4. Conditional Probability

P( | )

CSE 312

Autumn 2011

W.L. Ruzzo

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

"Conditioning on F"

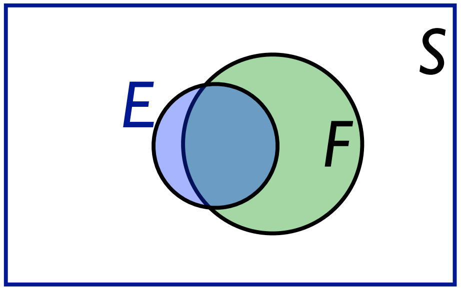
Written as P(E|F)

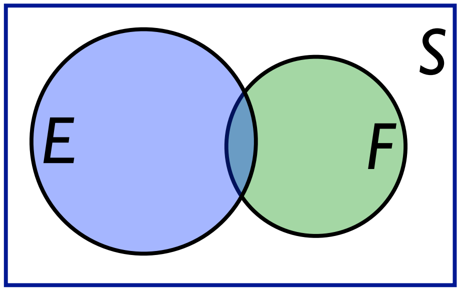
Means "P(E, given F observed)"

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,





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

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

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

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)

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

General definition of Chain Rule:

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

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

Holds even when outcomes are *not* equally likely.

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

Proof:

the idea is simple–the sample space contracts to F; dividing all (unconditional) probabilities by P(F) correspondingly re-normalizes the probability measure – see text for details; better yet, try it!

Ex: $P(A \cup B) \leq P(A) + P(B)$

$\therefore \; P(A \cup B \mid F) \leq P(A \mid F) + P(B \mid F)$

sending bit strings

Bit string with m 0's and n 1's sent on the network

- All distinct arrangements of bits equally likely
- E = first bit received is a 1
- F = k of first r bits received are 1's

What's P(E|F)?

Solution 1:

$$P(E) = \frac{n}{m+n} \qquad P(F) = \frac{\binom{n}{k}\binom{m}{r-k}}{\binom{m+n}{r}}$$

$$P(F \mid E) = \frac{\binom{n-1}{k-1}\binom{m}{r-k}}{\binom{m+n-1}{r-1}}$$

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

Bit string with m 0's and n 1's sent on the network

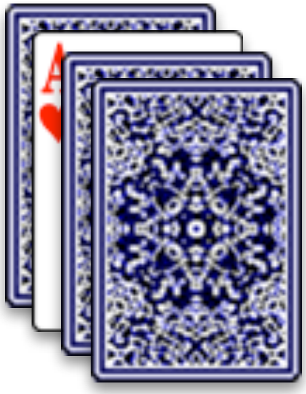
- All distinct arrangements of bits equally likely
- E = first bit received is a 1
- F = k of first r bits received are 1's

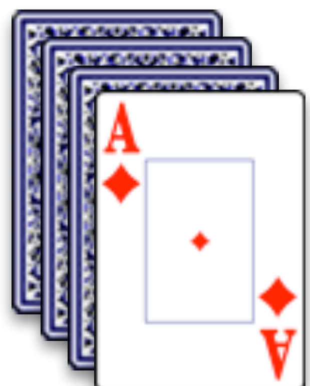
What's P(E|F)?

Solution 2:

- Observe:
  - P(E|F) = P(picking one of k 1's out of r bits)
- So:
  - P(E|F) = k/r

piling cards





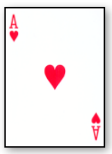
Deck of 52 cards randomly divided into 4 piles

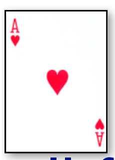
13 cards per pile

Compute P(each pile contains an ace)

Solution:

E_1 = {  in any one pile }

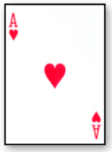
E_2 = {  &  in different piles }

E_3 = {    in different piles }

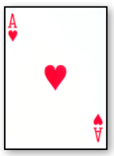
E_4 = { all four aces in different piles }

Compute $P(E_1\ E_2\ E_3\ E_4)$

piling cards

E_1 = {  in any one pile }

E_2 = {  &  in different piles }

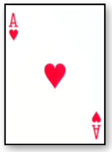
E_3 = {    in different piles }

E_4 = { all four aces in different piles }

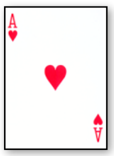
$$P(E_1E_2E_3E_4)$$

$$= P(E_1)\ P(E_2|E_1)\ P(E_3|E_1E_2)\ P(E_4|E_1E_2E_3)$$

piling cards

E_1 = {  in any one pile }

E_2 = {  &  in different piles }

E_3 = {    in different piles }

E_4 = { all four aces in different piles }

$P(E_1E_2E_3E_4) = P(E_1)\ P(E_2|E_1)\ P(E_3|E_1E_2)\ P(E_4|E_1E_2E_3)$

$P(E_1) = 1$

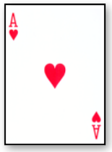
$P(E_2|E_1) = 39/51$ (39 of 51 slots not in AH pile)

$P(E_3|E_1E_2) = 26/50$ (26 not in AS, AH piles)

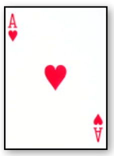
$P(E_4|E_1E_2E_3) = 13/49$ (13 not in AS, AH, AD piles)

A conceptual trick: what's randomized?

a) *randomize* cards, deal *sequentially* into piles
b) *sort* cards, aces first, deal *randomly* into piles.

E_1 = {  in any one pile }

E_2 = {  &  in different piles }

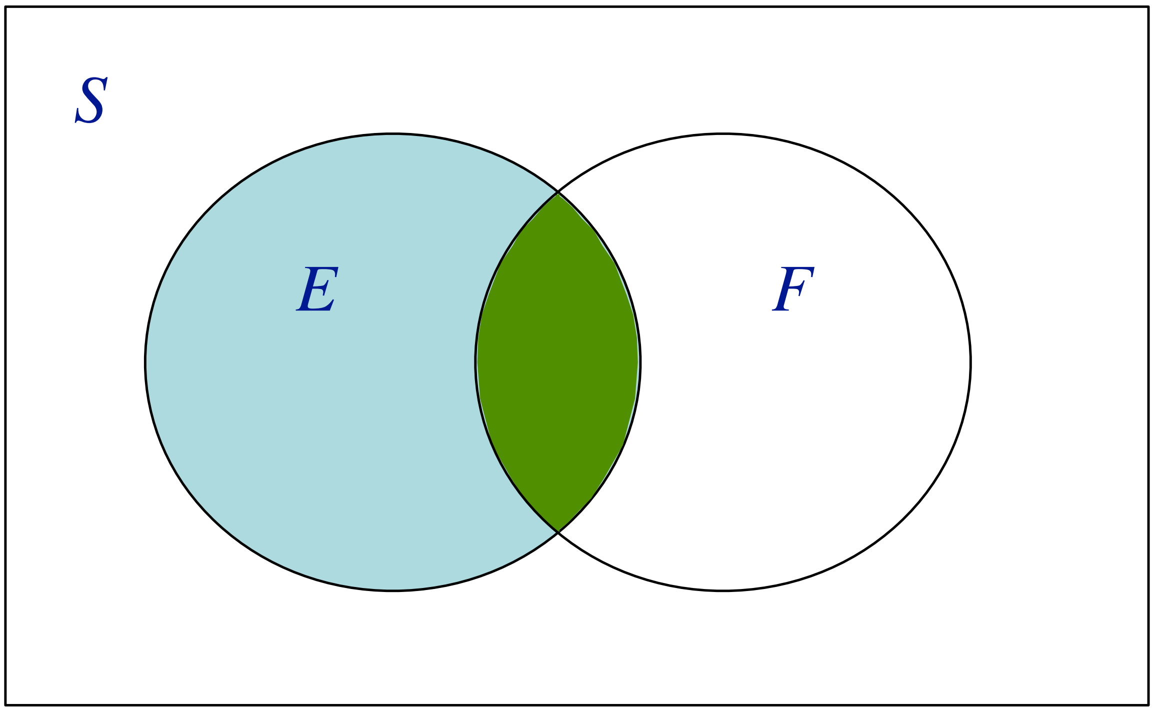
E_3 = {    in different piles }

E_4 = { all four aces in different piles }

$$P(E_1E_2E_3E_4)$$
$$= P(E_1)\ P(E_2|E_1)\ P(E_3|E_1E_2)\ P(E_4|E_1E_2E_3)$$
$$= (39 \bullet 26 \bullet 13)/(51 \bullet 50 \bullet 49)$$
$$\approx 0.105$$

E and F are events in the sample space S

$$E = EF \cup EF^c$$



$$EF \cap EF^c = \emptyset$$

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

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

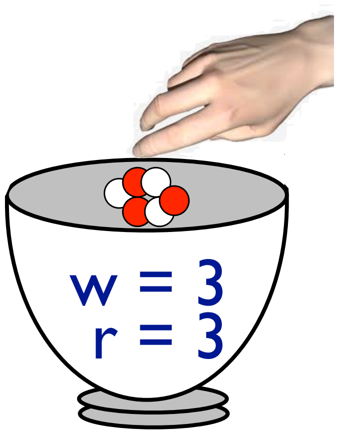
weighted average, conditioned on event F happening or not.

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

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

weighted average, conditioned on events F_i happening or not.

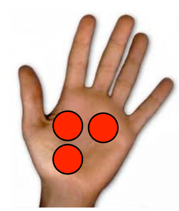
Bayes Theorem

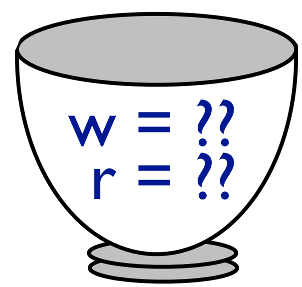


Probability of drawing 3 red balls, given 3 in urn ?

Rev. Thomas Bayes c. 1701-1761

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





Improbable Inspiration: The future of software may lie in the obscure theories of an 18th century cleric named Thomas Bayes

Los Angeles Times (October 28, 1996)
By Leslie Helm, Times Staff Writer

When Microsoft Senior Vice President Steve Ballmer [now CEO] first heard his company was planning a huge investment in an Internet service offering… he went to Chairman Bill Gates with his concerns…

Gates began discussing the critical role of "Bayesian" systems…

source: http://www.ar-tiste.com/latimes_oct-96.html

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:

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

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):

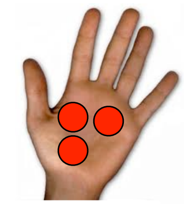
$$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E \mid F)P(F) + P(E \mid F^c)P(F^c)}$$

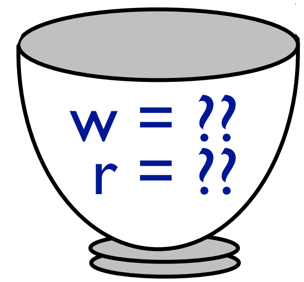
Why it's important:

- Reverse conditioning
- P(model|data) ~ P(data|model)
- Combine new evidence (E) with prior belief (P(F))
- Posterior vs prior

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

$$P(M \mid D) = P(D \mid M)P(M)/[P(D \mid M)P(M) + P(D \mid M^c)P(M^c)]$$

$$P(D \mid M) = (3 \text{ choose } 3)/(6 \text{ choose } 3) = 1/20$$

$$P(M \mid D) = (1/20)(3/4)/[(1/20)(3/4) + (1)(1/4)] = 3/23$$

prior = 3/4 ; *posterior* = 3/23

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:

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

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:

$$P(J \mid V) = \frac{P(V \mid J)P(J)}{P(V \mid J)P(J) + P(V \mid J^c)P(J^c)}$$

$$= \frac{(0.1)(0.6)}{(0.1)(0.6) + (0.01)(1 - 0.6)}$$

$$\approx 0.896$$

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 $P(M_1 \mid B_{A,a})$?

Solution:

$$P(M_1 \mid B_{Aa})$$

$$= \frac{P(B_{Aa} \mid M_1)P(M_1)}{P(B_{Aa} \mid M_1)P(M_1) + P(B_{Aa} \mid M_2)P(M_2)}$$

$$= \frac{1 \cdot p}{1 \cdot p + 0.5(1-p)} = \frac{2p}{1+p} > \frac{2p}{1+1} = p$$

i.e., data about child *raises* probability that M_1 is father

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:

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

	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 $P(F|E^C)$?

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

The *odds* of event E is $P(E)/(P(E^c)$

Example: A = any of 2 coin flips is H:

P(A) = 3/4, $P(A^c)$ = 1/4, so odds of A is 3
(or "3 to 1 in favor")

Example: odds of having HIV:

P(F) = .5% so $P(F)/P(F^c)$ = .005/.995
(or 1 to 199 *against*)

posterior odds from prior odds

F = some event of interest (say, "HIV+")

E = *additional* evidence (say, "HIV test was positive")

Prior odds of F: P(F)/P(F^c)

What are the *Posterior odds* of F: P(F|E)/P(F^c|E) ?

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

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

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

$$\begin{pmatrix}\text{posterior}\\\text{odds}\end{pmatrix} = \begin{pmatrix}\text{``Bayes}\\\text{factor''}\end{pmatrix} \cdot \begin{pmatrix}\text{prior}\\\text{odds}\end{pmatrix}$$

Let E = you test *positive* for HIV
Let F = you actually *have* HIV
What are the posterior odds?

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

$$(\text{posterior odds} = \text{“Bayes factor”} \cdot \text{prior odds})$$

$$= \frac{0.98}{0.01} \cdot \frac{0.005}{0.995}$$

More likely to *test* positive if you *are* positive, so Bayes factor >1; positive test *increases* odds 98-fold, to 2.03:1 against (vs prior of 199:1 against)

posterior odds from prior odds

Let E = you test *negative* for HIV
Let F = you actually *have* HIV
What is the *ratio* between P(F|E) and P(F^c|E) ?

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

$$(\text{posterior odds} = \text{“Bayes factor”} \cdot \text{prior odds})$$

$$= \frac{0.02}{0.99} \cdot \frac{0.005}{0.995}$$

Unlikely to test *negative* if you are *positive*, so Bayes factor <1; negative test *decreases* odds 49.5-fold, to 9850:1 against (vs prior of 199:1 against)

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 are posterior odds that a message containing "Viagra" is spam ?

Solution:

$$\begin{aligned}
\frac{P(J \mid V)}{P(J^c \mid V)} &= \frac{P(V \mid J)}{P(V \mid J^c)} \frac{P(J)}{P(J^c)} \\
(\text{posterior odds} &= \text{“Bayes factor”} \cdot \text{prior odds}) \\
15 &= \frac{0.10}{0.01} \cdot \frac{0.6}{0.4}
\end{aligned}$$