17: Beta

Lisa Yan
October 30, 2019

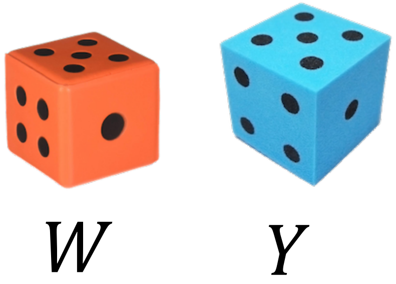
Conditional expectation

Review

The conditional expectation of X (discrete) given $Y = y$ is

$$E[X|Y=y] = \sum_x xP(X=x|Y=y) = \sum_x xp_{X|Y}(x|y)$$

Let W, Y be two RVs for the outcomes of two independent dice rolls, respectively. Let $X = W + Y$.



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

$$= y + \sum_w wP(W=w|Y=y) = y + \sum_w wP(W=w)$$

$$= y + E[W] = y + 3.5$$

👉 $E[X|Y]$ is a random variable.
It is a function of Y.

Properties of conditional expectation

1. LOTUS:

$$E[g(X)|Y=y] = \sum_x g(x) p_{X|Y}(x|y) \quad \text{or} \quad \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\, dx$$

2. Linearity of conditional expectation:

$$E\left[\sum_{i=1}^{n} X_i \mid Y=y\right] = \sum_{i=1}^{n} E[X_i|Y=y]$$

3. Law of total expectation:

$$E[X] = E\big[E[X|Y]\big]$$ 👉

Proof of Law of Total Expectation

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

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

$(g(Y) = E[X|Y])$

$$= \sum_y P(Y=y) \sum_x xP(X=x|Y=y)$$

(def of conditional expectation)

$$= \sum_y \left(\sum_x xP(X=x|Y=y)P(Y=y) \right) = \sum_y \left(\sum_x xP(X=x, Y=y) \right)$$

(chain rule)

$$= \sum_x \sum_y xP(X=x, Y=y) \quad = \sum_x x \sum_y P(X=x, Y=y)$$

(switch order of summations)

$$= \sum_x xP(X=x)$$

(marginalization)

$$= E[X]$$

Properties

1. LOTUS:

$$E[g(X)|Y=y] = \sum_x g(x) p_{X|Y}(x|y) \quad \text{or} \quad \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\, dx$$

2. Linearity of conditional expectation:

$$E\left[\sum_{i=1}^{n} X_i \mid Y = y\right] = \sum_{i=1}^{n} E[X_i|Y=y]$$

3. Law of total expectation:

$$\boxed{E[X] = E\big[E[X|Y]\big]}$$

For any RV X and **discrete** RV Y,

👉 $$E[X] = \sum_y E[X|Y=y]P(Y=y)$$

Analyzing recursive code

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

```
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let Y = return value of `recurse()`. What is $E[Y]$?

$$E[Y] = \underbrace{E[Y|X=1]}_{3}P(X=1) + E[Y|X=2]P(X=2) + E[Y|X=3]P(X=3)$$

When $X = 1$, return 3.

Analyzing recursive code

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

```
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let Y = return value of `recurse()`.

What is $E[Y]$?

$$E[Y] = \underset{3}{E[Y|X=1]}P(X=1) + \boxed{E[Y|X=2]}P(X=2) + E[Y|X=3]P(X=3)$$

When $X = 2$, return 5 + a future return value of `recurse()`.

What is $E[Y|X=2]$?

A. $E[5] + Y$

B. $E[5+Y] = 5 + E[Y]$

C. $E[5] + E[Y|X=2]$

Analyzing recursive code

If Y discrete

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

```
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let Y = return value of `recurse()`.

What is $E[Y]$?

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

3

When $X = 2$, return 5 + a future return value of `recurse()`.

What is $E[Y|X=2]$?

A. $E[5] + Y$

[✓ B.] $E[5+Y] = 5 + E[Y]$

C. $E[5] + E[Y|X=2]$

Analyzing recursive code

If Y discrete

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

```
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let Y = return value of `recurse()`.

What is $E[Y]$?

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

3

$5 + E[Y]$

When $X = 3$, return $7 +$ a future return value of `recurse()`.

$$E[Y|X=3] = 7 + E[Y]$$

Analyzing recursive code

If Y discrete

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

```
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let Y = return value of `recurse()`.

What is $E[Y]$?

$$E[Y] = \underset{\overset{\Uparrow}{3}}{E[Y|X=1]}P(X=1) + \underset{\overset{\Uparrow}{5+E[Y]}}{E[Y|X=2]}P(X=2) + \underset{\overset{\Uparrow}{7+E[Y]}}{E[Y|X=3]}P(X=3)$$

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

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

$$E[Y] = 15$$

Law of Total Expectation, a summary

Conditional expectation of X given Y:

- $E[X|Y]$ is a function of Y.
- To evaluate at $Y = y$, $E[X|Y = y] = \sum_x xP(X = x|Y = y)$

Law of total expectation:

$$\boxed{E[X] = E\big[E[X|Y]\big]}$$

- Helps us analyze recursive code.
- Pro tip: use this more in CS161

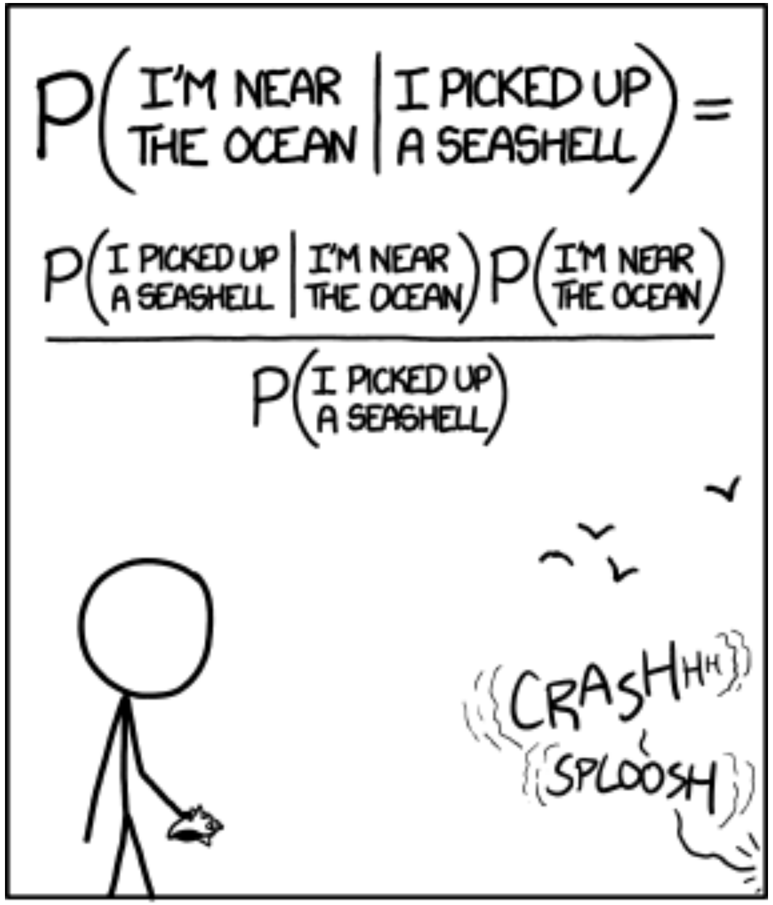
Today’s plan

Law of Total Expectation

➡ Mixing discrete and continuous random variables

Beta distribution

Bayes' on the waves



STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

Let's play a game

Roll a die twice:

- If either time you roll a 6, I win.
- Otherwise you win.

Let W = the event where you win. What is $P(W)$?

If the die is fair:

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

What if the probabilities of the die are unknown?

(demo)

Today's plan

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

We are going to think of probabilities as random variables.

Today’s plan

Law of Total Expectation

➡ Mixing discrete and continuous random variables

Beta distribution

Conditional distributions

Review

For discrete RVs X and Y, the **conditional PMF** of X given Y is

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

Bayes' Theorem:

$$p_{Y|X}(y|x) = \frac{p_{X|Y}(x|y)p_Y(y)}{p_X(x)}$$

For continuous RVs X and Y, the **conditional PDF** of X given Y is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Bayes' Theorem:

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

Conditioning with a continuous RV feels weird at first, but then it gets good

Mixing discrete and continuous

Let X be a continuous random variable, and
N be a discrete random variable.

The conditional PDF of X given N is:

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

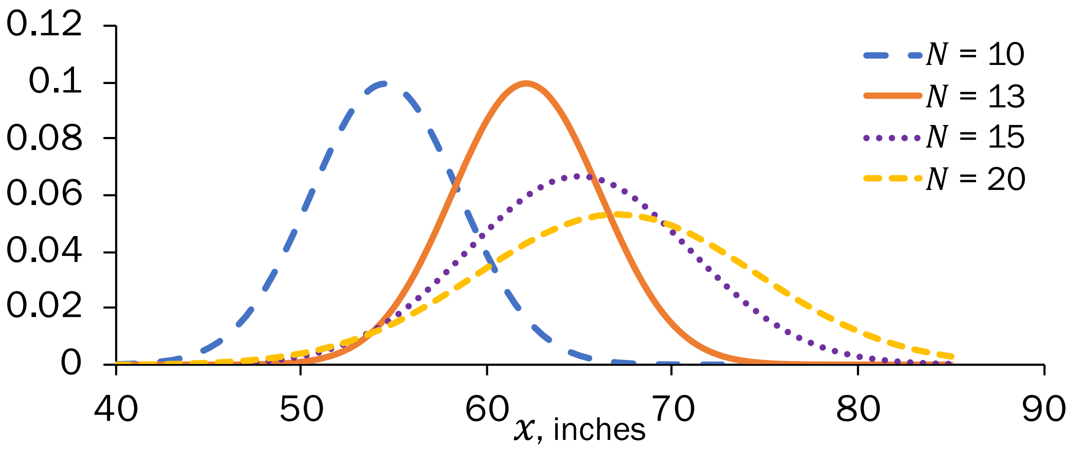
The conditional PMF of N given X is:

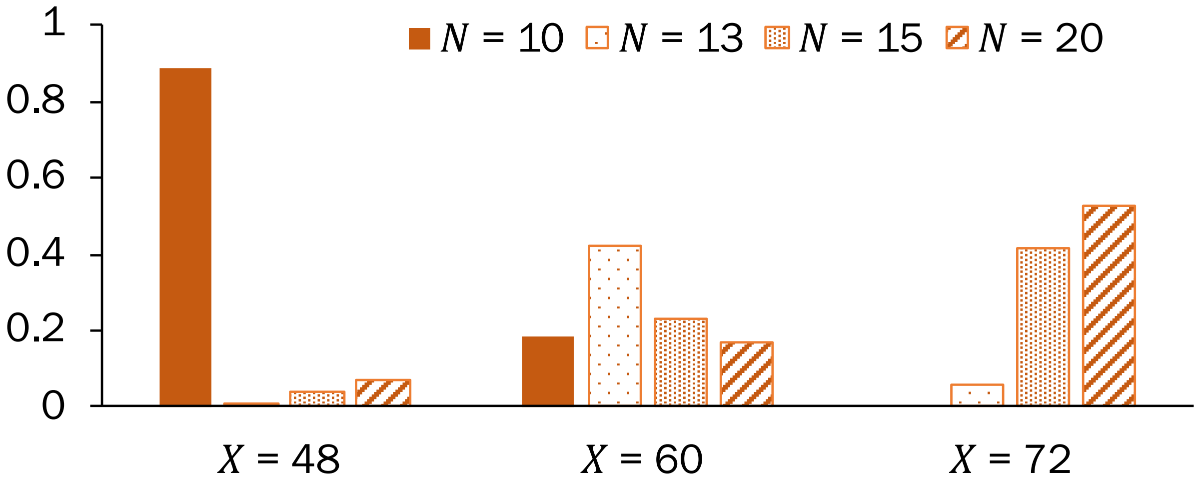
$$p_{N|X}(n|x)$$

Mixing discrete and continuous

Let X be a **continuous** random variable for person's height (inches), and
N be a **discrete** random variable for person's age (10, 13, 15, or 20).

Matching: A. $f_{X|N}(x|n)$, conditional PDF of X given N
B. $p_{N|X}\,(n|x)$, conditional PMF of N given X

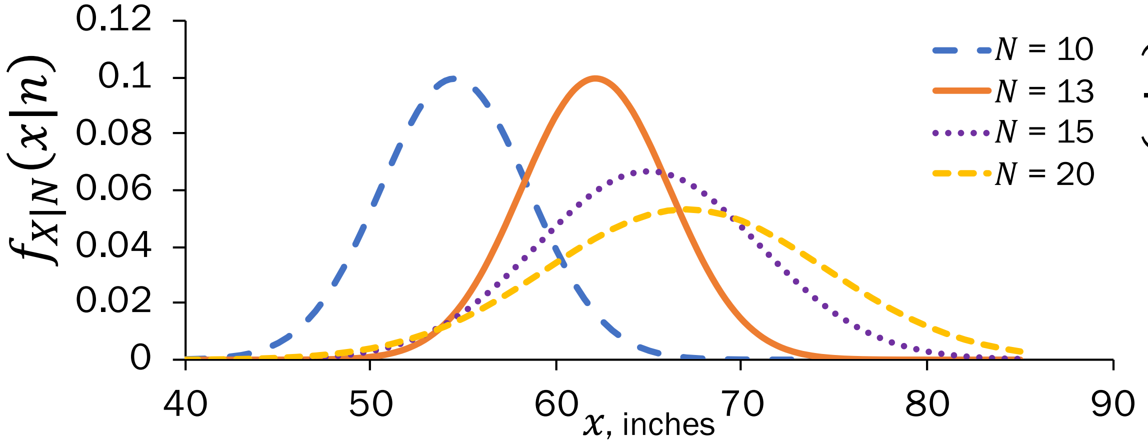




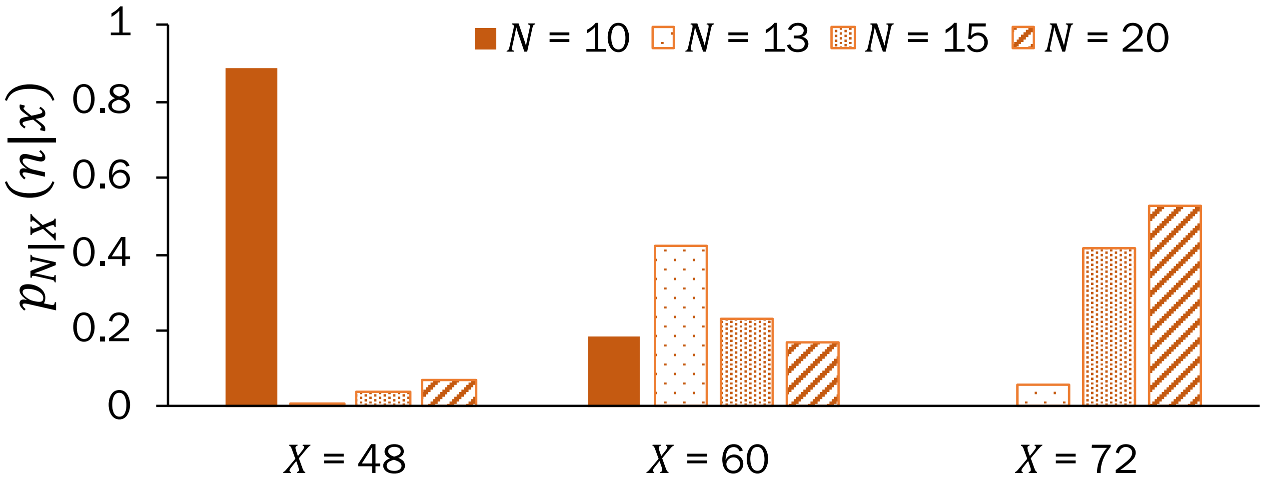
Mixing discrete and continuous

Let X be a **continuous** random variable for person's height (inches), and N be a **discrete** random variable for person's age (10, 13, 15, or 20).

Matching: A. $f_{X|N}(x|n)$, conditional PDF of X given N
B. $p_{N|X}(n|x)$, conditional PMF of N given X



A. conditional PDF of X given N



B. conditional PMF of N given X

Mixing discrete and continuous

Let X be a continuous random variable, and
N be a discrete random variable.

The conditional PDF of X given N is:

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

The conditional PMF of N given X is:

$$p_{N|X}(n|x)$$

Bayes' Theorem:

$$\boxed{f_{X|N}(x|n) = \frac{p_{N|X}(n|x) f_X(x)}{p_N(n)}}$$

Intuition:

$$P(X = x|N = n) = \frac{P(N = n|X = x)P(X = x)}{P(N = n)} \Leftrightarrow f_{X|N}(x|n)\varepsilon_X = \frac{p_{N|X}(n|x) \cdot f_X(x)\varepsilon_x}{p_N(n)}$$

All your Bayes are belong to us

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

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

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.

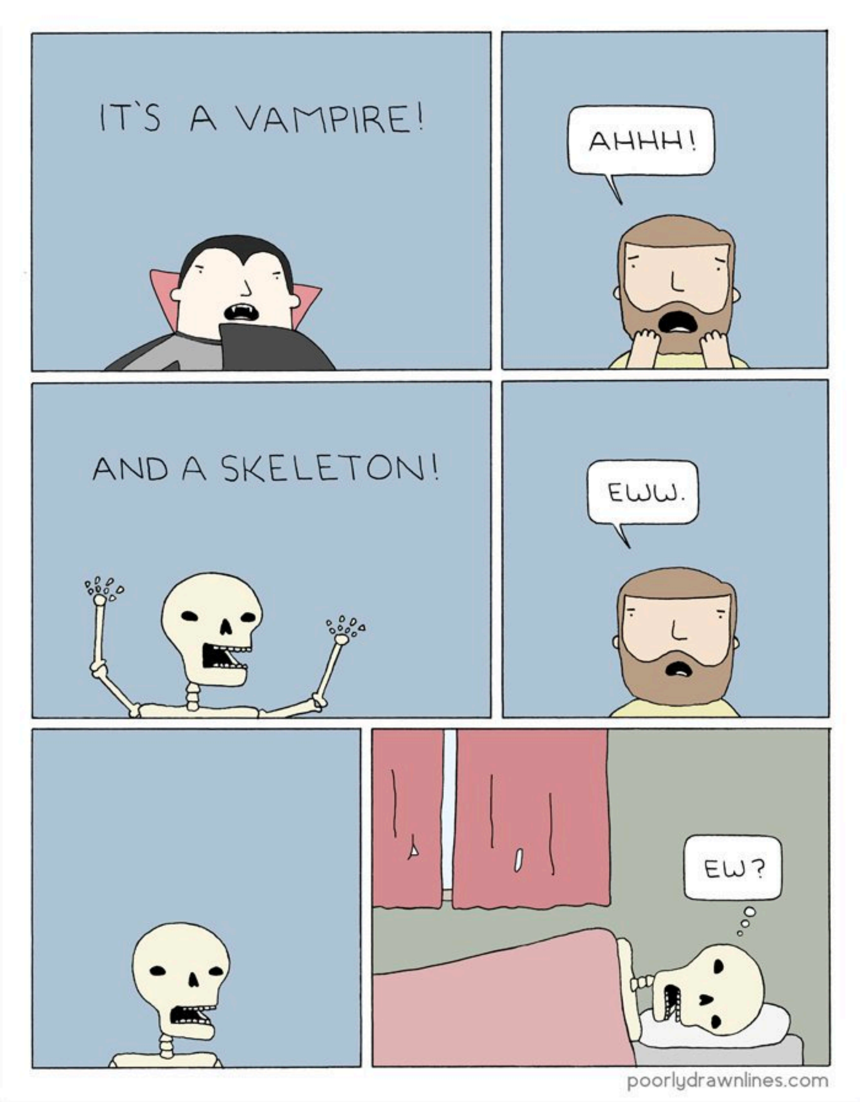
$$X = \lim_{n+m\to\infty} \frac{n}{n+m} \approx \frac{n}{n+m}$$

Bayesian

X is a **random variable**.

X's support: (0, 1)

Break for jokes/ announcements



Announcements

Midterm exam

It’s done! (refrain from posting to Piazza until Thursday)

Grades:	Friday 11/1
Solutions:	Friday 11/1

Concept checks

Week 5’s:	Today (10/31) 11:59pm

Problem Set 4

Due:	Wednesday 11/6
Covers:	Up to Law of Total Expectation

Late day reminder: No late days permitted past last day of the quarter, 12/7

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. Other
E. Don't know

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. Other
E. Don't know

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

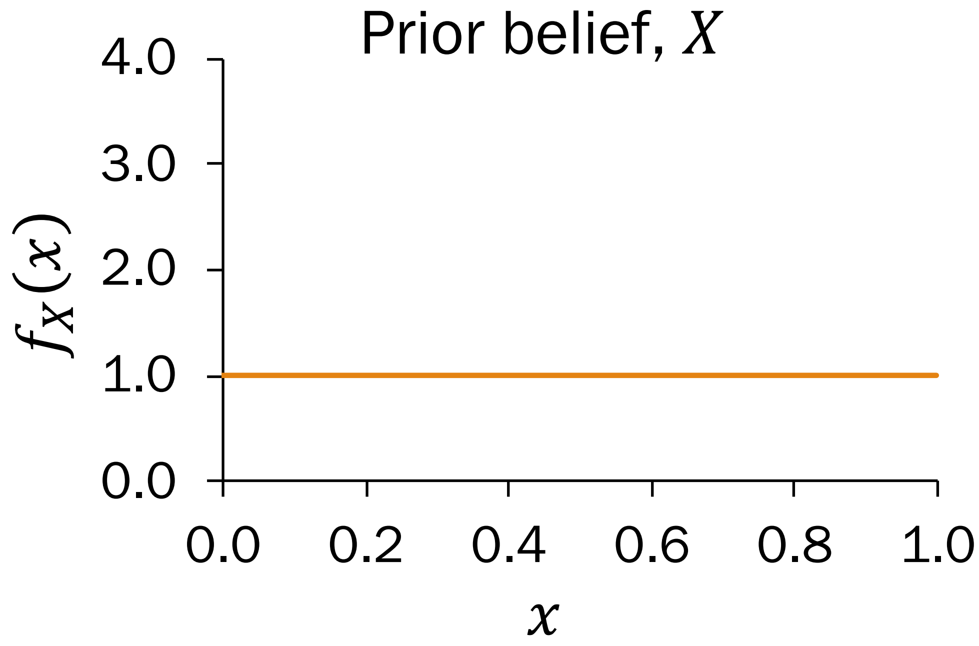
$$f_{X|N}(x|n) = \frac{p_{N|X}(n|x) f_X(x)}{p_N(n)} = \frac{\binom{n+m}{n} x^n (1-x)^m \cdot 1}{p_N(n)}$$

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

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:

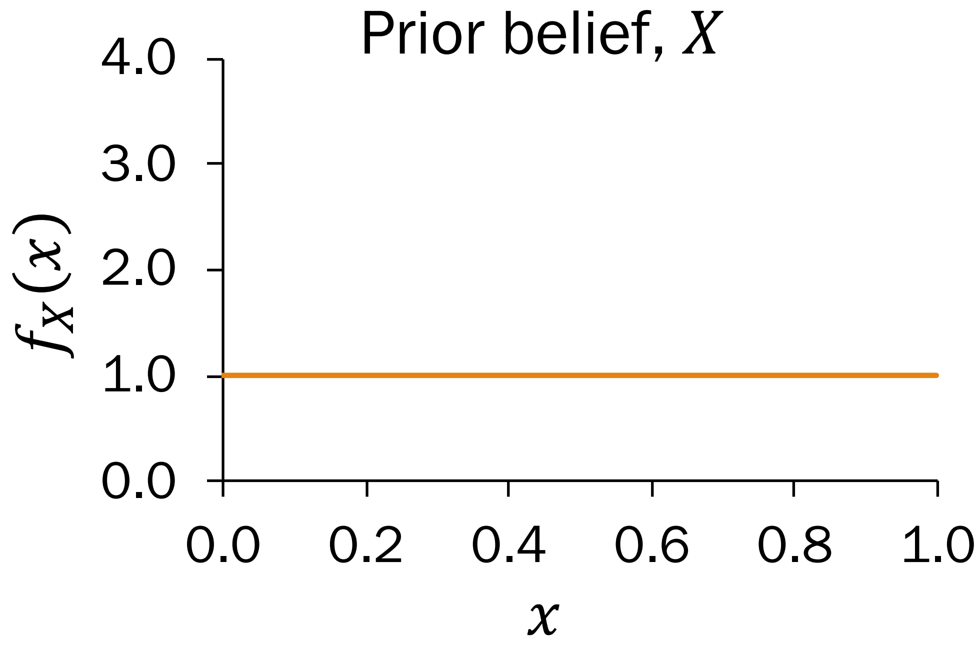
- $n = 7$ heads (successes)
- $m = 1$ tail (failure)

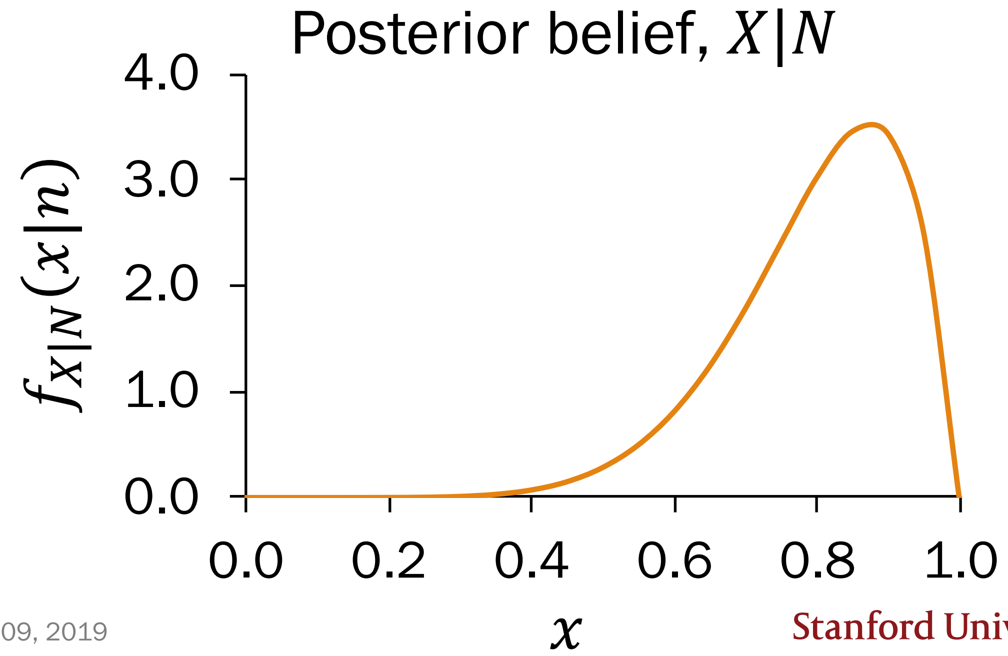
What is our posterior belief, $X|N$?

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:

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





Today's plan

Law of Total Expectation

Mixing discrete and continuous random variables

➡ Beta distribution

Beta random variable

def An **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 $$E[X] = \frac{a}{a+b}$$

Variance $$\text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Beta is a distribution for probabilities.

Beta is a distribution of probabilities

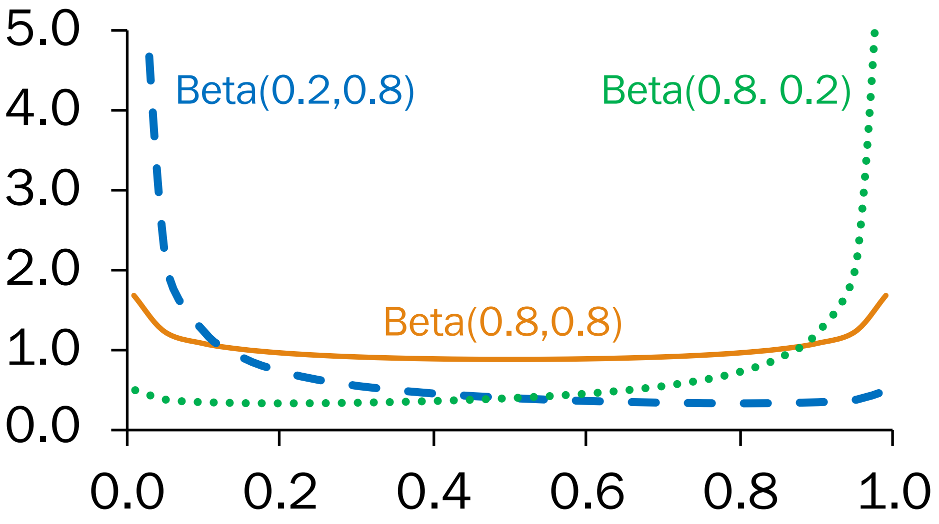
$$X \sim \text{Beta}(a, b)$$

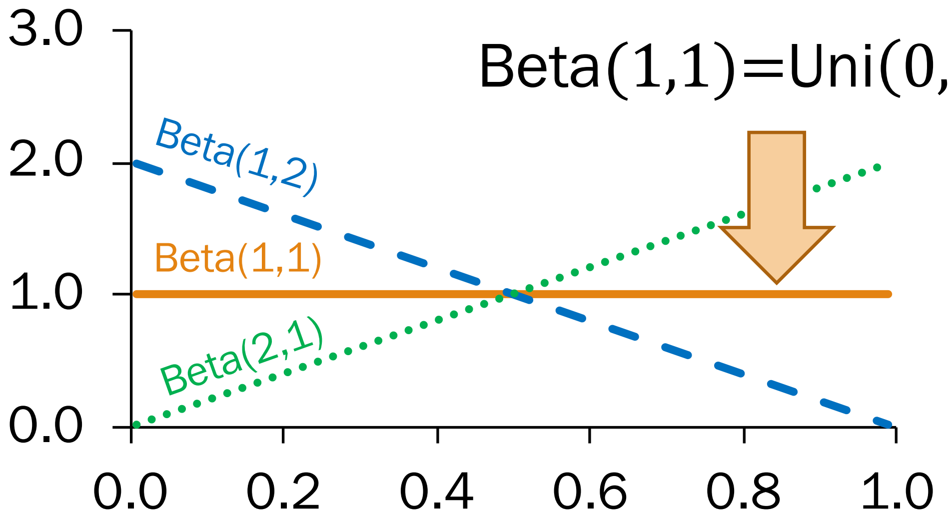
$$a > 0, b > 0$$

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



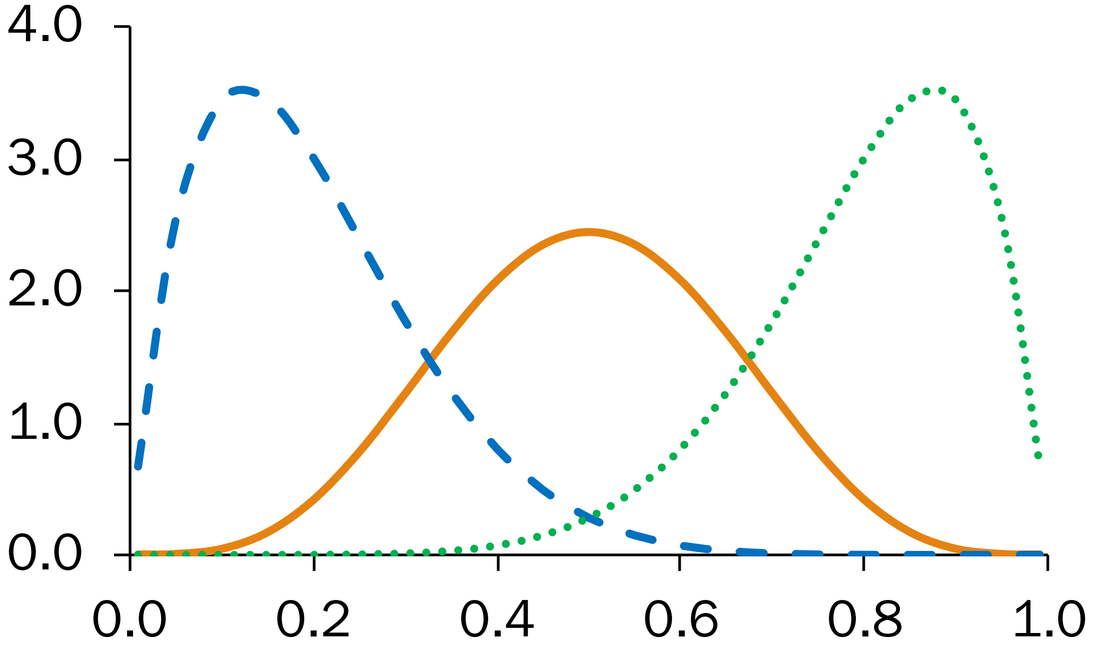


Beta(1,1)=Uni(0,1)

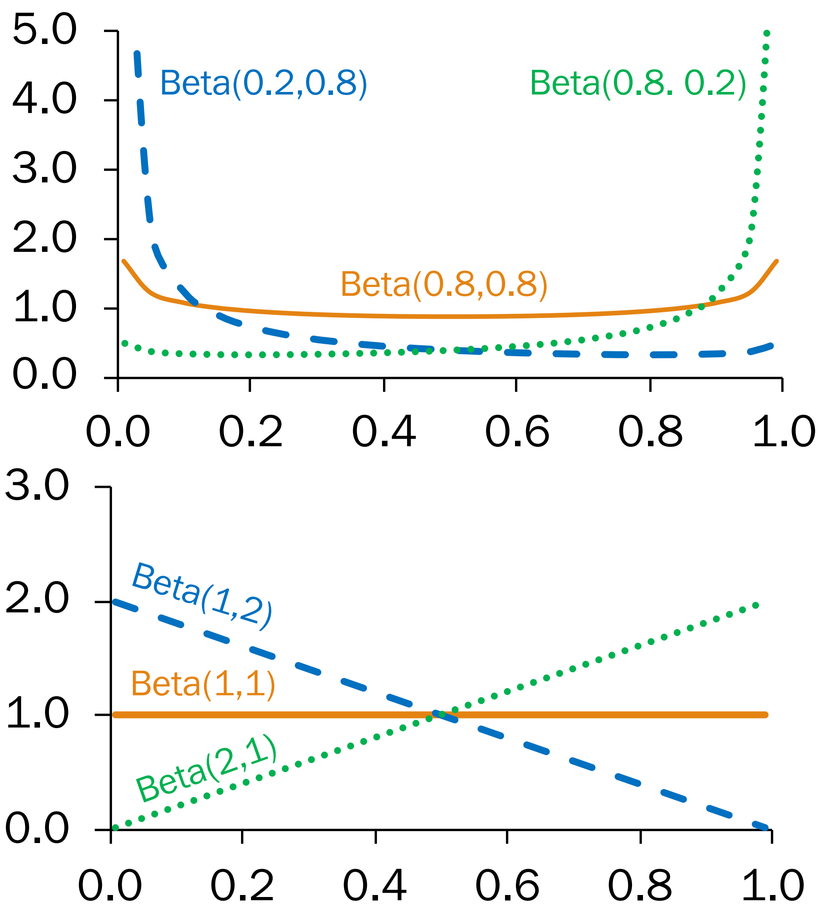
CS109 focus: Beta where a, b both positive integers

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

Match PDF to distribution:



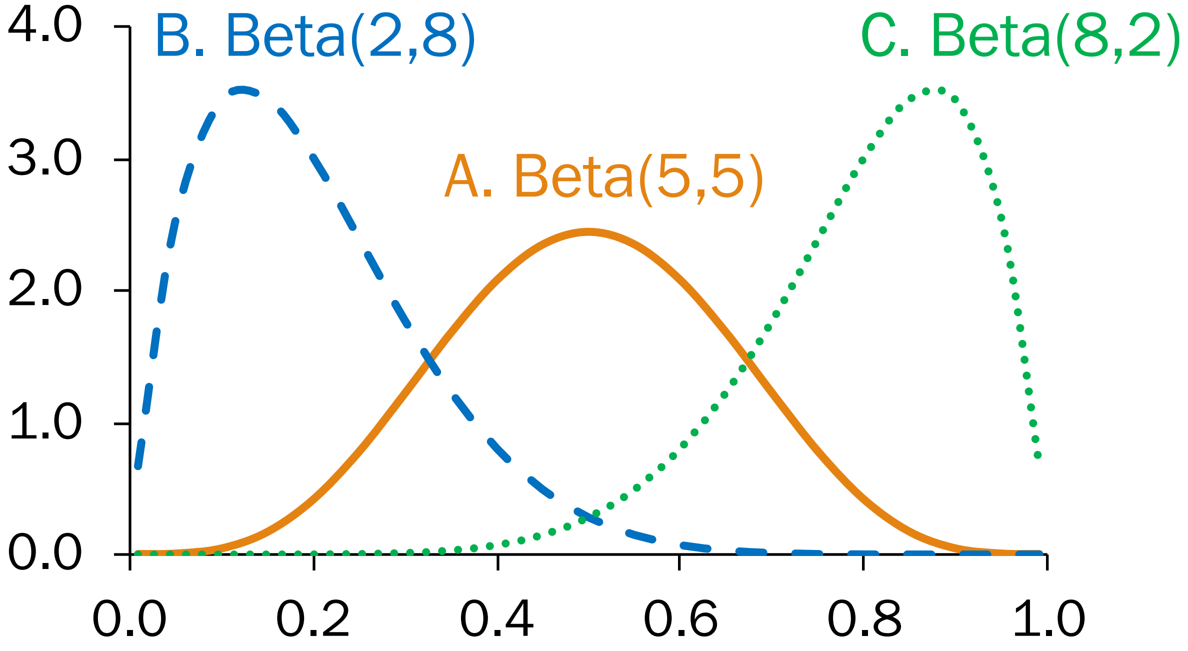
A. Beta(5,5)
B. Beta(2,8)
C. Beta(8,2)



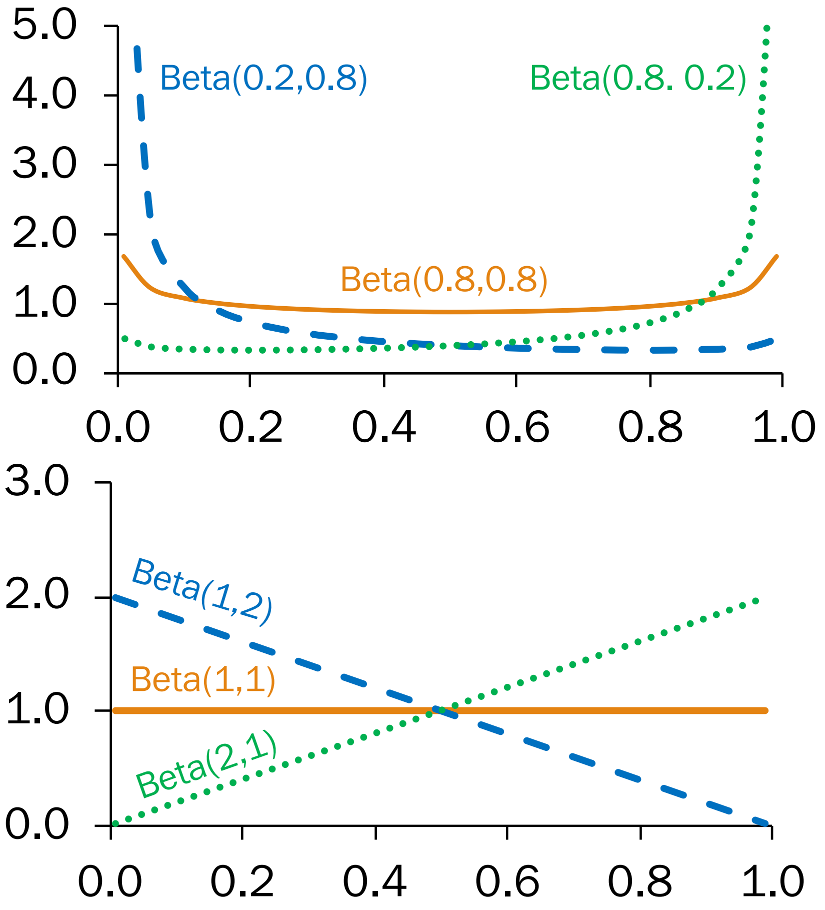
CS109 focus: Beta where a, b both positive integers

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

Match PDF to distribution:



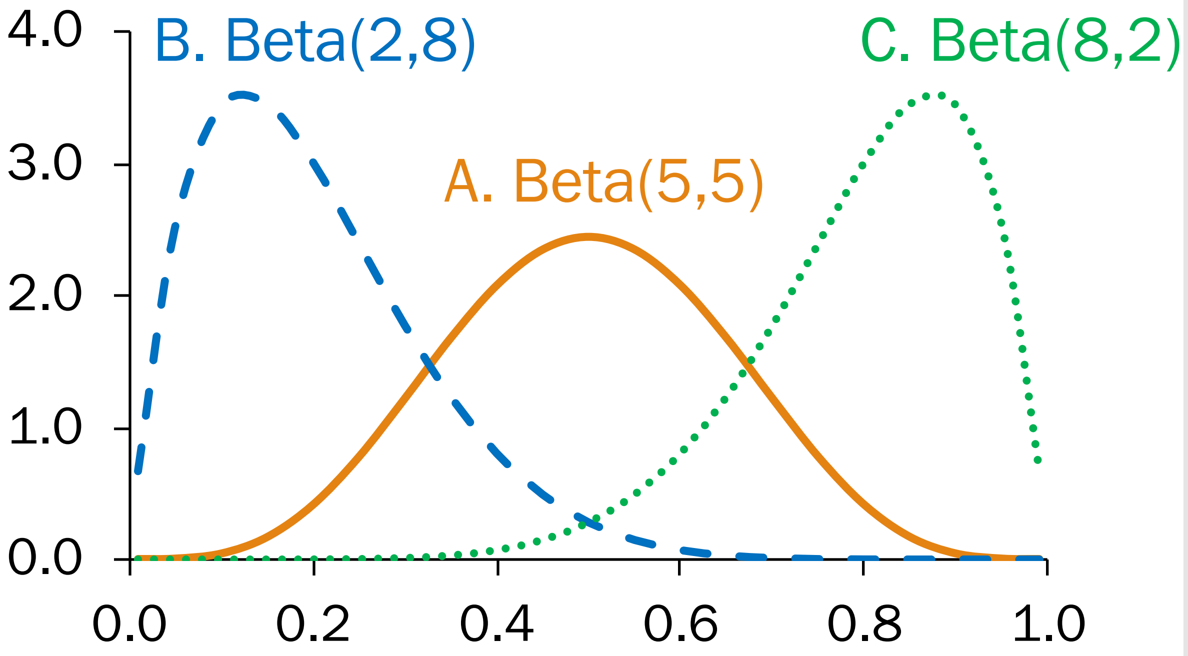
A. Beta(5,5)
B. Beta(2,8)
C. Beta(8,2)



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:

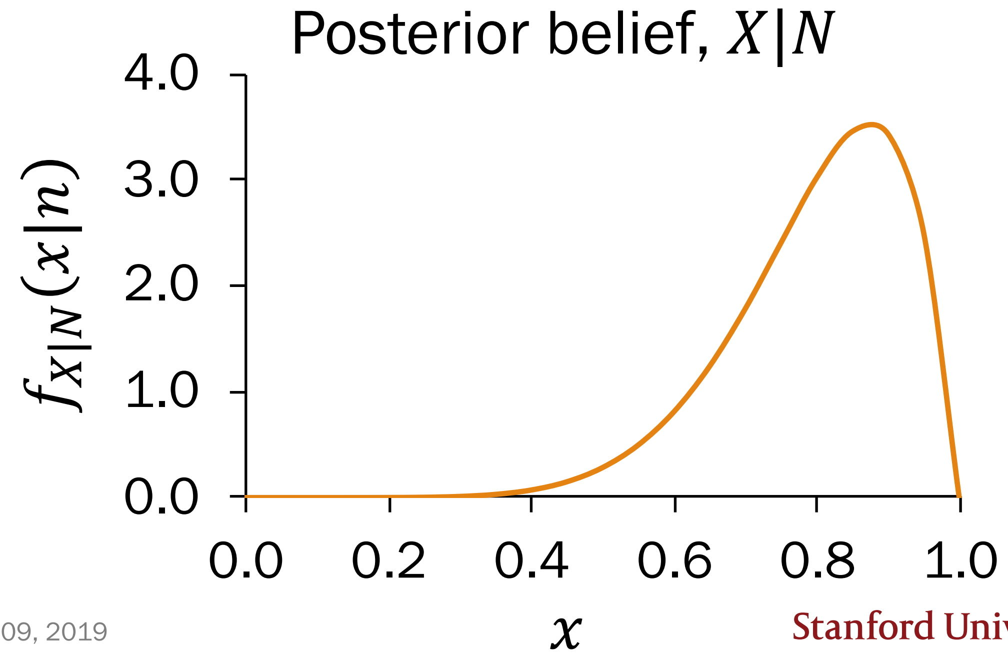
$$f_{X|N}(x|n) = \frac{1}{c}\, x^{7}(1-x)^{1}, \text{where } c = \int_0^1 x^{7}(1-x)^{1}dx$$

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:

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

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 and N

- 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 and N

- 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 and N

- 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.

- 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 and N

- 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 apriori.

- 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 and N

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

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.

The enchanted die

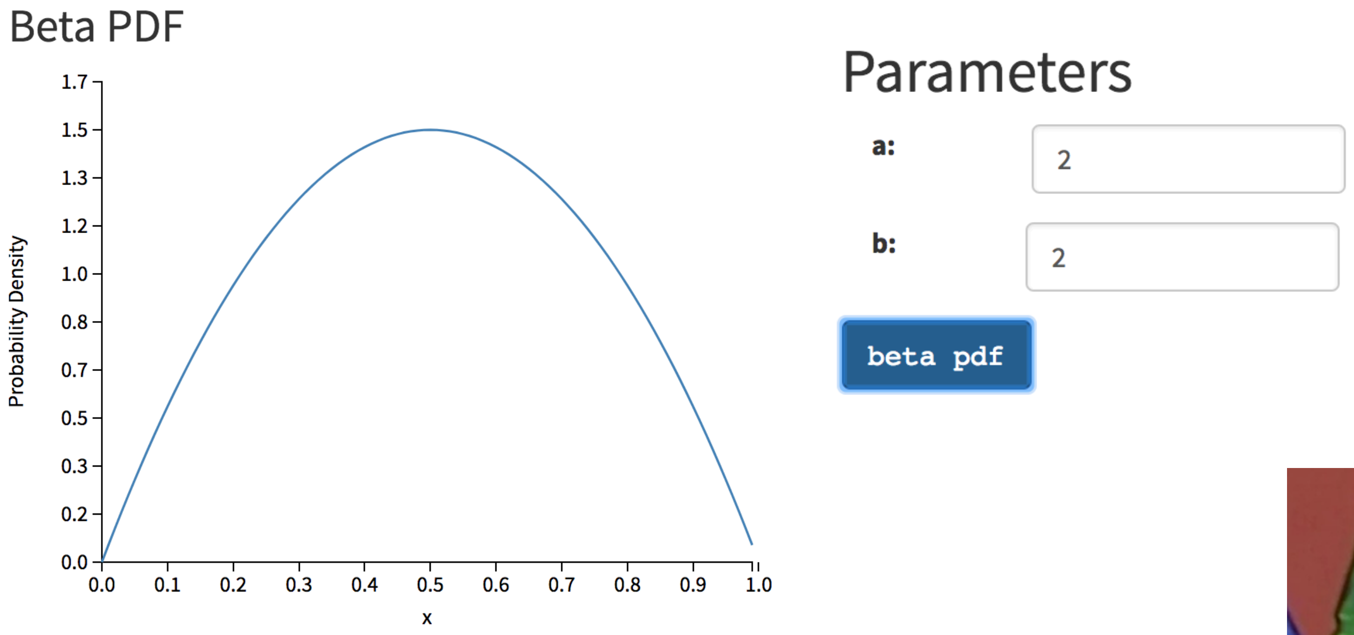
Prior	$\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$
Posterior	$\text{Beta}(a = n_{imag} + n + 1, b = m_{imag} + m + 1)$

Let X be the probability of rolling a 6 on Lisa’s die.

- Prior: Imagine 5 die rolls where only 6 showed up
- Observation: roll it a few times…

What is the updated distribution of X after our observation?

Check out the demo!



Parameters

a: 2

b: 2

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	Beta$(a = n_{imag} + 1, b = m_{imag} + 1)$
Posterior	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. $X\sim$Beta$(1, 1)$ = Uni$(0, 1)$

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

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

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

E. $X\sim$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.] $X \sim \text{Beta}(81, 21)$ Interpretation: 80 successes / 100 imaginary trials

[✓ E.] $X \sim \text{Beta}(5, 2)$ Interpretation: 4 successes / 5 imaginary trials 🤔

(you can choose either; we choose E on next slide)

Medicinal Beta

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

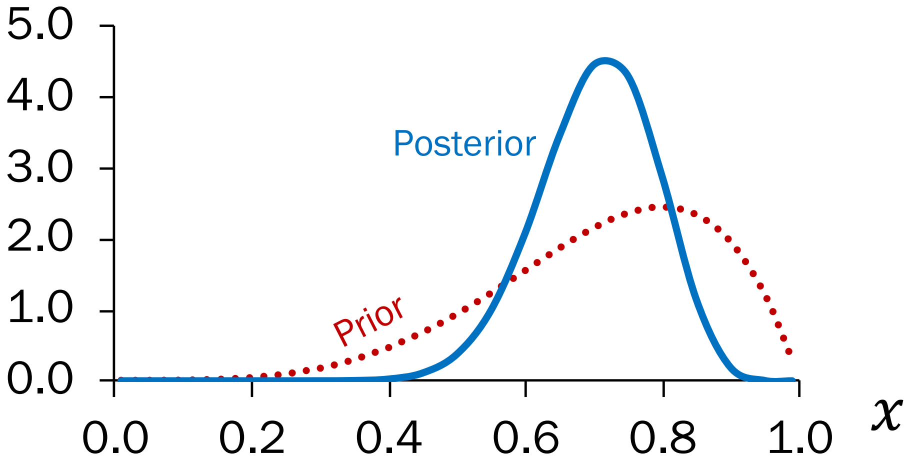
Posterior $\text{Beta}(a = n_{imag} + n + 1, b = m_{imag} + m + 1)$

- 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.

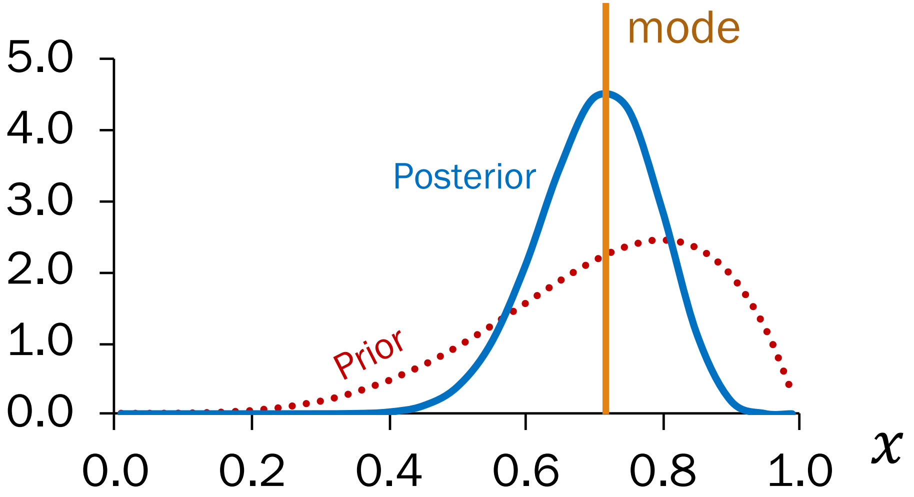
What is your new belief that the drug “works”?

(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

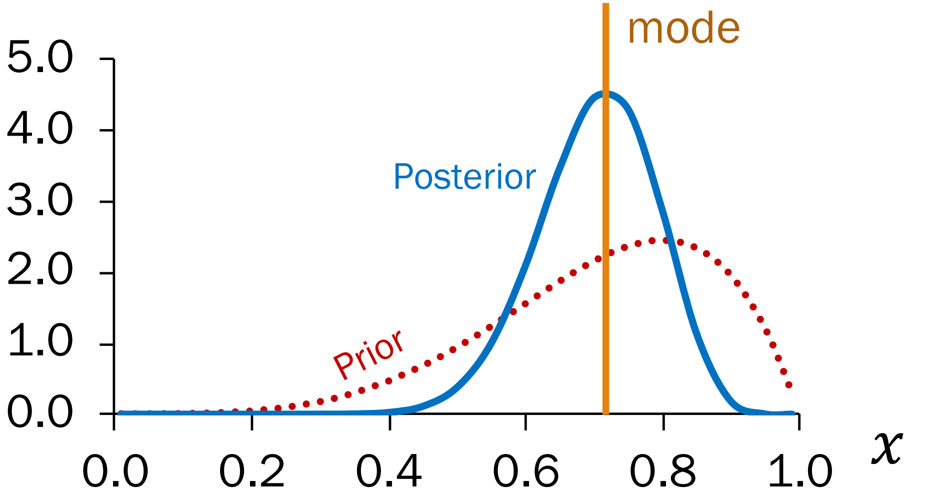
Prior $\text{Beta}(a = n_{imag} + 1, b = m_{imag} + 1)$

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:

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

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.

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