Continuous Variables

Will Song
CS109, Stanford University

1906 Earthquake
Magnitude 7.8

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

Review

Random Variable Notation!

Our random variable

Num trials

Probability of success on each trial

$$X \sim \mathrm{Bin}(n, p)$$

Is distributed as a

Binomial

With these parameters

This part must be an Event

Probability Mass Function for a Binomial

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

Probability that our variable takes on the value k

* This is also called the binomial term

Binomial Random Variable

```
(H, H, H, H, T, T, T, T, T, T)
(H, H, H, T, H, T, T, T, T, T)
(H, H, H, T, T, H, T, T, T, T)
(H, H, H, T, T, T, H, T, T, T)
(H, H, H, T, T, T, T, H, T, T)
(H, H, H, T, T, T, T, T, H, T)
(H, H, H, T, T, T, T, T, T, H)
(H, H, T, H, H, T, T, T, T, T)
(H, H, T, H, T, H, T, T, T, T)
(H, H, T, H, T, T, H, T, T, T)
(H, H, T, H, T, T, T, H, T, T)
(H, H, T, H, T, T, T, T, H, T)
(H, H, T, H, T, T, T, T, T, H)
(H, H, T, T, H, H, T, T, T, T)
(H, H, T, T, H, T, H, T, T, T)
(H, H, T, T, H, T, T, H, T, T)
(H, H, T, T, H, T, T, T, H, T)
(H, H, T, T, H, T, T, T, T, H)
```

The number of **successes**, in *n* independent **trials**, where each **trial** is a **success** with probability p:

Binomial Random Variable

Binomial Random Variable

Notation: $X \sim \text{Bin}(n, p)$

Description: Number of "successes" in n identical, independent experiments each with probability of success p.

Parameters: $n \in \{0, 1, \ldots\}$, the number of experiments.
$p \in [0, 1]$, the probability that a single experiment gives a "success".

Support: $x \in \{0, 1, \ldots, n\}$

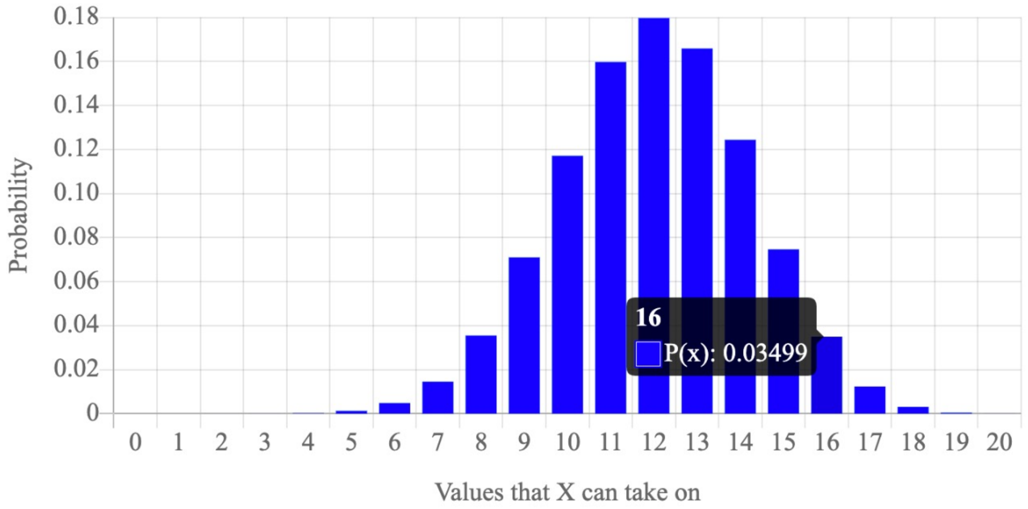
PMF equation: $\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

Expectation: $\text{E}[X] = n \cdot p$

Variance: $\text{Var}(X) = n \cdot p \cdot (1-p)$

PMF graph:

Parameter n: 20 Parameter p: 0.60



Bernoulli Random Variable

Notation: $X \sim \text{Bern}(p)$

Description: A boolean variable that is 1 with probability p

Parameters: p, the probability that $X = 1$.

Support: x is either 0 or 1

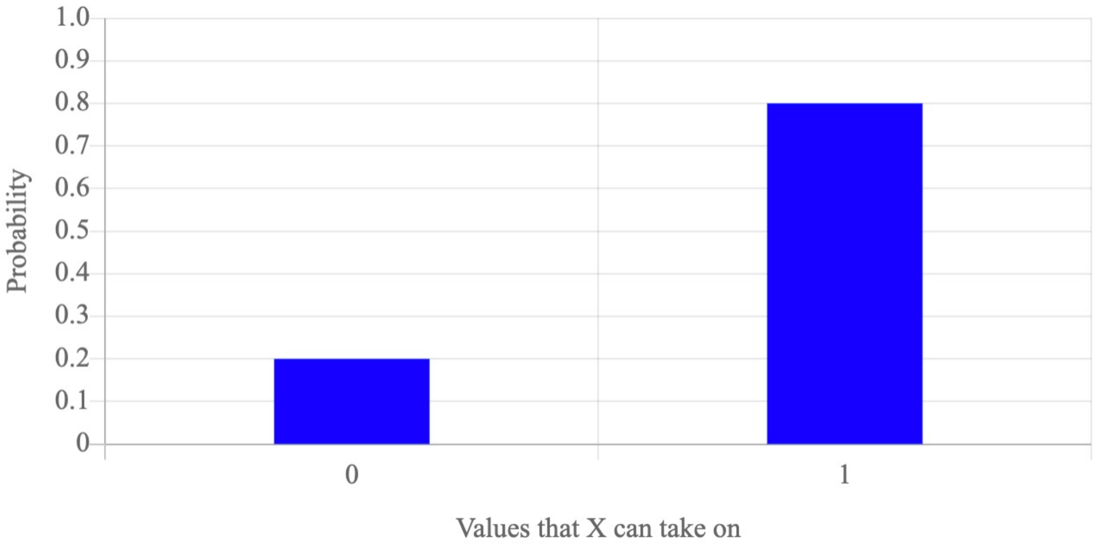
PMF equation: $\Pr(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$

Expectation: $\text{E}[X] = p$

Variance: $\text{Var}(X) = p(1-p)$

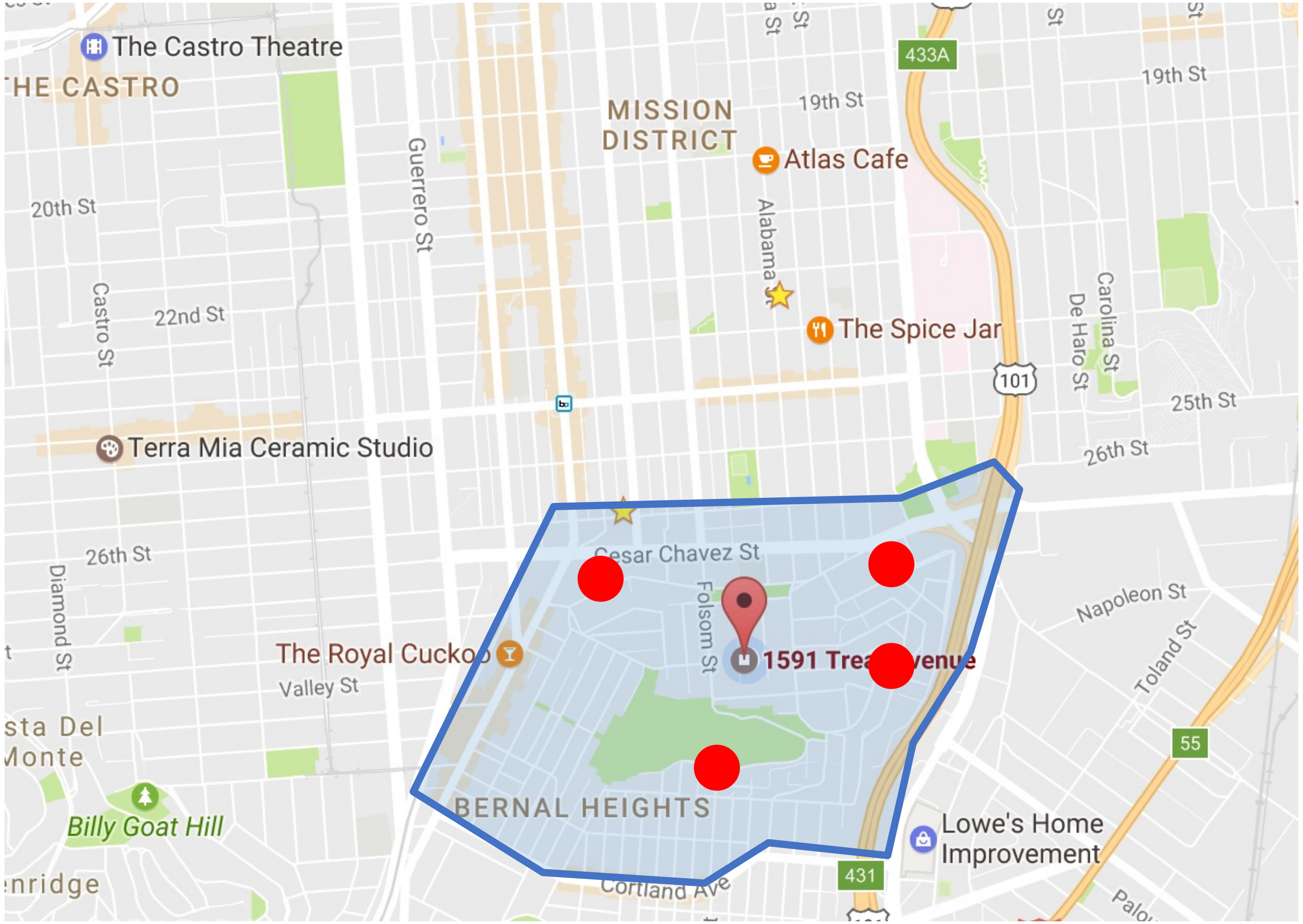
PMF graph:

Parameter p: 0.80



Poisson Random Variable

Probability of ***k* requests** from this area in the next 1 min



Poisson Random Variable

Poisson Random Variable

Notation: $X \sim \text{Poi}(\lambda)$

Description: Number of events in a fixed time frame if (a) the events occur with a constant mean rate and (b) they occur independently of time since last event.

Parameters: $\lambda \in \{0, 1, \ldots\}$, the constant average rate.

Support: $x \in \{0, 1, \ldots\}$

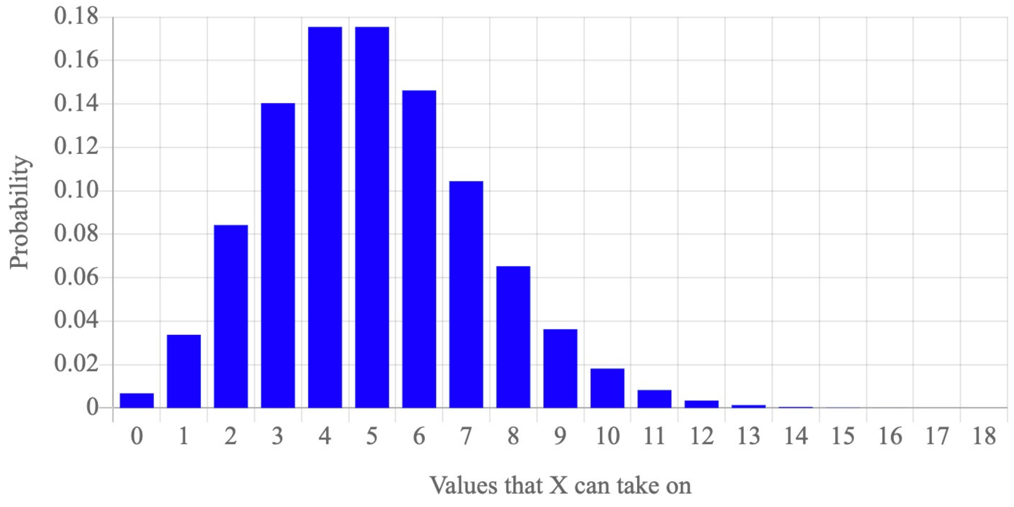
PMF equation: $\Pr(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$

Expectation: $\text{E}[X] = \lambda$

Variance: $\text{Var}(X) = \lambda$

PMF graph:

Parameter λ: 5



Geometric Random Variable

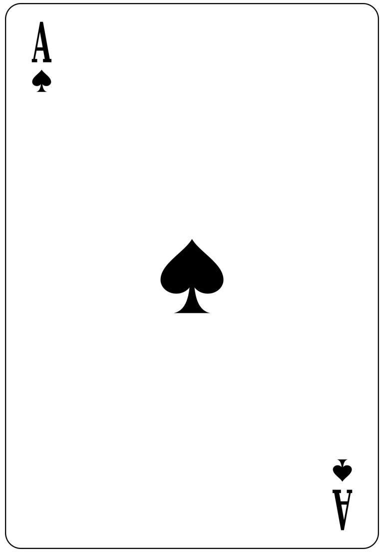
X is **Geometric** Random Variable: X ~ Geo(*p*)

- 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:

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

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

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



Negative Binomial Random Variable

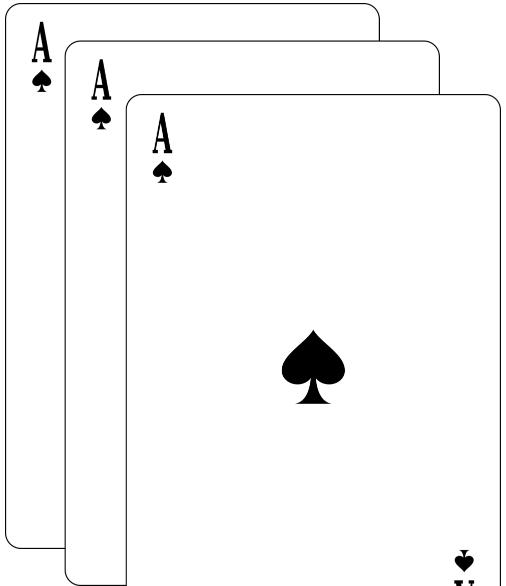
X is **<u>Negative Binomial</u>** RV: X ~ NegBin(r, p)

- 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)



Dating at Stanford

Each person you date has a 0.2 probability of being someone you spend your life with. What is the average number of people one will date? What is the standard deviation? **Your meta goal: what steps would you take to answer this question?**

Bitcoin Mining

SHA-256 Hash(Data *Fixed* Salt *Choice*

Number that looks like random bits

You "mine a bitcoin" if, for given data D, you find a salt number N such that Hash(D, N) produces a string that starts with g zeroes.

You "mine a bitcoin" if, for given data D, you find a number N such that Hash(D, N) produces a string that starts with g zeroes.

(a) What is the probability that the first number you try will produce a bit string which starts with g zeroes (in other words you mine a bitcoin)?

Let X be the number of zeros in the first g bits. $X \sim \text{Bin}(n = g, p = 0.5)$

$$P(X = 0) = \binom{g}{0} \frac{1}{2}^g = \frac{1}{2}^g$$

Call this answer p_a

(b) What is the probability that you will need under 100 attempts to mine 2 bit coins? (Feel free to leave in summation form)

Let Y be the number of tries until you mine 2 bitcoins. $Y \sim \text{NegBin}(r = 2, p = p_a)$

$$P(Y < 100) = \sum_{x=2}^{99} P(Y = x)$$
$$= \sum_{x=2}^{99} \binom{n-1}{r-1} p^r (1-p)^{x-r}$$

Call this answer p_b

Geometric and Negative Binomial

Geometric Random Variable

Notation: $X \sim \text{Geo}(p)$

Description: Number of experiments until a success. Assumes independent experiments each with probability of success p.

Parameters: $p \in [0, 1]$, the probability that a single experiment gives a "success".

Support: $x \in \{1, \ldots, \infty\}$

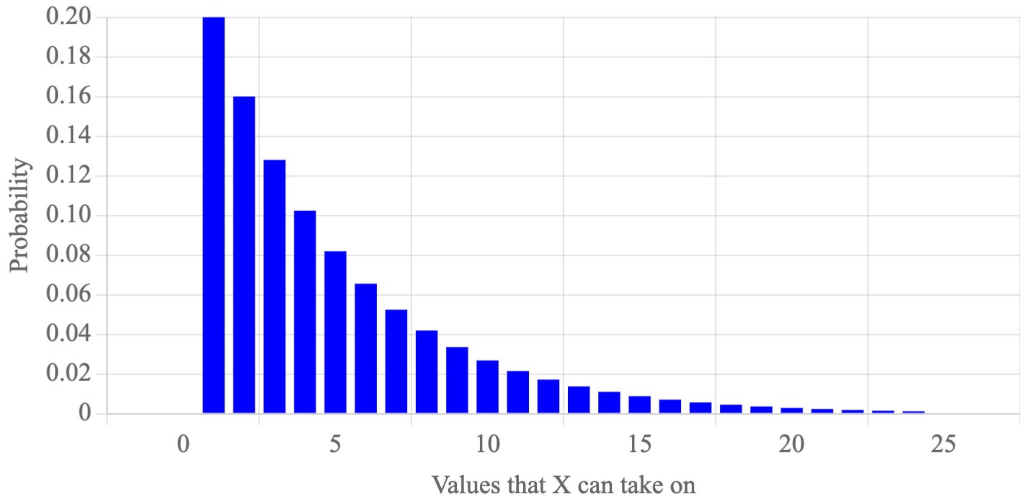
PMF equation: $\text{P}(X = x) = (1 - p)^{x-1} p$

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

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

PMF graph:

Parameter p: 0.20



Negative Binomial Random Variable

Notation: $X \sim \text{NegBin}(r, p)$

Description: Number of experiments until r successes. Assumes each experiment is independent with probability of success p.

Parameters: $r > 0$, the number of success we are waiting for.
$p \in [0, 1]$, the probability that a single experiment gives a "success".

Support: $x \in \{r, \ldots, \infty\}$

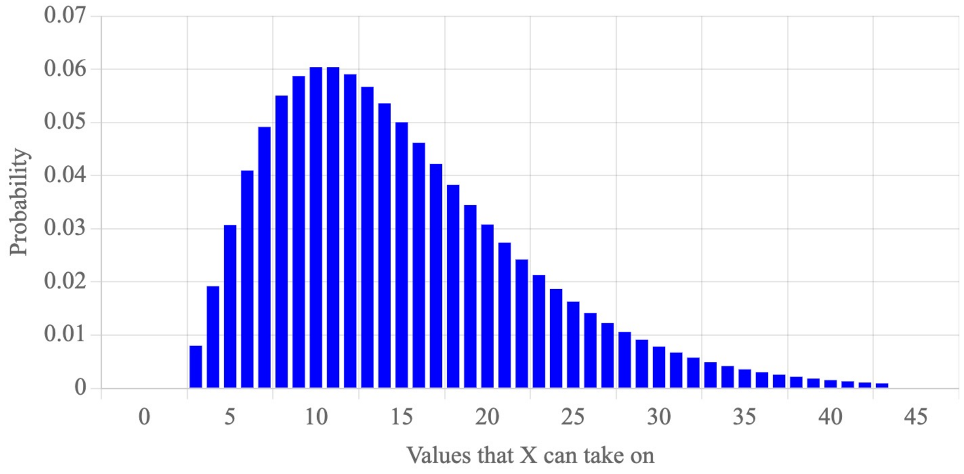
PMF equation: $\text{P}(X = x) = \binom{x-1}{r-1} p^r (1 - p)^{x-r}$

Expectation: $\text{E}[X] = \frac{r}{p}$

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

PMF graph:

Parameter r: 3 Parameter p: 0.20



Last Class!!

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

What We Will Learn Today!

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

Motivation for Continuous RVs

End of Review!

Learning Goals

1. Comfort using Continuous Random Variables
2. Integrate a density function (PDF) to get a probability
3. Use a cumulative function (CDF) to get a probability

Random Variables!

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

Goal: Be Able to Use a New Random Variable

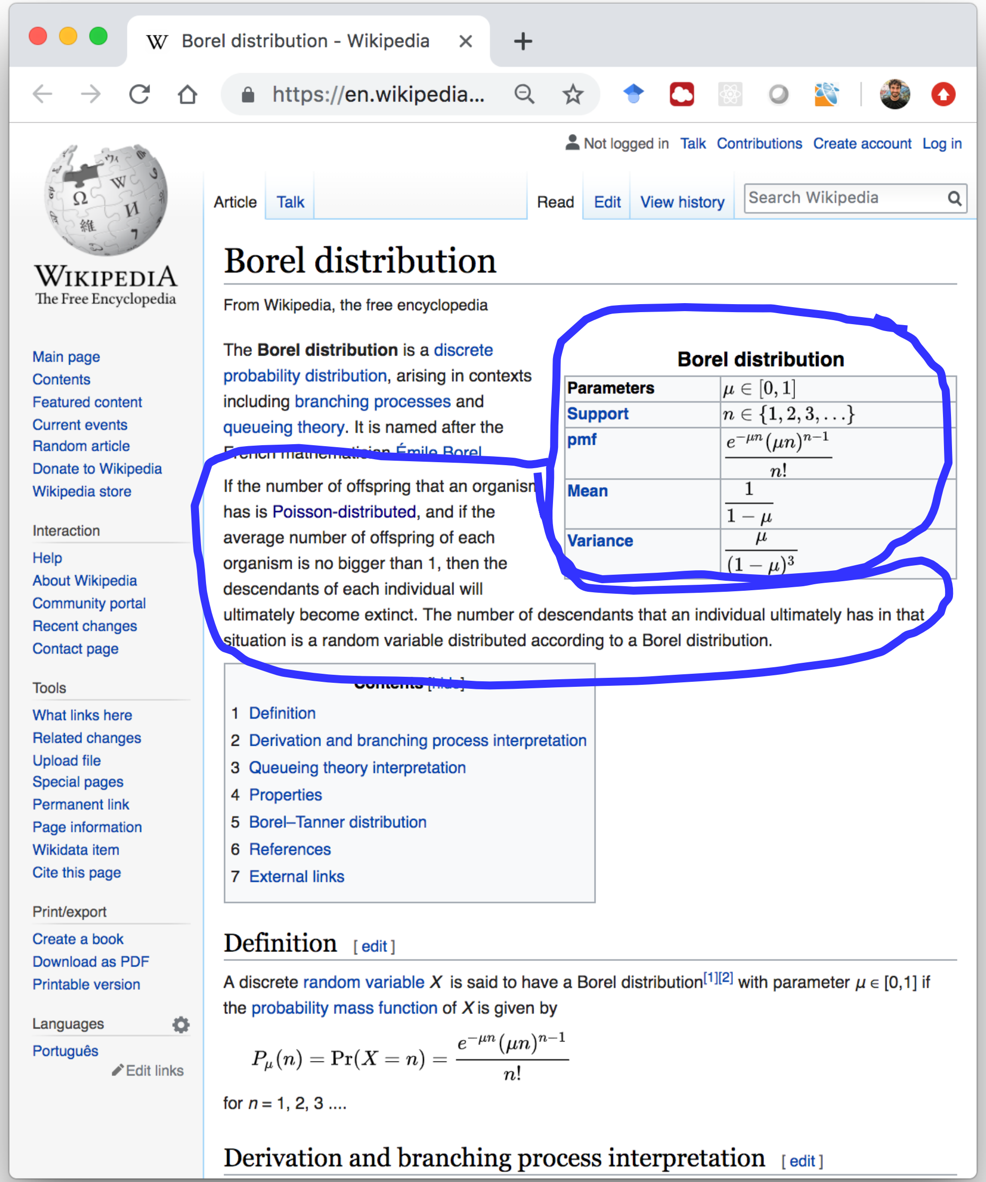
Don't have to derive all of the following distributions.
We want you to get a sense of how random variables work.

Goal: Be Able to Use a New Random Variable

You are learning about servers…

You read about the MD1 queue…

You find a paper that says the length of a server "busy period" is distributed as a Borel with parameter μ = 0.2 …



Borel distribution - Wikipedia

https://en.wikipedia...

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

Languages
Português
Edit links

Borel distribution

From Wikipedia, the free encyclopedia

The **Borel distribution** is a discrete probability distribution, arising in contexts including branching processes and queueing theory. It is named after the French mathematician Émile Borel.

If the number of offspring that an organism has is Poisson-distributed, and if the average number of offspring of each organism is no bigger than 1, then the descendants of each individual will ultimately become extinct. The number of descendants that an individual ultimately has in that situation is a random variable distributed according to a Borel distribution.

Borel distribution	
Parameters	$\mu \in [0,1]$
Support	$n \in \{1,2,3,\ldots\}$
pmf	$\frac{e^{-\mu n}(\mu n)^{n-1}}{n!}$
Mean	$\frac{1}{1-\mu}$
Variance	$\frac{\mu}{(1-\mu)^3}$

Contents [hide]

Definition [edit]

A discrete random variable X is said to have a Borel distribution[1][2] with parameter $\mu \in [0,1]$ if the probability mass function of X is given by

$$P_\mu(n) = \Pr(X = n) = \frac{e^{-\mu n}(\mu n)^{n-1}}{n!}$$

for $n = 1, 2, 3 \ldots$.

Derivation and branching process interpretation [edit]

Abstractions

More abstract

Is this quantity Random?	Die Rolls? Random Process?	Random Variable
What kind of Random?	Named Random Variables	Binomial, Poisson, Geometric, Exponential
What do I need to parameterize this randomness?	$X \sim \text{Poi}(3)$ $Y \sim \text{Bin}(20, 0.1)$ $Z \sim \text{Geo}(0.5)$	R.Vs with Specific Parameters
What's a specific event I care about?	Values	$P(X = 3)$ $P(Y = 3 \mid X = 3)$ $E[Y], \quad \text{Var}[Z]$

More Specific

Discrete Distributions

Bernoulli:
- indicator of coin flip X ~ Ber(p)

Binomial:
- # successes in n coin flips X ~ Bin(n, p)

Poisson:
- # successes in a fixed interval of time X ~ Poi(λ)

Geometric:
- # coin flips until success X ~ Geo(p)

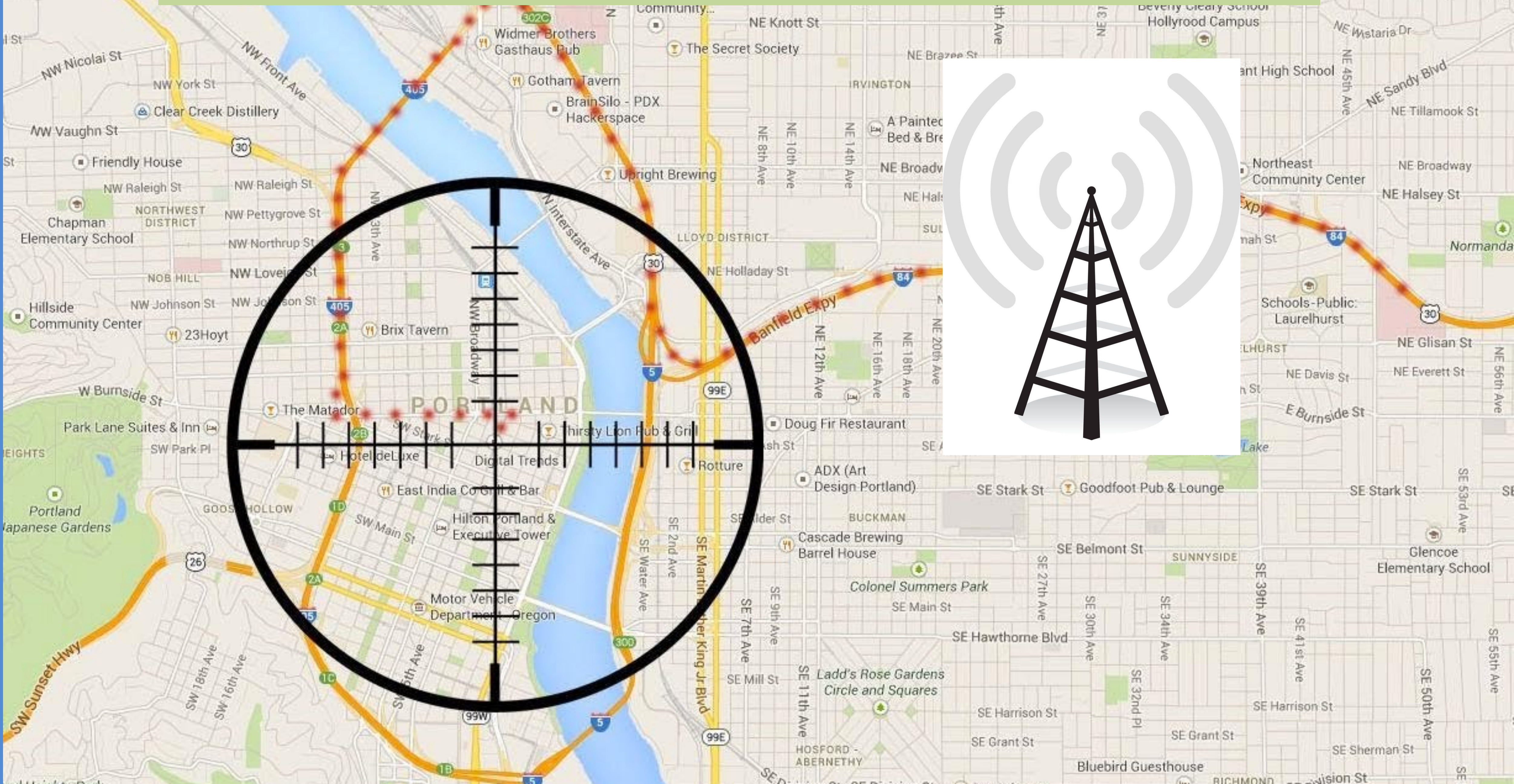
Negative Binomial:
- # trials until r successes X ~ NegBin(r, p)

Why Named Discrete?

Because they all take on integer values.

Big hole in our knowledge

Not all values are discrete



random()?

```
>>> import random
>>> random.random()
0.26735294788086206
>>> random.random()
0.07594786913253304
>>> random.random()
0.10915346833886497
>>> random.random()
0.22733701384493341
>>> random.random()
0.5386449238096968
```

Poisson

Say the average rate of earthquakes is 1 every 100 years.

We can talk about the probability distribution of different numbers of earthquakes next year.

We can't talk about the probability distribution of the amount of time until the next earthquake.

Riding the Marguerite

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 2:15pm.

What is P(wait < 5 minutes)?

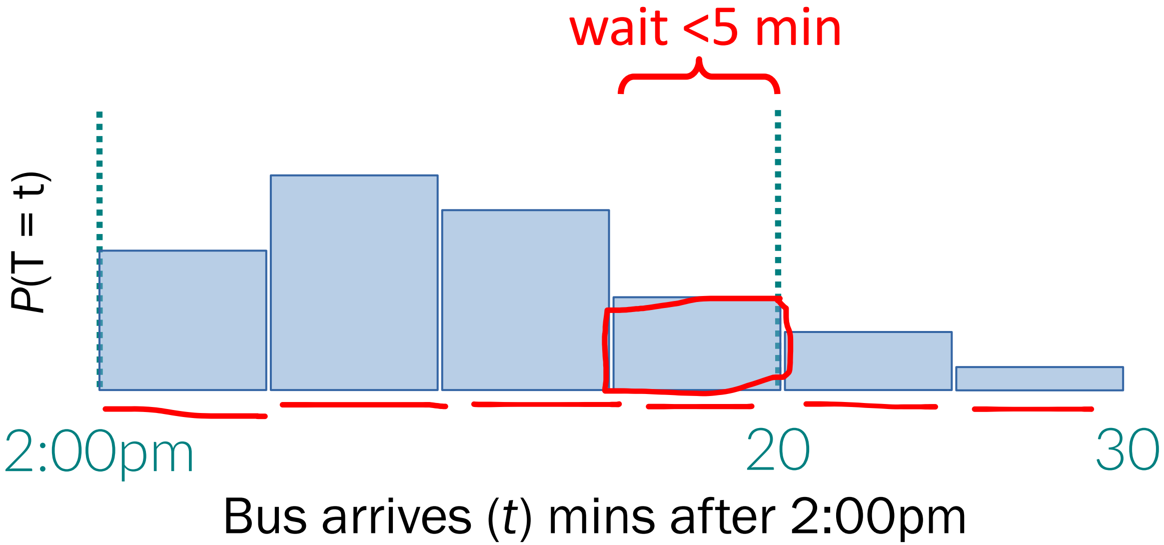
What is the probability that the bus arrives at:
2:17pm and 12.1233391102389234 seconds?

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 2:15pm.

What is P(wait < 5 minutes)?



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 2:15pm.

What is P(wait < 5 minutes)?

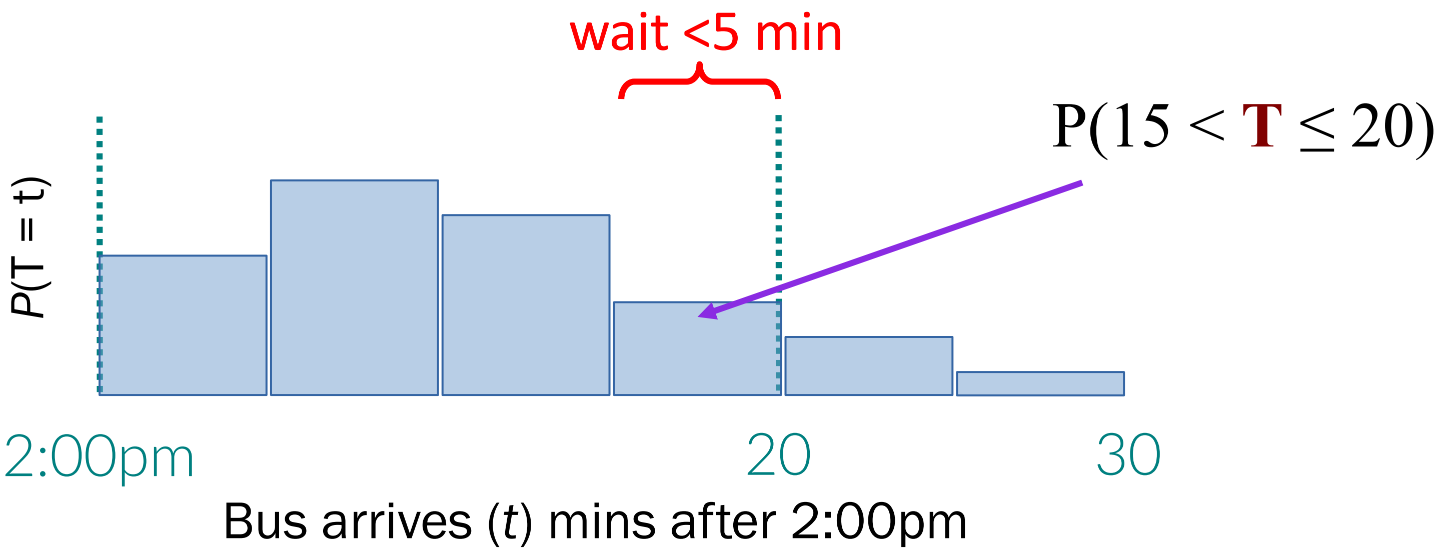


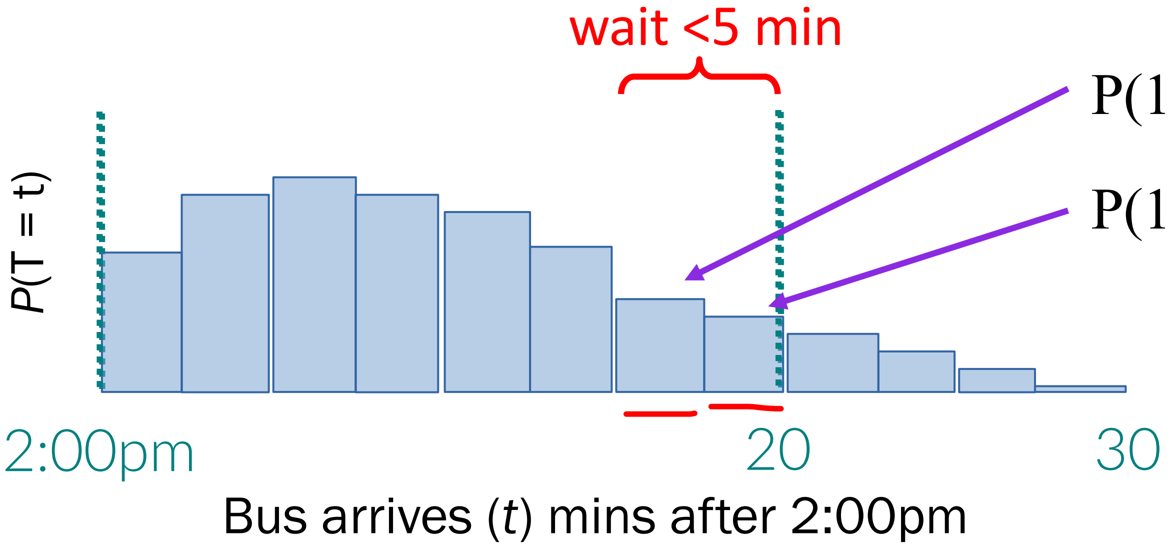
image: Haha169

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 2:15pm.

What is P(wait < 5 minutes)?



$$P(15 < \mathbf{X} \leq 17.5)$$
$$+$$
$$P(17.5 < \mathbf{X} \leq 20)$$

image: Haha169

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 2:15pm.

What is P(wait < 5 minutes)?

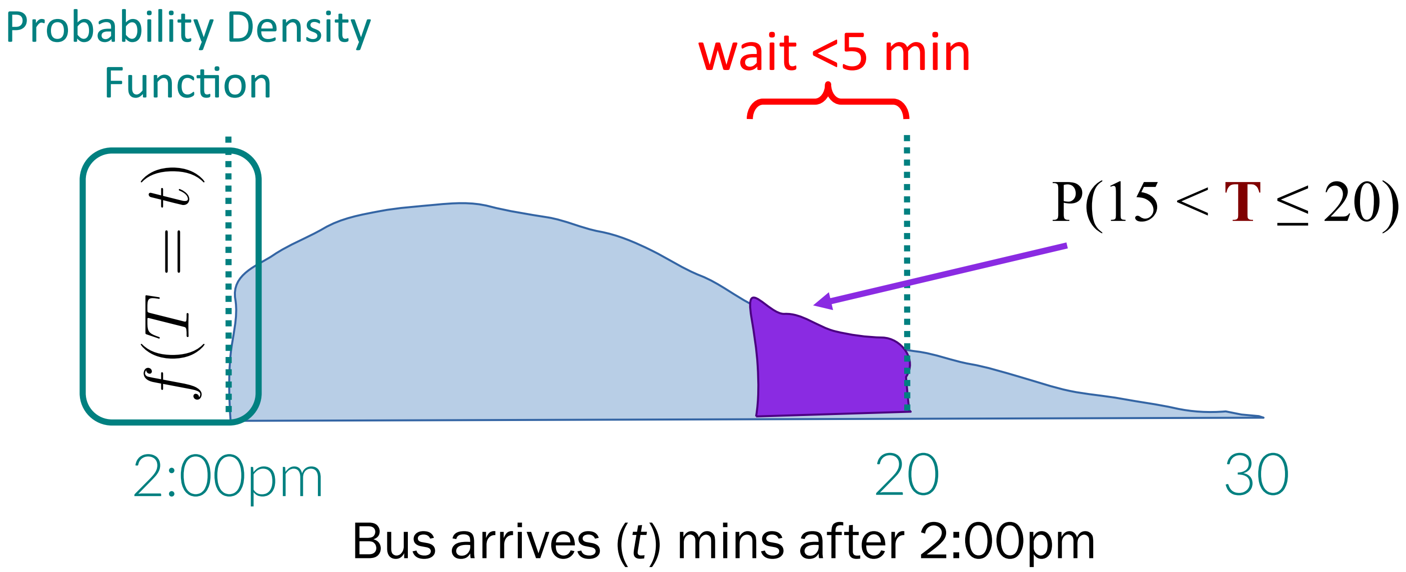


image: [Haha169]()

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

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!

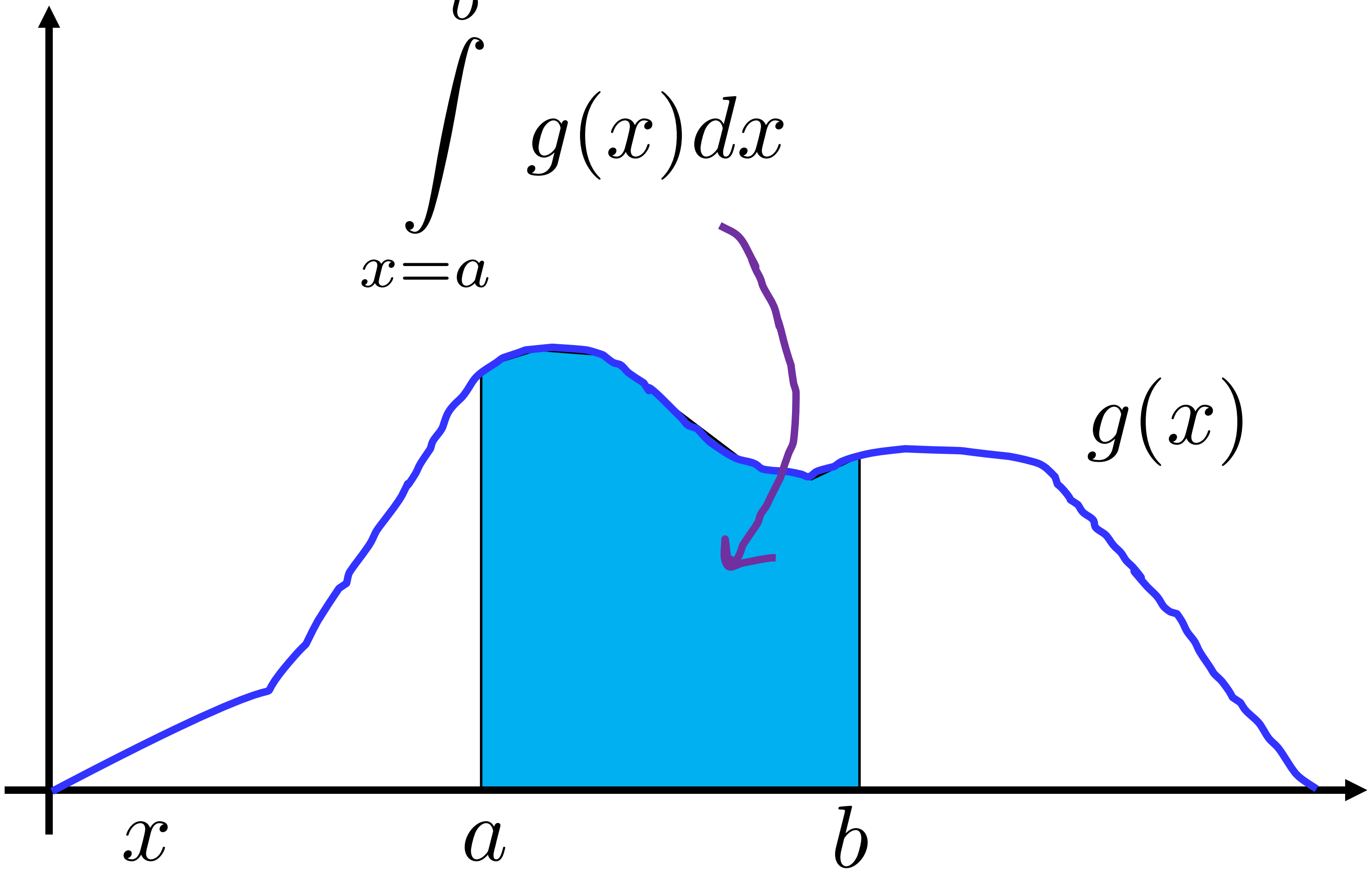
$$P(a < X < b) = \int_{x=a}^{b} f(X = x)\,dx$$

Integrals!

*loving, not scary

Integrals

$$\int_{x=a}^{b} g(x)dx$$



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 2:15pm.

What is P(wait < 5 minutes)?

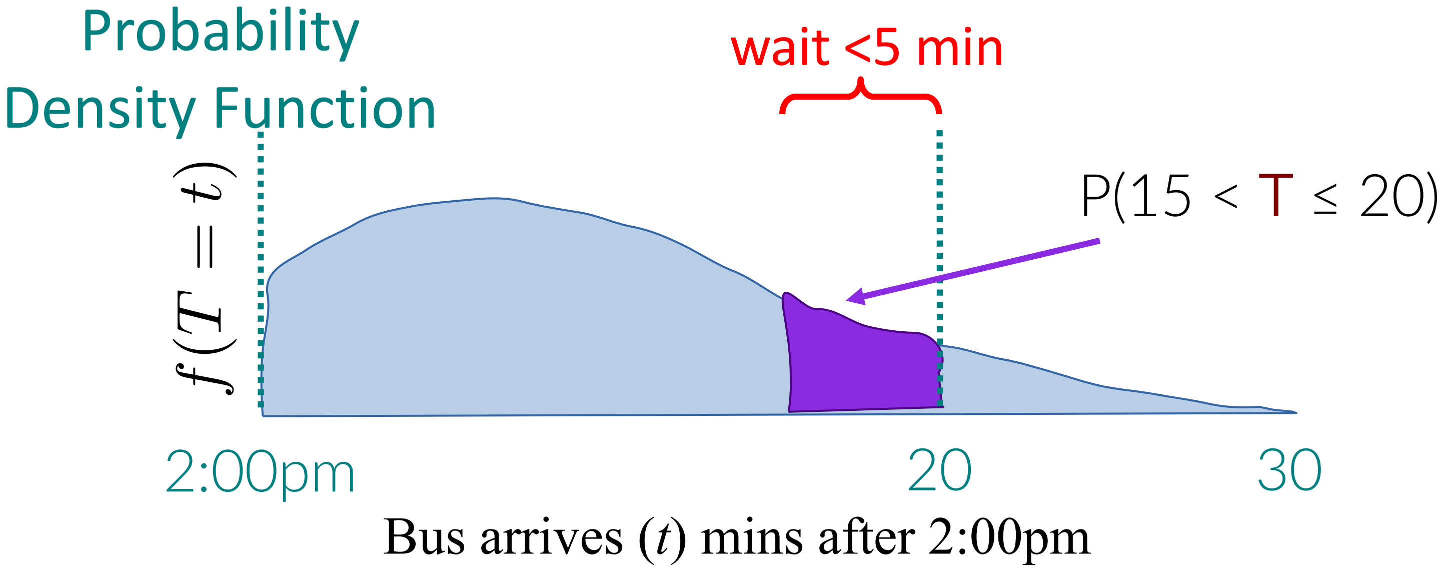


image: Haha169

Properties of PDFs

The integral of a PDF gives a probability. Thus:

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

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

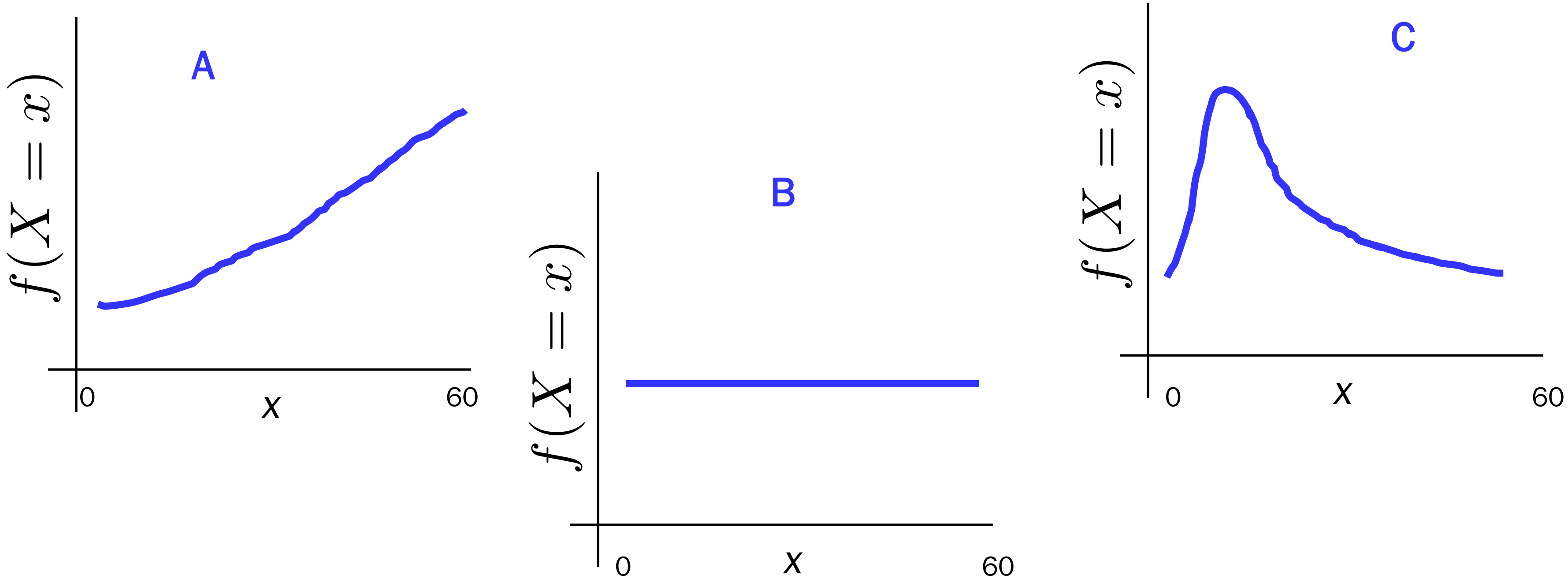
What do you get if you integrate over a probability *density* function?

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:

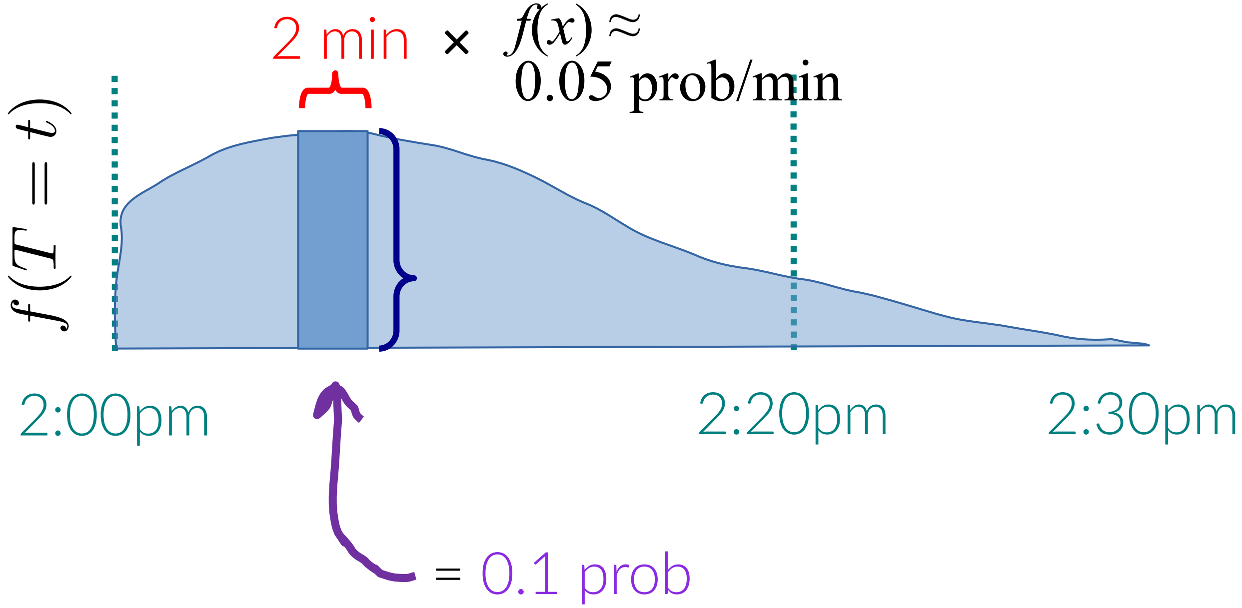


Which of these represent that you think the arrival is more likely to be close to 3:00pm

The ratio of probability densities is meaningful

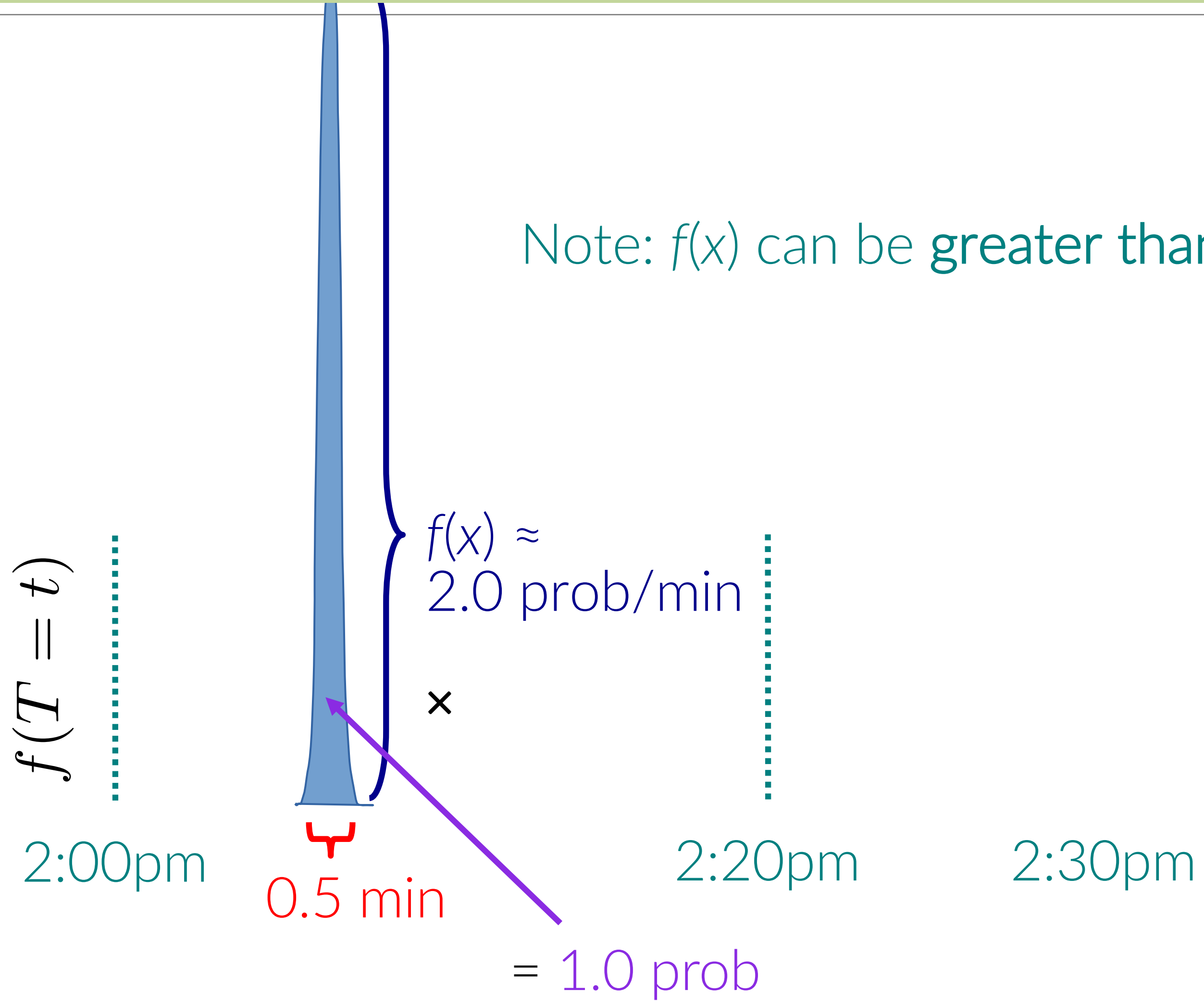
$f(X = x)$ is *Not* a Probability

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 at the value little *x*."

$$f(X = x)$$

Aka the PDF

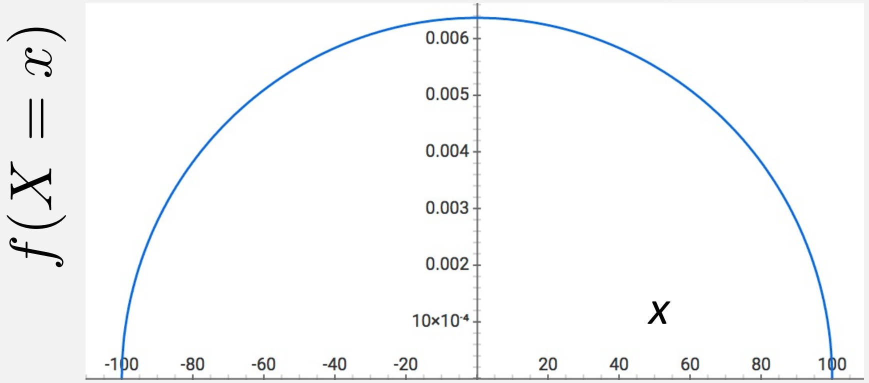
They are both measures of how likely X is to take on the value x.
Sometimes called the distribution function. Sometimes called the likelihood functions.

Rephrased as a Standard Continuous Problem

Let X be a continuous random variable[1]

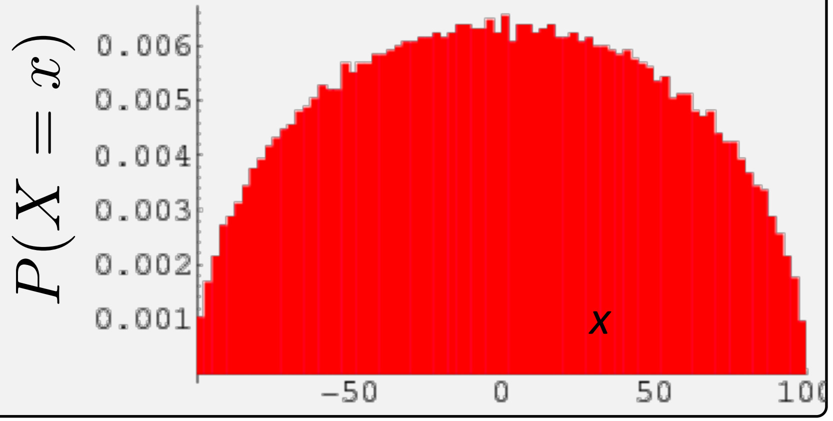
Theory

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



Practice

From simulations



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

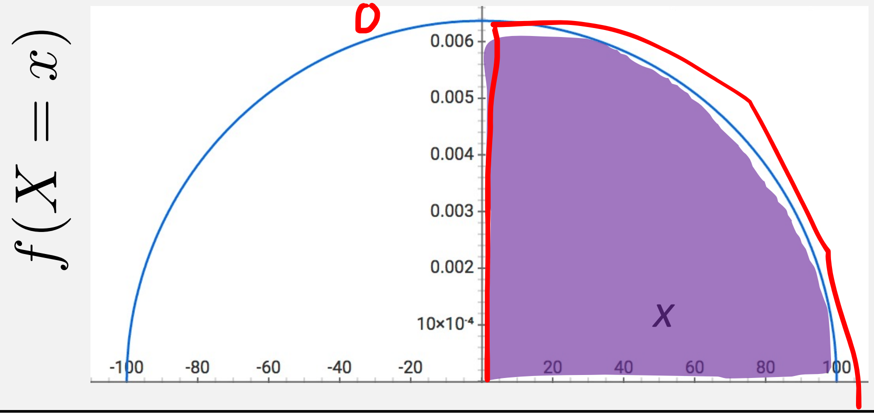
https://en.wikipedia.org/wiki/Wigner_semicircle_distribution

Simple Example from Quantum Physics

Let X be a continuous random variable[1]

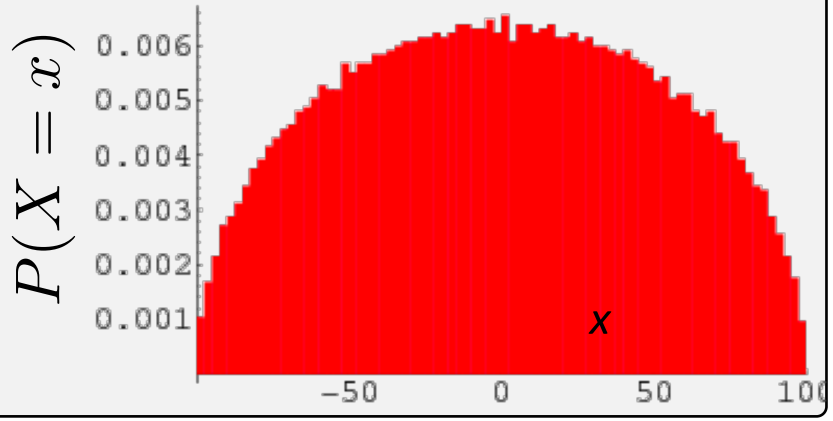
Theory

$$f(X = x) = \int_{0}^{100} \frac{1}{15708}\sqrt{100^2 - x^2}\, dx$$



Practice

From simulations



Approach #1: Integrate over the PDF

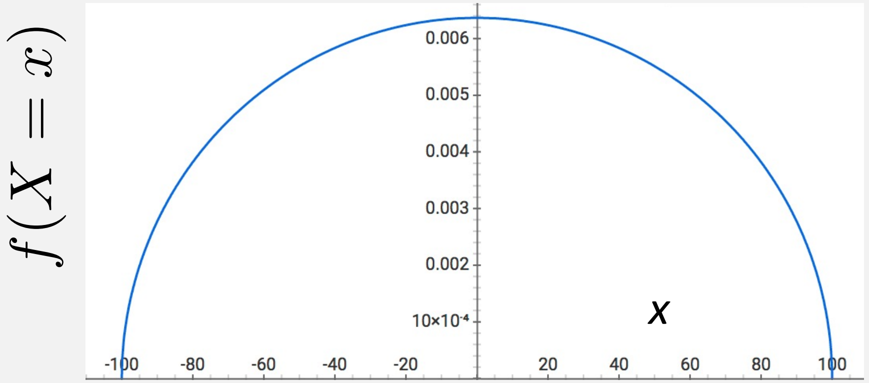
$$P(X > 0) = \int_{0}^{100} f(X = x)\, dx$$

Simple Example from Quantum Physics

Let X be a continuous random variable[1]

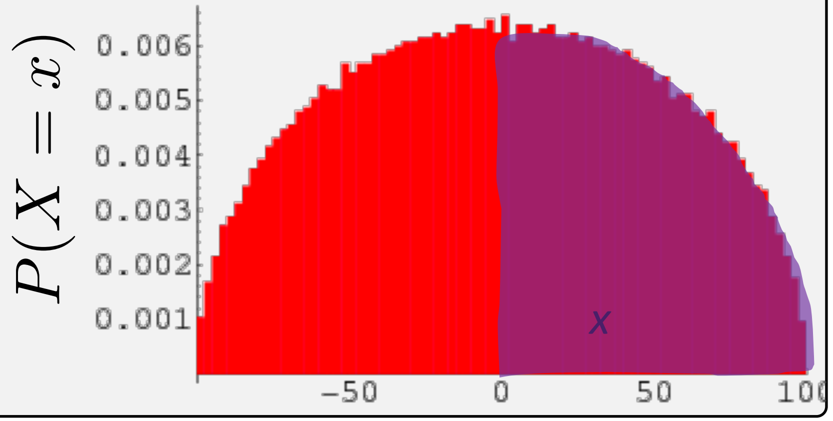
Theory

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



Practice

From simulations



Approach #2: Discrete Approximation

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

Simple Example from Quantum Physics

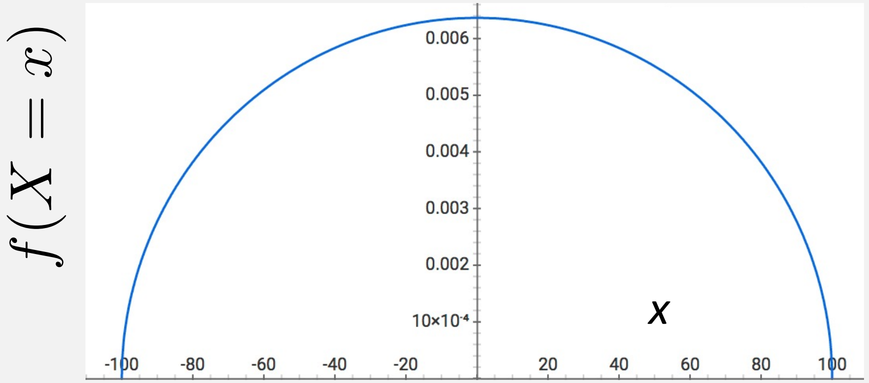
Let X be a continuous random variable[1]

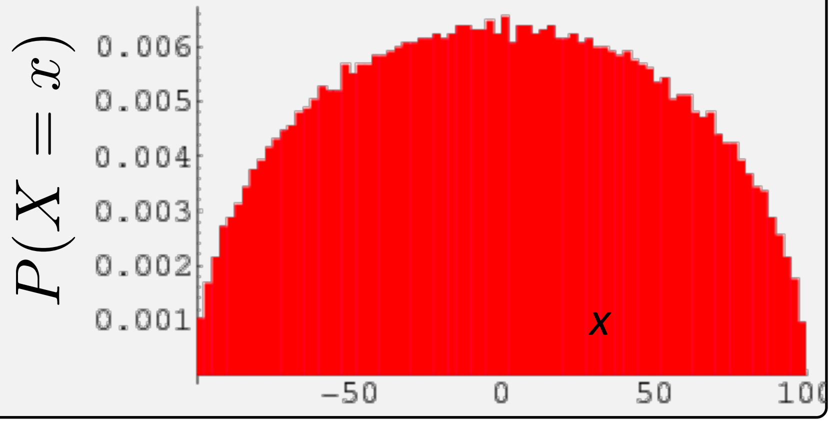
Theory

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

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!

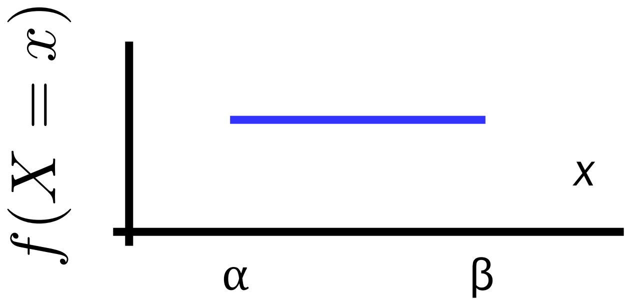
Uniform Random Variable

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

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

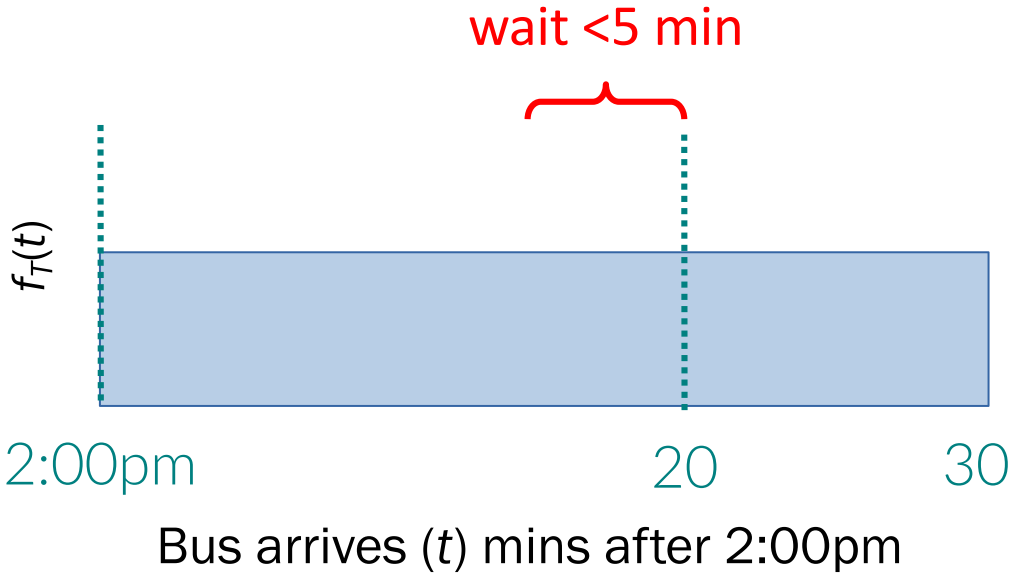
$$\mathrm{Var}(X) = \frac{(\beta - \alpha)^2}{12}$$

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)?

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



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

$$= \left. \frac{x}{\beta - \alpha} \right|_{15}^{20}$$

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

Exponential Random Variable

Consider an experiment that lasts a duration of time until success occurs.

def An **Exponential** random variable X is the amount of time until success.

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

Support: $[0, \infty)$

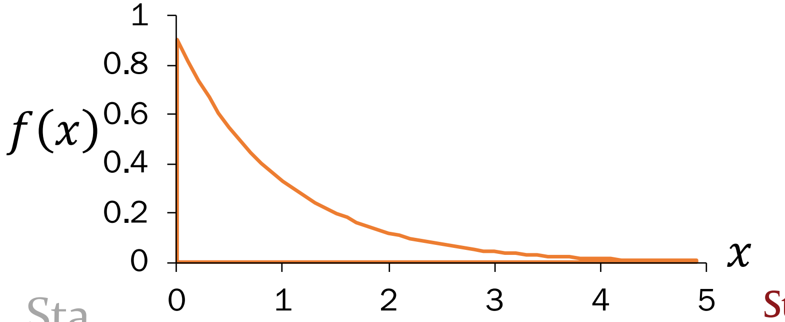
PDF $\quad f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$

Expectation $\quad E[X] = \dfrac{1}{\lambda}$

Variance $\quad \text{Var}(X) = \dfrac{1}{\lambda^2}$

Examples:

- Time until next earthquake
- Time for request to reach web server
- Time until end of cell phone contract



1906 Earthquake
Magnitude 7.8

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

ty

How Long Until the 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+

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

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

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

*In California, according to the USGS, 2015

ty

Integral Review

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

How Long Until the 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+

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

$$= 0.002^{-0.002y}$$

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

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

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

*In California, according to the USGS, 2015

How Long Until the 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+

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

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

How Long Until the 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