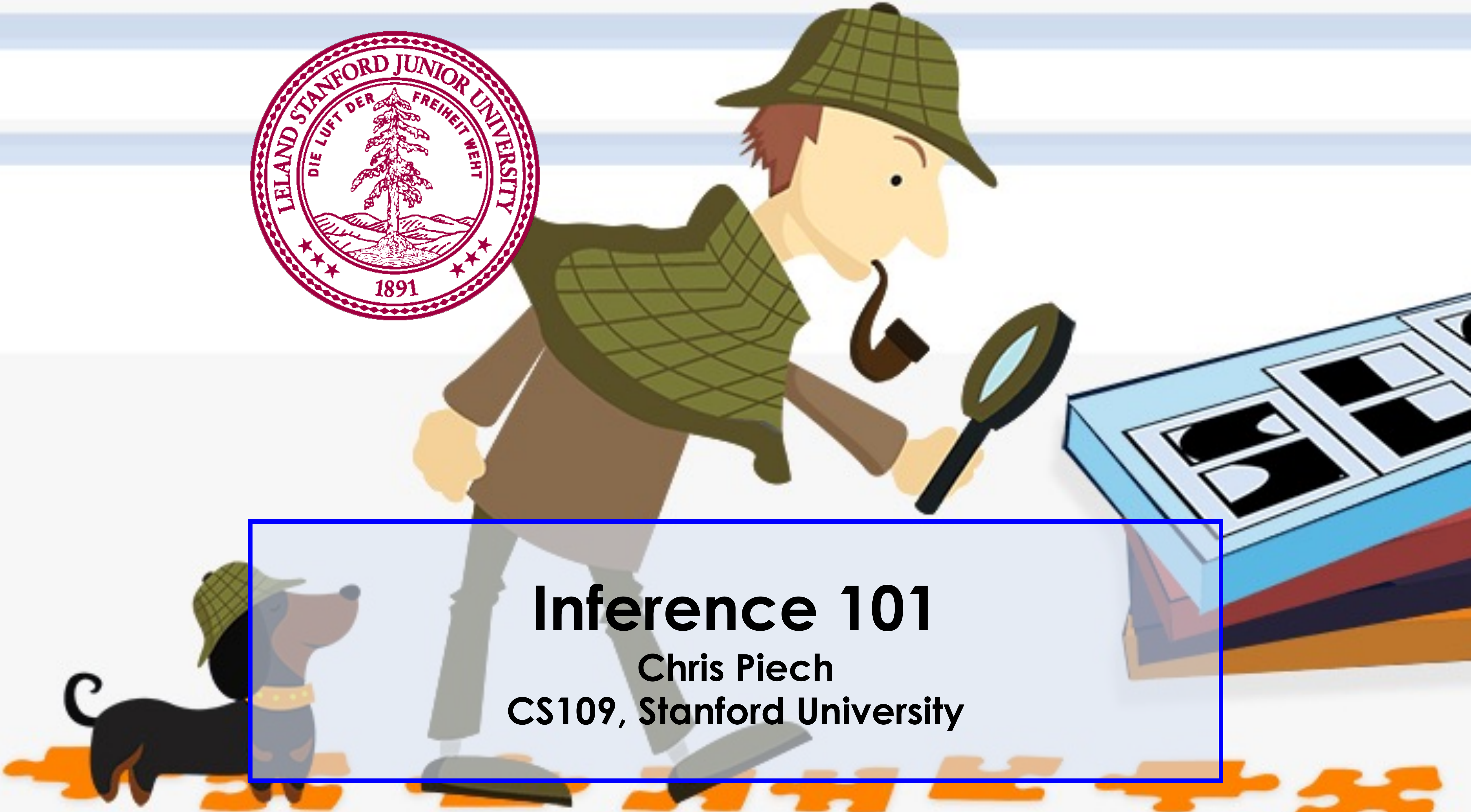


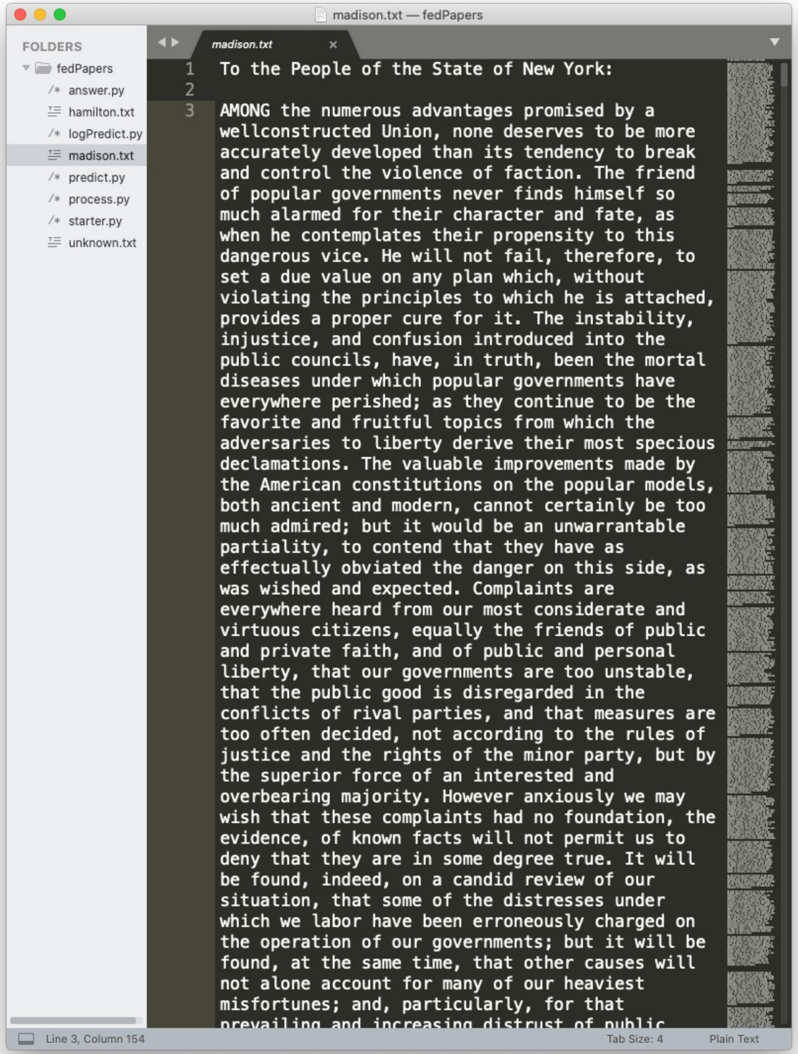
Inference 101

Chris Piech
CS109, Stanford University

Finishing Up

Who wrote Federalist Paper 53?

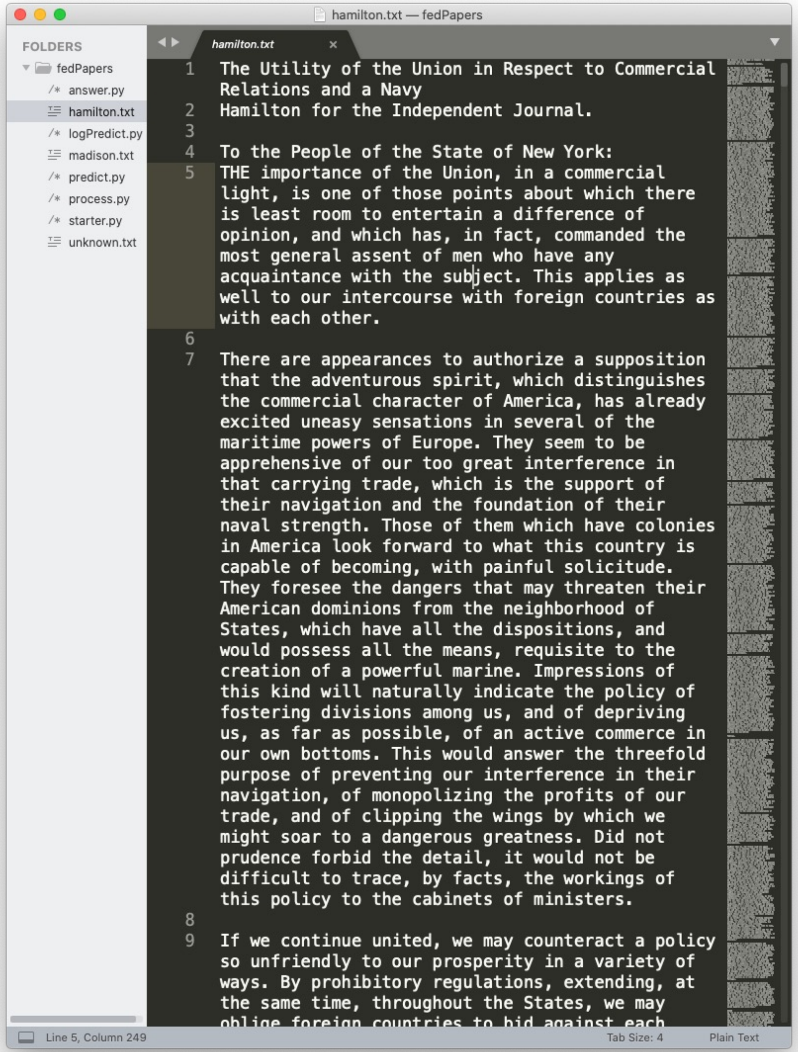
madison.txt



To the People of the State of New York:

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



The Utility of the Union in Respect to Commercial Relations and a Navy

Hamilton for the Independent Journal.

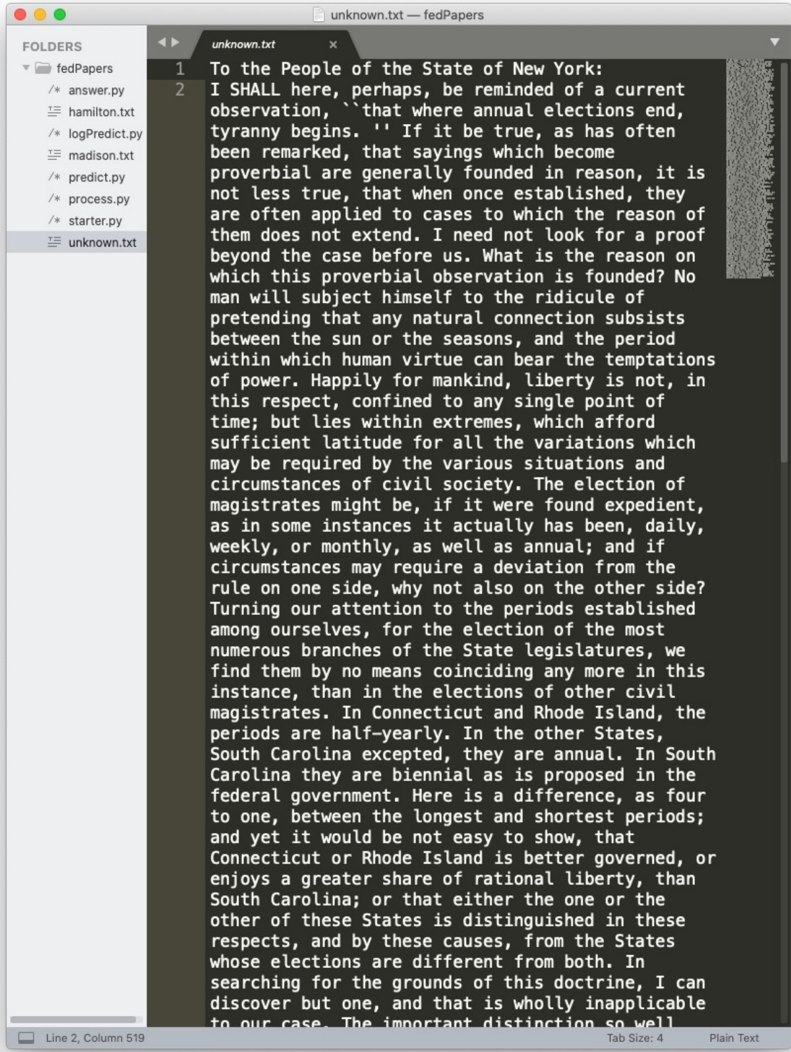
To the People of the State of New York:

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.

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.

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



To the People of the State of New York:

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

What we want to calculate

Prob Hamilton given Document

Probability Hamilton wrote word i

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

How often word i shows up in the doc

Prob Madison given Document

Probability Madison wrote word i

Use logs when probabilities become too small!

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

$$= \sum_i \log h_i^{c_i} - \sum_i \log m_i^{c_i}$$

$$= \sum_i c_i \cdot \log h_i - \sum_i c_i \log m_i$$

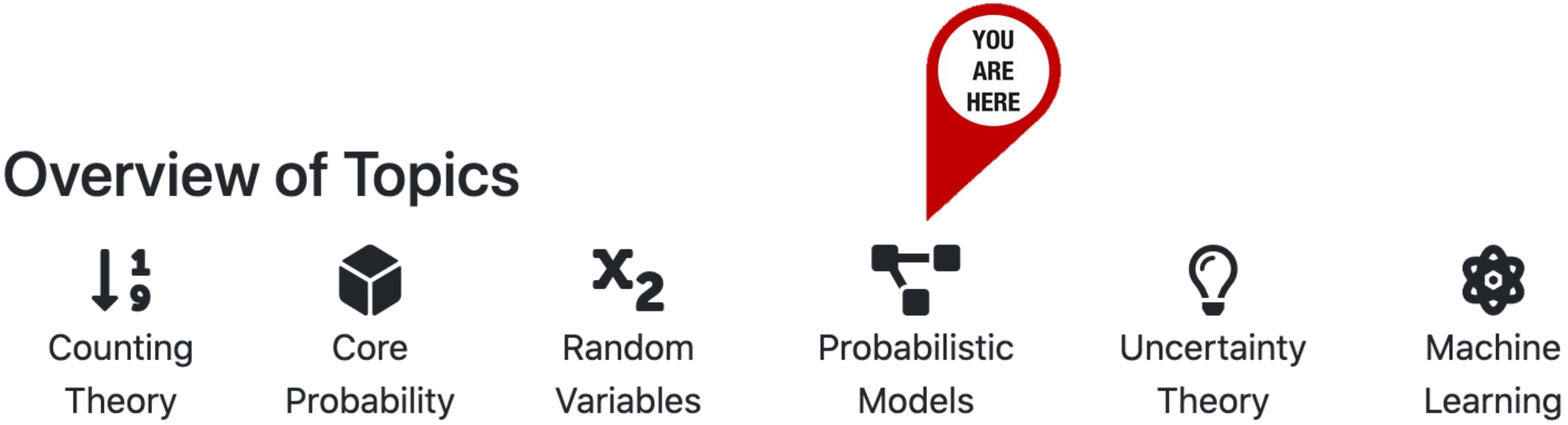
$$= -1344$$

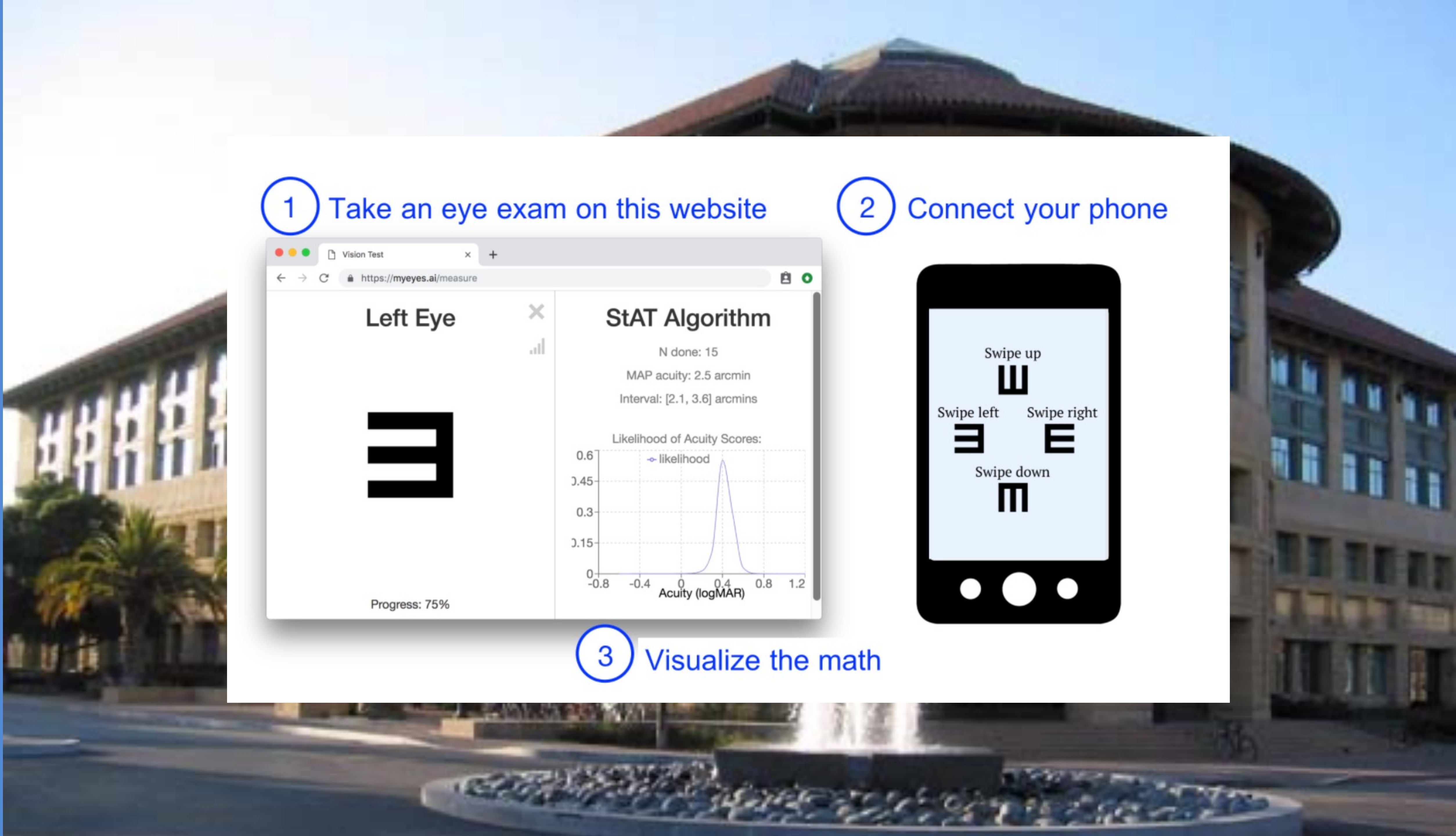
Hamilton Term	-12925
Madison Term	-11581

$$\frac{P(H|D)}{P(M|D)} < 1$$

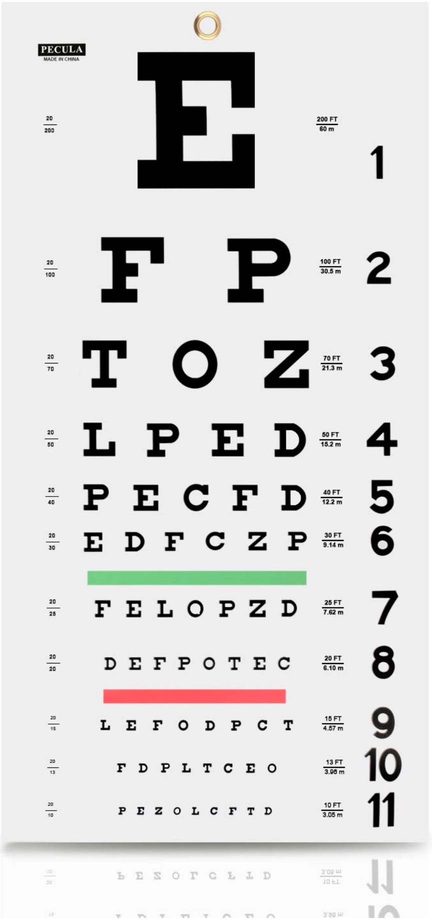
Madison wrote it!

Where are we in CS109?

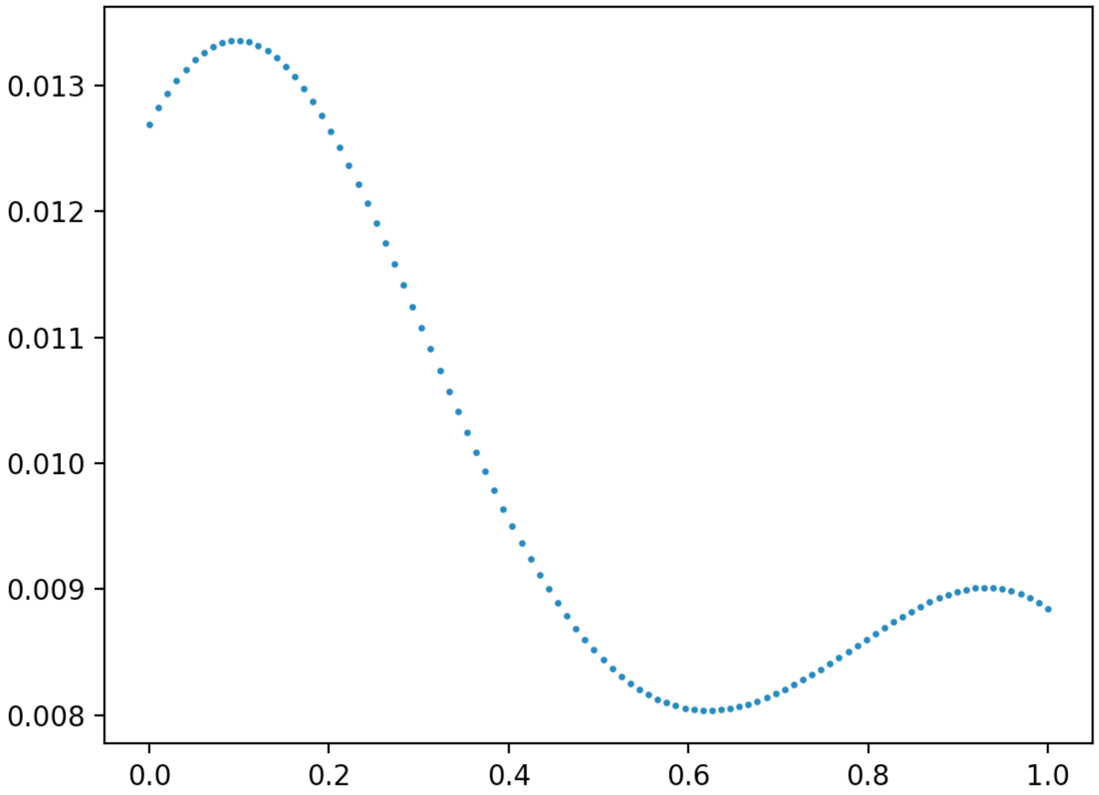



1 Take an eye exam on this website
Vision Test
https://myeyes.ai/measure
Left Eye
Progress: 75%
StAT Algorithm
N done: 15
MAP acuity: 2.5 arcmin
Interval: [2.1, 3.6] arcmins
Likelihood of Acuity Scores:
likelihood
Acuity (logMAR)
2 Connect your phone
Swipe up
Swipe left
Swipe right
Swipe down
3 Visualize the math

Belief in **Vision** Given **User Responses**

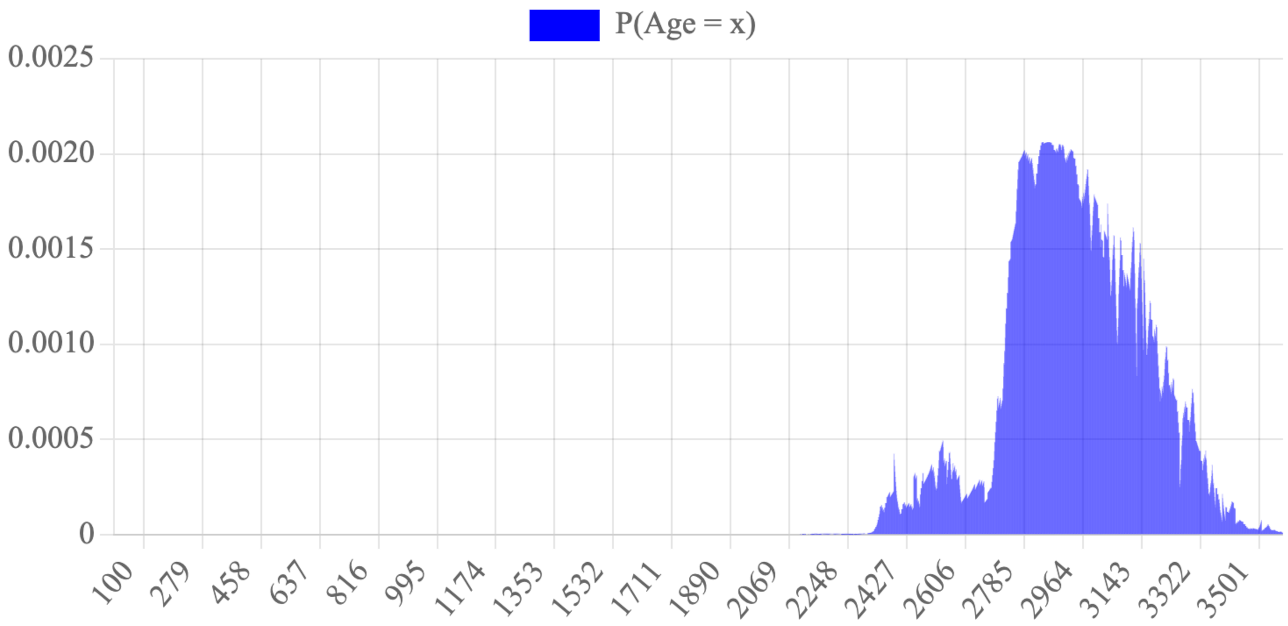


P(Ability to See | Observed Responses)



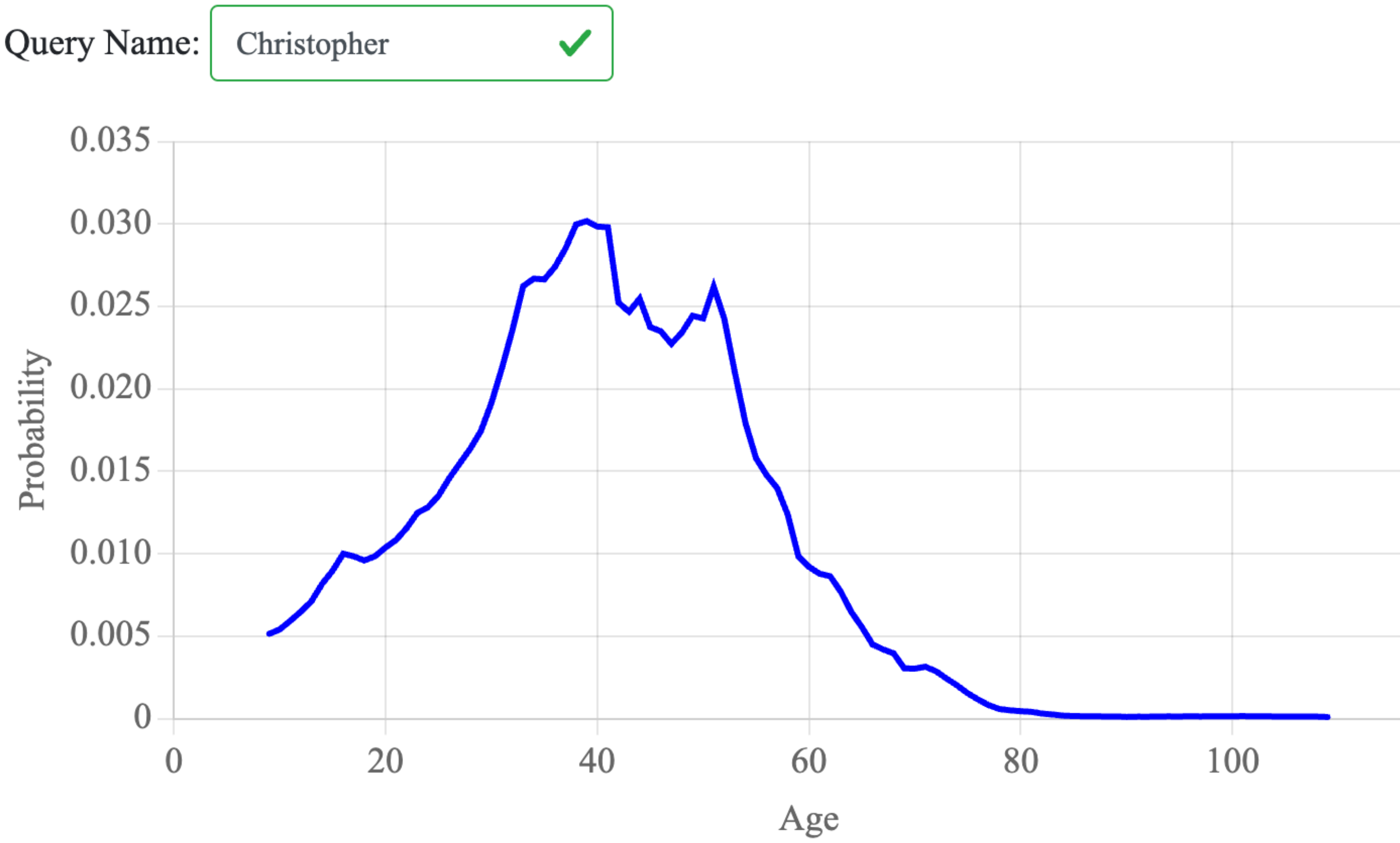
Belief in **Age** Given **Observed C14**

P(Age | Observed C14)



Belief in **Age** Given **Name**

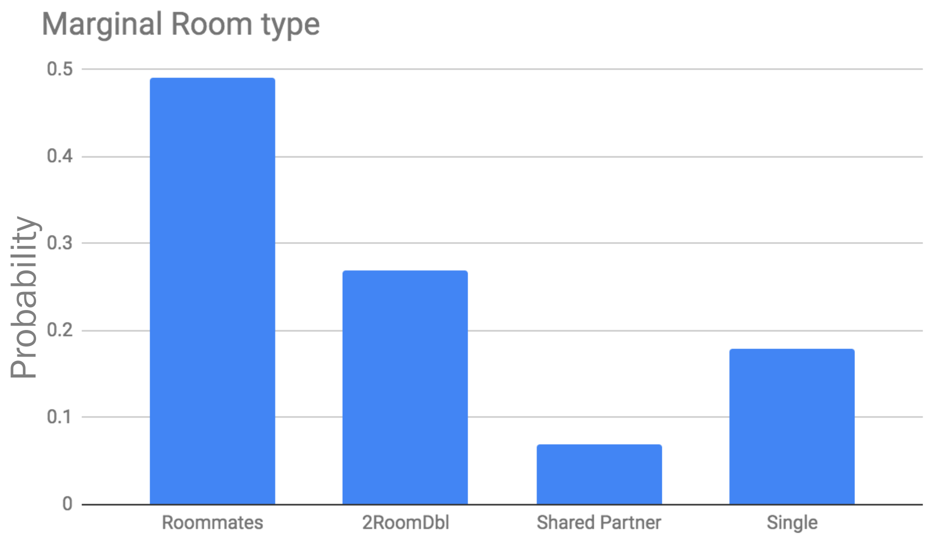
P(Age | Name)

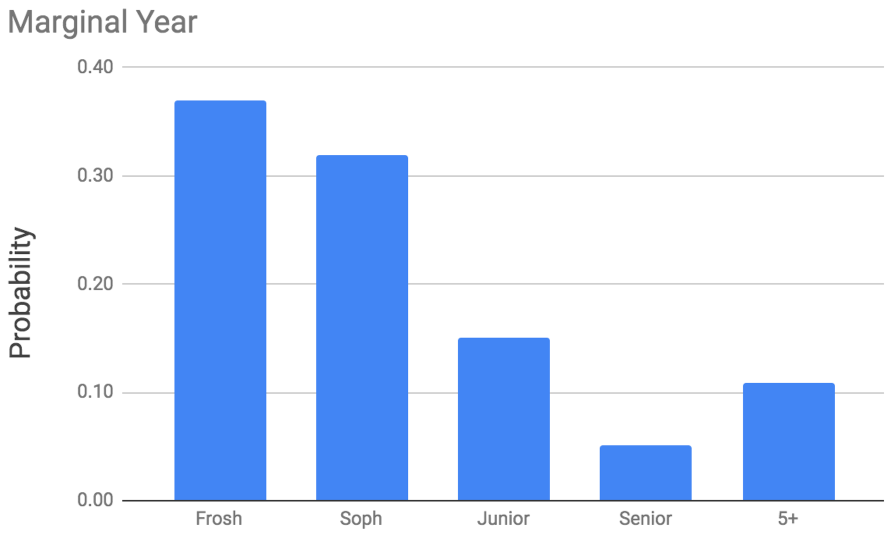


Review

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

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

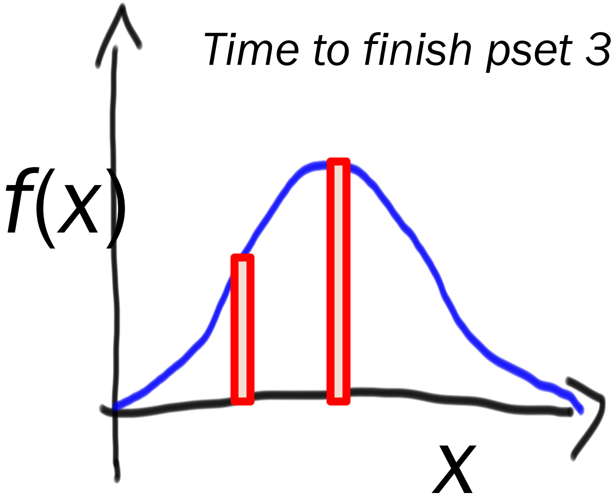
$$P_{X,Y}(x, y)$$

$$P(x, y)$$

Relative Probability of Continuous Variables

X = time to finish pset 3

X ~ N(μ = 10, σ^2 = 2)



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

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

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

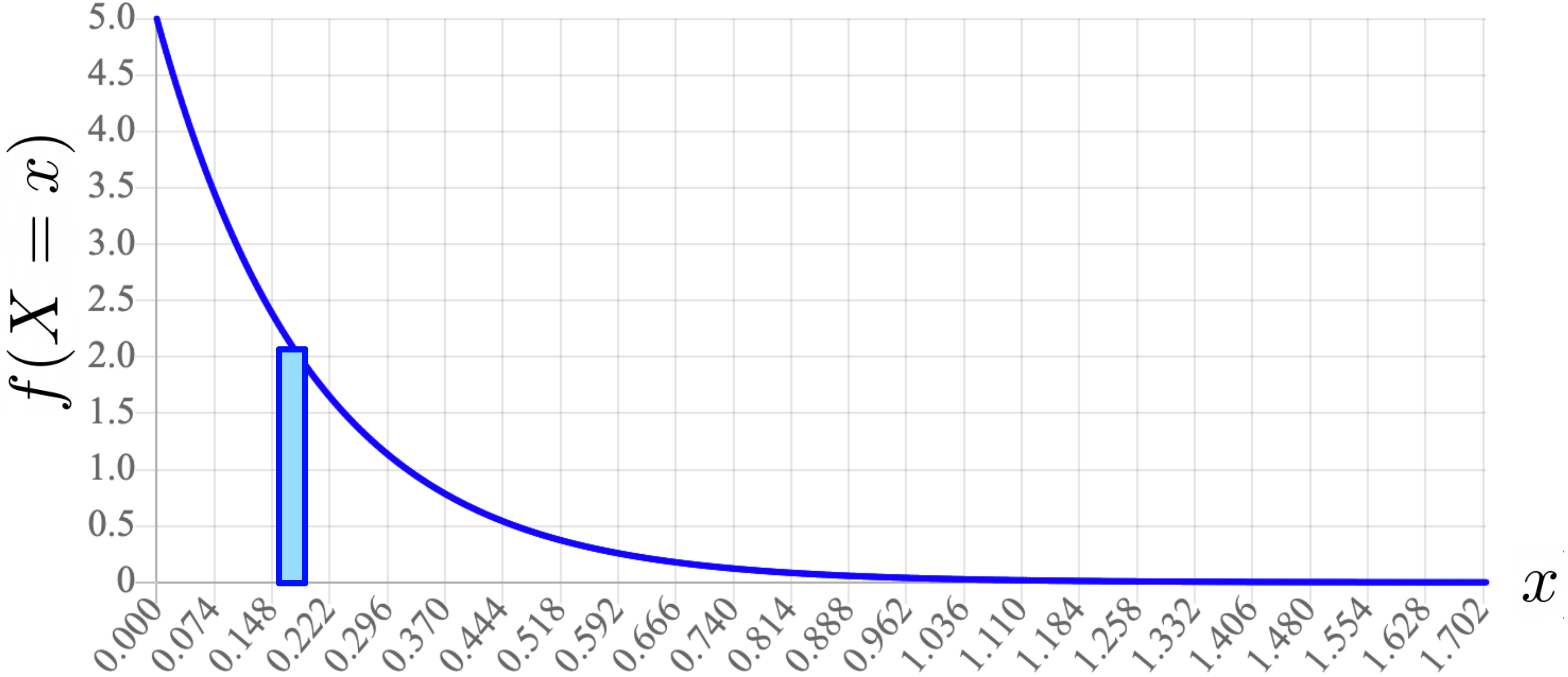
$$= \frac{\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(10-\mu)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(5-\mu)^2}{2\sigma^2}}}$$

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

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

Epsilon: Useful perspective

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



Normal Probability Density Function

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

"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

Get Ready…

Learning Goals

1. Update a Random Variable Belief given evidence
2. Apply Bayes Theorem with Discrete and Continuous variables

Today: Inference

Inference *noun*

Updating one's 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.

Bayes Theorem

Posterior belief

Likelihood of evidence

Prior belief

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

Normalization constant

Bayes with Discrete Random Variables

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

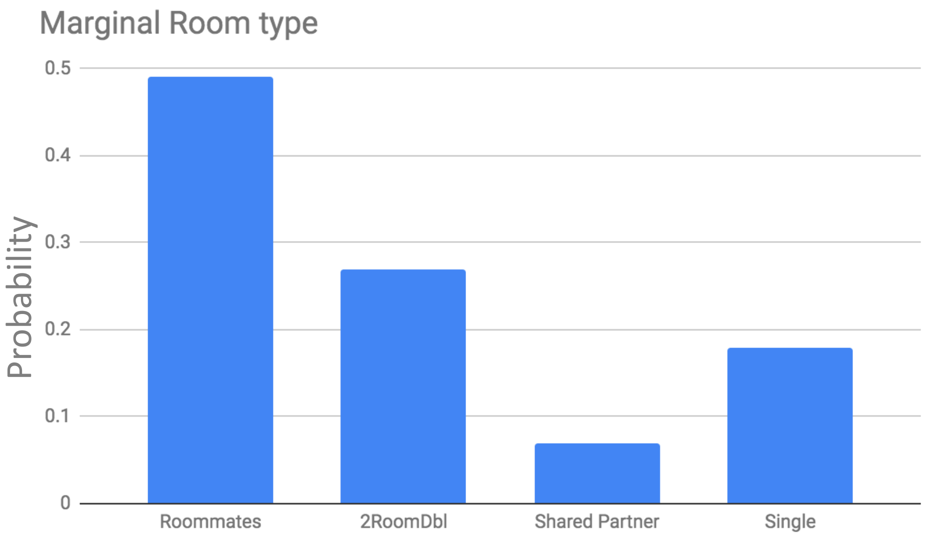
$$P(m|n) = \frac{P(n|m)P(m)}{P(n)}$$

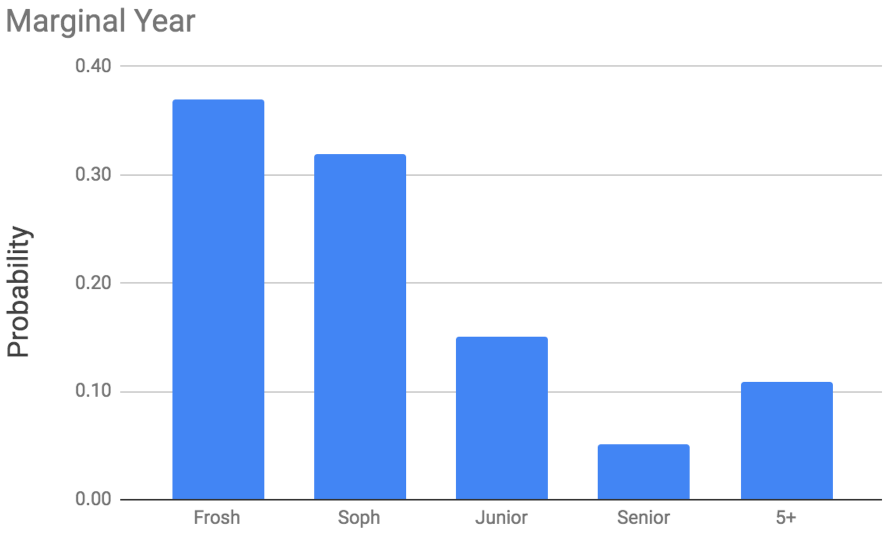
What is the probability distribution of rooms | student is a senior?

Joint ↳

	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

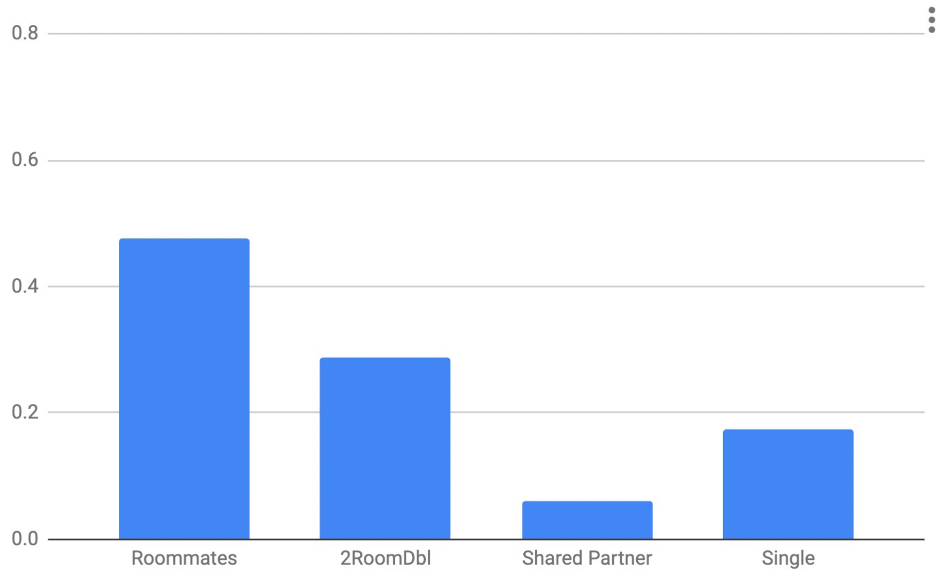
Marginals ↳



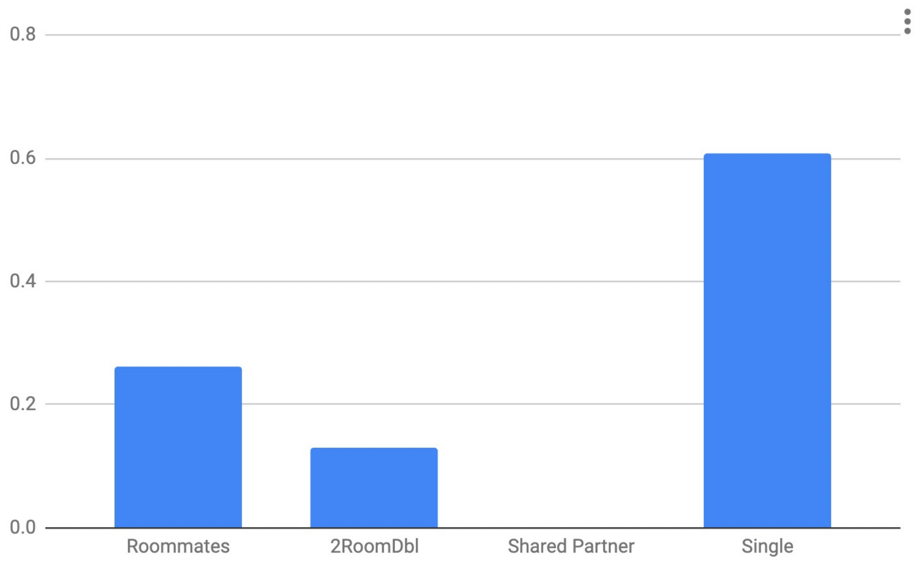


What is the probability distribution of rooms | student is a senior?

$$P(R = r)$$



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



Inference

Inference is hard for two reasons:

1. Mix continuous and Discrete
2. Result can be a PMF or PDF

Inference is hard for two reasons:

1. Mix continuous and Discrete
2. Result can be a PMF

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

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

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

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?

Y = 1 means the child can hear the sound

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

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

I Heard That with Continuous

Normal Assumption:

For babies who can hear sounds, we approximate their gaze movement after the sound is played as: $N(\mu = 15, \sigma^2 = 25)$.

For babies who can **not** hear sounds, we approximate gaze movement as $N(\mu = 8, \sigma^2 = 25)$.

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?

Equivalently

Normal Assumption: For babies who can hear sounds, we approximate their gaze movement after the sound is played as: $N(\mu = 15, \sigma^2 = 25)$.

For babies who can **not** hear sounds, we approximate gaze movement as $N(\mu = 8, \sigma^2 = 25)$.

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?

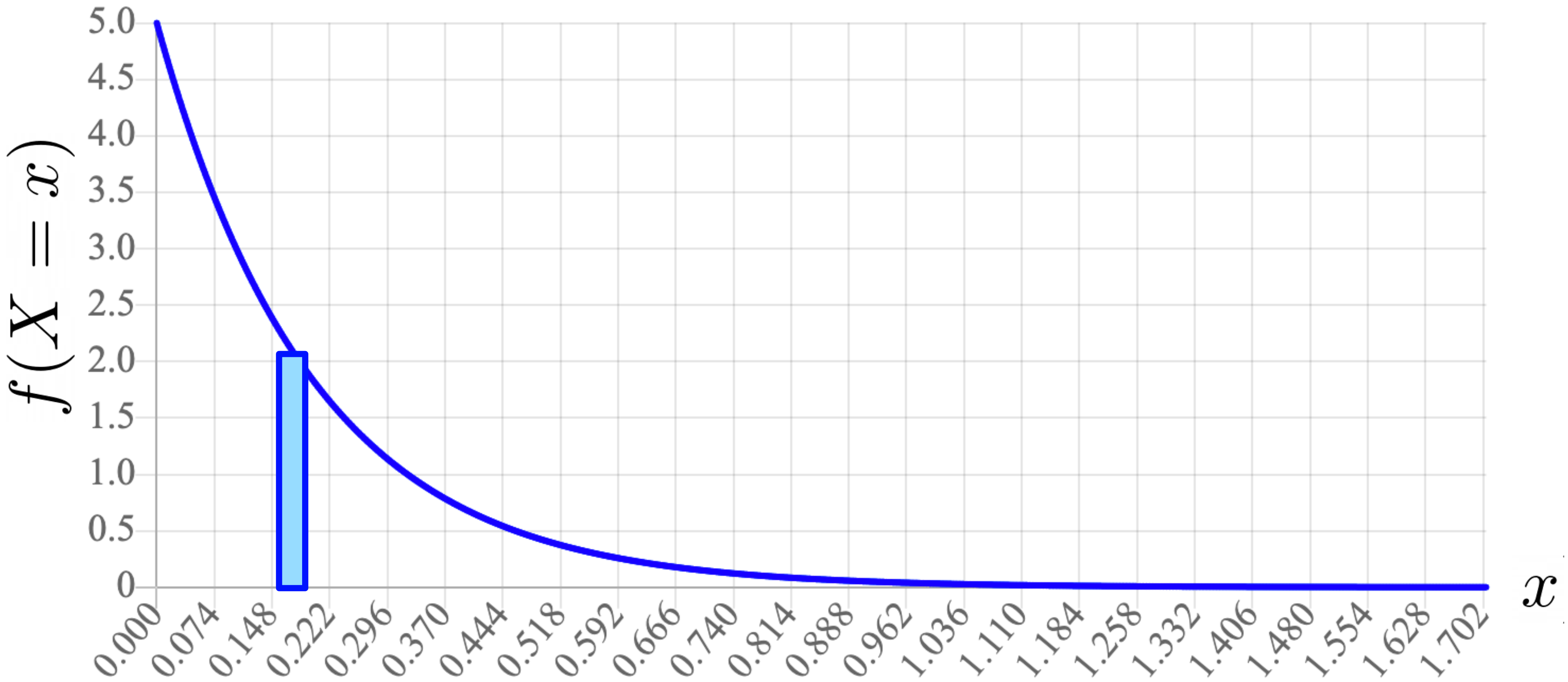
```
def sample ():
   # bernoulli sample
   can_hear = rand_bern(0.75)
   if can_hear == 0:
       # gaussian sample
       return rand_gauss (mu = 15 , std = 5)
   else:
       # gaussian sample
       return rand_gauss (mu = 8, std = 5)
```

The function sample returned the value 14. What is the probability that can_hear was 1?

Aside: Models with continuous RVs

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

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

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

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

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

Chain Rule

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

Law of total probability

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

End Aside

How can We Solve This Problem?

Normal Assumption: For babies who can hear sounds, we approximate their gaze movement after the sound is played as: $N(\mu = 15, \sigma^2 = 25)$.

For babies who can **not** hear sounds, we approximate gaze movement as $N(\mu = 8, \sigma^2 = 25)$.

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?

```
def sample ():
    # bernoulli sample
    can_hear = rand_bern(0.75)
    if can_hear == 0:
        # gaussian sample
        return rand_gauss (mu = 15 , std = 5)
    else:
        # gaussian sample
        return rand_gauss (mu = 8, std = 5)
```

The function sample returned the value 14. What is the probability that can_hear was 1?

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 $P(G = 1 \mid X = 163)$

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 $N(\mu = 165, \sigma^2 = 3^2)$

Pedagogical Pause

Inference is hard for two reasons:

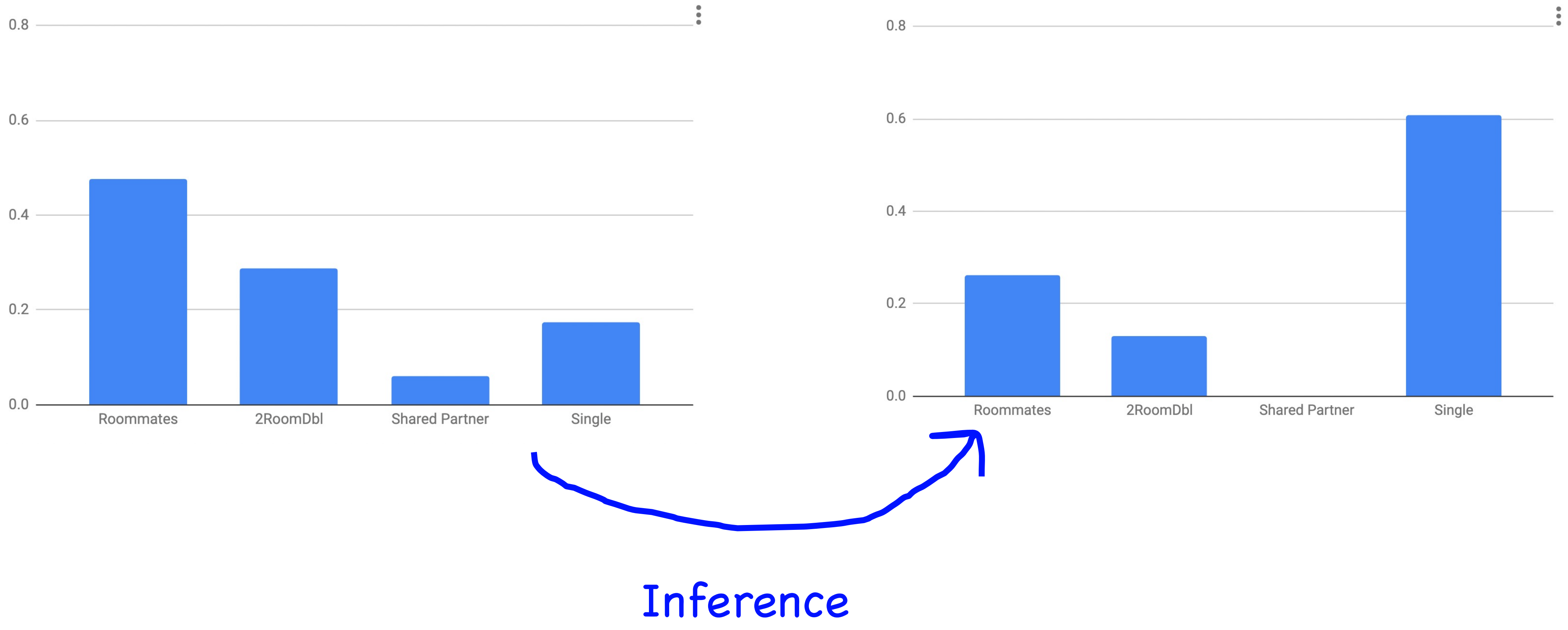
1. Mix continuous and Discrete
2. Result can be a PMF

Inference is hard for two reasons:

1. Mix continuous and Discrete
2. Result can be a PMF

What is the probability distribution of rooms | student is a senior?

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



Goal: Inference

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

*Note in the earlier examples, we were updating Bernoulli Random Variables

Lets Play Number of Function!

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)

Popularized in 2009:

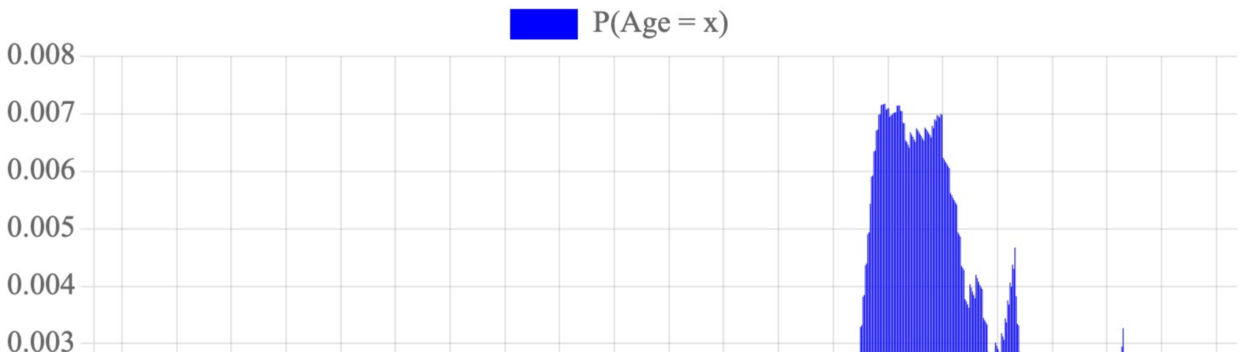
Bayesian Carbon Dating

We are able to know the age of ancient artefacts using a process called carbon dating. This process involves a lot of uncertainty! You observe a measurement of 90% of natural C14 molecules in a sample. What is your belief distribution over the age of the sample? This task requires probabilistic models because we have to think about two random variables together: the age A of the sample, and M the remaining C14 molecules.

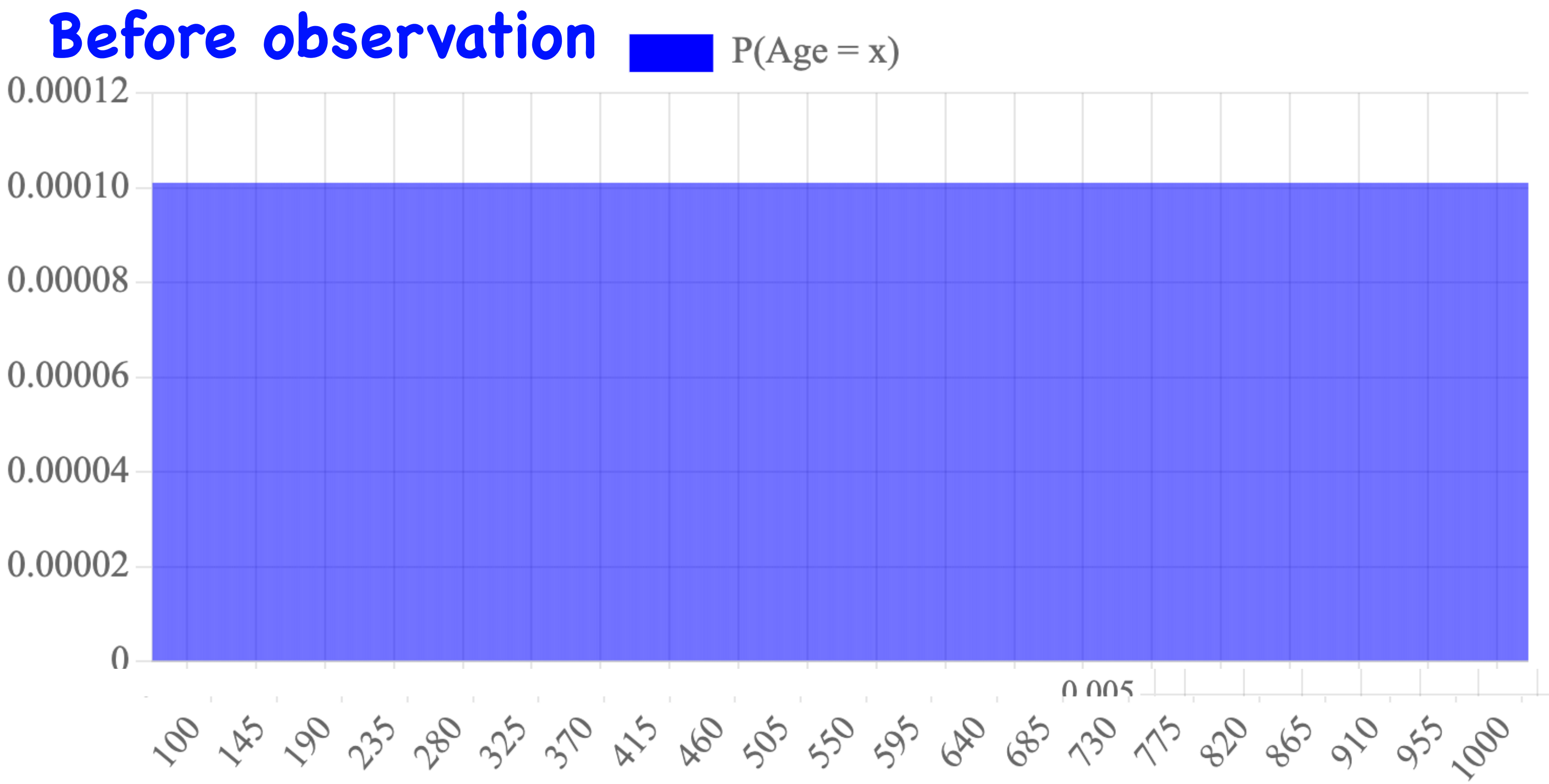
Carbon Dating Demo

Imagine you have just taken a sample from your artifact. For the sample size you took, a living organism would have had 1000 molecules of C14. Use this demo to explore the relationship between how much C14 is left and your belief ditribution for how old your artifact is.

Remaining C14: 900

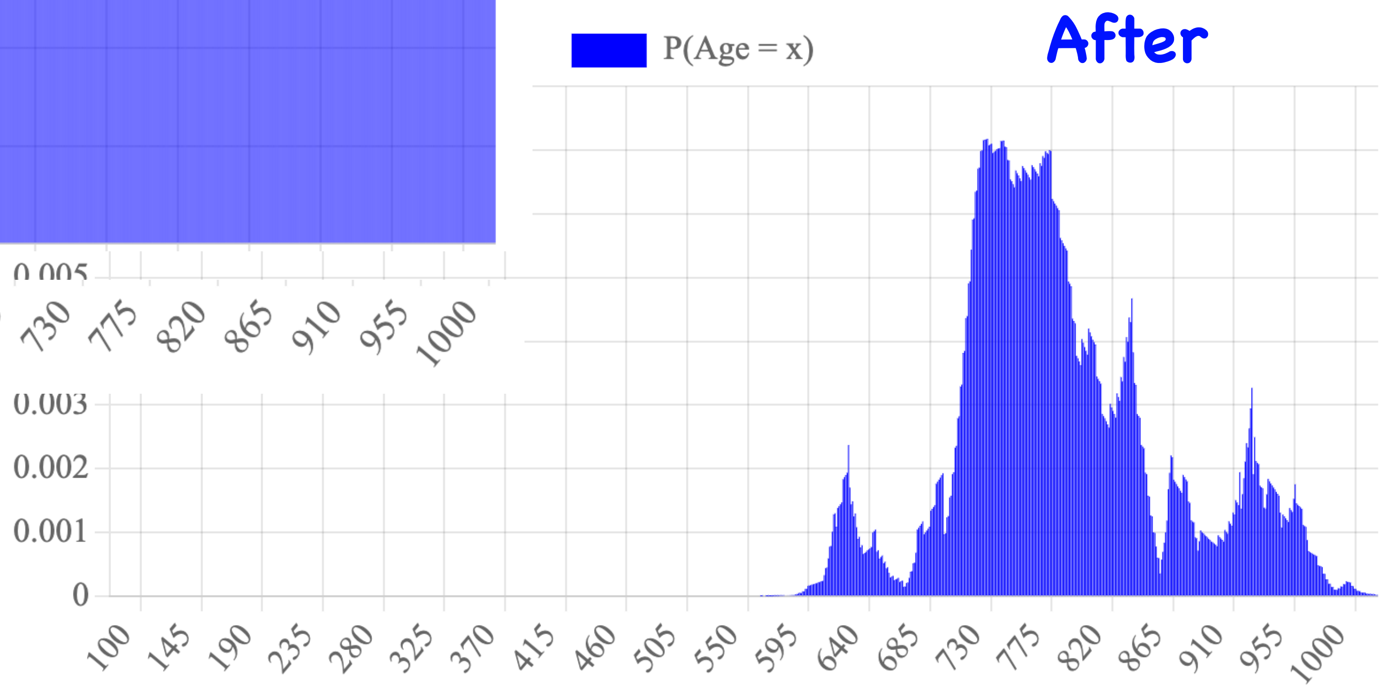


Update Belief PMF



Observation

Remaining C14: 900



Bayesian Carbon Dating: Inference Overview

Let A be how many years old the sample is (A = 100 means the sample is 100 years old)
Let M be the observed amount of C14 left in the sample

$$P(A = i | M = 900) = \frac{P(M = 900 | A = i) P(A = i)}{P(M = 900)}$$

$$= P(M = 900 | A = i) \cdot \frac{P(A = i)}{P(M = 900)}$$

$$= P(M = 900 | A = i) \cdot P(A = i) \cdot K$$

Understanding Through Code

$$P(A = i | M = 900) = P(M = 900 | A = i) \cdot P(A = i) \cdot K$$

```
def update_belief(m = 900):
    """
    Returns a dictionary A, where A[i] contains the
    corresponding probability, P(A = i| M = 900).
    m is the number of C14 molecules remaining and i
    is age in years. i is in the range 100 to 10000
    """
    A = {}
    n_years = 9901
    for i in range(100,10000+1):
        prior = 1 / n_years # P(A = i)
        likelihood = calc_likelihood(m, i) # P(M=m | A=i)
        A[i] = prior * likelihood
    # implicitly computes the normalization constant
    normalize(A)
    return A
```

Carbon Dating Specific Math

Probability of Having 900 Remain

$$P(M = 900 | A = i)$$

There were originally 1000 C14 molecules.
Each molecule remains independently with equal probability p_i
What is the probability that 900 remain?

$$M \sim \text{Bin}(n = 1000, p = p_i)$$

$$P(M = 900 | A = i) = \binom{1000}{900} (p_i)^{900} \cdot (1 - p_i)^{100}$$

Each molecules' time to live is exponential with λ = 1/8267
Let T be the time to decay for any one molecule

$$T \sim \text{Exp}(\lambda = 1/8267) \quad p_i = P(T > i) = 1 - P(T < i) = e^{-\frac{i}{8267}}$$

Probability of Having 900 Remain

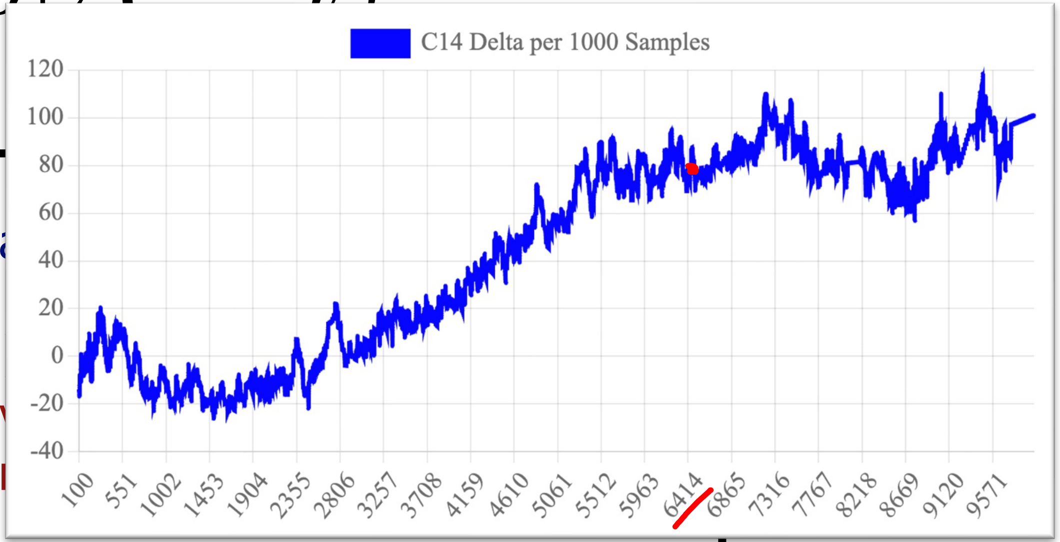
$$P(M = 900 | A = i)$$

```
def calc_likelihood(m = 900, age):
    """
    Computes P(M = m | A = age), the probability of
    having m molecules left given the sample is age
    years old. Uses the exponential decay of C14
    """
    n_original = 1000
    p_remain = math.exp(-age/C14_MEAN_LIFE)
    return stats.binom.pmf(m, n_original, p_remain)
```

Probability of Having 900 Remain

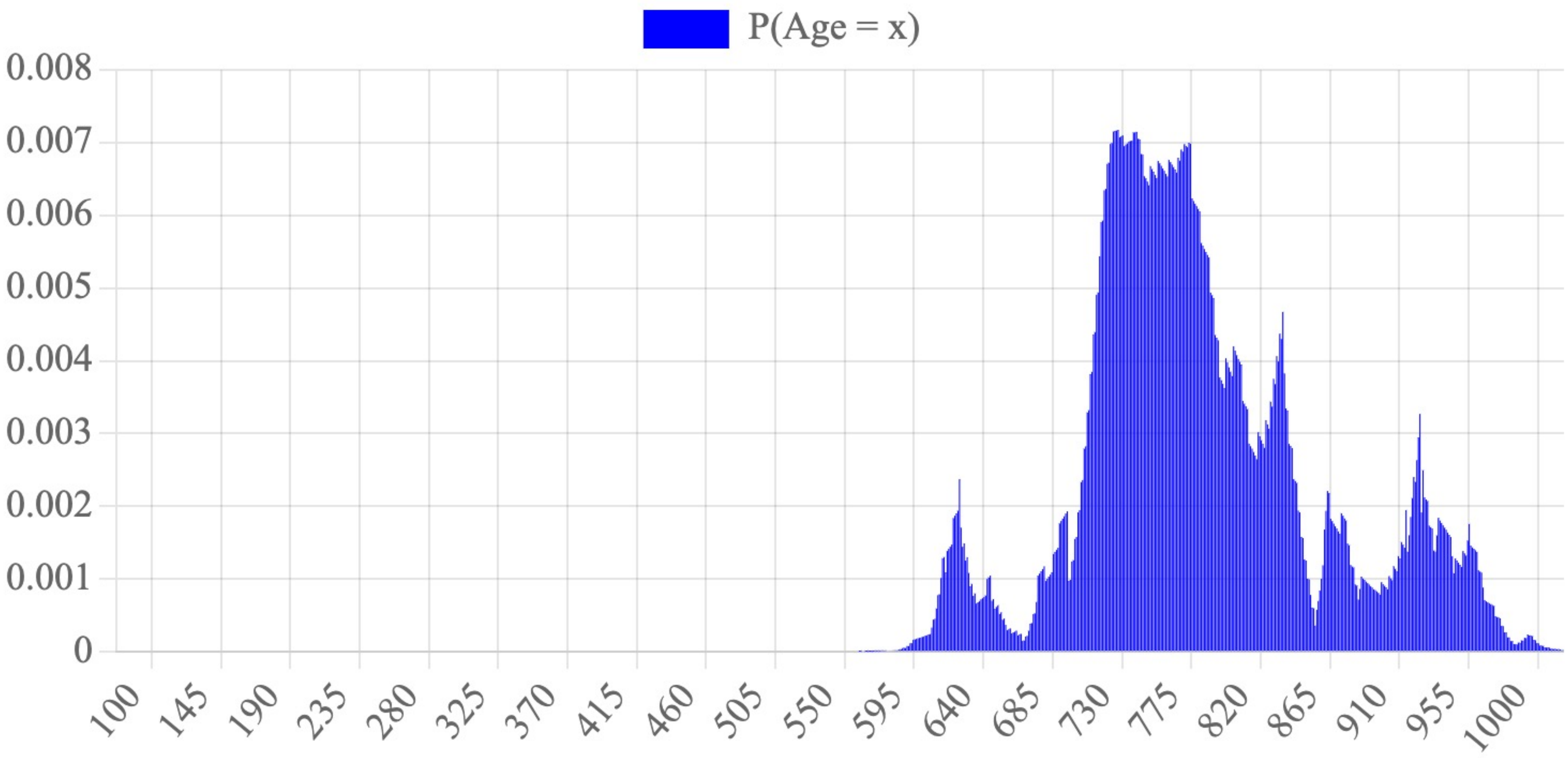
$$P(M = 900 | A = i)$$

```
def calc_likelihood(m = 900, a
    """
    Computes P(M = m | A = age
    having m molecules left giv
    years old. Uses the exponen
    """
    n_original = 1000 + delta_start(age)
    p_remain = math.exp(-age/C14_MEAN_LIFE)
    return stats.binom.pmf(m, n_original, p_remain)
```



Probability of Having 900 Remain

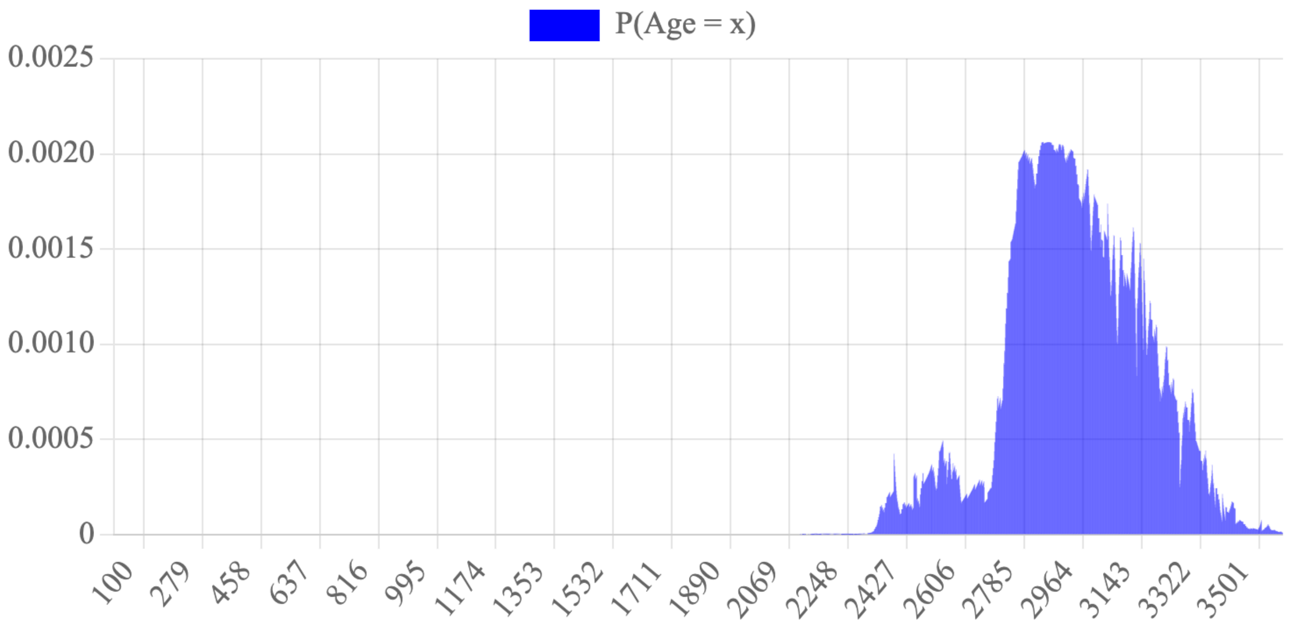
Remaining C14: 900



Come Back for More!

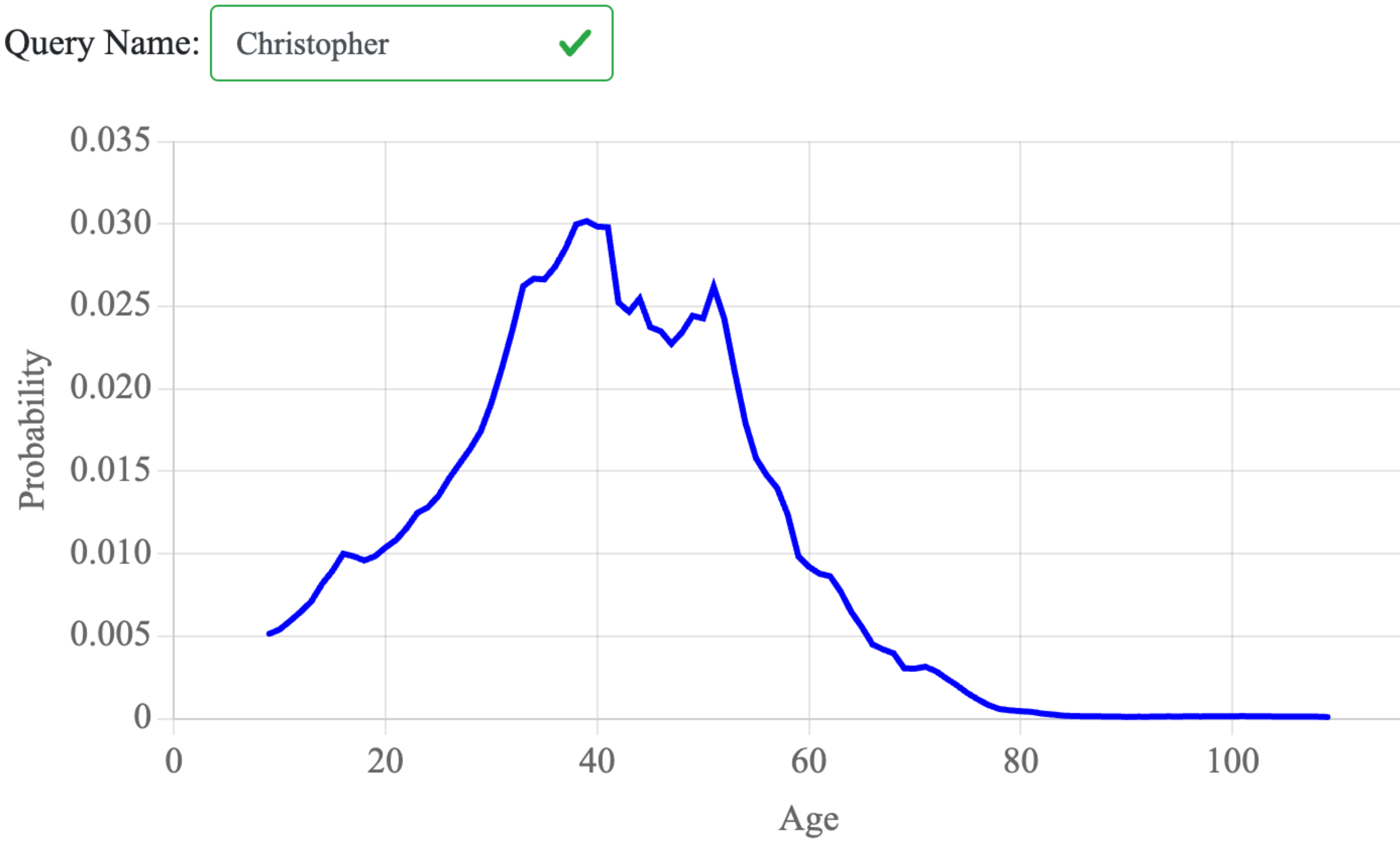
Belief in **Age** Given **Observed C14**

P(Age | Observed C14)

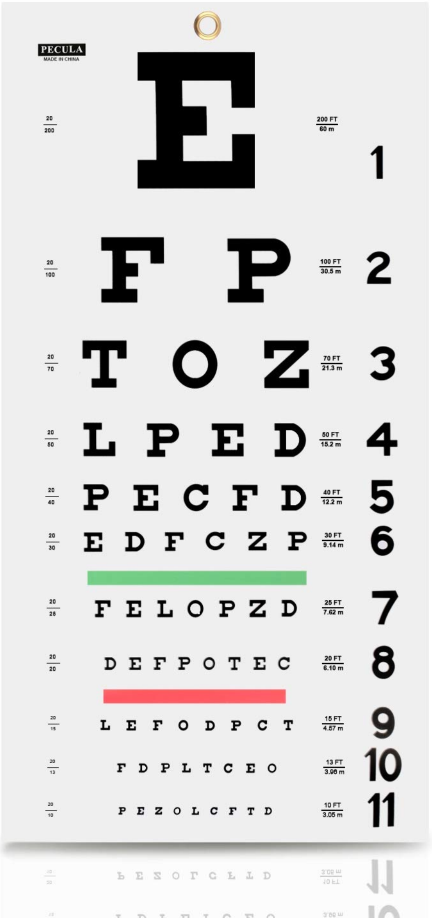


Belief in Age Given Name

P(Age | Name)



Belief in **Vision** Given **User Responses**



P(Ability to See | Observed Responses)