Lecture 15: More Probability.

Events, Conditional Probability, Independence, Bayes' Rule

Summary.

Modeling Uncertainty: Probability Space

1. Random Experiment
2. Probability Space: $\Omega; Pr[\omega] \in [0,1]; \sum_{\omega} Pr[\omega] = 1$.
3. Uniform Probability Space: $Pr[\omega] = 1/|\Omega|$ for all $\omega \in \Omega$.
4. Event: "subset of outcomes." $A \subseteq \Omega$. $Pr[A] = \sum_{w \in A} Pr[\omega]$
5. Some calculations.

CS70: Onwards.

Events, Conditional Probability, Independence, Bayes' Rule

1. Probability Basics Review
2. Events
3. Conditional Probability
4. Independence of Events
5. Bayes' Rule

Probability Basics Review

Setup:

- Random Experiment.
  Flip a fair coin twice.
- Probability Space.
  - **Sample Space:** Set of outcomes, Ω.
    $\Omega = \{HH, HT, TH, TT\}$
    (Note: Not $\Omega = \{H, T\}$ with two picks!)
  - **Probability:** $Pr[\omega]$ for all $\omega \in \Omega$.
    $Pr[HH] = \cdots = Pr[TT] = 1/4$
    1. $0 \le Pr[\omega] \le 1$.
    2. $\sum_{\omega \in \Omega} Pr[\omega] = 1$.

Probability: Events.

An *event* A in a probability space, $\Omega, Pr[\cdot]$, is $A \subseteq \Omega$.

The probability of an event A is $Pr[A] = \sum_{\omega in \Omega} Pr[\omega]$.

Don't sweat $Pr[A]$ or $Pr(A)$. Same deal.

Examples:

Flip two coins: Event A - exactly one heads.
$\Omega = \{HH, HT, TH, TT\}$.
$A = \{HT, TH\}$.

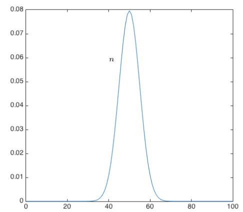
Deal a poker head: Event four aces.
$\Omega =$ all five card poker hands. $|\Omega| = \binom{52}{5}$
$A =$ the poker hands with four aces. $|A| = 48$.

Flip $2n$ coins: Event A - exactly n heads.
$\Omega = \{H, T\}^{2n}$. $|\Omega| = 2^{2n}$
A is set of outcomes with n heads. $|A| = \binom{2n}{n}$.
Approximation: roughly $1/\sqrt{\pi n}$.

$\implies$ not surprising to have something like $n + \sqrt{\pi n}/2$ heads

Probability of n heads in 100 coin tosses.

$\Omega = \{H, T\}^{100}$; $|\Omega| = 2^{100}$.



Event E_n = 'n heads'; $|E_n| = \binom{100}{n}$

$p_n := Pr[E_n] = \frac{|E_n|}{|\Omega|} = \frac{\binom{100}{n}}{2^{100}}$

Observe:

- Concentration around mean: Law of Large Numbers;
- Bell-shape: Central Limit Theorem.

Probability is Additive

Theorem

(a) If events A and B are disjoint, i.e., $A \cap B = \emptyset$, then

$$Pr[A \cup B] = Pr[A] + Pr[B].$$

(b) If events $A_1, \ldots, A_n$ are pairwise disjoint,
i.e., $A_k \cap A_m = \emptyset, \forall k \neq m$, then

$$Pr[A_1 \cup \cdots \cup A_n] = Pr[A_1] + \cdots + Pr[A_n].$$

Proof:

Obvious. Straightforward. Use definition of probability of events.

Consequences of Additivity

Theorem

(a) $Pr[A \cup B] = Pr[A] + Pr[B] - Pr[A \cap B]$;

(inclusion-exclusion property)

(b) $Pr[A_1 \cup \cdots \cup A_n] \leq Pr[A_1] + \cdots + Pr[A_n]$;

(union bound)

(c) If $A_1, \ldots A_N$ are a partition of Ω, i.e.,

pairwise disjoint and $\cup_{m=1}^{N} A_m = \Omega$, then

$$Pr[B] = Pr[B \cap A_1] + \cdots + Pr[B \cap A_N].$$

(law of total probability)

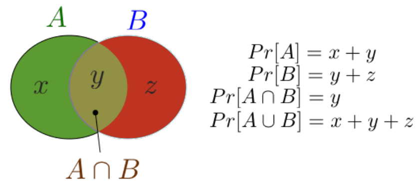
Proof:

(b) is obvious. Doh!
Add probabilities of outcomes once on LHS and at least once on RHS.

Proofs for (a) and (c)? Next...

Inclusion/Exclusion

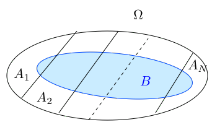
$$Pr[A \cup B] = Pr[A] + Pr[B] - Pr[A \cap B]$$



Another view. Any $\omega \in A \cup B$ is in $A \cap \overline{B}$, $A \cup B$, or $\overline{A} \cap B$. So, add it up.

Total probability

Assume that Ω is the union of the disjoint sets $A_1, \ldots, A_N$.



Then,

$$Pr[B] = Pr[A_1 \cap B] + \cdots + Pr[A_N \cap B].$$

Indeed, B is the union of the disjoint sets $A_n \cap B$ for $n = 1, \ldots, N$.

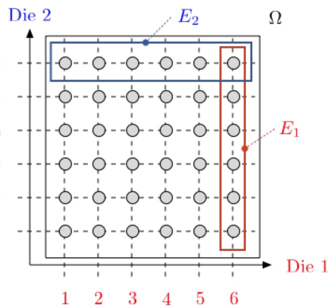
In "math": $\omega \in B$ is in exactly one of $A_i \cap B$.

Adding up probability of them, get $Pr[\omega]$ in sum.

..Did I say...

Add it up.

Roll a Red and a Blue Die.



$$|E_1 \cup E_2| = |E_1| + |E_2| - |E_1 \cap E_2|$$

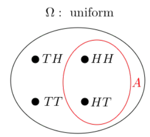
E_1 = 'Red die shows 6'; E_2 = 'Blue die shows 6'

$E_1 \cup E_2$ = 'At least one die shows 6'

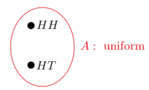
$$Pr[E_1] = \frac{6}{36}, Pr[E_2] = \frac{6}{36}, Pr[E_1 \cup E_2] = \frac{11}{36}.$$

Conditional probability: example.

Two coin flips. First flip is heads. Probability of two heads?
$\Omega = \{HH, HT, TH, TT\}$; Uniform probability space.
Event A = first flip is heads: $A = \{HH, HT\}$.



New sample space: A; uniform still.



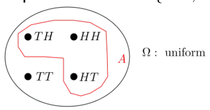
Event B = two heads.

The probability of two heads if the first flip is heads.
The probability of B given A is $1/2$.

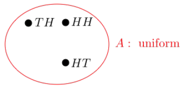
A similar example.

Two coin flips. At least one of the flips is heads.
→ Probability of two heads?

$\Omega = \{HH, HT, TH, TT\}$; uniform.
Event A = at least one flip is heads. $A = \{HH, HT, TH\}$.



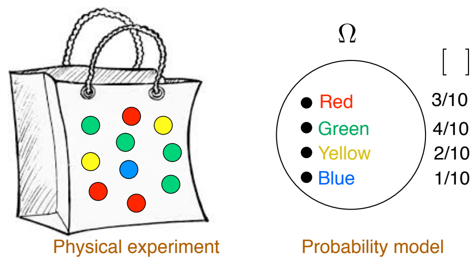
New sample space: A; uniform still.



Event B = two heads.

The probability of two heads if at least one flip is heads.
The probability of B given A is $1/3$.

Conditional Probability: A non-uniform example

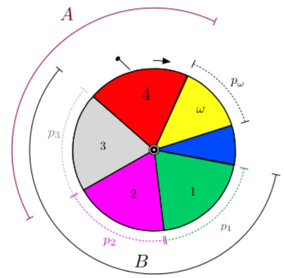


$\Omega = \{\text{Red, Green, Yellow, Blue}\}$

$$Pr[\text{Red}|\text{Red or Green}] = \frac{3}{7} = \frac{Pr[\text{Red} \cap (\text{Red or Green})]}{Pr[\text{Red or Green}]}$$

Another non-uniform example

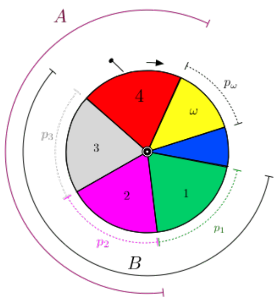
Consider $\Omega = \{1, 2, \ldots, N\}$ with $Pr[n] = p_n$.
Let $A = \{3, 4\}, B = \{1, 2, 3\}$.



$$Pr[A|B] = \frac{p_3}{p_1 + p_2 + p_3} = \frac{Pr[A \cap B]}{Pr[B]}.$$

Yet another non-uniform example

Consider $\Omega = \{1, 2, \ldots, N\}$ with $Pr[n] = p_n$.
Let $A = \{2, 3, 4\}, B = \{1, 2, 3\}$.

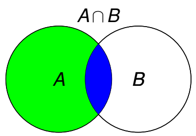


$$Pr[A|B] = \frac{p_2 + p_3}{p_1 + p_2 + p_3} = \frac{Pr[A \cap B]}{Pr[B]}.$$

Conditional Probability.

Definition: The **conditional probability** of B given A is

$$Pr[B|A] = \frac{Pr[A \cap B]}{Pr[A]}$$

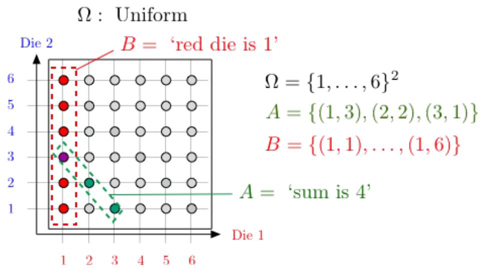


In A!
In B?
Must be in $A \cap B$.

$Pr[B|A] = \frac{Pr[A \cap B]}{Pr[A]}$.

More fun with conditional probability.

Toss a red and a blue die, sum is 4,
What is probability that red is 1?

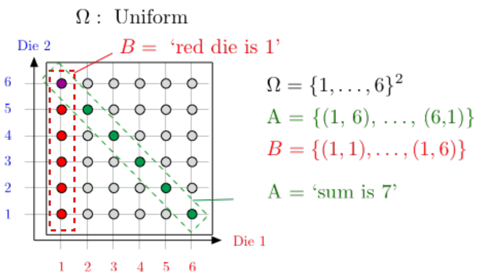


$Pr[B|A] = \frac{|B \cap A|}{|A|} = \frac{1}{3}$; versus $Pr[B] = 1/6$.

B is more likely given A.

Yet more fun with conditional probability.

Toss a red and a blue die, sum is 7,
what is probability that red is 1?

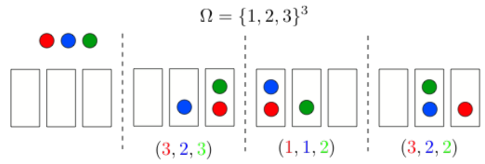


$Pr[B|A] = \frac{|B \cap A|}{|A|} = \frac{1}{6}$; versus $Pr[B] = \frac{1}{6}$.

Observing A does not change your mind about the likelihood of B.

Emptiness..

Suppose I toss 3 balls into 3 bins.
A ="1st bin empty"; B ="2nd bin empty." What is $Pr[A|B]$?



$\omega =$ (bin of red ball, bin of blue ball, bin of green ball)

$Pr[B] = Pr[\{(a,b,c) \mid a,b,c \in \{1,3\}] = Pr[\{1,3\}^3] = \frac{8}{27}$

$Pr[A \cap B] = Pr[(3,3,3)] = \frac{1}{27}$

$Pr[A|B] = \frac{Pr[A \cap B]}{Pr[B]} = \frac{(1/27)}{(8/27)} = 1/8$; vs. $Pr[A] = \frac{8}{27}$.

A is less likely given B:
Second bin is empty $\Longrightarrow$ first bin is more likely to contain ball(s).

Gambler's fallacy.

Flip a fair coin 51 times.
A = "first 50 flips are heads"
B = "the 51st is heads"
$Pr[B|A]$?

$A = \{HH \cdots HT, HH \cdots HH\}$
$B \cap A = \{HH \cdots HH\}$

Uniform probability space.

$Pr[B|A] = \frac{|B \cap A|}{|A|} = \frac{1}{2}$.

Same as $Pr[B]$.

The likelihood of 51st heads does not depend on the previous flips.

Product Rule

Recall the definition of conditional probability:

$$Pr[B|A] = \frac{Pr[A \cap B]}{Pr[A]}.$$

Hence,

$$Pr[A \cap B] = Pr[A]Pr[B|A].$$

Consequently,

$$\begin{aligned} Pr[A \cap B \cap C] &= Pr[(A \cap B) \cap C] \\ &= Pr[A \cap B]Pr[C|A \cap B] \\ &= Pr[A]Pr[B|A]Pr[C|A \cap B]. \end{aligned}$$

Product Rule

Theorem Product Rule
Let $A_1, A_2, \ldots, A_n$ be events. Then

$$Pr[A_1 \cap \cdots \cap A_n] = Pr[A_1]Pr[A_2|A_1] \cdots Pr[A_n|A_1 \cap \cdots \cap A_{n-1}].$$

Proof: By induction.
Assume the result is true for n. (It holds for $n = 2$.) Then,

$$\begin{aligned} & Pr[A_1 \cap \cdots \cap A_n \cap A_{n+1}] \\ &= Pr[A_1 \cap \cdots \cap A_n]Pr[A_{n+1}|A_1 \cap \cdots \cap A_n] \\ &= Pr[A_1]Pr[A_2|A_1] \cdots Pr[A_n|A_1 \cap \cdots \cap A_{n-1}]Pr[A_{n+1}|A_1 \cap \cdots \cap A_n], \end{aligned}$$

Thus, the result holds for $n+1$. □

Correlation

An example.
Random experiment: Pick a person at random.
Event A: the person has lung cancer.
Event B: the person is a heavy smoker.

Fact:

$$Pr[A|B] = 1.17 \times Pr[A].$$

Conclusion:

- Smoking increases the probability of lung cancer by 17%.
- Smoking causes lung cancer.

Correlation

Event A: the person has lung cancer.
Event B: the person is a heavy smoker.

$Pr[A|B] = 1.17 \times Pr[A]$.

A second look.

Note that

$$\begin{aligned} Pr[A|B] = 1.17 \times Pr[A] &\Leftrightarrow \frac{Pr[A \cap B]}{Pr[B]} = 1.17 \times Pr[A] \\ &\Leftrightarrow Pr[A \cap B] = 1.17 \times Pr[A]Pr[B] \\ &\Leftrightarrow \frac{Pr[A \cap B]}{Pr[A]} = 1.17 \times Pr[B]. \\ &\Leftrightarrow Pr[B|A] = 1.17 \times Pr[B]. \end{aligned}$$

Conclusion:

- Lung cancer increases the probability of smoking by 17%.
- Lung cancer causes smoking. Really?

Causality vs. Correlation

Events A and B are **positively correlated** if

$$Pr[A \cap B] > Pr[A]Pr[B].$$

(E.g., smoking and lung cancer.)

A and B being positively correlated does not mean that A causes B or that B causes A.

Other examples:

- Tesla owners are more likely to be rich. That does not mean that poor people should buy a Tesla to get rich.
- People who go to the opera are more likely to have a good career. That does not mean that going to the opera will improve your career.
- Rabbits eat more carrots and do not wear glasses. Are carrots good for eyesight?

Proving Causality

Proving causality is generally difficult. One has to eliminate external causes of correlation and be able to test the cause/effect relationship (e.g., randomized clinical trials).

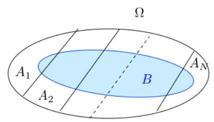
Some difficulties:

- A and B may be positively correlated because they have a common cause. (E.g., being a rabbit.)
- If B precedes A, then B is more likely to be the cause. (E.g., smoking.) However, they could have a common cause that induces B before A. (E.g., studious, CS70, Tesla.)

More about such questions later. For fun, check “N. Taleb: Fooled by randomness.”

Total probability

Assume that Ω is the union of the disjoint sets $A_1, \ldots, A_N$.



Then,

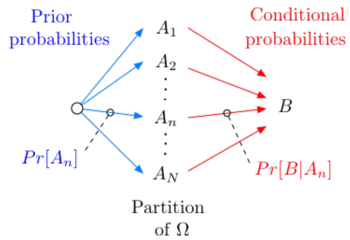
$$Pr[B] = Pr[A_1 \cap B] + \cdots + Pr[A_N \cap B].$$

Indeed, B is the union of the disjoint sets $A_n \cap B$ for $n = 1, \ldots, N$. Thus,

$$Pr[B] = Pr[A_1]Pr[B|A_1] + \cdots + Pr[A_N]Pr[B|A_N].$$

Total probability

Assume that Ω is the union of the disjoint sets $A_1, \ldots, A_N$.



$$Pr[B] = Pr[A_1]Pr[B|A_1] + \cdots + Pr[A_N]Pr[B|A_N].$$

Is you coin loaded?

Your coin is fair ($Pr[H] = 0.5$) w/prob $1/2$ or 'unfair' ($Pr[H] = 0.6$), otherwise.

You flip your coin and it yields heads.

What is the probability that it is fair?

Analysis:

$$A = \text{'coin is fair'}, B = \text{'outcome is heads'}$$

We want to calculate $P[A|B]$.

We know $P[B|A] = 1/2, P[B|\bar{A}] = 0.6, Pr[A] = 1/2 = Pr[\bar{A}]$

Now,

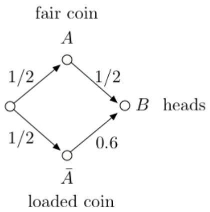
$$\begin{aligned} Pr[B] &= Pr[A \cap B] + Pr[\bar{A} \cap B] = Pr[A]Pr[B|A] + Pr[\bar{A}]Pr[B|\bar{A}] \\ &= (1/2)(1/2) + (1/2)0.6 = 0.55. \end{aligned}$$

Thus,

$$Pr[A|B] = \frac{Pr[A]Pr[B|A]}{Pr[B]} = \frac{(1/2)(1/2)}{(1/2)(1/2) + (1/2)0.6} \approx 0.45.$$

Is you coin loaded?

A picture:



Imagine 100 situations, among which
$m := 100(1/2)(1/2)$ are such that A and B occur and
$n := 100(1/2)(0.6)$ are such that $\bar{A}$ and B occur.

Thus, among the $m+n$ situations where B occurred, there are m where A occurred.

Hence,

$$Pr[A|B] = \frac{m}{m+n} = \frac{(1/2)(1/2)}{(1/2)(1/2)+(1/2)0.6}.$$

Independence

Definition: Two events A and B are **independent** if

$$Pr[A \cap B] = Pr[A]Pr[B].$$

Examples:

- When rolling two dice, A = sum is 7 and B = red die is 1 are independent; $Pr[A \cap B] = \frac{1}{36}$, $Pr[A]Pr[B] = (\frac{1}{6})(\frac{1}{6})$.
- When rolling two dice, A = sum is 3 and B = red die is 1 are not independent; $Pr[A \cap B] = \frac{1}{36}$, $Pr[A]Pr[B] = (\frac{2}{36})(\frac{1}{6})$.
- When flipping coins, A = coin 1 yields heads and B = coin 2 yields tails are independent; $Pr[A \cap B] = \frac{1}{4}$, $Pr[A]Pr[B] = (\frac{1}{2})(\frac{1}{2})$.
- When throwing 3 balls into 3 bins, A = bin 1 is empty and B = bin 2 is empty are not independent; $Pr[A \cap B] = \frac{1}{27}$, $Pr[A]Pr[B] = (\frac{8}{27})(\frac{8}{27})$.

Independence and conditional probability

Fact: Two events A and B are **independent** if and only if

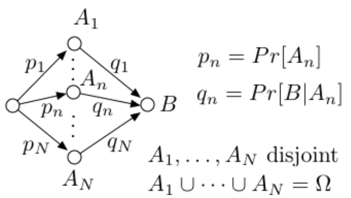
$$Pr[A|B] = Pr[A].$$

Indeed: $Pr[A|B] = \frac{Pr[A \cap B]}{Pr[B]}$, so that

$$Pr[A|B] = Pr[A] \Leftrightarrow \frac{Pr[A \cap B]}{Pr[B]} = Pr[A] \Leftrightarrow Pr[A \cap B] = Pr[A]Pr[B].$$

Bayes Rule

Another picture: We imagine that there are N possible causes $A_1, \ldots, A_N$.

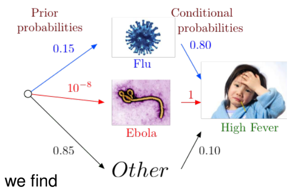


Imagine 100 situations, among which $100 p_n q_n$ are such that A_n and B occur, for $n = 1, \ldots, N$.
Thus, among the $100 \sum_m p_m q_m$ situations where B occurred, there are $100 p_n q_n$ where A_n occurred.

Hence,

$$Pr[A_n|B] = \frac{p_n q_n}{\sum_m p_m q_m}.$$

Why do you have a fever?



Using Bayes' rule, we find

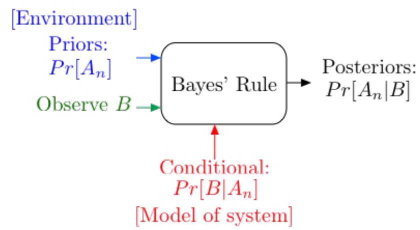
$$Pr[\text{Flu}|\text{High Fever}] = \frac{0.15 \times 0.80}{0.15 \times 0.80 + 10^{-8} \times 1 + 0.85 \times 0.1} \approx 0.58$$

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

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

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

Bayes' Rule Operations



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

Thomas Bayes

Thomas Bayes

Portrait used of Bayes in a 1936 book,[1] but it is doubtful whether the portrait is actually of him.[2] No earlier portrait or claimed portrait survives.

Born	c. 1701 London, England
Died	7 April 1761 (aged 59) Tunbridge Wells, Kent, England
Residence	Tunbridge Wells, Kent, England
Nationality	English
Known for	Bayes' theorem

Source: Wikipedia.

Thomas Bayes

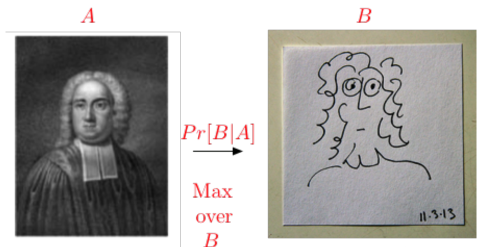


FIG. 3. *Joshua Bayes (1671–1746).*

Thomas Bayes?

A Bayesian picture of Thomas Bayes.

Testing for disease.

Let's watch TV!!
Random Experiment: Pick a random male.
Outcomes: $(test, disease)$
A - prostate cancer.
B - positive PSA test.

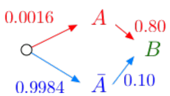
- $Pr[A] = 0.0016$, (.16 % of the male population is affected.)
- $Pr[B|A] = 0.80$ (80% chance of positive test with disease.)
- $Pr[B|\bar{A}] = 0.10$ (10% chance of positive test without disease.)

From http://www.cpcn.org/01_psa_tests.htm and http://seer.cancer.gov/statfacts/html/prost.html (10/12/2011.)

Positive PSA test (B). Do I have disease?

$$Pr[A|B]???$$

Bayes Rule.



Using Bayes' rule, we find

$$P[A|B] = \frac{0.0016 \times 0.80}{0.0016 \times 0.80 + 0.9984 \times 0.10} = .013.$$

A 1.3% chance of prostate cancer with a positive PSA test.

Surgery anyone?

Impotence...

Incontinence..

Death.

Summary

Events, Conditional Probability, Independence, Bayes' Rule

Key Ideas:

- Conditional Probability:

  $Pr[A|B] = \frac{Pr[A \cap B]}{Pr[B]}$
- Independence: $Pr[A \cap B] = Pr[A]Pr[B]$.
- Bayes' Rule:

  $$Pr[A_n|B] = \frac{Pr[A_n]Pr[B|A_n]}{\sum_m Pr[A_m]Pr[B|A_m]}.$$

  $Pr[A_n|B]$ = posterior probability; $Pr[A_n]$ = prior probability .
- All these are possible:

  $Pr[A|B] < Pr[A]; Pr[A|B] > Pr[A]; Pr[A|B] = Pr[A]$.