CSCI 3022

intro to data science with probability & statistics

Lecture 6
February 2, 2018

0. HW1 due by 5PM today to Moodle.

1. Bayes' Rule
2. Random variables and probability distributions

Department of Computer Science
UNIVERSITY OF COLORADO BOULDER

Dan Larremore

Bayes' Rule

- Recall the probability that two events intersect: $P(A \cap B) = P(A|B)P(B)$
- But we can write it the other way, too: $P(A \cap B) = P(B|A)P(A)$
- And then… $P(A|B)P(B) = P(B|A)P(A)$ $P(B) > 0$
- And this, is Bayes' Rule: $$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$
- [Bonus: suppose A and B are independent…]

$P(A|B) = P(A)$, $P(B|A) = P(B)$

$\cancel{P(A)} = \frac{\cancel{P(B)}\,\cancel{P(A)}}{\cancel{P(B)}}$

Bayes' Rule + Law of T. P.

- Use Law of Total Probability to rewrite the denominator:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

LTP

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

- Or, if B can be broken into K disjoint events:

$$P(A_1|B) = \frac{P(B|A_1)\,P(A_1)}{P(B)} \overset{\text{LTP}}{=} \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_K)\,P(A_K)}$$

$$= \frac{P(B|A_1)\,P(A_1)}{\sum_{k=1}^{K} P(B|A_k)\,P(A_k)}$$

sum notation

Bayes' Rule + Law of T. P.

- Use Law of Total Probability to rewrite the denominator:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}$$

- Or, if B can be broken into K disjoint events:

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots}$$

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{\sum_{k=1}^{K} P(B|A_k)P(A_k)}$$

Drugs, Prostates, TSA

- One of the most common applications of Bayes' rule is when we develop a test for something, but **the test is not always accurate**.

- Imagine that 1% of math professors are using a drug *Bayes Salts*. There is a test that detects Bayes Salts on the breath of professors 98% of the time, and only gives a false positive 1% of the time.

- Suppose we test Professor Charles Xavier… he's **positive** for Bayes Salts! What is the probability that Professor Xavier is **actually** on Bayes Salts?

Drugs, Prostates, TSA

2% is the False Negative Rate

1% is the False Positive Rate

- 1% of math professors *Bayes Salts* users.
- If on *Bayes Salts*, test says **+** 98% of the time
- If not on *Bayes Salts*, test says **+** only 1% of the time.
- What is probability that Prof. X is a user if he tests positive for *Bayes Salts*?

$$P(\text{user} \mid \text{test} \oplus) \overset{\text{Bayes}}{=} \frac{P(\text{test} \oplus \mid \text{user})\, P(\text{user})}{P(\text{test} \oplus)} \overset{\text{LTP}}{=} \frac{P(\text{test} \oplus \mid \text{user})\, P(\text{user})}{P(\text{test} \oplus \mid \text{user})\, P(\text{user}) + P(\text{test} \oplus \mid \text{not use})\, P(\text{not use})}$$

$$= \frac{0.98 \cdot \cancel{0.01}}{0.98 \cdot \cancel{0.01} + \cancel{0.01} \cdot 0.99} = \frac{0.98}{0.98 + 0.99} = \boxed{0.497}$$

Teach the controversy! 🫢

- Should we test men for prostate cancer?
- Bayes' Rule allows us to write down the probability that someone who tests positive for prostate cancer *actually has* prostate cancer.
- False positives may cause huge amounts of stress, heartache, and even unnecessary surgery!
- On the other hand, if you don't test for cancer, you may not discover it until it's too late.
- Things are slightly more complicated than this: age, PSA cutoffs, etc.

Flipping around a previous problem

- **Suppose I have two bags of marbles.** The first bag contains 6 white marbles and 4 black marbles. The second bag contains 3 white marbles and 7 black marbles. Now suppose I put the two bags in a box. If I close my eyes, grab a bag from the box, reach into the bag and pull out a white marble. What is the probability that I picked Bag 1?

$$P(\text{bag }1 \mid \text{white}) \overset{\text{Bayes}}{=} \frac{P(\text{white} \mid \text{bag }1)\,P(\text{bag }1)}{P(\text{white})}$$

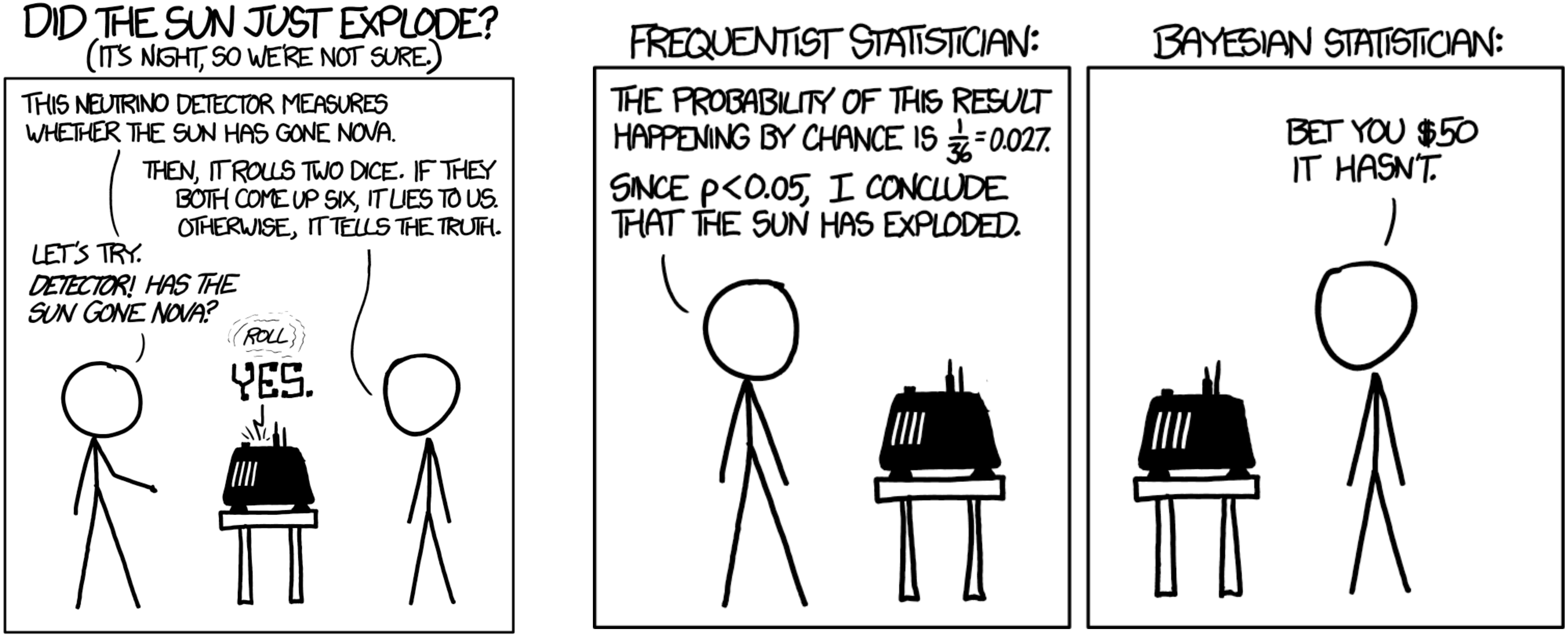
$$\overset{\text{LTP}}{=} \frac{P(\text{white} \mid \text{bag }1)\,P(\text{bag }1)}{P(\text{white} \mid \text{bag }1)\,P(\text{bag }1) + P(\text{white} \mid \text{bag }2)\,P(\text{bag }2)}$$

$$= \frac{6/10 \cdot 1/2}{6/10 \cdot 1/2 + 3/10 \cdot 1/2} = \frac{6}{9} = \boxed{\frac{2}{3}}$$

Bayes' Rule in the wild ☀️

xkcd
v funny

- Bayes' Rule is very helpful because it helps us incorporate our **prior knowledge** about probabilities into our conclusions.



Bayes' Rule in the wild

- Bayes' Rule is very helpful because it helps us incorporate our **prior knowledge** about probabilities into our conclusions.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

prior knowledge.

- When we calculated the probability that Professor X was on Bayes Salts, which one of these terms was our prior knowledge of the background rate of Salts use?

1% $\quad P(\text{Bayes Salts})$

Bayes' Rule in machine learning

- Often, we have a **model with parameters** M and we have **data** D.

- Our goal is to learn the parameters M from the data. Yet we also have some beliefs about the parameters, and no particular beliefs about the data.

Likelihood

Prior beliefs about parameters.

$$P(M|D) = \frac{P(D|M)\,P(M)}{P(D)}$$

posterior distribution

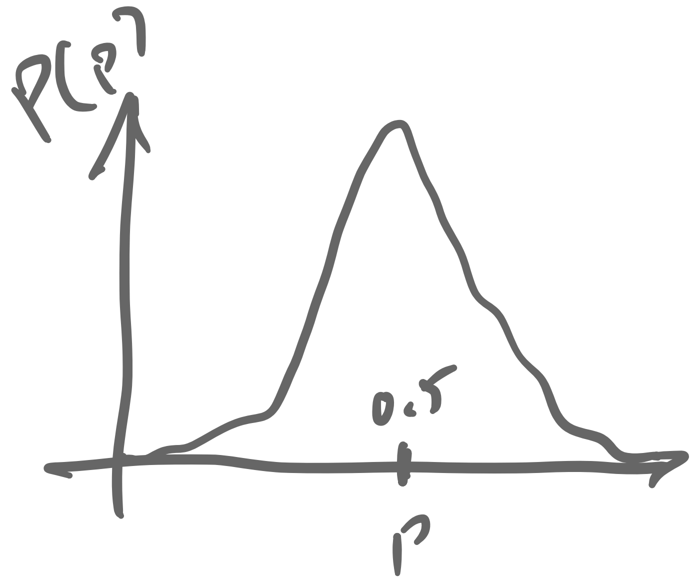
D data.

Flip coin 9H, 1T

Compute Likelihood data | p

but, I tell you, "I'm pretty sure it's not biased."

Prior suggest p=0.5



Random variables

- Say I roll two dice.
- What's the most likely outcome? all same
- What's the most likely sum?

  7

Random variables

$\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$

- Say I roll two dice. What's the sample space? What are the tables of sums, differences, and maxima?

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

sums

$P(\text{sum}=7) = \frac{6}{36} = \frac{1}{6}$

$P(\text{sum}=2) = \frac{1}{36}$

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

Random variables

- **The key**: the dice are random, so the sum is random!
- Let's *sidestep* the sample space entirely and just go straight to the thing we care about: the sum.
- We call the sum of the dice a *random variable*.

Discrete random variables

- **Definition**: a discrete random variable is a function that maps the elements of a sample space Ω to a finite number of values $a_1, a_2, \ldots, a_N$ or an infinite number of values $a_1, a_2, \ldots$

- Note: even if there are an infinite number, the values must be discrete.

- **Examples** of discrete random variables:
  - Sum of dice; difference of dice; maximum of dice.
  - Number of flips until we get a heads.

Probability mass functions

- **Definition**: a *probability mass function* is the map between the random variable's values and the probabilities of those values.

$$f(a) = P(X = a)$$

PMF Random variable outcome

- Called a "probability mass" function (PMF) because each of the random variable's values has some probability mass (or weight) attached to it.

- Since the PMF is a probability function, what is the sum of all the masses?

$P(\Omega) = \underline{1}$ (def'n of prob. function)

$a \in \Omega$

a is in the set omega

\in

$$\sum_{a \in \Omega} f(a) = 1$$

Probability mass functions

- **Question**: what is the probability mass function for the number of coin flips until a biased coin comes up heads?

Want: function f(n) that tells me the prob that it takes n flips to get a H.

$f(1) = p$

$f(2) = (1-p)p$

$f(3) = (1-p)(1-p)p = (1-p)^2 p$

$f(n) = (1-p)^{n-1} p$

Cumulative distribution functions

- **Definition**: a *cumulative distribution function* (CDF) is a function whose value at point *a* is the cumulative sum of probability masses up until *a*.

$$\underbrace{F}_{\text{CDF}}(a) = P(X \leq a)$$

- **Question**: what's the relationship between the PMF and the CDF?

$$\underbrace{F(a)}_{\text{CDF}} = \underbrace{\sum_{x=-\infty}^{a}}_{\text{everything up to } a} \underbrace{f(x)}_{\text{PMF}}$$

Cumulative distribution functions

- **Example**: What is the probability that I roll two dice and they add up to at most 9?

$$F(9) = P(X \leq 9) = \sum_{y=-\infty}^{9} f(y) = \sum_{y=2}^{9} f(y)$$

$$= f(2) + f(3) + f(4) + \ldots + f(9)$$

$= \ldots$ do this

$$= \frac{5}{6}$$

Cumulative distribution functions

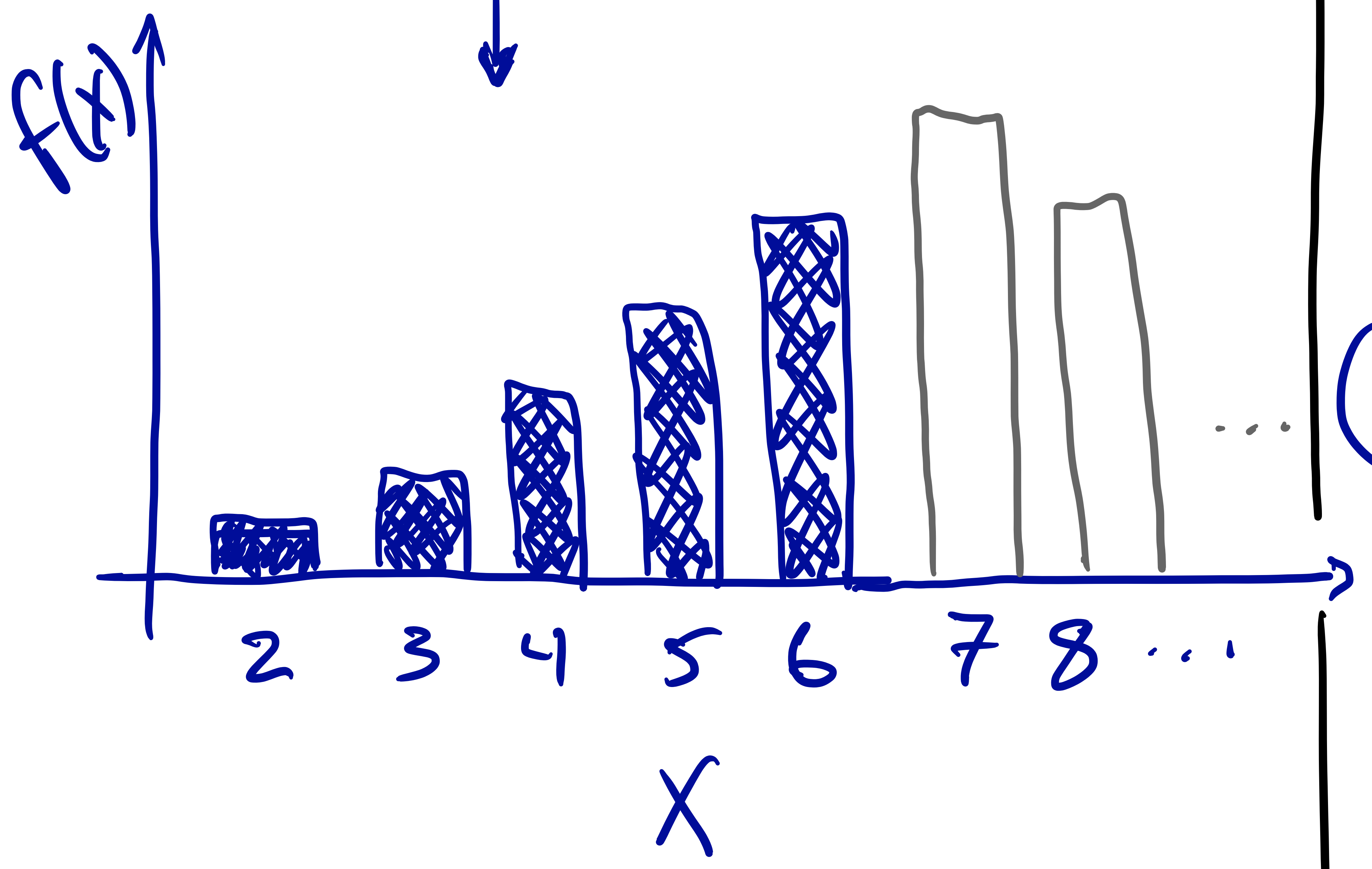
- **Example**: What is the probability that I roll two dice and they add up to between 3 and 6, inclusive?

Let X = the sum of two dice.

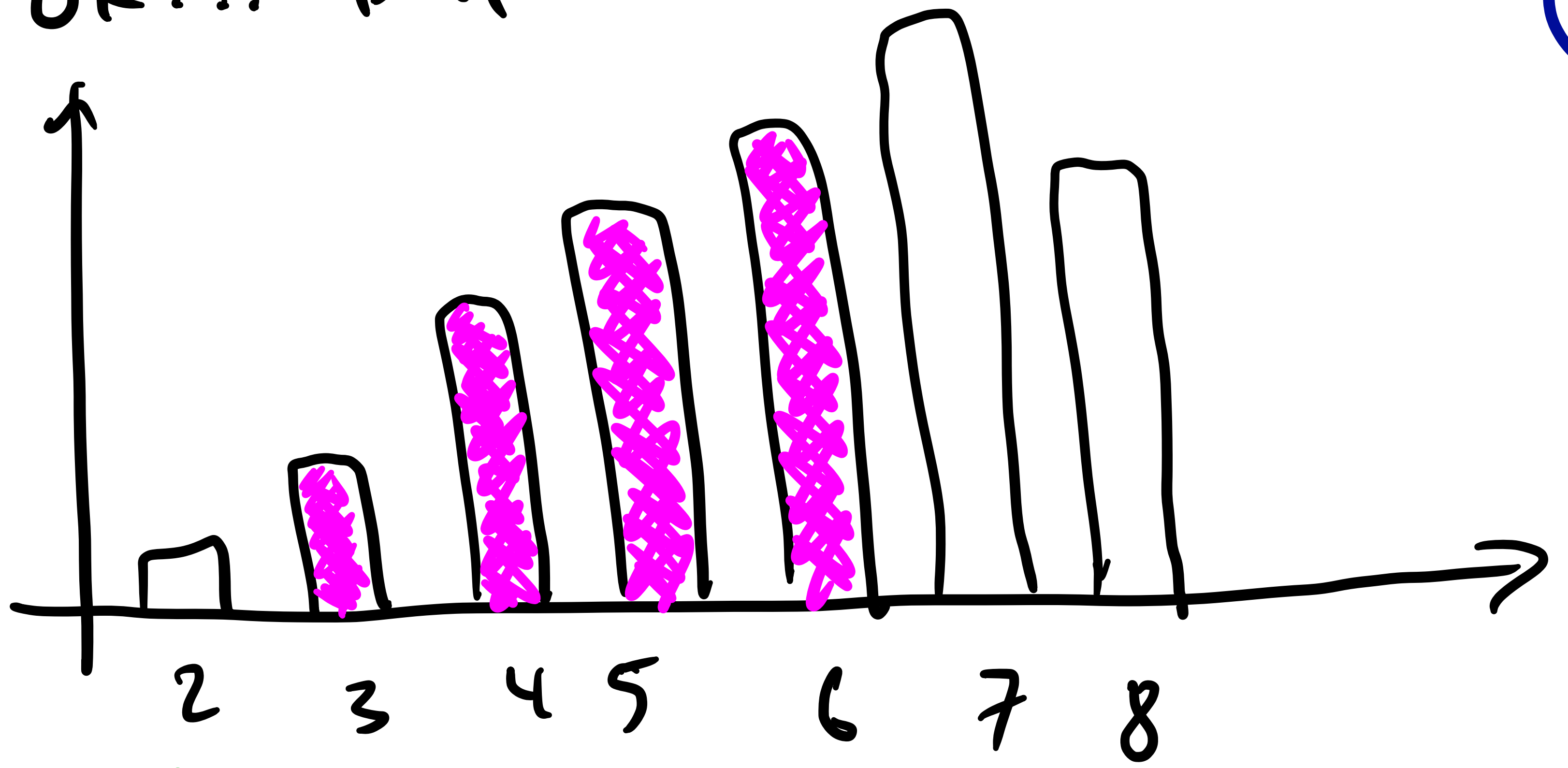
We want $P(X \le 6 \text{ and } X \ge 3)$

1 CDF gives us this:

$P(X \le 6) = F(6)$



ok... but we want this: 2



[More on next slide!]

3 so... we want $F(6)$, but minus this guy.

This guy is $F(2)$, however! so...

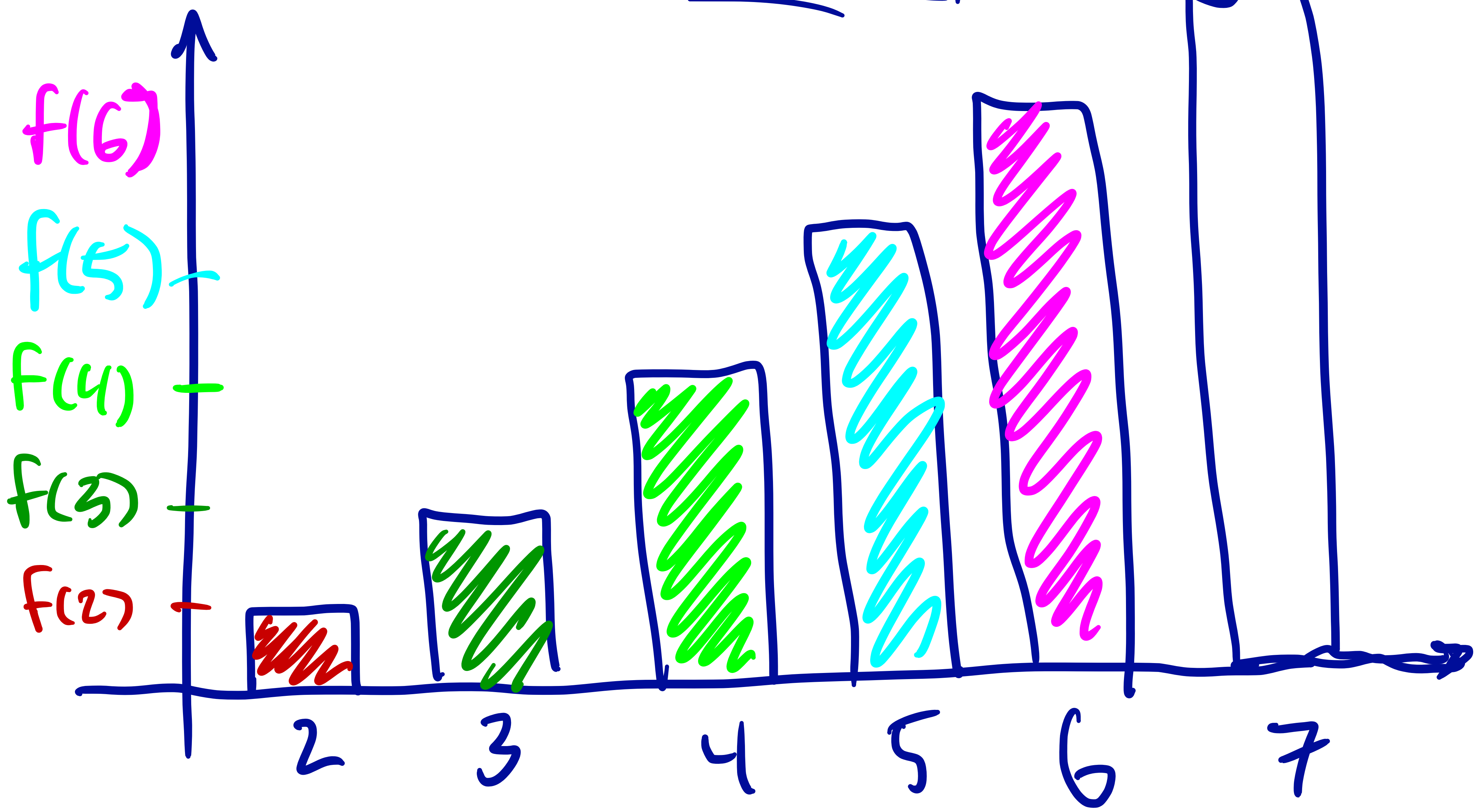
$F(6) - F(2)$ gives us what we want.

4 $P(X \le 6 \text{ and } X \ge 3) = F(6) - F(2)$

Cumulative distribution functions

- **Example**: What is the probability that I roll two dice and they add up to between 3 and 6, inclusive?

What are we actually doing?



$$F(6) = \sum_{k=2}^{6} f(k) = f(2) + f(3) + f(4) + f(5) + f(6)$$

$$F(2) = \sum_{k=2}^{2} f(k) = f(2)$$

So $F(6) - F(2)$ means that $f(2)$s cancel!

leaving $\underbrace{f(3) + f(4) + f(5) + f(6)}_{\text{this is what we wanted.}}$ ✓