Conditional Probability

P( | )

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

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)"

$EF \triangleq E \cap F$

$S \triangleq \Omega$

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,

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

2 random cards are selected from a deck of cards:

$|\Omega| = \binom{52}{2}$

- 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?

$\Pr(\text{both Aces} \mid \text{one of cards } A\spadesuit) = \dfrac{\Pr(\text{both Aces \& one is } A\spadesuit)}{\Pr(\text{one is } A\spadesuit)}$

$= \dfrac{|\text{both Aces \& one A spades}|}{|\text{one is } A\spadesuit|} = \dfrac{3}{51}$

$\Pr(\text{both Aces} \mid \text{at least one is } A) = \dfrac{|\overbrace{\text{both Aces \& at least one is Ace}}^{\text{both Aces}}|}{|\text{at least one is } A|}$

$= \dfrac{\binom{4}{2}}{\#\text{2 card hands} - \#\text{2 card hands with no Aces}} = \dfrac{\binom{4}{2}}{\binom{52}{2} - \binom{48}{2}}$

Gambler's fallacy

Flip a fair coin 51 times

$|\Omega| = 2^{51}$

A = "first 50 flips are heads"

B = "the 51st flip is heads"

Pr (B | A) = ?

$$= \frac{|B \cap A|}{|A|} = \frac{|\{\overbrace{HH\ldots H}^{51}\}|}{|\{\underbrace{HH\ldots H}_{50}H, \underbrace{HH\ldots H}_{50}T\}|} = \frac{1}{2}$$

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

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

General definition of Chain Rule:

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

$$\Pr(\underbrace{(E_1 \cap E_2)}_{F} \cap \underbrace{E_3}_{E}) = \Pr(E_3 \mid E_1 \cap E_2)\, \Pr(E_1 \cap E_2)$$

$$= \Pr(E_3 \mid E_1 \cap E_2)\, \Pr(E_2 \mid E_1)\, \Pr(E_1)$$

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.

nobody wins 3, 4

Otherwise, they play a second round and so on.

What is P($\overbrace{\text{Alice wins on n}^{\text{th}}\text{ round}}^{A_n}$)?

A_i: Alice wins on i^{th} round

N_i: nobody wins on i^{th} round

$Pr(A_n) = Pr(N_1 \cap N_2 \cap \dots \cap N_{n-1} \cap A_n)$

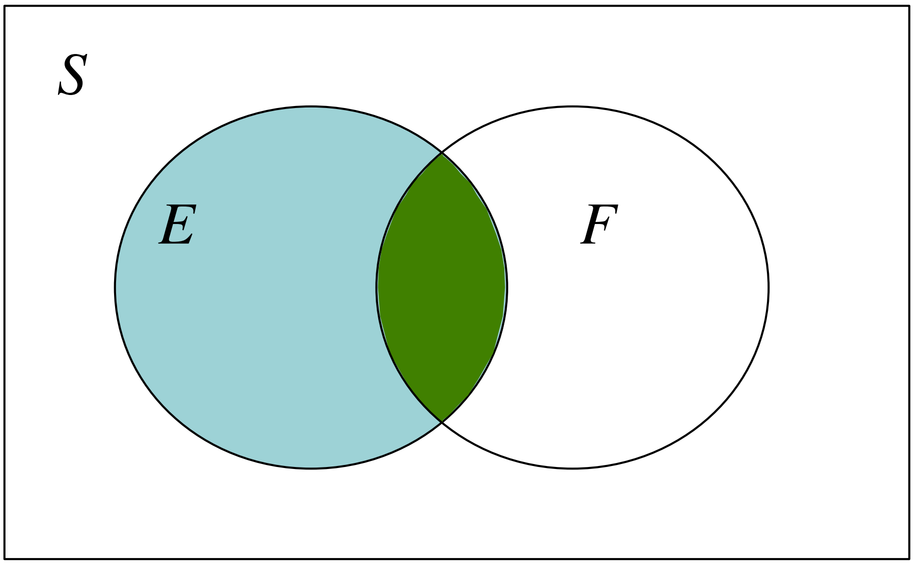
$= Pr(N_1)\, Pr(N_2 | N_1)\, Pr(N_3 | N_1 \cap N_2) \cdots Pr(N_{n-1} | N_1 \cap \dots \cap N_{n-2})$

$\cdot\, Pr(A_n | N_1 \dots N_{n-1})$

$\frac{1}{3} \quad \frac{1}{3} \cdots \frac{1}{3}\; \frac{1}{6} = \left(\frac{1}{3}\right)^{n-1} \frac{1}{6}$

E and F are events in the sample space S

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



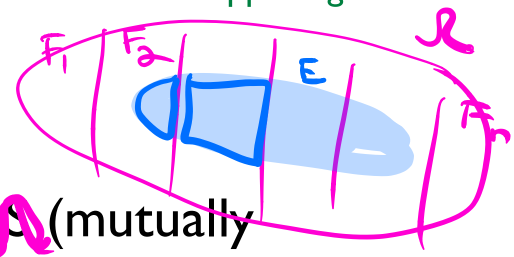
$$EF \cap EF^c = \emptyset$$

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

law of total probability

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

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

$$P(E) = \sum_i P(E \text{ and } F_i) = \sum_i P(E|F_i)\ P(F_i)$$

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.

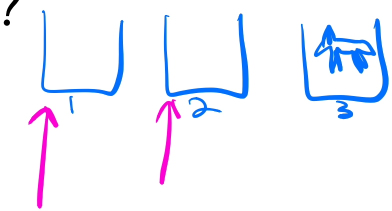
What is the probability that she gets an A?

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

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.

The Monty Hall Problem

Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the other, goats. You pick a door, say number 1, and the host, who knows what's behind the doors, opens another door, say number 3, which has a goat. He says to you, "Do you want to switch to door number 2?" Is it to your advantage to switch your choice of doors?



Assumptions:

The car is equally likely to be behind each of the doors.

The player is equally likely to pick each of the three doors, regardless of the car's location

After the player picks a door, the host ***must*** open a different door with a goat behind it and offer the player the choice of staying with the original door or switching

If the host has a choice of which door to open, then he is equally likely to select each of them.

Switch

$$\Pr(\text{win}) = \overbrace{\Pr(\text{win} \mid \$ \text{ behind } \underbrace{\text{door \#1}}_{\substack{= \text{ one} \\ \text{you picked}}})}^{0} \boxed{\overbrace{\Pr(\$ \text{ behind door } 1)}^{1/3}}$$

$$+ \overbrace{\Pr(\text{win} \mid \$ \to 2)}^{1} \boxed{\overbrace{\Pr(\$ \to 2)}^{1/3}}$$

$$+ \overbrace{\Pr(\text{win} \mid \$ \to 3)}^{1} \boxed{\Pr(\$ \to 3)}\ \tfrac{1}{3} = \frac{2}{3}$$

Independence

Intuition: E is independent of F if the chance of E occurring is not affected by whether F occurs.

Formally:

$$\underbrace{Pr(E|F)}_{= Pr(E\cap F)/Pr(F)} = Pr(E) \qquad \text{or} \qquad Pr(E \cap F) = Pr(E)Pr(F)$$

These two definitions are equivalent.

Draw a card from a shuffled deck of 52 cards. $|\Omega| = 52$

E: card is a spade $\Pr(E) = \frac{13}{52} = \frac{1}{4}$

F: card is an Ace $\Pr(F) = \frac{4}{52} = \frac{1}{13}$

Are E and F independent?

$\Pr(E \cap F) = \Pr(A\spadesuit) = \frac{1}{52} = \Pr(E)\Pr(F)$

Toss a coin 3 times. Each of 8 outcomes equally likely.
Define

A = {at most one T} = {HHH, HHT, HTH, THH} $\Pr(A) = \frac{1}{2}$

B = {at most 2 Heads}= $\{HHH\}^c$ $\Pr(B) = \frac{7}{8}$

Are A and B independent?

$\Pr(A \cap B) = \Pr(\{HHT, HTH, THH\}) = \frac{3}{8} \neq \Pr(A)\Pr(B)$

It is often convenient to **assume** independence.

People often assume it without noticing.

Example: A sky diver has two chutes. Let

E = {main chute doesn't open} Pr (E) = 0.02

F = {backup doesn't open} Pr (F) = 0.1

What is the chance that at least one opens assuming independence?

$$\Pr(\text{at least one opens}) = 1 - \Pr(\underbrace{\text{none open}}_{E \cap F})$$

$$= 1 - 0.02 \cdot 0.1$$

$$= .998$$

Independence as an assumption

It is often convenient to **assume** independence.

People often assume it without noticing.

Example: A sky diver has two chutes. Let

E = {main chute doesn't open} Pr (E) = 0.02

F = {backup doesn't open} Pr (F) = 0.1

What is the chance that at least one opens assuming independence?

Note: Assuming independence doesn't justify the assumption! Both chutes could fail because of the same rare event, e.g. freezing rain.

Using independence to define a probabilistic model

We can **define** our probability model via independence.

Example: suppose a biased coin comes up heads with probability 2/3, independent of other flips.

Sample space: sequences of 3 coin tosses.

$\Omega = \{H, T\}^3$

Pr (3 heads)=? $Pr(H_1 H_2 H_3) = Pr(H_1)Pr(H_2)Pr(H_3) = \left(\frac{2}{3}\right)^3$

Pr (3 tails) = ? $\left(\frac{1}{3}\right)^3$

Pr (2 heads) = ?

$Pr(\{HHT, HTH, THH\}) = 3\left(\frac{2}{3}\right)^2 \cdot \frac{1}{3}$

$\left(\frac{2}{3}\right)^{2}\cdot\frac{1}{3}$

Bayes Theorem

P( | )

Is your coin loaded?

Your coin is fair ($Pr(H) = 0.5$) with probability ½ or "unfair" ($Pr(H) = 0.6$) otherwise.

You flip the coin and it comes up Heads.

What is the probability that it is fair given that it came up Heads?

$Pr(\text{fair}) = \frac{1}{2}$

$Pr(H|\text{fair}) = \frac{1}{2}$

$Pr(\text{unfair}) = \frac{1}{2}$

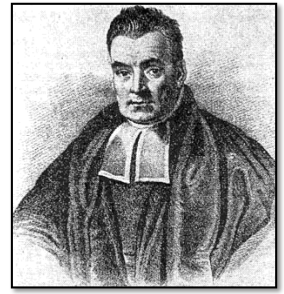
$Pr(H|\text{unfair}) = 0.6$

$$Pr(\text{fair}|H) = \frac{Pr(\text{fair} \cap H)}{Pr(H)} = \frac{Pr(H|\text{fair})Pr(\text{fair})}{Pr(H)}$$

$$Pr(H|\text{fair})Pr(\text{fair}) + Pr(H|\text{unfair})Pr(\text{unfair})$$

Most common form:

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



Rev. Thomas Bayes c. 1701-1761

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

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 $P(F|E)$?

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

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

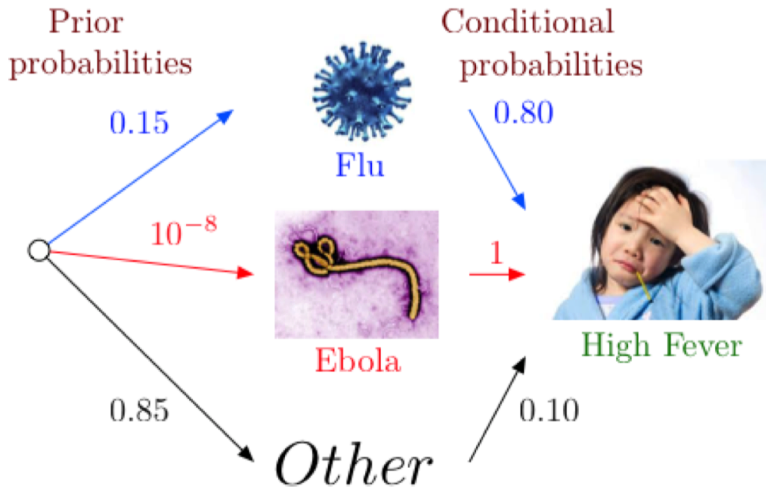
What is P(F|E) ?

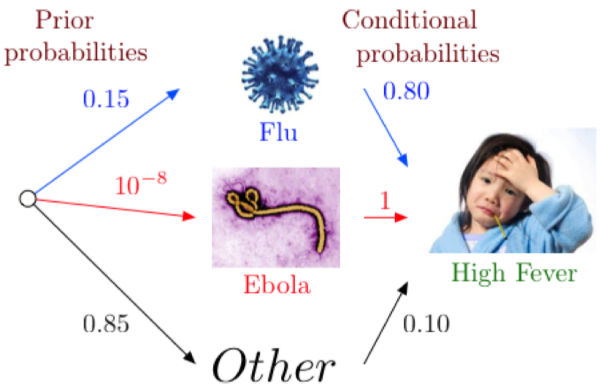
Solution:

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

Why do you have a fever?

Pr (flu | high fever)?





$$Pr[\text{Flu}|\text{High Fever}] = \frac{0.15 \times 0.80}{0.15 \times 0.80 + 10^{-8} \times 1 + 0.85 \times 0.1} \approx 0.58$$

$$Pr[\text{Ebola}|\text{High Fever}] = \frac{10^{-8} \times 1}{0.15 \times 0.80 + 10^{-8} \times 1 + 0.85 \times 0.1} \approx 5 \times 10^{-8}$$

$$Pr[\text{Other}|\text{High Fever}] = \frac{0.85 \times 0.1}{0.15 \times 0.80 + 10^{-8} \times 1 + 0.85 \times 0.1} \approx 0.42$$

These are the posterior probabilities. One says that ‘Flu’ is the Most Likely a Posteriori (MAP) cause of the high fever.

A lie detector is known to be 80% reliable when the person is guilty and 95% reliable when the person is innocent.

If a suspect is chosen from a group of suspects of which only 2% have ever committed a crime, and the test indicates that the person is guilty, what is the probability that he is innocent?

A lie detector is known to be 80% reliable when the person is guilty and 95% reliable when the person is innocent.

If a suspect is chosen from a group of suspects of which only 2% have ever committed a crime, and the test indicates that the person is guilty, what is the probability that he is innocent?

$$I: \text{ event that he is innocent}$$

$$G: \text{ event that the test indicates guilty}$$

$$Pr(I|G) = \frac{Pr(G|I)Pr(I)}{Pr(G)} = \frac{Pr(G|I)Pr(I)}{Pr(G|I)Pr(I) + Pr(G|I^c)Pr(I^c)}$$

$$= \frac{0.05 \cdot 0.98}{0.05 \cdot 0.98 + 0.8 \cdot 0.02}$$

Problem

There is a population of N people. The number of good guys among these people is i with probability p_i

Take a sample of n people from the population. Each subset of n is equally likely. What is the probability that there are j good guys in the population conditioned on the fact that there are k good guys in the sample.

Problem

There is a population of N people. The number of good guys among these people is i with probability p_i

Take a sample of n people from the population. What is the probability that there are j good guys in the population conditioned on the fact that there are k good guys in the sample.

$$E_i = \text{ event that there are } i \text{ good guys among } N$$

$$S_i = \text{ event that there are } i \text{ good guys in sample}$$

$$Pr(E_j|S_k) = \frac{Pr(S_k|E_j)Pr(E_j)}{Pr(S_k)} \qquad Pr(S_k|E_j) = \frac{\binom{j}{k}\binom{N-j}{n-k}}{\binom{N}{k}}$$

$$Pr(S_k) = \sum_j Pr(S_k|E_j)Pr(E_j) = \sum_j \frac{\binom{j}{k}\binom{N-j}{n-k}}{\binom{N}{k}} p_j$$

Conditional Probability

Satisfies usual axioms of probability

Example:

$\Pr(E \mid F) = 1 - \Pr (E^c \mid F)$

Conditional Probabilities yield a probability space

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

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

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

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

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

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

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.

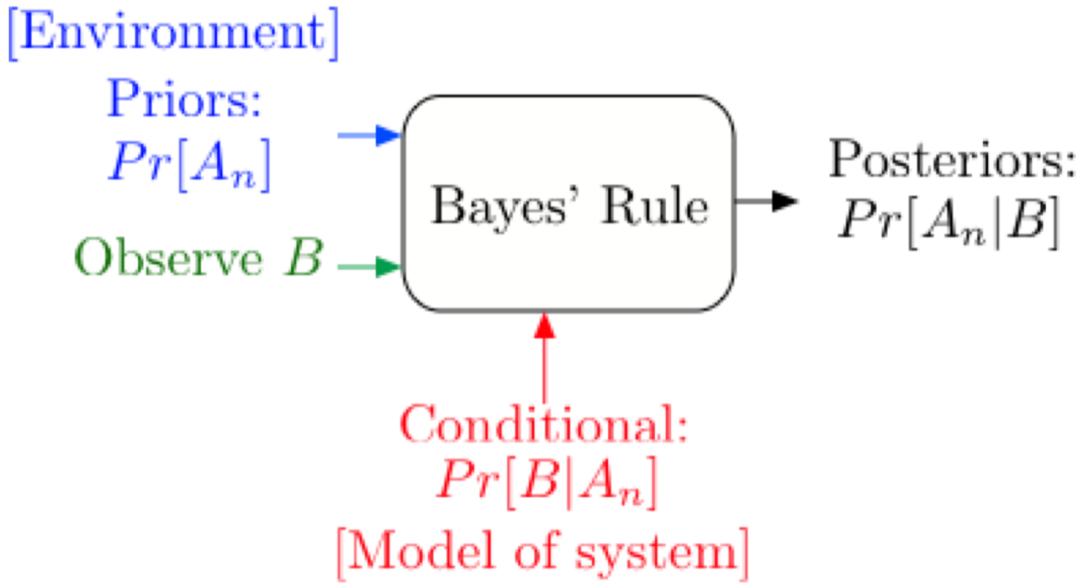
P(EF) = P(E|F) P(F)   ("the chain rule")

"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

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



Bayes' Rule is the canonical example of how information changes our opinions.

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

Let E^c = you test **negative** for HIV

Let F = you actually have HIV

What is $P(F|E^c)$?

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