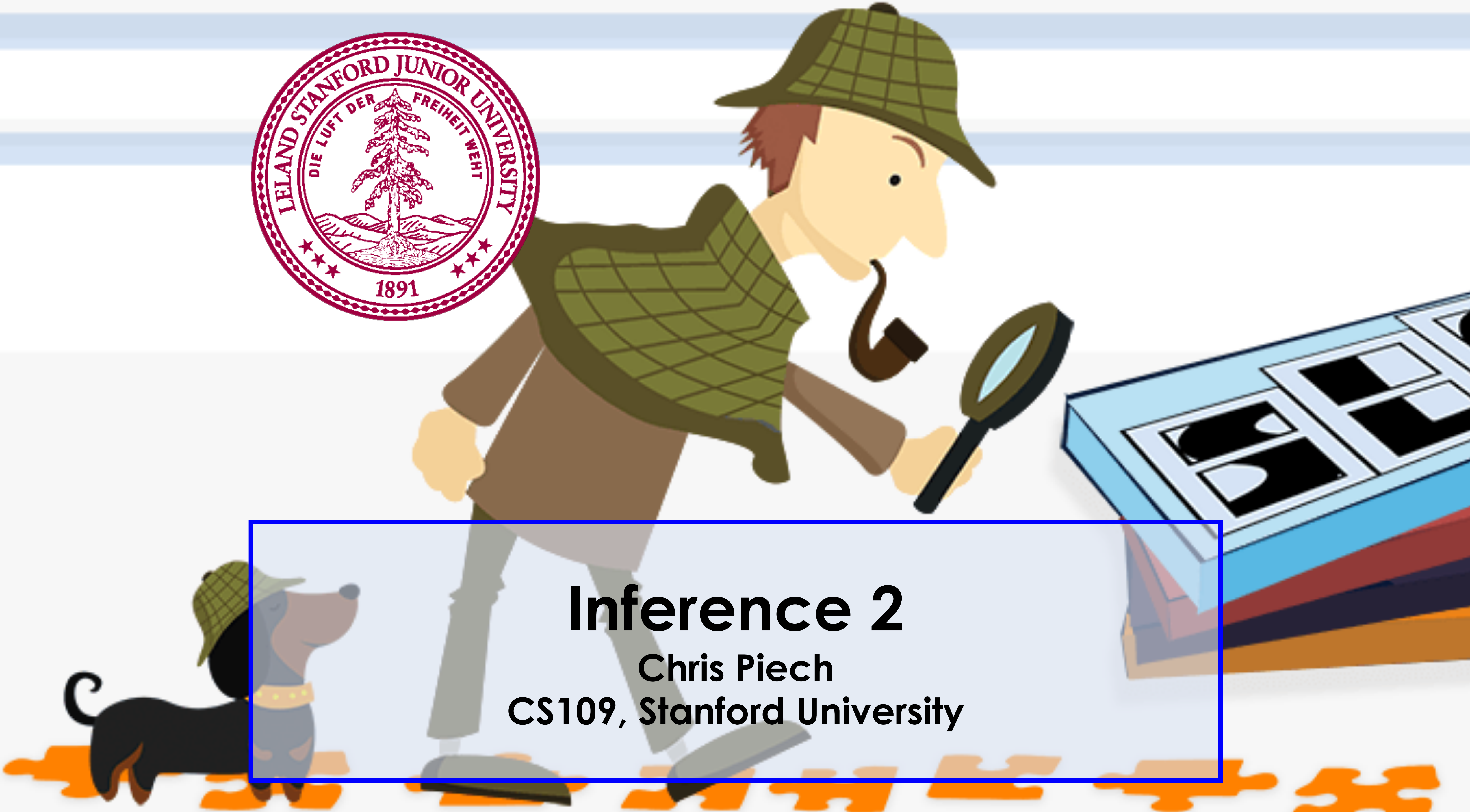
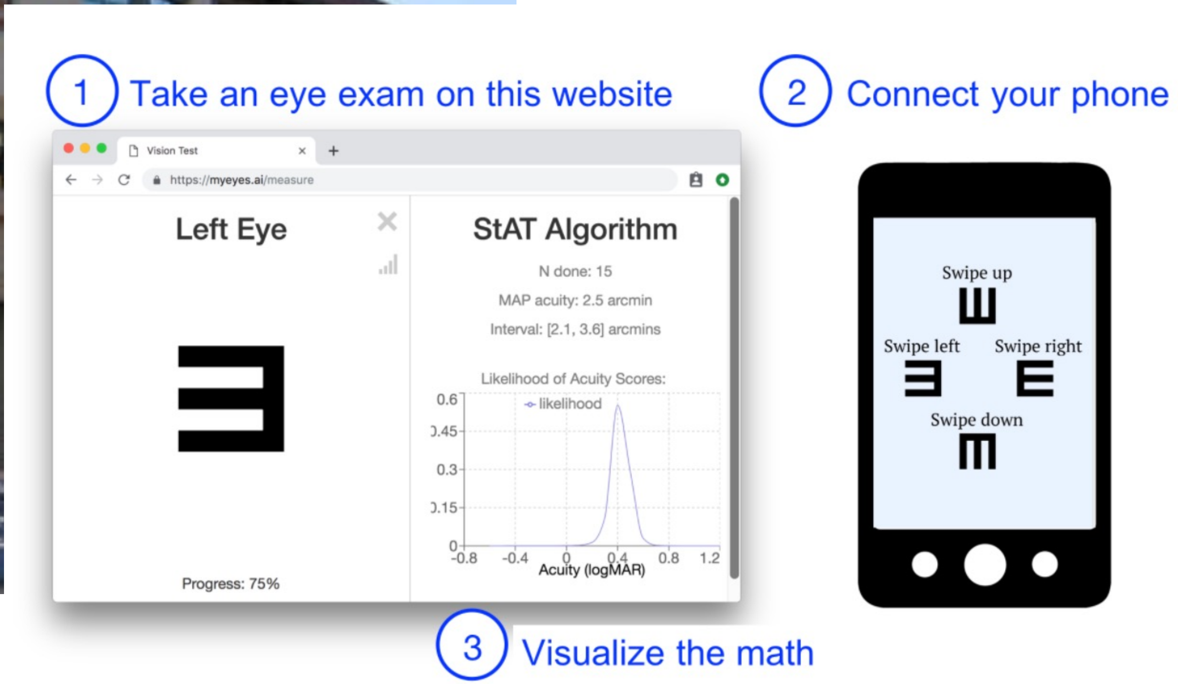


Inference 2

Chris Piech
CS109, Stanford University

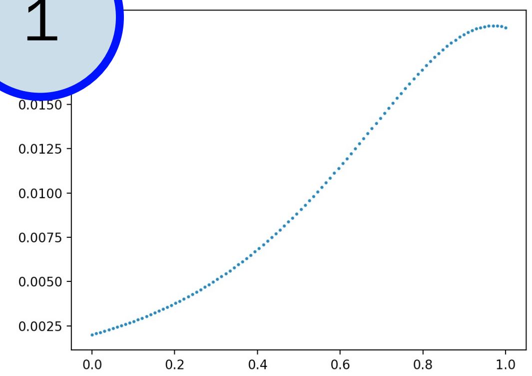
Today: Stanford Eye Test



I always wanted to make this a class demo

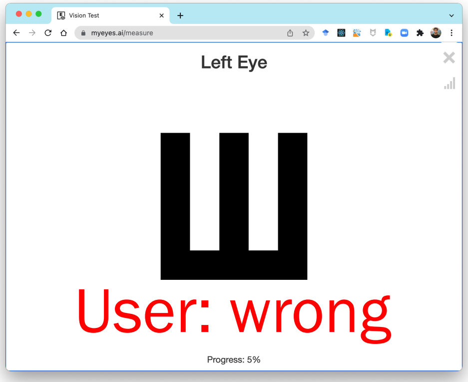
Stanford Eye Test

1



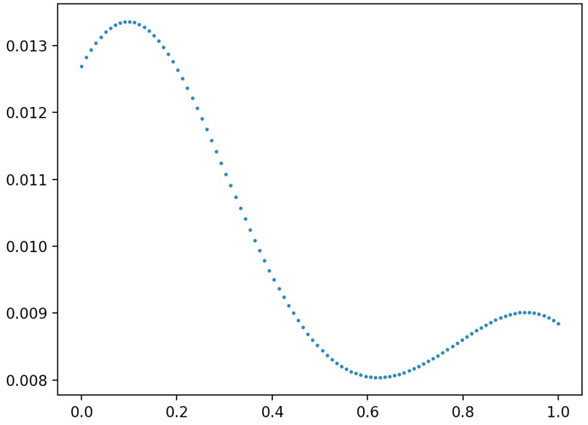
$$P(A = a)$$

2



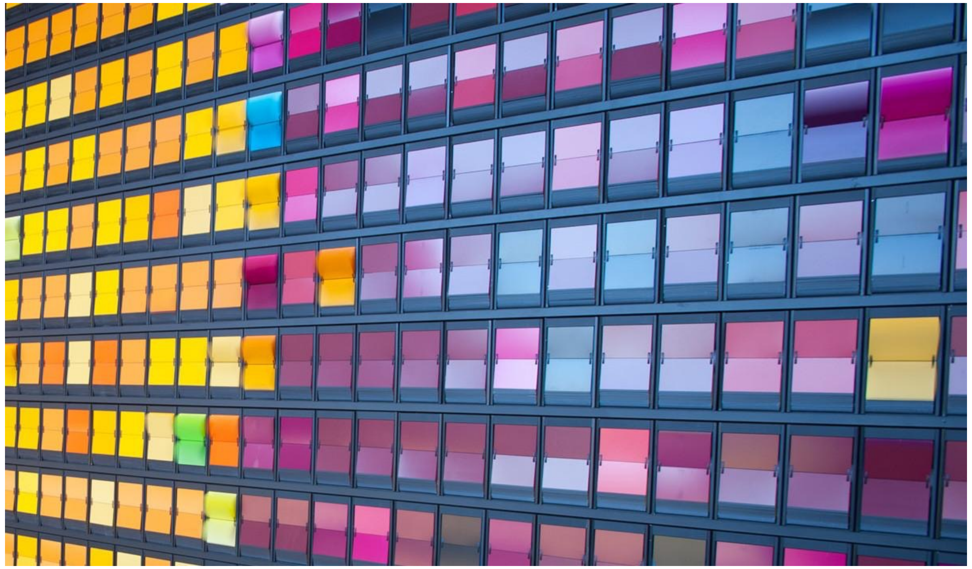
Observation E

3



$$P(A = a|E)$$

Midterm Location



CEMEX: 587 seats

109 students

Hewlett 200: 502 seats

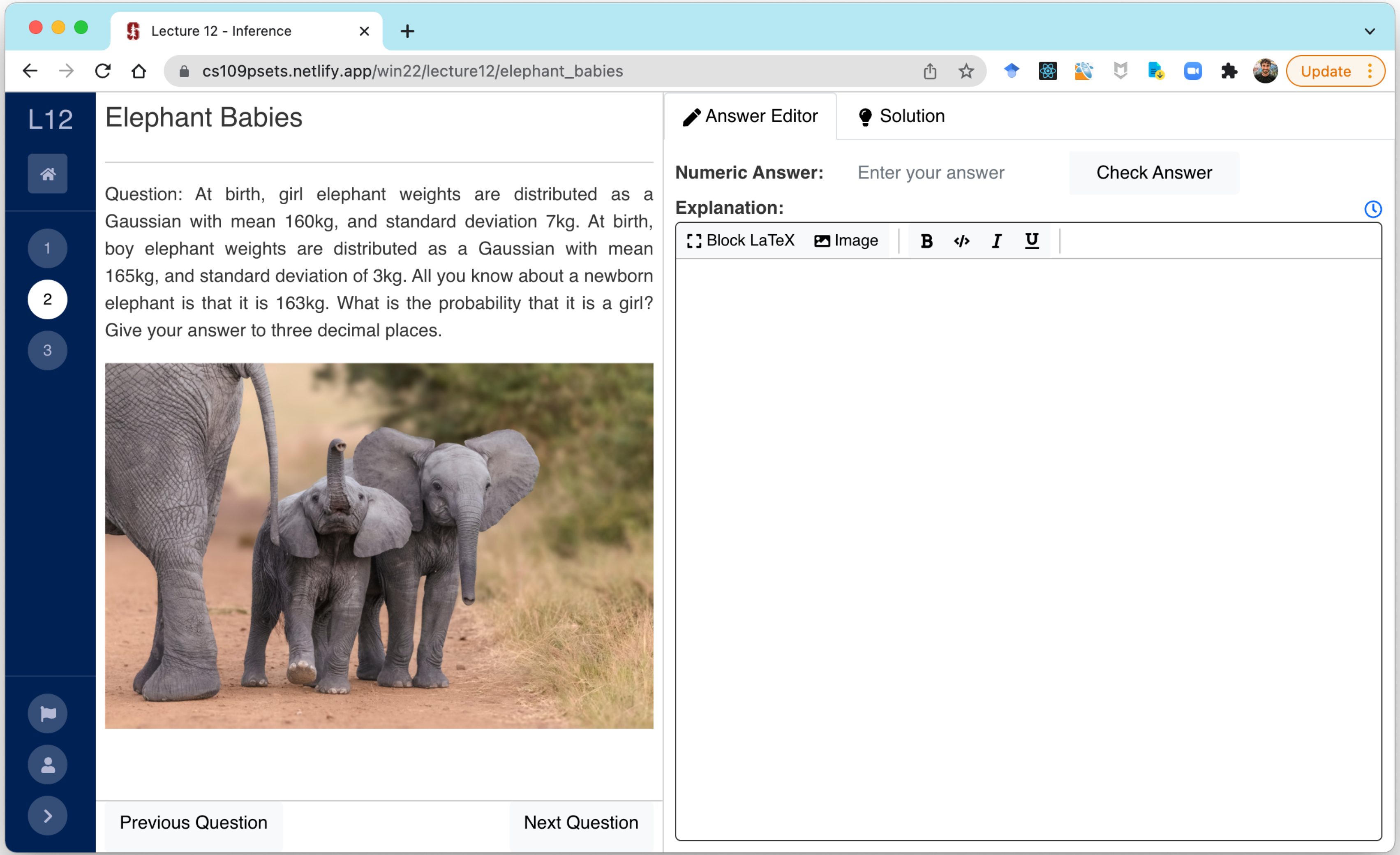
94 students

Hewlett 201: 188 seats

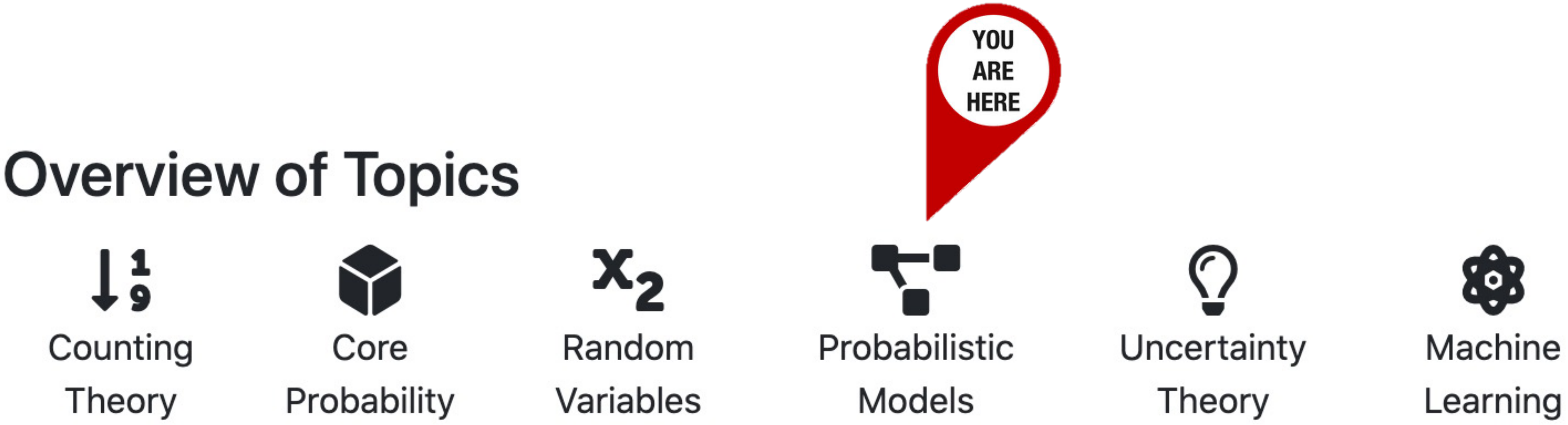
36 students

Look out for your assignment! Let us know if you are at risk

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



Where are we in CS109?



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. Combine Bayes Theorem and Random Variables

Review

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?

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?

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

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

End Review

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 $\text{Bern}(p = 0.5)$

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

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**

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

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

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

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, $\sigma^2 = 3^2$)

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

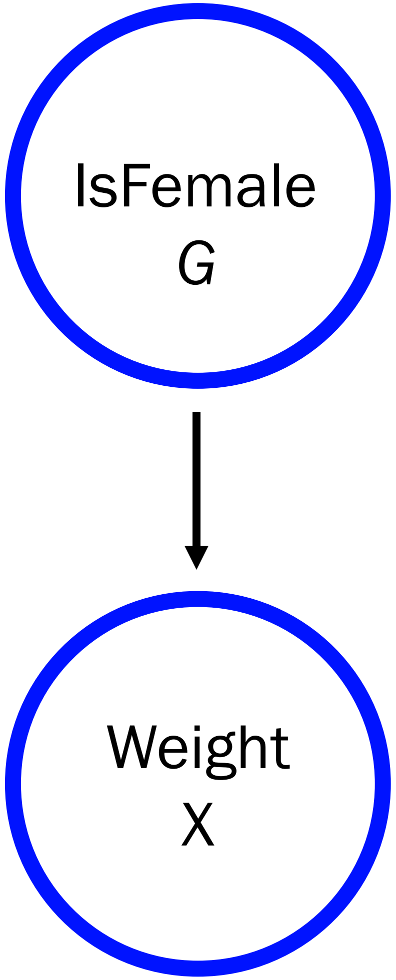
Let G be an indicator that the elephant is a girl. G is $\text{Bern}(p = 0.5)$

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 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?

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

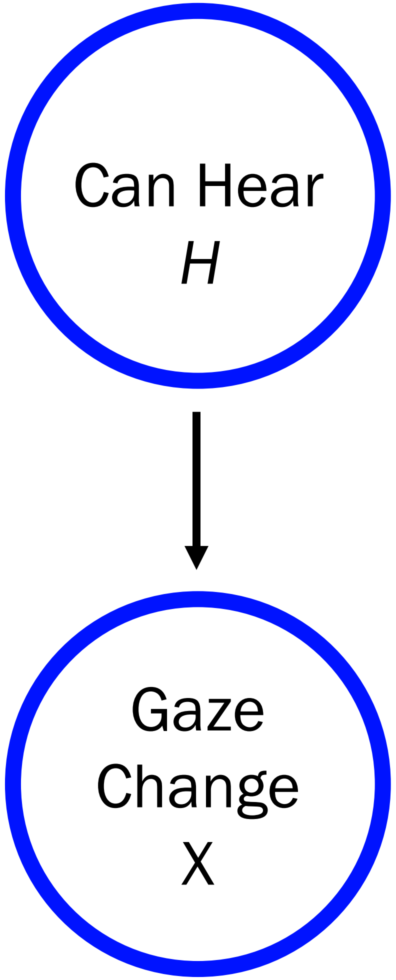
Q: What is P(G = 1 | X = 163)

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 $\text{Bern}(p = 0.75)$

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

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

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

Harder when neither random variable is a bernoulli

A Better Eye Test

https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)02149-8/fulltext

https://www.science.org/content/article/eye-robot-artificial-intelligence-dramatically-improves-accuracy-classic-eye-exam

https://ojs.aaai.org/index.php/AAAI/article/view/5384/5240

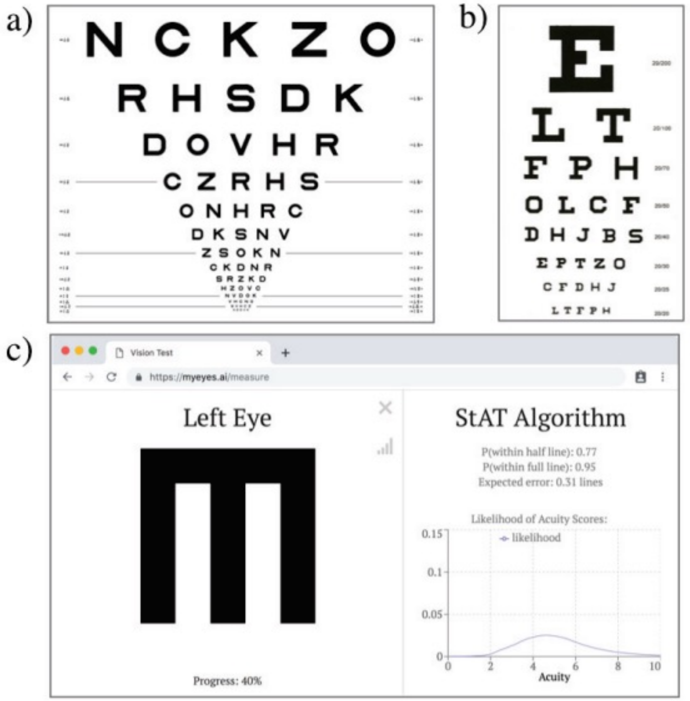
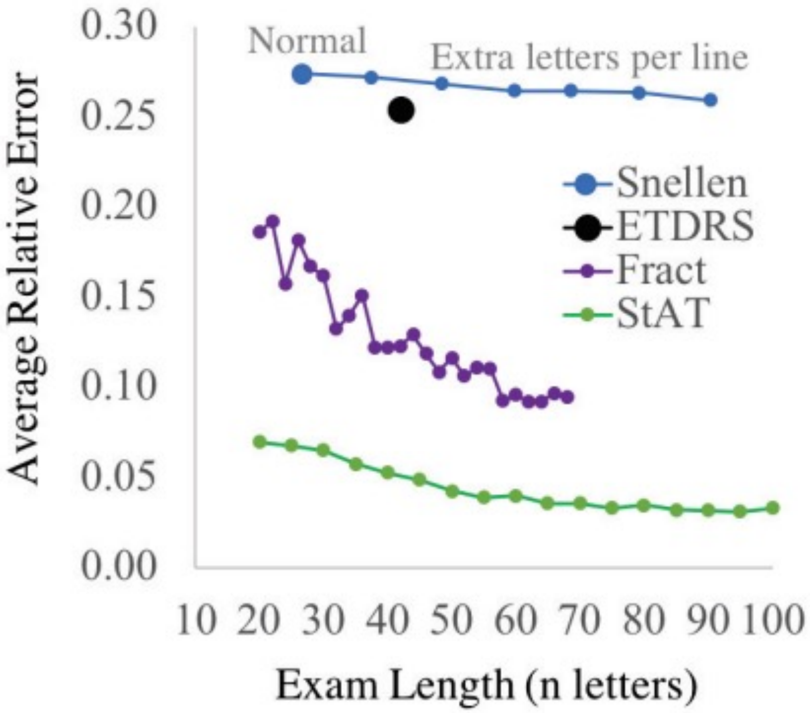
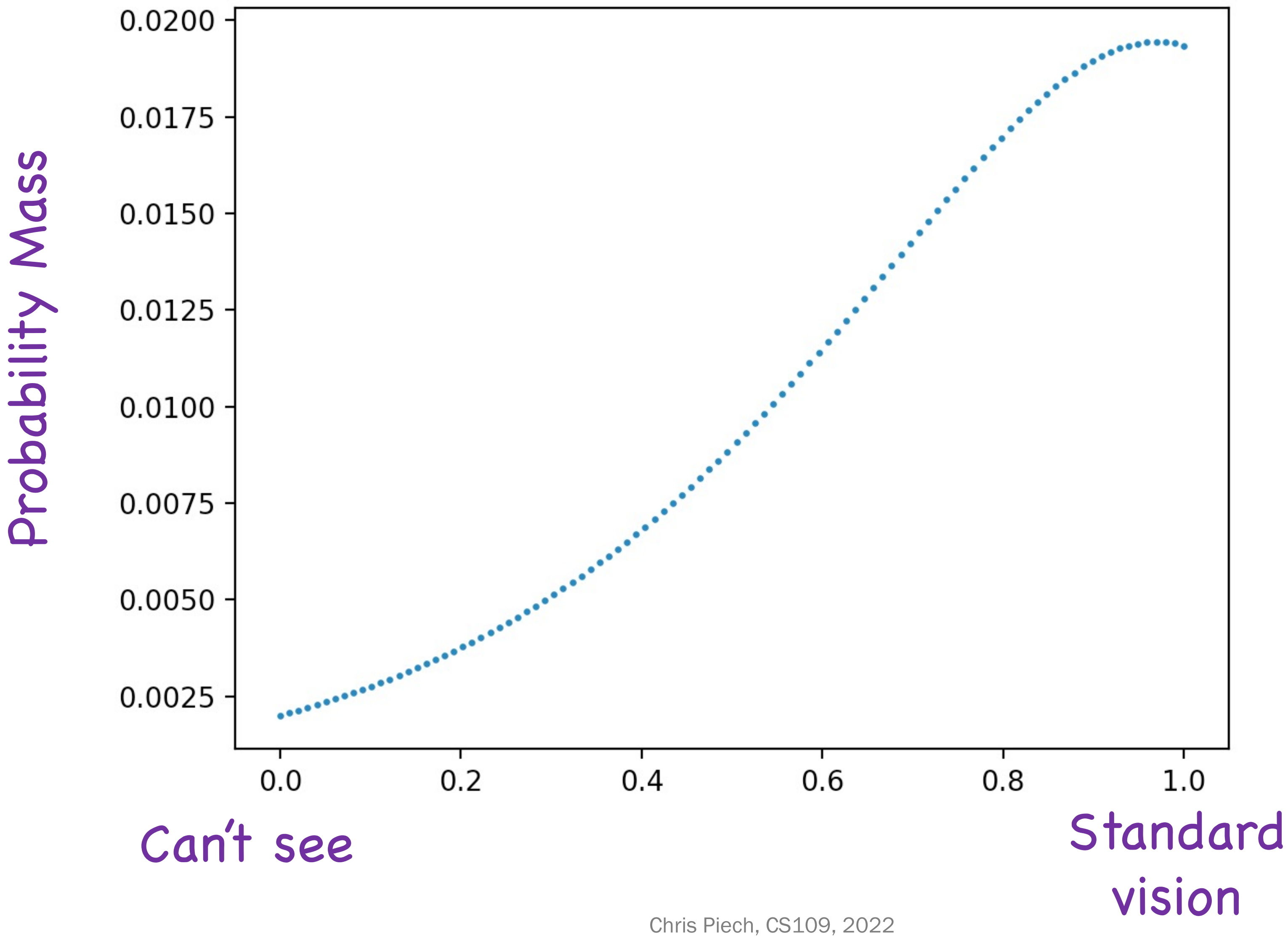


Figure 1: a) ETDRS, b) Snellen and c) StAT eye exams.



Prior Belief in Ability to See (Random Var A)



PMF is Actually Stored as a Dictionary

```python
def main():
    belief = get_prior_belief()
```

a	P(A=a)
0.00	0.00198
0.01	0.00205
0.02	0.00211
0.03	0.00218
0.04	0.00225
0.05	0.00233
0.06	0.0024
0.07	0.00248
0.08	0.00256
0.09	0.00264
0.10	0.00273
0.11	0.00281
0.12	0.0029
0.13	0.00299
0.14	0.00309
0.15	0.00319
0.16	0.00329
0.17	0.00339
0.18	0.0035
0.19	0.00361

a	P(A=a)
0.20	0.00372
0.21	0.00384
0.22	0.00396
0.23	0.00408
0.24	0.00421
0.25	0.00434
0.26	0.00447
0.27	0.00461
0.28	0.00475
0.29	0.00489
0.30	0.00504
0.31	0.00519
0.32	0.00535
0.33	0.00551
0.34	0.00567
0.35	0.00584
0.36	0.00601
0.37	0.00619
0.38	0.00637
0.39	0.00655

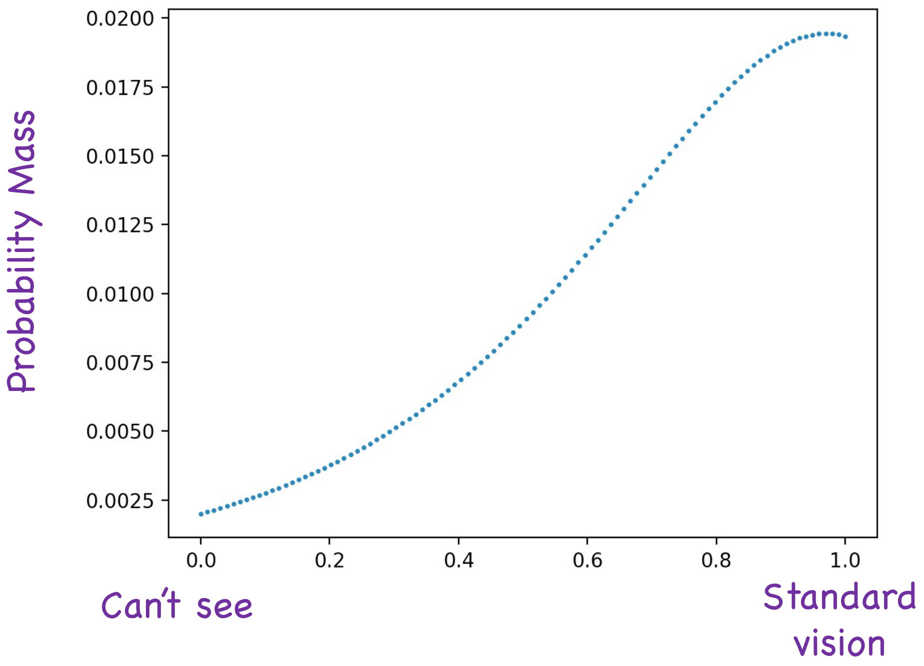
...

a	P(A=a)
0.80	0.01684
0.81	0.01708
0.82	0.01731
0.83	0.01753
0.84	0.01774
0.85	0.01795
0.86	0.01814
0.87	0.01832
0.88	0.01848
0.89	0.01864
0.90	0.01877
0.91	0.0189
0.92	0.019
0.93	0.01909
0.94	0.01916
0.95	0.01921
0.96	0.01924
0.97	0.01925
0.98	0.01924
0.99	0.01921

Prior Belief in Ability to See (Random Var A)

```
belief = get_prior_belief()
```

As a graph



As a dictionary

a	P(A=a)
0.00	0.00198
0.01	0.00205
0.02	0.00211
0.03	0.00218
0.04	0.00225
0.05	0.00233
0.06	0.0024
0.07	0.00248
0.08	0.00256
0.09	0.00264
0.10	0.00273
0.11	0.00281
0.12	0.0029
0.13	0.00299
0.14	0.00309
0.15	0.00319
0.16	0.00329
0.17	0.00339
0.18	0.0035
0.19	0.00361

a	P(A=a)
0.20	0.00372
0.21	0.00384
0.22	0.00396
0.23	0.00408
0.24	0.00421
0.25	0.00434
0.26	0.00447
0.27	0.00461
0.28	0.00475
0.29	0.00489
0.30	0.00504
0.31	0.00519
0.32	0.00535
0.33	0.00551
0.34	0.00567
0.35	0.00584
0.36	0.00601
0.37	0.00619
0.38	0.00637
0.39	0.00655

...

a	P(A=a)
0.80	0.01684
0.81	0.01708
0.82	0.01731
0.83	0.01753
0.84	0.01774
0.85	0.01795
0.86	0.01814
0.87	0.01832
0.88	0.01848
0.89	0.01864
0.90	0.01877
0.91	0.0189
0.92	0.019
0.93	0.01909
0.94	0.01916
0.95	0.01921
0.96	0.01924
0.97	0.01925
0.98	0.01924
0.99	0.01921

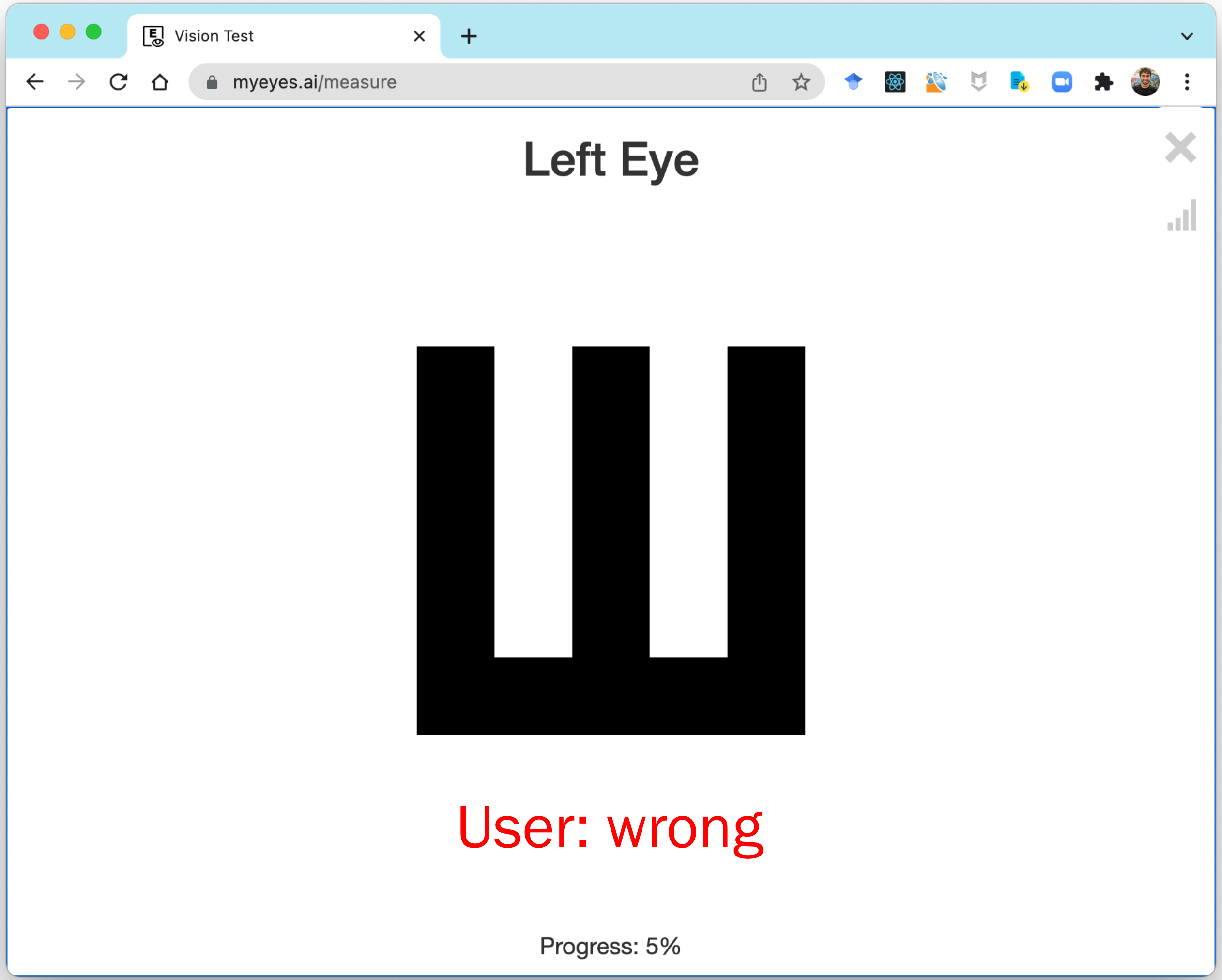
Today: I am going to simplify the units of vision

Normally doctors measure vision in logarithmic units. To make today's demo easier to understand:

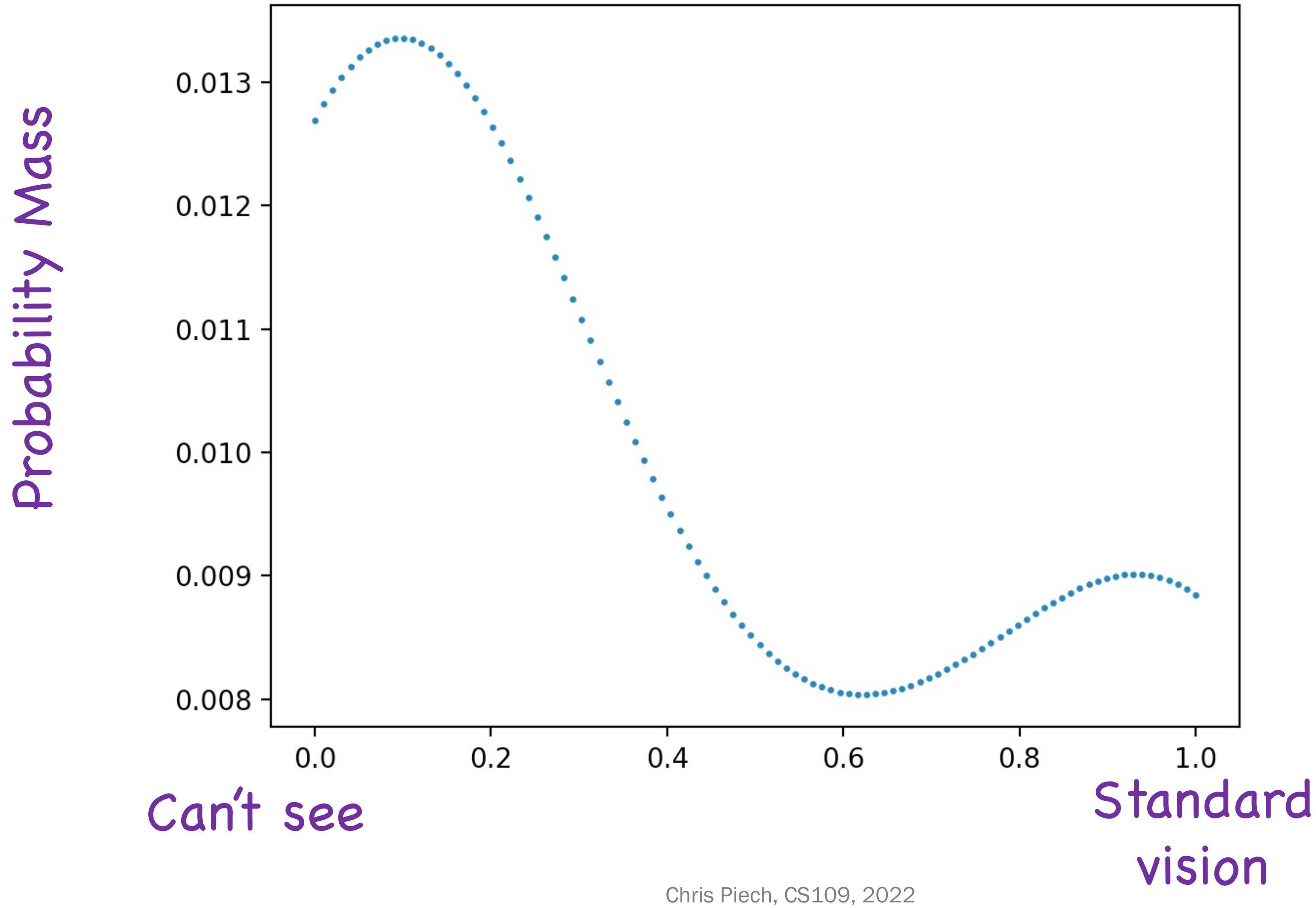
I have translated ability to see

onto a [0, 1] scale.

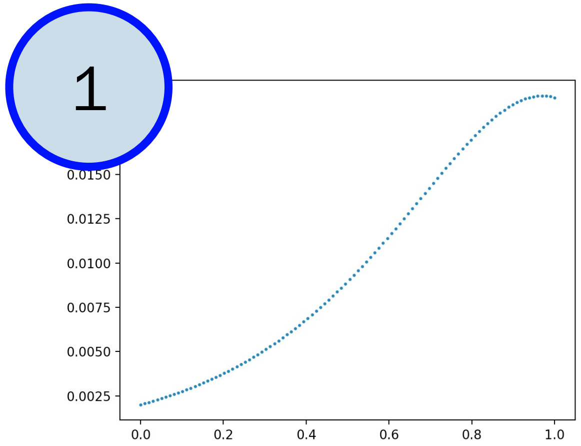
The Patient is Shown One Letter and They Get it Wrong



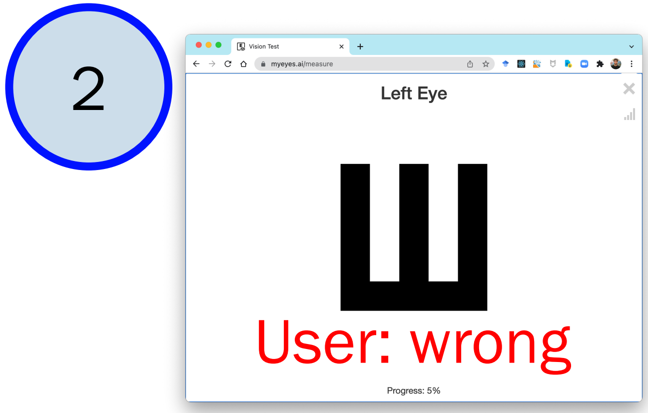
Posterior Belief in Ability to See (Random Var A)



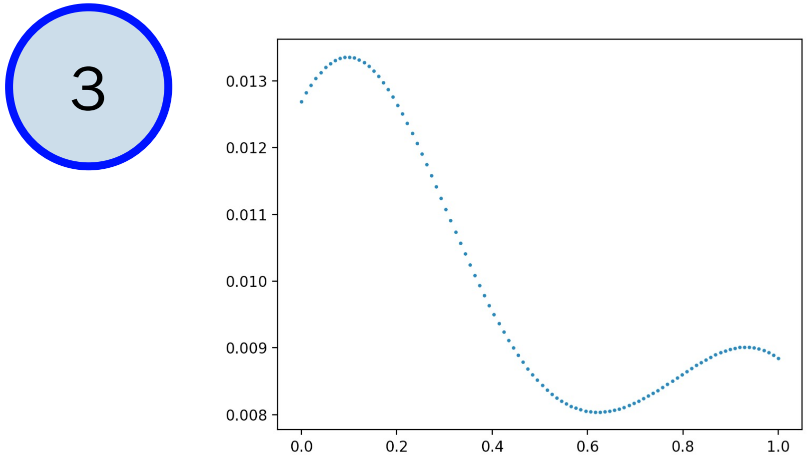
Bayes with Random Variables



$$P(A = a)$$



Observation E



$$P(A = a | E)$$

$$P(A = a|E) = \frac{P(E|A = a)P(P = a)}{P(E)}$$

Inference

In general Bayes theorem with a random variable is like the cellphone problem: multiple possible assignments to keep track of

Still true when some variables are continuous

Random Variables

Not all beliefs can be represented as a **function**. **Dictionary** / table is a great way to represent a random variable belief.

This is formally called non-parametric

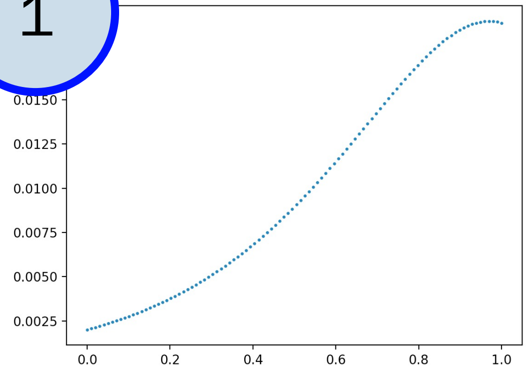
Representing Continuous Variables

Dictionary can also be used to represent a discretization of a continuous random var

I do it all the time! Yay compute!

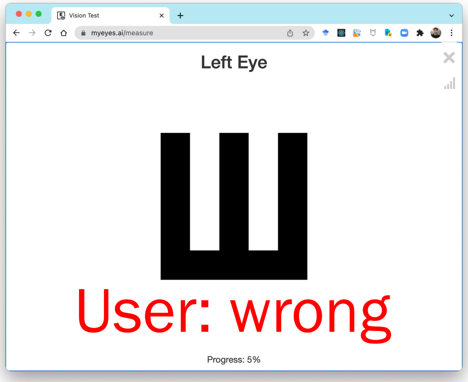
To the Code!!!

1



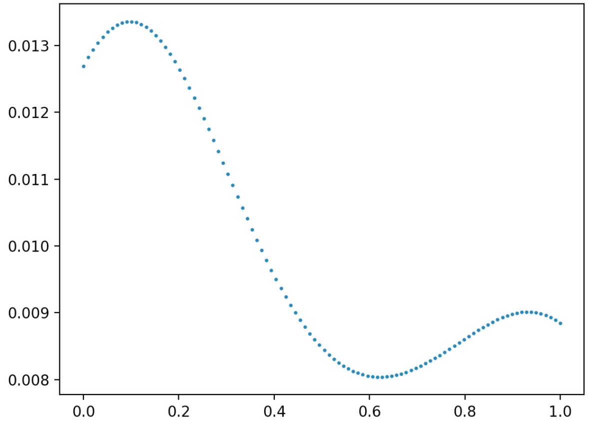
$$P(A = a)$$

2



Observation E

3



$$P(A = a|E)$$

Known function

$$P(A = a|E) = \frac{P(E|A = a)P(P = a)}{P(E)}$$

Inference

In general Bayes theorem with a random variable is like the cellphone problem: multiple possible assignments to keep track of

Still true when some variables are continuous

Random Variables

Not all beliefs can be represented as a **function**. **Dictionary** / table is a great way to represent a random variable belief.

This is formally called non-parametric

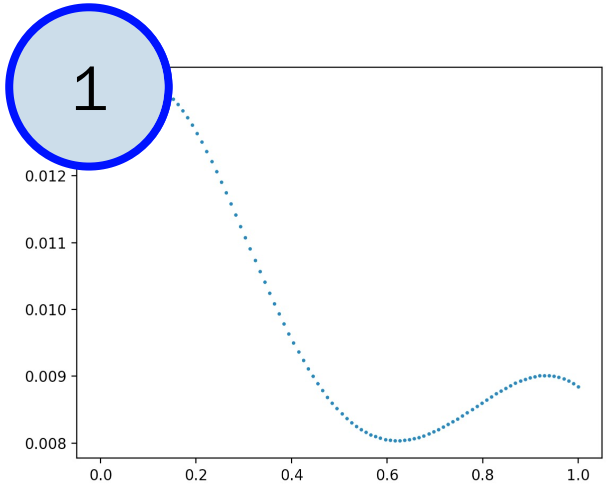
Representing Continuous Variables

Dictionary can also be used to represent a discretization of a continuous random var

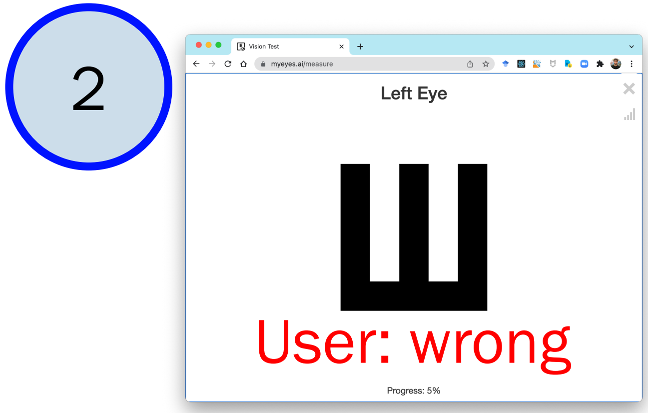
I do it all the time! Yay compute!

Multiple observations??

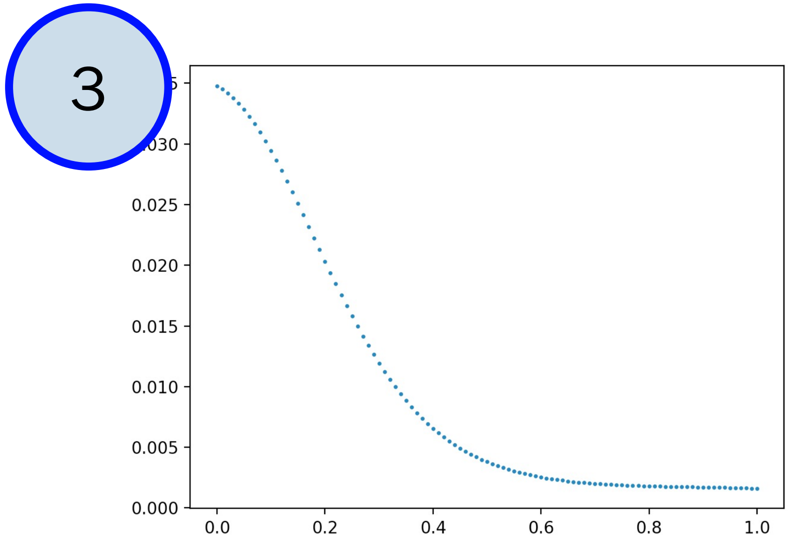
Posterior becomes new prior



$$P(A = a)$$



Observation E



$$P(A = a|E)$$

$$P(A = a|E) = \frac{P(E|A = a)P(P = a)}{P(E)}$$

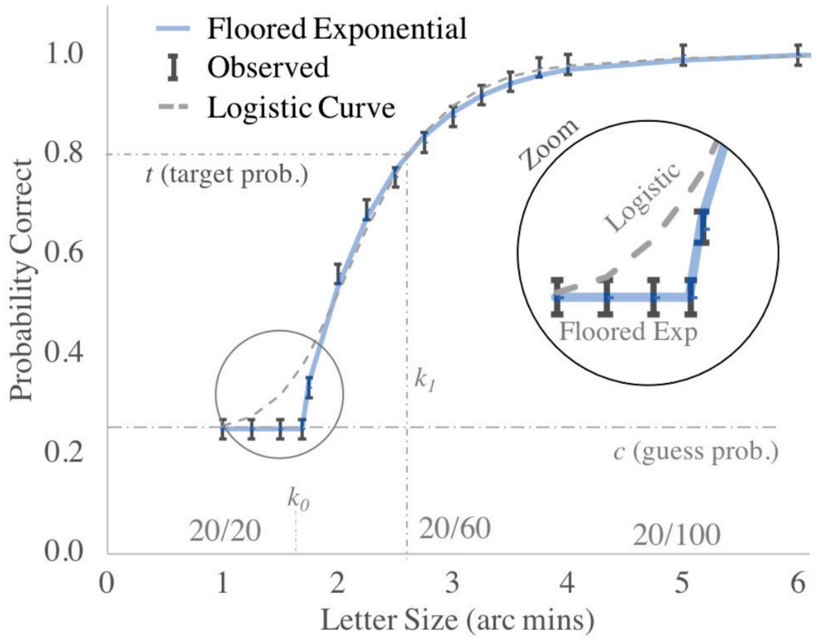
How to get

$$P(E|A = a)$$

How to get probability of a response given ability

$$P(E|A = a) \ ???$$

In Practice



In Theory

```
GUESS = 0.05
SLIP = 0.05
SCALING = 7
def p_correct_given_ability(ability, font_size):
   difficulty = 1 - font_size
   p_no_slip = sigmoid(SCALING * (ability - difficulty))
   p_with_guess = max(p_no_slip, GUESS)
   return SLIP * GUESS + (1-SLIP) * p_with_guess

def p_observation_given_ability(ability, observation):
   p = p_correct_given_ability(ability, observation['size'])
   if observation['correct']:
      return p
   else:
      return 1-p
```

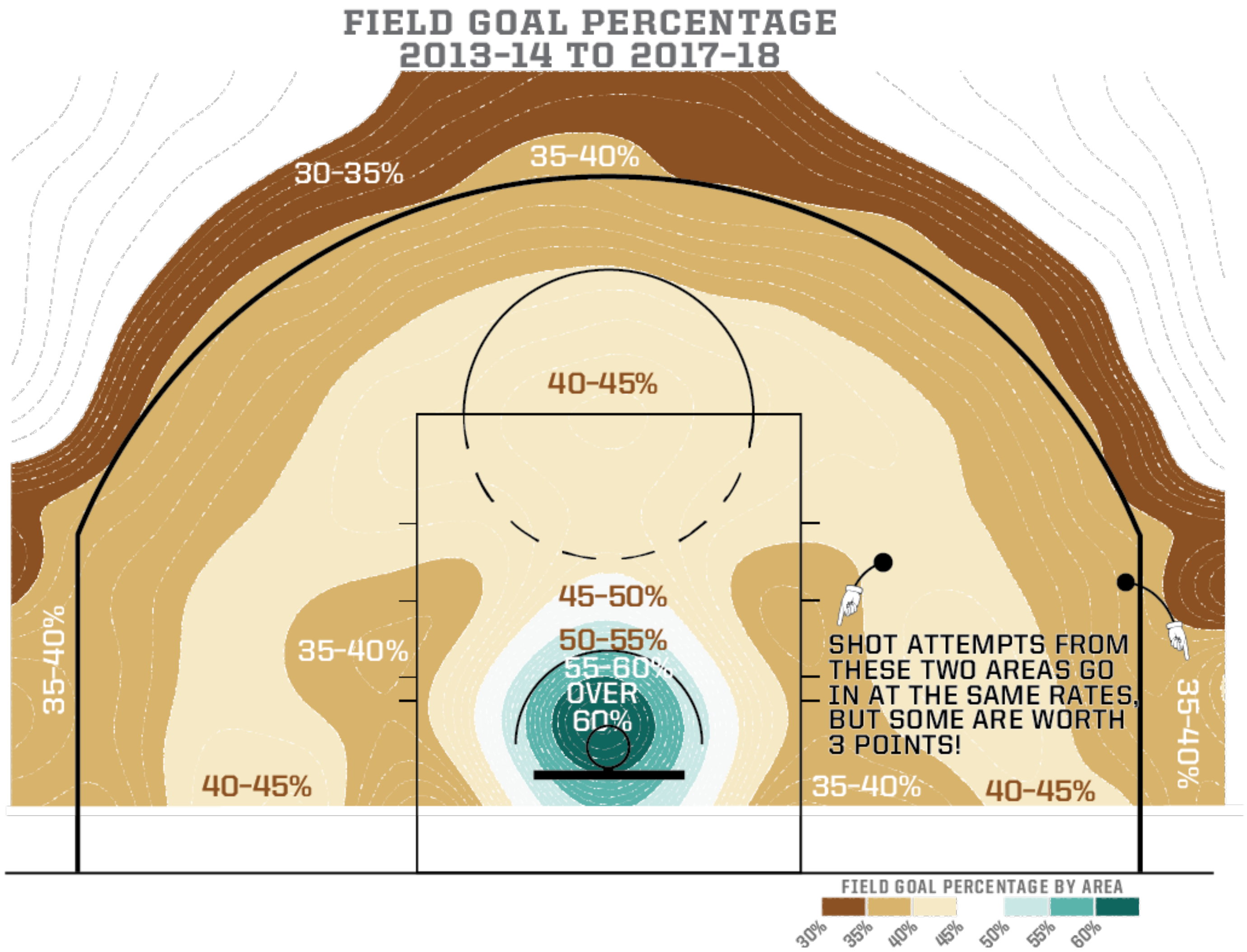
Aside, if time: Visualizing continuous joints

All the Bayes Belong to Us

X, Y are continuous

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

Probabilistic Models can have Continuous Random Vars



X location of the shot

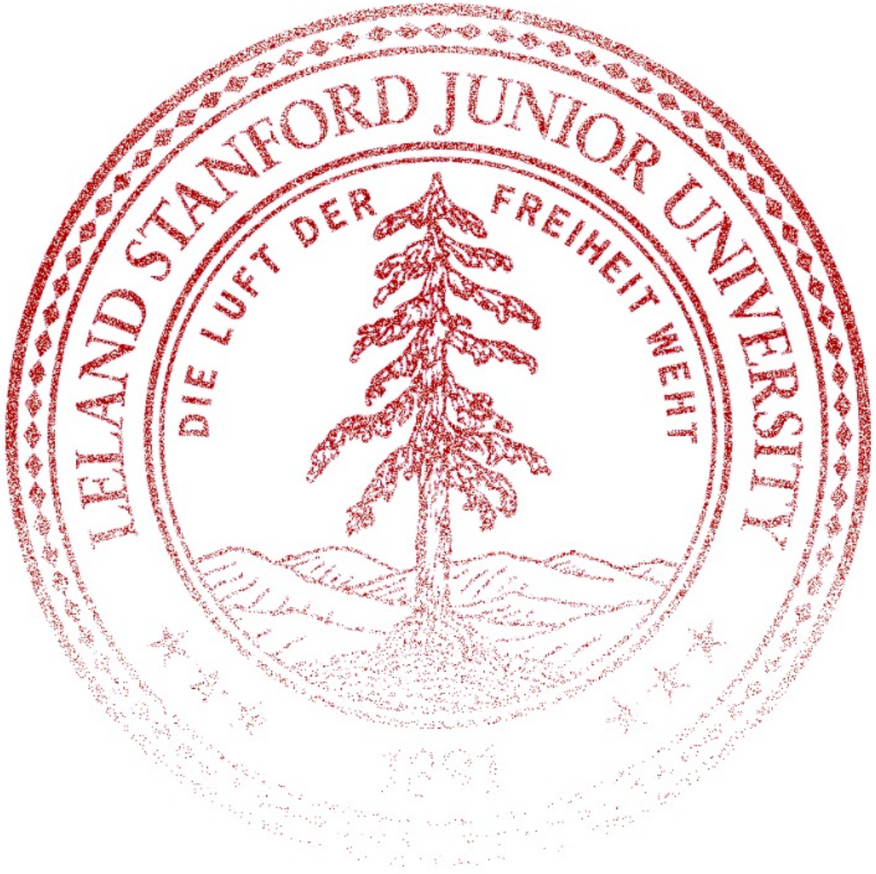
Y location of the shot

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

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

Probabilistic Models can have Continuous Random Vars

Dart Results



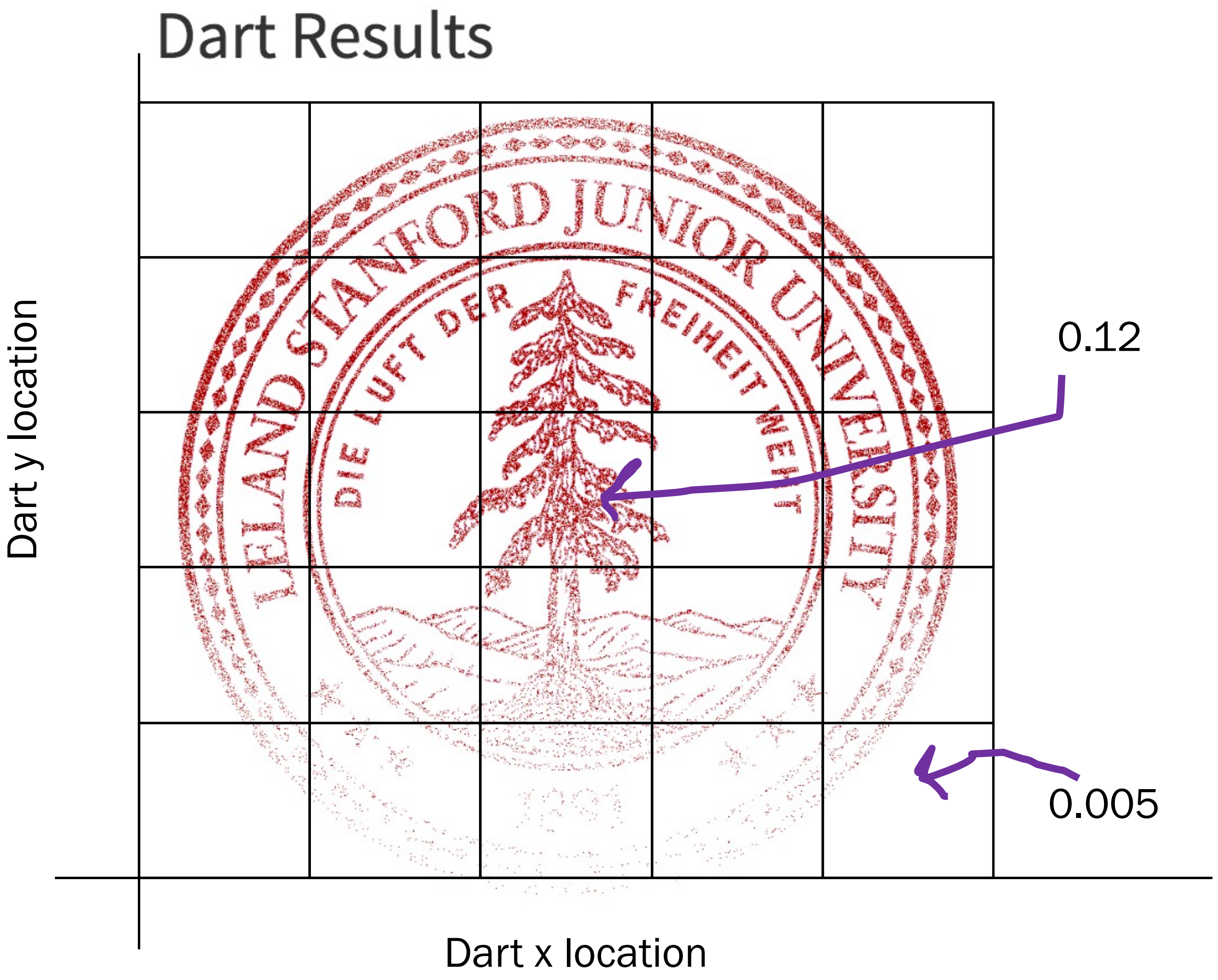
X location of the shot

Y location of the shot

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

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

Probabilistic Models can have Continuous Random Vars



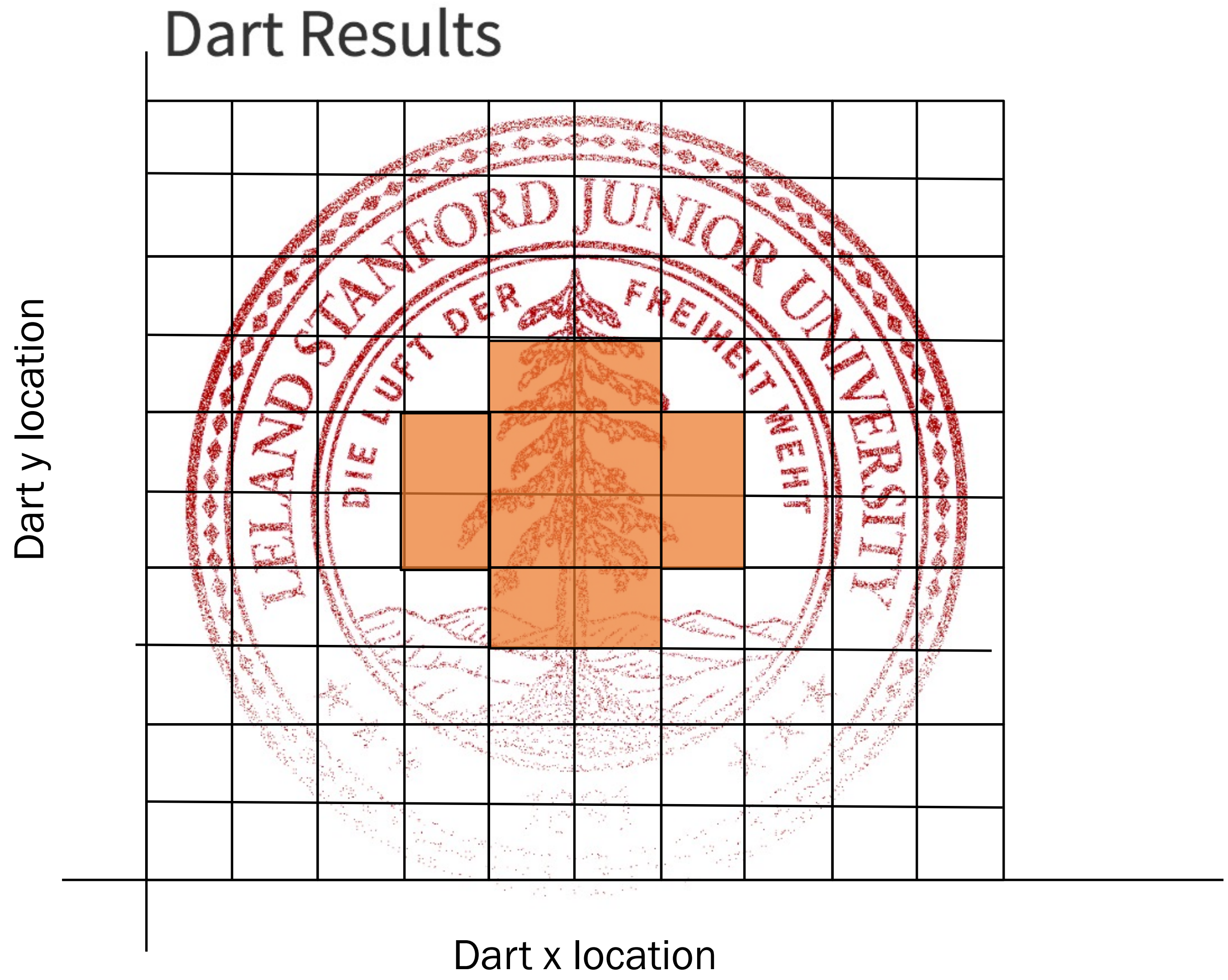
X location of the shot

Y location of the shot

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

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

Probabilistic Models can have Continuous Random Vars



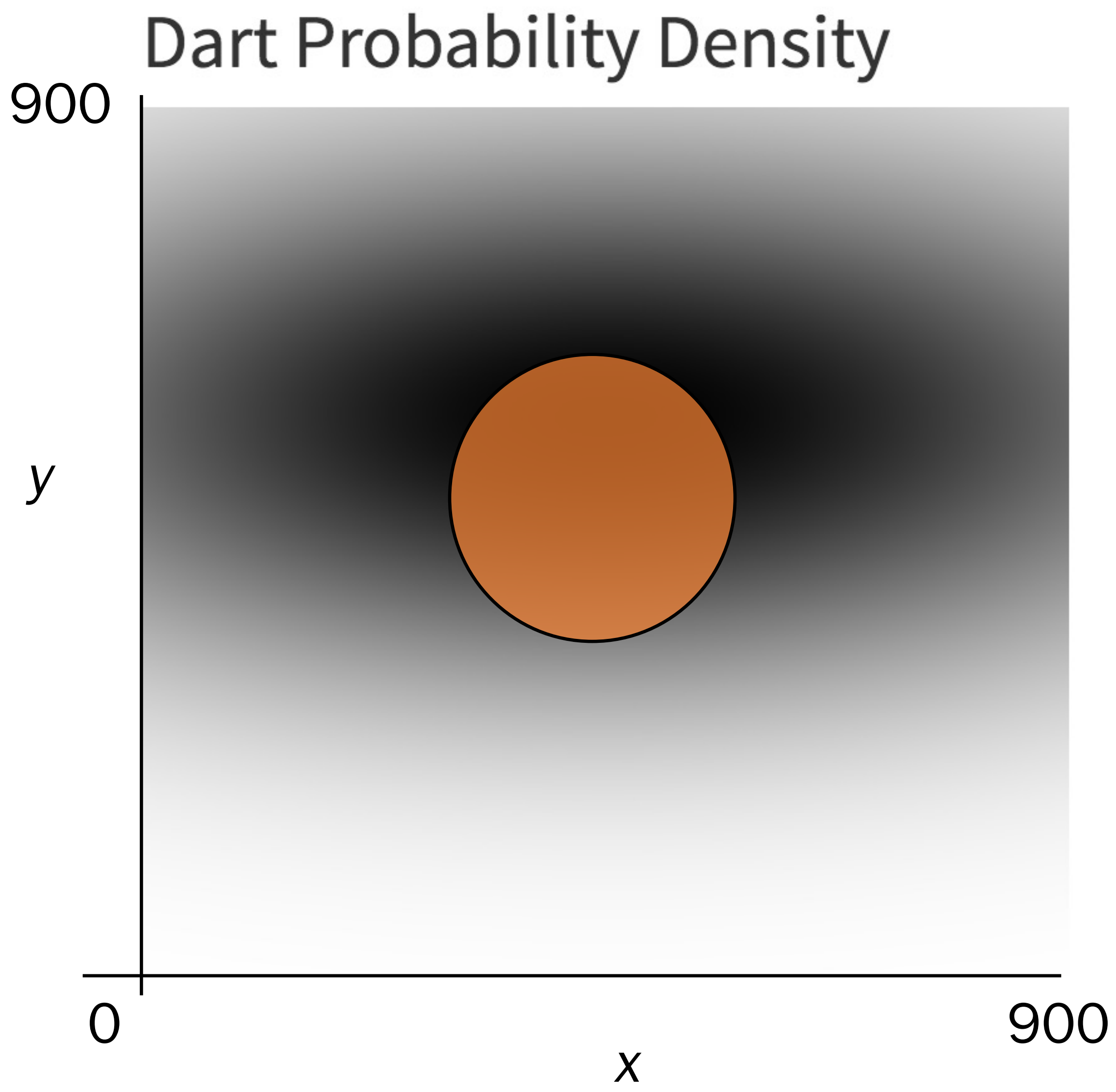
X location of the shot

Y location of the shot

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

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

Probabilistic Models can have Continuous Random Vars



X location of the shot

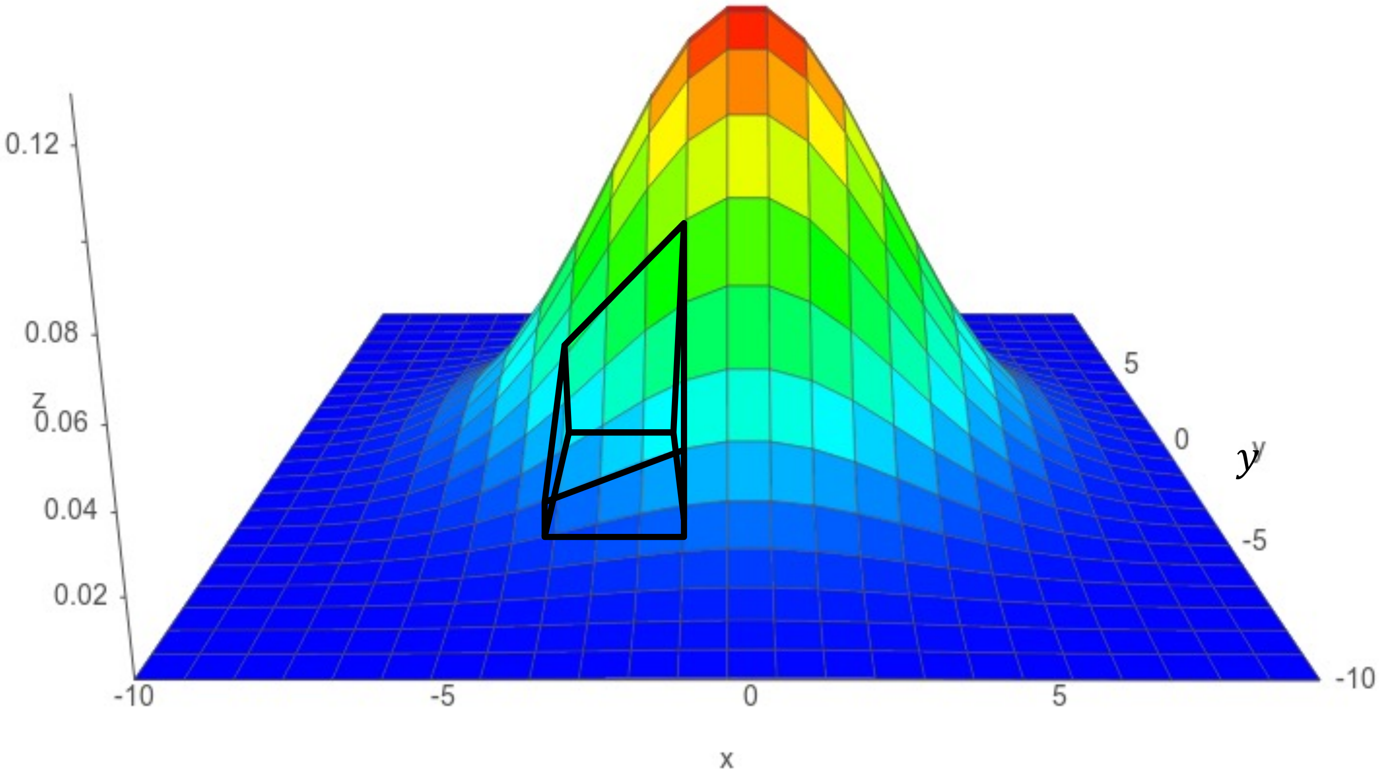
Y location of the shot

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

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

Joint Probability Density Function

$$P(a_1 < X < a_2, b_1 < Y < b_2) = \int_{x=a_1}^{a_2} \int_{y=b_1}^{b_2} f(X = x, Y = y)\ \partial y \partial x$$



plot by Academo

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

Focus on Inference

When there are a mixture of discrete and continuous (or multiple continuous) I want you to focus on inference

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

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

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

End Aside

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 random variables
- [ ] Use and find **expectation** of random variables