

# Inference 101

Chris Piech

CS109, Stanford University

# Announcements

---

- Midterm in 1 week and 1 day.
  - 7-9p in three rooms Hewlett 200, Hewlett 201, CEMEX. 33% the density of a normal exam.
  - Feel uncomfortable taking an exam in these rooms? Let us know by next Wednesday 9a. Get sick? Don't come! Let us know.
  - Need accommodations? Let us know before next Thursday 9a
  - Open notes, closed book, closed computer
  - Cumulative up until this Friday. Most emphasis will be on the topics in the first 3 psets, sections.


# https://cs109psets.netlify.app/win22/lecture12

Lecture 12 - Inference

cs109psets.netlify.app/win22/lecture12/elephant\_babies

## L12 Elephant Babies

Question: At birth, girl elephant weights are distributed as a Gaussian with mean 160kg, and standard deviation 7kg. At birth, boy elephant weights are distributed as a Gaussian with mean 165kg, and standard deviation of 3kg. All you know about a newborn elephant is that it is 163kg. What is the probability that it is a girl? Give your answer to three decimal places.



Previous Question      Next Question

Answer Editor      Solution

Numeric Answer: Enter your answer      Check Answer

Explanation:

Block LaTeX   Image   **B**   *I*   U

# Where are we in CS109?

---

## Overview of Topics



Counting  
Theory



Core  
Probability



Random  
Variables



Probabilistic  
Models



Uncertainty  
Theory



Machine  
Learning

# Where are we locally?

---

**Discrete  
Models:**

General Case,  
Multinomial

**Inference**

Conclusions  
from  
Observations

**Modelling:**

Make your own!

**General**

**Inference:**  
Use computers  
to infer

# Learning Goals

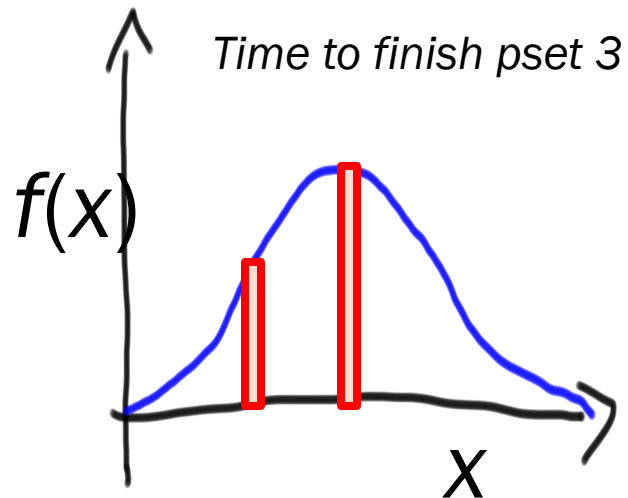
1. Know how to use a multinomial
2. Be able to calculate large bayes problems using a computer
3. Combine Bayes Theorem and Random Variables



# Relative Probability of Continuous Variables

$X =$  time to finish pset 3

$X \sim N(\mu = 10, \sigma^2 = 2)$

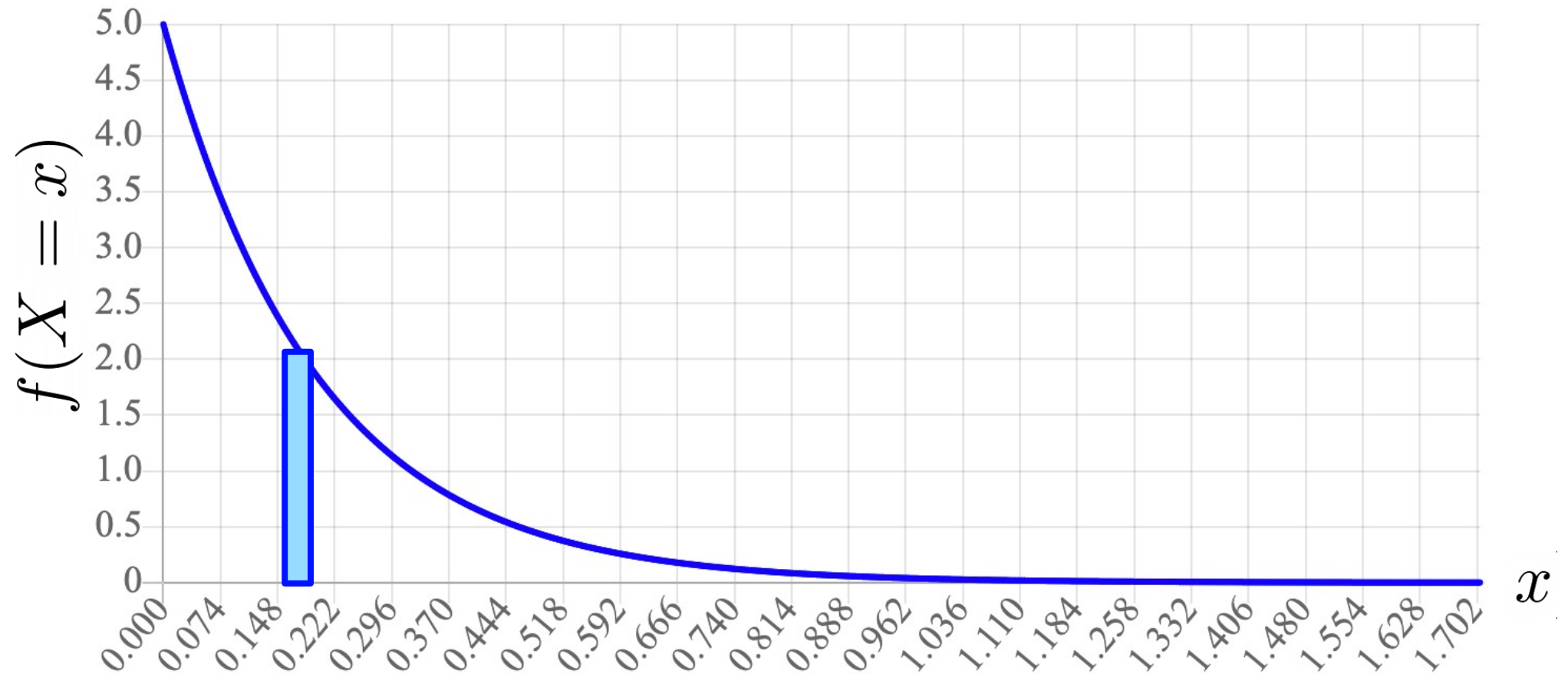


How much more likely are you to complete in 10 hours than in 5?

$$\begin{aligned} \frac{P(X = 10)}{P(X = 5)} &= \frac{\varepsilon f(X = 10)}{\varepsilon f(X = 5)} \\ &= \frac{f(X = 10)}{f(X = 5)} \\ &= \frac{\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(10-\mu)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(5-\mu)^2}{2\sigma^2}}} \\ &= \frac{\frac{1}{\sqrt{4\pi}} e^{-\frac{(10-10)^2}{4}}}{\frac{1}{\sqrt{4\pi}} e^{-\frac{(5-10)^2}{4}}} \\ &= \frac{e^0}{e^{-\frac{25}{4}}} = 518 \end{aligned}$$

# Calculation is based on this claim

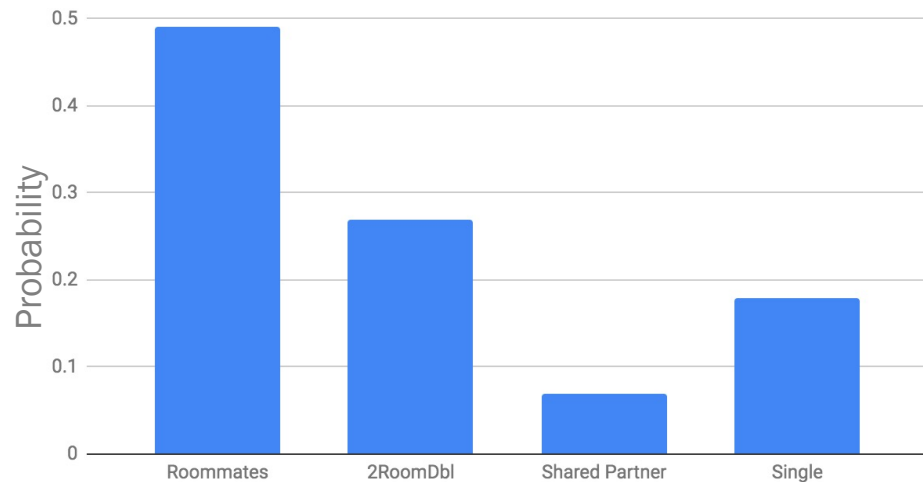
$$P(X = x) = f(X = x) \cdot \epsilon_x$$



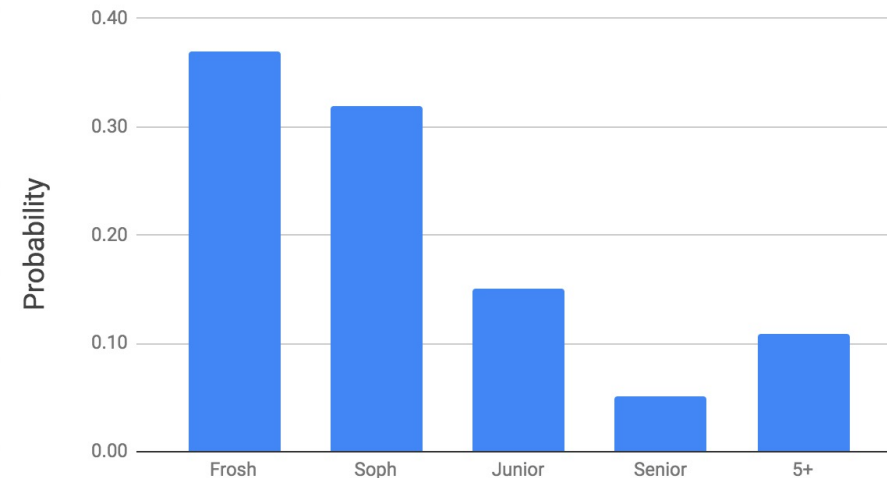
# Joint Probability Table

	Roommates	2RoomDbl	Shared Partner	Single	
Frosh	0.30	0.07	0.00	0.00	0.37
Soph	0.12	0.18	0.00	0.03	0.32
Junior	0.04	0.01	0.00	0.10	0.15
Senior	0.01	0.02	0.02	0.01	0.05
5+	0.02	0.00	0.05	0.04	0.11
	0.49	0.27	0.07	0.18	1.00

Marginal Room type



Marginal Year



# Last Week

---

## **Joint Distribution** *noun*

The probability of a simultaneous assignment to ***all*** the random variables in a probabilistic model.

***Eg:***

$$P(X = x, Y = y)$$

$$f(X = x, Y = y)$$

$$P(X = x, Y = y, \dots, Z = z)$$

Notation: These are all the same

---

$$P(X = x, Y = y)$$

$$P_{X,Y}(x, y)$$

$$P(x, y)$$

# The Multinomial

## Multinomial distribution

- $n$  independent trials of experiment performed
- Each trial results in one of  $m$  outcomes, with respective probabilities:  $p_1, p_2, \dots, p_m$  where
- $X_i =$  number of trials with outcome  $i$

$$\sum_{i=1}^m p_i = 1$$

$$P(X_1 = c_1, X_2 = c_2, \dots, X_m = c_m) = \binom{n}{c_1, c_2, \dots, c_m} p_1^{c_1} p_2^{c_2} \dots p_m^{c_m}$$

Joint distribution

Multinomial # ways of ordering the successes

Probabilities of each ordering are equal and mutually exclusive

where

$$\sum_{i=1}^m c_i = n$$

and

$$\binom{n}{c_1, c_2, \dots, c_m} = \frac{n!}{c_1! c_2! \dots c_m!}$$



End Review

Who wrote the federalist papers?



# Who wrote Federalist Paper 53?

## madison.txt

```
FOLDERS
  fedPapers
    /* answer.py
    hamilton.txt
    /* logPredict.py
    madison.txt
    /* predict.py
    /* process.py
    /* starter.py
    unknown.txt

1 To the People of the State of New York:
2
3 AMONG the numerous advantages promised by a
wellconstructed Union, none deserves to be more
accurately developed than its tendency to break
and control the violence of faction. The friend
of popular governments never finds himself so
much alarmed for their character and fate, as
when he contemplates their propensity to this
dangerous vice. He will not fail, therefore, to
set a due value on any plan which, without
violating the principles to which he is attached,
provides a proper cure for it. The instability,
injustice, and confusion introduced into the
public councils, have, in truth, been the mortal
diseases under which popular governments have
everywhere perished; as they continue to be the
favorite and fruitful topics from which the
adversaries to liberty derive their most specious
declamations. The valuable improvements made by
the American constitutions on the popular models,
both ancient and modern, cannot certainly be too
much admired; but it would be an unwarrantable
partiality, to contend that they have as
effectually obviated the danger on this side, as
was wished and expected. Complaints are
everywhere heard from our most considerate and
virtuous citizens, equally the friends of public
and private faith, and of public and personal
liberty, that our governments are too unstable,
that the public good is disregarded in the
conflicts of rival parties, and that measures are
too often decided, not according to the rules of
justice and the rights of the minor party, but by
the superior force of an interested and
overbearing majority. However anxiously we may
wish that these complaints had no foundation, the
evidence, of known facts will not permit us to
deny that they are in some degree true. It will
be found, indeed, on a candid review of our
situation, that some of the distresses under
which we labor have been erroneously charged on
the operation of our governments; but it will be
found, at the same time, that other causes will
not alone account for many of our heaviest
misfortunes; and, particularly, for that
prevailing and increasing distrust of public
```

## hamilton.txt

```
FOLDERS
  fedPapers
    /* answer.py
    hamilton.txt
    /* logPredict.py
    madison.txt
    /* predict.py
    /* process.py
    /* starter.py
    unknown.txt

1 The Utility of the Union in Respect to Commercial
Relations and a Navy
2 Hamilton for the Independent Journal.
3
4 To the People of the State of New York:
5 THE importance of the Union, in a commercial
light, is one of those points about which there
is least room to entertain a difference of
opinion, and which has, in fact, commanded the
most general assent of men who have any
acquaintance with the subject. This applies as
well to our intercourse with foreign countries as
with each other.
6
7 There are appearances to authorize a supposition
that the adventurous spirit, which distinguishes
the commercial character of America, has already
excited uneasy sensations in several of the
maritime powers of Europe. They seem to be
apprehensive of our too great interference in
that carrying trade, which is the support of
their navigation and the foundation of their
naval strength. Those of them which have colonies
in America look forward to what this country is
capable of becoming, with painful solicitude.
They foresee the dangers that may threaten their
American dominions from the neighborhood of
States, which have all the dispositions, and
would possess all the means, requisite to the
creation of a powerful marine. Impressions of
this kind will naturally indicate the policy of
fostering divisions among us, and of depriving
us, as far as possible, of an active commerce in
our own bottoms. This would answer the threefold
purpose of preventing our interference in their
navigation, of monopolizing the profits of our
trade, and of clipping the wings by which we
might soar to a dangerous greatness. Did not
prudence forbid the detail, it would not be
difficult to trace, by facts, the workings of
this policy to the cabinets of ministers.
8
9 If we continue united, we may counteract a policy
so unfriendly to our prosperity in a variety of
ways. By prohibitory regulations, extending, at
the same time, throughout the States, we may
oblige foreign countries to bid against each
```

## unknown.txt

```
FOLDERS
  fedPapers
    /* answer.py
    hamilton.txt
    /* logPredict.py
    madison.txt
    /* predict.py
    /* process.py
    /* starter.py
    unknown.txt

1 To the People of the State of New York:
2 I SHALL here, perhaps, be reminded of a current
observation, ``that where annual elections end,
tyranny begins.`` If it be true, as has often
been remarked, that sayings which become
proverbial are generally founded in reason, it is
not less true, that when once established, they
are often applied to cases to which the reason of
them does not extend. I need not look for a proof
beyond the case before us. What is the reason on
which this proverbial observation is founded? No
man will subject himself to the ridicule of
pretending that any natural connection subsists
between the sun or the seasons, and the period
within which human virtue can bear the temptations
of power. Happily for mankind, liberty is not, in
this respect, confined to any single point of
time; but lies within extremes, which afford
sufficient latitude for all the variations which
may be required by the various situations and
circumstances of civil society. The election of
magistrates might be, if it were found expedient,
as in some instances it actually has been, daily,
weekly, or monthly, as well as annual; and if
circumstances may require a deviation from the
rule on one side, why not also on the other side?
Turning our attention to the periods established
among ourselves, for the election of the most
numerous branches of the State legislatures, we
find them by no means coinciding any more in this
instance, than in the elections of other civil
magistrates. In Connecticut and Rhode Island, the
periods are half-yearly. In the other States,
South Carolina excepted, they are annual. In South
Carolina they are biennial as is proposed in the
federal government. Here is a difference, as four
to one, between the longest and shortest periods;
and yet it would be not easy to show, that
Connecticut or Rhode Island is better governed, or
enjoys a greater share of rational liberty, than
South Carolina; or that either the one or the
other of these States is distinguished in these
respects, and by these causes, from the States
whose elections are different from both. In
searching for the grounds of this doctrine, I can
discover but one, and that is wholly inapplicable
to our case. The important distinction so well
```

# Where to start?

---

We have words, we want to know probability of authorship. We also know probability of words given author...



Well hello again...

# Who wrote Federalist Paper 53?

---

Prob Document given Hamilton

Prior belief it was Hamilton

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Prob Hamilton given Document

Prob of the document???

# Who wrote Federalist Paper 53?

---

Model document as a multinomial where we care about count of words

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

# Who wrote Federalist Paper 53?

Loop over unique words

Prob hamilton would write word i

Number of times word i is in the doc

Prior belief it was Hamilton

Prob Hamilton given Document

Prob of the document???

$$P(H|D) = \frac{\binom{n}{c_1 \dots c_k} \cdot \prod_i h_i^{c_i} \cdot P(H)}{P(D)}$$

The diagram illustrates the components of the equation for the probability of Hamilton writing the document given the document's word counts. The binomial coefficient  $\binom{n}{c_1 \dots c_k}$  is linked to the text 'Number of times word i is in the doc'. The product  $\prod_i h_i^{c_i}$  is linked to 'Loop over unique words' and 'Prob hamilton would write word i'. The prior probability  $P(H)$  is linked to 'Prior belief it was Hamilton'. The denominator  $P(D)$  is linked to 'Prob of the document???' and 'Prob Hamilton given Document'.

# Who wrote Federalist Paper 53?

Prob that Hamilton wrote it

$$\begin{aligned} P(H|D) &= \frac{P(D|H)P(H)}{P(D)} \\ &= \frac{P(H) \cdot \binom{n}{c_1 \dots c_m} \cdot \prod_i h_i^{c_i}}{P(D)} \end{aligned}$$

Prob that Madison wrote it

$$\begin{aligned} P(M|D) &= \frac{P(D|M)P(M)}{P(D)} \\ &= \frac{P(M) \cdot \binom{n}{c_1 \dots c_m} \cdot \prod_i m_i^{c_i}}{P(D)} \end{aligned}$$

$$\begin{aligned} \frac{P(H|D)}{P(M|D)} &= \frac{P(H) \cdot \binom{n}{c_1 \dots c_m} \cdot \prod_i h_i^{c_i}}{P(M) \cdot \binom{n}{c_1 \dots c_m} \cdot \prod_i m_i^{c_i}} \\ &= \frac{\prod_i h_i^{c_i}}{\prod_i m_i^{c_i}} \end{aligned}$$

# To the code

---



What happened?

# All our probabilities are zero...



# Use logs when probabilities become too small!

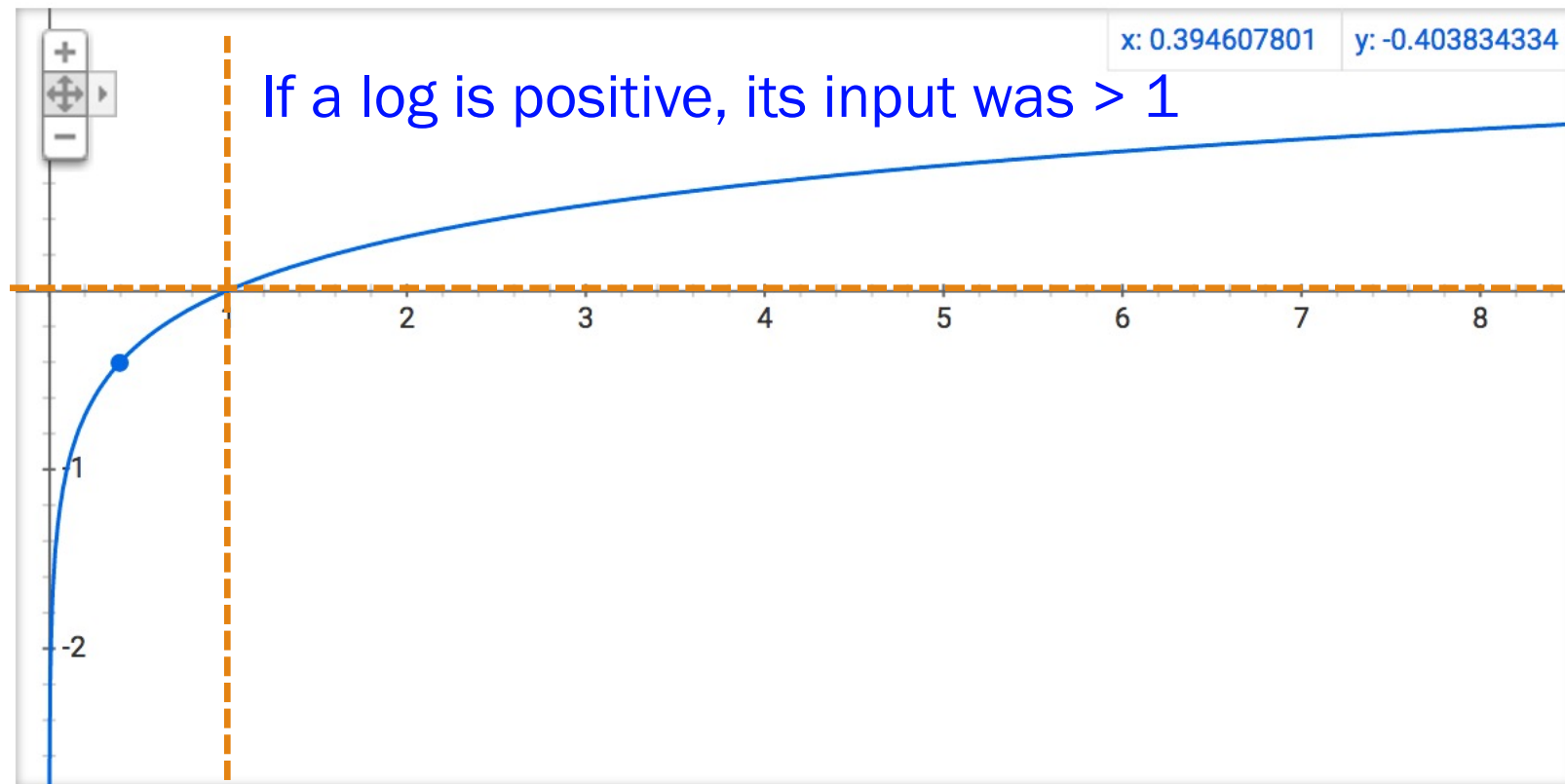
---

$$\frac{P(H|D)}{P(M|D)} = \frac{\prod_i m_i^{c_i}}{\prod_i h_i^{c_i}}$$

$$\begin{aligned}\log \frac{P(H|D)}{P(M|D)} &= \log \frac{\prod_i h_i^{c_i}}{\prod_i m_i^{c_i}} \\ &= \sum_i \log h_i^{c_i} - \sum_i \log m_i^{c_i} \\ &= \sum_i c_i \cdot \log h_i - \sum_i c_i \log m_i\end{aligned}$$

# What does it mean if a log value is positive / negative

Graph for  $\log(x)$



If a log is negative, its input was between 0 and 1

[More info](#)

woot

# Joint Random Variables



Use a joint table, or joint function to solve probability question



Think about **conditional** probabilities with joint variables (which might be continuous)



Use and find **independence** of random variables

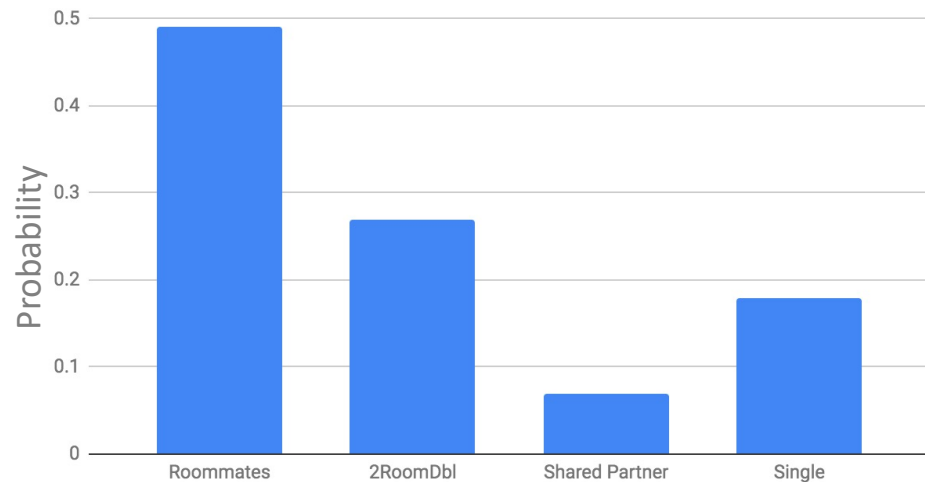


Use and find **expectation** of random variables

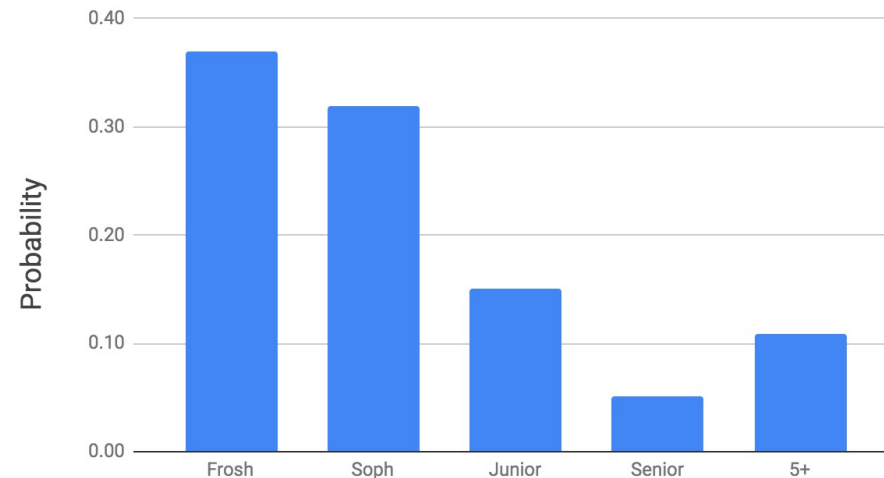
# Joint Probability Table

	Roommates	2RoomDbI	Shared Partner	Single	
Frosh	0.30	0.07	0.00	0.00	0.37
Soph	0.12	0.18	0.00	0.03	0.32
Junior	0.04	0.01	0.00	0.10	0.15
Senior	0.01	0.02	0.02	0.01	0.05
5+	0.02	0.00	0.05	0.04	0.11
	0.49	0.27	0.07	0.18	1.00

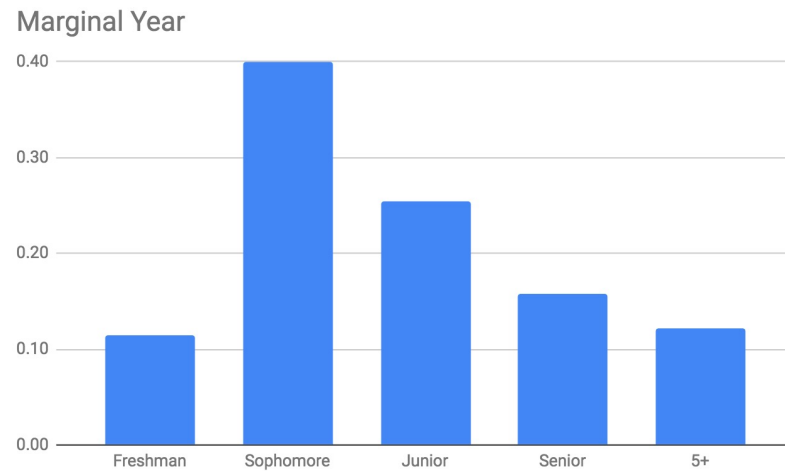
Marginal Room type



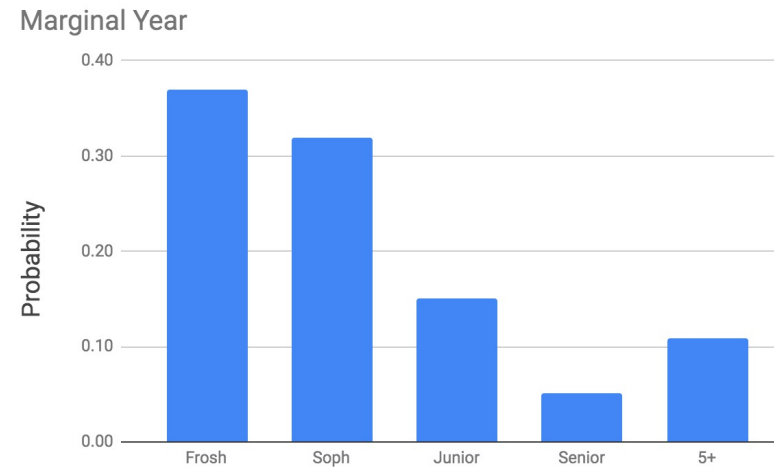
Marginal Year



# Change in Marginal Year



Fall quarter '18



Spr quarter '19

# Joint Random Variables



Use a joint table, or joint function to solve probability question



Think about **conditional** probabilities with joint variables (which might be continuous)



Use and find **independence** of random variables



Use and find **expectation** of random variables

# Today: Inference

---

## **Inference** *noun*

An updated belief about a random variable (or multiple) based on conditional knowledge regarding another random variable (or multiple) in a probabilistic model.

TLDR: conditional probability with random variables.

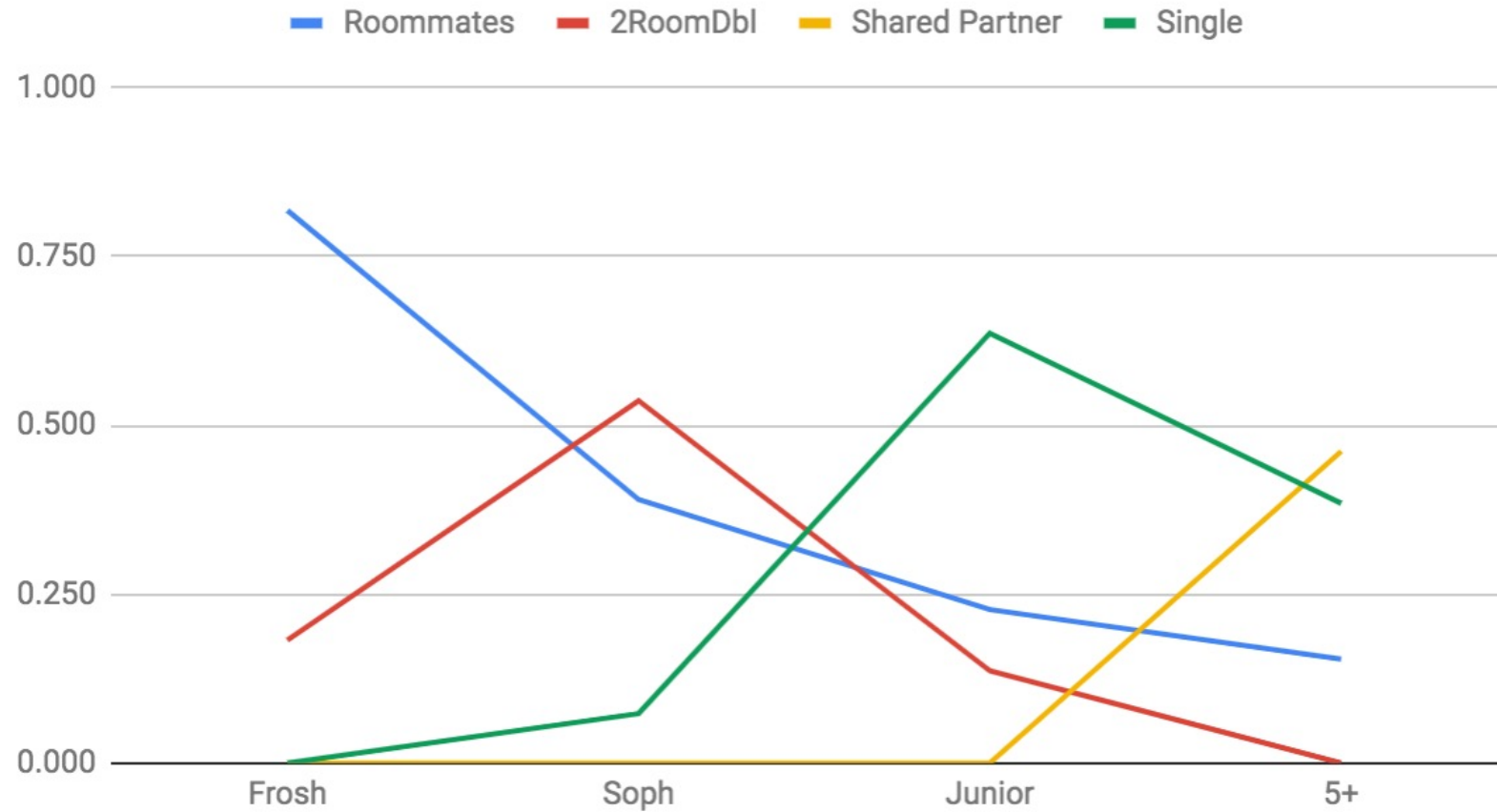
# Warmup Inference

---

Q: What is the probability that someone has a single, given that they are a senior?

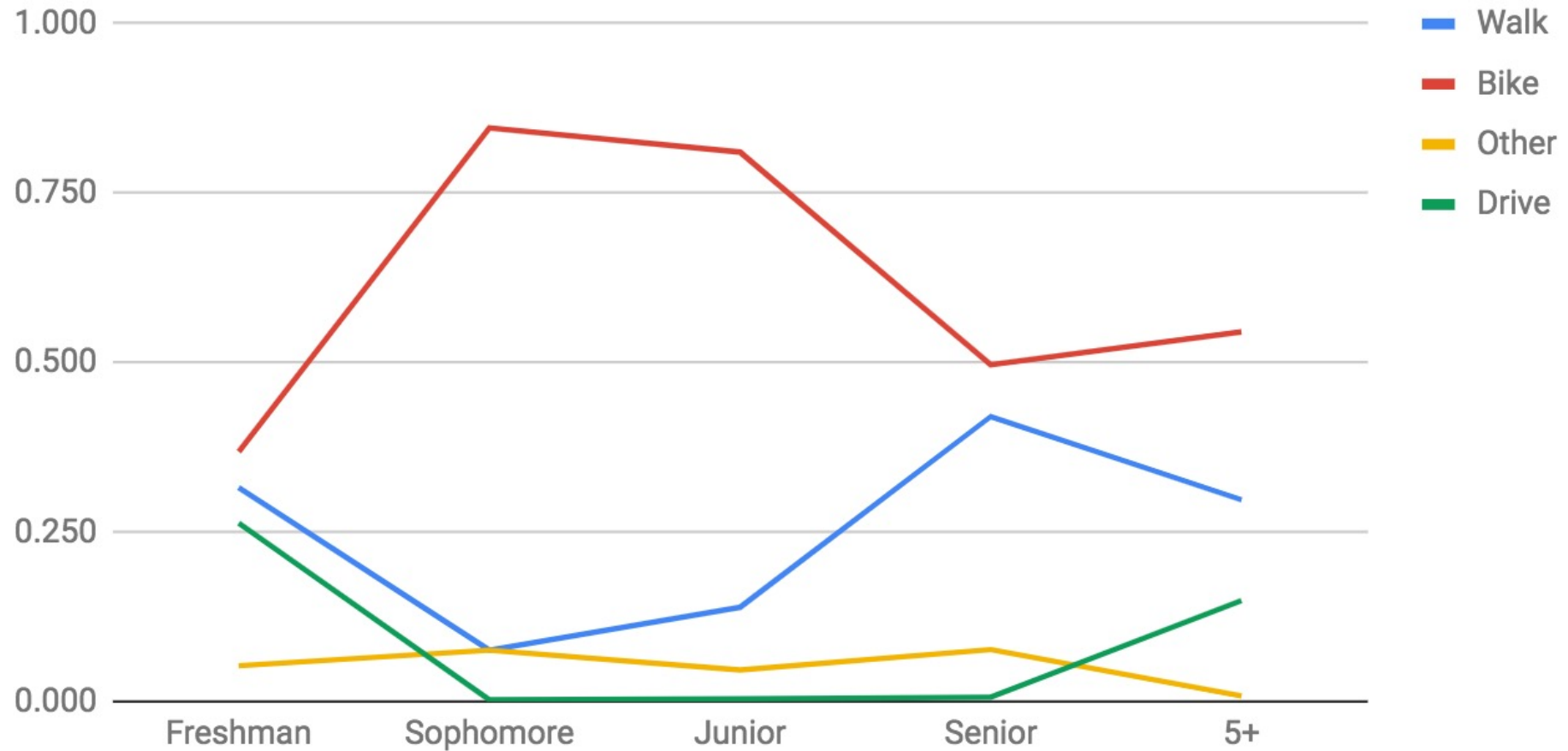
# Room | Year

$P(\text{Room} | \text{Year})$



# Transport | Year

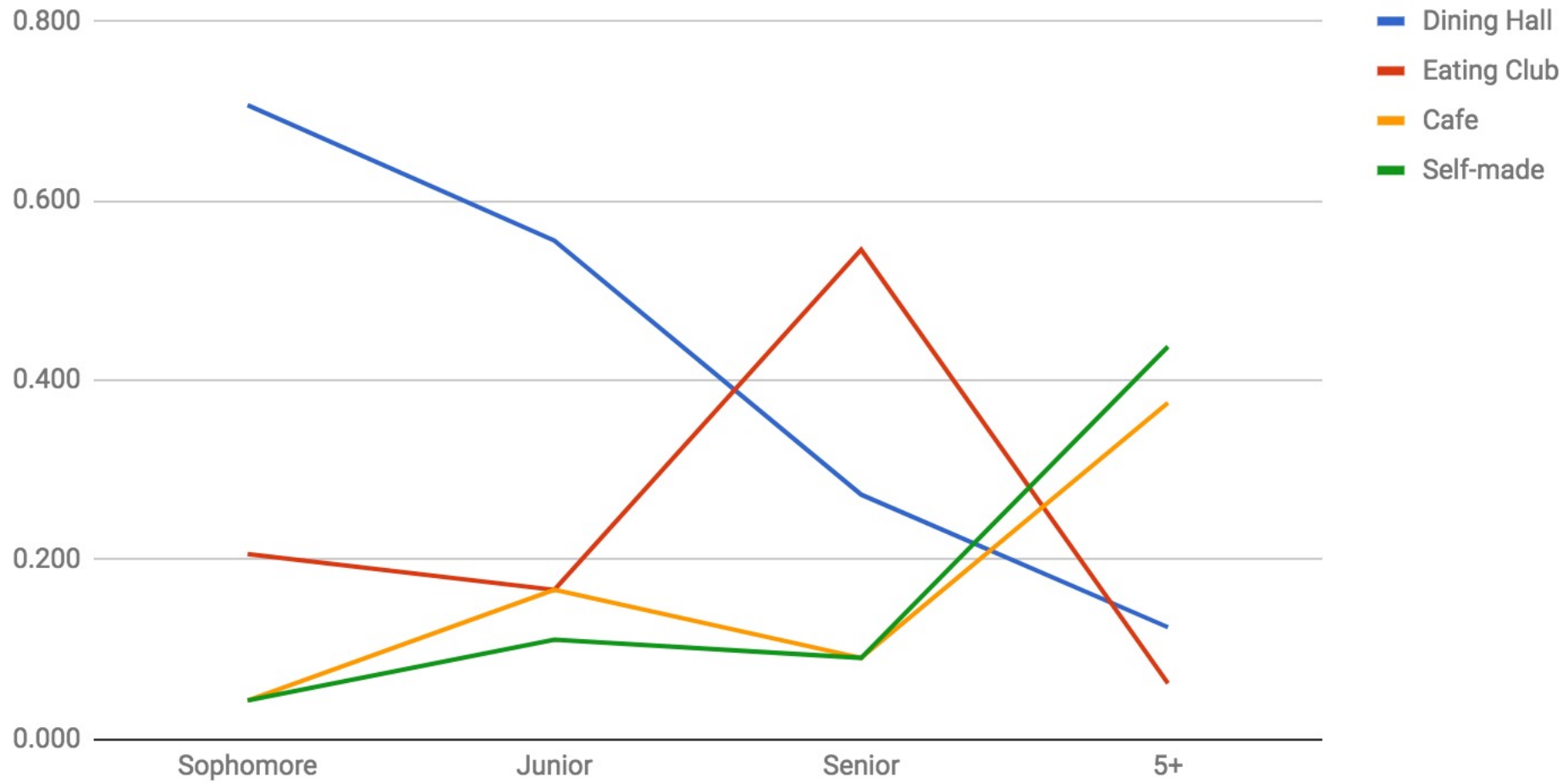
## Transport | Year



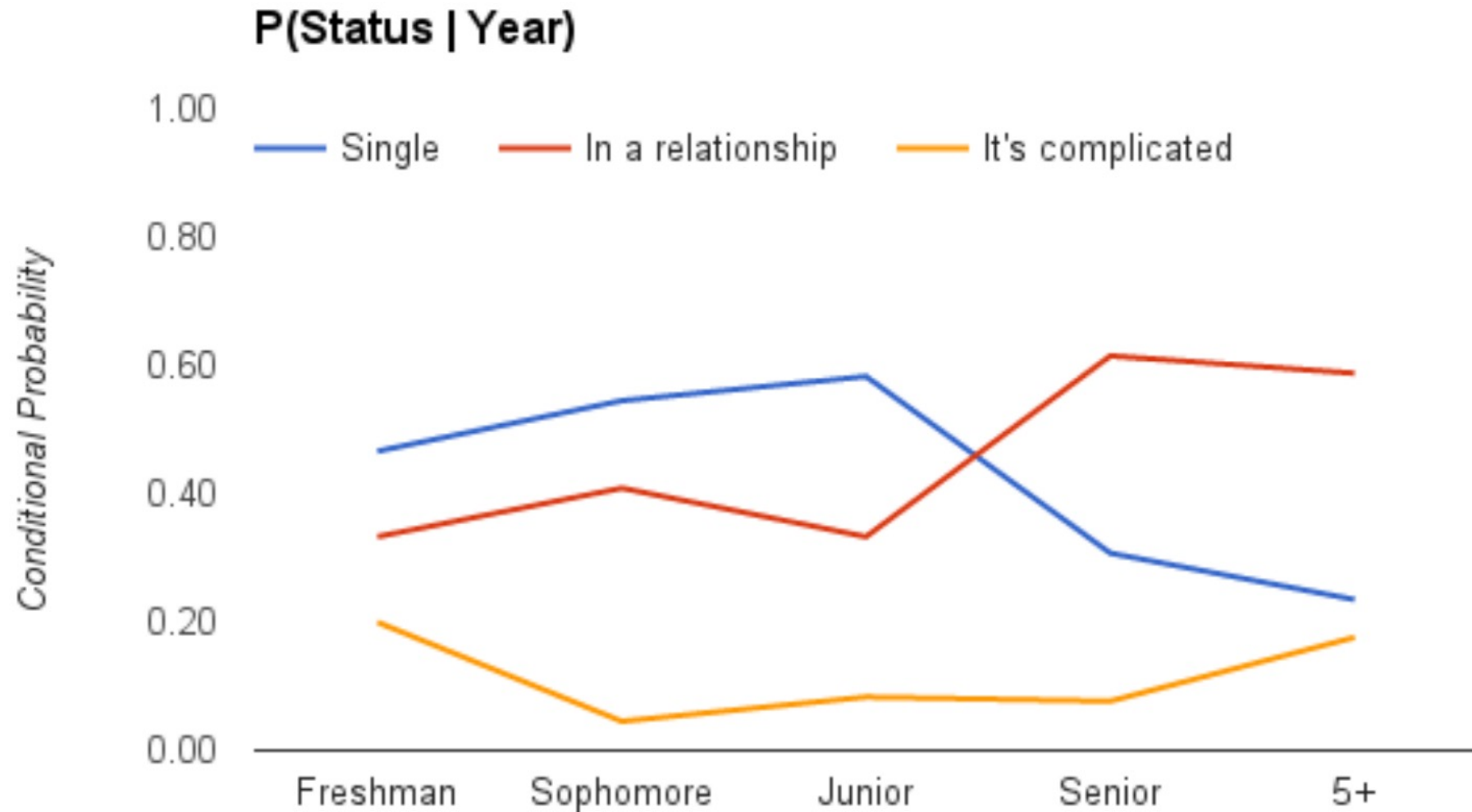
Conditional Probability Table

# Lunch | Year

## Lunch Type | Year



# Relationship Status | Year



# Number or Function?

$$P(X = 2 | Y = 5)$$

Number

# Number or Function?

$$P(X = x | Y = 2)$$

## Random Variable

(also a function or 1D table)

# Number or Function?

$$P(X = x | Y = y)$$

2D Function  
(or 2D table)

Pedagogical Pause

# Warmup: Bayes Revisited

$$P(B|E) = \frac{P(E|B) P(B)}{P(E)}$$

Posterior belief

Likelihood of evidence

Prior belief

Normalization constant

# Bayes Theorem with Discrete

Let  $M$  be a **discrete** random variable

Let  $N$  be a **discrete** random variable

$$P(M = 2|N = 3) = \frac{P(N = 3|M = 2)P(M = 2)}{P(N = 3)}$$

$$P(M = m|N = n) = \frac{P(N = n|M = m)P(M = m)}{P(N = n)}$$

More  
generally

Shorthand  
notation

$$P(m|n) = \frac{P(n|m)P(m)}{P(n)}$$



# I Heard That



Let  $X$  be the **change in gaze** (measured in degrees) over 3 seconds after a sound is played

Value of $X$	PMF of $X$ given Baby can hear the sound	PMF of $X$ given Baby can <b>not</b> hear the sound
0 to 5	0.08	0.40
5 to 10	0.15	0.30
10 to 15	0.35	0.12
15 to 20	0.20	0.08
20 to 25	0.12	0.05
Above 25	0.10	0.05

$$P(\text{can hear the sound}) = \frac{3}{4}$$

You observe  $X = 0$ . What is the probability the baby **can** hear the sound?

# I Heard That

Value of $X$	PMF of $X$ given Baby can hear the sound	PMF of $X$ given Baby can <b>not</b> hear the sound
0 to 5	0.08	0.40
5 to 10	0.15	0.30
10 to 15	0.35	0.12
15 to 20	0.20	0.08
20 to 25	0.12	0.05
Above 25	0.10	0.05

$$P(\text{can hear the sound}) = \frac{3}{4}$$

You observe  $X = 0$ . What is the probability the baby **can** hear the sound?

$$P(Y = 1|X = 0) = \frac{P(X = 0|Y)P(Y)}{P(X = 0|Y)P(Y) + P(X = 0|Y^C)P(Y^C)}$$

$$P(Y = 1|X = 0) = \frac{0.08 * 0.75}{0.08 * 0.75 + 0.40 * 0.25} = \frac{3}{8}$$

# Inference with Continuous

---

Q: At birth, girl elephant weights are distributed as a Gaussian with mean = 160kg, std = 7kg. At birth, boy elephant weights are distributed as a Gaussian with mean = 165kg, std = 3kg. All you know about a newborn elephant is that it is 163kg. What is the probability that it is a girl?



# Inference with Continuous



Q: At birth, girl elephant weights are distributed as a Gaussian with mean = 160kg, std = 7kg. At birth, boy elephant weights are distributed as a Gaussian with mean = 165kg, std = 3kg. All you know about a newborn elephant is that it is 163kg. What is the probability that it is a girl?

Model:

Let  $G$  be an indicator that the elephant is a girl.  $G$  is  $\text{Bern}(p = 0.5)$

Let  $X$  be the distribution of weight of the elephant.

$X \mid G = 1$  is  $N(\mu = 160, \sigma^2 = 7^2)$

$X \mid G = 0$  is  $N(\mu = 165, \sigma^2 = 3^2)$

# Inference with Continuous



Q: What is  $P(G = 1 \mid X = 163)$

Let  $G$  be an indicator that the elephant is a girl.  $G$  is  $\text{Bern}(p = 0.5)$

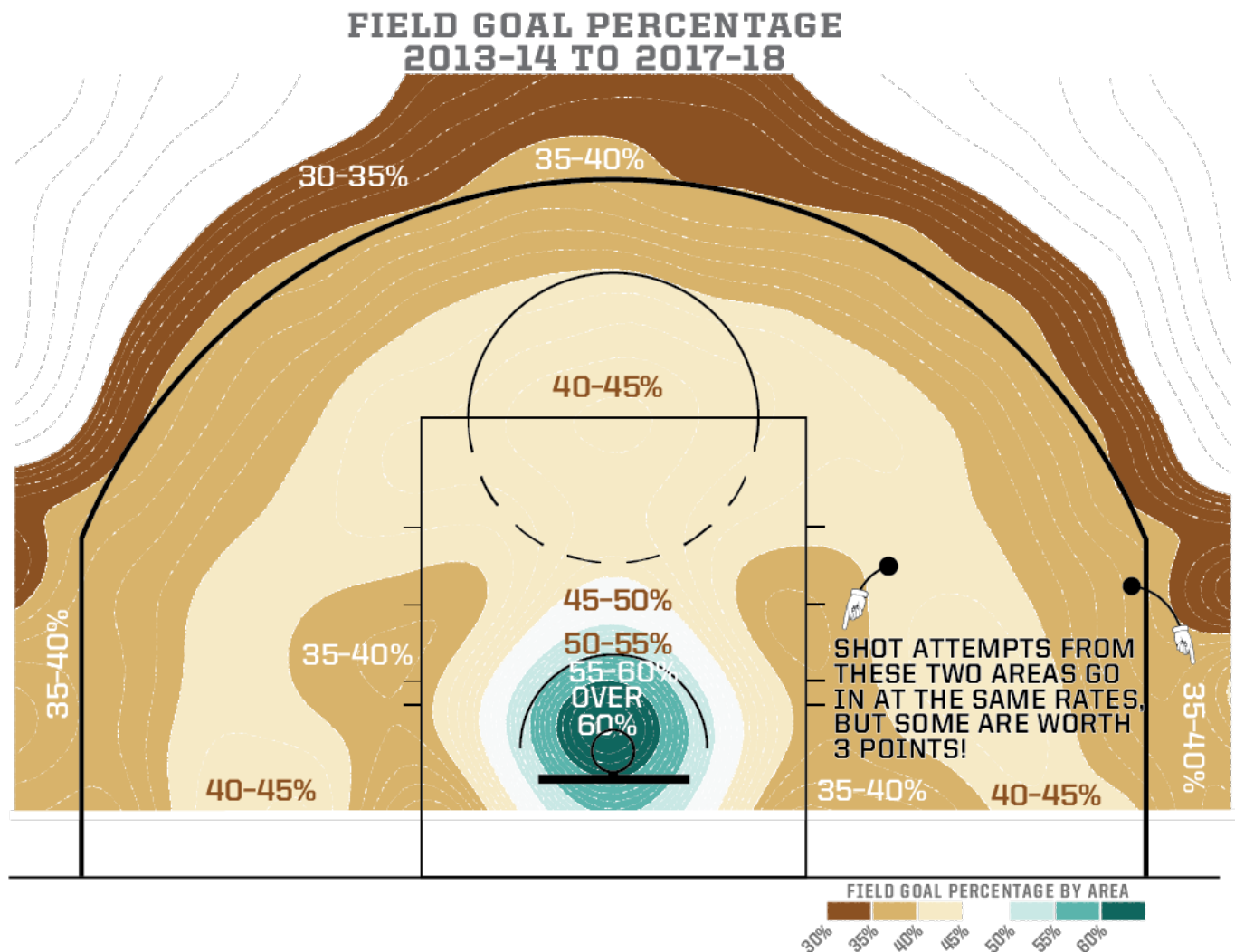
Let  $X$  be the distribution of weight of the elephant.

$X \mid G = 1$  is  $N(\mu = 160, \sigma^2 = 7^2)$

$X \mid G = 0$  is  $N(\mu = 165, \sigma^2 = 3^2)$

Aside: Models with continuous RVs

# Probabilistic Models can have Continuous Random Vars



X location  
of the shot

Y location  
of the shot

$$f(X = x, Y = y)$$

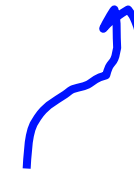
Joint: If any random variable is continuous, then we consider the joint a density

# Joint is Complete Information!

---



A joint distribution is complete information. It can be used to answer any probability question.



Still true when some variables are continuous

# Mixing Discrete and Continuous

Let  $X$  be a **continuous** random variable

Let  $N$  be a **discrete** random variable

$$P(N = n|X = x) = \frac{P(X = x|N = n)P(N = n)}{P(X = x)}$$

$$P(N = n|X = x) = \frac{f(X = x|N = n) \cdot \epsilon \cdot P(N = n)}{f(X = x) \cdot \epsilon}$$

$$P(N = n|X = x) = \frac{f(X = x|N = n) \cdot P(N = n)}{f(X = x)}$$

# Mixing Discrete and Continuous

Let  $X$  be a **continuous** random variable

Let  $N$  be a **discrete** random variable

$$P(X = x|N = n) = \frac{P(N = n|X = x)P(X = x)}{P(N = n)}$$

$$P(x|n) = \frac{P(n|x)P(x)}{P(n)}$$

$$f(x|n) \cdot \epsilon_x = \frac{P(n|x)f(x) \cdot \epsilon_x}{P(n)}$$

Change notation



$$f(x|n) = \frac{P(n|x)f(x)}{P(n)}$$

# All the Bayes Belong to Us

---

**M, N are discrete. X, Y are continuous**

OG Bayes

$$P(m|n) = \frac{P(n|m)P(m)}{P(n)}$$

Mix Bayes #1

$$f(x|n) = \frac{P(n|x)f(x)}{P(n)}$$

Mix Bayes #2

$$P(n|x) = \frac{f(x|n)P(n)}{f(x)}$$

$$f(x|y) = \frac{f(y|x)f(x)}{f(y)}$$

# LOTP? Chain Rule? You can play too!

---

**N is discrete. X is continuous**

$$f(N = n, X = x) = f(X = x | N = n)P(N = n)$$

$$f(X = x) = \sum_n f(X = x | N = n)P(N = n)$$

End Aside

# Inference with Continuous



Q: What is  $P(G = 1 \mid X = 163)$

Let  $G$  be an indicator that the elephant is a girl.  $G$  is  $\text{Bern}(p = 0.5)$

Let  $X$  be the distribution of weight of the elephant.

$X \mid G = 1$  is  $N(\mu = 160, \sigma^2 = 7^2)$

$X \mid G = 0$  is  $N(\mu = 165, \sigma^2 = 3^2)$

# Inference with Continuous



Q: What is  $P(G = 1 \mid X = 163)$

Let  $G$  be an indicator that the elephant is a girl.  $G$  is  $\text{Bern}(p = 0.5)$

Let  $X$  be the distribution of weight of the elephant.

$X \mid G = 1$  is  $N(\mu = 160, \sigma^2 = 7^2)$

$X \mid G = 0$  is  $N(\mu = 165, \sigma^2 = 3^2)$

Joint Distribution is Implied:

$$f(G = 1, X = 72.3) = f(X = 72.3 \mid G = 1)P(G = 1)$$

$$f(G = g, X = x) = f(X = x \mid G = g)P(G = g)$$

More generally

# Inference with Continuous



Q: What is  $P(G = 1 \mid X = 163)$

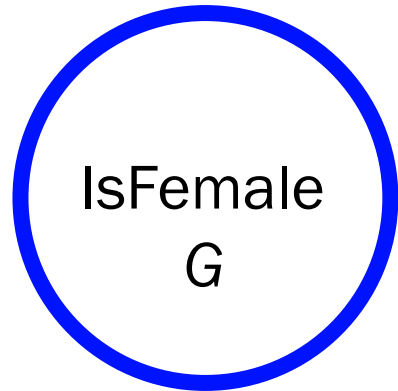
Let  $G$  be an indicator that the elephant is a girl.  $G$  is  $\text{Bern}(p = 0.5)$

Let  $X$  be the distribution of weight of the elephant.

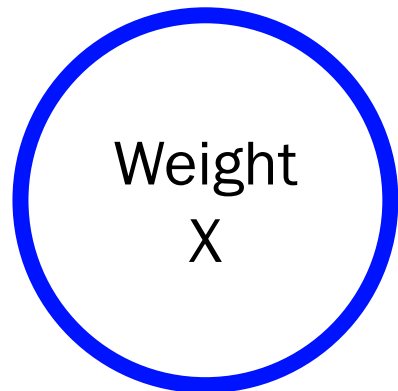
$X \mid G = 1$  is  $N(\mu = 160, \sigma^2 = 7^2)$

$X \mid G = 0$  is  $N(\mu = 165, \sigma^2 = 3^2)$

# Model Shown Graphically



$G = 1$  is  $\text{Bern}(p = 0.5)$



$X | G = 1$  is  $N(\mu = 160, \sigma^2 = 7^2)$

$X | G = 0$  is  $N(\mu = 165, \sigma^2 = 3^2)$

Does this define the joint?

$$f(G = g, X = x)$$

$$= f(X = x | G = g) P(G = g)$$

Q: What is  $P(G = 1 | X = 163)$

# I Heard That Redux

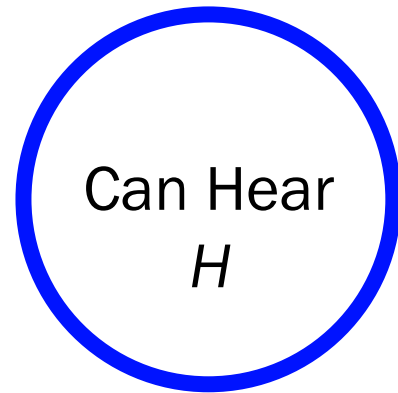
---

**Normal Assumption:** We choose to approximate eye movements with normal distributions. For babies who can hear sounds, we approximate their gaze movement after the sound is played as:  $N(\mu = 15, \sigma^2 = 50)$ . For babies who can **not** hear sounds, we approximate gaze movement as  $N(\mu = 8, \sigma^2 = 50)$ .

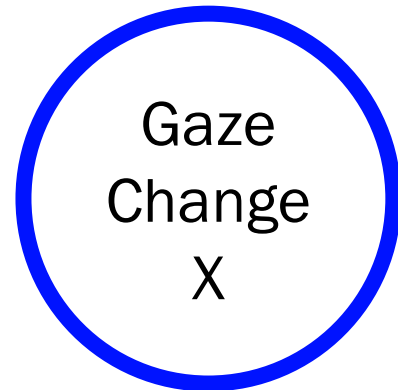
For a new baby we observe a 14 degree movement after the sound is played. What is your belief that a baby can hear, under The **Normal Assumption**?

# I Heard That Redux

---



$H = 1$  is  $\text{Bern}(p = 0.75)$



$X | H = 1$  is  $\text{N}(\mu = 15, \sigma^2 = 50)$

$X | H = 0$  is  $\text{N}(\mu = 8, \sigma^2 = 50)$

Q: What is  $P(H = 1 | X = 14)$

# Joint Random Variables



Use a joint table, or joint function to solve probability question



Think about **conditional** probabilities with joint variables (which might be continuous)



Use and find **independence** of random variables



Use and find **expectation** of random variables