

1. Probability

Event: a possible outcome or set of possible outcomes of an experiment or observation. Typically denoted by a capital letter: A, B etc.

E.g. The result of a coin toss

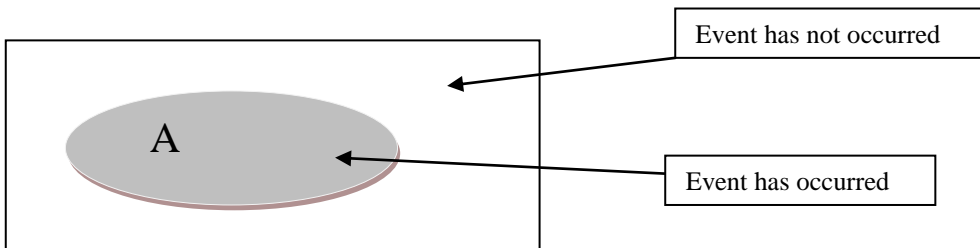
Probability of an event A : denoted by $P(A)$. Measured on a scale between 0 and 1 inclusive. If A is impossible $P(A) = 0$, if A is certain then $P(A)=1$.

E.g. $P(\text{result of a coin toss is heads})$.

If there a fixed number of equally likely outcomes $P(A)$ is the fraction of the outcomes that are in A .
E.g. for a coin toss there are two possible outcomes, Heads or Tails, so

$P(\text{result of a coin toss is heads}) = 1/2$.

Intuitive idea: $P(A)$ is the typical fraction of times A would occur if an experiment were repeated very many times.



Probability of a statement S : $P(S)$ denotes degree of belief that S is true.

E.g. $P(\text{tomorrow it will rain})$.

Probability theory gives a consistent way of reasoning about random events or uncertain propositions.

Conditional probability: $P(A|B)$ means the probability of A given that B has happened or is true.

e.g. $P(\text{result of coin toss is heads} \mid \text{the coin is fair}) = 1/2$

$P(\text{Tomorrow is Tuesday} \mid \text{it is Monday}) = 1$

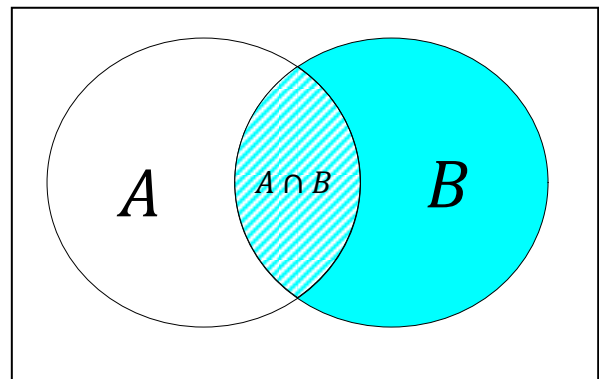
$P(\text{card is a heart} \mid \text{it is a red suit}) = 1/2$

Probabilities are always conditional on something, for example prior knowledge, but often this is left implicit when it is irrelevant or assumed to be obvious from the context.

In terms of $P(B)$ and $P(A \text{ and } B)$ we have

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

$P(B)$ gives the probability of an event in the B set. Given that the event is in B , $P(A|B)$ is the probability of also being in A . It is the fraction of the B outcomes that are also in A :



Rules of probability

The rules of probability generalize the rules of logic in a consistent way. You can check the rules are consistent with normal logic when $P(A)=1$ or 0 (true or false).

1. Complement Rule

Denote “all events that are not A ” as A^c . Since either A or not A must happen, $P(A) + P(A^c) = 1$. Hence

$$P(\text{Event happens}) = 1 - P(\text{Event doesn't happen})$$

or

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

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

E.g. when throwing a fair die, $P(\text{not } 6) = 1 - 1/6 = 5/6$.

2. Multiplication Rule

We can re-arrange the definition of the conditional probability

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

to obtain equivalent expressions for $P(A \text{ and } B)$:

$$P(A \cap B) = \begin{cases} P(A|B)P(B) \\ P(B|A)P(A) \end{cases}$$

You can often think of $P(A \text{ and } B)$ as being the probability of first getting A with probability $P(A)$, and then getting B with probability $P(B|A)$. This is the same as first getting B with probability $P(B)$ and then getting A with probability $P(A|B)$.

E.g. Drawing two random cards from a pack without replacement, the probability of getting two aces is

$$P(\text{first is an ace and second is an ace}) = P(\text{first is an ace}) \times P(\text{second is an ace} | \text{first is an ace})$$

$$= \frac{4}{52} \times \frac{3}{51} = \frac{1}{13} \times \frac{1}{17} = \frac{1}{221}$$

Special Multiplication Rule

If two events A and B are *independent* then $P(A|B) = P(A)$ and $P(B|A) = P(B)$: knowing that A has occurred does not affect the probability that B has occurred and vice versa. In that case

$$P(A \text{ and } B) = P(A \cap B) = P(A) P(B)$$

Probabilities for any number of independent events can be multiplied to get the joint probability. For example if you toss a fair coin twice, the outcome of the first throw shouldn't affect the outcome of the second throw, so the throws are independent.

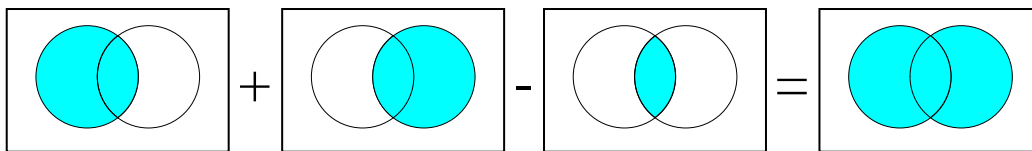
E.g. A fair coin is tossed twice, the chance of getting a head and then a tail is
 $P(H_1 \text{ and } T_2) = P(H_1)P(T_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

E.g. A die is thrown 3 times. The probability of getting the first six on the last throw is
 $P(\text{not } 6)P(\text{not } 6)P(6) = \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6} = \frac{25}{216} = 0.116..$

3. Addition Rule

For any two events A and B :

$$\begin{aligned} P(A \text{ or } B) &= P(A \cup B) \\ &= P(A) + P(B) - P(A \text{ and } B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$



Note: “ A or B ” = $A \cup B$ includes the possibility that both A and B occur.

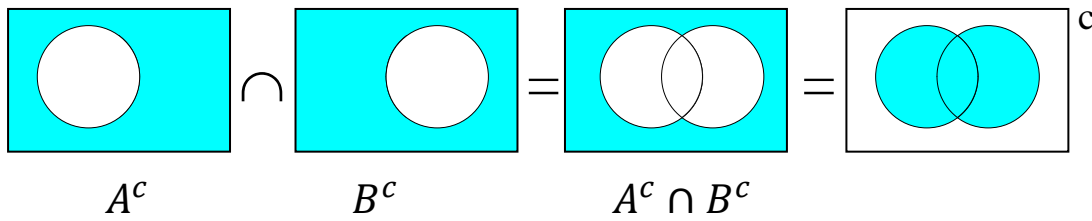
E.g. Throwing a fair die, let events be

A = get an odd number
 B = get a 5 or 6

$$P(A \text{ or } B) = P(A \cup B) = P(\text{odd}) + P(5 \text{ or } 6) - P(5) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$$

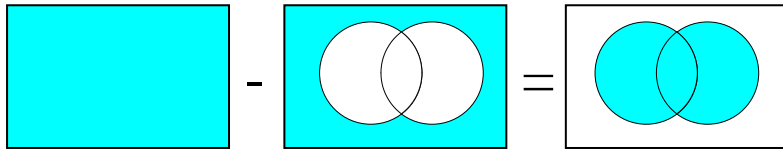
This is consistent, since $P(A \cup B) = P(\{1,3,5,6\}) = \frac{4}{6} = \frac{2}{3}$

Alternative: Note that $A^c \cap B^c = (A \cup B)^c$



So we could also calculate $P(A \cup B)$ using

$$P(A \cup B) = 1 - P(A^c \cap B^c)$$



E.g. As before, throwing a fair die let results of interest be $A = \text{get an odd number}$, $B = \text{get a 5 or 6}$

Then $A^c = \{2,4,6\}$, $B^c = \{1,2,3,4\}$ so $A^c \cap B^c = \{2,4\}$. Hence

$$P(A \text{ or } B) = 1 - P(A^c \cap B^c) = 1 - P(\{2,4\}) = 1 - \frac{1}{3} = \frac{2}{3}$$

This alternative form has the advantage of generalizing easily to lots of possible events:

$$P(A_1 \text{ or } A_2 \text{ or } \dots \text{ or } A_k) = 1 - P(A_1^c \cap A_2^c \cap \dots \cap A_k^c)$$

e.g. if there are three alternative routes A , B , or C to work, the probability of me not being able to get to work is the probability of all three being blocked. So the probability of me being able to get to work is $P(A \text{ clear or } B \text{ clear or } C \text{ clear}) = 1 - P(A \text{ blocked and } B \text{ blocked and } C \text{ blocked})$.

Special Addition Rule

If $P(A \cap B) = 0$, the events are mutually exclusive, so

$$P(A \text{ or } B) = P(A \cap B) = P(A) + P(B)$$

We will often consider mutually exclusive sets of outcomes, in which case the addition rule is very simple to apply:

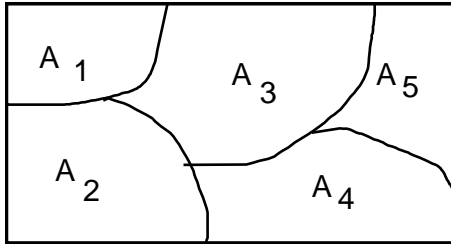
In general if several events A_1, A_2, \dots, A_k are mutually exclusive (i.e. at most one of them can happen in a single experiment) then

$$P(A_1 \text{ or } A_2 \text{ or } \dots \text{ or } A_k) = P(A_1 \cup A_2 \cup \dots \cup A_k) = P(A_1) + P(A_2) + \dots + P(A_k) = \sum_k P(A_k)$$

E.g. Throwing a fair die, $P(\text{getting } 4,5 \text{ or } 6) = P(4)+P(5)+P(6) = 1/6+1/6+1/6=1/2$.

Useful Results

Total Probability



If A_1, A_2, \dots, A_k form a partition (a mutually exclusive list of all possible outcomes) and B is any event then

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_k)P(A_k) = \sum_k P(B|A_k)P(A_k)$$

Proof: This follows since

$$\begin{aligned} P(B) &= P(B | A_1)P(A_1) + P(B | A_2)P(A_2) + \dots + P(B | A_k)P(A_k) \\ &= P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_k) \\ &= P(B \cap A_1 \text{ or } B \cap A_2 \text{ or } \dots \text{ or } B \cap A_k) \\ &= P(B \cap (A_1 \text{ or } A_2 \text{ or } \dots \text{ or } A_k)) \\ &= P(B) \end{aligned}$$

Bayes' Theorem

The multiplication rule gives $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$.
 Bayes' theorem follows by dividing through by $P(B)$ (assuming $P(B) > 0$):

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

This is an incredibly simple, useful and important result. If you have a model that tells you how likely X is given Y , Bayes' theorem allows you to calculate the probability of Y if you observe X . This is the key to learning about your model from statistical data.

Note: often the Total Probability rule is often used to evaluate $P(B)$:

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

Example: evidence in court

The cars in a city are 90% black and 10% grey. A witness to a bank robbery briefly sees the escape car, and says it is grey. Testing the witness under similar conditions shows the witness correctly identifies the colour 80% of the time (in either direction). What is the probability that the car was actually grey?

Solution: Let G = car is grey, B =car is black, W = Witness says car is grey.
 Bayes' theorem gives:

$$P(G|W) = \frac{P(W|G)P(G)}{P(W)}$$

Use total probability rule to write

$$P(W) = P(W|G)P(G) + P(W|B)P(B) = 0.8 \times 0.1 + 0.2 \times 0.9 = 0.26$$

Hence :

$$P(G|W) = \frac{0.8 \times 0.1}{0.26} \approx 0.31$$

Even though the witness is quite reliable, the high prior probability that the car is black makes this significantly more likely despite what the witness reported.

Example: coin tosses

An fair coin is tossed 7 times, and comes up heads all 7 times. What is the probability that the 8th toss is tails?

You meet a man in a bar who offers to bet on the outcome of a coin toss being heads. Being suspicious you think there's a 50% chance the coin is totally biased (has two heads!), but 50% that it is an honest bet. The man tosses the coin 7 times and it comes up heads all 7 times. What is the probability that the 8th toss is a tail?

Solution: A fair coin is by definition unbiased, and each toss is independent and with $P(\text{heads})=1/2$. So the 8th toss of a fair coin is still $P(\text{tails})= 1/2$.

Let B = coin is biased, F = coin is fair ($F = B^c$), $7H$ = seeing seven heads in first seven tosses.

Know $P(7H|B) = 1$, $P(7H|F) = \left(\frac{1}{2}\right)^7 = \frac{1}{128}$, $P(B) = P(F) = \frac{1}{2}$ hence

$$P(7H) = P(7H|B)P(B) + P(7H|F)P(F) = 1 \times 0.5 + \frac{1}{128} \times 0.5 = 0.504$$

So

$$P(B|7H) = \frac{P(7H|B)P(B)}{P(7H)} = \frac{1 \times 0.5}{0.504} = 0.992$$

So very likely biased. Let T_8 be getting a tail on the 8th toss. Using the total probability rule:

$$P(T_8|7H) = P(T_8|B, 7H)P(B|7H) + P(T_8|F, 7H)P(F|7H) = 0 \times \frac{1}{2} + \frac{1}{2} \times (1 - 0.992) = 0.004$$

Note that $P(A|B, C)$ means the probability of A given both B and C .

Example: three-card swindle

Suppose there are three cards:

- A *red card* that is red on both sides,
- A *white card* that is white on both sides, and
- A *mixed card* that is red on one side and white on the other.

All the cards are placed into a hat and one is pulled at random and placed on a table. The side facing up is red. What is the probability that the other side is also red?

Solution: Let R=red card, W = white card, M = mixed card. For a random draw $P(R)=P(W)=P(M)=1/3$. Let SR = see a red face. P(SR) is the probability of getting the red card plus 1/2 the probability of the mixed card.

$$P(SR) = P(SR|R)P(R) + P(SR|M)P(M) = 1 \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{3} = \frac{1}{2}$$

The probability we want is P(R|SR) since having the red card is the only way for the other side also to be red. This is

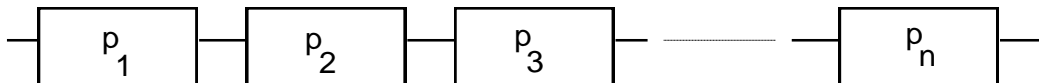
$$P(R|SR) = \frac{P(SR|R)P(R)}{P(SR)} = \frac{1 \times \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Intuition: 2/3 of the three red faces are on the red card.

Reliability of a system

General approach: bottom-up analysis. Need to break down the system into subsystems just containing elements in series or just containing elements in parallel. Find the reliability of each of these subsystems and then repeat the process at the next level up.

Series subsystem: in the diagram p_i = probability that element i fails, so $1 - p_i$ = probability that it does not fail.



The system only works if all n elements work. Failures of different elements are assumed to be independent (so the probability of Element 1 failing does alter after connection to the system).

i.e. P(System does not fail) =

P(Element 1 doesn't fail and Element 2 doesn't fail and ... and Element n doesn't fail)

= P(Element 1 doesn't fail)P(Element 2 doesn't fail) ... P(Element n doesn't fail)
 [Special multiplication rule; independence of failures]

$$= (1-p_1)(1-p_2) \dots (1-p_n) = \prod_{j=1}^n (1-p_j)$$

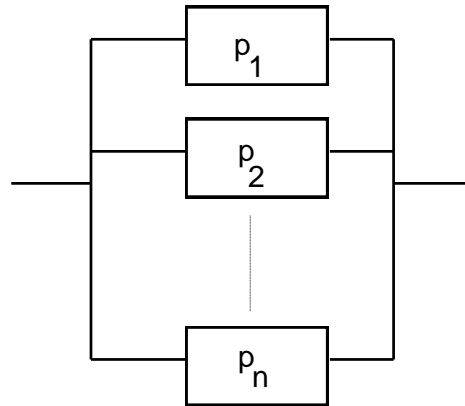
Parallel subsystem: the subsystem only fails if all the elements fail.

i.e. P(System fails) = P(Element 1 fails and Element 2 fails and ... and Element n fails)

= P(Element 1 fails)P(Element 2 fails) ... P(Element n fails)

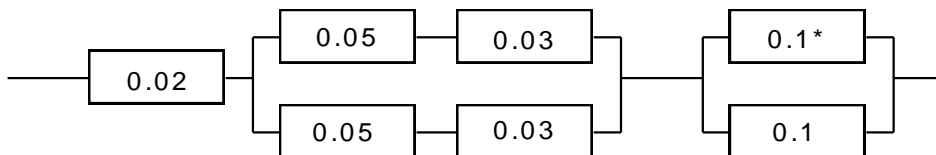
[Independence of failures]

$$= p_1 p_2 \dots p_n = \prod_{j=1}^n p_j$$



Example: reliability of a system

The reliability of a critical system has to be determined. An assessment has already been made of the reliability of components making up the system. The probabilities of failure of the various components in the next year are indicated in the diagram below. It can be assumed that components fail independently of one another.



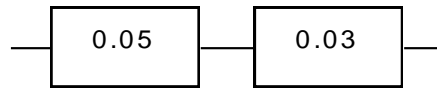
(a) What is the probability that the system does not fail in the next year?

(b) Find the probability that within one year the system does not fail but component * does fail.

(a) What is the probability that the system does not fail in the next year?

Solution

Subsystem 1:

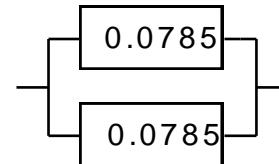


$$P(\text{Subsystem 1 doesn't fail}) = (1 - 0.05)(1 - 0.03) = 0.9215$$

$$P(\text{Subsystem 1 fails}) = 0.0785$$

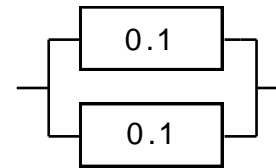
Subsystem 2: (two units of subsystem 1)

$$P(\text{Subsystem 2 fails}) = 0.0785 \times 0.0785 = 0.006162$$



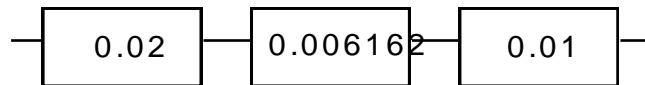
Subsystem 3:

$$P(\text{Subsystem 3 fails}) = 0.1 \times 0.1 = 0.01$$



System (summarised):

$$P(\text{System doesn't fail}) =$$



$$(1 - 0.02)(1 - 0.006162)(1 - 0.01) = 0.964$$

(b) Find P(System does not fail and component * does fail)

Solution

Let B = event that the system does not fail

Let C = event that component * does fail

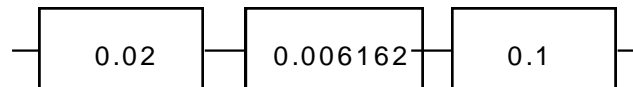
We need to find $P(B \text{ and } C)$.

Now, $P(C) = 0.1$.

Also, $P(B | C) = P(\text{system does not fail given component * has failed})$;

now if component * has failed, Subsystem 3 has probability of failing of 0.1 instead of 0.01, so that the final reliability diagram becomes:

$$\therefore P(B | C) = (1 - 0.02) \times (1 - 6.162 \times 10^{-3})(1 - 0.1) = 0.8766$$



$$\therefore P(B \text{ and } C) = P(B | C) P(C) = 0.8766 \times 0.1 = 0.08766$$

Combinatorics

Permutations - ways of ordering k items: $k!$

Factorials: for a positive integer k , $k! = k(k-1)(k-2) \dots 2 \cdot 1$

e.g. $3! = 3 \times 2 \times 1 = 6$.

By definition, $0! = 1$.

The first item can be chosen in k ways, the second in $k-1$ ways, the third, in $k-2$ ways, etc., giving $k!$ possible orders.

e.g. *ABC* can be arranged as *ABC, ACB, BAC, BCA, CAB* and *CBA*, a total of $3! = 6$ ways.

Ways of choosing k things from n , irrespective of ordering:

Binomial coefficient: for integers n and k where $n \geq k \geq 0$:

$$C_k^n = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Sometimes this is also called “ n choose k ”. Other notations include ${}_n C_k$ and variants.

Justification: Choosing k things from n there are n ways to choose the first item, $n-1$ ways to choose the second.... and $(n-k+1)$ ways to choose the last, so

$$n(n-1)(n-2) \dots (n-k+1) = \frac{n!}{(n-k)!}$$

ways. This is the number of different orderings of k things drawn from n . But there $k!$ orderings of k things, so only $1/k!$ of these is a distinct set, giving the C_k^n distinct sets.

E.g. *There are $3!/(2! \times 1!) = 3$ ways to choose 2 letters from 3 letters *ABC*: *AB, BC* and *AC*.*

E.g. *in the National Lottery, the numbers of ways of choosing 6 numbers from 49 (1, 2, ..., 49) is:*

$$C_6^{49} = \frac{49!}{6!43!} = \frac{49 \times 48 \times 47 \times 46 \times 45 \times 44}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 13,983,816$$

So the probability of winning with a given random ticket is about $1/(14 \text{ million})$.

E.g. *Tossing a fair coin 10 times, the probability of getting exactly 5 heads (in any order) is*

$$\frac{1}{2^5} \frac{1}{2^5} C_5^{10} = \frac{1}{1024} \frac{10 \times 9 \times 8 \times 7 \times 6}{5 \times 4 \times 3 \times 2 \times 1} = \frac{63}{256} = 0.246$$

Calculating factorials and C_k^n

Many calculators have a factorial button, but they become very large very quickly:

$15! = 1,307,674,368,000 \approx 1.3 \times 10^{12}$, so be careful they do not overflow.

Some calculators have a button for calculating C_k^n or you can calculate it directly using factorials or more manually using

$$\begin{aligned} \frac{n!}{k!(n-k)!} &= \frac{n(n-1)(n-2)\dots(n-k+1)(n-k)(n-k-1)\dots 1}{k!(n-k)(n-k-1)\dots 1} \\ &= \frac{n(n-1)\dots(n-k+1)}{k(k-1)(k-2)\dots 1} \end{aligned}$$

Beware that it can also become very large for large n and k , for example there are $100891344545564193334812497256 \approx 10^{29}$ ways to choose 50 items from 100.

For computer users: In MatLab the function is called “nchoosek”, in other systems like Maple and Mathematica it is called “binomial”.

Postscript

I assume most people are familiar with a standard pack of cards. For those that are not, the standard pack is 52 different cards. Each card has a ‘number’ (one of 13 possibilities Ace, 2, 3...10, Jack, Queen, King) and a suit (Hearts, Diamonds, Spades, Clubs). Hearts and Diamonds are red, Spades and Clubs are black. The pack consists of all the possible combinations, so there are four Aces, one in each suit, four 3s, one in each suit, etc, a total of $13 \times 4 = 52$ cards. If a card is picked at random it has probability $\frac{1}{4}$ of being in any given suite, and a probability $\frac{1}{13}$ of being any given number.