Beta: The Random Variable for Probabilities

Chris Piech
CS109, Stanford University

Which video are you more likely to like?

Davie504

10,000 50

Not Davie504

10 0

Which drug should you give if you are uncertain about *p*?

Drug A

Drug B

Which one do you give to a patient?

Philosophical Ponderings:
You ask about the probability of rain tomorrow.

Person A: My leg itches when it rains and its kind of itchy…. Uh, $p = .80$

Person B: I have done complex calculations and have seen 10,451 days like tomorrow… $p = 0.80$

What is the difference between the two estimates?

"Those who are able to represent what they do not know make better decisions"

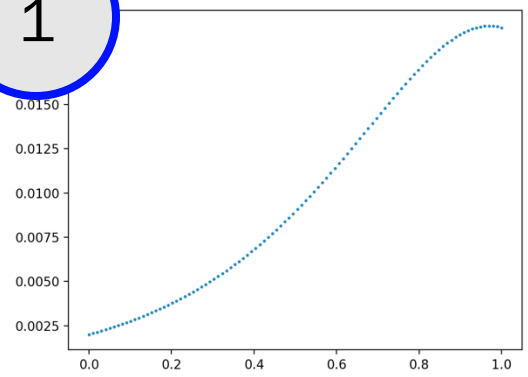
- CS109

Today we are going to learn something unintuitive, beautiful and useful

Review

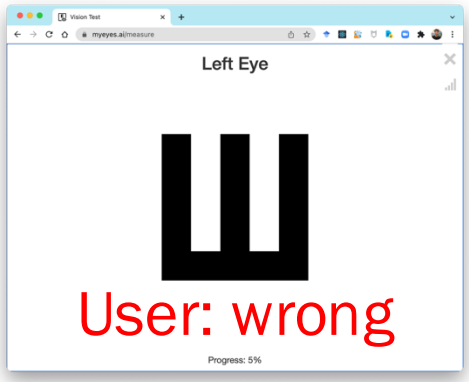
Inference on a non-bernoulli random variable

1



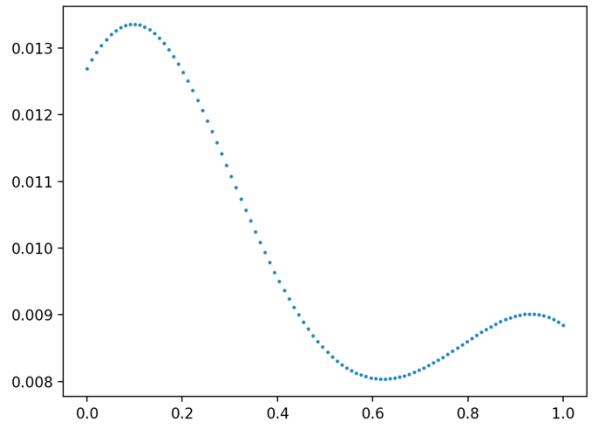
$P(A = a)$

2



Observation $Y = 0$

3



$P(A = a|Y = 0)$

We can perform **inference** when there are two random variables using Bayes!

Number or Dictionary?

belief

$$P(A = a|Y = 0) = \frac{P(Y = 0|A = a)P(A = a)}{P(Y = 0)}$$

belief

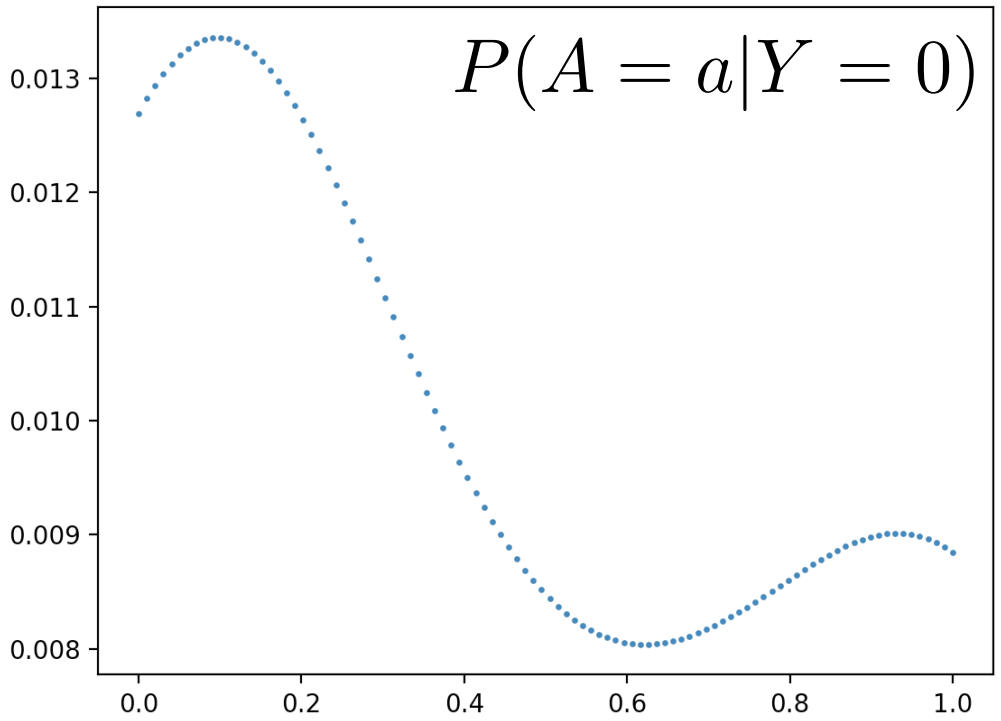
belief[0.02] = 0.001

Value of a

P(A=a)

Inference on a non-bernoulli random variable

In plain English: run bayes for each value of a




```
# RV bayes as code
def update(belief, obs):
    for a in support:
      prior_a = belief[a]
      likelihood = calc_likelihood(a, obs)
      belief[a] = prior_a * likelihood
    normalize(belief)
```

likelihood

$$P(A = a|Y = 0) = \frac{P(Y = 0|A = a)P(A = a)}{P(Y = 0)}$$

Normalize???

```
# RV bayes as code
def update(belief, obs):
    for a in support:
      prior_a = belief[a]
      likelihood = calc_likelihood(a, obs)
      belief[a] = prior_a * likelihood
    normalize(belief)
```

In plain English: this is the sum of all the things in belief

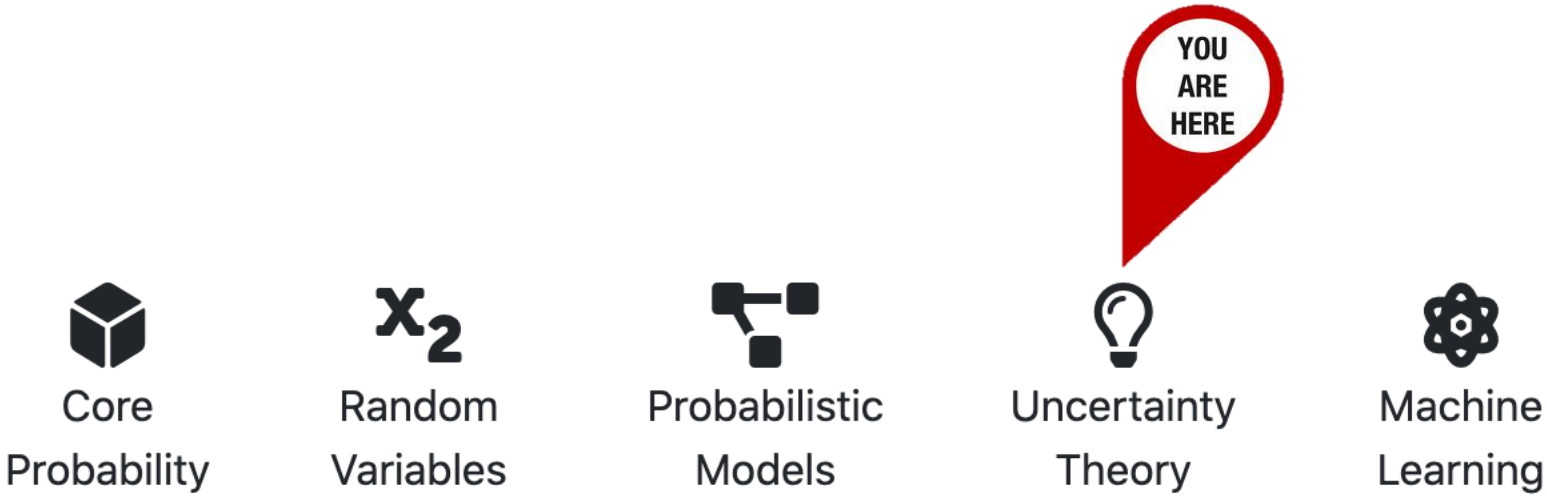
$$P(A=a|Y=0) = \frac{P(Y=0|A=a)P(A=a)}{P(Y=0)}$$

$$= \frac{P(Y=0|A=a)P(A=a)}{\sum_a P(Y=0, A=a)}$$

$$= \frac{P(Y=0|A=a)P(A=a)}{\sum_a P(Y=0|A=a)P(A=a)}$$

End Review

Where are we in CS109?



Let's play a game!

Flip a plate 5 times. If you get heads 3 times you win

Credit: Rembrandt via Dall E

$$P(X = 3) = \binom{5}{3} \cdot \frac{1}{2}^{3} \cdot \frac{1}{2}^{2}$$

$$= 0.3125$$

What if you don't know a probability?

What if you don't know a probability?

What is your belief that you flip a heads on my coin?

The parameter p to a binomial can be a random variable

9 Heads out of 10 Flips. What is your Belief in p?

$$p = \frac{9}{10}$$

9 Heads out of 10 Flips. What is your Belief in p?

Let X be our belief about the probability of heads:

Let H be our observed number of heads in 10 flips:

$$f(X = x|H = 9, T = 1)$$

$$= \frac{P(H = 9, T = 1|X = x) f(X = x)}{P(H = 9, T = 1)}$$

Binomial

Uniform?

9 Heads out of 10 Flips. What is your Belief in p?

Let X be our belief about the probability of heads:

Let H be our observed number of heads in 10 flips:

$$f(X = x|H = 9, T = 1)$$

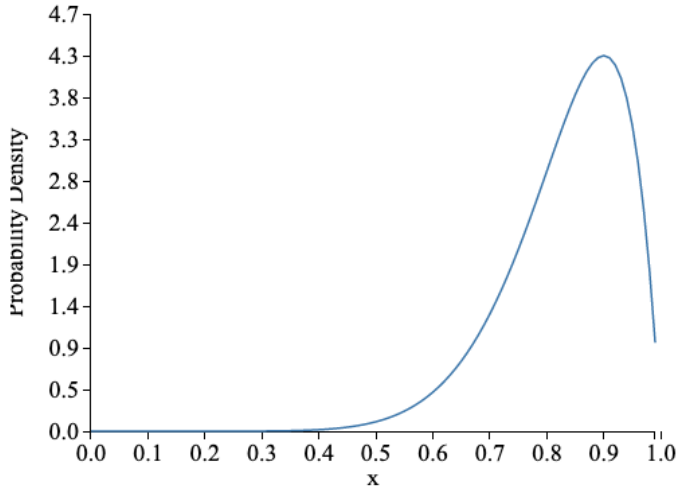
Binomial

$$= \frac{P(H = 9, T = 1|X = x)f(X = x)}{P(H = 9, T = 1)}$$

Uniform?

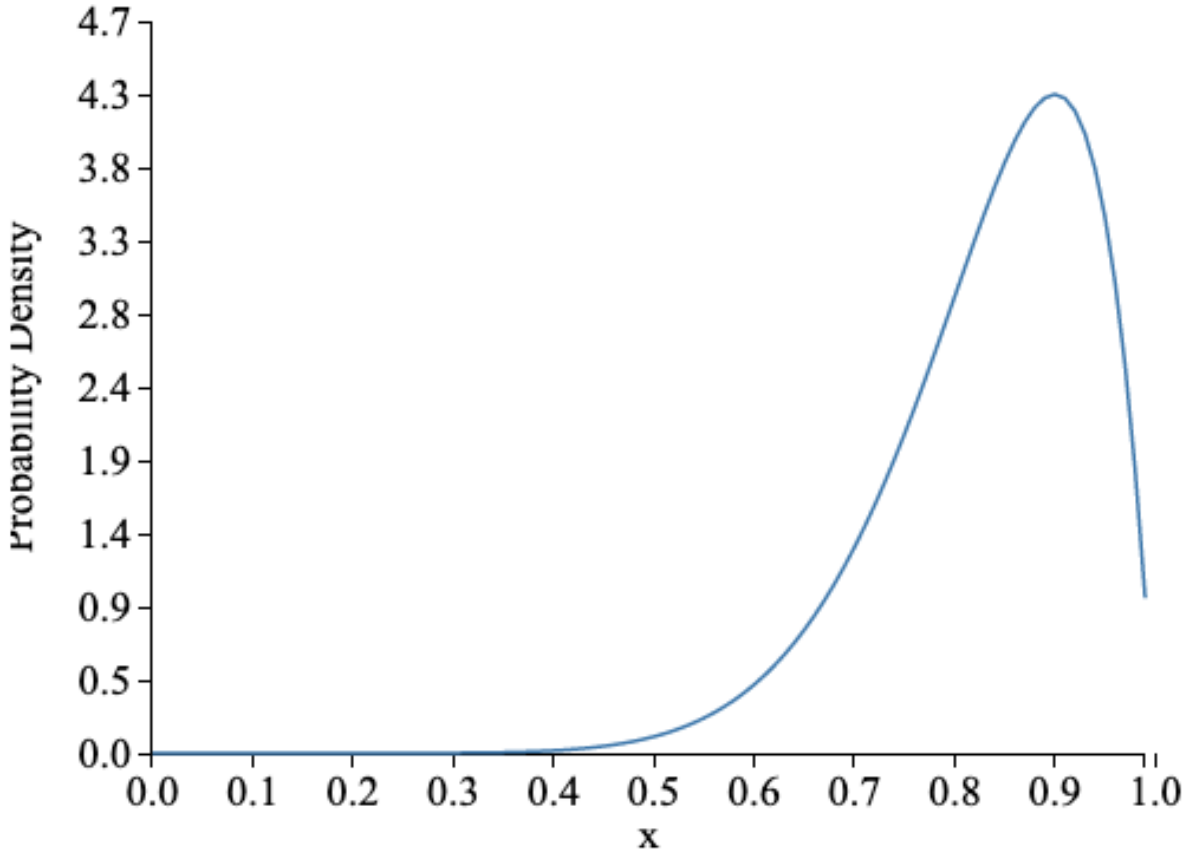
$$= \frac{\binom{10}{9}x^9(1-x)^1}{P(H = 9, T = 1)}$$

$$= K \cdot x^9(1-x)^1$$



9 Heads out of 10 Flips. What is your Belief in p?

$$f(X = x | H = 9, T = 1)$$



Two different perspectives:

Flip a coin n times, comes up with h heads (let $H = h$ be the event of flipping n times and getting h heads)

- We don't know probability X that coin comes up heads

Frequentist (never prior)

$$X = \lim_{k \to \infty} \frac{\text{count}(\text{heads in } k \text{ flips})}{k}$$
$$\approx \frac{h}{n}$$

Bayesian (prior is great)

$$f(X = x | H = h) = \frac{P(H = h | X = x) \cdot f(X = x)}{P(H = h)}$$

Flip a coin n times, comes up with h heads and t tails

- We don't know probability X that coin comes up heads
- Our belief before flipping coins is that: X ~ Uni(0, 1)
- Let H = number of heads
- Given X = x, coin flips independent:

$$f(X = x|H = h) = \frac{P(H = h|X = x) \cdot f(X = x)}{P(H = h)}$$

Bayesian "posterior" probability distribution

Bayesian "prior" probability distribution

Flip a coin n times, comes up with h heads and t tails

- We don't know probability X that coin comes up heads
- Our belief before flipping coins is that: X ~ Uni(0, 1)
- Let H = number of heads
- Given X = x, coin flips independent:

Binomial

$$f(X = x|H = h) = \frac{P(H = h|X = x) \cdot f(X = x)}{P(H = h)} \quad 1$$

$$= \frac{\binom{n}{h} x^h (1-x)^t \cdot 1}{P(H = h)}$$

$$= \frac{\binom{n}{h}}{P(H = h)} \cdot x^h (1-x)^t$$

Move terms around

$$= \frac{1}{c} \cdot x^h (1-x)^t$$

$$c = \int_0^1 x^h (1-x)^t \, \partial x$$

Flip a coin with unknown probability!

If you start with a $X \sim \text{Uni}(0, 1)$ prior over probability, and observe:

n "successes" and

m "failures"…

Your new belief about the probability is:

$$f_X(x) = \frac{1}{c} \cdot x^n (1-x)^m$$

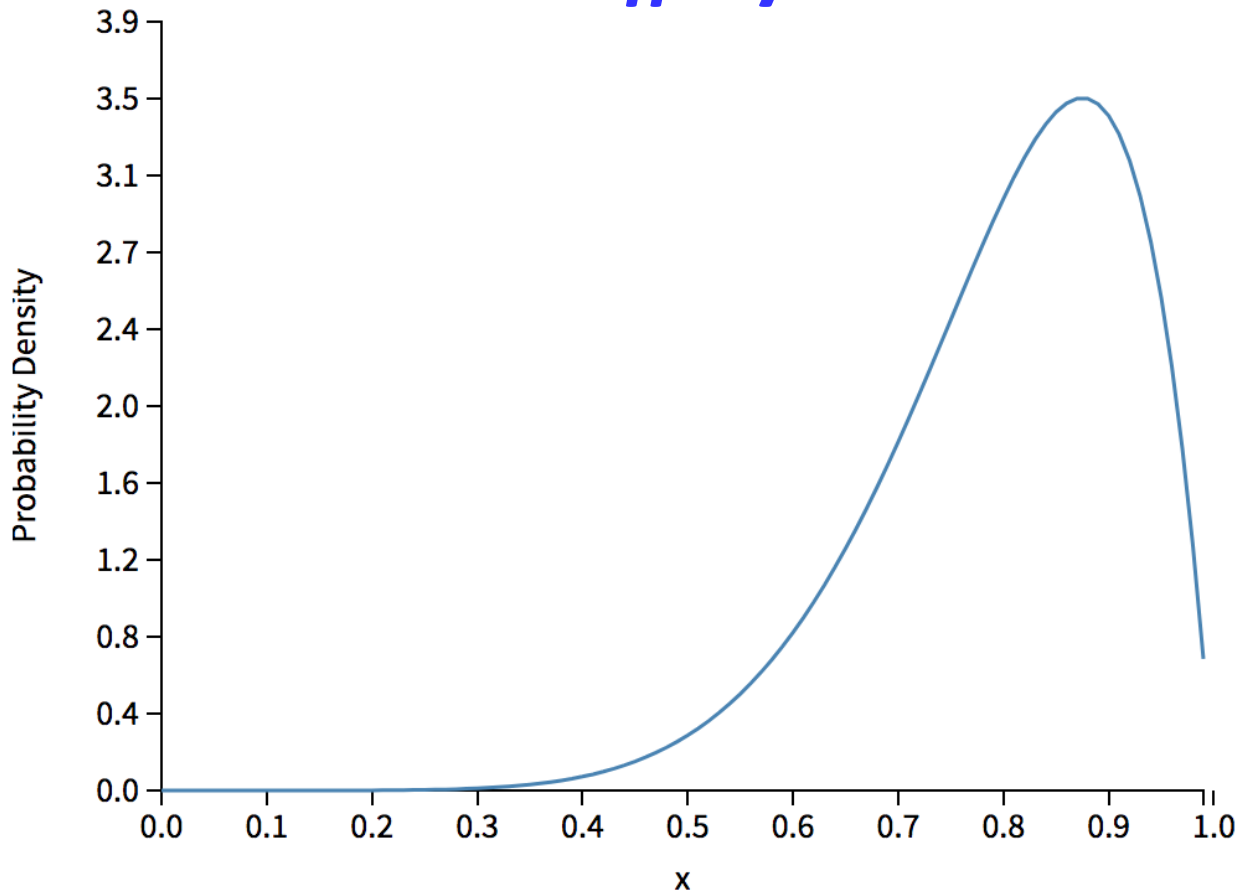
where $c = \int_0^1 x^n (1-x)^m$

Belief after 7 success and 1 fail

$$f(X = x|H = h) = \frac{1}{c} \cdot x^h(1-x)^t$$

$h = 7$

$t = 1$



Equivalently!

If you start with a $X \sim \text{Uni}(0, 1)$ prior over probability, and observe:

let a = num "successes" + 1

let b = num "failures" + 1

Your new belief about the probability is:

$$f_X(x) = \frac{1}{c} \cdot x^{a-1}(1-x)^{b-1}$$

where $$c = \int_0^1 x^{a-1}(1-x)^{b-1}$$

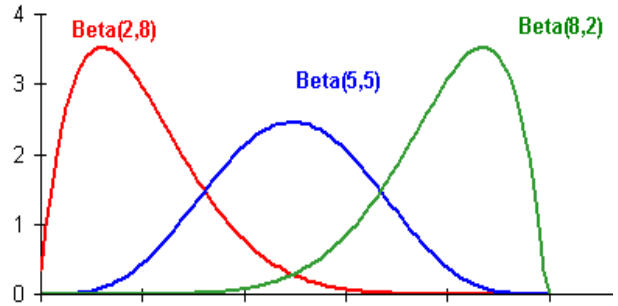
Beta Random Variable

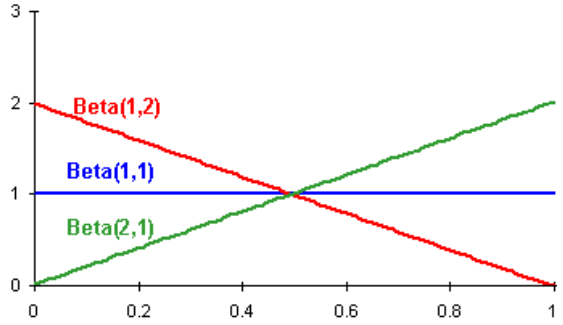
X is a **Beta Random Variable**: X ~ Beta(a, b)

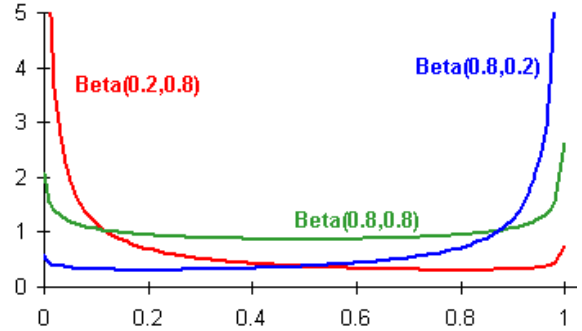
- Probability Density Function (PDF): (where $a, b > 0$)

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$$



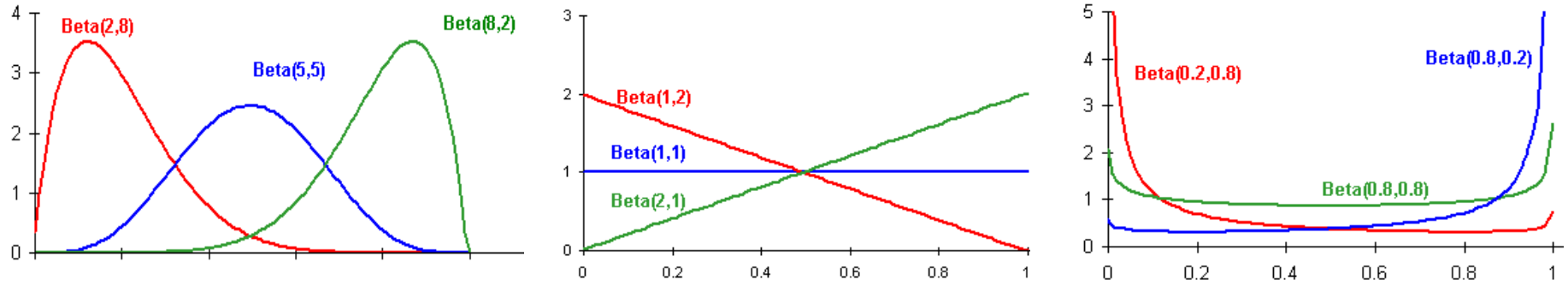




- Symmetric when $a = b$

$$E[X] = \frac{a}{a+b} \qquad Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Beta is the Random Variable for Probabilities

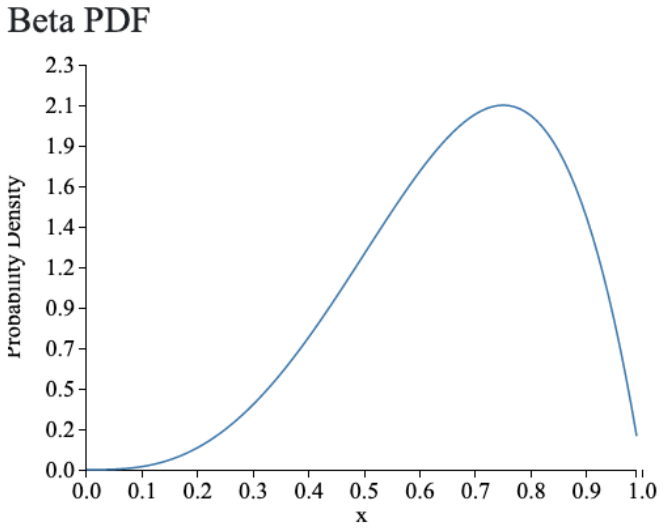


Used to represent a distributed belief of a probability

Think about the difference between a **point estimate** and a **distribution**

$p = 0.75$

$p =$



Beta is a distribution for probabilities. Its range is values between 0 and 1

Beta Parameters *can* come from experiments:

a = “successes” + 1

b = “failures” + 1

If the Prior was Beta?

X is our random variable for probability

If our prior belief about X was beta

$$f(X = x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

What is our posterior belief about X after observing *n* heads (and *m* tails)?

$$f(X = x|N = n) = ???$$

If the Prior was Beta?

$$
\begin{aligned}
f(X = x|N = n) &= \frac{P(N = n|X = x)f(X = x)}{P(N = n)} \\
&= \frac{\binom{n+m}{n} x^n (1-x)^m \, f(X = x)}{P(N = n)} \\
&= \frac{\binom{n+m}{n} x^n (1-x)^m \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}}{P(N = n)} \\
&= K_1 \cdot \binom{n+m}{n} x^n (1-x)^m \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \\
&= K_3 \cdot x^n (1-x)^m x^{a-1}(1-x)^{b-1} \\
&= K_3 \cdot x^{n+a-1}(1-x)^{m+b-1} \\
X|N &\sim \text{Beta}(n+a, m+b)
\end{aligned}
$$

A beta understanding

- If "Prior" distribution of X (before seeing flips) is Beta

- Then "Posterior" distribution of X (after flips) is Beta

Beta is a **conjugate** distribution for Beta

- Prior and posterior parametric forms are the same!
- Practically, conjugate means easy update:
  - Add number of "heads" and "tails" seen to Beta parameters

Laplace Smoothing

One imagined heads

Prior: $X \sim \text{Beta}(a = 2, b = 2)$

One imagined tail

Fancy name. Simple prior

Check this out, Boss

- Beta$(a = 1, b = 1)$ =?

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} = \frac{1}{B(a,b)} x^{0}(1-x)^{0}$$

$$= \frac{1}{\int_0^1 1\,dx} 1 = 1 \quad \text{where} \quad 0 < x < 1$$

- Beta$(a = 1, b = 1)$ = Uni$(0, 1)$

Mystery Plate

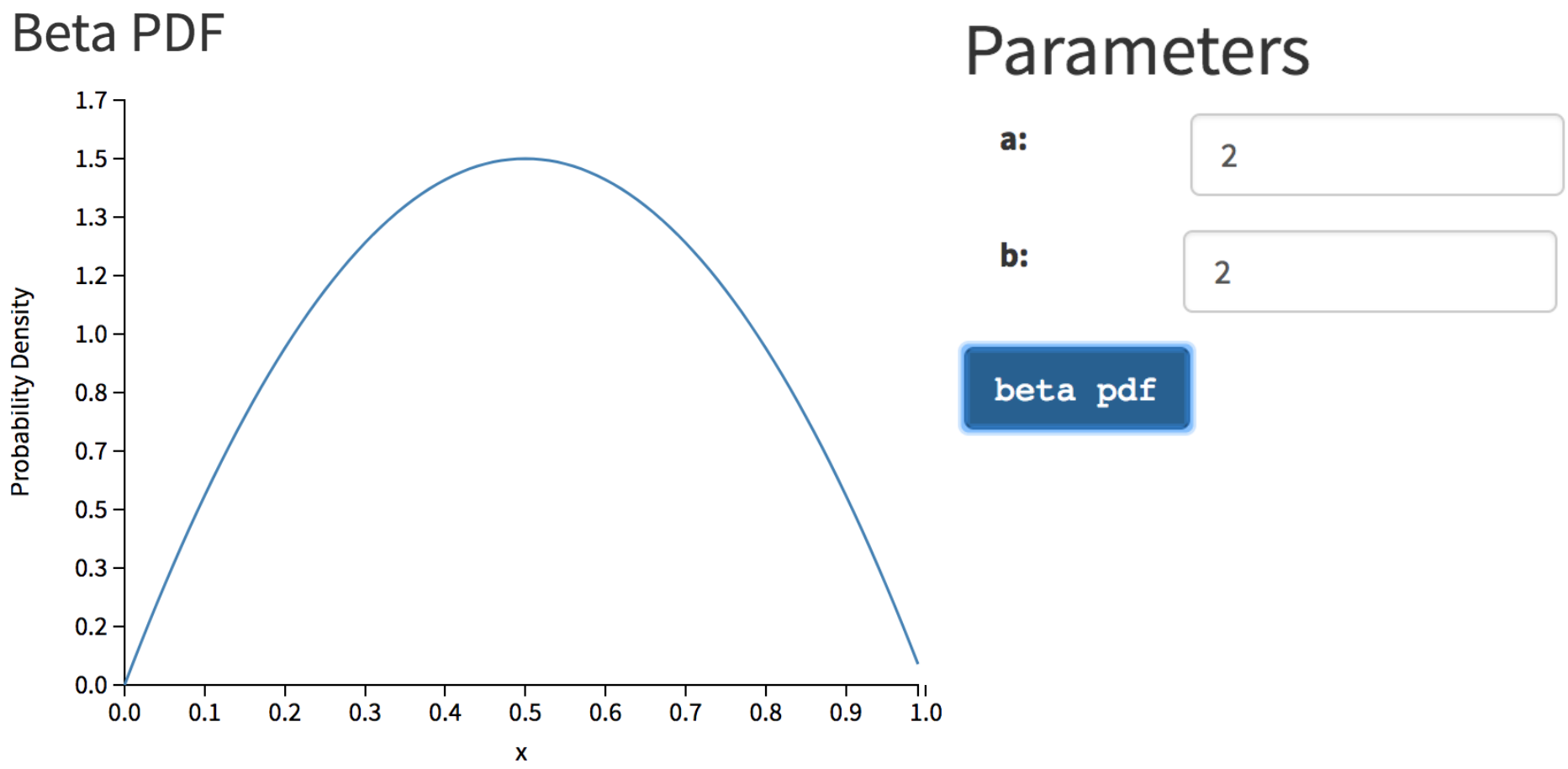
Let X be the probability of getting a heads when flipping a plate.

Prior: Imagine 5 coin flips that were heads

Observation: Flip it a few times…

What is the updated probability density function of X after our observations?

Check out the Demo!



Damn

A beta example

Before being tested, a medicine is believed to “work” about 80% of the time. The medicine is tried on 20 patients. It “works” for 14 and “doesn’t work” for 6. What is your new belief that the drug works?

Frequentist:

$$p \approx \frac{14}{20} = 0.7$$

A beta example

Before being tested, a medicine is believed to "work" about 80% of the time. The medicine is tried on 20 patients. It "works" for 14 and "doesn't work" for 6. What is your new belief that the drug works?

Bayesian: $X \sim \text{Beta}$

Prior:	Interpretation:
$X \sim \text{Beta}(a = 81, b = 21)$	80 successes / 100 trials
$X \sim \text{Beta}(a = 9, b = 3)$	8 successes / 10 trials
$X \sim \text{Beta}(a = 5, b = 2)$	4 successes / 5 trials

A beta example

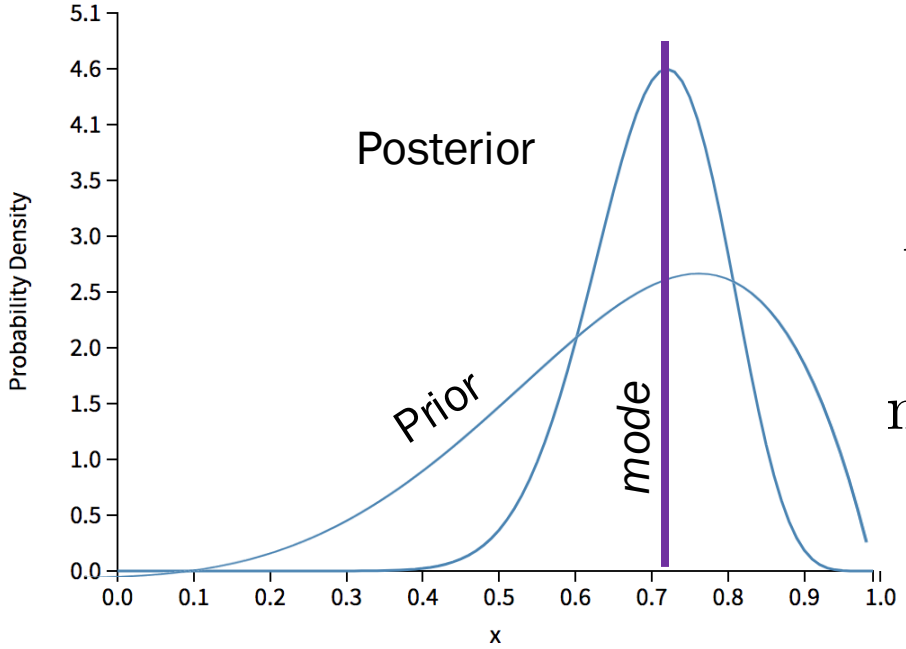
Before being tested, a medicine is believed to "work" about 80% of the time. The medicine is tried on 20 patients. It "works" for 14 and "doesn't work" for 6. What is your new belief that the drug works?

Bayesian: $X \sim \text{Beta}$

Prior: $X \sim \text{Beta}(a = 5, b = 2)$

Posterior: $X \sim \text{Beta}(a = 5 + 14, b = 2 + 6)$

$\sim \text{Beta}(a = 19, b = 8)$



$$E[X] = \frac{a}{a+b} = \frac{19}{19+8} \approx 0.70$$

$$\text{mode}(X) = \frac{a-1}{a+b-2}$$

$$= \frac{19}{18+7} \approx 0.72$$

Which video are you more likely to like?

10,000 50

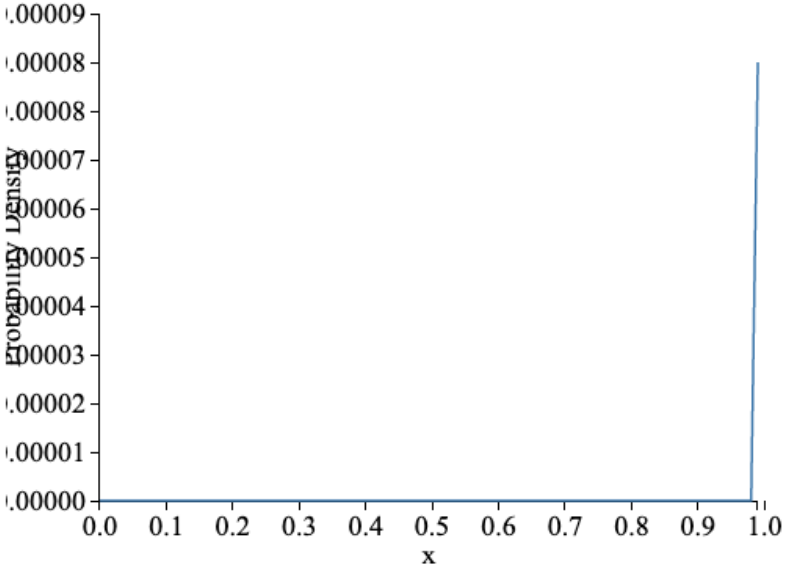
10 0

Which video are you more likely to like?

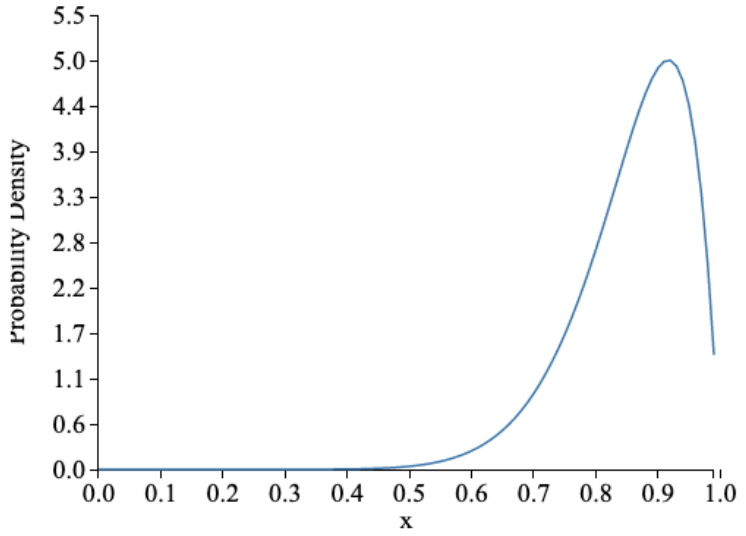
10,000 50

10 0

Beta PDF (Using Laplace prior)



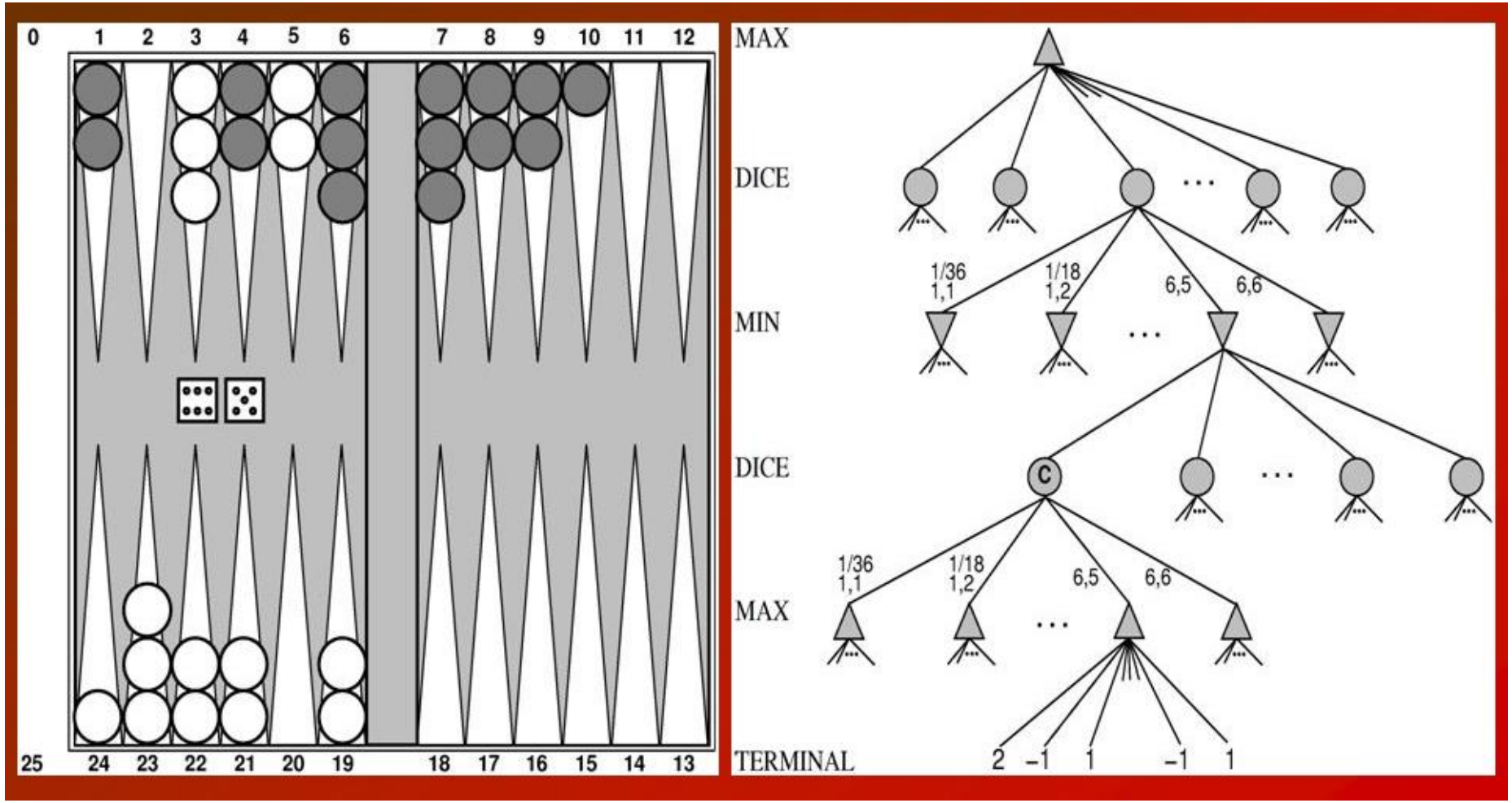
Beta PDF (Using Laplace prior)



Next level?

Alpha GO mixed deep learning and core reasoning under uncertainty

Multi Armed Bandit



Multi Armed Bandit

Drug A

Drug B

Which one do you give to a patient?

Lets Play!

Drug A

Drug B

Which one do you give to a patient?

Lets Play!

sim.py

```
1  import pickle
2  import random
3
4  def main():
5    X1, X2  = pickle.load(open('probs.pkl', 'rb'))
6
7    print("Welcome to the drug simulator. There are two drugs")
8
9    while True:
10     choice = getChoice()
11     prob = X1 if choice == "a" else X2
12     success = bernoulli(prob)
13     if success:
14       print('Success. Patient lives!')
15     else:
16       print('Failure. Patient dies!')
17     print('')
18
```

Optimal Decision Making

You try drug B, 5 times. It is successful 2 times.

If you had a uniform prior, what is your posterior belief about the likelihood of success?

---

2 successes

3 failures

$$X \sim \text{Beta}(a = 3, b = 4)$$

Optimal Decision Making

You try drug B, 5 times. It is successful 2 times.
X is the probability of success.

$$X \sim \text{Beta}(a = 3, b = 4)$$

What is expectation of X?

$$E[X] = \frac{a}{a+b} = \frac{3}{3+4} \approx 0.43$$

Optimal Decision Making

You try drug B, 5 times. It is successful 2 times.
X is the probability of success.

$$X \sim \text{Beta}(a = 3, b = 4)$$

What is the probability that $X > 0.6$

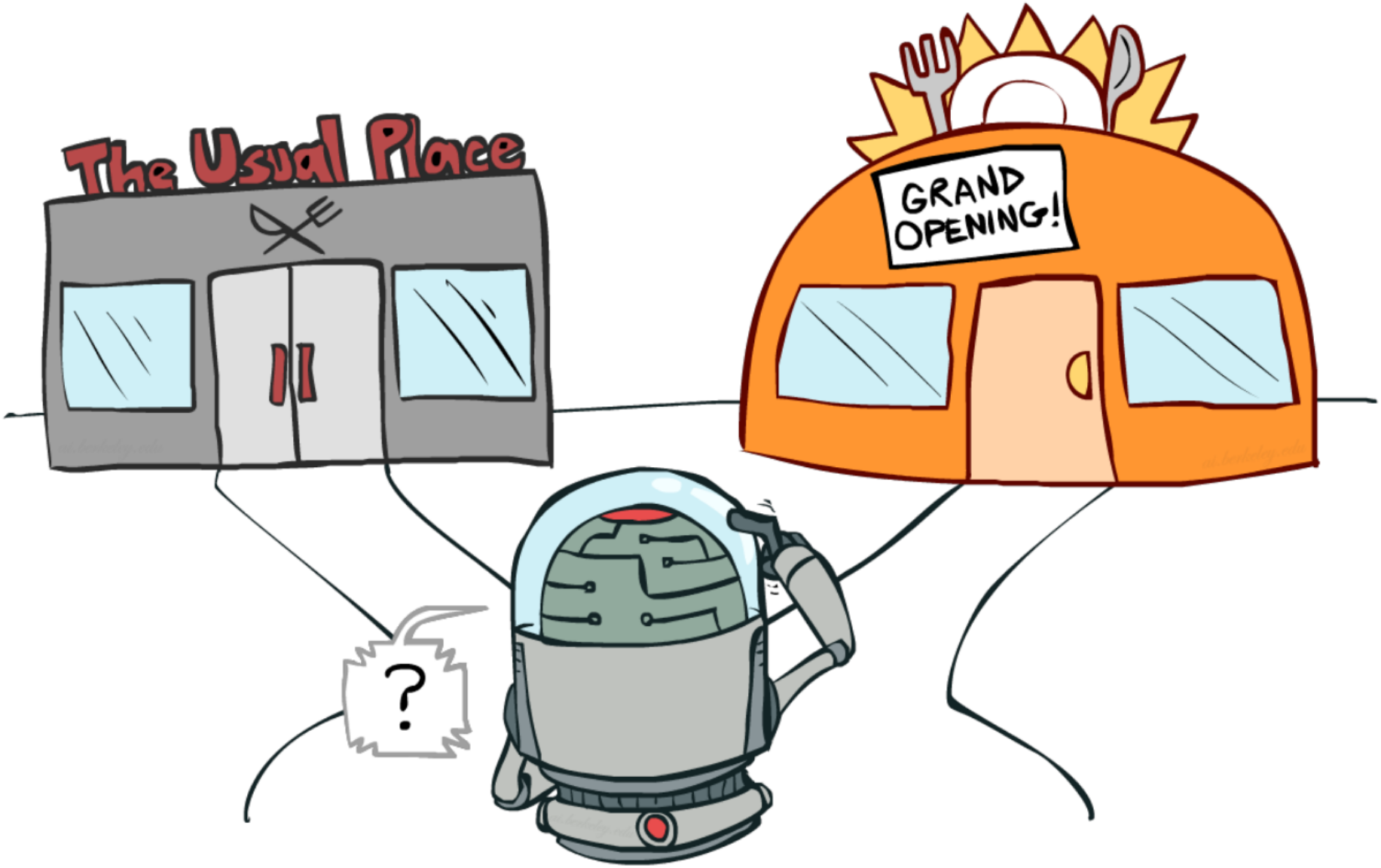
$$P(X > 0.6) = 1 - P(X < 0.6) = 1 - F_X(0.6)$$

Wait what? Chris are you holding out on me?

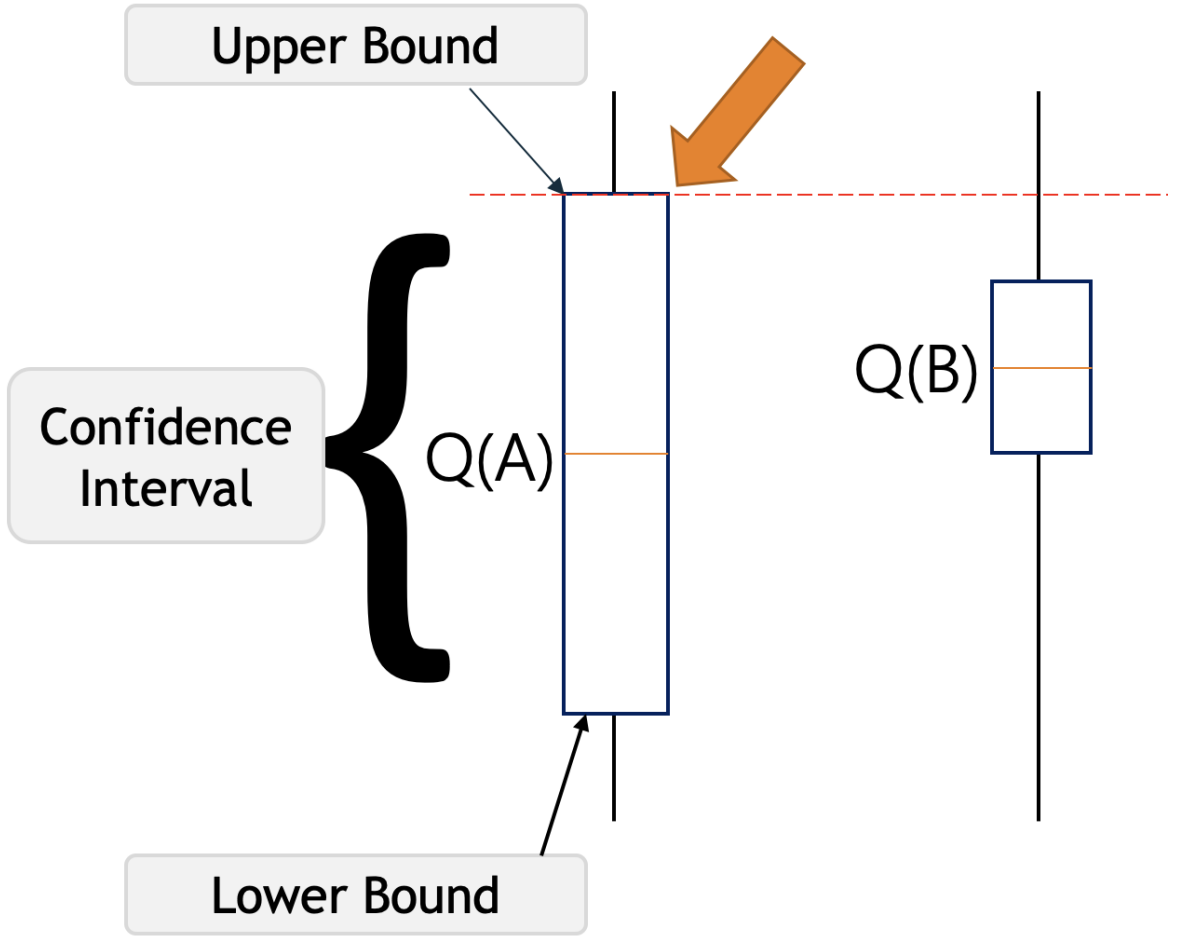
```
stats.beta.cdf(x, a, b)
```

$$P(X > 0.6) = 1 - F_X(0.6) = 0.1792$$

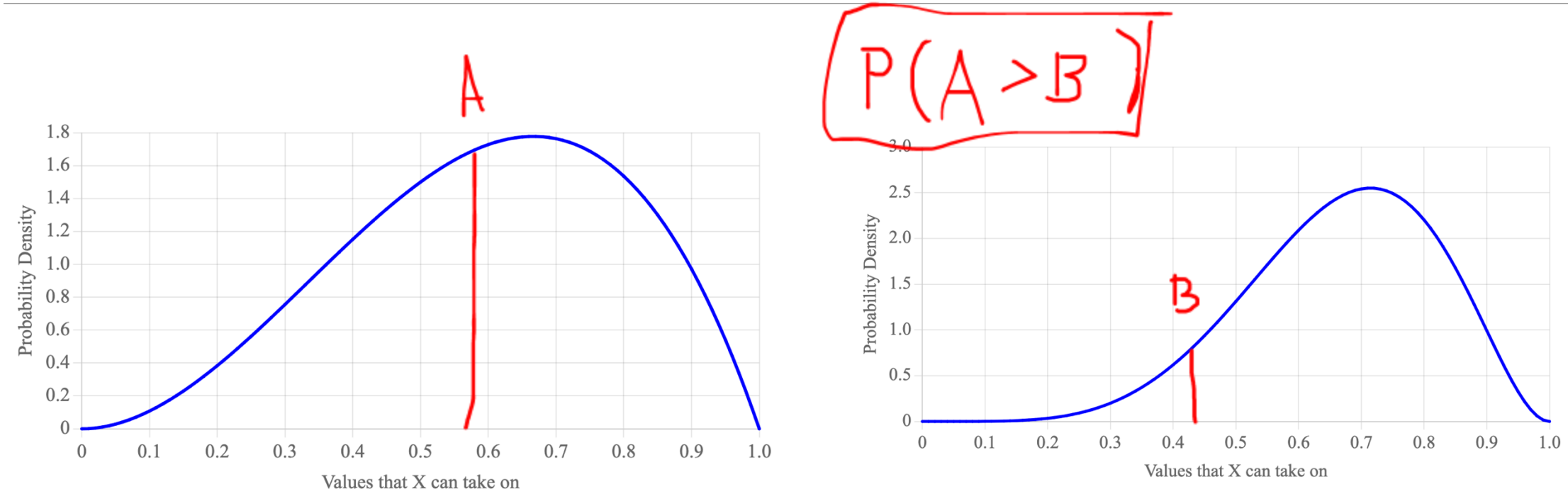
Explore something new? Or go for what looks good now?



One option: Upper Confidence Bound



Amazing option: Thompson Sampling



1. Chose a sample from each drug's beta

   A = .58 B = .42

2. Select the drug with the higher sample

   A

Beta:
The probability density for probabilities

Beta is a distribution for probabilities

Beta Distribution

If you start with a $X \sim \text{Uni}(0, 1)$ prior over probability, and observe:

let a = num "successes" + 1

let b = num "failures" + 1

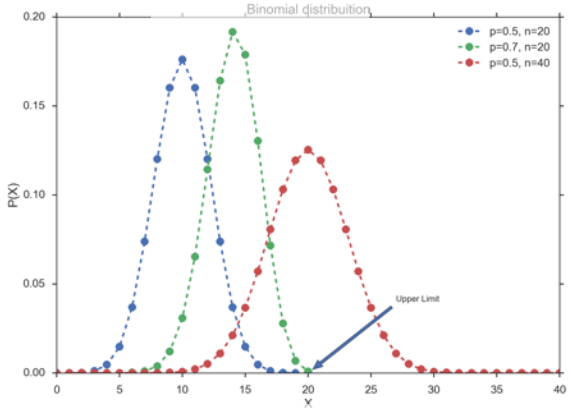
Your new belief about the probability is:

$$f_X(x) = \frac{1}{c} \cdot x^{a-1}(1-x)^{b-1}$$

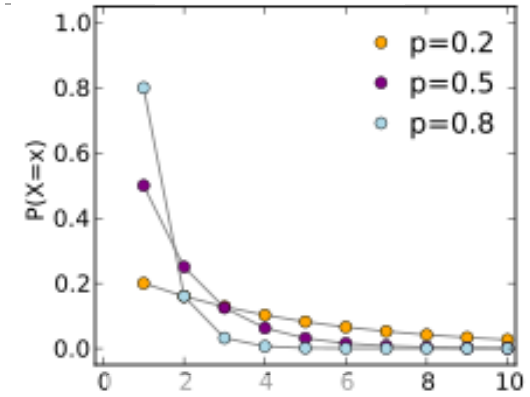
$$\text{where} \quad c = \int_0^1 x^{a-1}(1-x)^{b-1}$$

Distributions

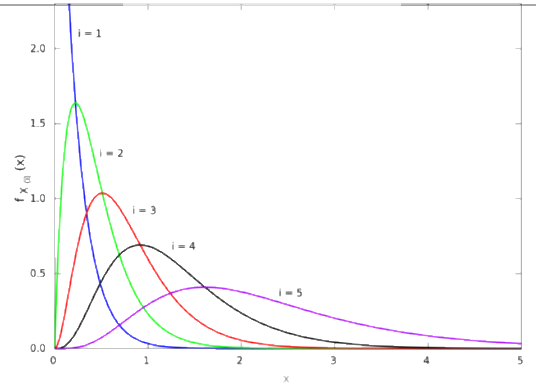
Binomial



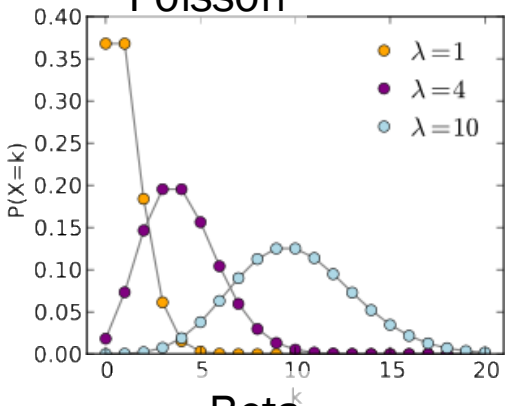
Geometric



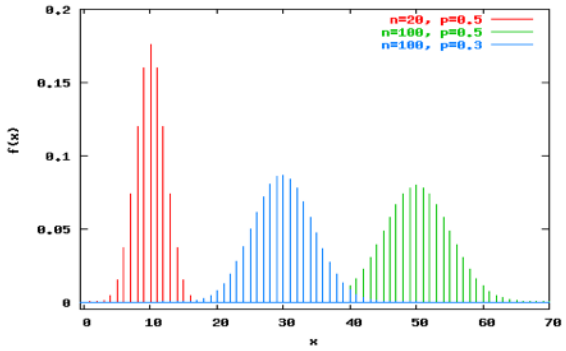
Exponential



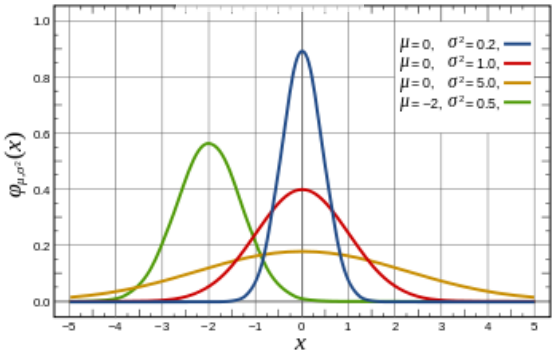
Poisson



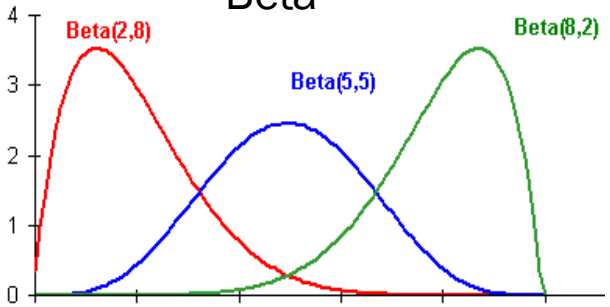
Neg Binomial



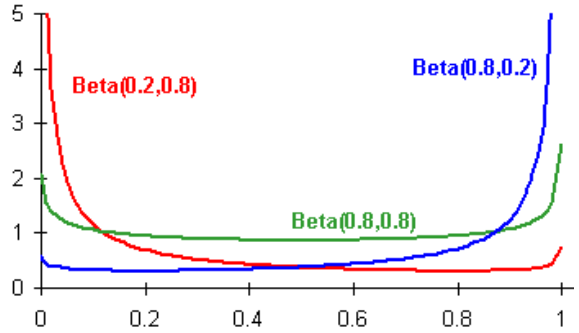
Normal



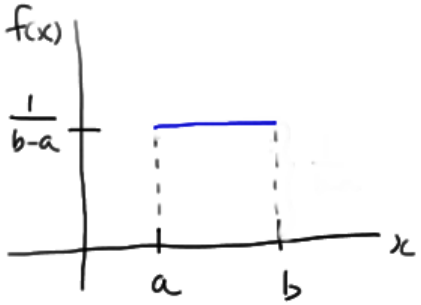
Beta



Beta



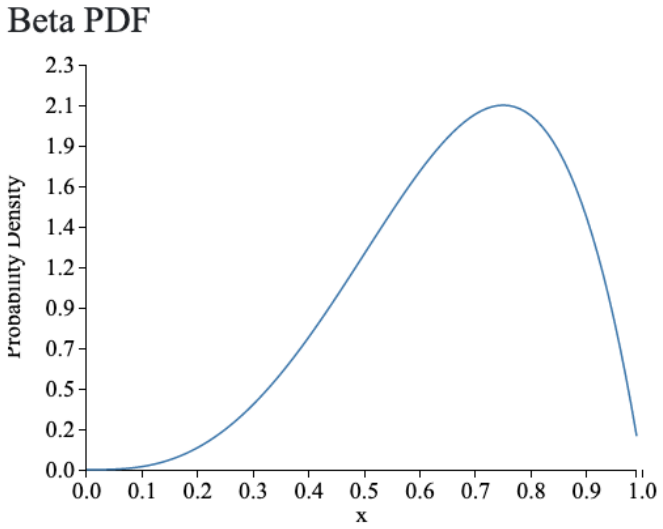
Uniform



Think about the difference between a **point estimate** and a **distribution**

$p = 0.75$

$p =$



Problem with a point estimate:

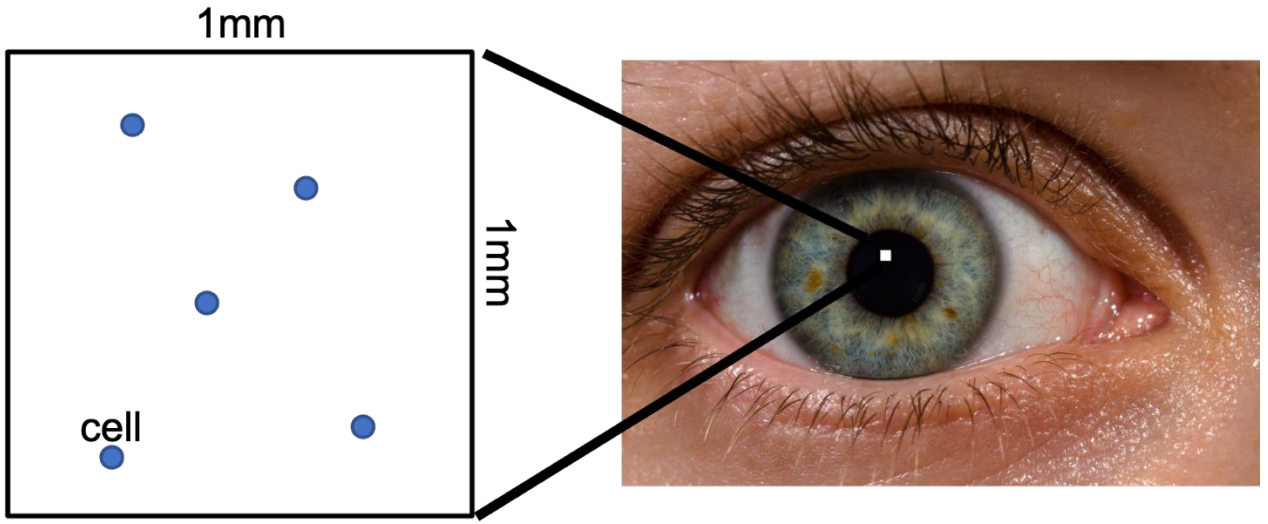
Person A: My leg itches when it rains and its kind of itchy…. Uh, p = .80

Person B: I have done complex calculations and have seen 10,451 days like tomorrow… p = 0.80

Give me the uncertainty!!!

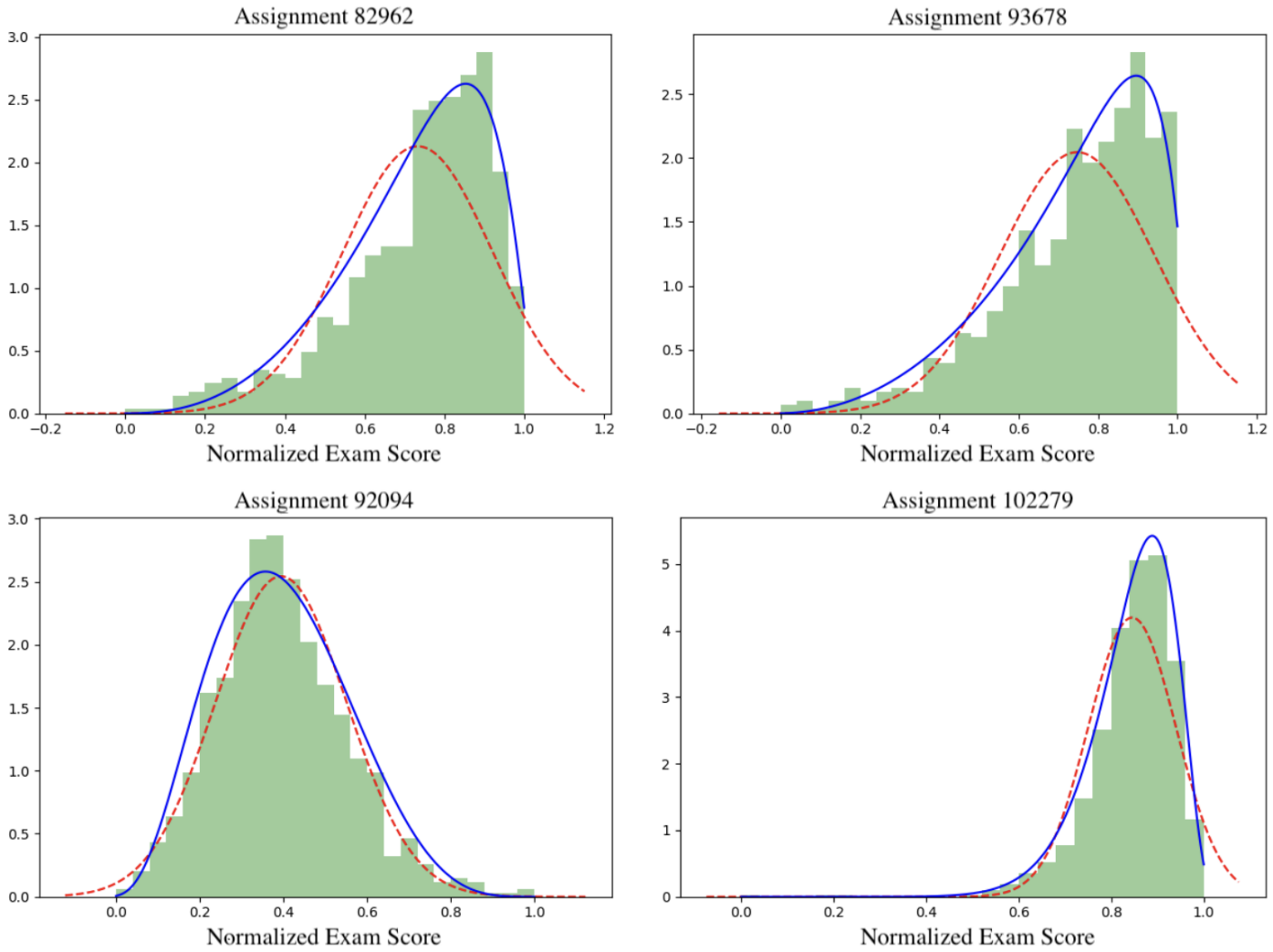
Any parameter for a “parameterized” random variable can be thought of as a random variable.

Eg:

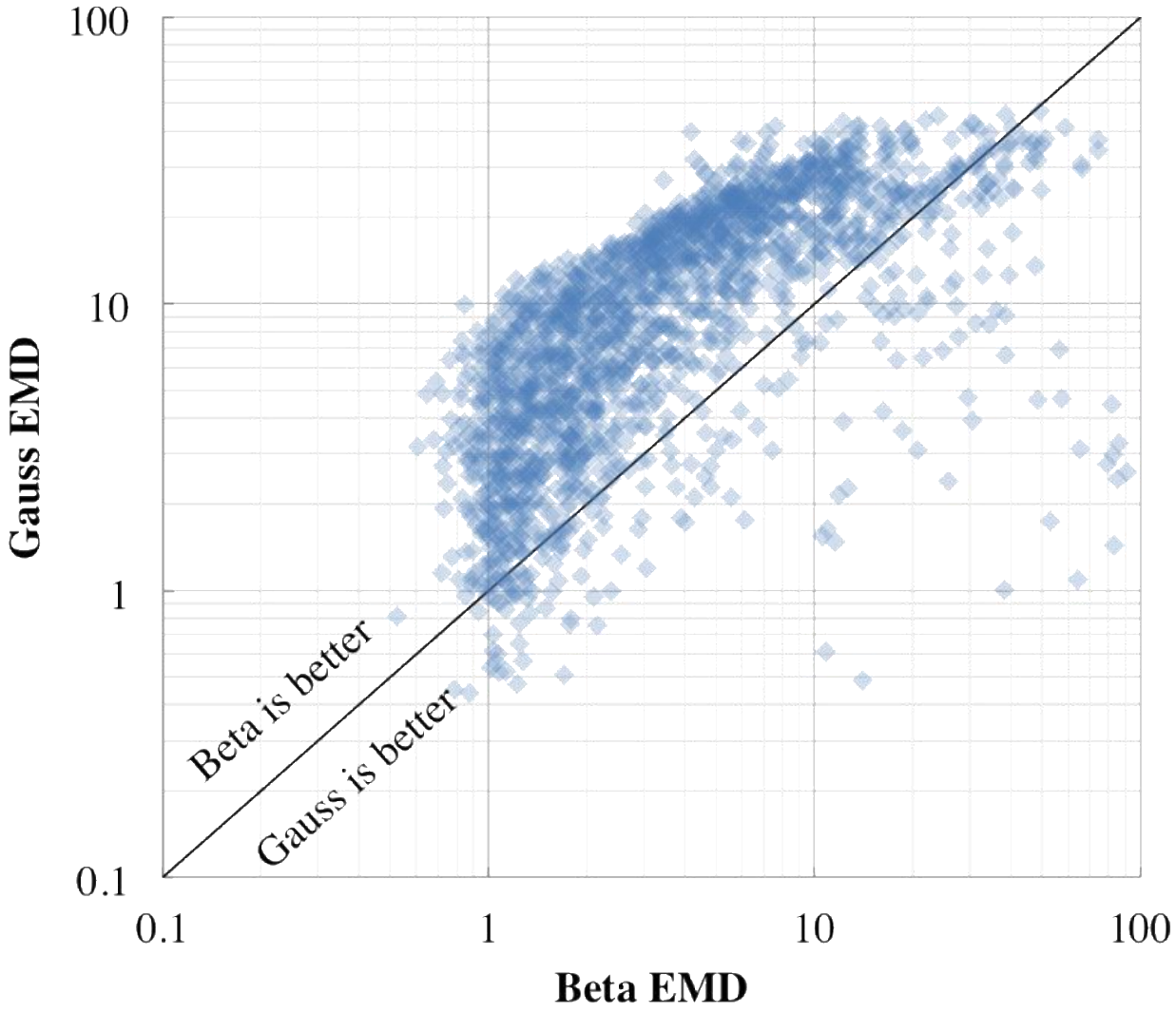


$$P(\Lambda = \lambda | N = 5)$$

Grades are Not Normal



Grades are Not Normal



Based on 4000 classes on Gradescope