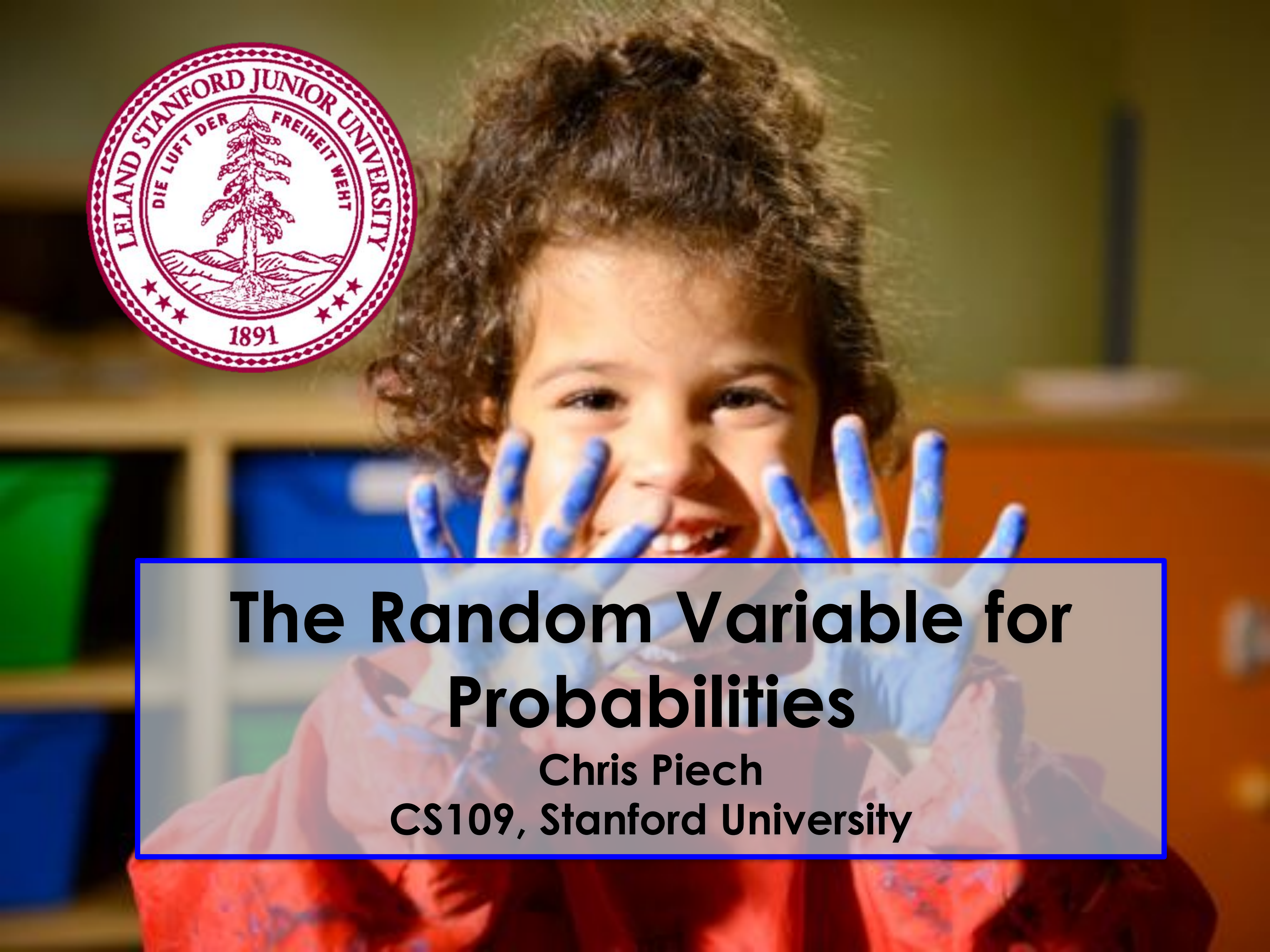


The Random Variable for Probabilities

Chris Piech
CS109, Stanford University

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

- CS109

Finishing up Expectation

Conditional Expectation Functions

This is a number:

$$E[X]$$

This is a function of rv Y:

$$E[X|Y = y]$$

$E[X = 5]$ Doesn't make sense. Take expectation of random variables, not events

Law of Total Expectation

For any random variable X and any **discrete** random variable Y

$$E[X] = \sum_y E[X|Y=y]P(Y=y)$$

Analyzing Recursive Code

```
int Recurse() {
   int x = randomInt(1, 3);  // Equally likely values

   if (x == 1) return 3;
   else if (x == 2) return (5 + Recurse());
   else return (7 + Recurse());
}
```

- Let Y = value returned by `Recurse()`. What is E[Y]?

$$E[Y] = E[Y \mid X=1]P(X=1) + E[Y \mid X=2]P(X=2) + E[Y \mid X=3]P(X=3)$$

$$E[Y \mid X=1] = 3$$

$$E[Y \mid X=2] = E[5+Y] = 5 + E[Y]$$

$$E[Y \mid X=3] = E[7+Y] = 7 + E[Y]$$

$$E[Y] = 3(1/3) + (5+E[Y])(1/3) + (7+E[Y])(1/3) = (1/3)(15 + 2E[Y])$$

$$E[Y] = 15$$

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

Review

Conditioning with a continuous random variable is odd at first. But then it gets fun.

Its like snorkeling…

Continuous Conditional Distributions

- Let X be continuous random variable
- Let E be an event:

$$
\begin{aligned}
P(E|X=x) &= \frac{P(X=x,E)}{P(X=x)} \\
&= \frac{P(X=x|E)P(E)}{P(X=x)} \\
&= \frac{f_X(x|E)P(E)\epsilon_x}{f_X(x)\epsilon_x} \\
&= \frac{f_X(x|E)P(E)}{f_X(x)}
\end{aligned}
$$

Continuous Conditional Distributions

- Let X be a measure of time to answer a question
- Let E be the event that the user is a human:

$$
\begin{aligned}
P(E|X = x) &= \frac{P(X = x, E)}{P(X = x)} \\
&= \frac{P(X = x|E)P(E)}{P(X = x)} \\
&= \frac{f_X(x|E)P(E)\epsilon_x}{f_X(x)\epsilon_x} \\
&= \frac{f_X(x|E)P(E)}{f_X(x)}
\end{aligned}
$$

Biometric Keystroke

- Let X be a measure of time to answer a question
- Let E be the event that the user is a human
- What if you don't know normalization term?:

Normal pdf

Prior

$$P(E|X = x) = \frac{f_X(x|E)P(E)}{f_X(x)}$$

???

$$\frac{P(E|X = x)}{P(E^C|X = x)}$$

End Review

Lets play a game

Roll a dice twice. If either time you roll a 6, I win.
Otherwise you win.

$$P(W) = \left(\frac{5}{6}\right)^2 \approx 0.69$$

Flip a Coin With Unknown Probability

Demo

We are going to think of probabilities as random variables!!!

Flip a Coin With Unknown Probability

- Flip a coin (n + m) times, comes up with n heads
  - We don't know probability X that coin comes up heads

Frequentist

$$X = \lim_{n+m \to \infty} \frac{n}{n+m} \approx \frac{n}{n+m}$$

X is a single value

Bayesian

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

X is a random variable

What is your belief that you successfully roll a 6 on my die?

Flip a Coin With Unknown Probability

- Flip a coin (n + m) times, comes up with n heads
  - We don't know probability X that coin comes up heads
  - Our belief before flipping coins is that: X ~ Uni(0, 1)
  - Let N = number of heads
  - Given X = x, coin flips independent: $(N \mid X) \sim \text{Bin}(n+m, x)$

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

Bayesian "posterior" probability distribution

Bayesian "prior" probability distribution

Flip a Coin With Unknown Probability

- Flip a coin (n + m) times, comes up with n heads
  - We don't know probability X that coin comes up heads
  - Our belief before flipping coins is that: X ~ Uni(0, 1)
  - Let N = number of heads
  - Given X = x, coin flips independent: (N | X) ~ Bin(n + m, x)

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

Binomial 1

$$= \frac{\binom{n+m}{n} x^n (1-x)^m}{P(N=n)}$$

$$= \frac{\binom{n+m}{n}}{P(N=n)} x^n (1-x)^m$$

Move terms around

$$= \frac{1}{c} \cdot x^n (1-x)^m \qquad \text{where } c = \int_0^1 x^n (1-x)^m dx$$

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:

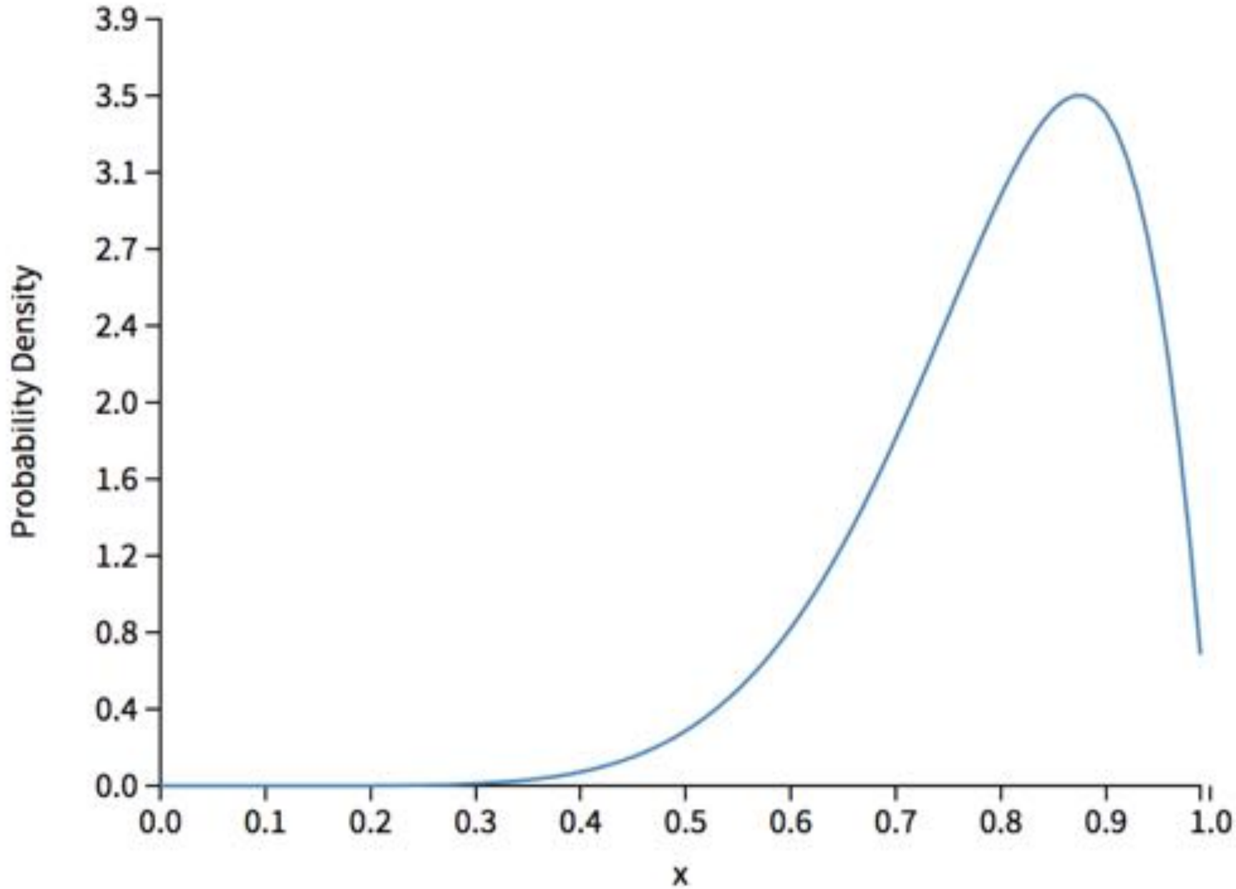
$$f_X(x) = \frac{1}{c} \cdot x^n (1 - x)^m$$

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

Belief after 7 "success" and 1 "fail"

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

$n = 7$ $m = 1$



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:

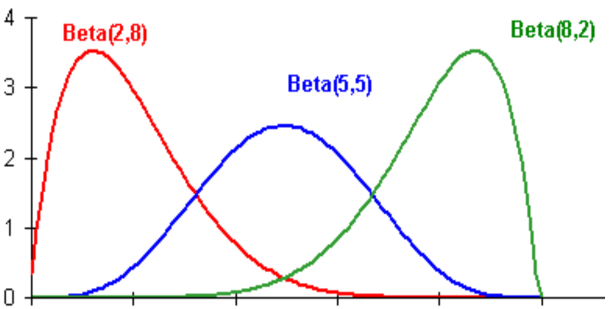
$$f_X(x) = \frac{1}{c} \cdot x^{a-1}(1-x)^{b-1}$$

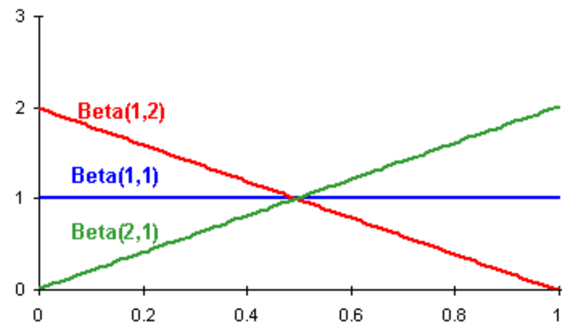
$$\text{where} \quad c = \int_0^1 x^{a-1}(1-x)^{b-1}$$

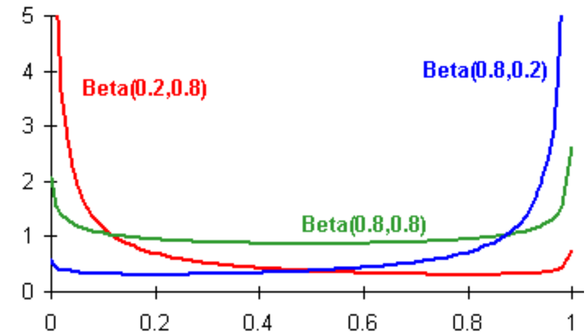
Beta Random Variable

- X is a **Beta Random Variable**: X ~ Beta(*a*, *b*)
  - Probability Density Function (PDF): (where *a*, *b* > 0)

$$f(x)=\begin{cases}\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1} & 0<x<1\\ 0 & \text{otherwise}\end{cases}\quad \text{where}\quad B(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}dx$$

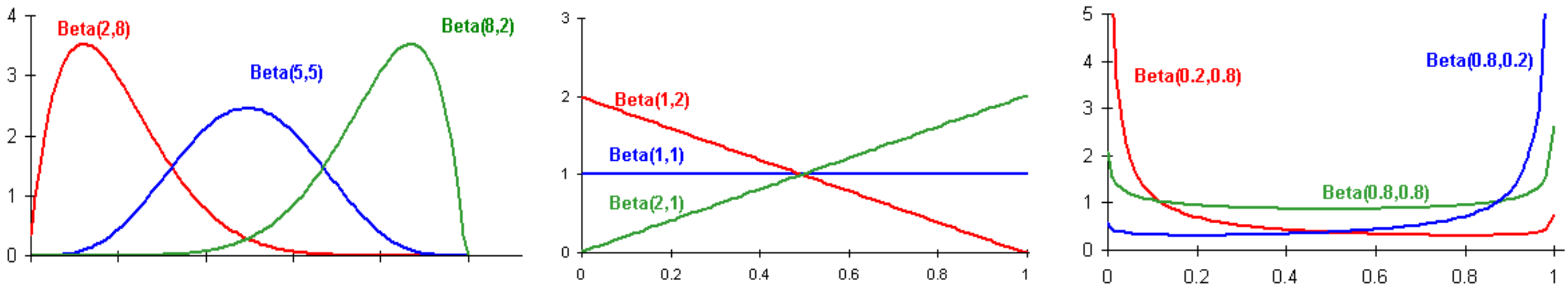






  - Symmetric when *a* = *b*
  - $E[X]=\dfrac{a}{a+b}$ $\quad Var(X)=\dfrac{ab}{(a+b)^2(a+b+1)}$

Meta Beta



Used to represent a
distributed belief of a probability

Beta is a distribution for probabilities

Beta Parameters *can* come from experiments:

a = "successes" + 1

b = "failures" + 1

Back to flipping coins

- Flip a coin (n + m) times, comes up with n heads
  - We don't know probability X that coin comes up heads
  - Our belief before flipping coins is that: X ~ Uni(0, 1)
  - Let N = number of heads
  - Given X = x, coin flips independent: $(N \mid X) \sim \text{Bin}(n+m, x)$

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

$$= \frac{\binom{n+m}{n} x^n (1-x)^m}{P(N=n)}$$

$$= \frac{\binom{n+m}{n}}{P(N=n)} x^n (1-x)^m$$

$$= \frac{1}{c} \cdot x^n (1-x)^m \qquad \text{where } c = \int_0^1 x^n (1-x)^m dx$$

Understanding Beta

- X | (N = n, M = m) ~ Beta(a = n + 1, b = m + 1)
  - Prior X ~ Uni(0, 1)
  - 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)
  - So, prior X ~ Beta(a = 1, b = 1)

N successes

M failures

If the Prior was a 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 a Beta...

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

$$= \frac{\binom{n+m}{n} x^n (1 - x)^m f(X = x)}{P(N = n)}$$

$$= \frac{\binom{n+m}{n} x^n (1 - x)^m \frac{1}{B(a,b)} x^{a-1} (1 - x)^{b-1}}{P(N = n)}$$

$$= K_1 \cdot \binom{n+m}{n} x^n (1 - x)^m \frac{1}{B(a, b)} x^{a-1} (1 - x)^{b-1}$$

$$= K_3 \cdot x^n (1 - x)^m x^{a-1} (1 - x)^{b-1}$$

$$= K_3 \cdot x^{n+a-1} (1 - x)^{m+b-1}$$

$$X|N \sim \text{Beta}(n + a, m + b)$$

Understanding Beta

- 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

Further Understanding Beta

- Can set X ~ Beta(*a*, *b*) as prior to reflect how biased you think coin is apriori
  - This is a subjective probability!
  - Prior probability for X based on seeing (*a* + *b* – 2) “imaginary” trials, where

    (*a* – 1) of them were heads.

    (b – 1) of them were tails.
  - Beta(1, 1) = Uni(0, 1) → we haven’t seen any “imaginary trials”, so apriori know nothing about coin
- Update to get posterior probability
  - X | (n heads and m tails) ~ Beta(a + n, b + m)

Enchanted Die

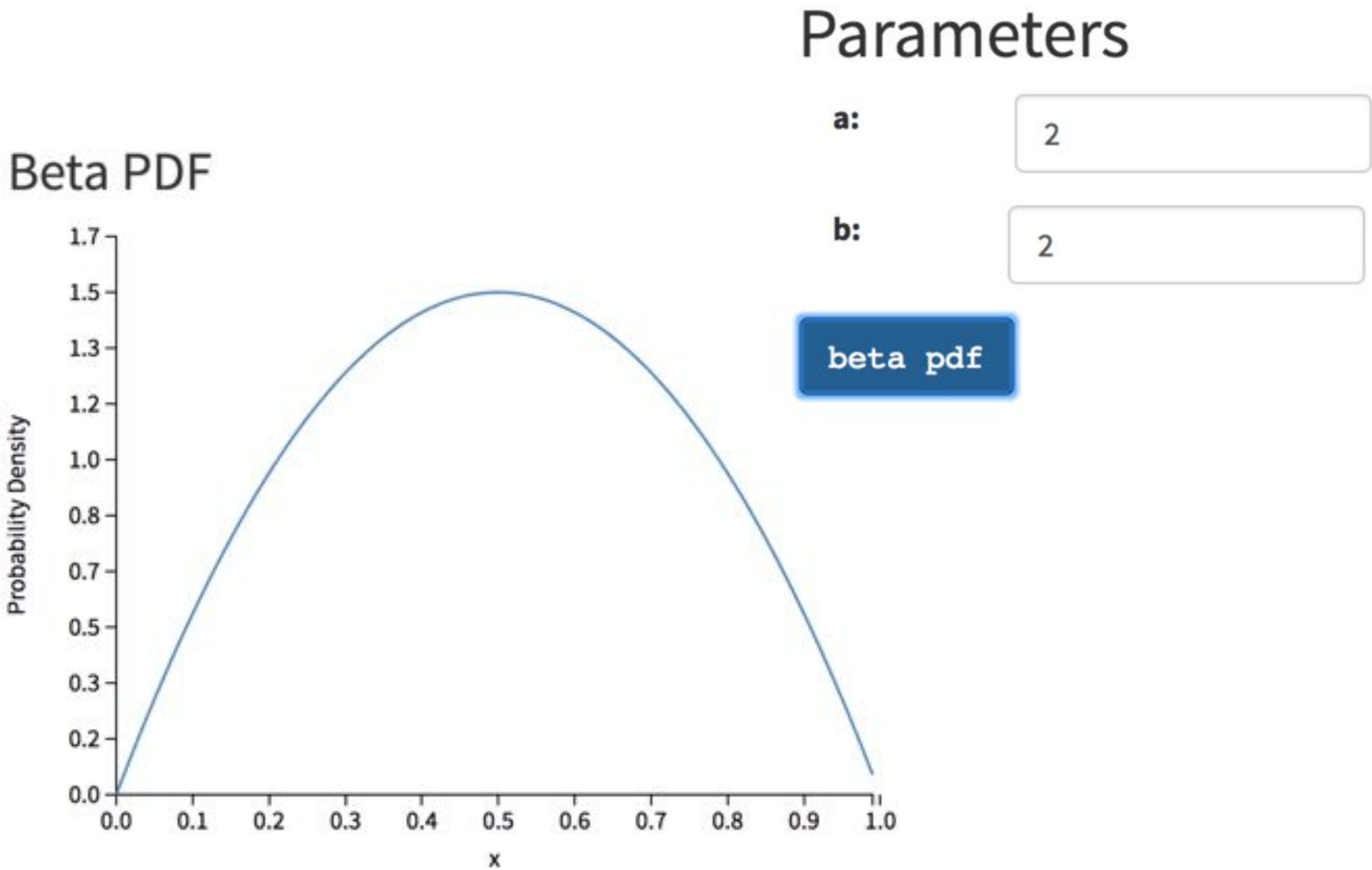
Let X be the probability of rolling a "6" on Chris' die.

Prior: Imagine 5 die rolls where only showed up as a "6"

Observation: Roll it a few times…

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

Check out Demo!



Damn

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:

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

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

Beta Example

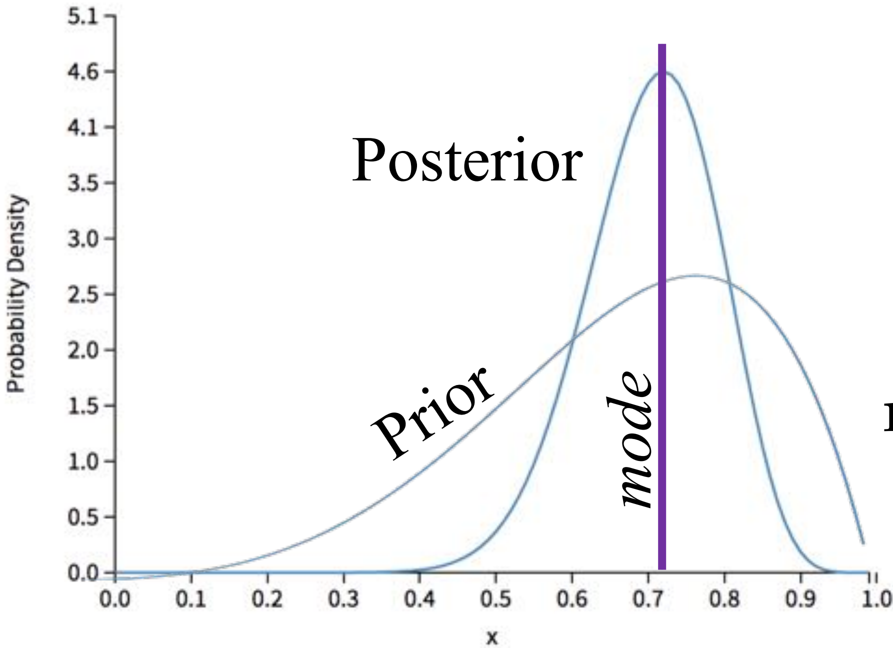
Before being tested, a medicine is believed to "work" about 80% of the time. The medicine is tried on 20 patients. It "works" for 14 and "doesn't work" for 6. What is your new belief that the drug works?

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

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

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

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



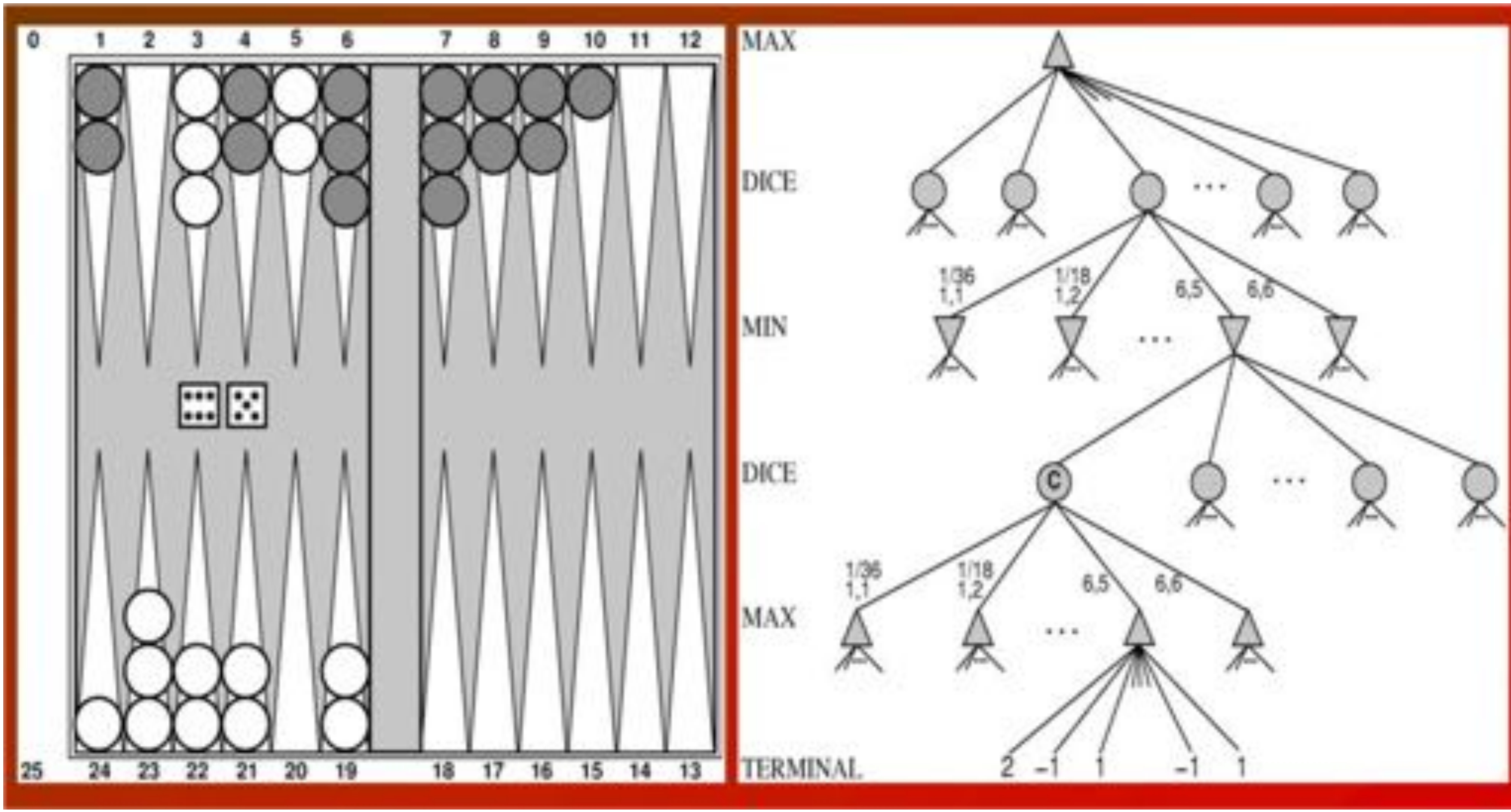
$$E[X] = \frac{a}{a+b} = \frac{19}{19+8} \approx 0.70$$

$$\text{mode}(X) = \frac{a-1}{a+b-2} = \frac{19}{18+7} \approx 0.72$$

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

```python
import pickle
import random

def main():
  X1, X2  = pickle.load(open('probs.pkl', 'rb'))

  print("Welcome to the drug simulator. There are two drugs")

  while True:
    choice = getChoice()
    prob = X1 if choice == "a" else X2
    success = bernoulli(prob)
    if success:
      print('Success. Patient lives!')
    else:
      print('Failure. Patient dies!')
    print('')

```

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)
```

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

Challenge for you

Send me your strategies sometime before Friday

Beta:
The probability density for probabilities

Beta is a distribution for probabilities

Beta Distribution

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

let a = num "successes" + 1

let b = num "failures" + 1

Your new belief about the probability is:

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

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

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

Eg: $X \sim N(\mu, \sigma^2)$

That’s all!