6 Event-based Independence and Conditional Probability

Let X be the result

with "sneak peek"

Example 6.1. Roll a dice...

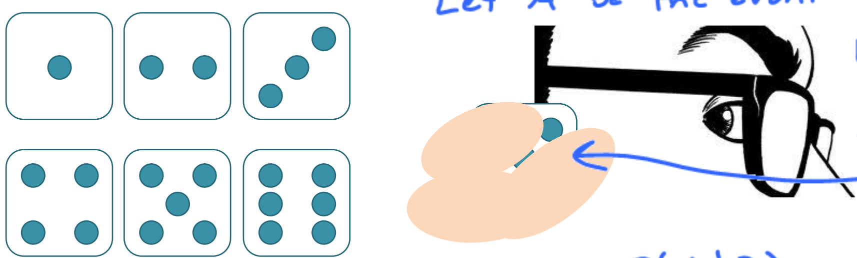
Let A be the event that $X = 2$.

Let B be the event that you observe

Find the probability that $X = 2$

$P[X=2] = \frac{1}{6}$

$P(A) = \frac{1}{6}$



~~$P[X=2] = 0$~~

$P(A|B) = 0$

Figure 10: Conditional Probability Example: Sneak Peek

"set minus" operator

$P(A \setminus B) = P(A \cap B^c)$

$P(A|B) = \frac{P(A \cap B)}{P(B)}$

Example 6.2 (Slides)**.** Diagnostic Tests.

6.1 Event-based Conditional Probability

Definition 6.3. ***Conditional Probability***: The conditional probability $P(A|B)$ of event A given that event $B \neq \emptyset$ occurred, is given by

$P(A^c|B^c) = \frac{P(A^c \cap B^c)}{P(B^c)}$

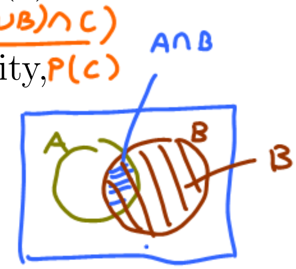
$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{6}$$

condition

$P(A \cup B|C) = \frac{P((A \cup B) \cap C)}{P(C)}$

$P(A) = \frac{P(A)}{1} = \frac{P(A)}{P(\Omega)}$

- Some ways to say[23] or express the conditional probability, $P(A|B)$, are:
  - the "(conditional) probability of A, given B"
  - the "(conditional) probability of A, knowing B"
  - the "(conditional) probability of A happening, knowing B has already occurred"
  - the "(conditional) probability of A, given that B occurred"
  - the "(conditional) probability of an event A under the knowledge that the outcome will be in event B"



ratio of areas

[23]Note also that although the symbol $P(A|B)$ itself is practical, it phrasing in words can be so unwieldy that in practice, less formal descriptions are used. For example, we refer to "the probability that a tested-positive person has the disease" instead of saying "the conditional probability that a randomly chosen person has the disease given that the test for this person returns positive result."

If you can write $P(A|B)$, then you implicitly assume that $P(B) > 0$.

- Defined only when $P(B) > 0$.
  - If $P(B) = 0$, then it is illogical to speak of $P(A|B)$; that is $P(A|B)$ is not defined.

6.4. *Interpretation*: It is sometimes useful to interpret $P(A)$ as our knowledge of the occurrence of event A *before* the experiment takes place. Conditional probability[24] $P(A|B)$ is the ***updated probability*** of the event A given that we now know that B occurred (but we still do not know which particular outcome in the set B did occur). In general, $P(A)$ and $P(A|B)$ are not the same.

Definition 6.5. Sometimes, we refer to $P(A)$ as	$P(A\|B)$
• **a priori probability**, or	a posteriori probability
• the **prior probability** of A, or	posterior probability
• the **unconditional probability** of A.	conditional probability

Example 6.6. Back to Example 6.1. Roll a dice. Let X be the outcome.

$|\Omega| = 24 = 6 \times 4$

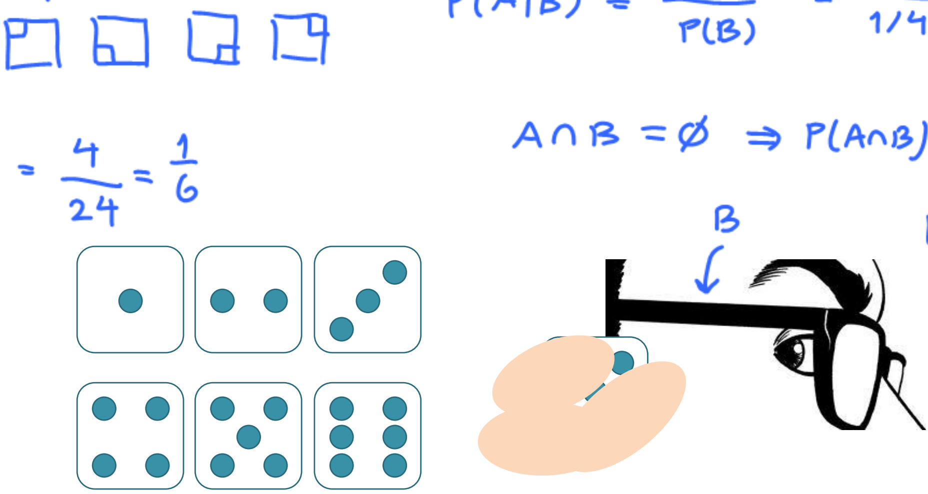
"$X = 2$"

$P(A) = \frac{|A|}{|\Omega|} = \frac{4}{24} = \frac{1}{6}$

$P(B) = \frac{2+4}{24} = \frac{6}{24} = \frac{1}{4}$

$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{1/4} = 0$

$A \cap B = \emptyset \Rightarrow P(A \cap B) = 0$



Let C be the event that $X = 5$

$P(C|B) = \frac{P(B \cap C)}{P(B)}$

Figure 11: Sneak Peek: A Revisit

$|B \cap C| = 4 \quad P(B \cap C) = \frac{4}{24} = \frac{1}{6}$

$P(C|B) = \frac{1/6}{1/4} = \frac{4}{6} = \frac{2}{3}$

[24]In general, $P(A)$ and $P(A|B)$ are not the same. However, in the next section (Section 6.2), we will consider the situation in which they are the same.

Example 6.7. In diagnostic tests Example 6.2, we learn whether we have the disease from test result. Originally, before taking the test, the probability of having the disease is 0.01%. Being tested positive from the 99%-accurate test ***updates*** the probability of having the disease to about 1%.

actually

More specifically, let D be the event that the testee has the disease and T_P be the event that the test returns positive result.

- Before taking the test, the probability of having the disease is $P(D) = 0.01\%$. $\frac{1}{10{,}000}$
- Using 99%-accurate test means

test negative

$$P(T_P|D) = 0.99 \text{ and } P(T_P^c|D^c) = 0.99.$$

- Our calculation shows that $P(D|T_P) \approx 0.01$.

6.8. "Prelude" to the concept of "independence":
If the occurrence of B does not give you more information about A, then

$$P(A|B) = P(A) \tag{7}$$

and we say that A and B are ***independent***.

- Meaning: "learning that event B has occurred does not change the probability that event A occurs."

We will soon define "independence" in Section 6.2. Property (7) can be regarded as a "practical" definition for independence. However, there are some "technical" issues[25] that we need to deal with when we actually define independence.

6.9. When Ω is finite and all outcomes have equal probabilities, $\Rightarrow$ classical probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{|A \cap B| / \cancel{|\Omega|}}{|B| / \cancel{|\Omega|}} = \frac{|A \cap B|}{|B|}.$$

This formula can be regarded as the classical version of conditional probability.

[25] Here, the statement assume $P(B) > 0$ because it considers $P(A|B)$. The concept of independence to be defined in Section 6.2 will not rely directly on conditional probability and therefore it will include the case where $P(B) = 0$.

Exercise 6.10. Someone has rolled a fair dice twice. You know that one of the rolls turned up a face value of six. What is the probability that the other roll turned up a six as well?
Ans: $\frac{1}{11}$ (not $\frac{1}{6}$). [21, Example 8.1, p. 244]

Example 6.11. Consider the following sequences of 1s and 0s which summarize the data obtained from 15 testees.

person #1 person #2 person #15

```
D:  0 1 1 0 0 0 0 1 1 1 1 0 1 0 1
TP: 1 0 0 1 1 0 0 0 0 0 1 1 0 1 1
```

The "D" row indicates whether each of the testees actually has the disease under investigation. The "TP" row indicates whether each of the testees is tested positive for the disease.

Numbers "1" and "0" correspond to "True" and "False", respectively.

Suppose we randomly pick a testee from this pool of 15 persons. Let D be the event that this selected person actually has the disease. Let T_P be the event that this selected person is tested positive for the disease.

Find the following probabilities.

(a) $P(D) = \frac{8}{15}$ Among the 15 testees, 8 have the disease

(b) $P(D^c) = \frac{7}{15}$ $1 - P(D)$

(c) $P(T_P) = \frac{7}{15}$

(d) $P(T_P^c) = 1 - P(T_p) = \frac{8}{15}$

(e) $P(T_P|D) = \frac{2}{8} = \frac{1}{4}$ Among the 8 testees who have the disease, 2 test positive.

(f) $P(T_P|D^c) = \frac{P(T_p \cap D^c)}{P(D^c)} = \frac{5/15}{7/15} = \frac{5}{7}$

(g) $P(T_P^c|D) = 1 - P(T_p|D) = 1 - \frac{1}{4} = \frac{3}{4}$

(h) $P(T_P^c|D^c) = 1 - P(T_p|D^c) = 1 - \frac{5}{7} = \frac{2}{7}$

$P(A^c|B) = 1 - P(A|B)$

$P(A^c|B^c) = 1 - P(A|B^c)$

6.12. Similar properties to the three probability axioms:

≥ 0

(a) Nonnegativity: $P(A|B) \geq 0$ $\quad = \dfrac{P(A\cap B)}{P(B)}$ $\quad \geq 0$, $\neq 0$

(b) Unit normalization: $P(\Omega|B) = 1$.

$\dfrac{P(\Omega \cap B)}{P(B)} = \dfrac{P(B)}{P(B)}$

In fact, for any event A such that $B \subset A$, we have $P(A|B) = 1$.

This implies

$$P(\Omega|B) = P(B|B) = 1.$$

(c) Countable additivity: For every countable sequence $(A_n)_{n=1}^{\infty}$ of disjoint events,

disjoint union

$$P\left(\bigcup_{n=1}^{\infty} A_n \,\middle|\, B\right) = \sum_{n=1}^{\infty} P\left(A_n | B\right).$$

Finite additivity

$$P\left(\bigcup_{n=1}^{N} A_n \,\middle|\, B\right) = \sum_{n=1}^{N} P(A_n|B)$$

- In particular, if $A_1 \perp A_2$, (disjoint)

$$P\left(A_1 \cup A_2 | B\right) = P(A_1|B) + P(A_2|B)$$

$\perp$ disjoint (CH2)

$\perp\!\!\!\perp$ independent (Sec. 6.2)

Ш sequence of delta func. (ECS332)

6.13. More Properties:

- $P(A|\Omega) = P(A)$
- $P(A^c|B) = 1 - P(A|B)$

  $P(A^c|B^c) = 1 - P(A|B^c)$

99% accurate test: $P(T_p|D) = 0.99 \Rightarrow P(T_p^c|D) = 1 - 0.99 = 0.01$

$P(T_p^c|D^c) = 0.99 \Rightarrow P(T_p|D^c) = 1 - 0.99 = 0.01$

- $P(A \cap B|B) = P(A|B)$
- $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$.
- $P(A \cap B) \leq P(A|B)$

$P(A|B) \equiv \frac{P(A\cap B)}{P(B)}$ ①

marginal probabilities

6.14. Probability of compound events

general "multiplication rule"

(a) $\underbrace{P(A\cap B)}_{\text{joint event}} \overset{②}{=} P(A)P(B|A) \overset{①}{=} P(B)P(A|B)$

joint probability

marginal events

$A\cap B$

$P(B|A) \equiv \frac{P(B\cap A)}{P(A)}$ ②

(b) $P(A\cap B\cap C) = P(A\cap B)\times P(C|A\cap B)$

(c) $P(A\cap B\cap C) = P(A)\times P(B|A)\times P(C|A\cap B)$

When we have many sets intersected in the conditioning part, we often use "," instead of "∩".

Example 6.15. Most people reason as follows to find the probability of getting two aces ($A\cap B$) when two cards are selected at random from an ordinary deck of cards:

(a) The probability of "getting an ace on the first card" (A) is 4/52. $P(A)$

(b) Given that one ace is gone from the deck, the probability of "getting an ace on the second card" (B) is 3/51. $P(B|A)$

(c) The desired probability is therefore

$$\frac{4}{52}\times\frac{3}{51}.$$

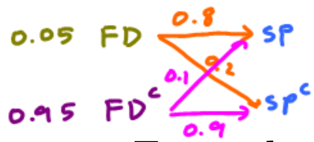
$P(A\cap B) = P(A)P(B|A)$

$\frac{4\times 3}{52\times 51}$

[21, p 243]

Question: What about the unconditional probability $P(B)$?

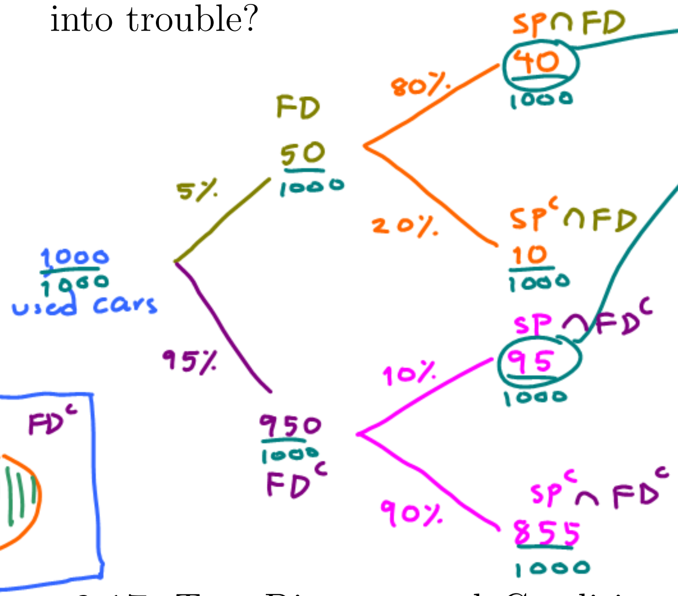
compact version :



$P(SP) = (0.05 \times 0.8) + (0.95 \times 0.1)$

P(FD) P(SP|FD) FD SP $P(SP|FD^c)$

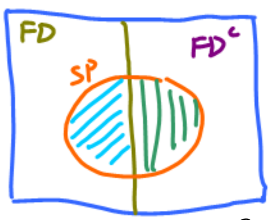
Example 6.16. You know that roughly 5% of all used cars have been flood-damaged and estimate that 80% of such cars will later develop serious engine problems, whereas only 10% of used cars that are not flood-damaged develop the same problems. Of course, no used car dealer worth his salt would let you know whether your car has been flood damaged, so you must resort to probability calculations. What is the probability that your car will later run into trouble?



$P(SP) = \frac{40+95}{1000} = 0.135$

$P(SP) = P(SP \cap FD) + P(SP \cap FD^c)$

$= \underbrace{P(SP|FD)}_{0.8}\underbrace{P(FD)}_{0.05} + \underbrace{P(SP|FD^c)}_{0.1}\underbrace{P(FD^c)}_{0.95}$



6.17. Tree Diagram and Conditional Probability: Conditional probabilities can be represented on a tree diagram as shown in Figure 12.

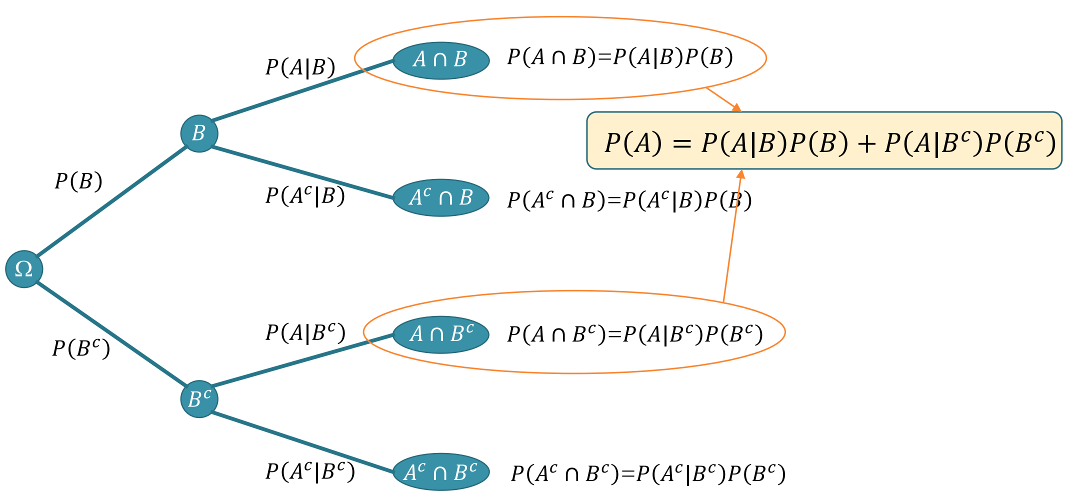


Figure 12: Tree Diagram and Conditional Probabilities

A more compact representation is shown in Figure 13.

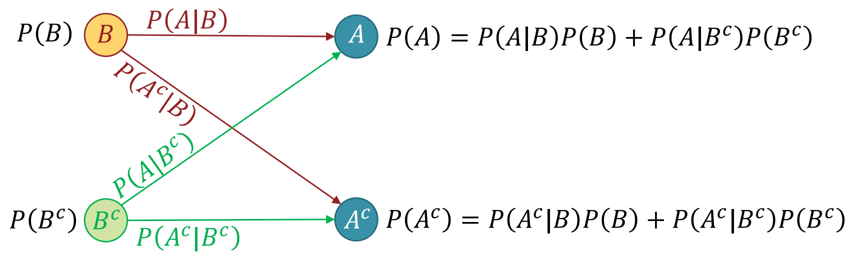


Figure 13: Compact Diagram for Conditional Probabilities

Example 6.18. A simple digital communication channel called **binary symmetric channel** (BSC) is shown in Figure 6.58. This channel can be described as a channel that introduces random bit errors with probability p.

A = the event that "Y = 1"

B = the event that "X = 1"

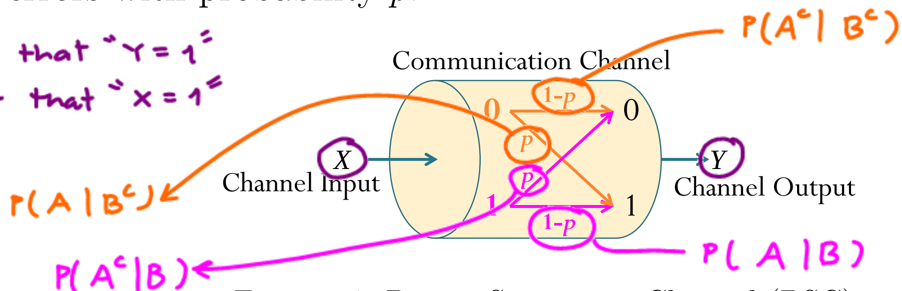
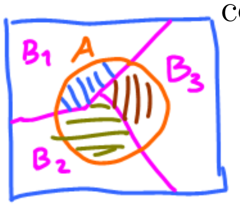


Figure 14: Binary Symmetric Channel (BSC)

6.19. ***Total Probability Theorem***: If a (finite or infinitely) countable collection of events $\{B_1, B_2, \ldots\}$ is a partition of Ω, then



① $B_1, B_2, \ldots$ are disjoint

$$P(A) = \sum_i P(A|B_i)P(B_i). \tag{8}$$

② $\bigcup_k B_k = \Omega$

This is a formula[26] for computing the probability of an event that can occur in different ways. Observe that it follows directly from 5.21 and Definition 6.3.

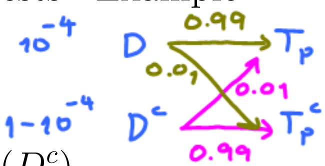
[26]The tree diagram is useful for helping you understand the process. However, when the number of possible cases is large (many B_i for the partition), drawing the tree diagram may be too time-consuming and therefore you should also learn how to apply the total probability theorem directly without the help of the tree diagram.

$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$

$= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)$

- Special case: $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$.
  This gives exactly the same calculation as what we discussed in Example 6.16.

Example 6.20. Continue from the "Diagnostic Tests" Example 6.2 and Example 6.7.



$$\begin{aligned} P\left(T_P\right) &= P\left(T_P \cap D\right) + P\left(T_P \cap D^c\right) \\ &= P\left(T_P \left| D\right.\right) P\left(D\right) + P\left(T_P \left| D^c\right.\right) P\left(D^c\right). \end{aligned}$$

For conciseness, we define

$P(T_p) = 10^{-4}(0.99) + (1-10^{-4})(0.01)$

$$p_d = P(D)$$

and

$$p_{TE} = P(T_P|D^c) = P(T_P^c|D).$$

T_p D

$P(A|B)$

$P(B|A)$

Then,

$$P\left(T_P\right) = (1 - p_{TE})p_D + p_{TE}(1 - p_D).$$

$= 0.99 \times 10^{-4} + 0.01(1-10^{-4})$

≈ 0.01

6.21. Bayes' Theorem:

(a) Form 1:

$P(B|A) \equiv \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}$

$$P(B|A) = P(A|B)\frac{P(B)}{P(A)}.$$

$P(D \mid T_p) = \underbrace{P(T_p|D)}_{0.99} \frac{\overbrace{P(D)}^{10^{-4}}}{P(T_p)}$ ≈ 0.01

(b) Form 2: If a (finite or infinitely) countable collection of events $\{B_1, B_2, \ldots\}$ is a partition of Ω, then

$$P(B_k|A) = P(A|B_k)\frac{P(B_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{\sum_i P(A|B_i)P(B_i)}.$$

- Extremely useful for making inferences about phenomena that cannot be observed directly.

total probability theorem

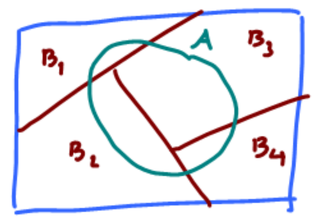
- Sometimes, these inferences are described as "reasoning about causes when we observe effects".

6.22. Summary: Def: $P(B|A) = \dfrac{P(A\cap B)}{P(A)}$

(a) An easy but crucial property:

$$P(A\cap B) = P(B|A)P(A) = P(A|B)P(B)$$

(b) Key setup: find a $\underbrace{\text{partition}}_{\{B_1, B_2, \dots, B_4\}}$ of the sample space



(c) Total probability theorem:

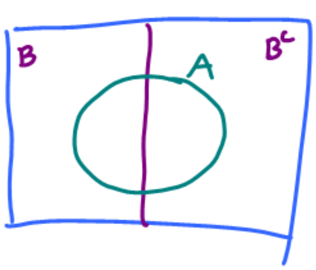
$$P(A) = P(A\cap B_1) + P(A\cap B_2) + \cdots + P(A\cap B_4)$$

$$= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_4)P(B_4)$$

(d) Bayes' theorem: (form #1) (form #2)

$$P(B_2|A) \overset{\text{Defn.}}{=} \frac{P(A\cap B_2)}{P(A)} = \frac{P(A|B_2)P(B_2)}{P(A)} = \frac{P(A|B_2)P(B_2)}{\bigcirc}$$

- Special case: When there are only two cases: B_1 and B_2, we can think of them as B and B^c, respectively:



  - $P(A) = P(A\cap B) + P(A\cap B^c) = P(A|B)P(B) + P(A|B^c)P(B^c)$
  - $P(B|A) = \dfrac{P(B\cap A)}{P(A)} = \dfrac{P(A|B)P(B)}{P(A)} = \dfrac{P(A|B)P(B)}{}$
  - $P(B|A^c) = \dfrac{P(A^c|B)P(B)}{P(A^c)} = \dfrac{(1-P(A|B))P(B)}{1-P(A)}$

Example 6.23. Suppose $\Omega = \{a, b, c, d, e\}$. Define four events

$$A = \{a, b, c\},\ B = \{a, b\},\ C = \{c, d\},\ \text{and } D = \{e\}.$$

Let

$$P(\{a\}) = P(\{b\}) = 0.2, \quad \text{and} \quad P(\{c\}) = P(\{d\}) = 0.1.$$

Calculate the following probabilities:

(a) $P(\{e\}) = 1 - P(\{e\}^c) = 1 - P(\{a,b,c,d\}) = 1 - (0.2+0.2+0.1+0.1) = 0.4$

(b) $P(B) = P(\{a,b\}) = 0.2 + 0.2 = 0.4$, $P(C) = P(\{c,d\}) = 0.1 + 0.1 = 0.2$

$P(D) = P(\{e\}) = 0.4$ from part (a).

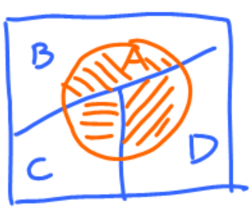
(c) $P(A|B) = \dfrac{P(A\cap B)}{P(B)} = \dfrac{P(\{a,b\})}{0.4} = \dfrac{0.4}{0.4} = 1$

$P(A|C) = \dfrac{P(A\cap C)}{P(C)} = \dfrac{P(\{c\})}{0.2} = \dfrac{0.1}{0.2} = \dfrac{1}{2}$

$P(A|D) = \dfrac{P(A\cap D)}{P(D)} = \dfrac{P(\emptyset)}{0.4} = \dfrac{0}{0.4} = 0$

(d) $P(A) = P(\{a,b,c\}) = 0.2 + 0.2 + 0.1 = 0.5$

Check: Observe that the collection $\{B, C, D\}$ partitions Ω. Use the total probability theorem to find $P(A)$.



A happens with D

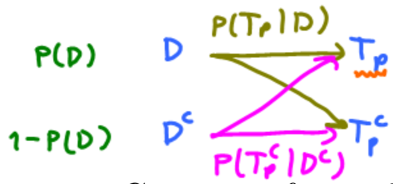
$$P(A) = P(A\cap B) + P(A\cap C) + P(A\cap D)$$
$$= P(A|B)P(B) + P(A|C)P(C) + P(A|D)P(D)$$
$$= 1 \times 0.4 + 1/2 \times 0.2 + 0 \times 0.4 = 0.5$$

A happens with B

A happens with C

(e) $P(B|A) = \dfrac{P(A|B)\,P(B)}{P(A)} = 1 \times \dfrac{0.4}{0.5} = \dfrac{4}{5}$

Bayes' theorem



Example 6.24. Continue from the "Disease Testing" Examples 6.2, 6.7, and 6.20:

Bayes' theorem

$$P(D|T_P) = \frac{P(D \cap T_P)}{P(T_P)} = \frac{P(T_P|D)\,P(D)}{P(T_P)} = \frac{P(T_p|D)P(D)}{P(D)P(T_p|D)+P(D^c)P(T_p|D^c)}$$

0.99, 0.99, 0.01

$$= \frac{(1-p_{TE})p_D}{(1-p_{TE})p_D + p_{TE}(1-p_D)}$$

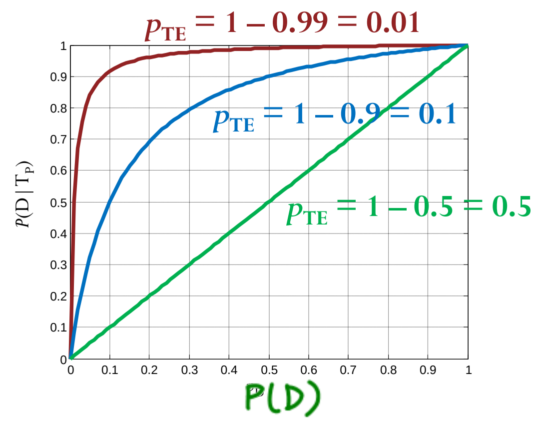


Figure 15: Probability $P(D|T_P)$ that a person will have the disease given that the test result is positive. The conditional probability is evaluated as a function of P_D which tells how common the disease is. Thee values of test error probability p_{TE} are shown.

Example 6.25. Medical Diagnostic: Because a new medical procedure has been shown to be effective in the early detection of an illness, a medical screening of the population is proposed. The probability that the test correctly identifies someone with the illness as positive is 0.99. and the probability that the test correctly identifies someone without the illness as negative is 0.95. The incidence of the illness in the general population is 0.0001. You take the test, and the result is positive. What is the probability that you have the illness? [15, Ex. 2-37]

$P(T_p|D)$

$P(T_p^c|D^c) = 0.95$

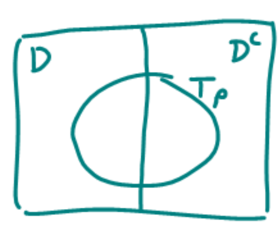
$P(D) = 10^{-4}$

PPV

$$P(D|T_p) = \frac{P(D\cap T_p)}{P(T_p)} = \frac{P(T_p|D)P(D)}{P(T_p)} = \frac{0.99\times 10^{-4}}{0.0501} \approx 0.0020$$

$$P(T_p) = P(T_p\cap D) + P(T_p \cap D^c) = P(T_p|D)P(D) + P(T_p|D^c)P(D^c)$$

$$= 0.99\times 10^{-4} + (1-0.95)(1-10^{-4}) = 0.0501$$



Example 6.26. Bayesian networks are used on the Web sites of high-technology manufacturers to allow customers to quickly diagnose problems with products. An oversimplified example is presented here.

A printer manufacturer obtained the following probabilities from a database of test results. Printer failures are associated with three types of problems: hardware (H), software (S), and other (O) (such as connectors), with probabilities 0.1, 0.6, and 0.3, respectively. The probability of a printer (F) failure given a hardware problem is 0.9, given a software problem is 0.2, and given any other type of problem is 0.5. If a customer enters the manufacturers Web site to diagnose a printer failure, what is the most likely cause of the problem?

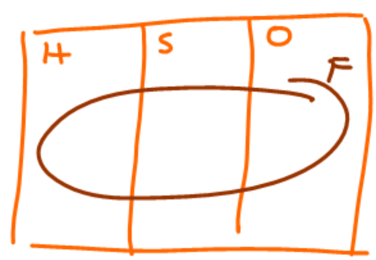
Let the events H, S, and O denote a hardware, software, or other problem, respectively, and let F denote a printer failure.

$P(H)=0.1$ $\quad P(F|H)=0.9$

$P(S)=0.6$ $\quad P(F|S)=0.2$

$P(O)=0.3$ $\quad P(F|O)=0.5$

$$P(F)=P(F\cap H)+P(F\cap S)+P(F\cap O)$$
$$=P(F|H)P(H)+P(F|S)P(S)+P(F|O)P(O)$$
$$=0.36$$

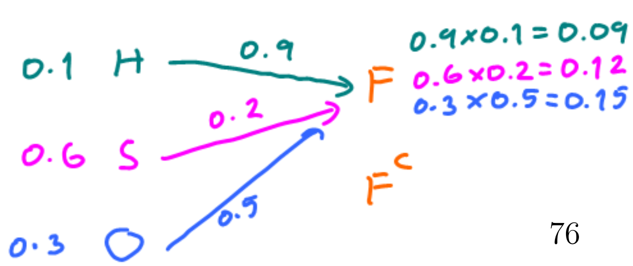


$$P(H|F)=\frac{P(H\cap F)}{P(F)}=\frac{P(F|H)P(H)}{P(F)}=\frac{0.9\times0.1}{0.36}=\frac{1}{4}$$

$$P(S|F)=\frac{P(S\cap F)}{P(F)}=\frac{P(F|S)P(S)}{P(F)}=\frac{0.2\times0.6}{0.36}=\frac{1}{3}$$

$$P(O|F)=\frac{P(O\cap F)}{P(F)}=\frac{P(F|O)P(O)}{P(F)}=\frac{0.5\times0.3}{0.36}=\frac{5}{12}$$

$P(O|F)$ is largest. This case is the mostly scenario.



Example 6.27 (Slides)**.** The Murder of Nicole Brown

6.28. Chain rule of conditional probability [9, p 58]:

$$P(A \cap B|C) = P(B|C)P(A|B \cap C).$$

6.29. In practice, here is how we use the total probability theorem and Bayes' theorem:

Usually, we work with a system, which of course has input and output. There can be many possibilities for inputs and there can be many possibilities for output. Normally, for deterministic system, we may have a specification that tells what would be the output given that a specific input is used. Intuitively, we may think of this as a table of mapping between input and output. For system with random component(s), when a specific input is used, the output is not unique. This mean we needs conditional probability to describe the output (given an input). Of course, this conditional probability can be different for different inputs.

We will assume that there are many cases that the input can happen. The event that the ith case happens is denoted by B_i. We assume that we consider all possible cases. Therefore, the union of these B_i will automatically be Ω. If we also define the cases so that they do not overlap, then the B_i partitions Ω.

Similarly, there are many cases that the output can happen. The event that the jth case happens is depenoted by A_j. We assume that the A_j also partitions Ω.

In this way, the system itself can be described by the conditional probabilities of the form $P(A_j|B_i)$. This replace the table mentioned above as the specification of the system. Note that even when this information is not available, we can still obtain an approximation of the conditional probability by repeating trials of inputting B_i in to the system to find the relative frequency of the output A_j.

Now, when the system is used in actual situation. Different input cases can happen with different probabilities. These are described by the prior probabilities $P(B_i)$. Combining this with the conditional probabilities $P(A_j|B_i)$ above, we can use the total probability theorem to find the probability of occurrence for output and, even more importantly, for someone who cannot directly observe the input, Bayes' theorem can be used to infer the value (or the probability) of the input from the observed output of the system.

In particular, total probability theorem deals with the calculation of the output probabilities $P(A_j)$:

$$P(A_j) = \sum_i P(A_j \cap B_i) = \sum_i P(A_j|B_i) P(B_i).$$

Bayes' theorem calculates the probability that B_k was the input event when the observer can only observe the output of the system and the observed value of the output is A_j:

$$P(B_k|A_j) = \frac{P(A_j \cap B_k)}{P(A_j)} = \frac{P(A_j|B_k)\,P(B_k)}{\sum_i P(A_j|B_i)\,P(B_i)}.$$

Example 6.30. In the early 1990s, a leading Swedish tabloid tried to create an uproar with the headline "Your ticket is thrown away!". This was in reference to the popular Swedish TV show "Bingolotto" where people bought lottery tickets and mailed them to the show. The host then, in live broadcast, drew one ticket from a large mailbag and announced a winner. Some observant reporter noticed that the bag contained only a small fraction of the hundreds of thousands tickets that were mailed. Thus the conclusion: Your ticket has most likely been thrown away!

Let us solve this quickly. Just to have some numbers, let us say that there are a total of $N = 100,000$ tickets and that $n = 1,000$ of them are chosen at random to be in the final drawing. If the drawing was from all tickets, your chance to win would be $1/N = 1/100,000$. The way it is actually done, you need to both survive the first drawing to get your ticket into the bag and then get your ticket drawn from the bag. The probability to get your entry into the bag is $n/N = 1,000/100,000$. The conditional probability to be drawn from the bag, given that your entry is in it, is $1/n = 1/1,000$. Multiply to get $1/N = 1/100,000$ once more. There were no riots in the streets. [17, p 22]

Example 6.31. Suppose your professor tells the class that there will be a surprise quiz next week. On one day, Monday-Friday, you will be told in the morning that a quiz is to be given on that day. You quickly realize that the quiz will not be given on Friday; if it was, it would not be a surprise because it is the last possible day to get the quiz. Thus, Friday is ruled out, which leaves Monday-Thursday. But then Thursday is impossible also, now having become the last possible day to get the quiz. Thursday is ruled out, but then Wednesday becomes impossible, then Tuesday, then Monday, and you conclude: There is no such thing as a surprise quiz! But the professor decides to give the quiz on Tuesday, and come Tuesday morning, you are surprised indeed.

This problem, which is often also formulated in terms of surprise fire drills or surprise executions, is known by many names, for example, the "hangman's paradox" or by serious philosophers as the "prediction paradox." To resolve it, let's treat it as a probability problem. Suppose that the day of the quiz is chosen randomly among the five days of the week. Now start a new school week. What is the probability that you get the test on Monday? Obviously 1/5 because this is the probability that Monday is chosen. If the test was not given on Monday. what is the probability that it is given on Tuesday? The probability that Tuesday is chosen to start with is 1/5, but we are now asking for the conditional probability that the test is given on Tuesday, given that it was not given on Monday. As there are now four days left, this conditional probability is 1/4. Similarly, the conditional probabilities that the test is given on Wednesday, Thursday, and Friday conditioned on that it has not been given thus far are 1/3, 1/2, and 1, respectively.

We could define the "surprise index" each day as the probability that the test is not given. On Monday, the surprise index is therefore 0.8, on Tuesday it has gone down to 0.75, and it continues to go down as the week proceeds with no test given. On Friday, the surprise index is 0, indicating absolute certainty that the test will be given that day. Thus, it is possible to give a surprise test but not in a way so that you are equally surprised each day, and it is never possible to give it so that you are surprised on Friday. [17, p 23–24]

Example 6.32. Today Bayesian analysis is widely employed throughout science and industry. For instance, models employed to determine car insurance rates include a mathematical function describing, per unit of driving time, your personal probability of having zero, one, or more accidents. Consider, for our purposes, a simplified model that places everyone in one of two categories: high risk, which includes drivers who average at least one accident each year, and low risk, which includes drivers who average less than one.

If, when you apply for insurance, you have a driving record that stretches back twenty years without an accident or one that goes back twenty years with thirty-seven accidents, the insurance company can be pretty sure which category to place you in. But if you are a new driver, should you be classified as low risk (a kid who obeys the speed limit and volunteers to be the designated driver) or high risk (a kid who races down Main Street swigging from a half-empty $2 bottle of Boone's Farm apple wine)?

Since the company has no data on you, it might assign you an equal prior probability of being in either group, or it might use what it knows about the general population of new drivers and start you off by guessing that the chances you are a high risk are, say, 1 in 3. In that case the company would model you as a hybrid–one-third high risk and two-thirds low risk–and charge you one-third the price it charges high-risk drivers plus two-thirds the price it charges low-risk drivers.

Then, after a year of observation, the company can employ the new datum to reevaluate its model, adjust the one-third and two-third proportions it previously assigned, and recalculate what it ought to charge. If you have had no accidents, the proportion of low risk and low price it assigns you will increase; if you have had two accidents, it will decrease. The precise size of the adjustment is given by Bayes's theory. In the same manner the insurance company can periodically adjust its assessments in later years to reflect the fact that you were accident-free or that you twice had an accident while driving the wrong way down a one-way street, holding a cell phone with your left hand and a doughnut with your right. That is why insurance companies can give out "good driver" discounts: the absence of accidents elevates the posterior probability that a driver belongs in a low-risk group. [14, p 111-112]