17: Beta

David Varodayan

February 14, 2020

Adapted from slides by Lisa Yan

Bayes in all its forms

Review

Let X, Y be continuous and M, N be discrete random variables.

OG Bayes: $$p_{M|N}(m|n) = \frac{p_{N|M}(n|m)p_M(m)}{p_N(n)}$$

Mix Bayes #1: $$f_{X|N}(x|n) = \frac{p_{N|X}(n|x)f_X(x)}{p_N(n)}$$

Mix Bayes #2: $$p_{N|X}(n|x) = \frac{f_{X|N}(x|n)p_N(n)}{f_X(x)}$$

All continuous: $$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$$

Today's Plan

We are going to learn something unintuitive, beautiful, and useful!

We are going to think of probabilities as random variables.

Mixing discrete and continuous random variables, combined with Bayes' Theorem, allows us to reason about probabilities as random variables.

Today's plan

➡ Thinking of probabilities as random variables

Beta distribution

A new definition of probability

Flip a coin $n+m$ times, comes up with n heads.

We don't know the probability X that the coin comes up with heads.

The world's first coin

Frequentist

X is a single value.

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

Bayesian

X is a **random variable**.

X's support: (0, 1)

Flip a coin with unknown probability

Flip a coin $n + m$ times, comes up with n heads.

- Before our experiment, X (the probability that the coin comes up heads) can be any probability. $f_X(x)$
- Let N = number of heads.
- Given $X = x$, coin flips are independent. $p_{N|X}(n|x)$

What is our updated belief of X after we observe $N = n$? $f_{X|N}(x|n)$

What are the distributions of the following?

1. X
2. $N|X$
3. $X|N$

A. Uni(0,1)
B. Bin$(n + m, x)$
C. Use Bayes'
D. Subjective opinion

Flip a coin with unknown probability

Flip a coin $n+m$ times, comes up with n heads.

- Before our experiment, X (the probability that the coin comes up heads) can be any probability. $f_X(x)$
- Let N = number of heads.
- Given $X = x$, coin flips are independent. $p_{N|X}(n|x)$

What is our updated belief of X after we observe $N = n$? $f_{X|N}(x|n)$

What are the distributions of the following?

1. X Bayesian prior $X \sim \text{Uni}(0,1)$
2. $N|X$ Likelihood $N|X \sim \text{Bin}(n+m, x)$
3. $X|N$ Bayesian posterior. Use Bayes'

A. Uni$(0,1)$
B. Bin$(n+m, x)$
C. Use Bayes'
D. Subjective opinion

Flip a coin with unknown probability

Flip a coin $n + m$ times, comes up with n heads.

- Before our experiment, X (the probability that the coin comes up heads) can be any probability.
- Let N = number of heads.
- Given $X = x$, coin flips are independent.

What is our updated belief of X after we observe $N = n$?

Prior:
$X \sim \text{Uni}(0,1)$

Likelihood:
$N|X \sim \text{Bin}(n+m, x)$

Posterior: $f_{X|N}(x|n)$

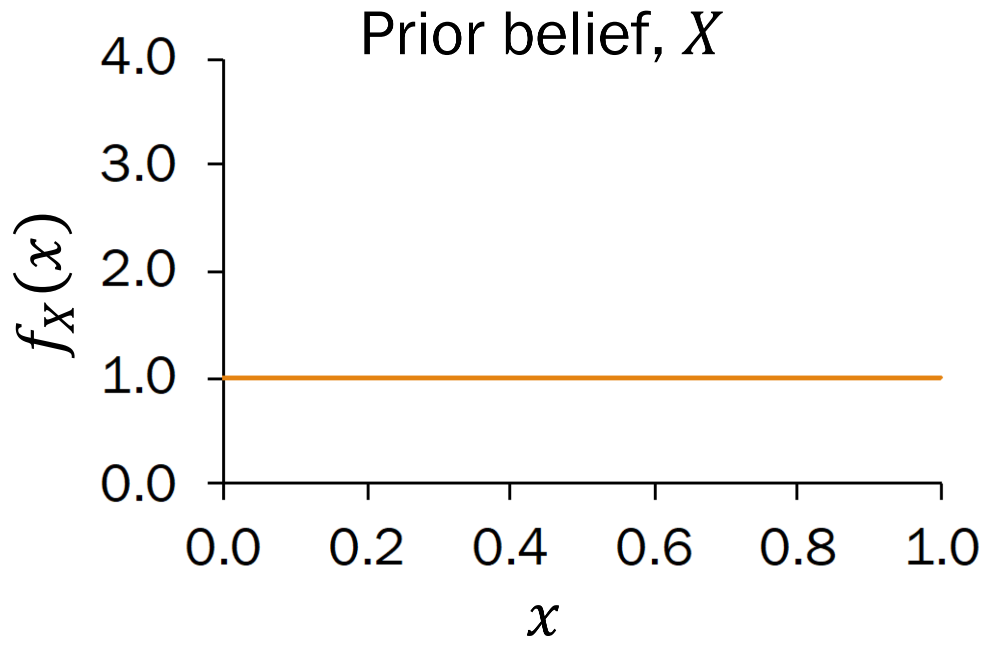
$$f_{X|N}(x|n) = \frac{p_{N|X}(n|x) f_X(x)}{p_N(n)} = \frac{\binom{n+m}{n} x^n (1-x)^m \cdot 1}{p_N(n)}$$

$$= \underbrace{\frac{\binom{n+m}{n}}{p_N(n)}}_{\text{constant, doesn't depend on } x} x^n (1-x)^m = \frac{1}{c} x^n (1-x)^m, \text{ where } c = \int_0^1 x^n (1-x)^m dx$$

Flip a coin with unknown probability

- Start with a $X \sim \text{Uni}(0,1)$ over probability
- Observe n successes and m failures
- Your new belief about the probability of X is:

$$f_{X|N}(x|n) = \frac{1}{c}\, x^n(1-x)^m, \text{where } c = \int_0^1 x^n(1-x)^m dx$$



Suppose our experiment is 8 flips of a coin. We observe:

- $n = 7$ heads (successes)
- $m = 1$ tail (failure)

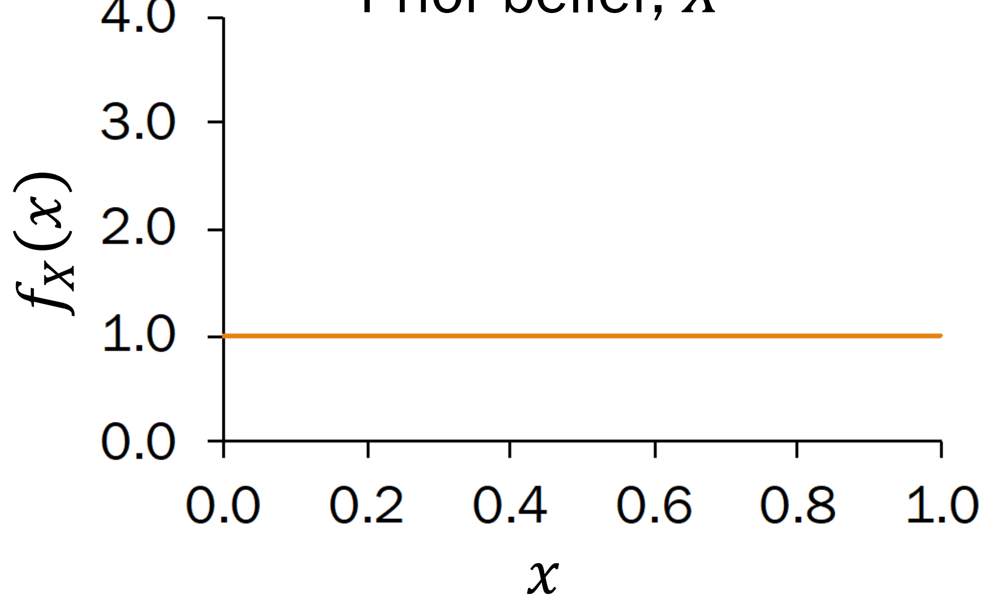
What is our posterior belief, $X|N$?

Flip a coin with unknown probability

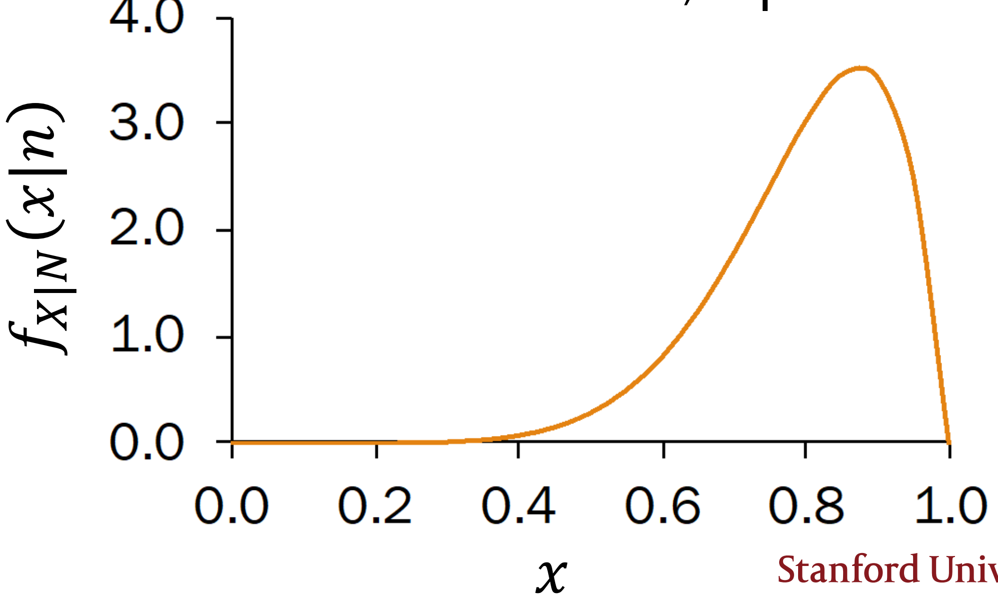
- Start with a $X \sim \text{Uni}(0,1)$ over probability
- Observe $n = 7$ successes and $m = 1$ failures
- Your new belief about the probability of X is:

$$f_{X|N}(x|n) = \frac{1}{c}\, x^{7}(1-x)^{1}, \text{where } c = \int_0^1 x^{7}(1-x)^{1} dx$$

Prior belief, X



Posterior belief, $X|N$



Announcements

Problem Set 4

Due: Wednesday 2/19

Late Day Reminder

No late days permitted past last day of the quarter, 3/13

Announcement: CS109 contest

Do something cool and creative with probability

Genuinely optional extra credit

Due Monday 3/9, 11:59pm

Today's plan

Thinking of probabilities as random variables

➡ Beta distribution

Beta random variable

<u>def</u> A **Beta** random variable X is defined as follows:

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

$$a > 0, b > 0$$

Support of X: $(0, 1)$

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

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

Expectation $$E[X] = \frac{a}{a+b}$$

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

Beta is a distribution for probabilities.

Beta is a distribution of probabilities

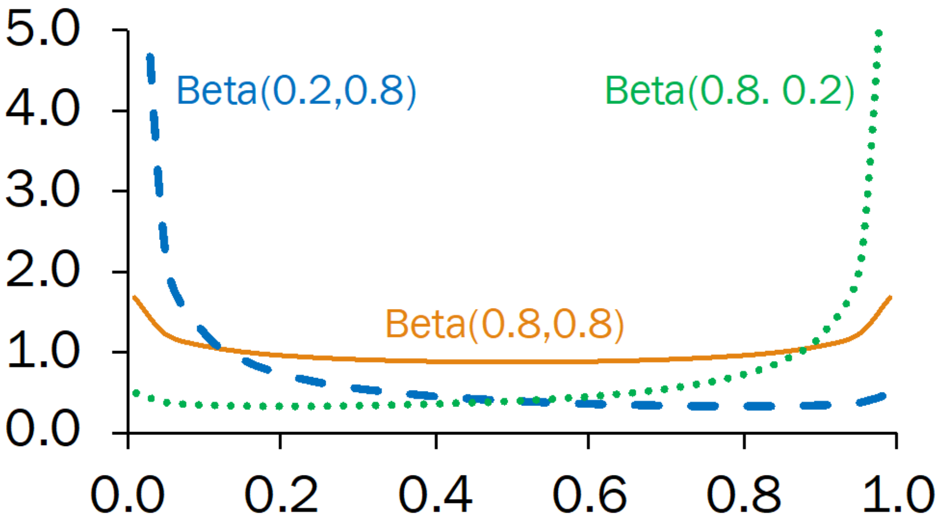
$$X \sim \text{Beta}(a, b)$$

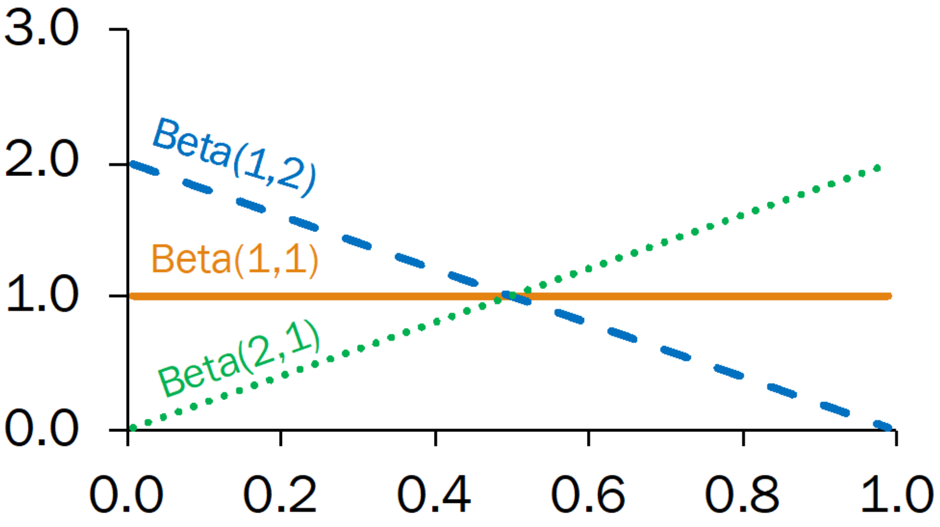
$$a > 0, b > 0$$

Support of X: $(0, 1)$

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

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

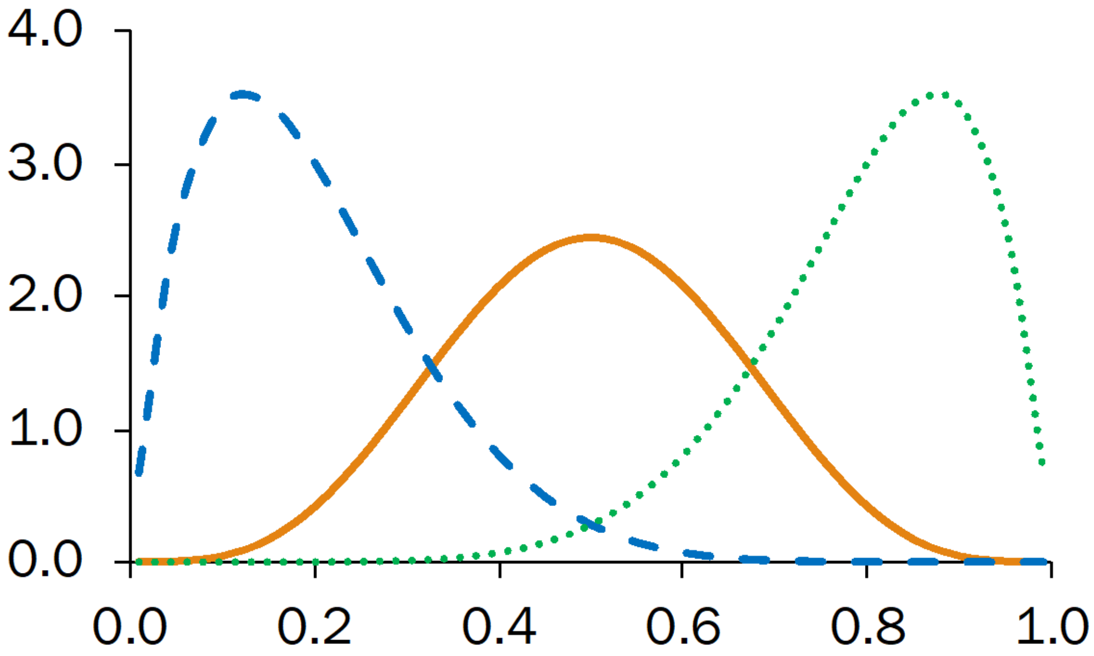




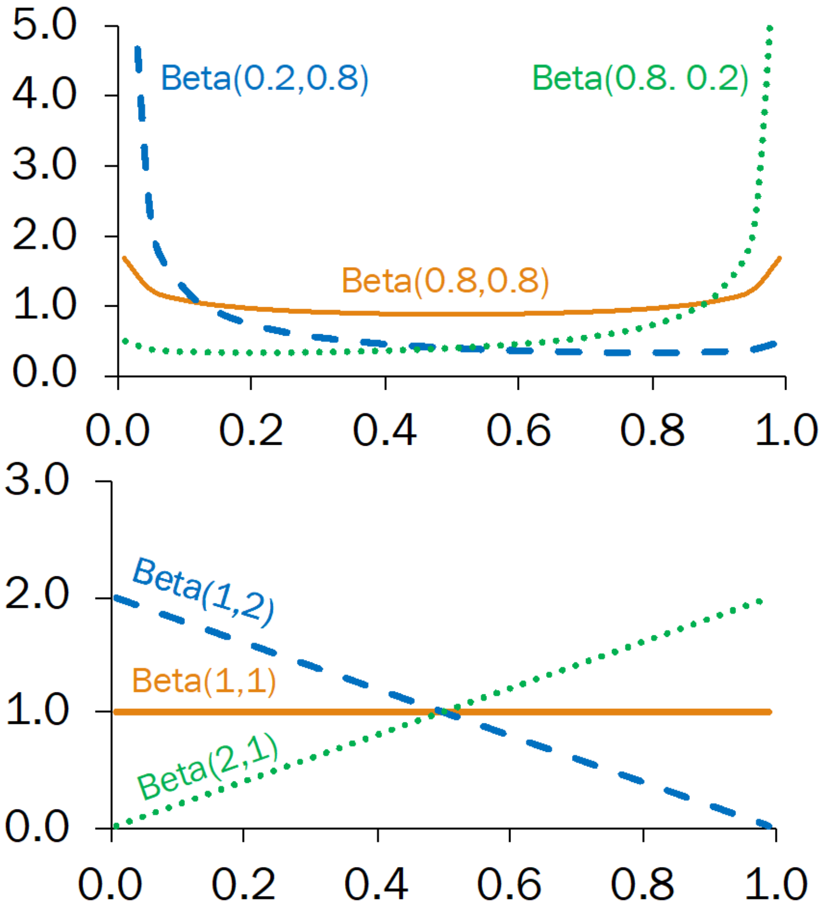
CS109 focus: Beta where a, b both positive integers

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

Match PDF to distribution:



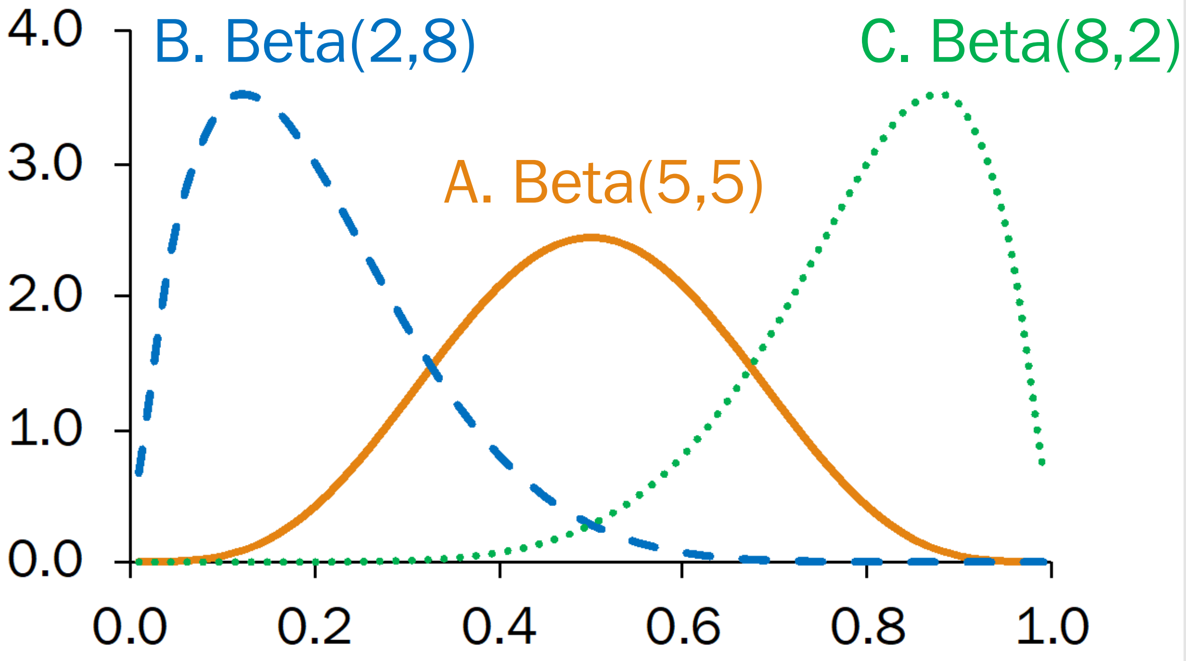
A. Beta(5,5)
B. Beta(2,8)
C. Beta(8,2)



CS109 focus: Beta where a, b both positive integers

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

Match PDF to distribution:



A. Beta(5,5)
B. Beta(2,8)
C. Beta(8,2)

Beta parameters a, b *could* come from an experiment:

$$a = \text{“successes”} + 1$$
$$b = \text{“failures”} + 1$$

Back to flipping coins

- Start with a $X \sim \text{Uni}(0,1)$ over probability
- Observe $n = 7$ successes and $m = 1$ failures
- Your new belief about the probability of X is:

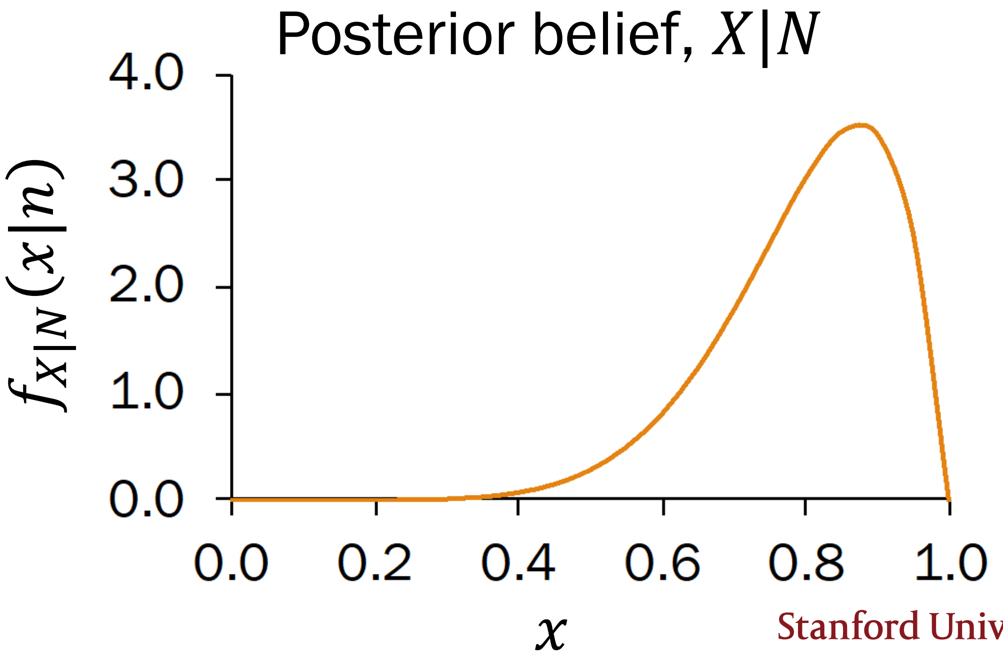
$$f_{X|N}(x|n) = \frac{1}{c}\, x^{7}(1-x)^{1}, \text{where } c = \int_0^1 x^{7}(1-x)^{1}dx$$

Posterior belief, $X|N$:

Beta$(a = 8, b = 2)$

$$f_{X|N}(x|n) = \frac{1}{c}\, x^{8-1}(1-x)^{2-1}$$

Beta$(a = n + 1, b = m + 1)$



Understanding Beta

- Start with a $X \sim \text{Uni}(0,1)$ over probability
- Observe n successes and m failures
- Your new belief about the probability of X is:

$$X|N \sim \text{Beta}(a = n + 1, b = m + 1)$$

Understanding Beta

- Start with a $X \sim \text{Uni}(0,1)$ over probability
- Observe n successes and m failures
- Your new belief about the probability of X is:

$$X|N \sim \text{Beta}(a = n + 1, b = m + 1)$$

Check this out:

Beta$(a = 1, b = 1)$ has PDF:

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} = \frac{1}{B(a,b)} x^0(1-x)^0 = \frac{1}{\int_0^1 1dx}$$

where $0 < x < 1$

So our prior $X \sim \text{Beta}(a = 1, b = 1)$!

If the prior is a Beta…

Let X be our random variable for probability of success

- If our **prior belief** about X is beta: $X \sim \text{Beta}(a, b)$

likelihood

- …and if we observe n successes and m failures: $N|X \sim \text{Bin}(n + m, x)$
- …then our **posterior belief** about X is also beta. $X|N \sim \text{Beta}(a + n, b + m)$

This is the main takeaway of today.

If the prior is a Beta…

Let X be our random variable for probability of success

- If our **prior belief** about X is beta: $X \sim \text{Beta}(a, b)$
- …and if we observe n successes and m failures: likelihood $N|X \sim \text{Bin}(n+m, x)$
- …then our **posterior belief** about X is also beta. $X|N \sim \text{Beta}(a+n, b+m)$

Proof:

$$f_{X|N}(x|n) = \frac{p_{N|X}(n|x) f_X(x)}{p_N(n)} = \frac{\binom{n+m}{m} x^n (1-x)^m \cdot \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}}{p_N(n)}$$

constants that don't depend on x

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

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

If the prior is a Beta…

Let X be our random variable for probability of success

- If our **prior belief** about X is beta: $X \sim \text{Beta}(a, b)$
- likelihood …and if we observe n successes and m failures: $N|X \sim \text{Bin}(n + m, x)$
- …then our **posterior belief** about X is also beta. $X|N \sim \text{Beta}(a + n, b + m)$

Beta is a conjugate distribution for Binomial.

- Prior and posterior parametric forms are the same
- Practically, conjugate means easy update:
  Add number of “heads” and “tails” seen
  to Beta parameter.

If the prior is a Beta…

Let X be our random variable for probability of success

- If our **prior belief** about X is beta: $X \sim \text{Beta}(a, b)$

likelihood

- …and if we observe n successes and m failures: $N|X \sim \text{Bin}(n+m, x)$
- …then our **posterior belief** about X is also beta. $X|N \sim \text{Beta}(a+n, b+m)$

You can set the prior to reflect how biased you think the coin is a priori.

- This is a subjective probability!
- $X \sim \text{Beta}(a, b)$: have seen $(a+b-2)$ **imaginary trials**, where $(a-1)$ are heads, $(b-1)$ tails
- Then $\text{Beta}(1,1) = \text{Uni}(0,1)$ means we haven't seen any imaginary trials

If the prior is a Beta…

Let X be our random variable for probability of success

- If our prior belief about X is beta: $X \sim \text{Beta}(a, b)$
- …and if we observe n successes and m failures: $N|X \sim \text{Bin}(n+m, x)$

likelihood

- …then our posterior belief about X is also beta. $X|N \sim \text{Beta}(a+n, b+m)$

Prior $\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$

Posterior $\text{Beta}\left(a = n_{imag} + n + 1, b = m_{imag} + m + 1\right)$

This is the main takeaway of Beta.

Medicinal Beta

- Before being tested, a medicine is believed to “work” 80% of the time.
- The medicine is tried on 20 patients.
- It “works” for 14, “doesn’t work” for 6.

What is your new belief that the drug “works”?

Frequentist

Bayesian

Medicinal Beta

- Before being tested, a medicine is believed to "work" 80% of the time.
- The medicine is tried on 20 patients.
- It "works" for 14, "doesn't work" for 6.

What is your new belief that the drug "works"?

Frequentist

Let p be the probability your drug works.

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

Bayesian

A frequentist view will not incorporate prior/expert belief about probability.

Medicinal Beta

- Before being tested, a medicine is believed to "work" 80% of the time.
- The medicine is tried on 20 patients.
- It "works" for 14, "doesn't work" for 6.

What is your new belief that the drug "works"?

Frequentist

Let p be the probability your drug works.

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

Bayesian

Let X be the probability your drug works.

X is a random variable.

Medicinal Beta

Prior $\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$

Posterior $\text{Beta}(a = n_{imag} + n + 1, b = m_{imag} + m + 1)$

- Before being tested, a medicine is believed to "work" 80% of the time.
- The medicine is tried on 20 patients.
- It "works" for 14, "doesn't work" for 6.

What is your new belief that the drug "works"? (Bayesian interpretation)

What is the prior distribution of X? (select all that apply)

A. $X \sim \text{Beta}(1, 1) = \text{Uni}(0, 1)$

B. $X \sim \text{Beta}(81, 101)$

C. $X \sim \text{Beta}(80, 20)$

D. $X \sim \text{Beta}(81, 21)$

E. $X \sim \text{Beta}(5, 2)$

Medicinal Beta

Prior $\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$

Posterior $\text{Beta}(a = n_{imag} + n + 1, b = m_{imag} + m + 1)$

- Before being tested, a medicine is believed to “work” 80% of the time.
- The medicine is tried on 20 patients.
- It “works” for 14, “doesn’t work” for 6.

What is your new belief that the drug “works”? (Bayesian interpretation)

What is the prior distribution of X? (select all that apply)

A. $X \sim \text{Beta}(1, 1) = \text{Uni}(0, 1)$

B. $X \sim \text{Beta}(81, 101)$

C. $X \sim \text{Beta}(80, 20)$

[✓ D.] $X \sim \text{Beta}(81, 21)$

[✓ E.] $X \sim \text{Beta}(5, 2)$

Which one to pick? Depends on how strong your belief is.
http://web.stanford.edu/class/cs109/demos/beta.html
(We choose E on next slide)

Medicinal Beta

Prior $\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$

Posterior $\text{Beta}(a = n_{imag} + n + 1, b = m_{imag} + m + 1)$

- Before being tested, a medicine is believed to "work" 80% of the time.
- The medicine is tried on 20 patients.
- It "works" for 14, "doesn't work" for 6.

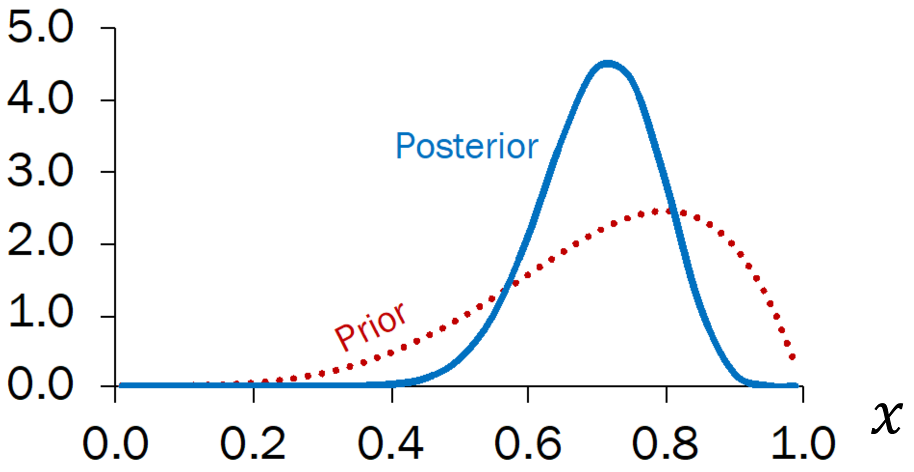
What is your new belief that the drug "works"?

(Bayesian interpretation)

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

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

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



Medicinal Beta

Prior	$\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$
Posterior	$\text{Beta}(a = n_{imag} + n + 1, b = m_{imag} + m + 1)$

- Before being tested, a medicine is believed to “work” 80% of the time.
- The medicine is tried on 20 patients.
- It “works” for 14, “doesn’t work” for 6.

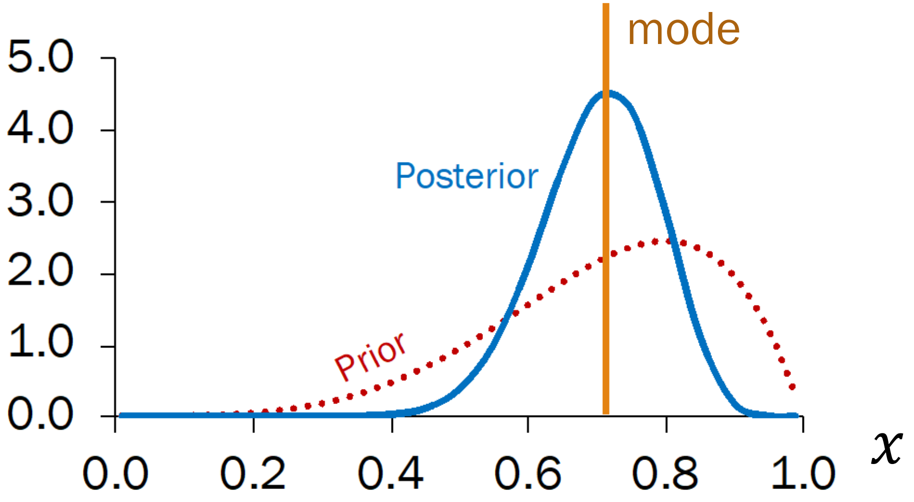
What is your new belief that the drug “works”?

(Bayesian interpretation)

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

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

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



What do you report to pharmacists?

A. Expectation of posterior
B. Mode of posterior
C. Distribution of posterior
D. Nothing

Medicinal Beta

Prior $\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$

Posterior $\text{Beta}(a = n_{imag} + n + 1, b = m_{imag} + m + 1)$

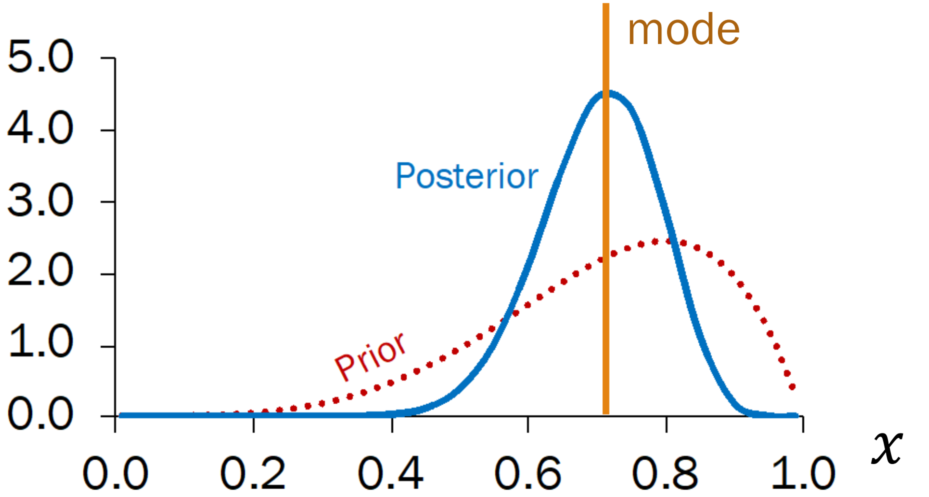
- Before being tested, a medicine is believed to "work" 80% of the time.
- The medicine is tried on 20 patients.
- It "works" for 14, "doesn't work" for 6.

What is your new belief that the drug "works"? (Bayesian interpretation)

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

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

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



What do you report to pharmacists?

[✓ A.] Expectation of posterior
[✓ B.] Mode of posterior
C. Distribution of posterior
D. Nothing

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

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

Food for thought

In this lecture:

$$Y \sim \mathrm{Ber}(p)$$

If we don't know the **parameter** p,
Bayesian statisticians will:

- Treat the parameter as a random variable X with a Beta distribution
- Perform an experiment
- Based on experiment outcomes, update the distribution of X

Food for thought:

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

$$Y \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

Next time: Central Limit Theorem!

Consider n independent and identically distributed (i.i.d.) variables $X_1, X_2, \ldots, X_n$ with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$.

As $n \to \infty$

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

The sum of n i.i.d. random variables is normally distributed with mean $n\mu$ and variance $n\sigma^2$.