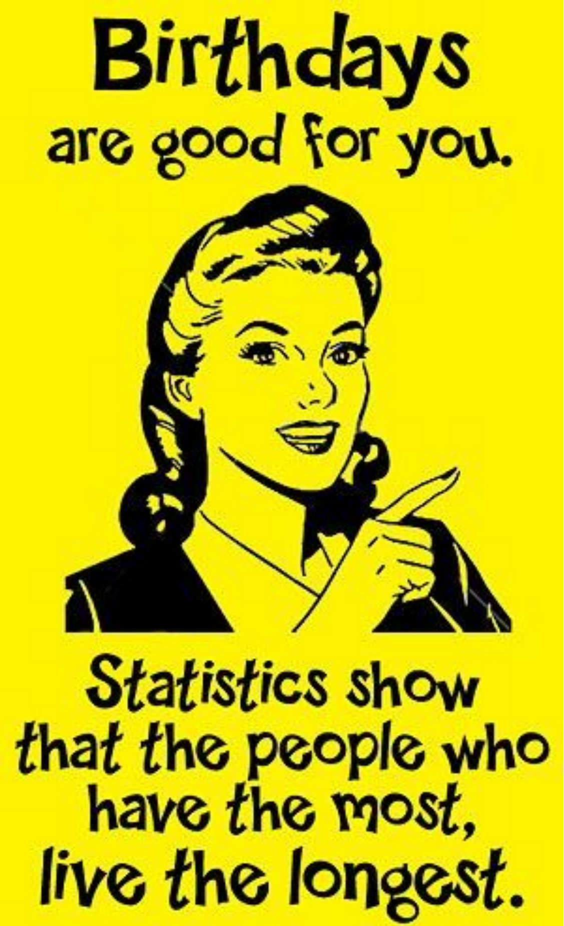
Lecture 6: LTP, Bayes' Theorem, and Random Variables



Announcements and reminders

- HW 2 posted. Due Friday 15 February at 5 PM

Previously, on CSCI 3022…

Conditional probability: The probability that A occurs **given** that C has occurred is

$$P(A \mid C) = \frac{P(A \cap C)}{P(C)}$$

$P(B1 \cap R) = P(R \cap B1)$

Product rule: $P(A \cap C) = P(A \mid C)\, P(C)$

$= P(R \mid B1) P(B1)$

Independence: events A and B are independent if and only if

1) $P(A \mid B) = P(A)$
2) $P(B \mid A) = P(B)$
3) $P(A \cap B) = P(A)\, P(B)$

Law of total probability: If $C_1, C_2, \ldots, C_m$ are disjoint events such that $C_1 \cup C_2 \cup \ldots C_m = \Omega$. Then the probability of an arbitrary event A can be expressed as:

$$P(A) = P(A \mid C_1)\, P(C_1) + P(A \mid C_2)\, P(C_2) + \ldots + P(A \mid C_m)\, P(C_m)$$

Let's flip things around…

Example: S'pose we have two boxes filled with green and red balls.

Box 1: 2 green balls, 7 red balls. Box 2: 4 green balls, 3 red balls

Paul selects a ball by first choosing one of the two boxes at random. He then selects one of the balls in this box at random. If Paul has selected a red ball, what is the probability that he picked from Box 1?

$P(B1) = 1/2$ (w/out any color information)

$$\rightarrow P(B1|R) \overset{\text{Def. of cond. prob.}}{=} \frac{P(B1 \cap R)}{P(R)} \overset{\text{Product Rule}}{=} \frac{P(R|B1)\,P(B1)}{P(R)}$$

$$\overset{\text{LTP}}{=} \frac{P(R|B1)\,P(B1)}{P(R|B1)P(B1) + P(R|B2)P(B2)}$$

Let’s flip things around…

Example: S’pose we have two boxes filled with green and red balls.

Box 1: 2 green balls, 7 red balls. Box 2: 4 green balls, 3 red balls

Paul selects a ball by first choosing one of the two boxes at random. He then selects one of the balls in this box at random. If Paul has selected a red ball, what is the probability that he picked from Box 1?

We want: $P(B1 \mid R)$ … which def. of cond. prob. gives as: $$P(B1 \mid R) = \frac{P(B1 \cap R)}{P(R)}$$

$\rightarrow$ $P(B1 \cap R) = P(R \mid B1)\, P(B1)$, and the LTP gives $P(R)$ as:

$$P(R) = P(R \mid B1)\, P(B1) + P(R \mid B2)\, P(B2)$$

Let's flip things around…

$P(B1) = 1/2$

Example: S'pose we have two boxes filled with green and red balls.

Box 1: 2 green balls, 7 red balls. Box 2: 4 green balls, 3 red balls

Paul selects a ball by first choosing one of the two boxes at random. He then selects one of the balls in this box at random. If Paul has selected a red ball, what is the probability that he picked from Box 1?

Putting it all together:

$$P(B1 \mid R) = \frac{P(B1 \cap R)}{P(R)} = \frac{\underline{P(R \mid B1)P(B1)}}{P(R \mid B1)P(B1) + P(R \mid B2)P(B2)}$$

$$= \frac{7/9 \cdot 1/2}{7/9 \cdot 1/2 + 3/7 \cdot 1/2} = \frac{7/18}{7/18 + 3/14} = \frac{7/18}{152/252} = \frac{49}{76}$$

$> 1/2$

Bayes' Theorem

The notion of using evidence (ball is Red) to update our belief about an event (that Paul picked from Box 1) is the cornerstone of **Bayesian Reasoning**. It is powerful magic.

The formulas we derived in the previous example are called **Bayes' Rule** or **Bayes' Theorem**:

$$P(A \mid C) = \frac{P(C \mid A)\ P(A)}{P(C)} = \frac{P(C \mid A)\ P(A)}{P(C \mid A)\ P(A) + P(C \mid A^c)\ P(A^c)}$$

LTP on denom.

$$P(A|C) = \frac{P(C|A)}{P(C)} \cdot P(A)$$

updated belief about A, GIVEN the data C

Bayesian update—assimilates the data C, & updates your belief about A

your belief about A w/out the data, C

Bayes' Theorem

The notion of using evidence (ball is Red) to update our belief about an event (that Paul picked from Box 1) is the cornerstone of **Bayesian Reasoning**. It is powerful magic.

The formulas we derived in the previous example are called **Bayes' Rule** or **Bayes' Theorem**:

$$P(A \mid C) = \frac{P(C \mid A)\ P(A)}{P(C)} = \frac{P(C \mid A)\ P(A)}{P(C \mid A)\ P(A) + P(C \mid A^c)\ P(A^c)}$$

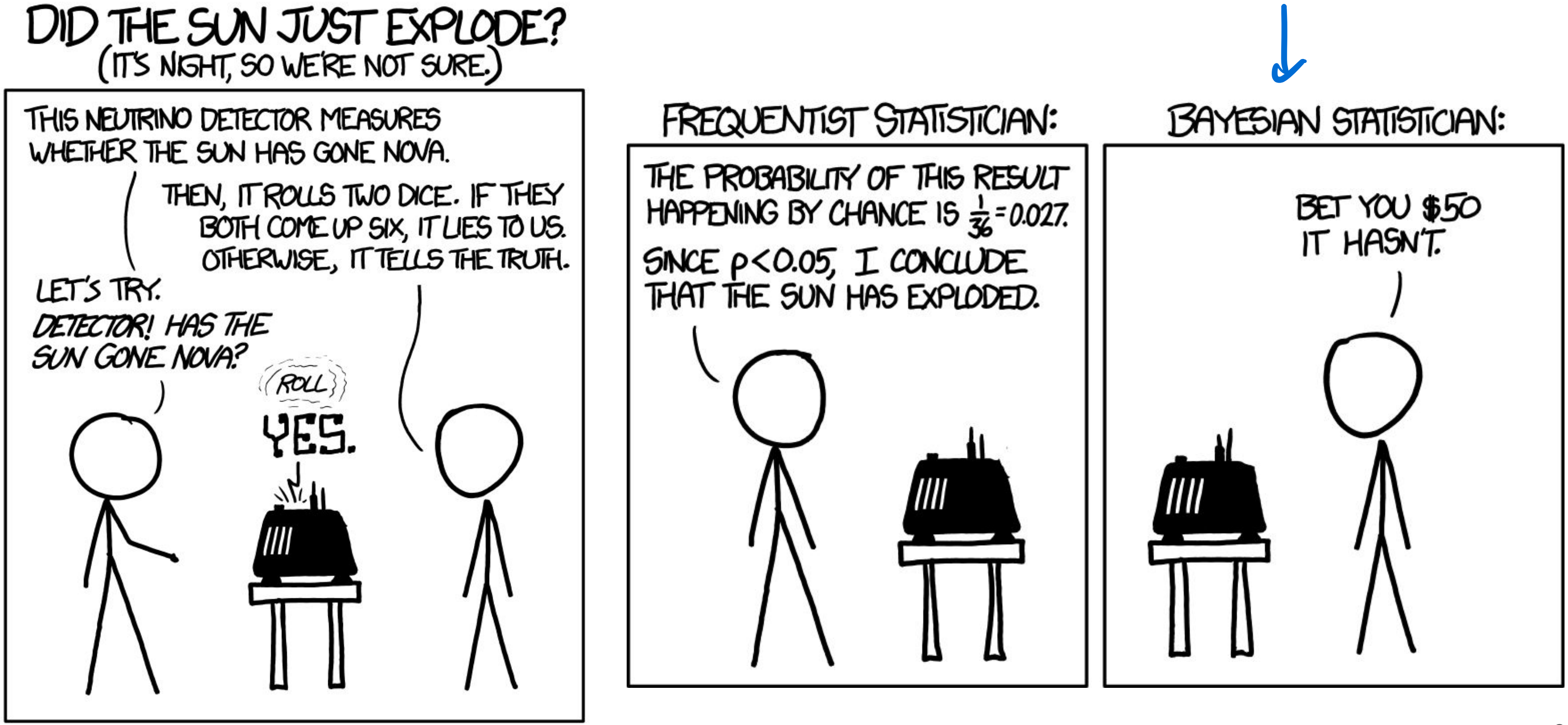
P(A | C) = **posterior distribution**

P(A) = **prior distribution** (before you have knowledge of C)

$\boxed{\text{P(C | A)}}$ = **likelihood function** how likely is the data C if event A was known?

$\boxed{\text{P(C) = evidence}}$ ← usually use LTP to get this

Two flavors of statistics



Bayes' Theorem

Has applications all over science. For example…

- Should we test men for prostate cancer? / women for breast cancer?
- Allows us to write down the probability that someone who tests positive for cancer ***actually has*** cancer… which is super, duper important!
- False positives causes stress, heartache and pain
  - ***Stuff You Should Know:*** podcast link ← if you aren't familiar, **fix that**
- But not diagnosing cancer until it's too late is bad too

Bayes' Theorem

C = event person actually has cancer

+ = event they test pos.

$P(C) = 0.01$ (prior)

Example: S'pose that 1% of men over the age of 40 have prostate cancer. Also s'pose that a test for prostate cancer exists with the following properties:

- 90% of people who have cancer will test positive → $P(+|C) = 0.90$
- 8% of people who do not have cancer will also test positive → $P(+|C^c) = 0.08$

What is the probability that a person who tests positive for cancer **actually has cancer**?

→ want: $P(C|+)$? (posterior prob.)

$$P(C|+) \overset{\text{Bayes}}{=} \frac{P(+|C)\,P(C)}{P(+)} \overset{\text{LTP}}{=} \frac{P(+|C)\,P(C)}{P(+|C)P(C) + P(+|C^c)\,P(C^c)}$$

$$= \frac{0.90 \times 0.01}{0.90 \times 0.01 + 0.08 \times 0.99} = 0.10$$

(10%)

$P(C) + P(C^c) = 1$

$\rightarrow P(C^c) = 1 - P(C)$

$= 0.99$

Bayes' Theorem

Example: S'pose that 1% of men over the age of 40 have prostate cancer. Also s'pose that a test for prostate cancer exists with the following properties:

- 90% of people who have cancer will test positive $P(+|c) = 0.9$
- 8% of people who do not have cancer will also test positive $P(+|c') = 0.8$

What is the probability that a person who tests positive for cancer **actually has cancer**?

$$P(c|+) = \frac{P(+|c)\,P(c)}{P(+)} = \frac{P(+|c)\,P(c)}{P(+|c)\,P(c) + P(+|c')\,P(c')}$$

$$P(c') = 1 - P(c) = 1 - 0.01 = 0.99$$

$$= \frac{0.9 \times 0.01}{(0.9 \times 0.01) + (0.08)(0.99)}$$

Random variables

fair

S'pose I roll two dice

- What is the most likely combination?
- What is the most likely sum?

What is the sample space?

$\Omega = W_1 \times W_2 = 36 = \{(1,1), (1,2), (1,3) \ldots (6,6)\}$

1

Random variables

S'pose I roll two dice

- What is the most likely combination?
- What is the most likely sum?

What is the sample space? Let ω_1 = outcome of 1st roll, and ω_2 = outcome of 2nd roll

$$\Omega = \omega_1 \times \omega_2 = \{(1,1), (1,2), (1,3), \ldots \text{ (36 total outcomes)}\}$$

$$P((1,3)) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

$$P(\omega_1 = 1) \times P(\omega_2 = 3)$$

die 2 (ω_2) \ die 1 (ω_1)	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

outcomes for X = r.v. for the sum

Random variables

What is the sample space? Let ω_1 = outcome of 1st roll, and ω_2 = outcome of 2nd roll

The dice are random, so the sum is also random.

So we can sidestep the sample space entirely and just go straight for the thing we care about:

→ the sum!

We call the sum of the dice a **random variable**

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

Examples: - sum of the dice, difference of the dice, maximum of the dice, …

- number of coin flips until we get a heads, number of heads in n flips, …

Probability mass function

Definition: A **probability mass function** (pmf) is the map between the random variable's values and the probabilities of those values.

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

- Called a "probability mass function" because each of the random variables' values has some **probability mass** (or weight) associated with it

- Because the pmf is a probability function, the sum of all the masses must be… ???

Probability mass function

Definition: A **probability mass function** (pmf) is the map between the random variable's values and the probabilities of those values.

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

- Called a "probability mass function" because each of the random variables' values has some **probability mass** (or weight) associated with it

- Because the pmf is a probability function, the sum of all the masses must be… ???

$$\sum_{i=1}^{n} f(a_i) = 1$$

one of these outcomes must've happened for X

Probability mass function

Example: What is the pmf for the number of coin flips until a biased coin ($P(H) = p$) comes up heads?

Need *f* that maps possible random variable values to probabilities

Let the random variable X = # flips before a Heads

Keep track in a chart!

X	$f(X)$
1	p
2	$(1-p)p$
3	$(1-p)^2 p$
4	
$\vdots$	
n	$(1-p)^{n-1} p$

pmf for X

$f(X=1) = P(H \text{ on first flip}) = p$

$f(X=2) = P(T \text{ on flip } 1 \cap H \text{ on flip } 2)$

$= P(T) \times P(H) = (1-p)p$

$f(X=3) = P(T) \times P(T) \times P(H)$

Probability mass function

Example: What is the pmf for the number of coin flips until a biased coin (P(H) = p) comes up heads?

Need *f* that maps possible random variable values to probabilities

Let the random variable X = # flips before a Heads

State space: $\Omega = \{H, TH, TTH, TTTH, \dots\}$

Associated r.v.: $X = 1, \ 2, \ 3, \ \ 4, \dots$

pmf: $f(X) = p, \ (1-p)p, \ (1-p)^2p, \ (1-p)^3p \dots$

Cumulative distribution function

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

$$F(a) = P(X \leq a)$$

$f(x=1) = 1/6$

$f(2) = 1/6$

$\vdots$

Example: If I roll a single fair die, what is the cdf?

$F(x=1) = P(X \leq 1) = 1/6$

$F(x=2) = P(X \leq 2) = f(x=1) + f(x=2) = 1/6 + 1/6 = 2/6$

$\vdots$

$F(x=6) = P(X \leq 6) = f(x=1) + f(x=2) + \ldots + f(x=6) = \ldots = 1$

Cumulative distribution function

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

$$F(a) = P(X \leq a)$$

Example: If I roll a single fair die, what is the cdf?

$F(1) = 1/6$ $F(2) = 2/6$ $F(3) = 3/6$ … $F(6) = 6/6 = 1$

(with probability 1, you will roll a number ≤ 6)

Question: What is the relationship between the pmf and the cdf?

HERE

Cumulative distribution function

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

$$F(a) = P(X \leq a)$$

Example: If I roll a single fair die, what is the cdf?

F(1) = 1/6   F(2) = 2/6   F(3) = 3/6   …   F(6) = 6/6 = 1

(with probability 1, you will roll a number ≤ 6)

Question: What is the relationship between the pmf and the cdf?

$$F(a) = \sum_{x \leq a} f(x)$$

CDF

Cumulative distribution function

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

X = r.v. for sum of the two dice

$$P(X \geq 9) = P(X=9) + P(X=10) + P(X=11) + P(X=12)$$
$$= \ldots$$

Make it easier!

Know: $1 = \underbrace{P(X < 9)}_{F(8)} + P(X \geq 9)$

$$\rightarrow P(X \geq 9) = 1 - F(8)$$

survival function = 1 - CDF

Cumulative distribution function

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

Solution:

Asking for $P(X = 9 \text{ or } 10 \text{ or } 11 \text{ or } 12) = P(X \geq 9)$

$= 1 - P(X < 9)$

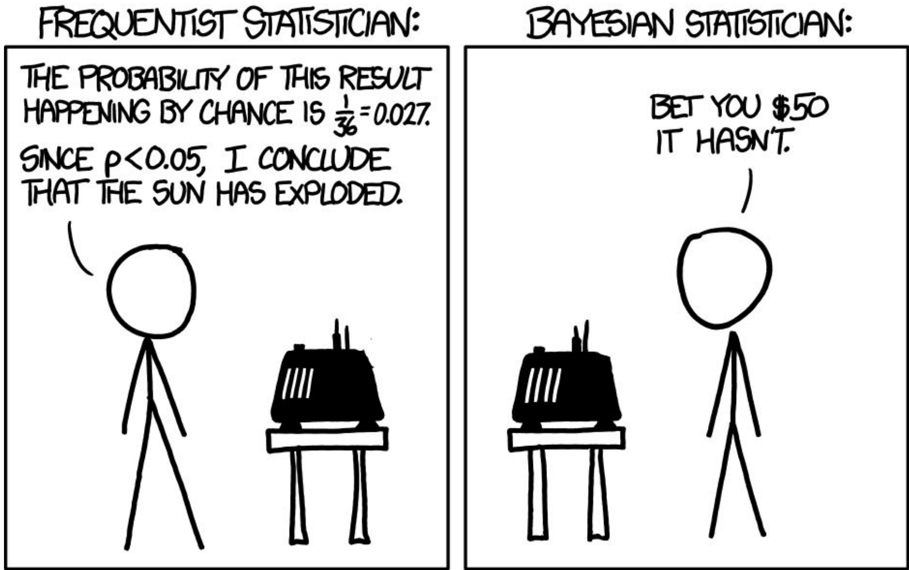
$= 1 - P(X \leq 8)$

$= 1 - F(8)$

$= \ldots = 10/36$

What just happened?

- We learned about **Bayes' Theorem**! It is **important** and it is **awesome.**
- And distributions!
  - Probability mass function: what's the probability of ***this*** event?
  - Cumulative distribution function: what's the **total** probability of everything up to this event?



Next time:

- We learn about some **specific** probability distributions
- Important so we can **model** and **simulate** things!