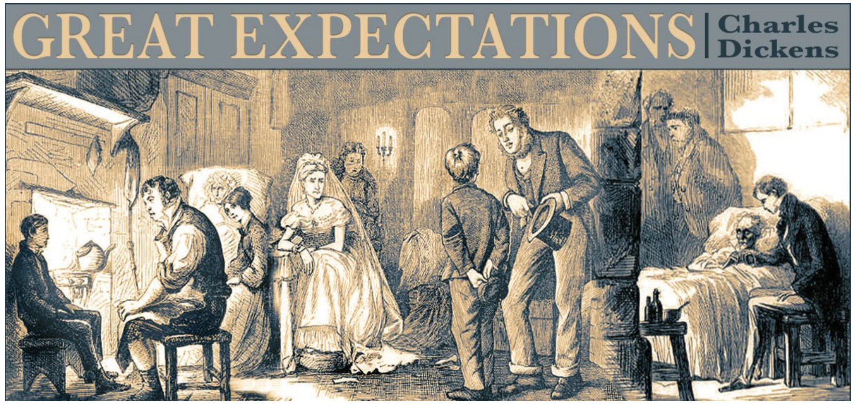


16: Great Expectations

David Varodayan
February 12, 2020
Adapted from slides by Lisa Yan

Today's plan

More expectations of sums of random variables

Conditional Expectation

Law of Total Expectation

Mixing discrete and continuous random variables

Properties of Expectation

1. Linearity:

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

2. Expectation of a sum = sum of expectation:

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

3. Unconscious statistician (LOTUS):

$$E[g(X)] = \sum_{x} g(x)p_X(x)$$

These properties hold **regardless of dependency** of random variables!

Indicator Random Variables

Let $E_1, E_2, \ldots, E_n$ be events with indicator random variables X_i:

- If event E_i occurs, then $X_i = 1$. Else $X_i = 0$.

Recall: $E[X_i] = P(E_i)$

Proof:

$$E[X_i] = 0 \cdot \left(1 - P(E_i)\right) + 1 \cdot P(E_i)$$

From expectation of a sum result:

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} P(E_i)$$

Coupon collector's problem

The coupon collector's problem in probability theory:

- You buy n boxes of cereal.
- There are k different types of coupons
- For each box you buy, you "collect" a coupon of type i.

Servers

n requests

k servers

request to server i

* 52% of Amazon profits
** more profitable than Amazon's North America commerce operations

What is the expected number of utilized servers after n requests?

Computer cluster utilization

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

Consider a computer cluster with k servers. We send n requests.

- Requests independently go to server i with probability p_i
- Let X = # servers that receive ≥ 1 request.

What is $E[X]$?

1. Define additional random variables.

Let: A_i = event that server i receives ≥ 1 request

X_i = indicator for A_i

$$P(A_i) = 1 - P(\text{no requests to } i) = 1 - (1 - p_i)^n$$

Note: A_i are dependent!

2. Solve.

$$E[X_i] = P(A_i) = 1 - (1 - p_i)^n$$

$$E[X] = E\left[\sum_{i=1}^{k} X_i\right] = \sum_{i=1}^{k} E[X_i] = \sum_{i=1}^{k} (1 - (1 - p_i)^n)$$

$$= \sum_{i=1}^{k} 1 - \sum_{i=1}^{k} (1 - p_i)^n = k - \sum_{i=1}^{k} (1 - p_i)^n$$

Coupon collector's problem – Hash tables

The coupon collector's problem in probability theory:

- You buy n boxes of cereal.
- There are k different types of coupons
- For each box you buy, you "collect" a coupon of type i.

Servers	Hash Tables
n requests	n strings
k servers	k buckets
request to server i	hashed to bucket i

What is the expected number of strings to hash until each bucket has ≥ 1 string?

Hash Tables

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

Consider a hash table with n buckets.

- Strings are equally likely to get hashed into any bucket (independently).
- Let Y = # strings to hash until each bucket ≥ 1 string.

What is $E[Y]$?

1. Define additional random variables.
2. Solve.

Let: Y_i = # of trials to get success after i-th success

- Success: hash string to previously empty bucket
- If i non-empty buckets:

$$P(\text{success}) = \frac{k-i}{k}$$

$$P(Y_i = n) = \left(\frac{i}{k}\right)^{n-1}\left(\frac{k-i}{k}\right)$$

Equivalently, $Y_i \sim \text{Geo}\left(p = \frac{k-i}{k}\right)$

$$E[Y_i] = \frac{1}{p} = \frac{k}{k-i}$$

Hash Tables

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

Consider a hash table with k buckets.

- Strings are equally likely to get hashed into any bucket (independently).
- Let Y = # strings to hash until each bucket ≥ 1 string.

What is $E[Y]$?

1. Define additional random variables.

Let: Y_i = # of trials to get success after i-th success

- Success: hash string to previously empty bucket

$$Y_i \sim \text{Geo}\left(p = \frac{k-i}{k}\right)$$

$$E[Y_i] = \frac{1}{p} = \frac{k}{k-i}$$

2. Solve.

$$Y = Y_0 + Y_1 + \cdots + Y_{n-1}$$

Y_i's are dependent!

$$E[Y] = E[Y_0] + E[Y_1] + \cdots + E[Y_{n-1}]$$

$$= \frac{k}{k} + \frac{k}{k-1} + \frac{k}{k-2} + \cdots + \frac{k}{1}$$

$$= k\left[\frac{1}{k} + \frac{1}{k-1} + \cdots + 1\right] = O(k \log k)$$

Announcements

Midterm exam

It's done!

Grades: Friday 2/14

Solutions: Friday 2/14

Problem Set 4

Due: Wednesday 2/19

Midquarter feedback (optional but appreciated)

Link posted in announcement on CS109 webpage

https://forms.gle/6JC6a4oyrH5hEGTy7

Closes: Today, 11:59pm

Today's plan

More expectations of sums of random variables

Conditional Expectation

Law of Total Expectation

Mixing discrete and continuous random variables

Conditional expectation

Recall the the conditional PMF of X given $Y = y$:

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

The conditional expectation of X given $Y = y$ is

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

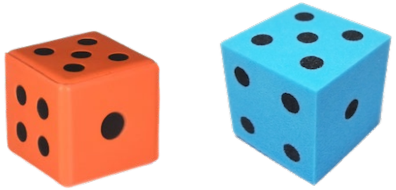
For continuous random variables:

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

$$E[X|Y = y] = \int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx$$

It’s been so long, our dice friends

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



- Roll two 6-sided dice, yielding values D_1 and D_2.
- Let X = value of $D_1 + D_2$
  Y = value of D_2

1. What is $E[X|Y = 6]$?

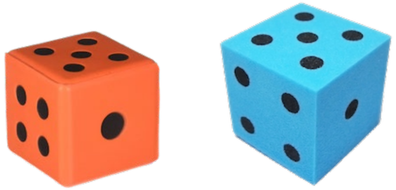
Quick check

A. number
B. function of Y, $g(Y)$
C. function of X, $g(X)$
D. function of X and Y, $g(X, Y)$
E. doesn't make sense

1. $E[X]$
2. $E[X, Y]$
3. $E[X|Y]$
4. $E[X|Y = 6]$
5. $E[Y|X]$
6. $E[X = 1]$

It's been so long, our dice friends

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



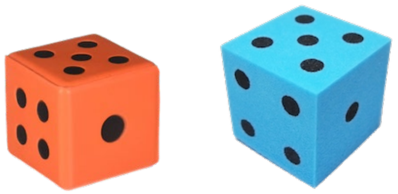
- Roll two 6-sided dice, yielding values D_1 and D_2.
- Let X = value of $D_1 + D_2$
  Y = value of D_2

1. What is $E[X|Y=6]$? $E[X|Y=6] = 9.5$
2. What is $E[X|Y]$?

A. A function of Y
B. A function of X
C. A number

It's been so long, our dice friends

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



- Roll two 6-sided dice, yielding values D_1 and D_2.
- Let X = value of $D_1 + D_2$
  Y = value of D_2

1. What is $E[X|Y=6]$? $E[X|Y=6] = 9.5$
2. What is $E[X|Y]$?

A. A function of Y
B. A function of X
C. A number

Let W = value of D_1. W and Y are independent.

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

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

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

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

$$E[X|Y] = 3.5 + Y$$

Today's plan

More expectations of sums of random variables

Conditional Expectation

Law of Total Expectation

Mixing discrete and continuous random variables

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:

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

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

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

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)

$$= E[X]$$

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 $E[Y]$?

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 $E[Y]$?

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

$$E[Y|X = 1] = 3$$

When $X = 1$, return 3.

Analyzing recursive code

If Y discrete

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

```python
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 $E[Y]$?

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

$E[Y|X = 1] = 3$

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

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

A. $E[5] + Y$

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

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 $E[Y]$?

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

$E[Y|X=1] = 3$

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

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

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

Analyzing recursive code

If Y discrete

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

```
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 $E[Y]$?

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

$E[Y|X=1] = 3$ $\quad$ $E[Y|X=2] = 5 + E[Y]$ $\quad$ $E[Y|X=3] = 7 + E[Y]$

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

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

$$E[Y] = 15$$

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:

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

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

Today's plan

More expectations of sums of random variables

Conditional Expectation

Law of Total Expectation

Mixing discrete and continuous random variables

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:

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

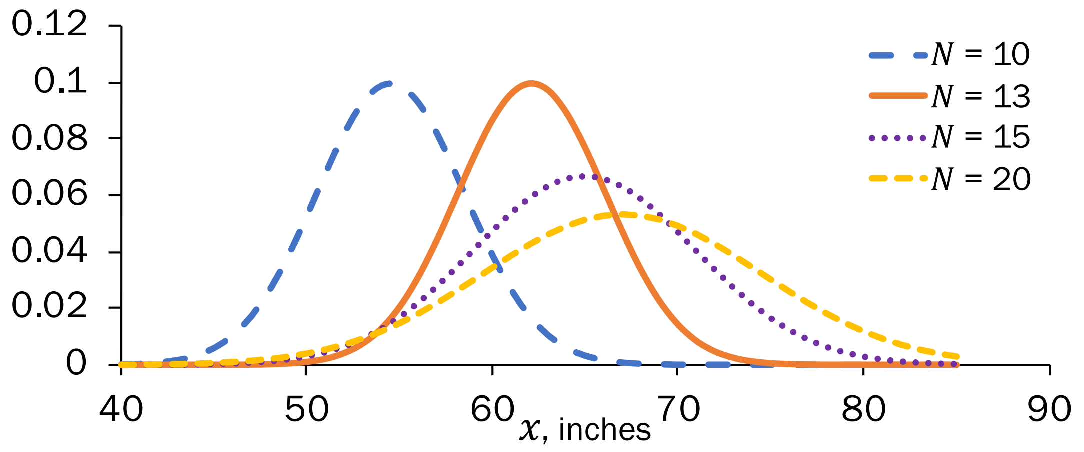
The conditional PMF of N given X is:

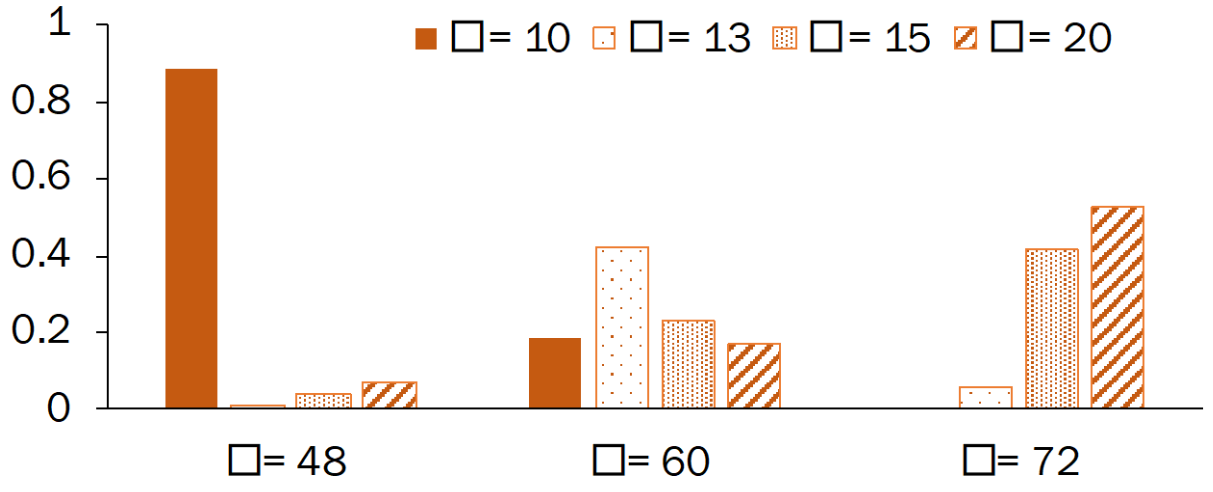
$$p_{N|X}(n|x)$$

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, 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:

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

Intuition:

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

Bayes in all its forms

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

Preview for next time

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.

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. $\text{Uni}(0,1)$
B. $\text{Bin}(n+m, x)$
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: $f_{X|N}(x|n)$

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

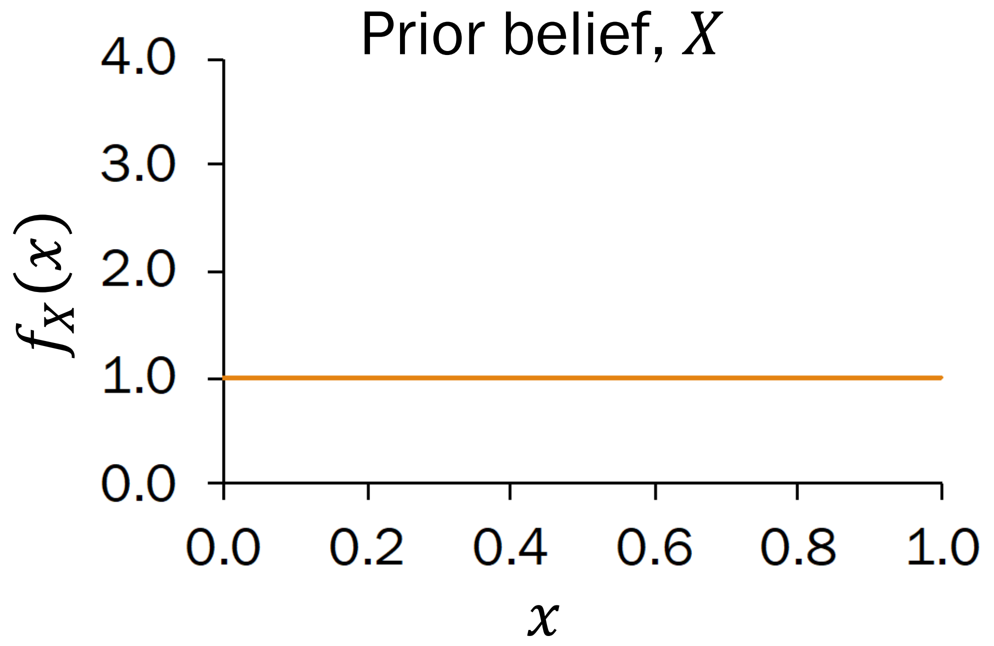
$$= \frac{\binom{n+m}{n}}{p_N(n)} x^n (1-x)^m = \frac{1}{c} x^n (1-x)^m, \text{ where } c = \int_0^1 x^n (1-x)^m dx$$

constant, doesn't depend on x

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:

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