \int_{\alpha}^{\beta} x \cdot \frac{1}{\beta - \alpha}\; dx$$

$$= \frac{1}{\beta - \alpha}\left[\frac{1}{2}x^2\right]_{\alpha}^{\beta}$$

$$= \frac{1}{\beta - \alpha}\left[\frac{\beta^2}{2} - \frac{\alpha^2}{2}\right]$$

$$= \frac{1}{2}\frac{1}{\beta - \alpha}(\beta + \alpha)(\beta - \alpha)$$

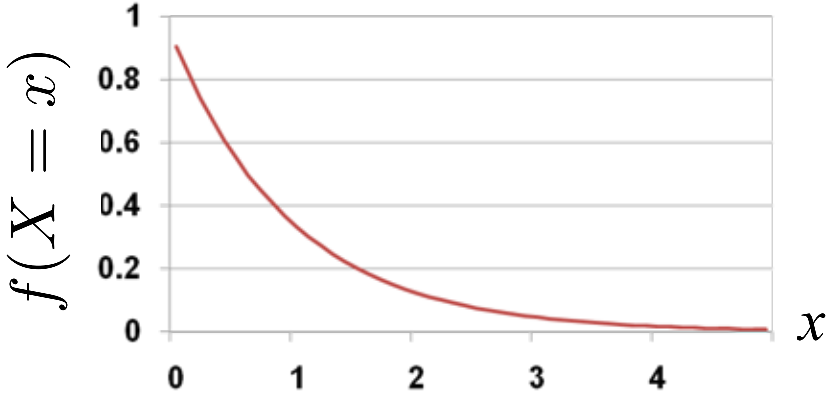
just average the start and end!

$$= \frac{1}{2}(\alpha + \beta)$$

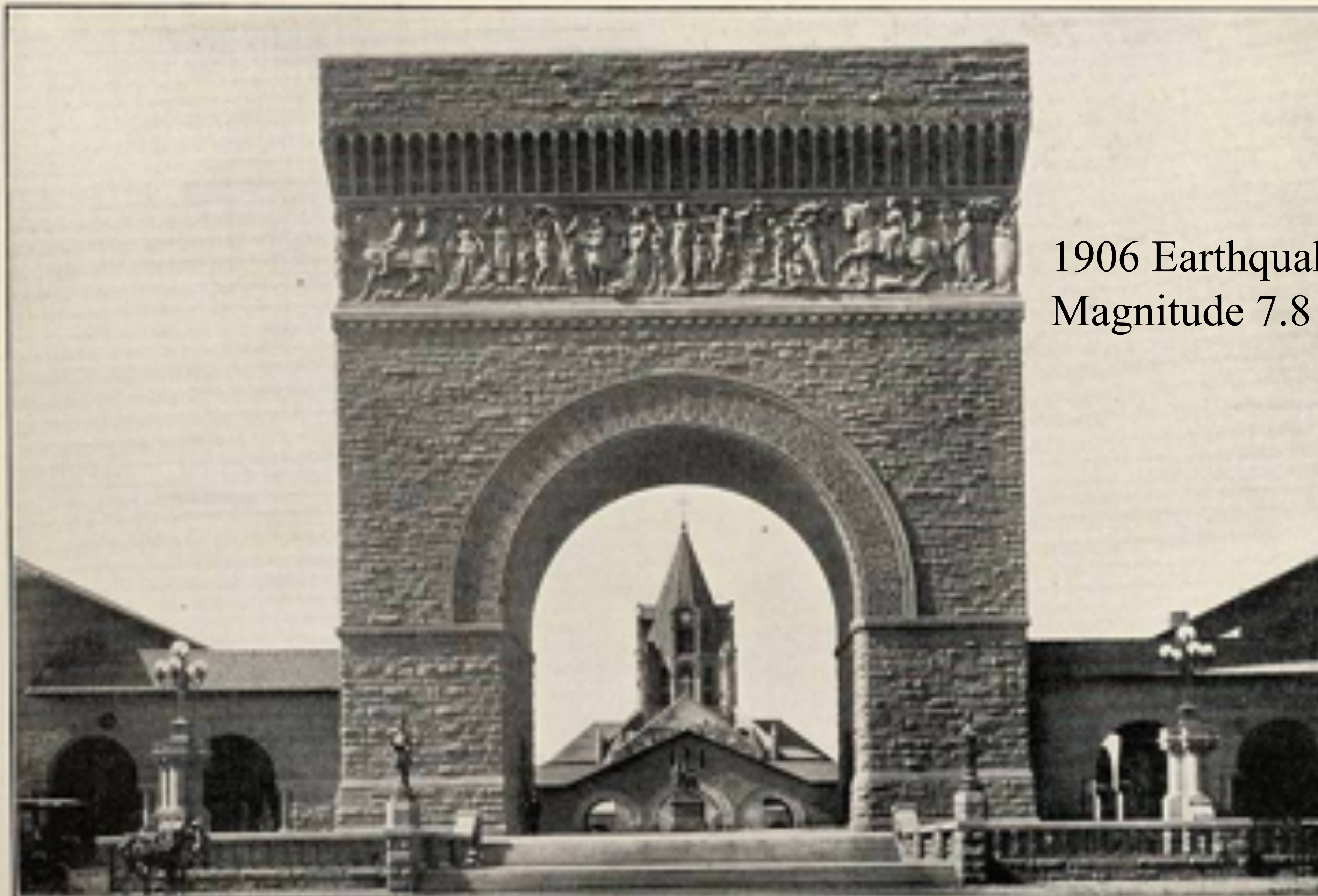
Exponential Random Variable

- X is an **Exponential RV**: $X \sim \text{Exp}(\lambda)$ Rate: $\lambda > 0$
  - Probability Density Function (PDF):

$$f(X = x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$



  - $E[X] = \dfrac{1}{\lambda}$
  - $\text{Var}(X) = \dfrac{1}{\lambda^2}$
  - Support: $0 \leq x \leq \infty$
  - Represents time until some event
    - Earthquake, request to web server, end cell phone contract, etc.



ILL. No. 65. MEMORIAL ARCH, WITH CHURCH IN BACKGROUND, STANFORD UNIVERSITY, SHOWING TYPES OF CARVED WO
WITH THE SANDSTONE.

How Many Earthquakes

Based on historical data, major earthquakes (magnitude 8.0+) happen at a **rate of 0.002** per year*. What is the probability of **zero major earthquakes magnitude next year?**

X = Number of major earthquakes next year

$$X \sim \text{Poi}(\lambda = 0.002)$$

$$P(X = 0) = \frac{\lambda^0 e^{-\lambda}}{0!} = \frac{0.002^0 e^{-0.002}}{0!} \approx 0.998$$

*In California, according to the USGS, 2015

How Long Until Next Earthquake

Based on historical data, major earthquakes (magnitude 8.0+) happen at a **rate of 0.002** per year*. What is the probability of **a major earthquake in the next 30 years?**

Y = Years until the next earthquake of magnitude 8.0+

$$Y \sim \text{Exp}(\lambda = 0.002)$$

$$f_Y(y) = \lambda e^{-\lambda y}$$
$$= 0.002^{-0.002y}$$

$$P(Y < 30) = \int_0^{30} 0.002e^{-0.002y} \, dy$$

*In California, according to the USGS, 2015

Integral Review

$$\int e^{cx} \; dx = \frac{1}{c} e^{cx}$$

How Long Until Next Earthquake

Based on historical data, major earthquakes (magnitude 8.0+) happen at a **rate of 0.002** per year*. What is the probability of **a major earthquake in the next 30 years?**

Y = Years until the next earthquake of magnitude 8.0+

$$Y \sim \mathrm{Exp}(\lambda = 0.002)$$

$$f_Y(y) = \lambda e^{-\lambda y}$$
$$= 0.002^{-0.002y}$$

$$P(Y < 30) = \int_0^{30} 0.002e^{-0.002y}\ dy$$

$$= 0.002\Big[- 500e^{-0.002y}\Big]_0^{30}$$

$$= \frac{500}{500}(-e^{-0.06} + e^0) \qquad \approx 0.06$$

*In California, according to the USGS, 2015

How Long Until Next Earthquake

Based on historical data, major earthquakes (magnitude 8.0+) happen at a **rate of 0.002** per year*. What is the **expected number of years until the next earthquake?**

Y = Years until the next earthquake of magnitude 8.0+

$$Y \sim \text{Exp}(\lambda = 0.002)$$

$$E[Y] = \frac{1}{\lambda} = \frac{1}{0.002} = 500$$

*In California, according to the USGS, 2015

How Long Until Next Earthquake

Based on historical data, major earthquakes (magnitude 8.0+) happen at a **rate of 0.002** per year*. What is the **standard deviation of years until the next earthquake?**

Y = Years until the next earthquake of magnitude 8.0+

$$Y \sim \text{Exp}(\lambda = 0.002)$$

$$\text{Var}(Y) = \frac{1}{\lambda^2} = \frac{1}{0.002^2} = 250,000 \text{ years}^2$$

$$\text{Std}(Y) = \sqrt{\text{Var}(X)} = 500 \text{ years}$$

*In California, according to the USGS, 2015

Is there a way to avoid integrals?

Cumulative Distribution Function

A cumulative density function (CDF) is a “closed form” equation for the probability that a random variable is less than a given value

$$F(x) = P(X < x)$$

If you learn how to use a cumulative density function, you can avoid integrals!

$$F_X(x)$$

This is also shorthand notation for the CDF

Cumulative Distribution Function

$$F(x) = P(X < x)$$

$$x = 2$$

$$0.03125$$

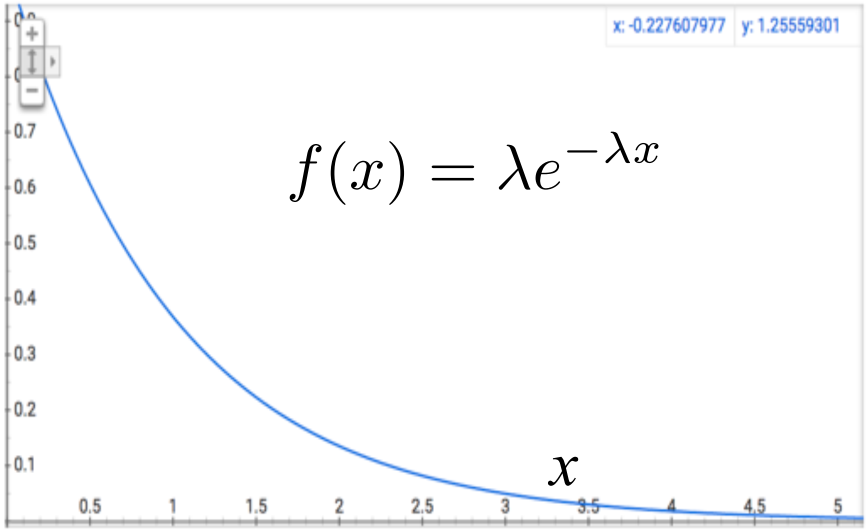
CDF of an Exponential

$$F_X(x) = 1 - e^{-\lambda x}$$

$$\begin{aligned}
P(X < x) &= \int_{y=-\infty}^{x} f(y)\, dy \\
&= \int_{y=0}^{x} \lambda e^{-\lambda y}\, dy \\
&= \frac{\lambda}{\lambda}\Big[-e^{-\lambda y}\Big]_0^x \\
&= [-e^{-\lambda x}] - [-e^{\lambda 0}] \\
&= 1 - e^{-\lambda x}
\end{aligned}$$

CDF: X ~ Exp(λ = 1)

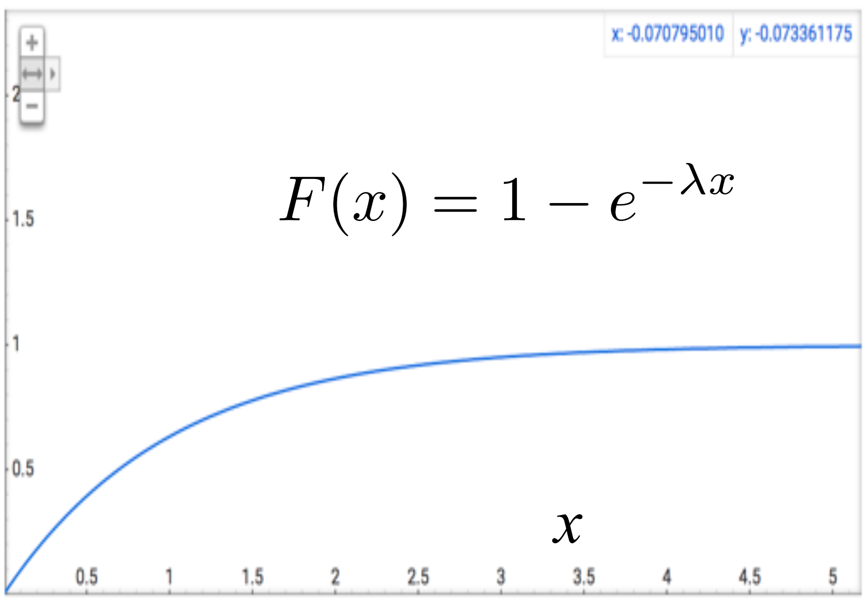
Probability
density
function



Cumulative
distribution
function

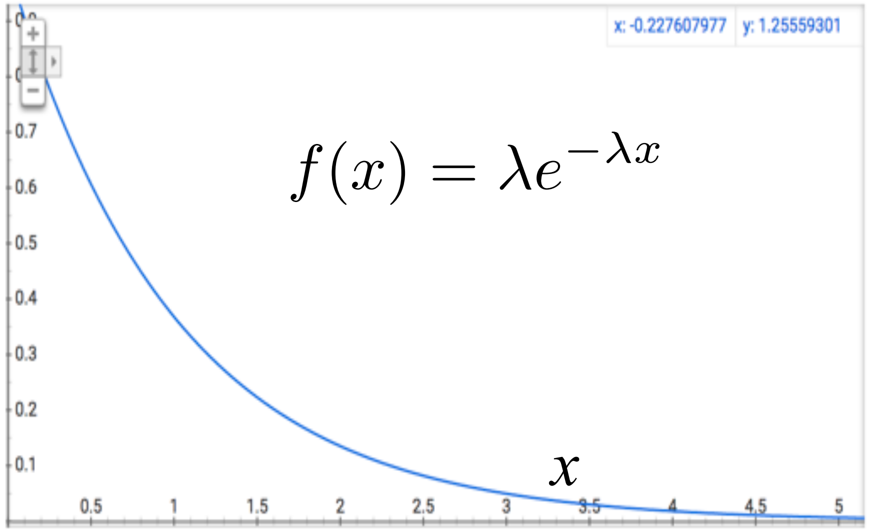
$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^{x} f(y)\ dy$$



CDF: X ~ Exp(λ = 1)

Probability density function



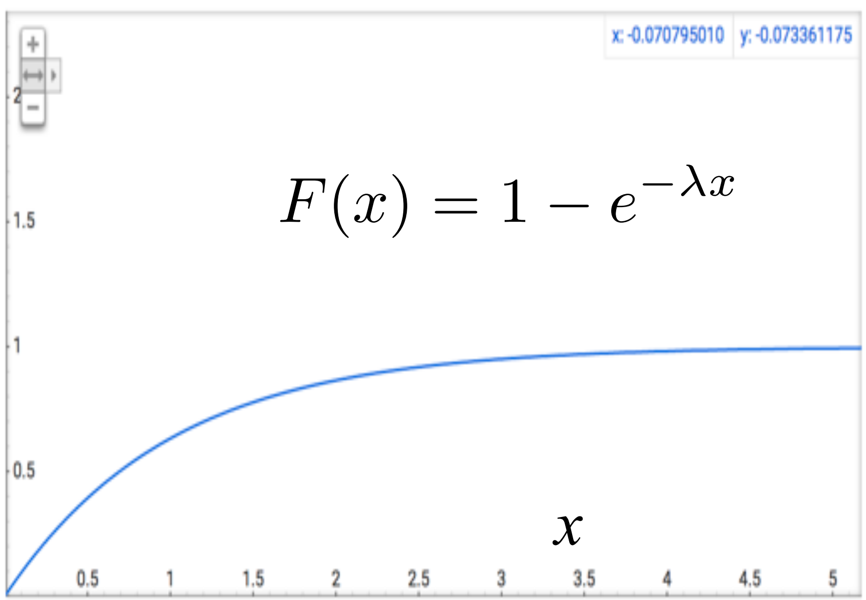
P(X < 2)

____

Cumulative distribution function

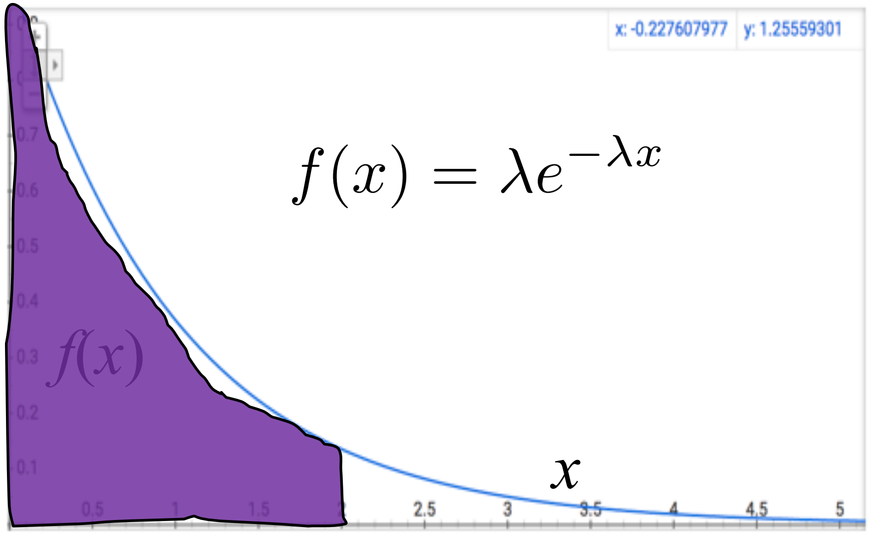
$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^{x} f(y)\, dy$$



CDF: X ~ Exp(λ = 1)

Probability density function



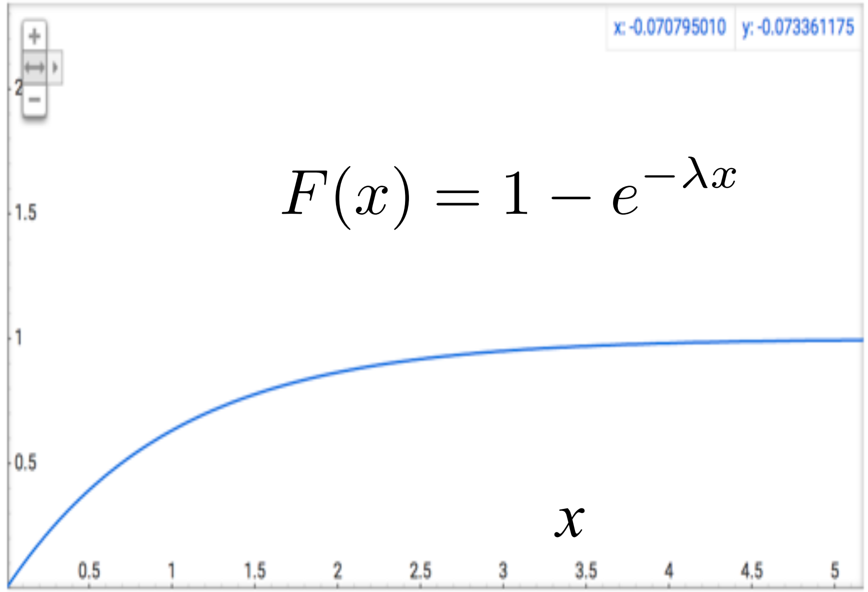
P(X < 2)

$$= \int_{x=-\infty}^{2} f(x)\ dx$$

Cumulative distribution function

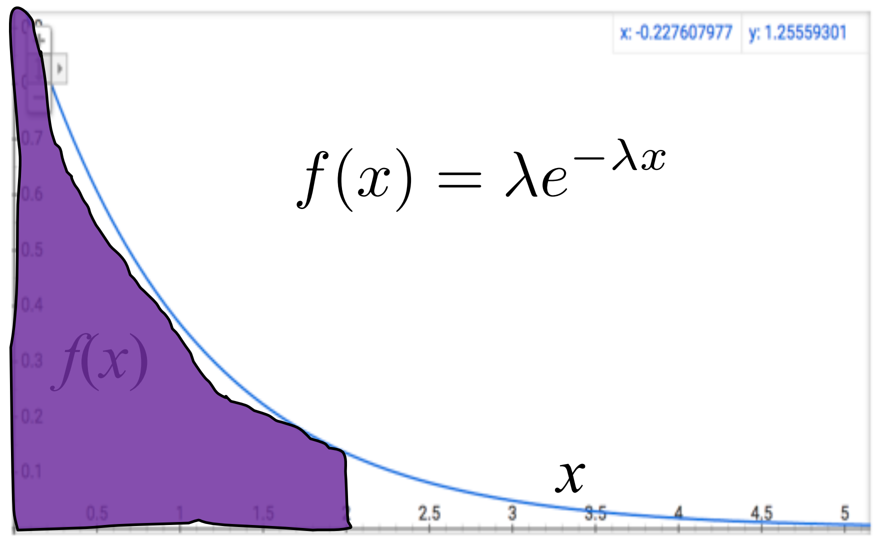
$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^{x} f(y)\ dy$$



CDF: X ~ Exp(λ = 1)

Probability density function



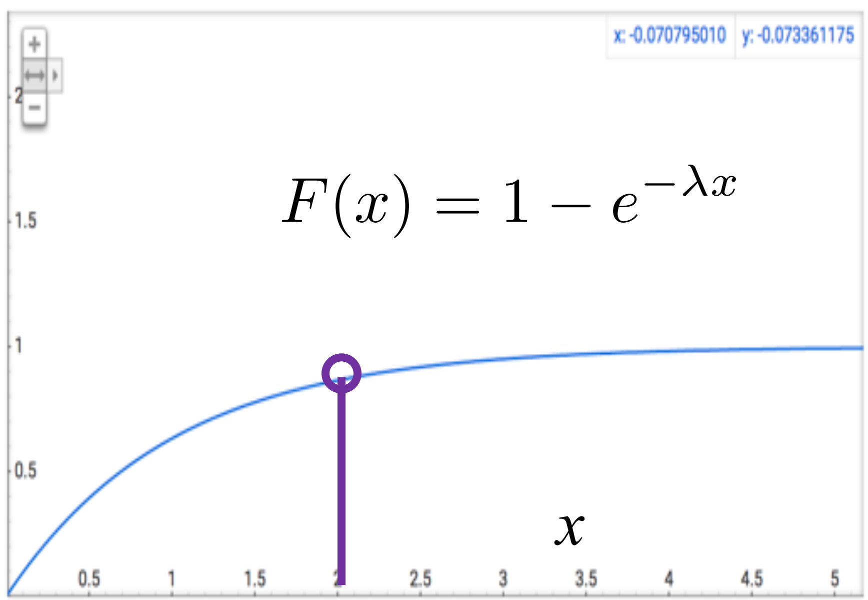
P(X < 2)

$$= \int_{x=-\infty}^{2} f(x)\ dx$$

Cumulative distribution function

$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^{x} f(y)\ dy$$



or

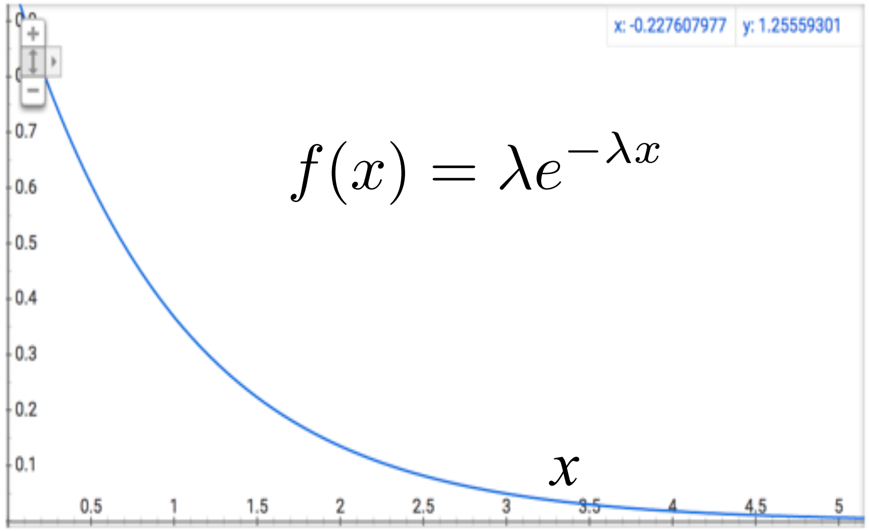
$$= F(2)$$

$$= 1 - e^{-2}$$

$$\approx 0.84$$

CDF: X ~ Exp(λ = 1)

Probability density function

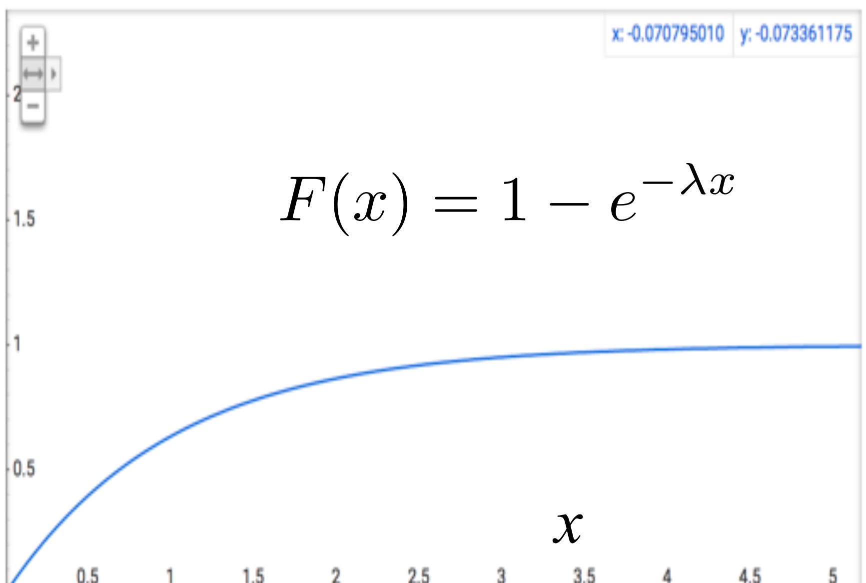


P(X > 1)

Cumulative distribution function

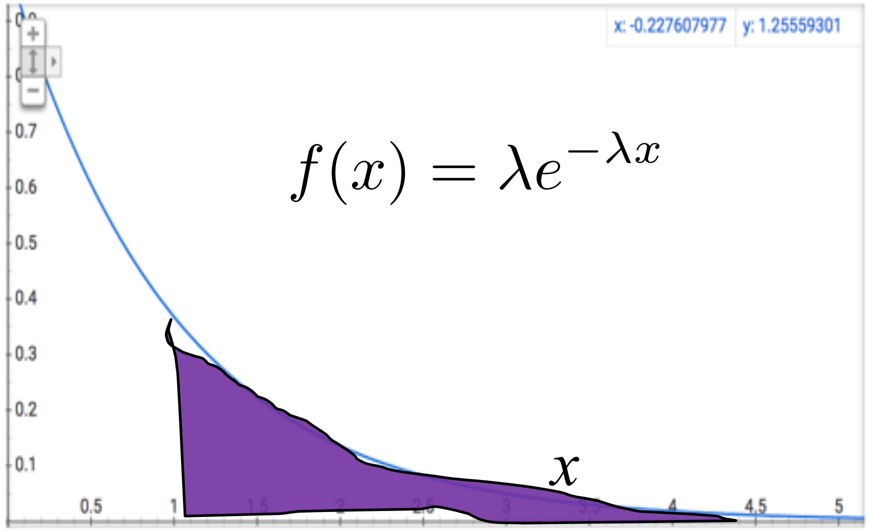
$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^{x} f(y)\, dy$$



CDF: X ~ Exp(λ = 1)

Probability density function



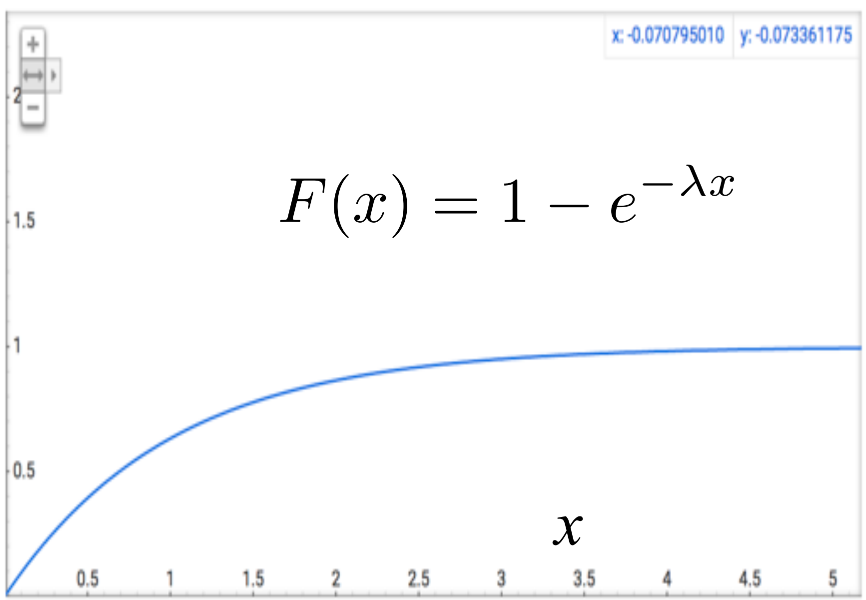
P(X > 1)

$$= \int_{x=1}^{\infty} f(x)\, dx$$

Cumulative distribution function

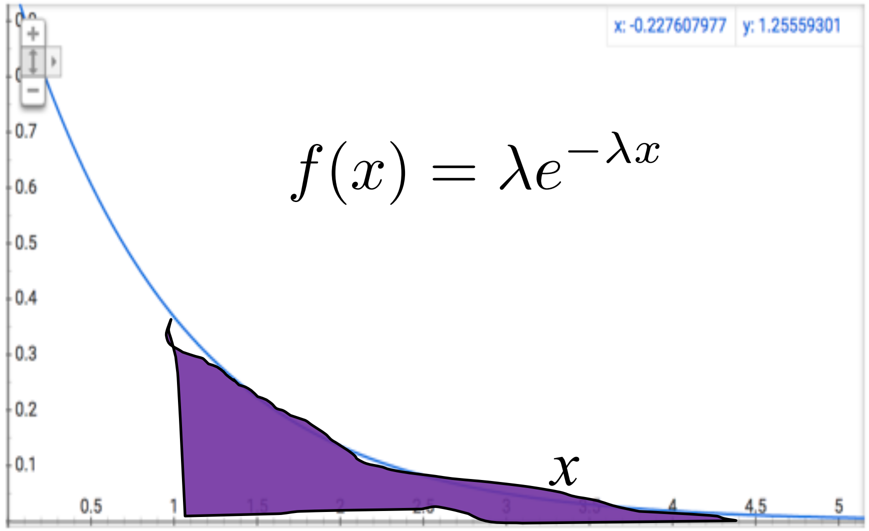
$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^{x} f(y)\, dy$$



CDF: X ~ Exp(λ = 1)

Probability density function



$$\frac{P(X > 1)}{}$$

$$= \int_{x=1}^{\infty} f(x)\, dx$$

or

$$= 1 - F(1)$$

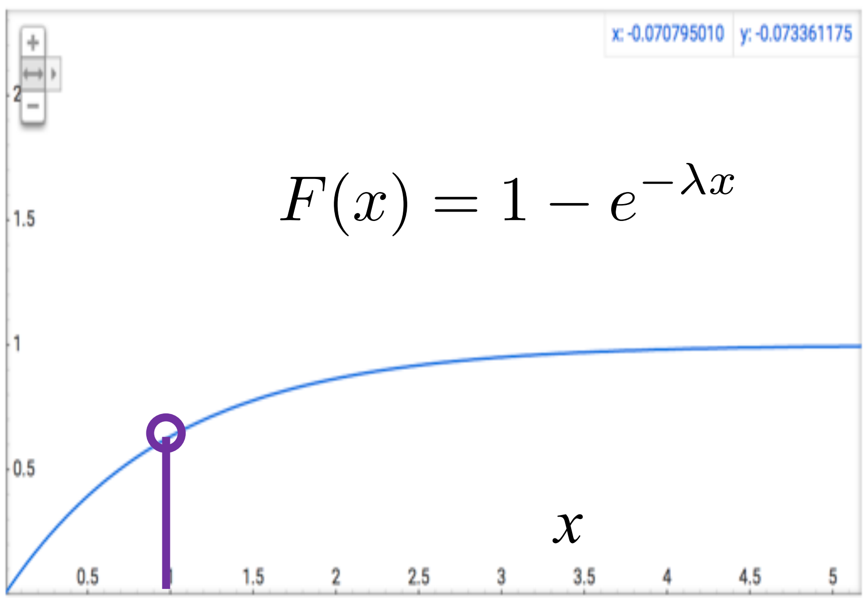
$$= e^{-1}$$

$$\approx 0.37$$

Cumulative distribution function

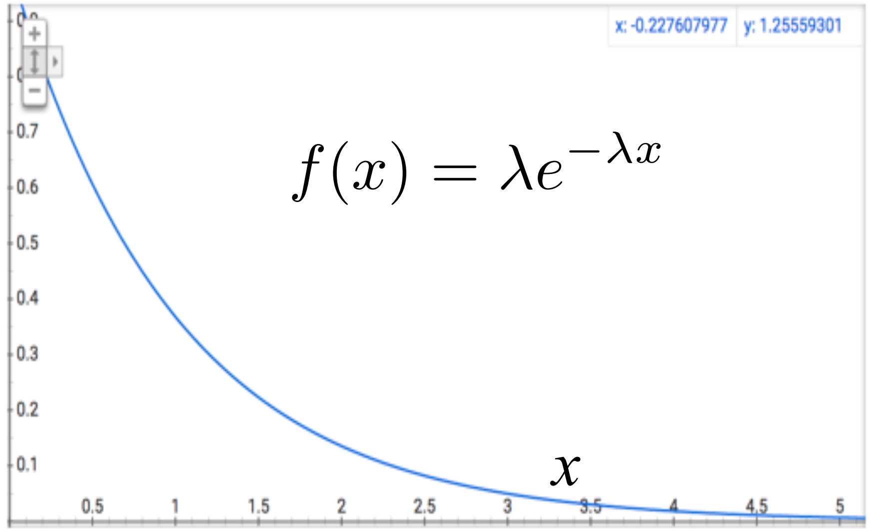
$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^{x} f(y)\, dy$$



CDF: X ~ Exp(λ = 1)

Probability density function

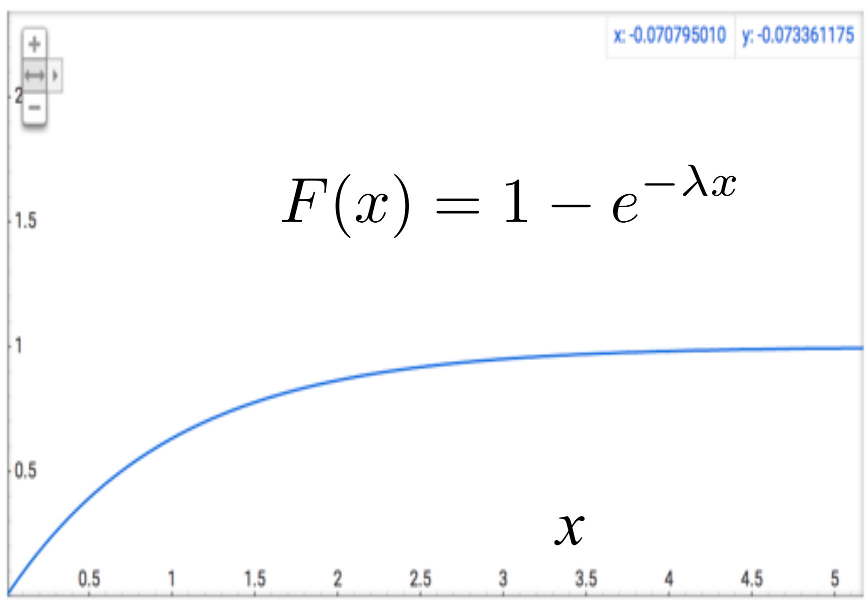


$P(1 < X < 2)$

Cumulative distribution function

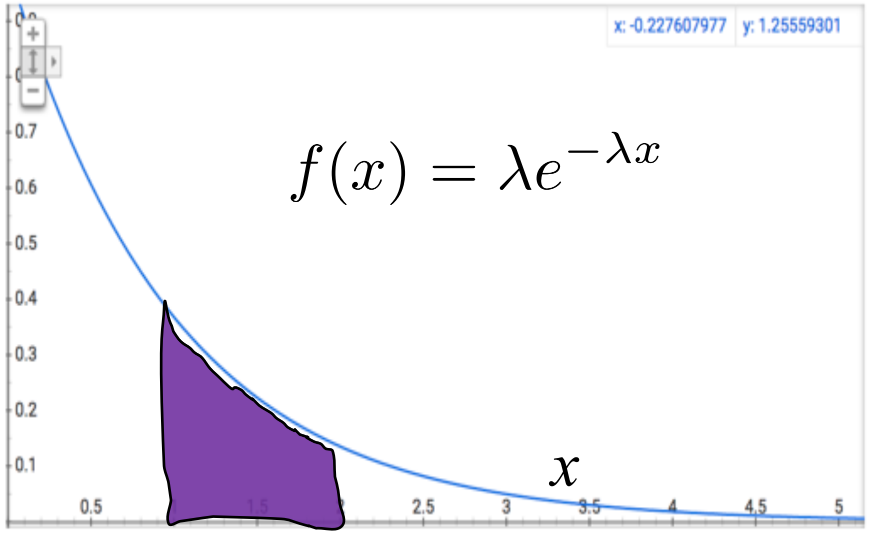
$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^{x} f(y)\, dy$$



CDF: X ~ Exp(λ = 1)

Probability density function

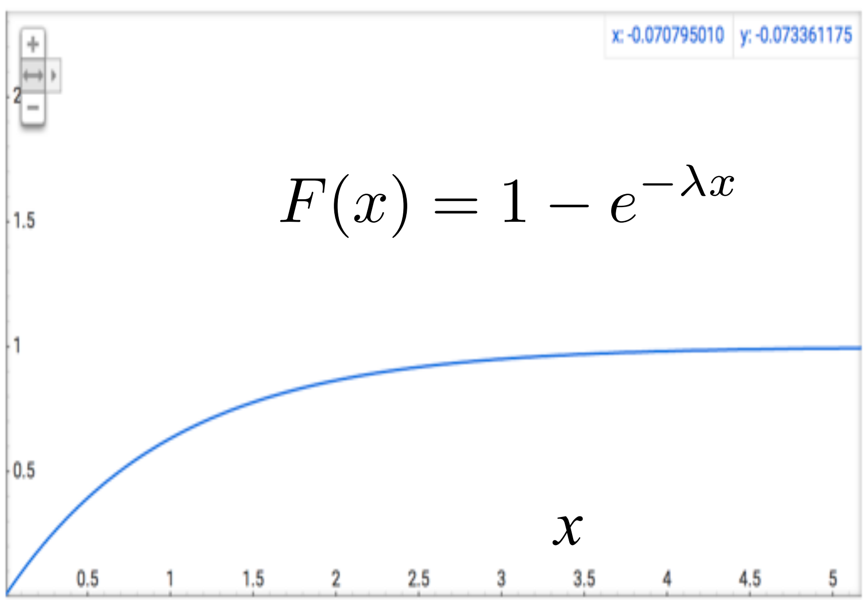


$$P(1 < X < 2) = \int_{x=1}^{2} f(x)\, dx$$

Cumulative distribution function

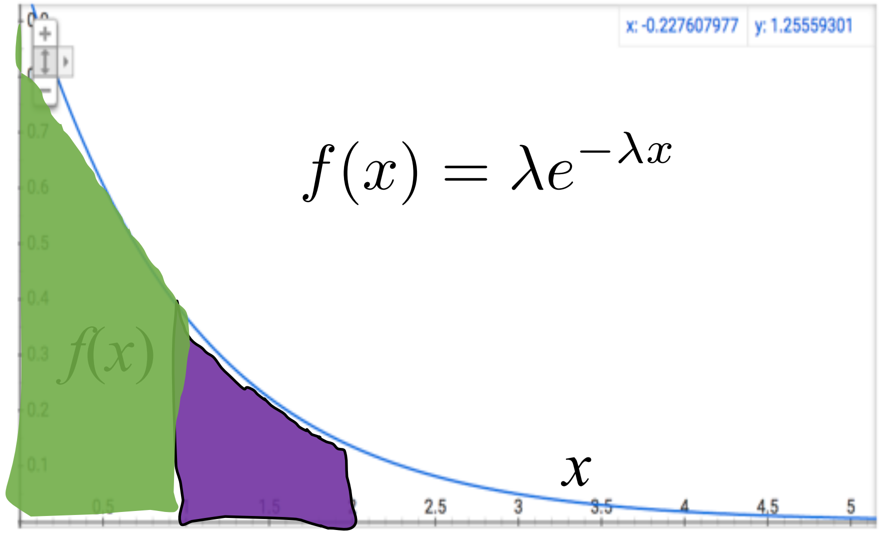
$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^{x} f(y)\, dy$$



CDF: X ~ Exp(λ = 1)

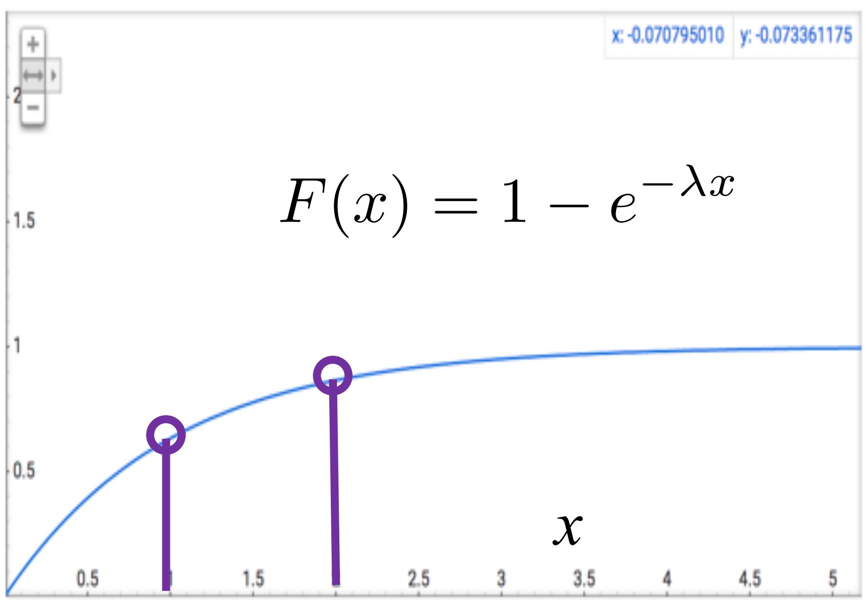
Probability density function



Cumulative distribution function

$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^{x} f(y)\, dy$$



P(1 < X < 2)

$$= \int_{x=1}^{2} f(x)\, dx$$

or

$$= F(2) - F(1)$$

$$= (1 - e^{-2}) - (1 - e^{-1})$$

$$\approx 0.23$$

Probability of Earthquake in Next 4 Years?

Based on historical data, earthquakes of magnitude 8.0+ happen at a **rate of 0.002** per year*. What is the probability of **a major earthquake in the next 4 years?**

Y = Years until the next earthquake of magnitude 8.0+

$$Y \sim \text{Exp}(\lambda = 0.002) \qquad F(y) = 1 - e^{-0.002y}$$

$$\begin{aligned} P(Y < 4) &= F(4) \\ &= 1 - e^{-0.002 \cdot 4} \\ &\approx 0.008 \end{aligned}$$

Feeling lucky?

*According to USGS, 2015

Exponential vs. Poisson

It happens that the previous problem can be solved with a Poisson as well. Let's make sure we get the same answer!

Y = Years until the next earthquake of magnitude 8.0+

X = Number of magnitude 8.0+ earthquakes in next 4 years

$$Y \sim \text{Exp}(\lambda = 0.002)$$

$$\begin{aligned} P(Y < 4) &= F(4) \\ &= 1 - e^{-0.002 \cdot 4} \\ &\approx 0.008 \end{aligned}$$

$$X \sim Poi(\lambda = 0.002 * 4)$$

$$\begin{aligned} P(X > 0) &= 1 - P(X = 0) \\ &= 1 - \frac{\lambda^0 e^{-\lambda}}{0!} \\ &= 1 - e^{-0.002 \cdot 4} \end{aligned}$$

Exponential vs. Poisson

Solvable by Exponential OR Poisson:

- *probability of a major earthquake in the next 4 years*
- *probability of no major earthquakes in the next 4 years*
- *probability that the next major earthquake is between 3 and 5 years from now* (much harder with Poisson)

Solvable by Poisson ONLY:

- *probability of at least 3 major earthquakes in the next 4 years*
- *probability of exactly 3 major earthquakes in the next 4 years*

Solvable by Exponential ONLY:

- *expected number of years until next major earthquake*
- *any problem that involves working with a belief distribution over the amount of time until the next earthquake.*

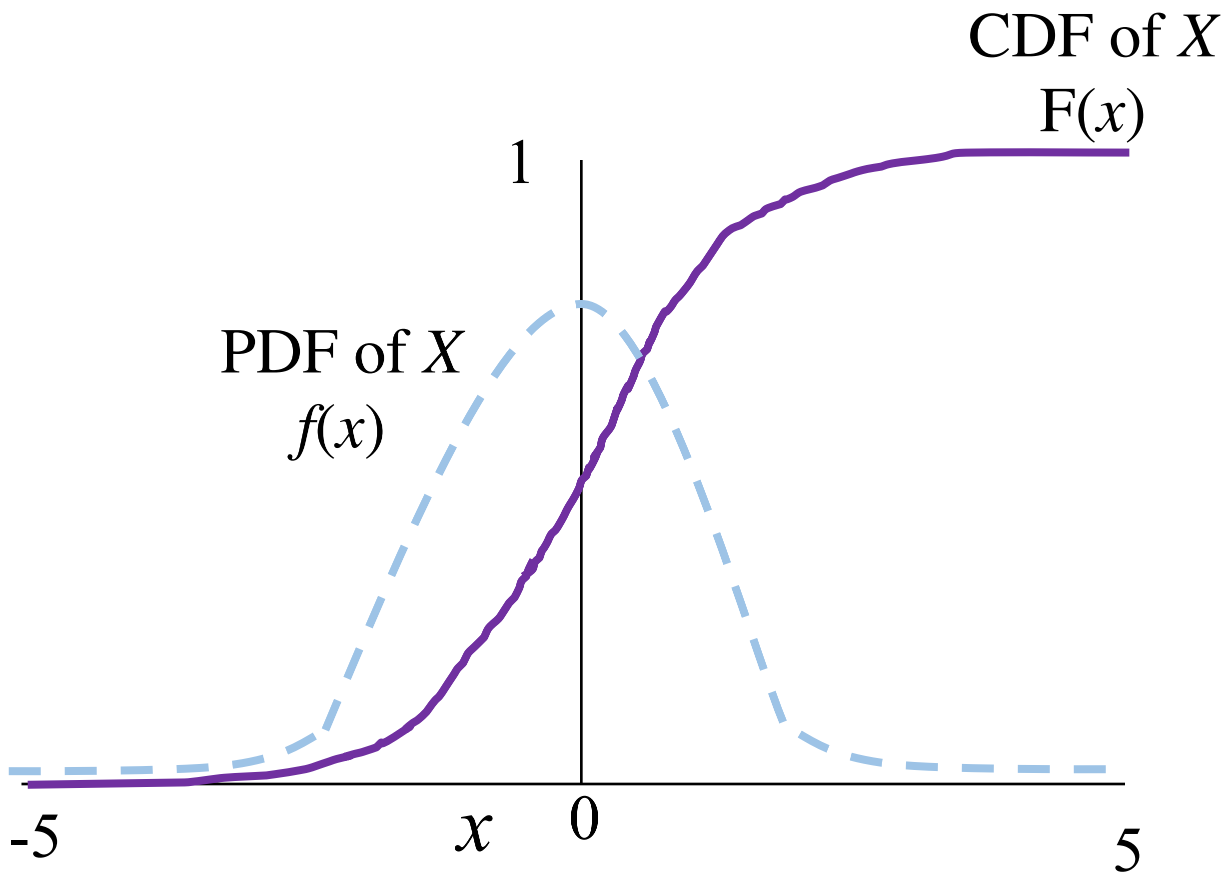
Notation

$p(a)$ or $p_X(a)$ Probability Mass Function (discrete) $P(X = a)$

$f(a)$ or $f_X(a)$ Probability Density Function (continuous) $f(X = a)$

$F(a)$ or $F_X(a)$ Cumulative Distribution Function $P(X \leq a)$

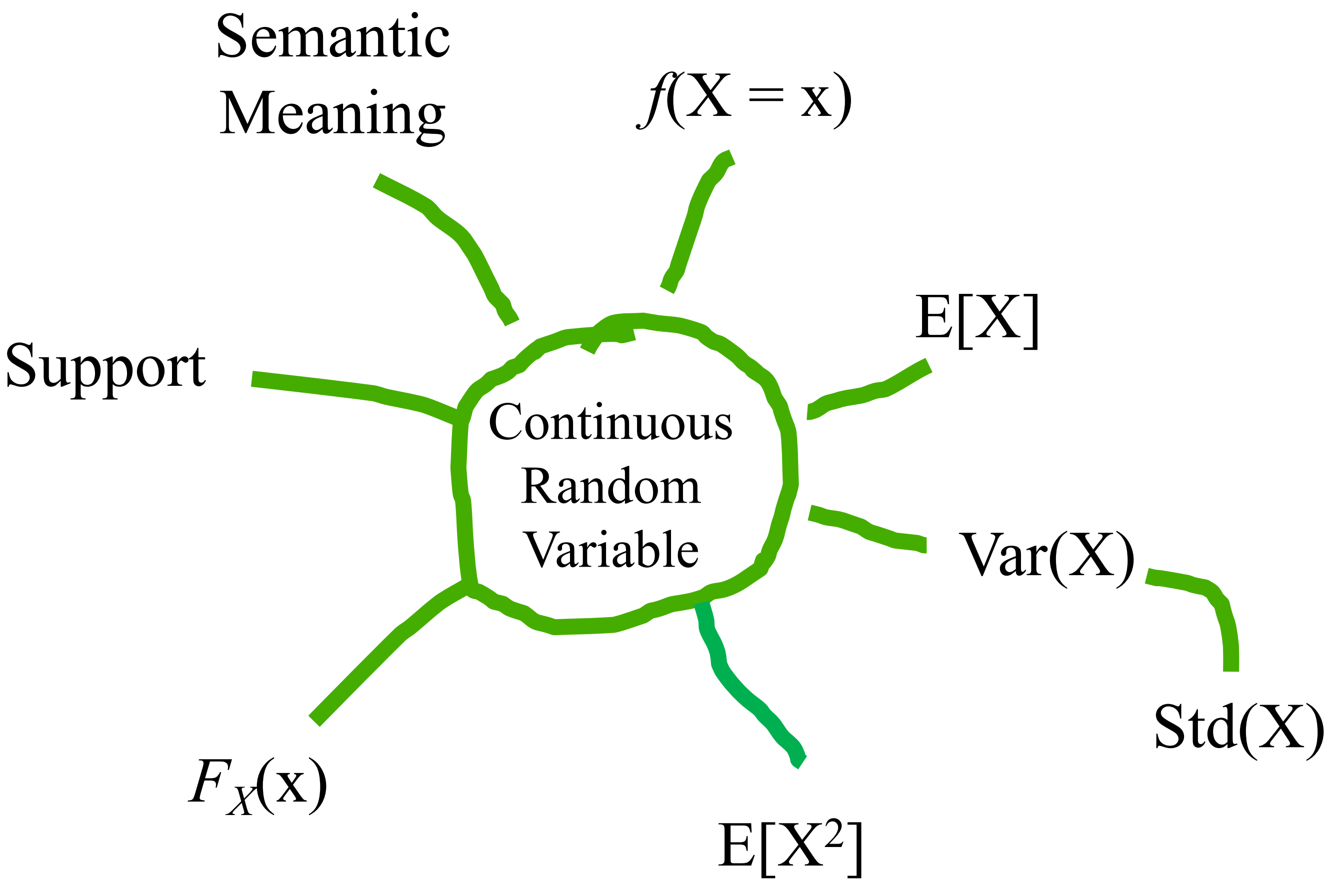
Density vs Cumulative



$f(x)$ = derivative of probability

$F(x) = P(X < x)$

Properties for Continuous Random Variables



Extra Problems

Replacing Your Laptop

- X = # hours of use until your laptop dies
  - On average, laptops die after 5000 hours of use
  - X ~ Exp(λ = 1/5000), since E[X] = 1/λ = 5000
  - You use your laptop 5 hours/day.
  - What is P(your laptop lasts 4 years)?
  - That is: P(X > (5)(365)(4) = 7300)

$$P(X > 7300) = 1 - F(7300) = 1 - (1 - e^{-7300/5000}) = e^{-1.46} \approx 0.2322$$

  - Better plan ahead... especially if you are coterming:

$$P(X > 9125) = 1 - F(9125) = e^{-1.825} \approx 0.1612 \quad \text{(5 year plan)}$$

$$P(X > 10950) = 1 - F(10950) = e^{-2.19} \approx 0.1119 \quad \text{(6 year plan)}$$

Exponential is Memoryless

- X = time until some event occurs
  - X ~ Exp(λ)
  - What is P(X > s + t | X > s)?

$$P(X > s+t \mid X > s) = \frac{P(X > s+t \text{ and } X > s)}{P(X > s)} = \frac{P(X > s+t)}{P(X > s)}$$

$$\frac{P(X > s+t)}{P(X > s)} = \frac{1-F(s+t)}{1-F(s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = 1-F(t) = P(X > t)$$

$$\text{So, } P(X > s+t \mid X > s) = P(X > t)$$

- After initial period of time *s*, P(X > t | •) for waiting another *t* units of time until event is same as at start
- "Memoryless" = no impact from preceding period *s*

Disk Crashes

- X = days of use before your disk crashes

$$f(x) = \begin{cases} \lambda e^{-x/100} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

  - First, determine λ to have actual PDF
    - Good integral to know: $\int e^u du = e^u$

$$1 = \int \lambda e^{-x/100} dx = -100\lambda \int \frac{-1}{100} e^{-x/100} dx = -100\lambda e^{-x/100} \Big|_0^{\infty} = 100\lambda \Rightarrow \lambda = \frac{1}{100}$$

  - What is P(50 < X < 150)?

$$F(150) - F(50) = \int_{50}^{150} \frac{1}{100} e^{-x/100} dx = -e^{-x/100} \Big|_{50}^{150} = -e^{-3/2} + e^{-1/2} \approx 0.383$$

  - What is P(X < 10)?

$$F(10) = \int_{0}^{10} \frac{1}{100} e^{-x/100} dx = -e^{-x/100} \Big|_{0}^{10} = -e^{-1/10} + 1 \approx 0.095$$


Stretch!
[redacted]

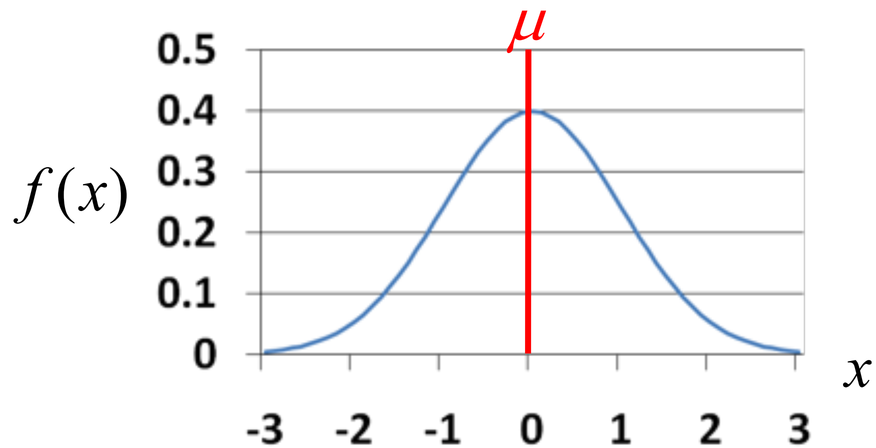
Gaussian

Noah Arthurs
CS109, Stanford University

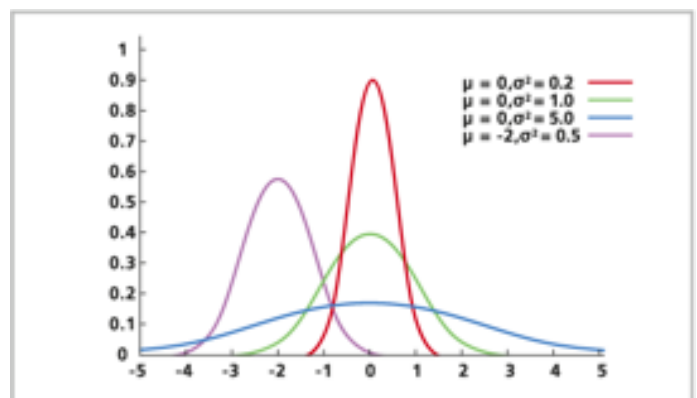
The Normal Distribution

- X is a **<u>Normal Random Variable</u>**: $X \sim N(\mu, \sigma^2)$
  - Probability Density Function (PDF):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad \text{where} \quad -\infty < x < \infty$$



  - $E[X] = \mu$
  - $Var(X) = \sigma^2$
  - Also called "Gaussian"
  - Note: $f(x)$ is symmetric about μ



Why use the normal?

- Common for natural phenomena: heights, weights, etc.
- Often results from the sum of multiple variables
- Most noise is Normal.
- Sample means are distributed normally.

Or that is what they want you to believe

But I Encourage you to be Critical

These are log-normal

- Common for natural phenomena: heights, weights, etc.

Only if they are equally weighted

- Often results from the sum of multiple variables

Most noise is assumed normal

- Most noise is Normal.

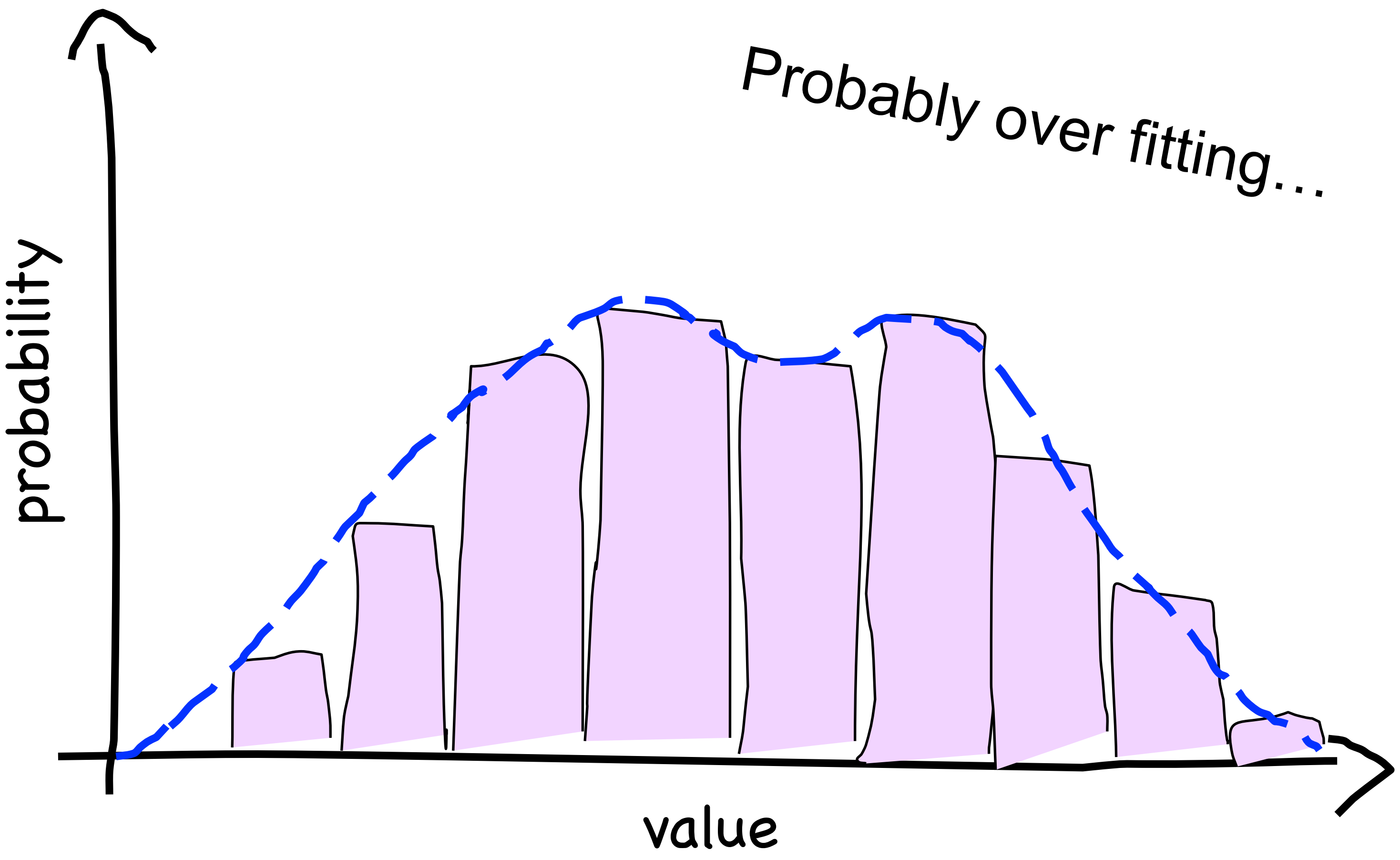
- Sample means are distributed normally.

It is the most important distribution

Because of a deeper truth…

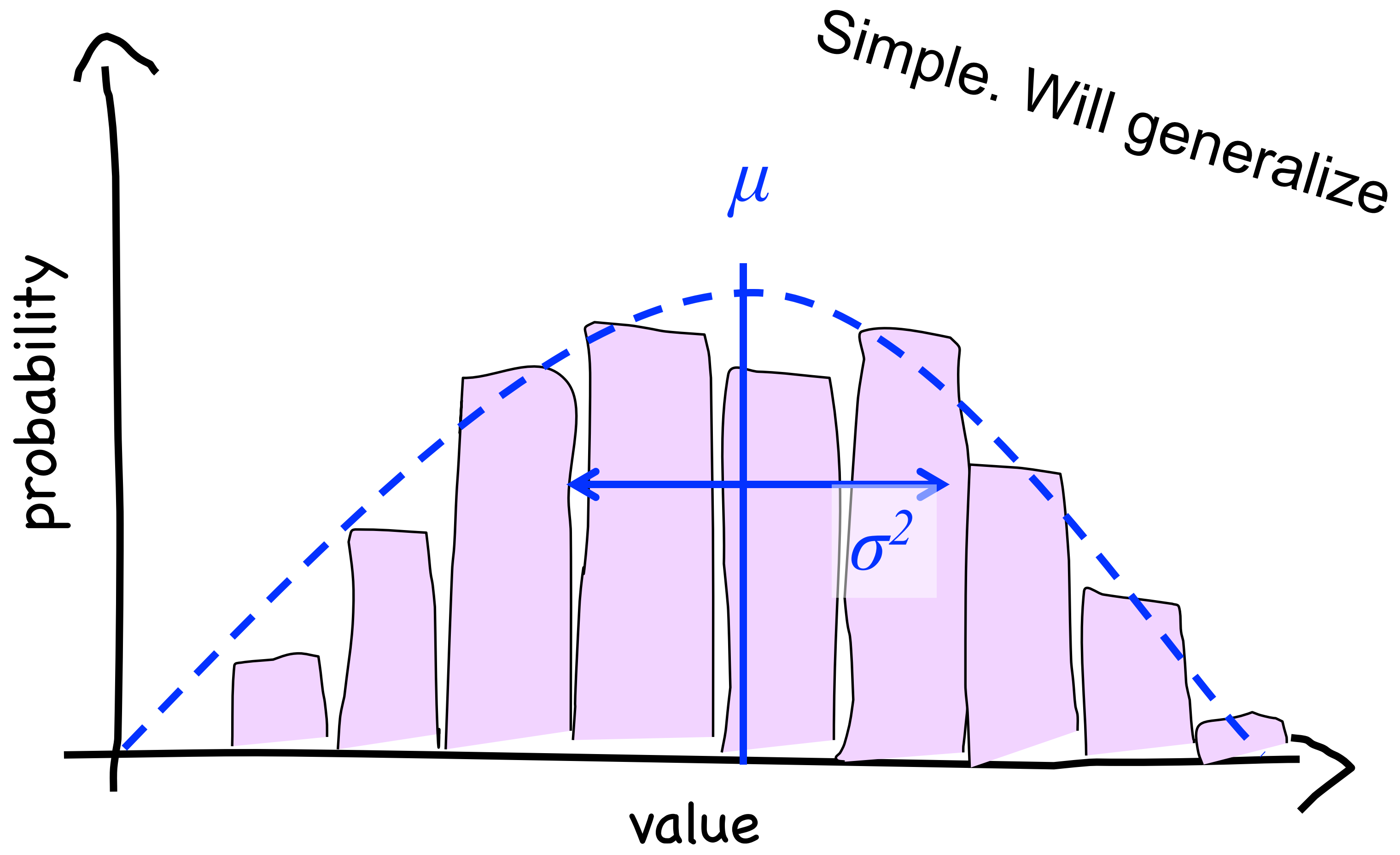
"The simplest explanation is usually the best one"

Complexity is Tempting



* That describes the training data, but will it generalize?

Simplicity is Humble



* A Gaussian maximizes entropy for a given mean and variance

Carl Friedrich Gauss

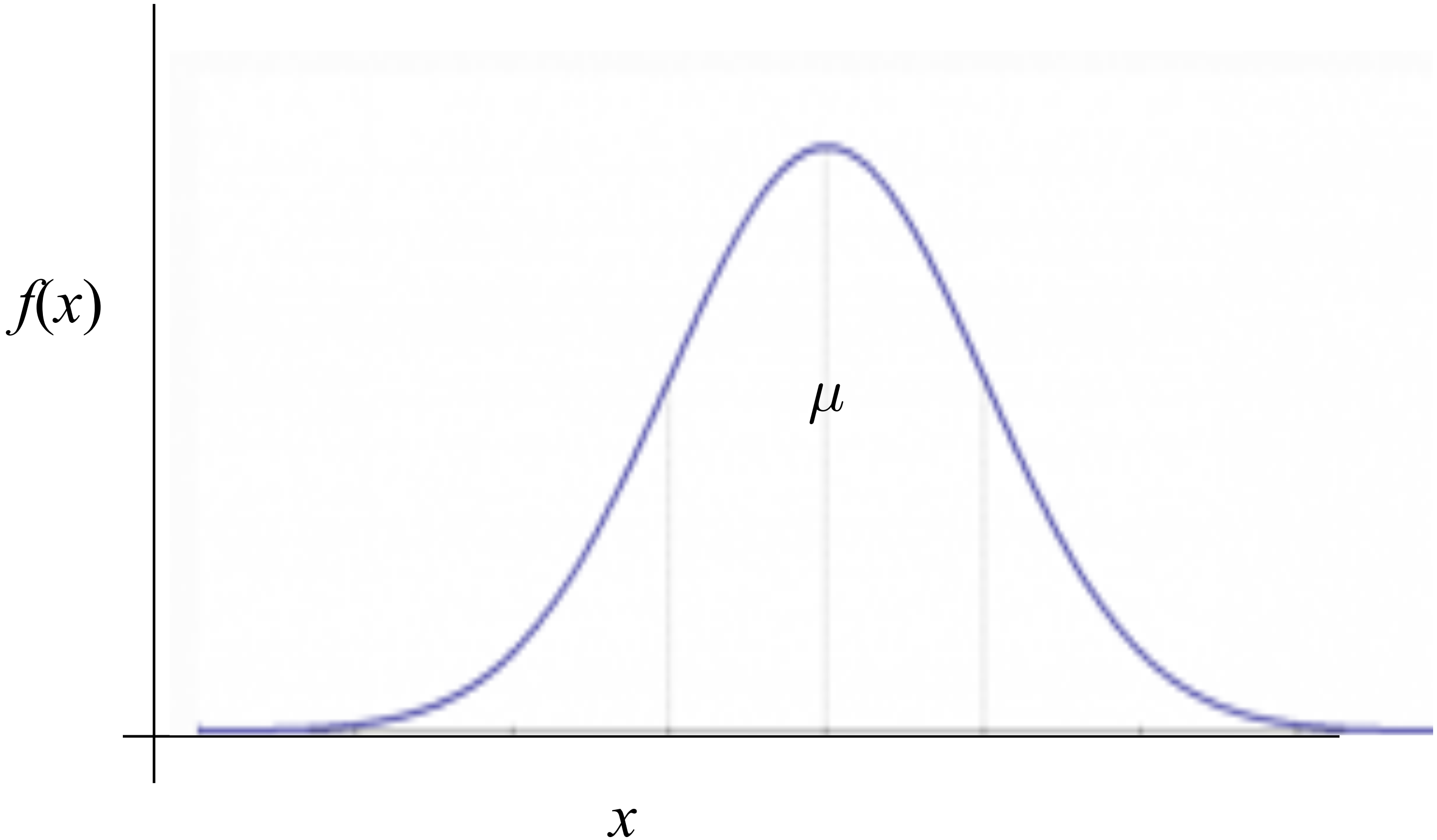
- Carl Friedrich Gauss (1777-1855) was a remarkably influential German mathematician

- Started doing groundbreaking math as teenager
  - Did not invent Normal distribution, but popularized it
- He looked like Martin Sheen
  - Who is, of course, Charlie Sheen’s father

Probability Density Function

$$\mathcal{N}(\mu, \sigma^2)$$

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



Anatomy of a beautiful equation

$\mathcal{N}(\mu, \sigma^2)$

"exponential"

the distance to the mean

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

probability density at x

a constant

sigma shows up twice

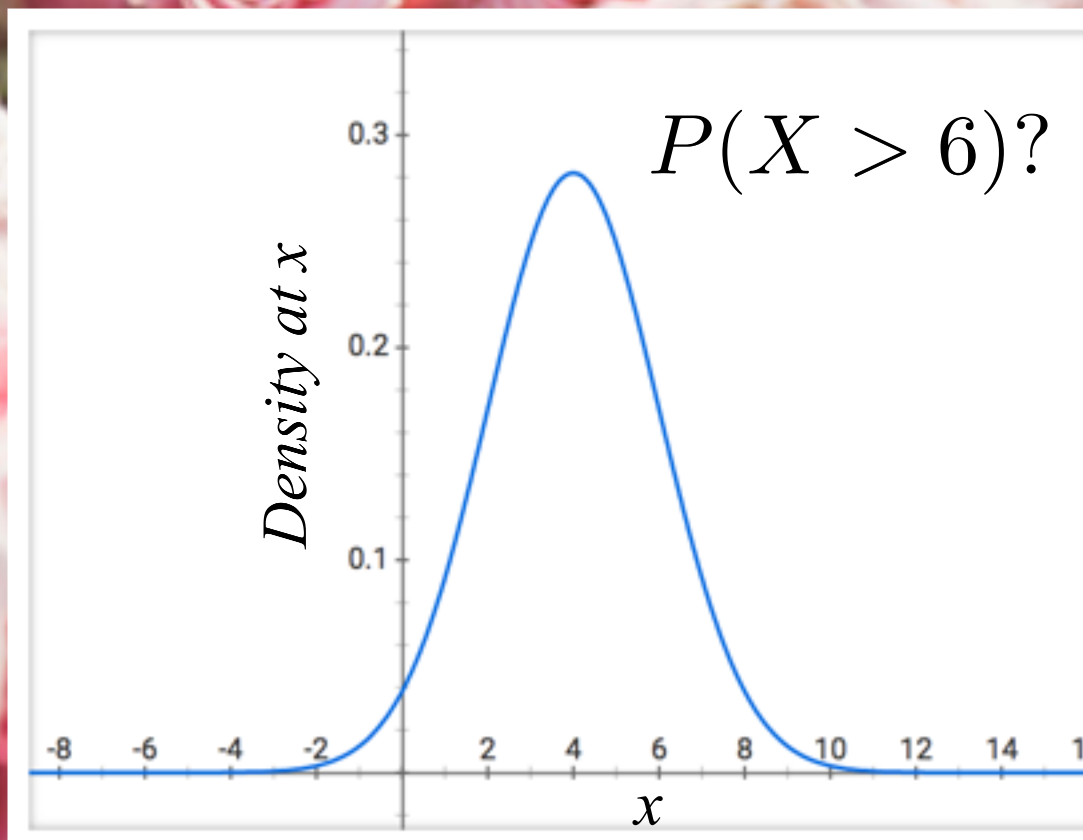
Flowers on a Rose Bush

$$X \sim N(\mu = 4, \sigma^2 = 2)$$

Partial credit for a partial rose

Flowers on a Rose Bush

$$X \sim N(\mu = 4, \sigma^2 = 2)$$



Let's try to integrate it!

$$P(a \leq X \leq b) =$$

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

* Call me if you find an equation for this

No closed form for the integral

No closed form for $F(x)$

Spoiler Alert

$\mathcal{N}(\mu, \sigma^2)$

A function that has been solved for numerically

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

The cumulative density function of any normal

* We are going to spend the next few slides getting here

Linear Transform of Normal is Normal

Let $X \sim \mathcal{N}(\mu, \sigma^2)$

If $Y = aX + b$ then Y is also Normal

$$\begin{aligned} E[Y] &= E[aX + b] \\ &= aE[X] + b \\ &= a\mu + b \end{aligned}$$

$$\begin{aligned} Var(Y) &= Var(aX + b) \\ &= a^2 Var(X) \\ &= a^2\sigma^2 \end{aligned}$$

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

Special Linear Transform

If $Y = aX + b$ then Y is also Normal

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

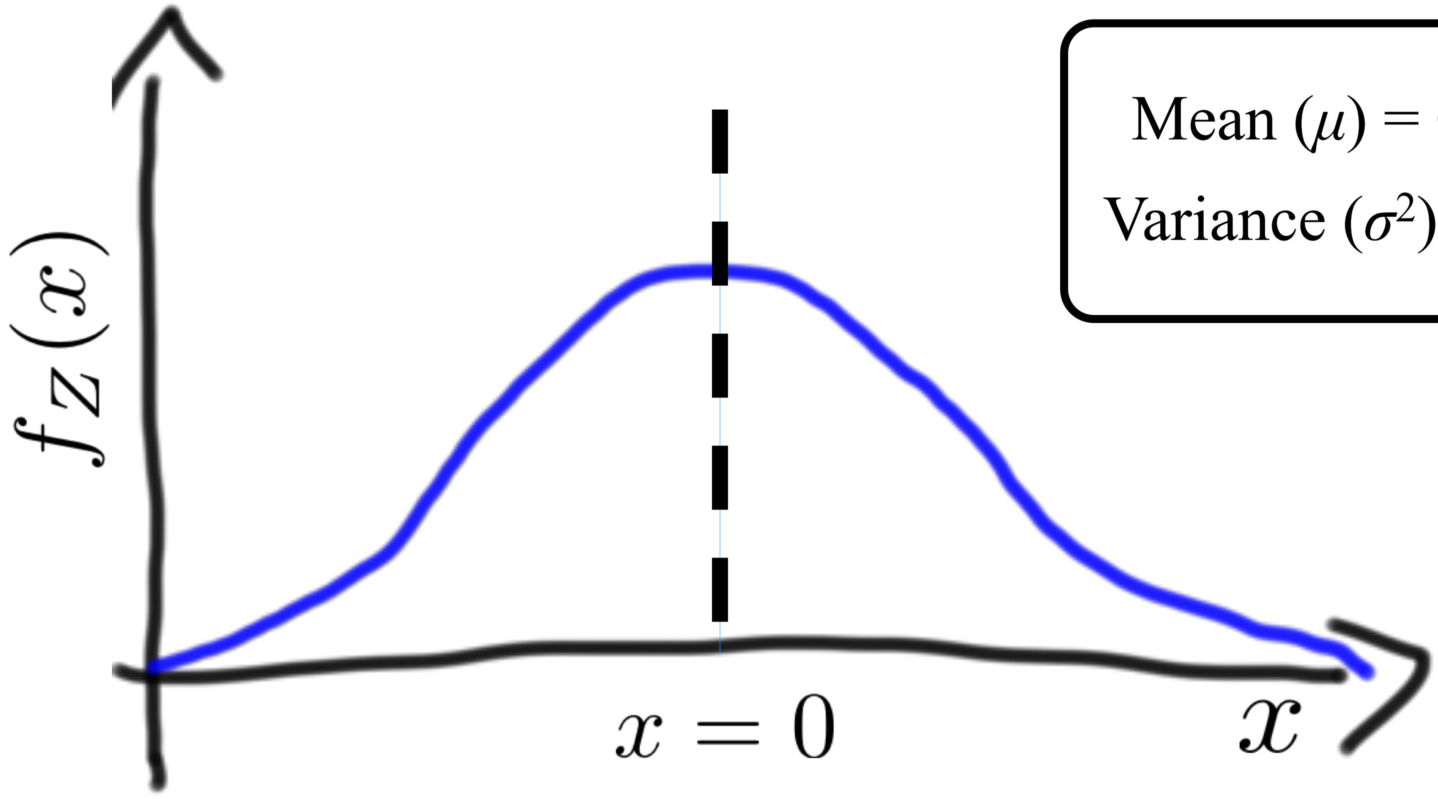
There is a special case of linear transform for any X:

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma} \qquad a = \frac{1}{\sigma} \quad b = -\frac{\mu}{\sigma}$$

$$\begin{aligned} Z &\sim \mathcal{N}(a\mu + b, a^2\sigma^2) \\ &\sim \mathcal{N}(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{\sigma^2}{\sigma^2}) \\ &\sim \mathcal{N}(0, 1) \end{aligned}$$

The Standard Normal

$$Z \sim N(\mu = 0, \sigma^2 = 1)$$



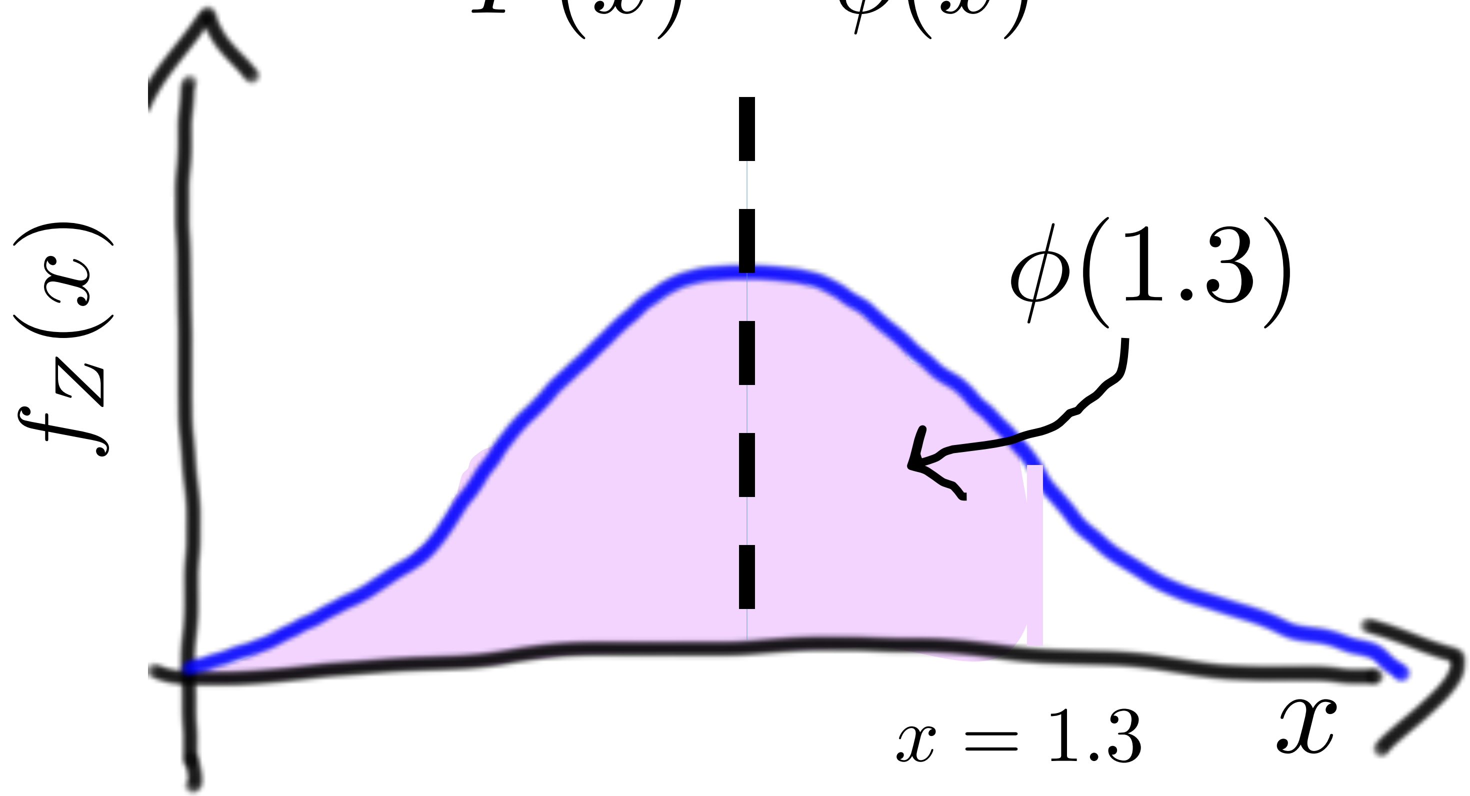
Mean (μ) = 0

Variance (σ^2) = 1

*This is the probability density function for the standard normal

Phi

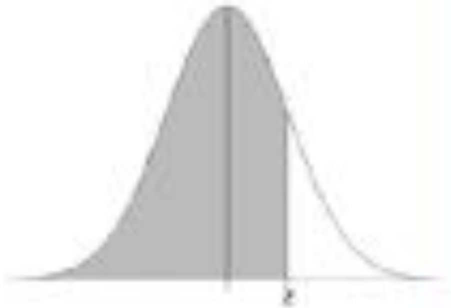
$$Z \sim N(\mu = 0, \sigma^2 = 1)$$
$$F(x) = \phi(x)$$



*This is the probability density function for the standard normal

Using Table of ϕ

Standard Normal Cumulative Probability Table



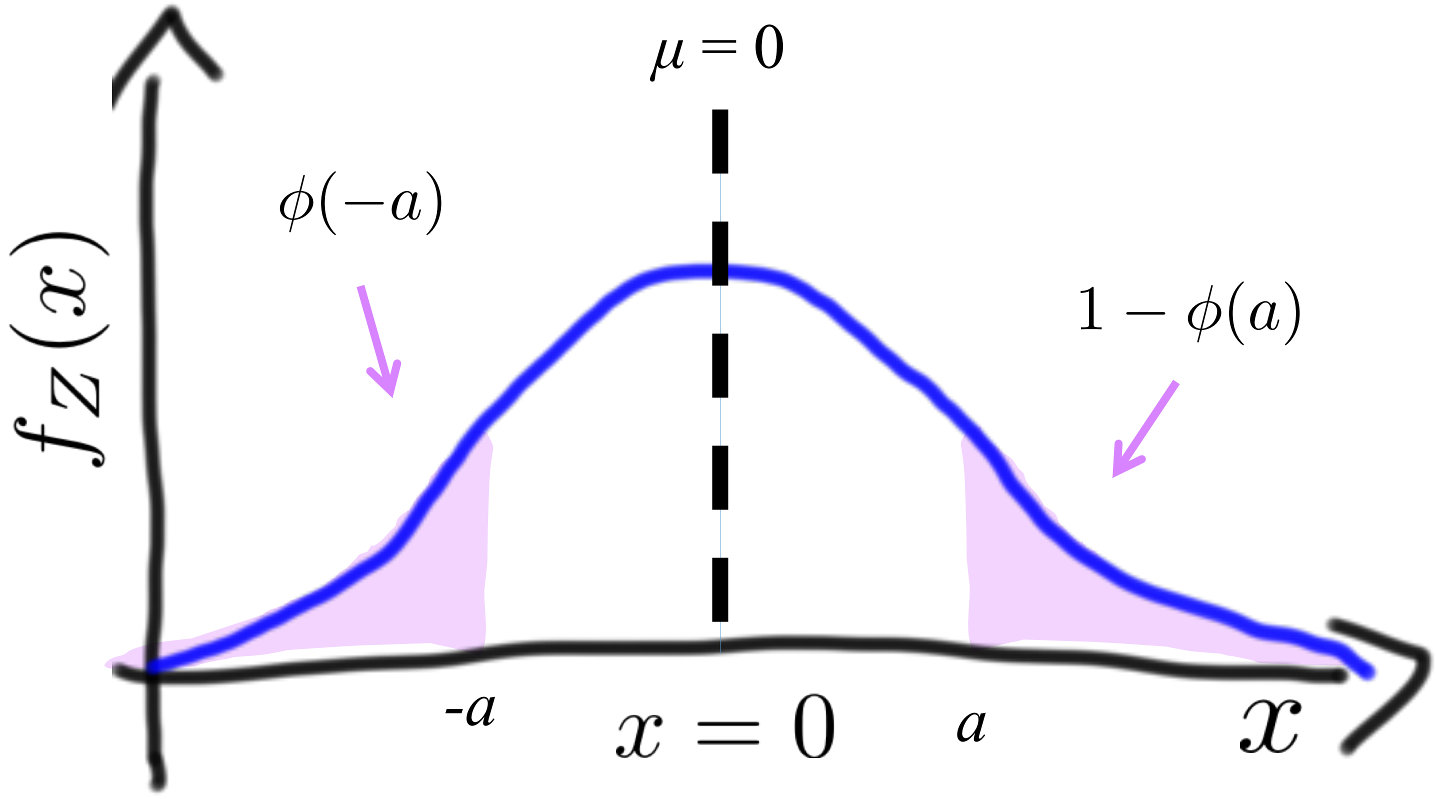
$\Phi(1.31) = 0.7054$

Cumulative probabilities for POSITIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

Symmetry of Phi

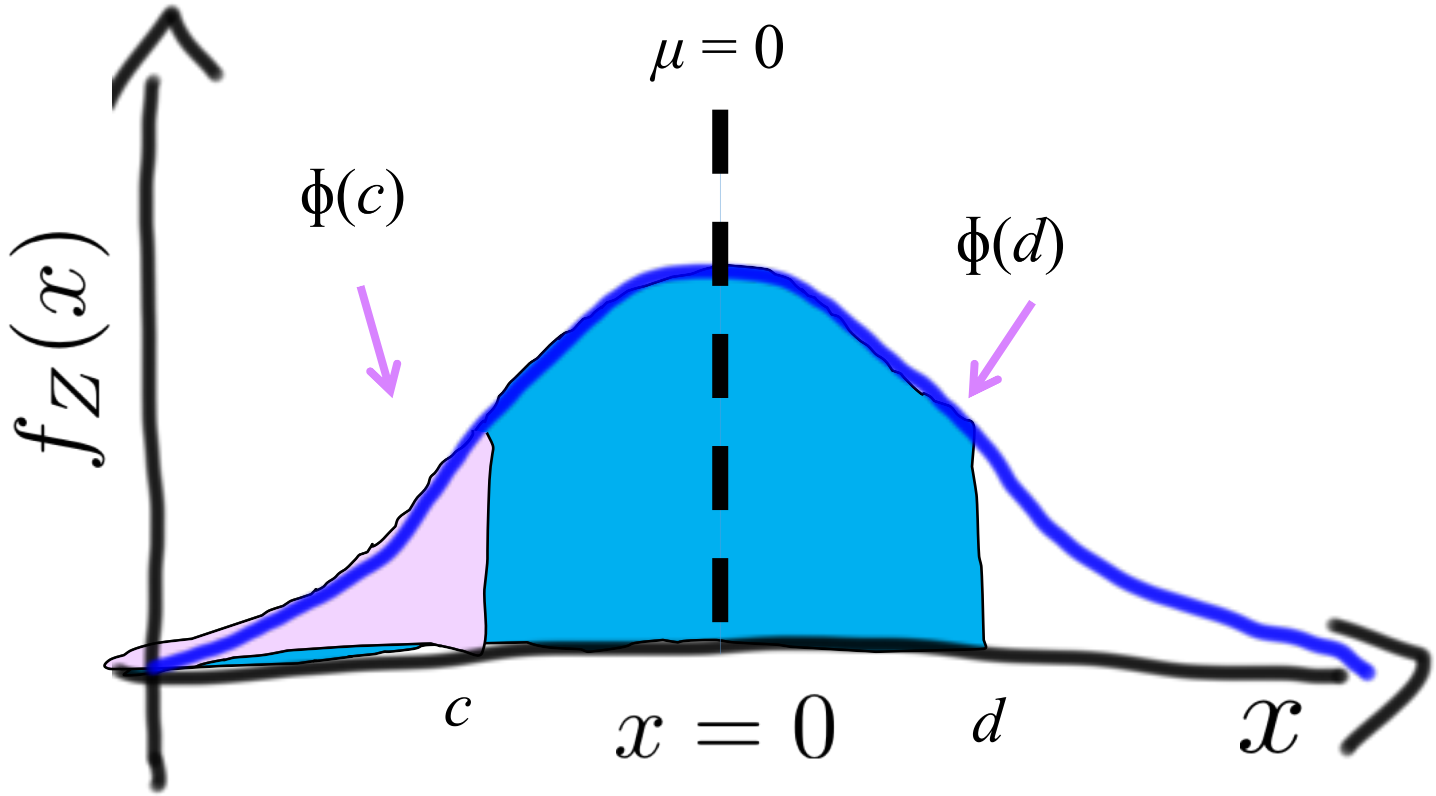
$$\varphi(-a) = 1 - \varphi(a)$$



*This is the probability density function for the standard normal

Interval of Phi

$$P(c < Z < d) = \phi(d) - \phi(c)$$



Compute F(x) via Transform

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ $\qquad$ $Z = \dfrac{X - \mu}{\sigma}$

Use Z to compute $F(x)$

$$
\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P(X - \mu \leq x - \mu) \\
&= P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) \\
&= P\left(Z \leq \frac{x - \mu}{\sigma}\right)
\end{aligned}
$$

$$
= \Phi\left(\frac{x - \mu}{\sigma}\right)
$$

For normal distribution, F(x) is computed using the phi transform.

And here we are

$$\mathcal{N}(\mu, \sigma^2)$$

CDF of Standard Normal: A function that has been solved for numerically

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

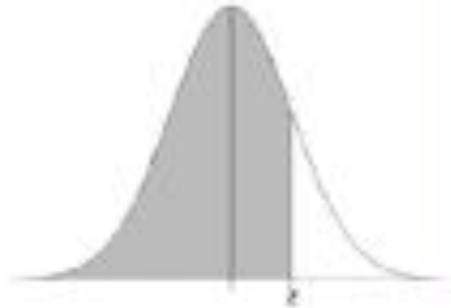
The cumulative density function (CDF) of any normal

Table of $\Phi(z)$ values in textbook, p. 201 and handout

Using Table of ϕ

Standard Normal Cumulative Probability Table

$\Phi(0.54) = 0.7054$



Cumulative probabilities for POSITIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

Table is kinda old school

Using Programming Library

Every modern programming language has a normal library

```
norm.cdf(x, mean, std)
```

$$= P(X < x) \text{ where } X \sim \mathcal{N}(\mu, \sigma^2)$$

$$= \Phi\left(\frac{x - \mu}{\sigma}\right)$$

* This is from Python's scipy library

We have one for you