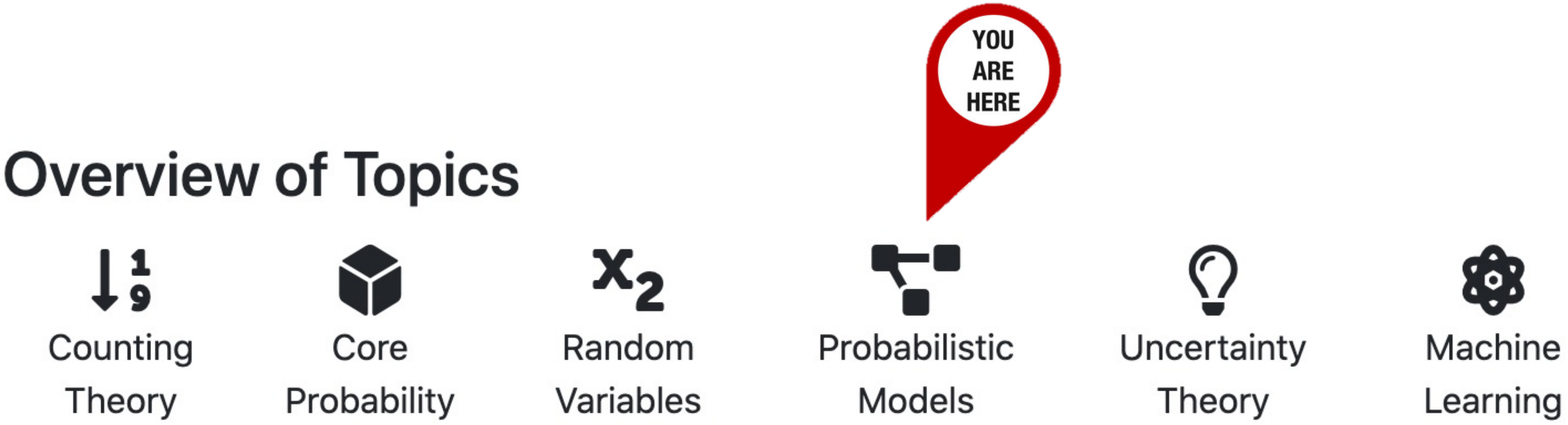
Intro to Inference

Will Song
CS109, Stanford University

Where are we in CS109?



Where are we locally?

Discrete Models: Joints, Multinomial

Inference Change RV belief from Observations

Modelling: Make your own!

General Inference: Use computers to infer

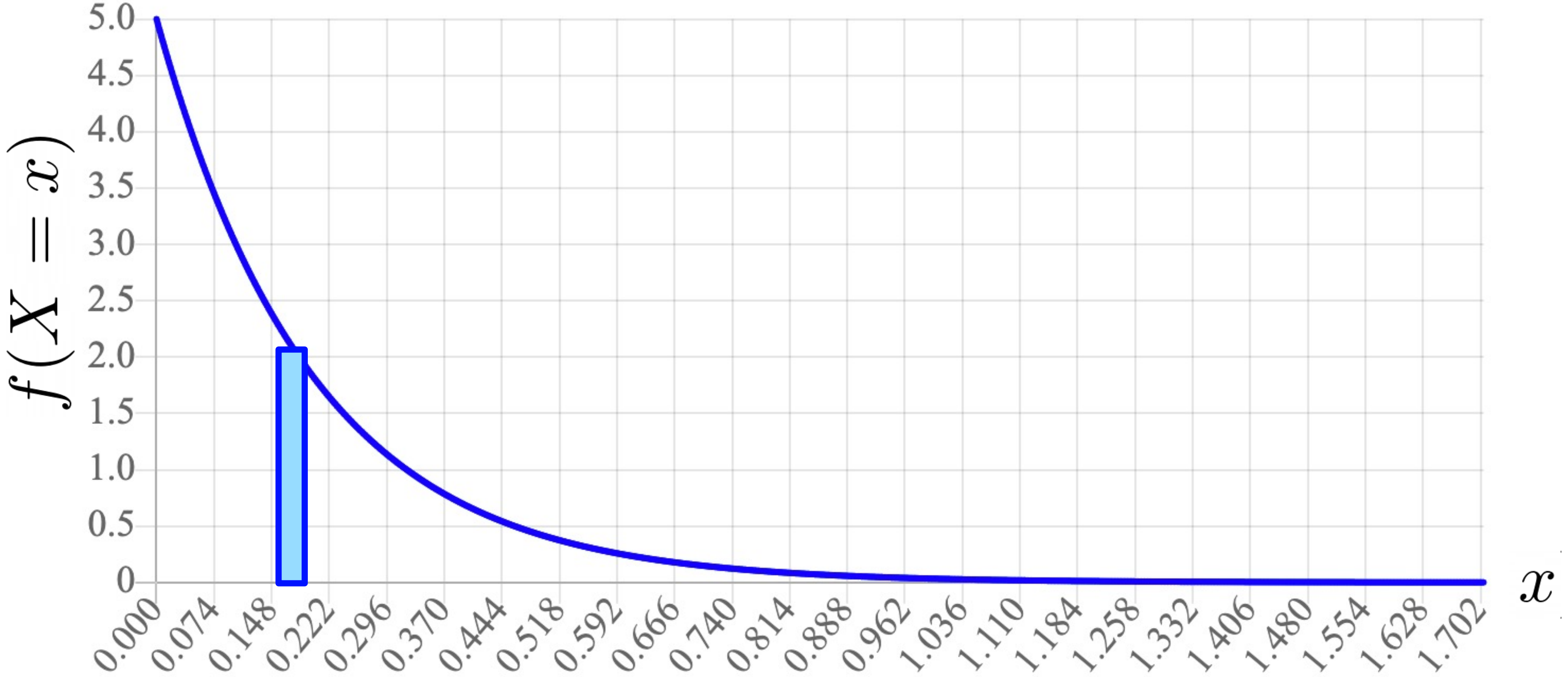
Learning Goals

1. More Familiarity with Multiple Random Variables!
2. Use Multiple Random Variables for Probability Problems!
3. Combine Bayes Theorem and Random Variables!

Review

Epsilon: Useful perspective

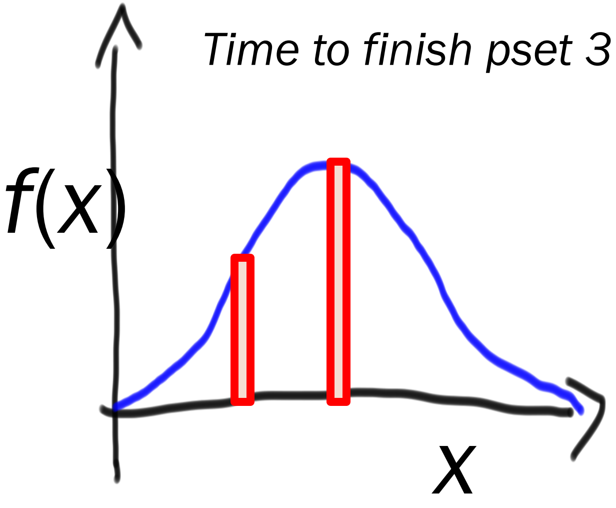
$$P(X = x) = f(X = x) \cdot \epsilon_x$$



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?

$$\frac{P(X=10)}{P(X=5)} = \frac{\varepsilon f(X=10)}{\varepsilon f(X=5)}$$

$$= \frac{f(X=10)}{f(X=5)}$$

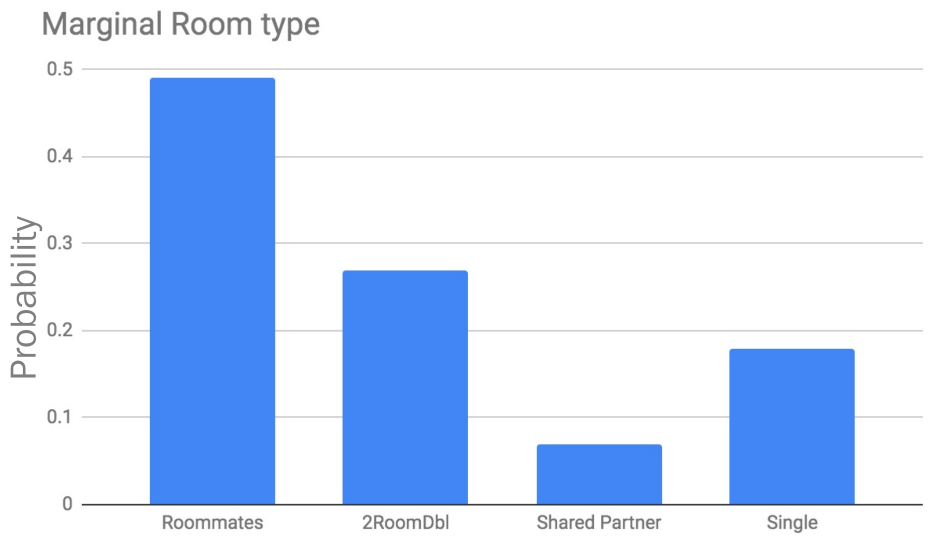
$$= \frac{\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(10-\mu)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(5-\mu)^2}{2\sigma^2}}}$$

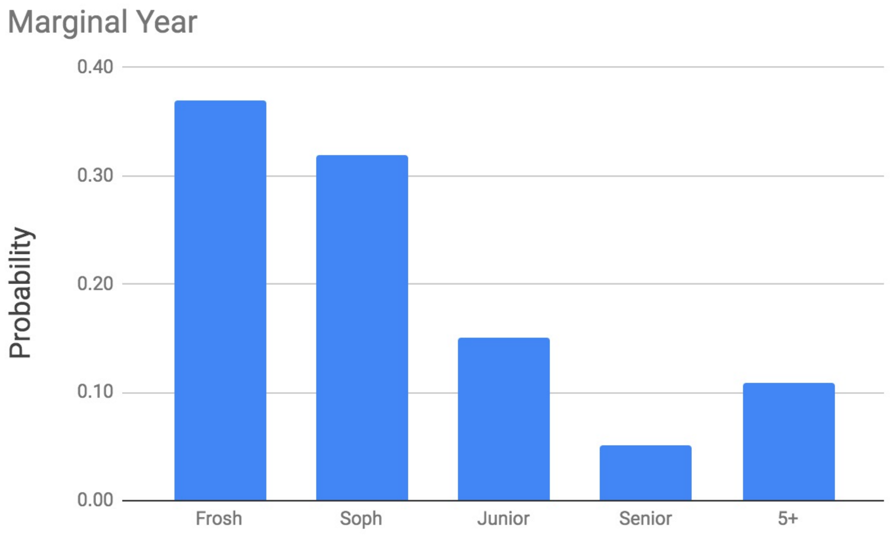
$$= \frac{\frac{1}{\sqrt{4\pi}} e^{-\frac{(10-10)^2}{4}}}{\frac{1}{\sqrt{4\pi}} e^{-\frac{(5-10)^2}{4}}}$$

$$= \frac{e^0}{e^{-\frac{25}{4}}} = 518$$

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





Last Week

Joint Distribution *noun*

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

Eg:

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

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: $p_1, p_2, \ldots, p_m$ where $\sum_{i=1}^{m} p_i = 1$
- X_i = number of trials with outcome *i*

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

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

Probabilistic Text Analysis

According to the Global Language Monitor there are 988,968 words in the english language used on the internet.

Who wrote Federalist Paper 53?

Loop over unique words

Prob hamilton would write word i

Prior belief it was Hamilton

Number of times word i is in the doc

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

Prob Hamilton given Document

Prob of the document???

Who wrote Federalist Paper 53?

Prob that Hamilton wrote it

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

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

Prob that Madison wrote it

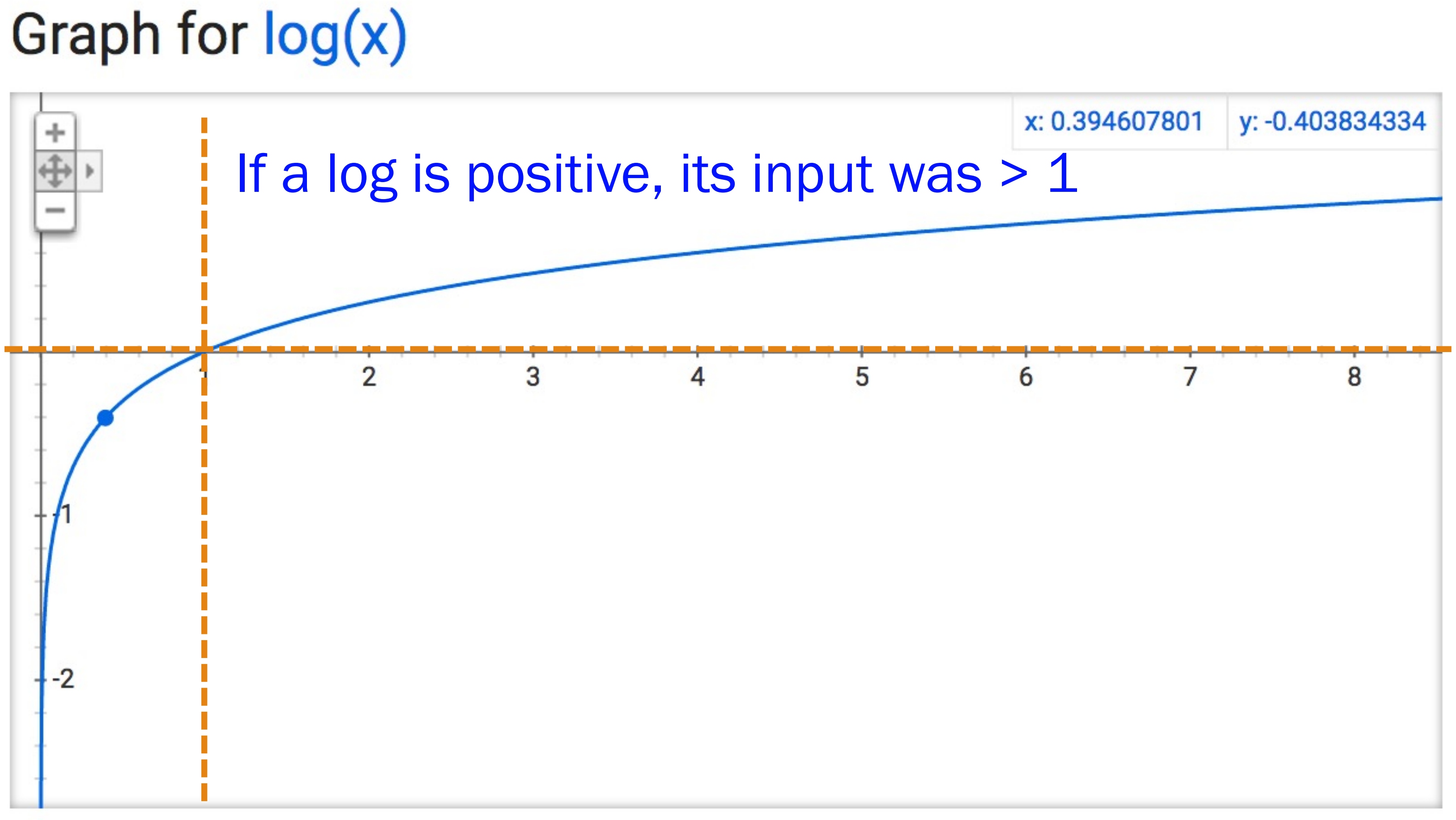
$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}$$

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

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

$$= \frac{\prod_i h_i^{c_i}}{\prod_i m_i^{c_i}}$$

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



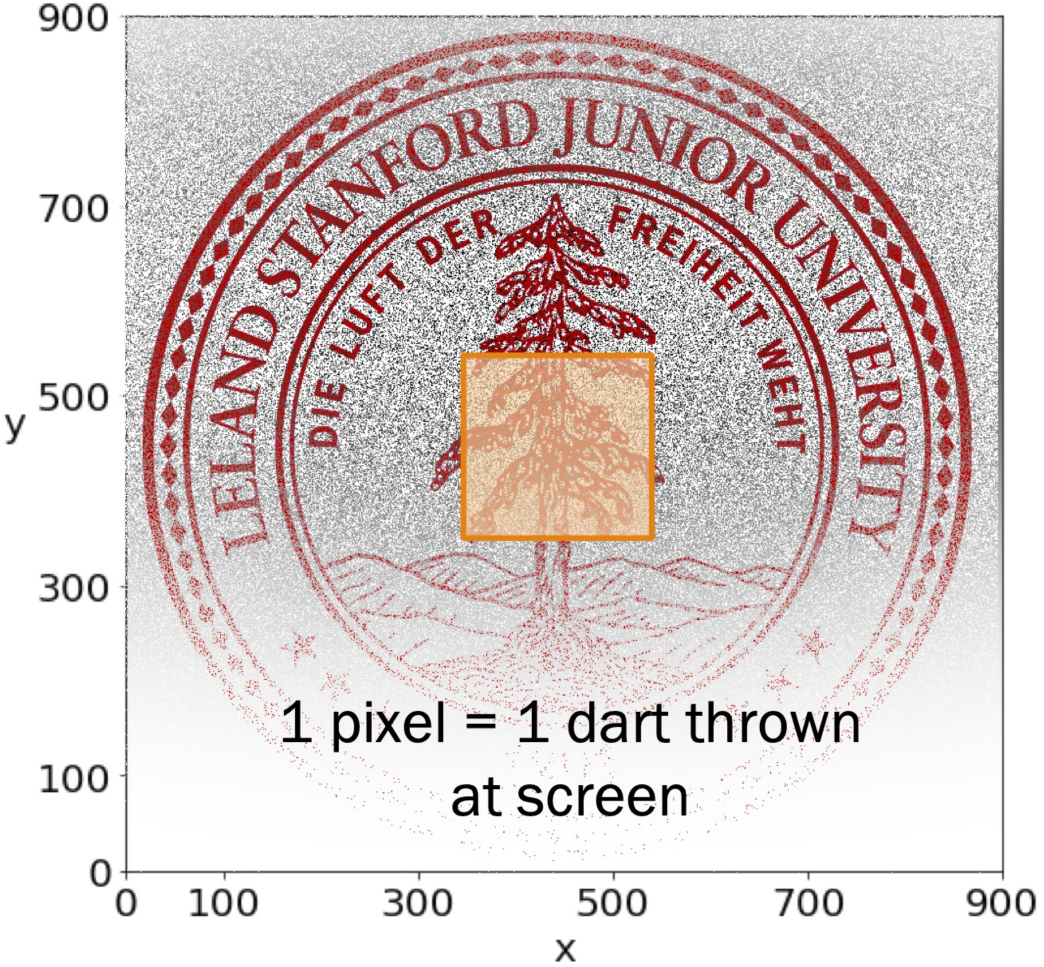
If a log is negative, its input was between 0 and 1

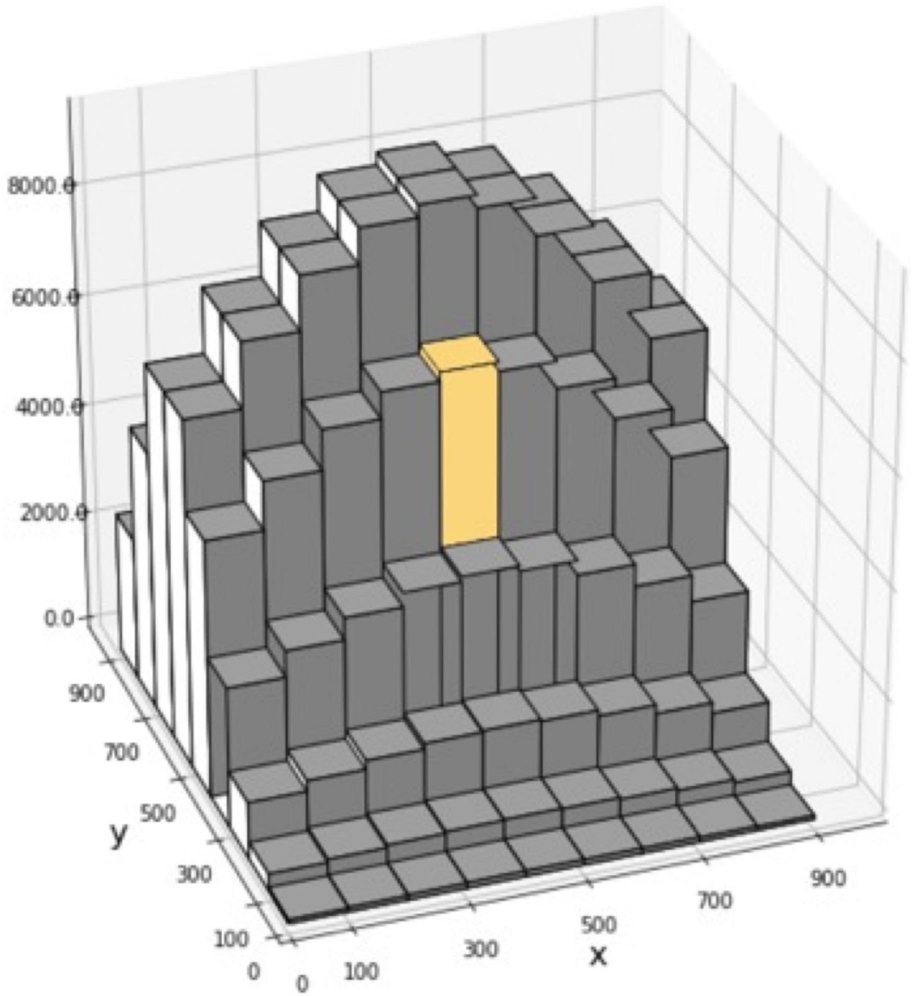
End Review

Joint PMF

Darts?

P(dart hits within r pixels of center)?

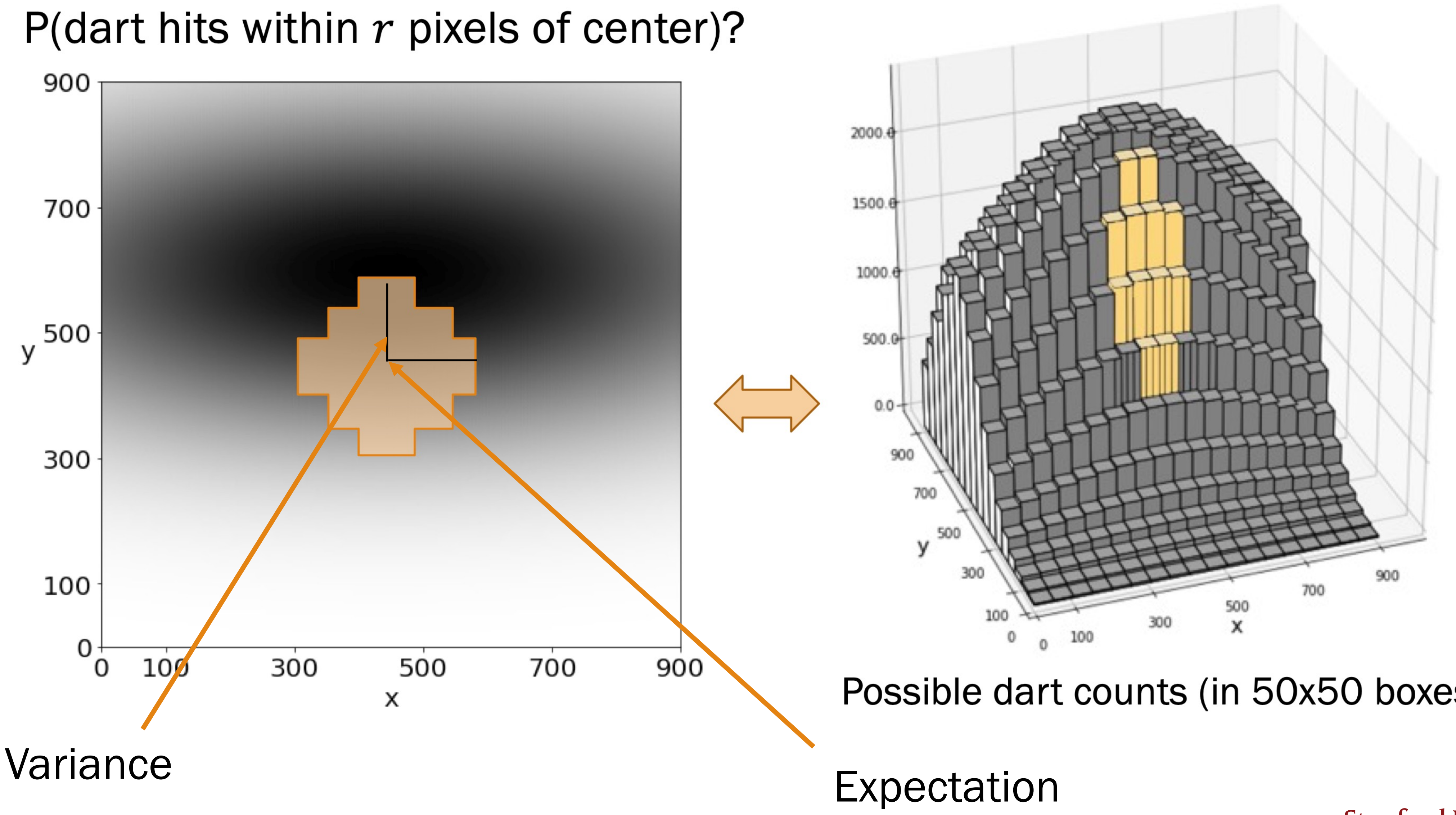




Possible dart counts (in 100x100 boxes)

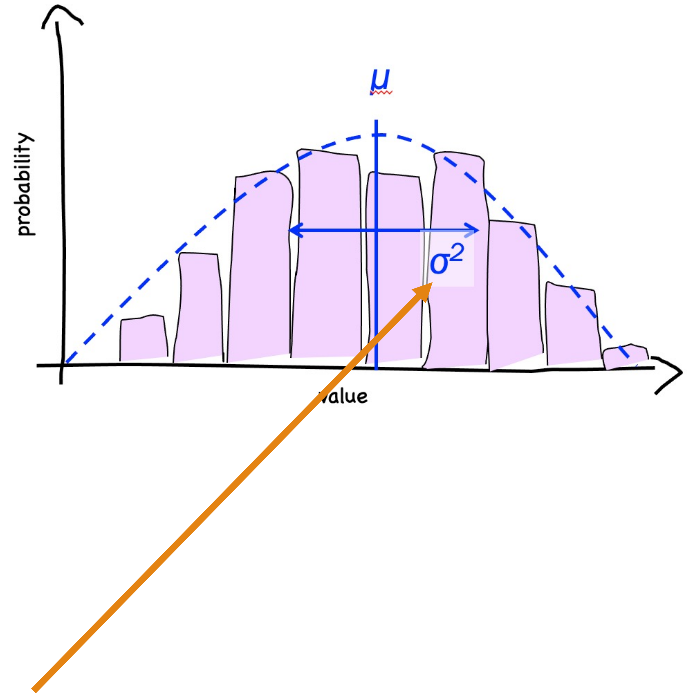
Darts?

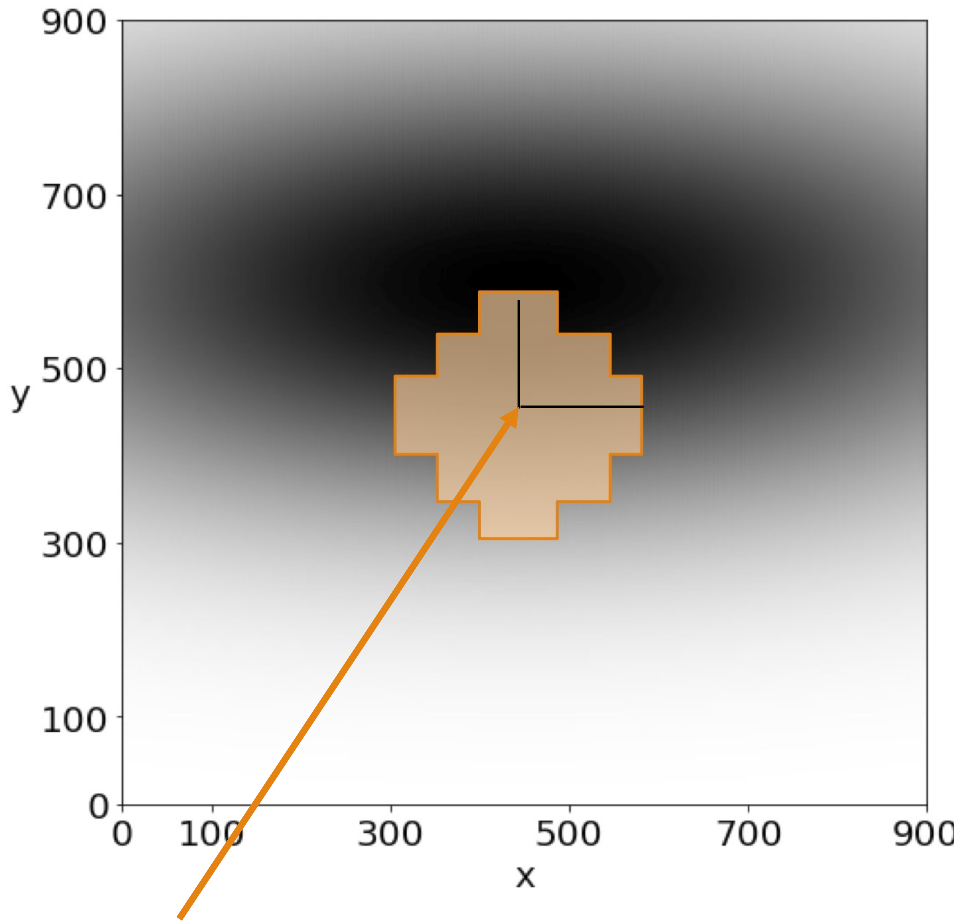
P(dart hits within r pixels of center)?



Possible dart counts (in 50x50 boxes)

Variance





Variance: A width

Variance: Describes an ellipse.

So We Have a Joint PMF

Multinomials are usually described by Multiple Random Variables $X_1, X_2, \dots X_m$. What are some summary things we can say about it?

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

Expectation of Multinomial?

$$E\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_m \end{bmatrix} = \begin{bmatrix} np_1 \\ np_2 \\ np_3 \\ \vdots \\ np_m \end{bmatrix}$$

Multi dim point

Variance of Multinomial? – We have a Covariance Matrix

$$\begin{bmatrix} Var(X_1) & \cdots & Cov(X_1, X_2) \\ \vdots & \ddots & \vdots \\ Cov(X_1, X_2) & \cdots & Var(X_2) \end{bmatrix}$$

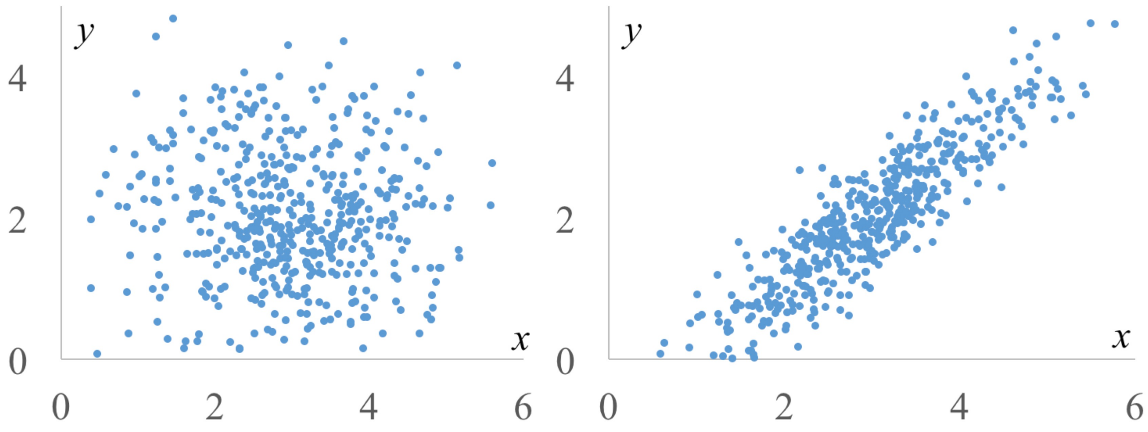
Multi dim ellipse

Another Summary Statistic

We have a special summary statistic that we use when talking about two random variables.

$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$
$$= E[XY] - E[X]E[Y]$$

Generally tells us how two distributions vary with each other.



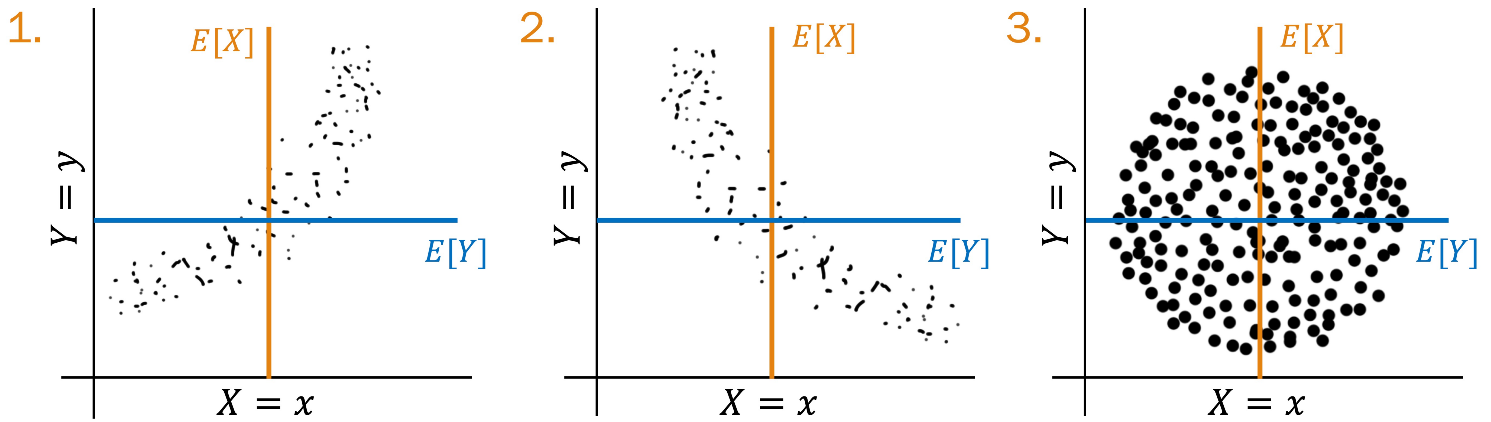
Some Practice

$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$
$$= E[XY] - E[X]E[Y]$$

Covariance measures how one random variable varies with a second.

- Outside temperature and utility bills have a negative covariance.
- Handedness and musical ability have near zero covariance.
- Product demand and price have a positive covariance.

Is the covariance positive, negative, or zero?



positive negative zero

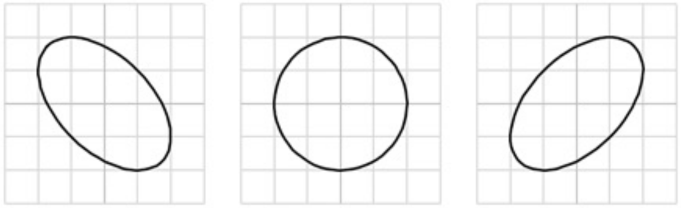
Some Practice

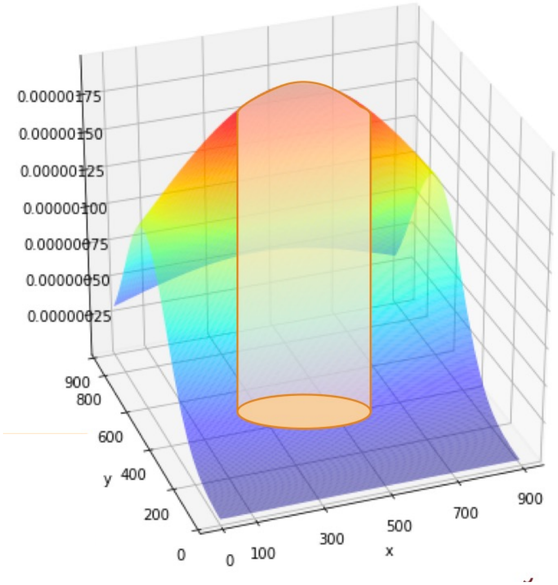
Variance of Multinomial? – We have a Covariance Matrix

$$\begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_1, X_2) & Var(X_2) \end{bmatrix}$$

What can we describe with it?

$$\begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix} \quad \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} \quad \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$$





Describes the ”Spread” of a Joint distribution as a “plate.”

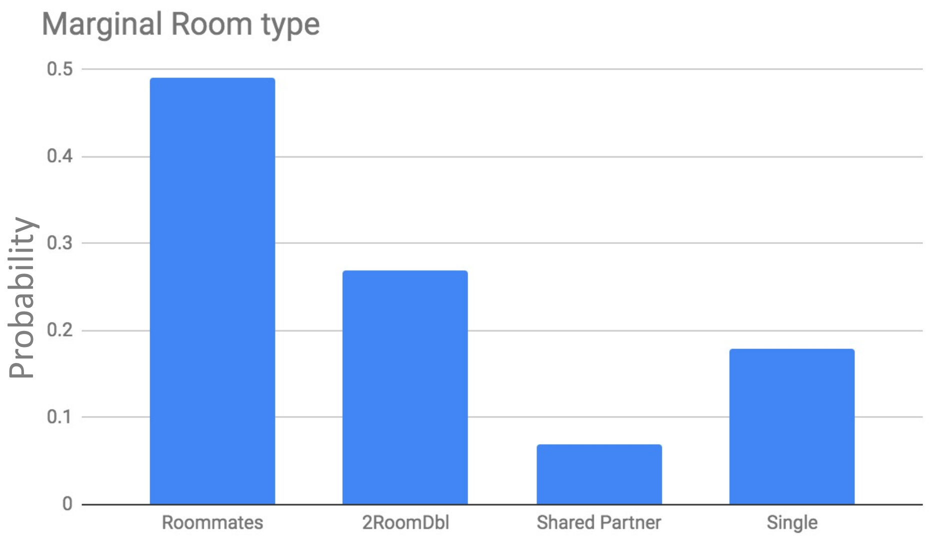
Back to Applications!

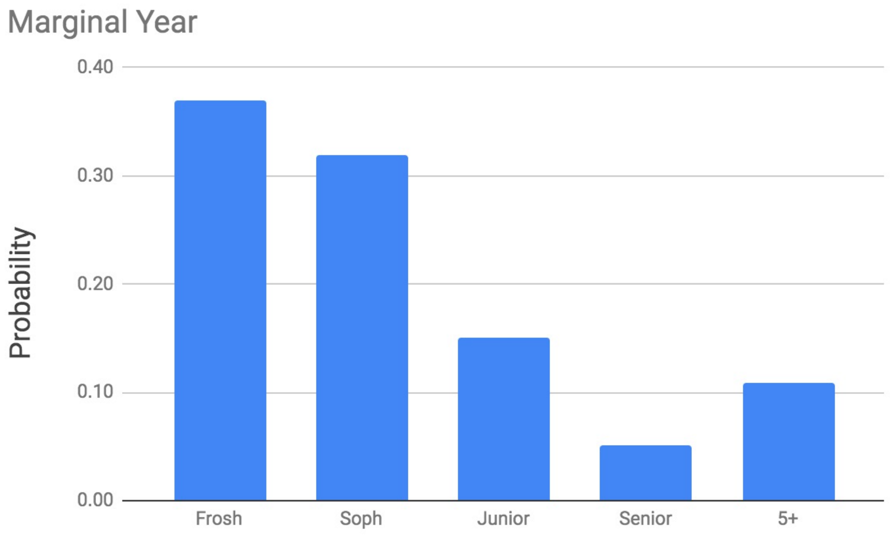
Joint Random Variables

- [ ] 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	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





Today: Introduction to 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.

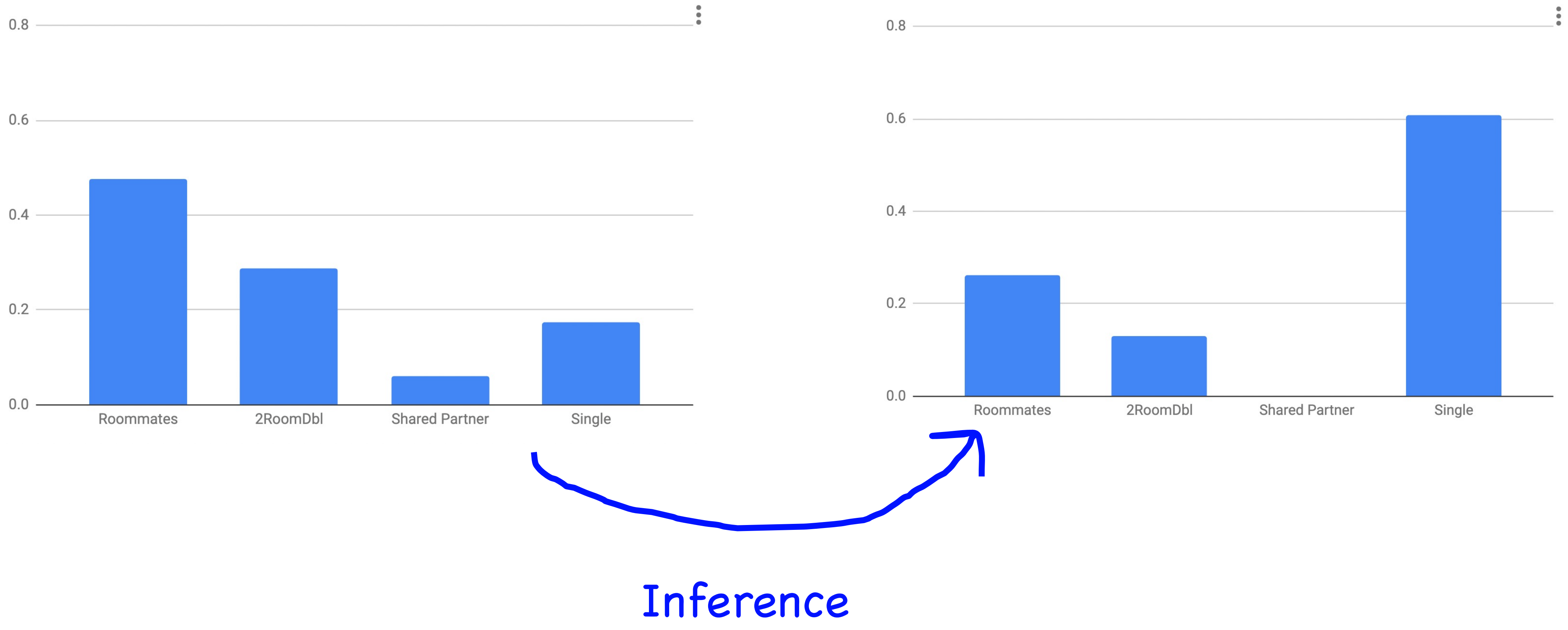
Warmup Inference

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

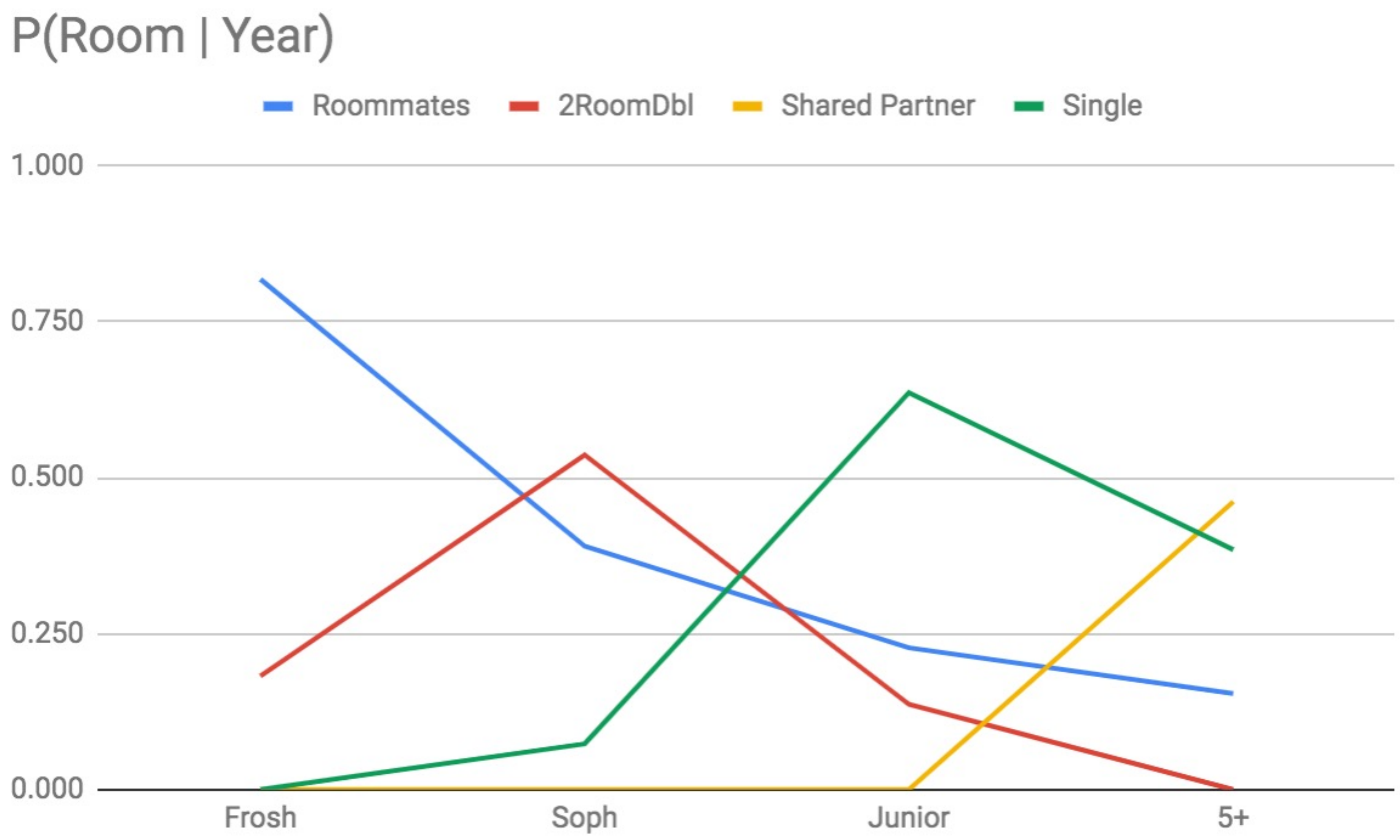
Room | Year

$$P(R = r)$$

$$P(R = r | Y = \text{junior})$$

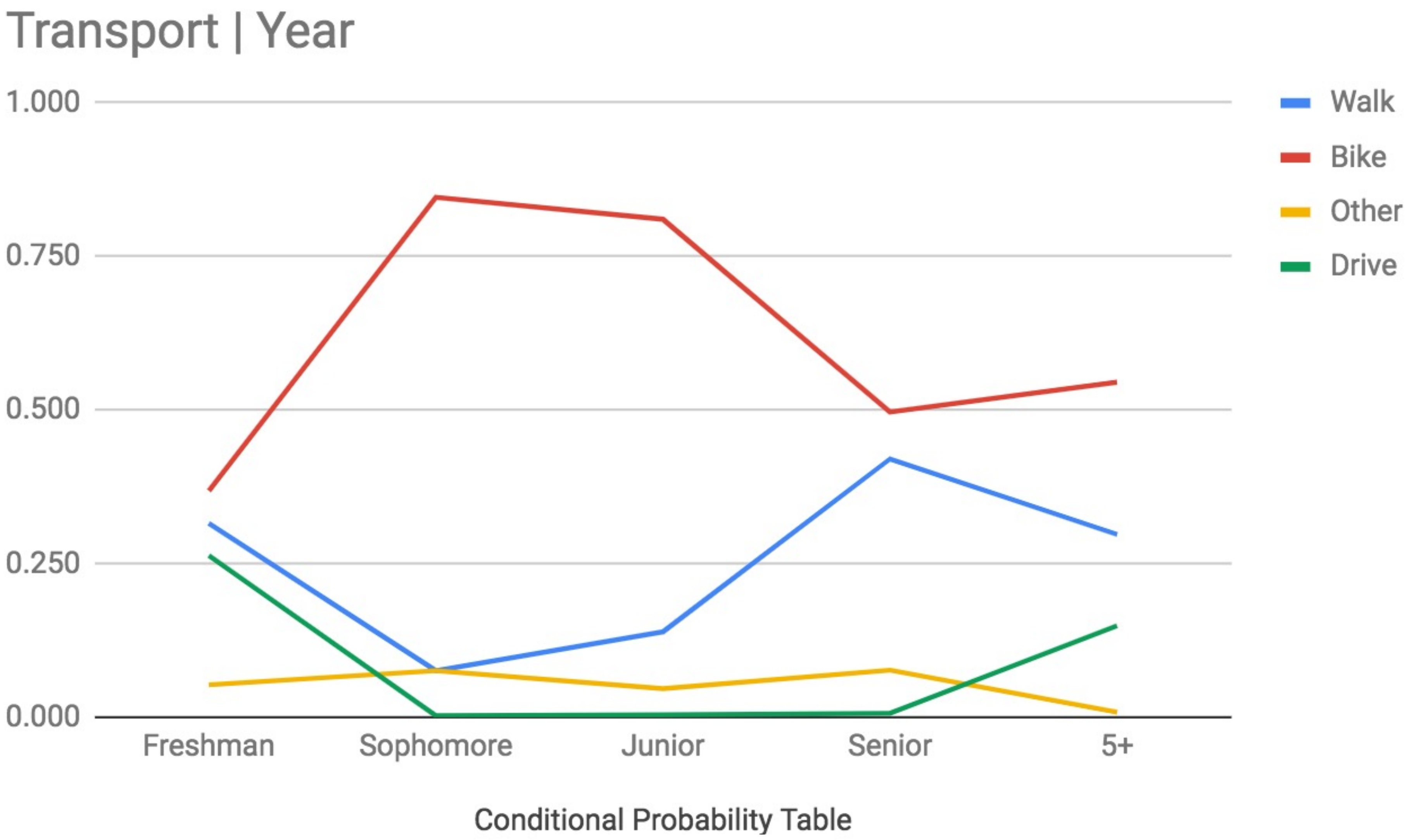


Room | Year

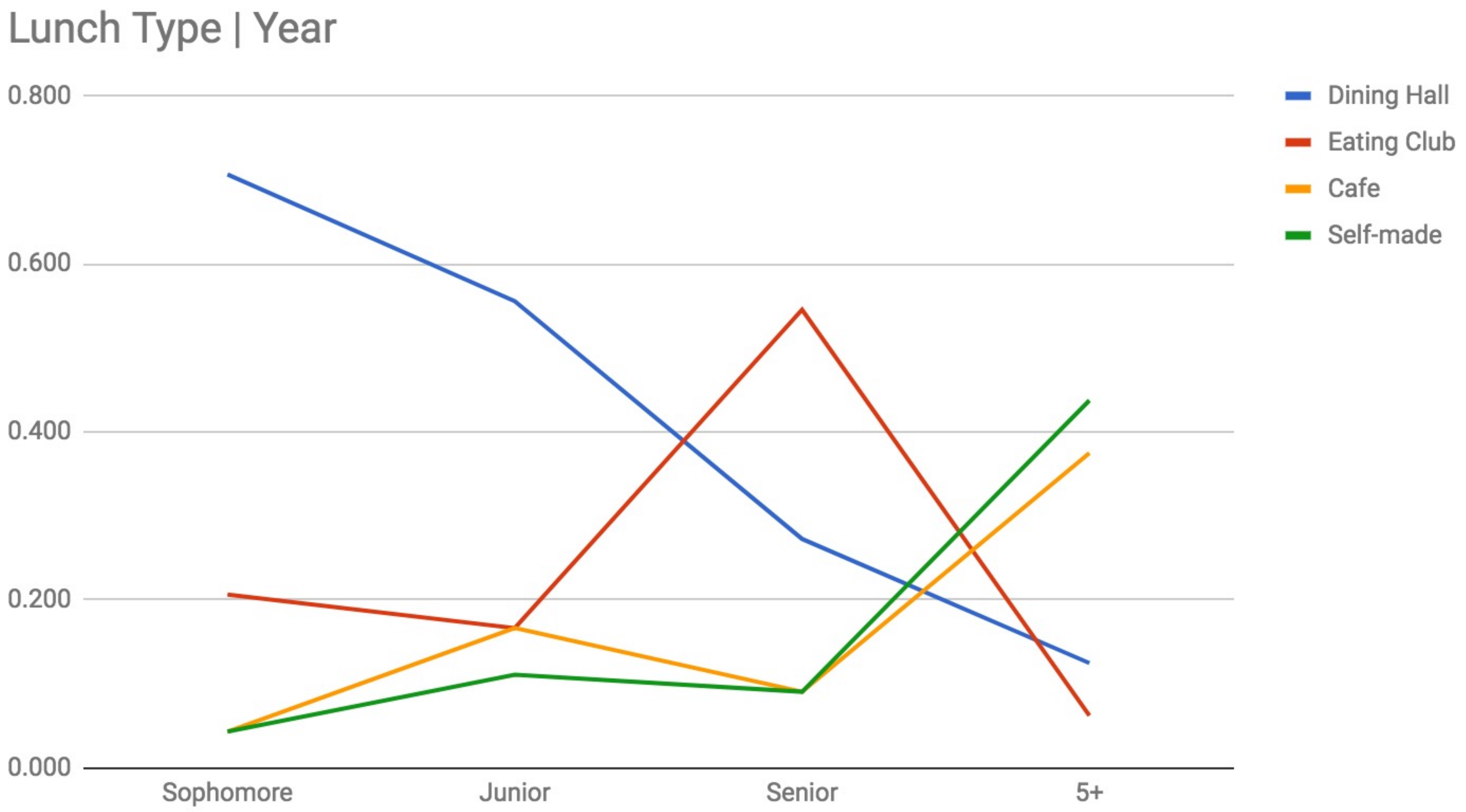


P(Room | Year)
Roommates
2RoomDbl
Shared Partner
Single
1.000
0.750
0.500
0.250
0.000
Frosh
Soph
Junior
5+

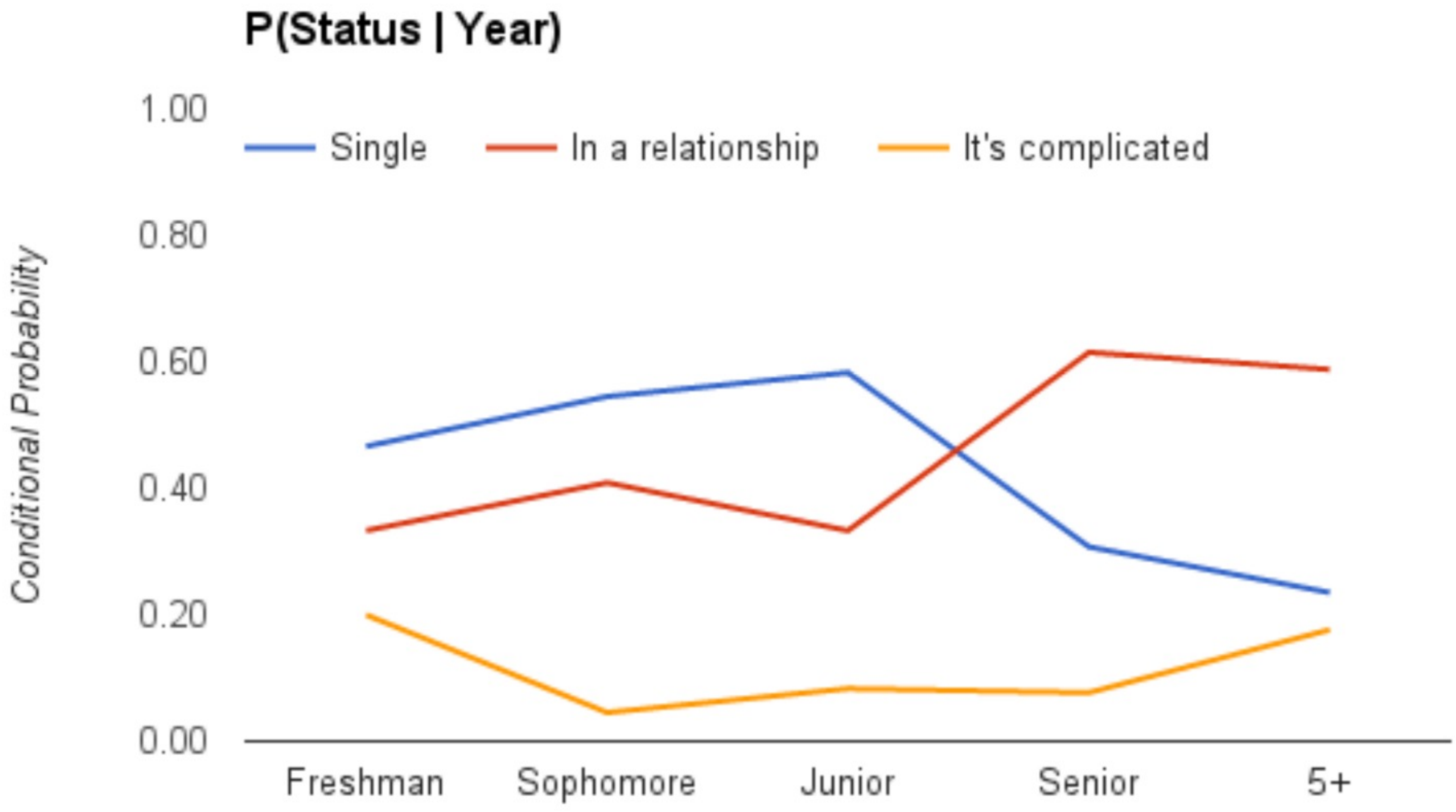
Transport | Year



Lunch | Year


Lunch Type | Year
0.800
0.600
0.400
0.200
0.000
Sophomore
Junior
Senior
5+
Dining Hall
Eating Club
Cafe
Self-made

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)

Warmup: Bayes Revisited

Posterior belief

Likelihood of evidence

Prior belief

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

Normalization constant

Bayes Theorm with Discrete

Let M be a **discrete** random variable

Let N be a **discrete** random variable

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

More generally

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

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 **not** 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

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

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

Question: Have I Been Given the Joint?

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 **not** 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

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

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 **not** 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

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

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

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

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

Time to mix discrete and continuous

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 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 | X = 163)

Let G be an indicator that the elephant is a girl. G is Bern(p = 0.5)

Let X be the distribution of weight of the elephant.

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

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

Aside: Models with continuous RVs

Joint is Complete Information!

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

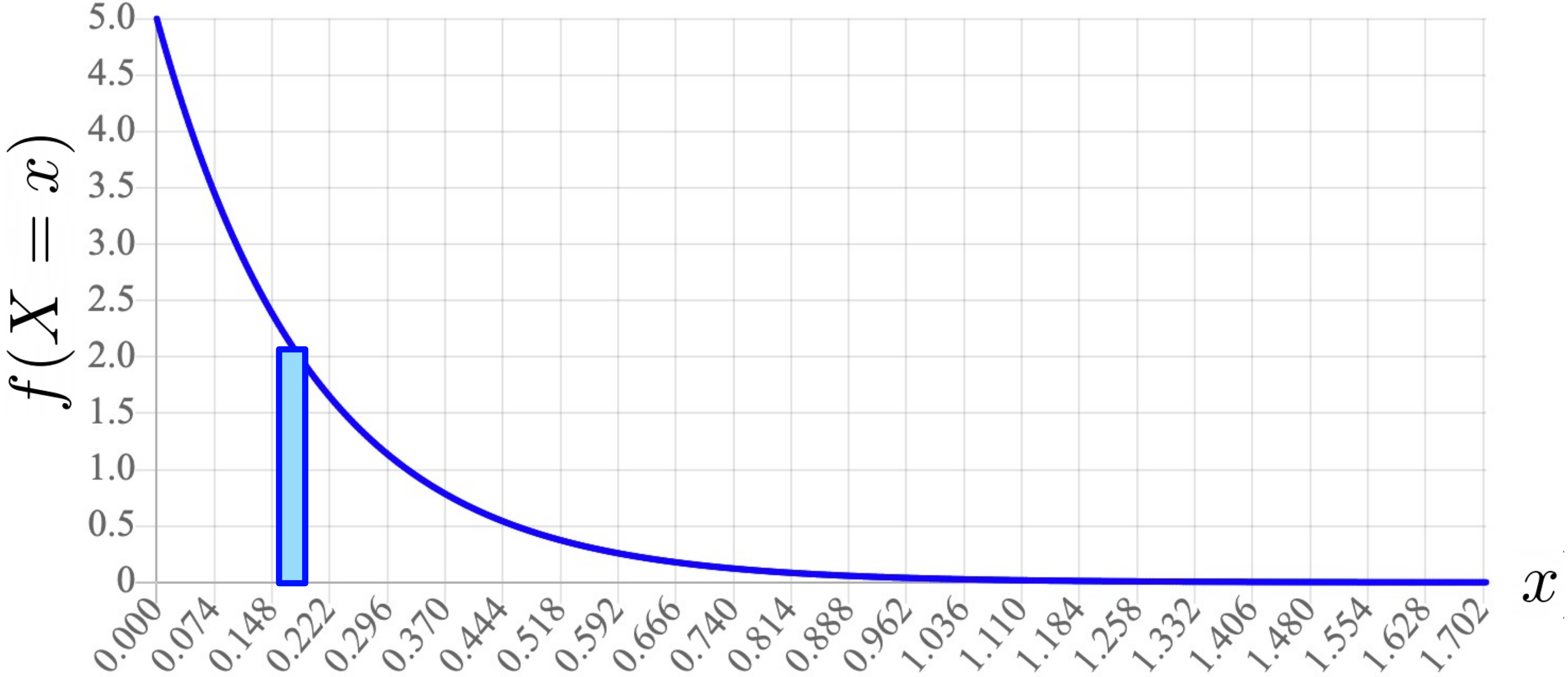
Still true when some variables are continuous

Goal: Inference

Change your belief
distribution
(Joint, PMF, or PDF)
of random variables,
based on
observations

Epsilon: Useful perspective

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



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

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

Change notation

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

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

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

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 | X = 163)

Let G be an indicator that the elephant is a girl. G is 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 | X = 163)

Let G be an indicator that the elephant is a girl. G is 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:

$$f(G = 1, X = 72.3) = f(X = 72.3 | G = 1) P(G = 1)$$

More generally

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

Inference with Continuous

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

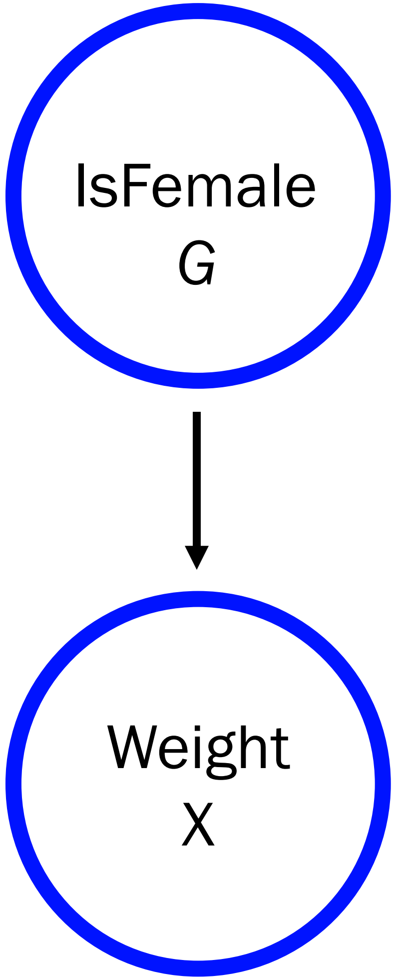
Let G be an indicator that the elephant is a girl. G is Bern(p = 0.5)

Let X be the distribution of weight of the elephant.

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

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

Model Shown Graphically



G = 1 is Bern(p = 0.5)

Does this define the joint?

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

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

X | G = 1 is N(μ = 160, $\sigma^2 = 7^2$)
X | G = 0 is N(μ = 165, $\sigma^2 = 3^2$)

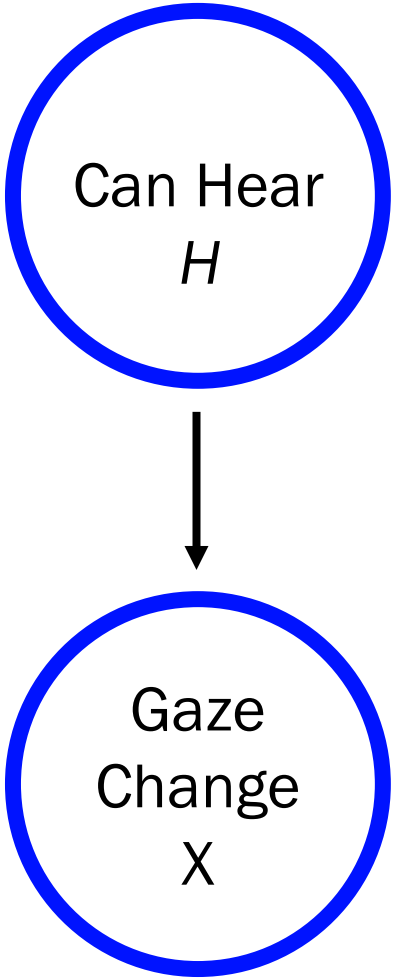
Q: What is P(G = 1 | X = 163)

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 $N(\mu = 8, \sigma^2 = 50)$. Recall $P(H = 1) = ¾$.

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 Bern(p = 0.75)

X | H = 1 is $N(\mu = 15, \sigma^2 = 50)$
X | H = 0 is $N(\mu = 8, \sigma^2 = 50)$

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

Joint Random Variables

- [x] Use a joint table, or joint function to solve probability question
- [x] Think about **conditional** probabilities with joint variables (which might be continuous)
- [ ] Use and find **independence** of joint random variables
- [ ] Use and find **expectation** of joint random variables

Stanford Acuity Test (StAT)