Conditional Probability

P( | )

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:

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

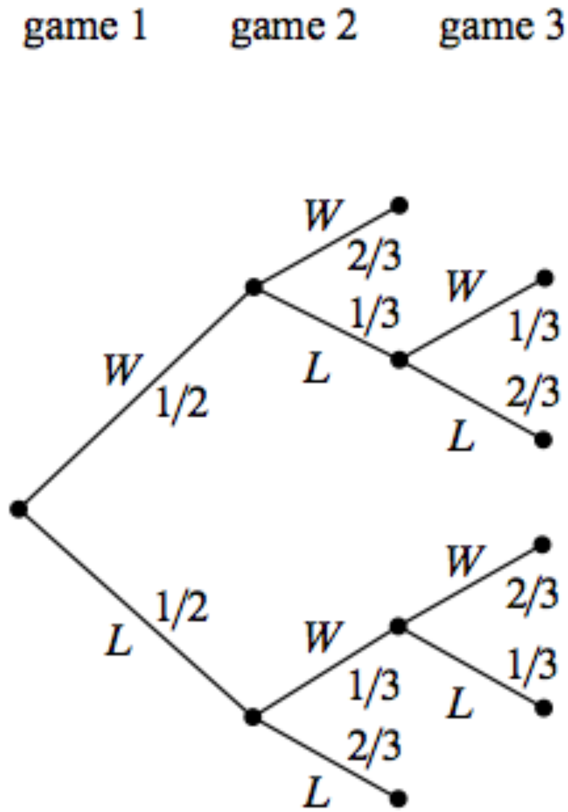
Best of 3 tournament between local team and other team

First game: local team wins with probability ½

Given that they win current game, win the next game with probability 2/3

Given that they lose current game, win next game with probability 1/3

What is the probability that they win a best of 3 tournament given that they win the first game.



outcome	event A: win the series	event B: win game 1	outcome probability
WW	✓	✓	1/3
WLW	✓	✓	1/18
WLL		✓	1/9
LWW	✓		1/9
LWL			1/18
LL			1/3

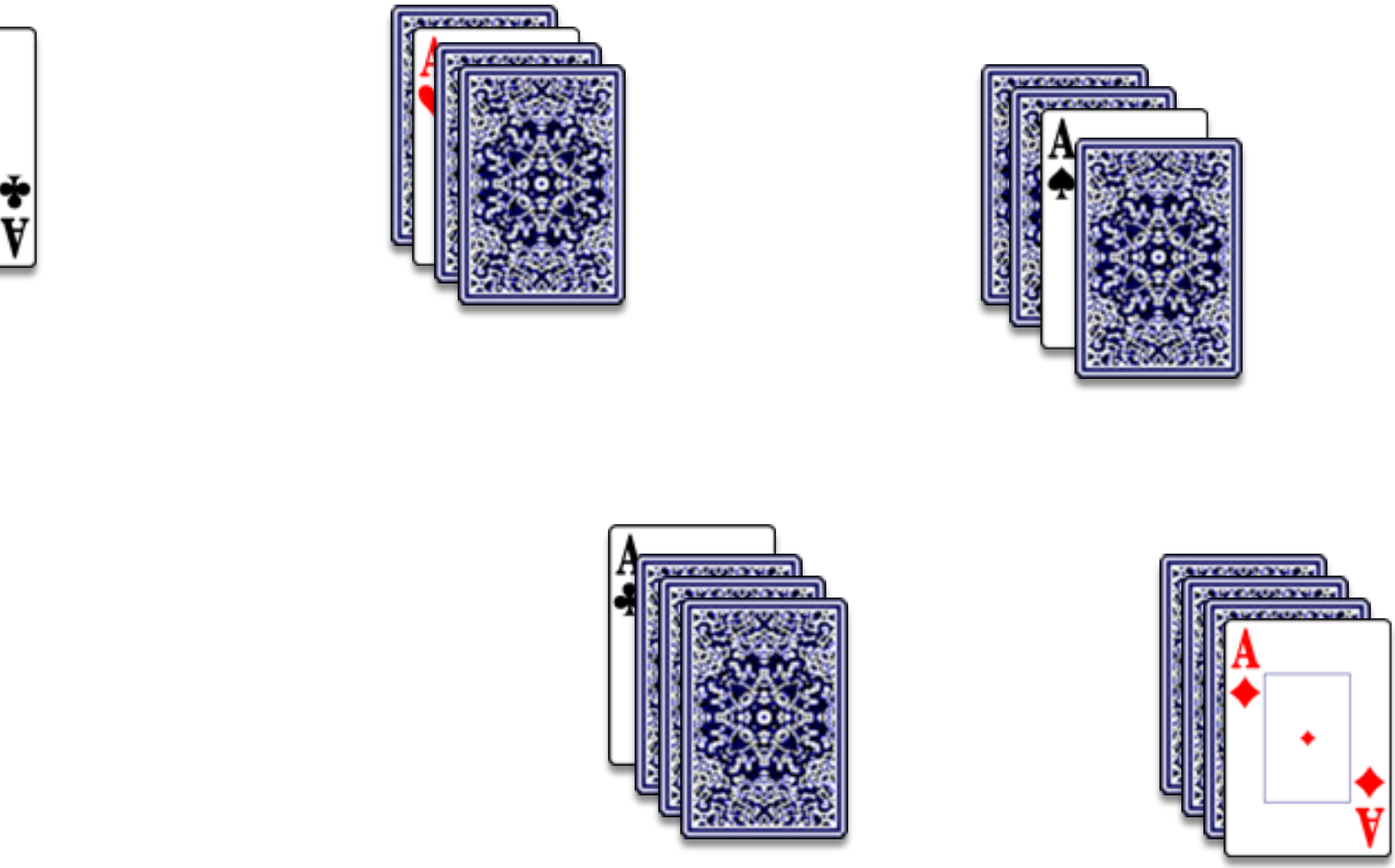
Tree diagrams are useful tools for reasoning about probabilities.

1. Find sample space
2. Define events of interest
3. Determine outcome probabilities, by labeling edges
4. Compute event probabilities

Mathematical justification:

- Edge probabilities are conditional probabilities
- Chain rule

piling cards



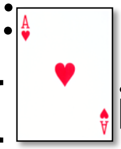
piling cards

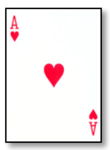
Deck of 52 cards randomly divided into 4 piles

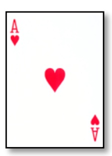
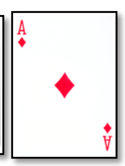
13 cards per pile

Compute P(each pile contains an ace)

Solution:

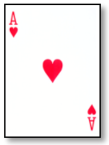
E_1 = {  in any one pile }

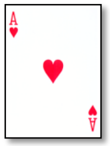
E_2 = {  &  in different piles }

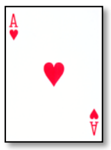
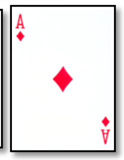
E_3 = {    in different piles }

E_4 = { all four aces in different piles }

Compute $P(E_1 E_2 E_3 E_4)$

E_1 = {  in any one pile }

E_2 = {  &  in different piles }

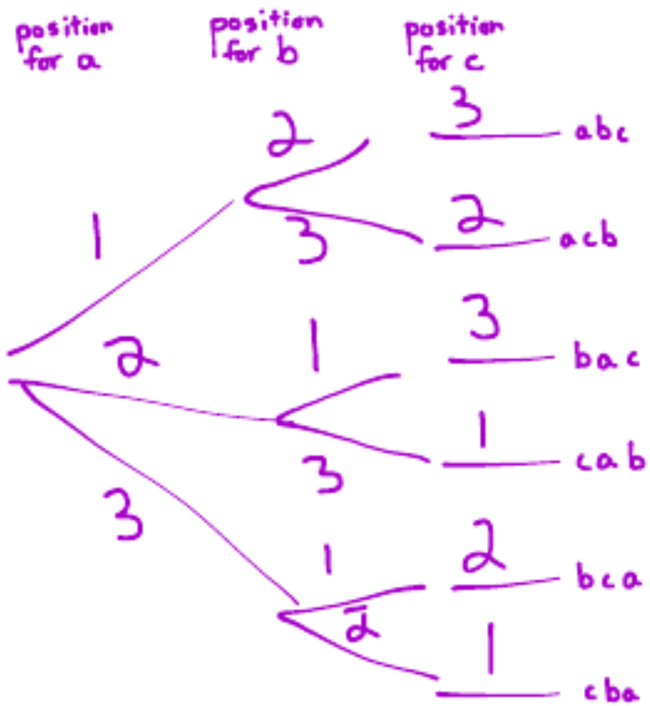
E_3 = {    in different piles }

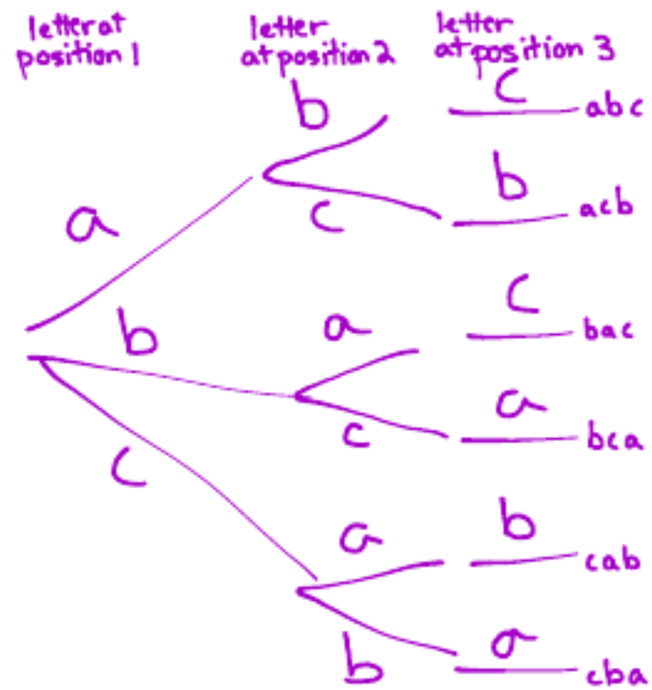
E_4 = { all four aces in different piles }

$P(E_1E_2E_3E_4)$

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

Permutations of {a,b,c}

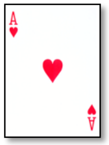


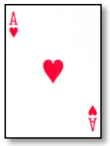


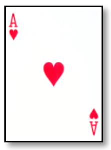
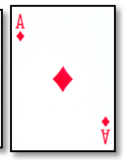
A conceptual trick: what's randomized?
a) *randomize* cards, deal *sequentially*
b) *sort* cards, deal *randomly* into empty slots

positions: 1 2 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)

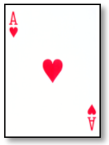
$P(E_4|E_1E_2E_3) = 13/49$ (13 not in A♥, A♠, A♦ piles)

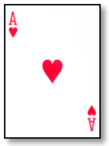
A conceptual trick: what's randomized?

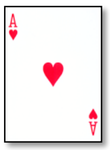
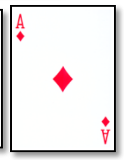
a) *randomize* cards, deal *sequentially* into 4 piles

b) *sort* cards, aces first, deal *randomly* into empty slots among 4 piles.

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 }

$P(E_1E_2E_3E_4)$

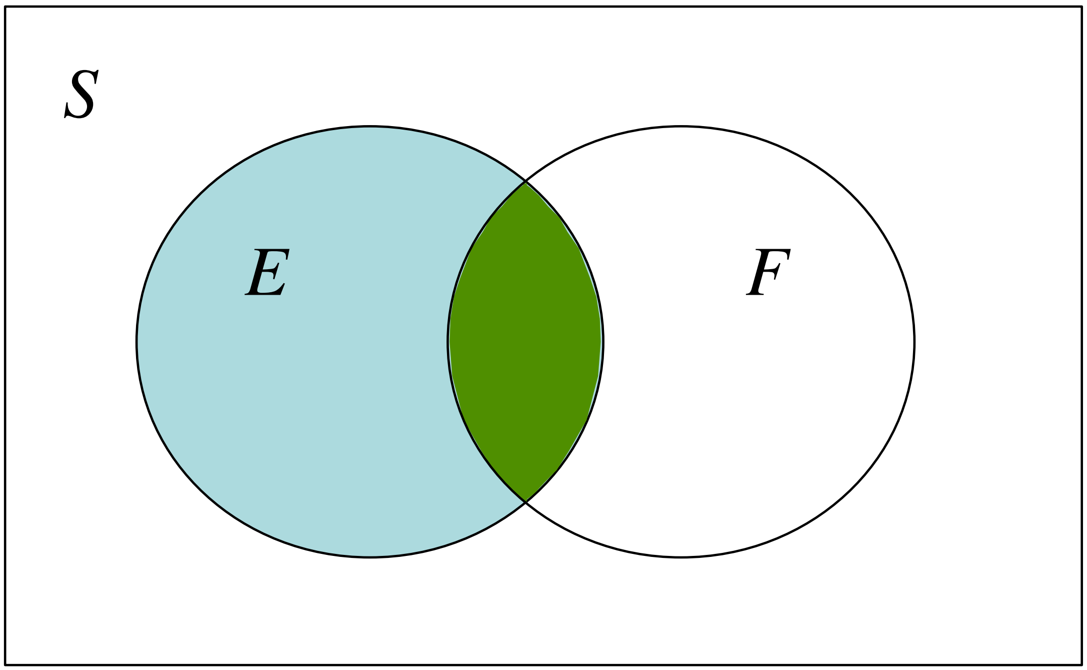
$= P(E_1)\ P(E_2|E_1)\ P(E_3|E_1E_2)\ P(E_4|E_1E_2E_3)$

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

≈ 0.105

E and F are events in the sample space S

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



$$EF \cap EF^c = \varnothing$$

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

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

weighted average, conditioned on event F happening or not.

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

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

weighted average, conditioned on events F_i happening or not.

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

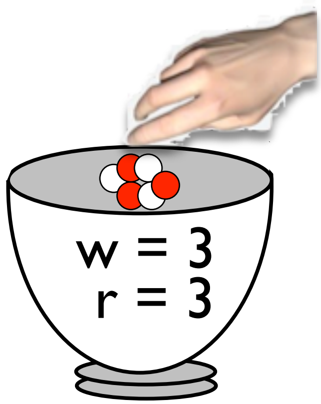
Sally has 1 elective left to take: either Phys or Chem. She will get A with probability 3/4 in Phys, with prob 3/5 in Chem. She flips a coin to decide which to take.

What is the probability that she gets an A?

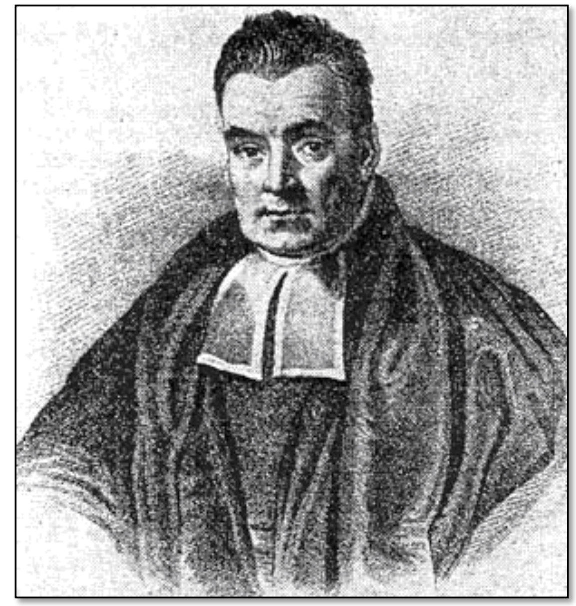
$$P(A) = P(A|Phys)P(Phys) + P(A|Chem)P(Chem)$$
$$= (3/4)(1/2)+(3/5)(1/2)$$
$$= 27/40$$

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.

6 red or white balls in an urn

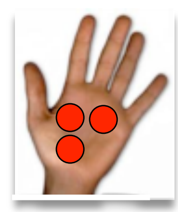


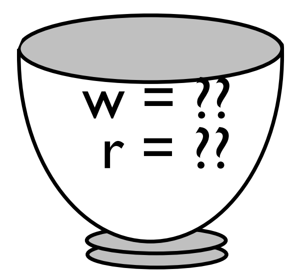
Probability of drawing 3 red balls, given 3 in urn ?



Rev. Thomas Bayes c. 1701-1761

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

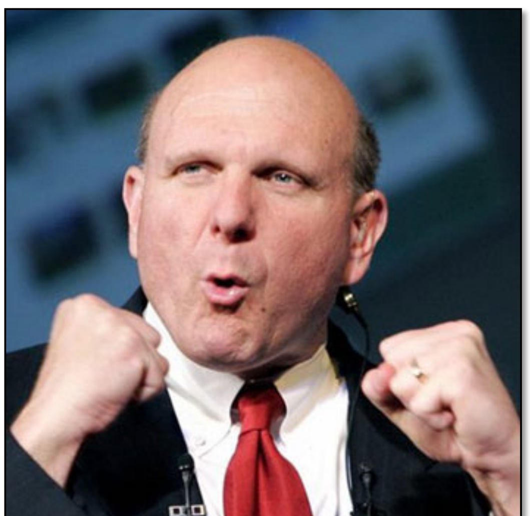




Improbable Inspiration: The future of software may lie in the obscure theories of an 18^{th} century cleric named Thomas Bayes

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

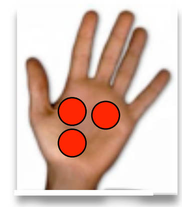
P(model | data) ~ P(data | model)

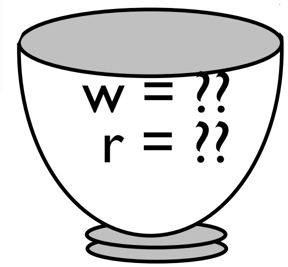
Combine new evidence (E) with prior belief (P(F))

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

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

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

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

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:

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

$$= \frac{(0.9)(0.6)}{(0.9)(0.6) + (0.2)(0.4)}$$

$$\approx 0.871$$