4. Conditional Probability

BT 1.3, 1.4

P( | )

CSE 312

Autumn 2013

W.L. Ruzzo

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

"Conditioning on F"

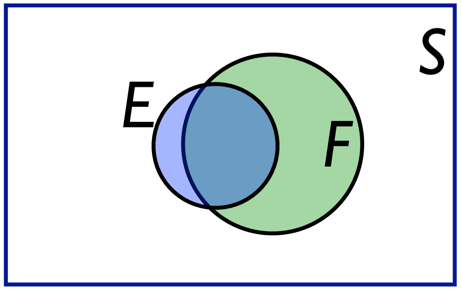
Written as P(E|F)

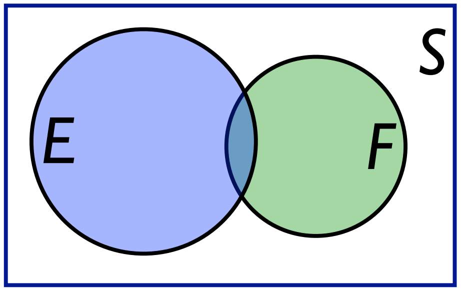
Means "P(E happened, given F observed)"

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:





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

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

Roll one fair die. What is the probability that the outcome is 5 given that it's odd?

$E = \{5\}$ event that roll is 5

$F = \{1, 3, 5\}$ event that roll is odd

Way 1 (from counting):

$P(E \mid F) = |EF| / |F| = |E| / |F| = 1/3$

Way 2 (from probabilities):

$P(E \mid F) = P(EF) / P(F) = P(E) / P(F) = (1/6) / (1/2) = 1/3$

Way 3 (from restricted sample space):

All outcomes are equally likely. Knowing F occurred doesn't distort relative likelihoods of outcomes *within* F, so they remain equally likely. There are only 3 of them, one being E, so

$P(E \mid F) = 1/3$

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 $A = \{HH, HT, TH\}$

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

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

  Let $G = \{TH, HT, TT\}$

  $P(B|G) = P(BG)/P(G) = P(\emptyset)/P(G) = 0/P(G) = 0$

slicing up the spam

24 emails are sent, 6 each to 4 users.
10 of the 24 emails are spam.
All possible outcomes equally likely.

E = user #1 receives 3 spam emails

What is P(E) ?

$$P(E) = \frac{|E|}{|S|} = \frac{\binom{10}{3}\binom{14}{3}\cancel{\binom{18}{6}}\cancel{\binom{12}{6}}\cancel{\binom{6}{6}}}{\binom{24}{6}\cancel{\binom{18}{6}}\cancel{\binom{12}{6}}\cancel{\binom{6}{6}}} \approx 0.3245$$

24 emails are sent, 6 each to 4 users.
10 of the 24 emails are spam.
All possible outcomes equally likely

E = user #1 receives 3 spam emails
F = user #2 receives 6 spam emails

What is P(E|F) ?

$$P(E \mid F) = \frac{|EF|}{|F|} = \frac{\cancel{\binom{10}{6}}\binom{4}{3}\binom{14}{3}\cancel{\binom{12}{6}}\cancel{\binom{6}{6}}}{\cancel{\binom{10}{6}}\binom{18}{6}\cancel{\binom{12}{6}}\cancel{\binom{6}{6}}} \approx 0.0784$$

slicing up the spam

24 emails are sent, 6 each to 4 users.
10 of the 24 emails are spam.
All possible outcomes equally likely

E = user #1 receives 3 spam emails
F = user #2 receives 6 spam emails
G = user #3 receives 5 spam emails

What is P(G|F) ?

$$P(G \mid F) = \frac{|GF|}{|F|} = \frac{\binom{10}{6}\overbrace{\boxed{\binom{4}{5}}}^{=0}\binom{14}{1}\binom{12}{6}\binom{6}{6}}{\binom{10}{6}\binom{18}{6}\binom{12}{6}\binom{6}{6}} = 0$$

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

Holds even when outcomes are *not* equally likely.

Example: S = {# of heads in 2 coin flips} = {0, 1, 2}

NOT equally likely outcomes: P(0)=P(2)=1/4, P(1)=1/2

Q. What is prob of 2 heads (E) given at least 1 head (F)?

A. P(EF)/P(F) = P(E)/P(F) = (1/4)/(1/4+1/2) = 1/3

Same as earlier formulation of this example (of course!)

BT p. 24

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

Holds even when outcomes are *not* equally likely.

What if P(F) = 0?

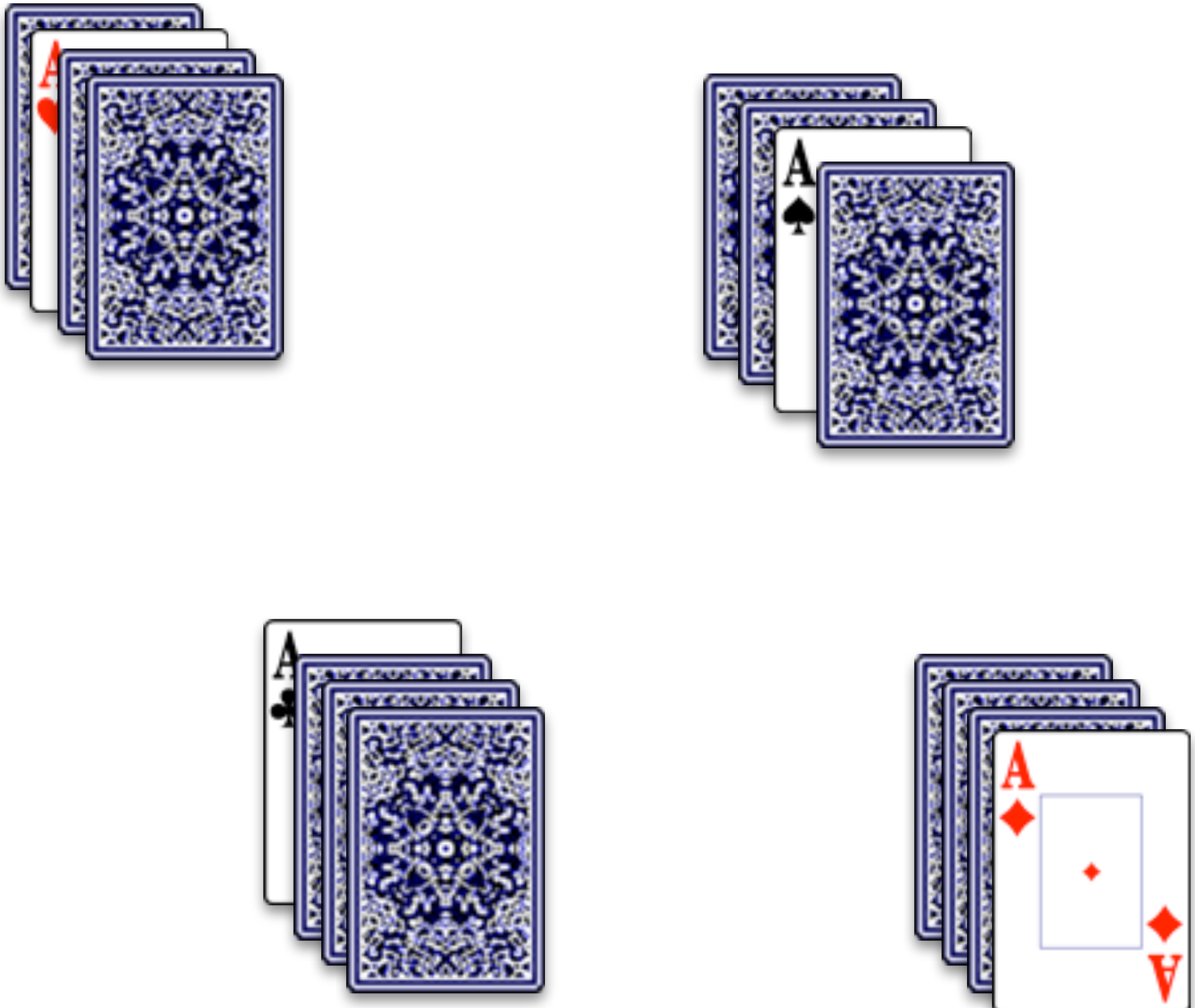
P(E|F) undefined: (you can't observe the impossible)

Implies *(when (PF)>0)*: P(EF) = P(E|F) P(F) (*"the chain rule"*)

General definition of Chain Rule:

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

piling cards

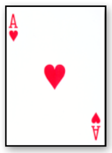


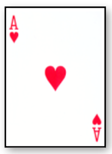
Deck of 52 cards randomly divided into 4 piles

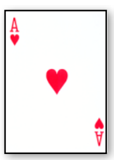
13 cards per pile

Compute P(each pile contains an ace)

Solution:

E_1 = {  in any one pile }

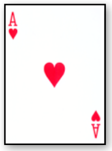
E_2 = {  &  in different piles }

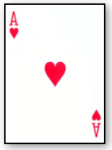
E_3 = {    in different piles }

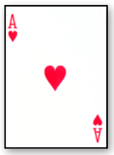
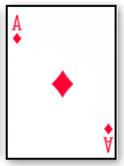
E_4 = { all four aces in different piles }

Compute $P(E_1 E_2 E_3 E_4)$

piling cards

E_1 = {  in any one pile }

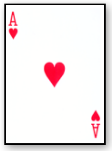
E_2 = {  &  in different piles }

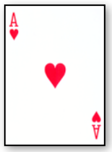
E_3 = {    in different piles }

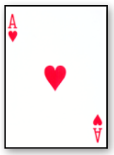
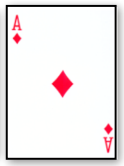
E_4 = { all four aces in different piles }

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

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 }

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

$P(E_1) = 52/52 = 1$ (A♥ can go anywhere)

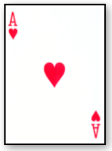
$P(E_2|E_1) = 39/51$ (39 of 51 slots not in A♥ pile)

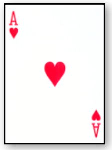
$P(E_3|E_1E_2\,) = 26/50$ (26 not in A♥, A♠ piles)

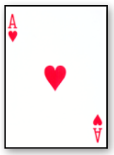
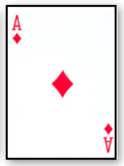
$P(E_4|E_1E_2E_3) = 13/49$ (13 not in A♥, A♠, A♦ piles)

A conceptual trick: what's randomized?

a) *randomize* cards, deal *sequentially* into 4 piles
b) *sort* cards, aces first, deal *randomly* into empty slots among 4 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 }

$$P(E_1E_2E_3E_4)$$

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

$$= (52/52)\cdot(39/51)\cdot(26/50)\cdot(13/49)$$

$$\approx 0.105$$

BT p. 19

"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: $P(A \cup B) \quad \leq P(A) \quad + P(B)$

$\therefore \quad P(A \cup B|F) \leq P(A|F) + P(B|F)$

Ex: $P(A) \quad = 1\text{-}P(A^C)$

$\therefore \quad P(A|F) = 1\text{-}P(A^C|F)$

etc.

sending bit strings

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

All distinct arrangements of bits equally likely

E = first bit received is a 0

F = k of first r bits received are 0's

What's P(E|F)?

Solution 1 ("restricted sample space"):

Observe:

P(E|F) = P(picking one of k 0's out of r bits)

So:

P(E|F) = k/r

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

All distinct arrangements of bits equally likely

E = first bit received is a 0

F = k of first r bits received are 0's

What's P(E|F)?

Solution 2 (counting):

EF = { (n+m)-bit strings | 1st bit = 0 & (k-1)0's in the next (r-1) }

$$|EF| = \binom{r-1}{k-1}\binom{n+m-r}{n-k}$$

$$|F| = \binom{r}{k}\binom{n+m-r}{n-k}$$

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

One of the many binomial identities

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

All distinct arrangements of bits equally likely

E = first bit received is a 0

F = k of first r bits received are 0's

What's P(E|F)?

Solution 3 (more fun with conditioning):

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

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

A generally useful trick: Reversing conditioning

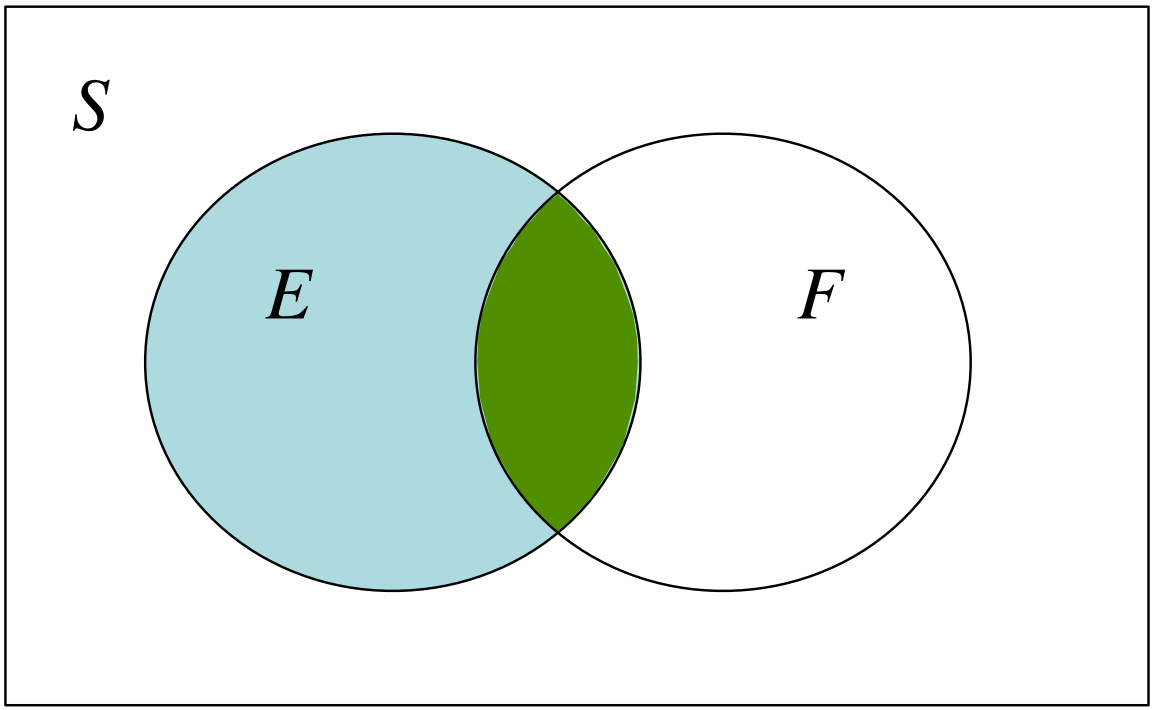
Above eqns, plus the same binomial identity twice.

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

BT p. 28

E and F are events in the sample space S

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



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

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

BT p. 28

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)\, P(F) + P(E|F^c)\, P(F^c) \\ &= P(E|F)\, P(F) + P(E|F^c)\, (1-P(F)) \end{aligned}$$

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

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

weighted average, conditioned on events F_i happening or not.

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

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?

$$
\begin{aligned}
P(A) &= P(A|Phys)P(Phys) + P(A|Chem)P(Chem) \\
&= (3/4)(1/2)+(3/5)(1/2) \\
&= 27/40
\end{aligned}
$$

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.

BT pg. 63

2 Gamblers: Alice & Bob.

A has i dollars; B has (N-i)

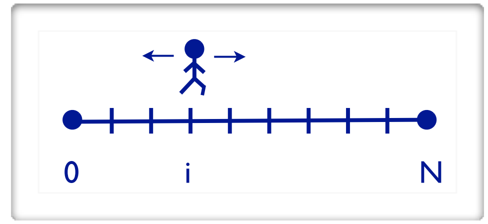
Flip a coin. Heads – A wins \$1; Tails – B wins \$1

Repeat until A or B has all N dollars

What is P(A wins)?

Let E_i = event that A wins starting with \$i

Approach: Condition on 1st flip; H = heads



aka "Drunkard's Walk"

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

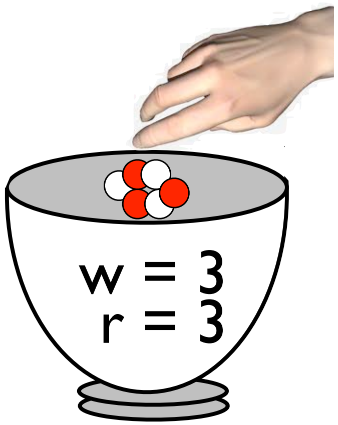
$$\begin{aligned}
p_i &= P(E_i) = P(E_i \mid H)P(H) + P(E_i \mid T)P(T) \\
p_i &= \frac{1}{2}(p_{i+1} + p_{i-1}) \\
2p_i &= p_{i+1} + p_{i-1} \\
p_{i+1} - p_i &= p_i - p_{i-1} \\
p_2 - p_1 &= p_1 - p_0 = p_1, \ \text{since } p_0 = 0
\end{aligned}$$

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

Bayes Theorem

BT p. 1.4

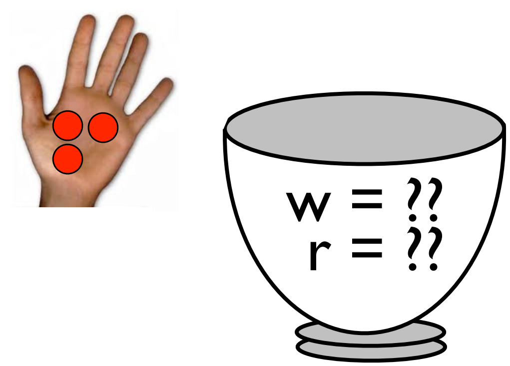
6 balls in an urn,
some red, some white

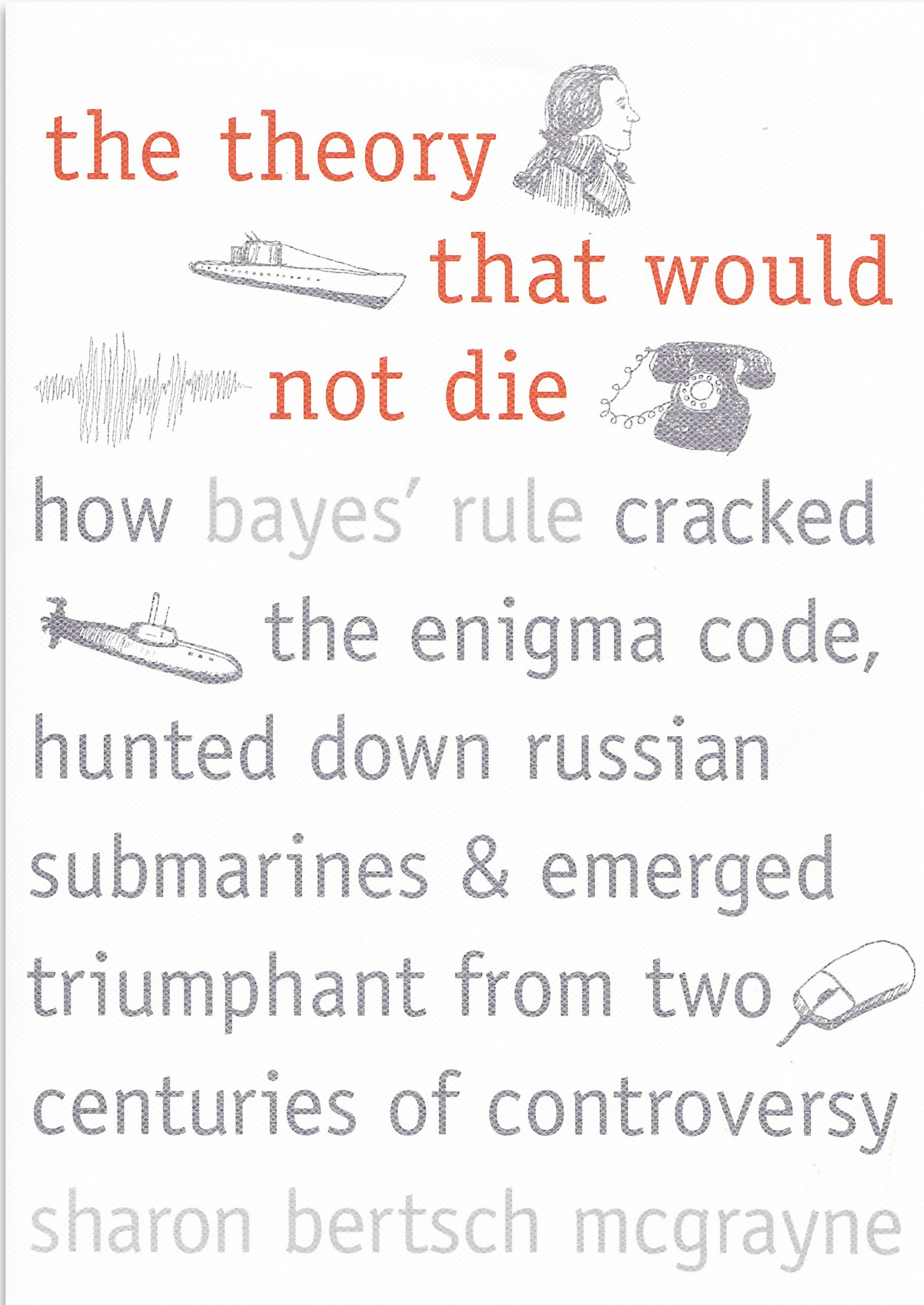


Probability of drawing 3 red balls, given 3 in urn ?

Rev. Thomas Bayes c. 1701-1761

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





Yale University Press, 2011

ISBN-13: 978-0300188226

http://www.amazon.com/Theory-That-Would-Not-Die/dp/0300188226/

"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:

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

Most common form:

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

Expanded form (using law of total probability):

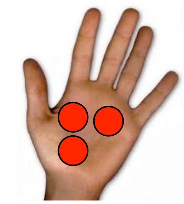
$$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E \mid F)P(F) + P(E \mid F^c)P(F^c)}$$

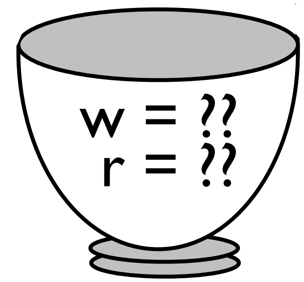
Why it's important:

- 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

$$P(D \mid M) = \binom{3}{3} \Big/ \binom{6}{3} = \frac{1}{20}$$

$$P(M \mid D) = \frac{P(D \mid M)P(M)}{P(D \mid M)P(M) + P(D \mid M^c)P(M^c)}$$

$$= \frac{(\frac{1}{20})(\frac{3}{4})}{(\frac{1}{20})(\frac{3}{4}) + (1)(1 - \frac{3}{4})} = \frac{3}{23}$$

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:

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

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

Exercises:

What if M_2 were (A,A)?

What if child were (A,A)?

Solution:

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

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

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

E.g., $1/2 \rightarrow 2/3$

I.e., the given 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:

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

P(E) ≈ 1.5%

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

	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 *probabiliy* of event E is P(E).

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*; this is close, but not equal to, P(F)=1/200)

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: $P(F|E)/P(F^c|E)$?

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

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

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

$$\begin{pmatrix}\text{posterior}\\ \text{odds}\end{pmatrix} = \begin{pmatrix}\text{“Bayes}\\ \text{factor”}\end{pmatrix} \cdot \begin{pmatrix}\text{prior}\\ \text{odds}\end{pmatrix}$$

In a way, nothing new here, versus prior results, but the simple form, and the simple interpretation are convenient.

	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 = 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 in this case, to 2.03:1 against (vs prior of 199:1 against)

Let E = you test *negative* for HIV

Let F = you actually *have* HIV

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

What is the *ratio* between $P(F|E)$ and $P(F^c|E)$?

$$\begin{aligned}
\frac{P(F \mid E)}{P(F^c \mid E)} &= \frac{P(E \mid F)}{P(E \mid F^c)} \frac{P(F)}{P(F^c)} \\
(\text{posterior odds} &= \text{“Bayes factor”} \cdot \text{prior odds}) \\
&= \frac{0.02}{0.99} \cdot \frac{0.005}{0.995}
\end{aligned}$$

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:

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

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

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

prior, posterior, odds, prior odds, posterior odds, Bayes factor