Conditional Probability

Varun Mahadevan

1

conditional probability

Conditional probability of E given F: probability that E occurs *given* that F has occurred.

"Conditioning on F"

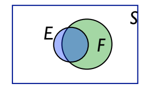
Written as P(E|F)

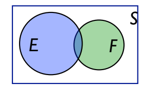
Means "P(E, given F observed)"

Sample space S reduced to those elements consistent with F (i.e. $S \cap F$)

Event space E reduced to those elements consistent with F (i.e. $E \cap F$)

With equally likely outcomes,





$$P(E \mid F) = \frac{\#\text{ of outcomes in } E \text{ consistent with } F}{\#\text{ of outcomes in } S \text{ consistent with } F} = \frac{|EF|}{|SF|} = \boxed{\frac{|EF|}{|F|}}$$

$$P(E \mid F) = \frac{|EF|}{|F|} = \frac{|EF|/|S|}{|F|/|S|} = \boxed{\frac{P(EF)}{P(F)}}$$

2

coin flipping

Suppose you flip two coins & all outcomes are equally likely.

What is the probability that both flips land on heads if…

- The first flip lands on heads?

Let B = {HH} and F = {HH, HT}

P(B|F) = P(BF)/P(F) = P({HH})/P({HH, HT})

= (1/4)/(2/4) = 1/2

- At least one of the two flips lands on heads?

Let A = {HH, HT, TH}, BA = {HH}

P(B|A) = |BA|/|A| = 1/3

- At least one of the two flips lands on tails?

Let G = {TH, HT, TT}

P(B|G) = P(BG)/P(G) = P(∅)/P(G) = 0/P(G) = 0

3

Exampls

2 random cards are selected from a deck of cards:

- What is the probability that both cards are aces given that one of the cards is the ace of spades?
- What is the probability that both cards are aces given that at least one of the cards is an ace?

4

conditional probability: the chain rule

General defn: $P(E \mid F) = \frac{P(EF)}{P(F)}$ where P(F) > 0

Holds even when outcomes are *not* equally likely.

What if P(F) = 0?

P(E|F) undefined: (you can't observe the impossible)

For equally likely outcomes:

$$P(E \mid F) = \frac{|EF|}{|F|} = \frac{|EF|/|S|}{|F|/|S|} = \frac{P(EF)}{P(F)}$$

5

Conditional Probability

Satisfies usual axioms of probability

Example:

Pr(E | F) = 1- Pr (E^c | F)

6

Conditional Probabilities yield a probability space

Suppose that $(S, Pr(\cdot))$ is a probability space.

Then $(S, Pr(\cdot|F))$ is a probability space for $F \subset S$ with $Pr(F) > 0$

$$0 \leq Pr(w|F) \leq 1$$

$$\sum_{w \in S} Pr(w|S) = 1$$

$E_1, E_2, \ldots, E_n$ disjoint implies

$$Pr(\cup_{i=1}^{n} E_i | F) = \sum_{i=1}^{n} Pr(E_i|F)$$

7

Chain rule application

Draw 2 balls at random without replacement from an urn with 8 red balls and 4 white ones. What is the probability that both balls are red?

8

Chain rule example

Alice and Bob play a game as follows: A die is thrown, and each time it is thrown, regardless of the history, it is equally likely to show any of the six numbers.

If it shows 5, Alice wins.

If it shows 1, 2 or 6, Bob wins.

Otherwise, they play a second round and so on.

What is P(Alice wins on n^{th} round)?

9

Keys

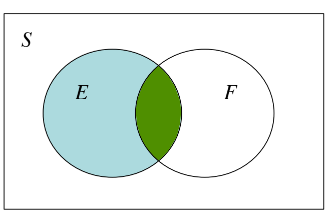
I have n keys, one of which opens a locked door. Trying keys at random without replacement, what is the chance of opening the door on the k^{th} try?

10

law of total probability

E and F are events in the sample space S

$E = EF \cup EF^c$



$EF \cap EF^c = \varnothing$

$\Rightarrow P(E) = P(EF) + P(EF^c)$

11

conditional probability: the chain rule

General defn: $P(E \mid F) = \dfrac{P(EF)}{P(F)}$ where P(F) > 0

Implies: P(EF) = P(E|F) P(F) (*"the chain rule"*)

General definition of Chain Rule:

$$P(E_1 E_2 \cdots E_n) = P(E_1)P(E_2 \mid E_1)P(E_3 \mid E_1, E_2) \cdots P(E_n \mid E_1, E_2, \ldots, E_{n-1})$$

12

law of total probability

$P(E) = P(EF) + P(EF^c)$
$= P(E|F)\,P(F) + P(E|F^c)\,P(F^c)$
$= P(E|F)\,P(F) + P(E|F^c)\,(1-P(F))$

weighted average, conditioned on event F happening or not.

More generally, if $F_1, F_2, \ldots, F_n$ partition S (mutually exclusive, $\bigcup_i F_i = S$, $P(F_i)>0$), then

$P(E) = \sum_i P(E|F_i)\,P(F_i)$

weighted average, conditioned on events F_i happening or not.

(Analogous to reasoning by cases; both are very handy.)

13

total probability

Sally has 1 elective left to take: either Phys or Chem. She will get A with probability 3/4 in Phys, with prob 3/5 in Chem. She flips a coin to decide which to take.

What is the probability that she gets an A?

P(A) = P(A|Phys)P(Phys) + P(A|Chem)P(Chem)
= (3/4)(1/2)+(3/5)(1/2)
= 27/40

Note that conditional probability was a means to an end in this example, not the goal itself. One reason conditional probability is important is that this is a common scenario.

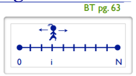
14

gamblers ruin

BT pg. 63

2 Gamblers: Alice & Bob.
A has i dollars; B has (N-i)
Flip a coin. Heads – A wins \$1; Tails – B wins \$1
Repeat until A or B has all N dollars
What is P(A wins)?

Let Ei = event that A wins starting with \$i
Approach: Condition on i^{th} flip



aka "Drunkard's Walk"

nice example of the utility of conditioning: future decomposed into two crisp cases instead of being a blurred superposition thereof

How does p_i vary with i?

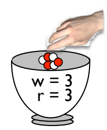
$$
\begin{aligned}
p_i &= \boxed{P(E_i) = P(E_i \mid H)P(H) + P(E_i \mid T)P(T)} \\
p_i &= \frac{1}{2}(p_{i+1} + p_{i-1}) \\
2p_i &= p_{i+1} + p_{i-1} \\
p_{i+1} - p_i &= p_i - p_{i-1} \\
p_2 - p_1 &= p_1 - p_0 = p_1, \text{ since } p_0 = 0
\end{aligned}
$$

So:
$$
\begin{aligned}
p_2 &= 2p_1 \\
&\cdots \\
p_i &= ip_1 \\
p_N &= Np_1 = 1 \\
\boxed{p_i} &\boxed{= i/N}
\end{aligned}
$$

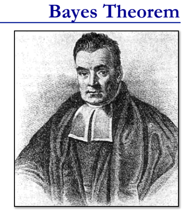
15

Bayes Theorem

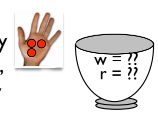
6 red or 3 red/3 white balls in an urn



Probability of drawing 3 red balls, given 3 in urn ?


Rev. Thomas Bayes c. 1701-1761

Probability of only 3 red balls in urn, given that I drew three?



16

Bayes Theorem

Most common form:

$$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E)}$$

Expanded form (using law of total probability):

$$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E \mid F)P(F) + P(E \mid F^c)P(F^c)}$$

Proof:
$$P(F \mid E) = \frac{P(EF)}{P(E)} = \frac{P(E \mid F)P(F)}{P(E)}$$

17

Bayes Theorem

Most common form:

$$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E)}$$

Expanded form (using law of total probability):

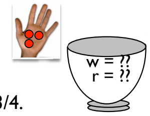
$$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E \mid F)P(F) + P(E \mid F^c)P(F^c)}$$

Why it's important:
Reverse conditioning
P(model | data) ~ P(data | model)
Combine new evidence (E) with prior belief (P(F))
Posterior vs prior

18

Bayes Theorem

An urn contains 6 balls, either 3 red + 3 white or all 6 red.
You draw 3; all are red.
Did urn have only 3 red?



Can't tell

Suppose it was 3 + 3 with probability p=3/4.
Did urn have only 3 red?

M = urn has 3 red + 3 white
D = I drew 3 red

prior = 3/4 ; posterior = 3/23

P(M | D) = P(D | M)P(M)/[P(D | M)P(M)+ P(D | M^c)P(M^c)]
P(D | M) = (3 choose 3)/(6 choose 3) = 1/20
P(M | D) = (1/20)(3/4)/[(1/20)(3/4) + (1)(1/4)] = 3/23

simple spam detection

Say that 60% of email is spam
- 90% of spam has a forged header
- 20% of non-spam has a forged header
- Let F = message contains a forged header
- Let J = message is spam

What is $P(J|F)$?

Solution:

$$\begin{aligned} P(J \mid F) &= \frac{P(F \mid J)P(J)}{P(F \mid J)P(J) + P(F \mid J^c)P(J^c)} \\ &= \frac{(0.9)(0.6)}{(0.9)(0.6) + (0.2)(0.4)} \\ &\approx 0.871 \end{aligned}$$

prior = 60%
posterior = 87%

simple spam detection

Say that 60% of email is spam
- 10% of spam has the word "Viagra"
- 1% of non-spam has the word "Viagra"
- Let V = message contains the word "Viagra"
- Let J = message is spam

What is $P(J|V)$?

Solution:

$$\begin{aligned} P(J \mid V) &= \frac{P(V \mid J)P(J)}{P(V \mid J)P(J) + P(V \mid J^c)P(J^c)} \\ &= \frac{(0.1)(0.6)}{(0.1)(0.6) + (0.01)(1-0.6)} \\ &\approx 0.9375 \end{aligned}$$

prior = 60%
posterior = 94%

DNA paternity testing

Child is born with (A,a) gene pair (event $B_{A,a}$)
- Mother has (A,A) gene pair
- Two possible fathers: M_1 = (a,a), M_2 = (a,A)
- P(M_1) = p, P(M_2) = 1-p

What is P(M_1 | $B_{A,a}$) ?

Solution:

$$\begin{aligned} &P(M_1 \mid B_{Aa}) \\ &= \frac{P(B_{Aa} \mid M_1)P(M_1)}{P(B_{Aa} \mid M_1)P(M_1) + P(B_{Aa} \mid M_2)P(M_2)} \\ &= \frac{1 \cdot p}{1 \cdot p + 0.5(1-p)} = \frac{2p}{1+p} \geq \frac{2p}{1+1} = p \end{aligned}$$

E.g., 1/2 → 2/3

I.e., the given data about child *raises* probability that M_1 is father

HIV testing

Suppose an HIV test is 98% effective in detecting HIV, i.e., its "false negative" rate = 2%. Suppose furthermore, the test's "false positive" rate = 1%.
- 0.5% of population has HIV
- Let E = you test positive for HIV
- Let F = you actually have HIV

What is P(F|E) ?

Solution:

$$\begin{aligned} P(F \mid E) &= \frac{P(E \mid F)P(F)}{P(E \mid F)P(F) + P(E \mid F^c)P(F^c)} \\ &= \frac{(0.98)(0.005)}{(0.98)(0.005) + (0.01)(1-0.005)} \\ &\approx 0.330 \end{aligned}$$

P(E) ≈ 1.5%

Note difference between conditional and joint probability: P(F|E) = 33% ; P(FE) = 0.49%

23

why testing is still good

	HIV+	HIV-
Test +	0.98 = P(E|F)	0.01 = P(E|F^c)
Test -	0.02 = P(E^c|F)	0.99 = P(E^c|F^c)

Let E^c = you test **negative** for HIV
Let F = you actually have HIV

What is P(F|E^c) ?

$$\begin{aligned} P(F \mid E^c) &= \frac{P(E^c \mid F)P(F)}{P(E^c \mid F)P(F) + P(E^c \mid F^c)P(F^c)} \\ &= \frac{(0.02)(0.005)}{(0.02)(0.005) + (0.99)(1-0.005)} \\ &\approx 0.0001 \end{aligned}$$

summary

Conditional probability

P(E|F): Conditional probability that E occurs *given* that F has occurred.

Reduce event/sample space to points consistent w/ F ($E \cap F$; $S \cap F$)

$$P(E \mid F) = \frac{P(EF)}{P(F)} \qquad (P(F) > 0)$$

$P(E \mid F) = \frac{|EF|}{|F|}$, if equiprobable outcomes.

P(EF) = P(E|F) P(F) ("the chain rule")

"P(- | F)" is a probability law, i.e., satisfies the 3 axioms

P(E) = P(E|F) P(F) + P(E|F^c) (1-P(F)) ("the law of total probability")

Bayes theorem

$$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E)}$$

25