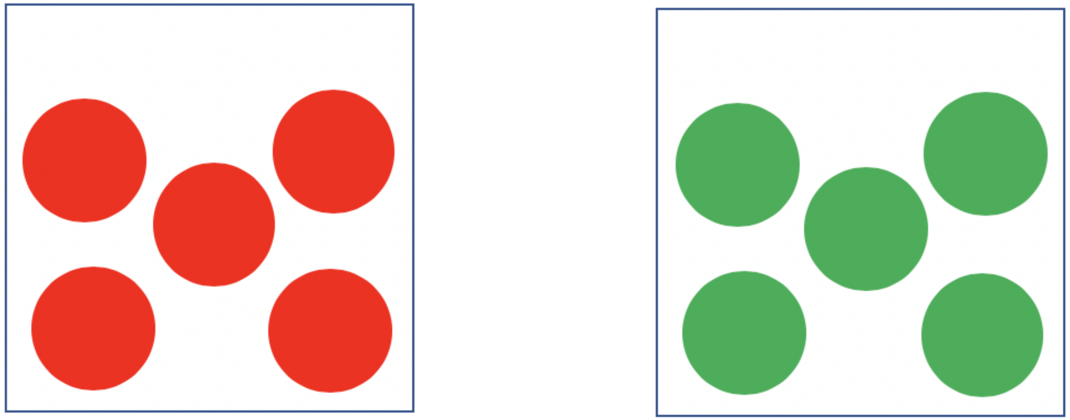
Conditional Probability (Ghahramani 3.1)

If A and H are two events, and it is known that event H has occurred, what effect does this information have on the probability of occurrence of A?



Example

Toss two fair dice. The sample space is

$$\Omega = \{(1,1), \cdots, (6,6)\}.$$

It contains 36 outcomes.

Let A be the event that the sum of the two dice is 8. Then

$$A = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}.$$

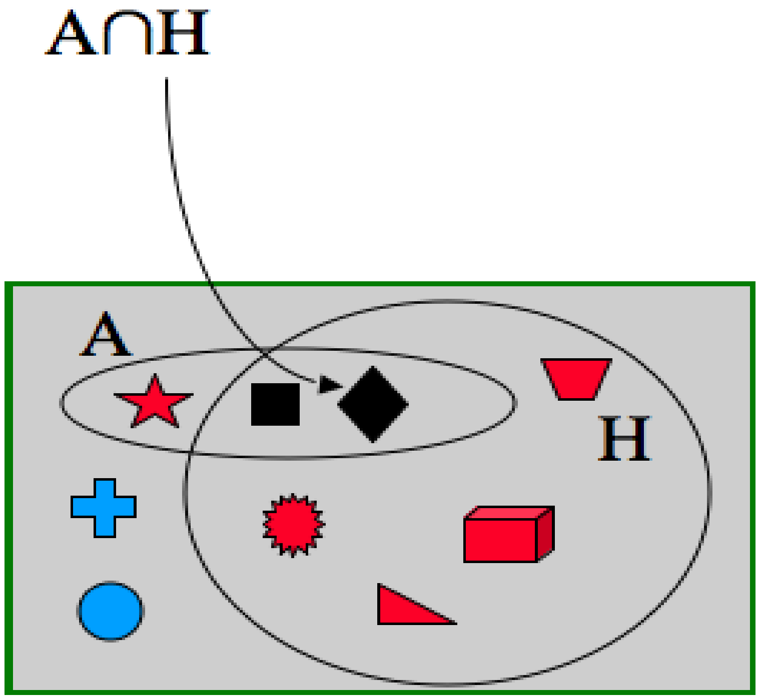
Therefore, $\mathbb{P}(A) = \frac{5}{36}$.

Suppose we know that the first die is a 3. What is the probability of A *given this information*?

Given the first die is a 3 there are six outcomes of interest, $\{(3,1),(3,2),\ldots,(3,6)\}$. Since the dice are fair, each of these outcomes has the same probability of occurring. There is only one outcome $(3,5)$ within these six outcomes producing a sum of 8. Therefore, given that the first die is a 3, the probability of the sum being 8 is $\frac{1}{6}$.

If A denotes "sum of the dice is 8" and H denotes "first die is a 3" the probability we have calculated is called the *conditional probability* of A given H and is denoted $\mathbb{P}(A|H)$.

Illustration in classical probability model



In the case of classical probability model, $\mathbb{P}(A|H)$ is the proportion of outcomes in A among all outcomes in H:

$$\mathbb{P}(A|H) = 2/6.$$

Definition of Conditional Probability

In general, imagine we conduct the experiment n times, and observe H happening for n_1 times and $A \cap H$ happening for n_2 times. The proportion of times A occurs among the n_1 experiments when H occurs is

$$\frac{n_2}{n_1} = \frac{n_2/n}{n_1/n} \to \frac{\mathbb{P}(A \cap H)}{\mathbb{P}(H)} \text{ as } n \to \infty.$$

Hence the probability of A given H should be defined as the probability of $A \cap H$ relative to the probability of H:

$$\mathbb{P}(A|H) = \frac{\mathbb{P}(A \cap H)}{\mathbb{P}(H)} \quad \text{if} \quad \mathbb{P}(H) > 0.$$

Multiplication Theorem (Ghahramani 3.2)

Sometimes we know $\mathbb{P}(H)$ and $\mathbb{P}(A|H)$ but not $\mathbb{P}(A \cap H)$. If this is the case we can use the definition of conditional probability to express the probability of $A \cap H$, that is

$$\mathbb{P}(A \cap H) = \mathbb{P}(H)\mathbb{P}(A|H).$$

Examples

- Toss a fair die. Let $A = \{2\}$ and $H = \{x : x \text{ is even}\}$, then

$$\mathbb{P}(A) = 1/6, \;\; \mathbb{P}(A|H) = 1/3.$$

- Toss two fair dice. Let $A = \{(i, j) : |i - j| = 1\}$ and $H = \{(i, j) : i + j = 7\}$, then

$$\mathbb{P}(A) = 5/18, \;\; \mathbb{P}(A|H) = 1/3.$$

In both examples, we have $\mathbb{P}(A|H) > \mathbb{P}(A)$. That is, it is more likely that A will occur if we know that H has occurred than if we know nothing about H.

We say that there exists a *positive relationship* between A and H if $\mathbb{P}(A|H) > \mathbb{P}(A)$ and a *negative relationship* between A and H if $\mathbb{P}(A|H) < \mathbb{P}(A)$.

If there is a positive relationship between A and H, then the occurrence of H will increase the chance of A occuring. If there is a negative relationship between A and H, then the occurrence of H will decrease the chance of A occuring.

Example

You are one of seven applicants for three jobs.

Consider the random experiment that occurs when the decision is made to offer the jobs to three applicants. The outcome space Ω consists of all combinations of three appointees from seven applicants. There are $\binom{7}{3} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = 35$ of these.

Another one of the applicants (Ms X) and you are the only applicants with a particular skill. You think that it is likely that the employer will want the skill and employ one of you, but unlikely that both of you will be employed. Specifically you make the assessments listed on the following slide about the likelihood of the various possible combinations.

- the five combinations where both you and Ms X get the job have equal probability $1/60$,
- the ten combinations where you get the job and Ms X doesn't have equal probability $1/24$,
- the ten combinations where Ms X gets the job and you don't have equal probability $1/24$, and
- the ten combinations where neither of you get the job have equal probability $1/120$.

You find out on the grapevine that Ms X has not got the job. How does your assessment of the probability that you will get the job change?

Solution

Let the A be the event "you are selected" and the event H "Ms X is not selected".

More rigorously A is the subset of Ω that consists of those outcomes in which you are selected and H is the subset of Ω that consists of those outcomes in which Ms X is not selected.

Then

$$\mathbb{P}(A) = \frac{5}{60} + \frac{10}{24} = \frac{1}{2},$$

$$\mathbb{P}(H) = \frac{10}{120} + \frac{10}{24} = \frac{1}{2},$$

and

$$\mathbb{P}(A \cap H) = \frac{10}{24}.$$

Therefore

$$\mathbb{P}(A|H) = \frac{10/24}{1/2} = 5/6,$$

and so if you know that Ms X did not get the job, your probability of getting the job is quite high.

There is a positive relationship between the events A and H.

Independence of Events (Ghahramani 3.5)

Question: What about the situation when $\mathbb{P}(A|H) = \mathbb{P}(A)$?

- Knowing whether H occurs does not influence the probability of occurrence for A.

We observe that

$$\mathbb{P}(A|H) = \mathbb{P}(A) \iff \frac{\mathbb{P}(A \cap H)}{\mathbb{P}(H)} = \mathbb{P}(A) \iff \frac{\mathbb{P}(A \cap H)}{\mathbb{P}(A)} = \mathbb{P}(H)$$
$$\iff \mathbb{P}(H|A) = \mathbb{P}(H).$$

- It is equivalently true that knowing whether A occurs does not influence the probability of occurrence for H.

As we have just seen $\mathbb{P}(A|B) = \mathbb{P}(A)$ and $\mathbb{P}(B|A) = \mathbb{P}(B)$ are algebraically equivalent to

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

This equation is taken as the mathematical definition of the *independence* of two events. It is a special case of the general multiplication theorem

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

Two events that are not independent are said to be *dependent*.

Example

Consider tossing a fair six-sided dice once and define events $A = \{2, 4, 6\}$, $B = \{1, 2, 3\}$ and $C = \{1, 2, 3, 4\}$, then A and C are independent while A and B are not independent.

Independence vs Disjointness

Independence of events A and B is a different concept from A and B being *disjoint*. You can test for disjointness simply by inspecting the outcomes in A and B to see if there are any in common, even before any probability function is defined (say by inspecting the Venn diagram).

But you cannot test independence without knowing the probabilities.

Unless one or both have probability zero, disjoint events A and B *cannot* be independent, since

$$\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0 < \mathbb{P}(A)\mathbb{P}(B).$$

In fact, the events A and B are negatively related: the occurence of A excludes the occurence of B.

Exercise

If A and B are independent events then so are

- A^c and B
- A and B^c
- A^c and B^c

Independence of $n > 2$ events (Ghahramani 3.5)

Now let's think about extending the idea of independence to more than two events. We talk of the "mutual" independence of $n > 2$ events. But can we cover all dependencies by checking for pairwise independence of all possible pairs? Consider the random experiment of tossing two fair coins and the following three events:

- A: First coin is H
- B: Second coin is H
- C: Exactly one coin is H

Events $A_1, A_2, \ldots, A_n$ are said to be *mutually independent* if for any subcollection $\{j_1, j_2, \ldots, j_m\} \subseteq \{1, 2, \ldots, n\}$

$$\mathbb{P}(A_{j_1} \cap A_{j_2} \cap \ldots \cap A_{j_m}) = \mathbb{P}(A_{j_1})\mathbb{P}(A_{j_2}) \ldots \mathbb{P}(A_{j_m})$$

- How many equations?

If the events $A_1, A_2, \ldots, A_n$ are mutually independent then the following derived collections of events are also mutually independent

- $A_1^c, A_2, \ldots, A_n$
- $A_1^c, A_2^c, A_3, \ldots, A_n$
- $A_1 \cap A_2, A_3, \ldots, A_n$
- $A_1 \cup A_2, A_3, \ldots, A_n$

An important and frequently-applied consequence of mutual independence:

If the events $A_1, A_2, \ldots, A_n$ are mutually independent then

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2)\ldots\mathbb{P}(A_n).$$

NB: The converse is not true. Why?

Network reliability

We can calculate the reliability of a network of mutually independent components connected in series and parallel.

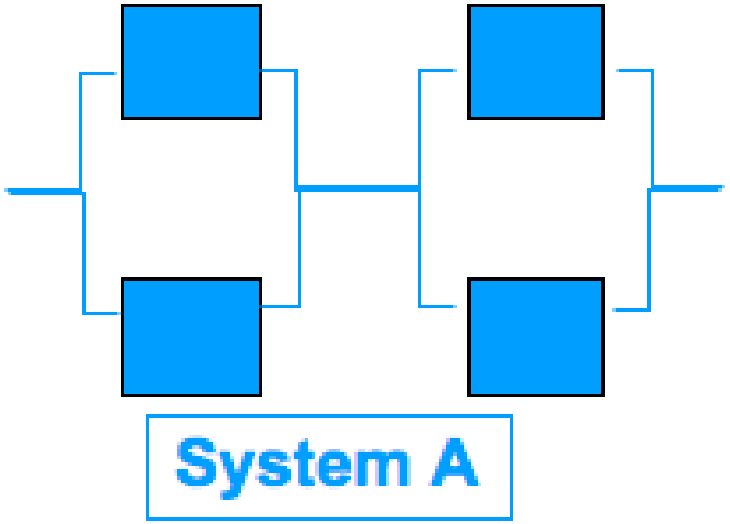
Each component has a particular probability of failure. If a component fails, current cannot flow through it.

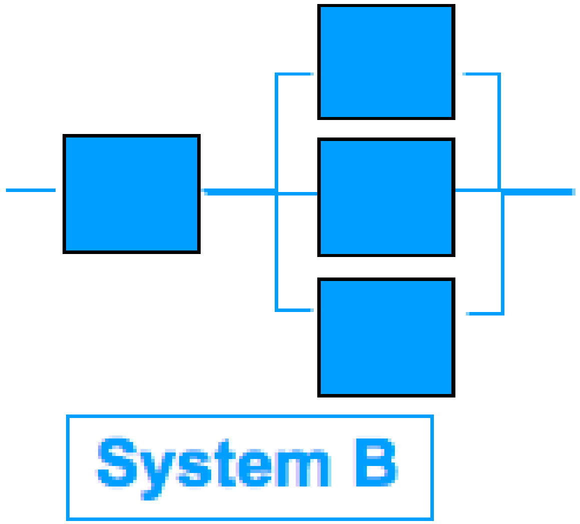
A system/network functions normally if the current can flow from the beginning to the end.

Reliability of the system is the probability that the system functions normally.

Example

Two systems in which components operate independently, each with a probability of failure equal to 0.01. Which system is more reliable?





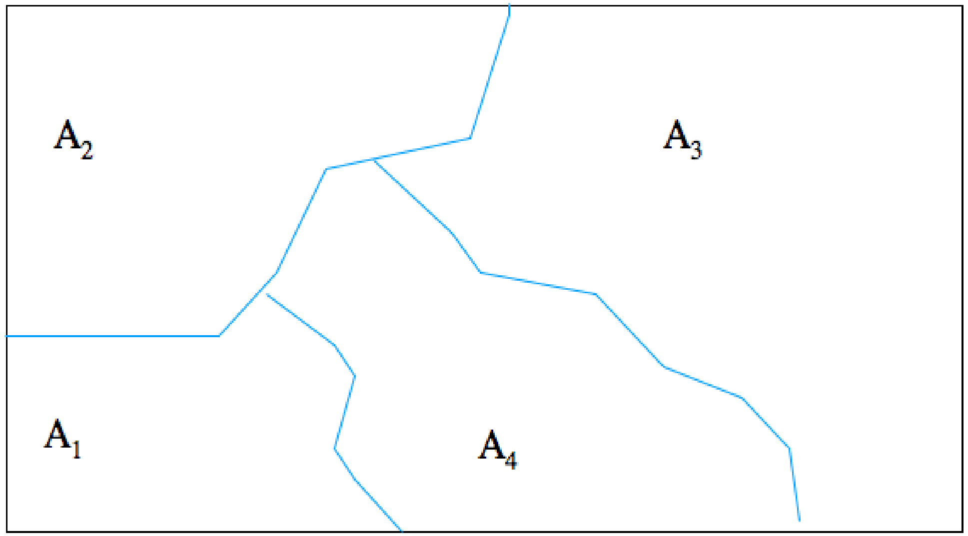
Law of Total Probability (Ghahramani 3.3)

General idea:

- Think of an event as the effect/result due to one of several distinct causes/reasons.
- Compute the probability of this event by conditioning on each of the causes and adding over all these possible causes.

A *partition* of the outcome space Ω is a collection of disjoint and exhaustive events $(A_1, A_2, \ldots)$. That is, for all i and j, $A_i \cap A_j = \emptyset$ and

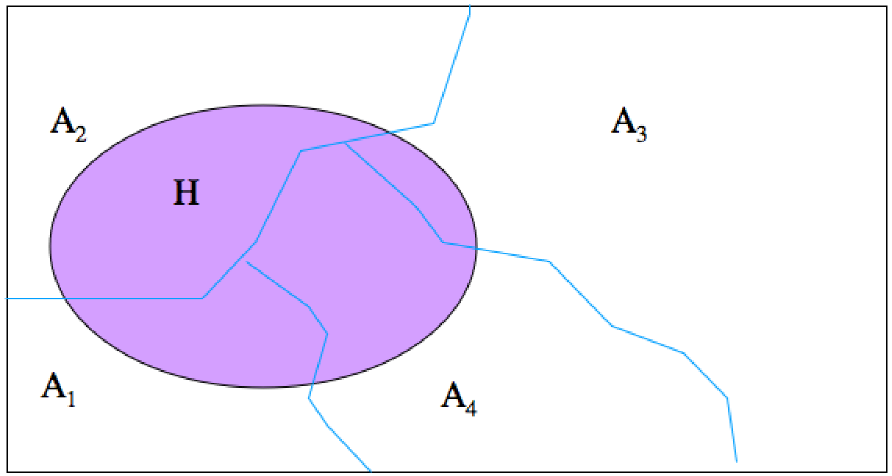
$$\bigcup_i A_i = \Omega$$



Now, for any event H,

$$\begin{aligned} H &= H \cap \Omega \\ &= H \cap \left(\bigcup_i A_i \right) \\ &= \bigcup_i (H \cap A_i) \end{aligned}$$

where the last equation follows from the distributive law.



Using probability axiom 3 and the multiplication formula, we have

$$\begin{aligned}\mathbb{P}(H) &= \sum_i \mathbb{P}\left(H \cap A_i\right) \\ &= \sum_i \mathbb{P}(H|A_i)\mathbb{P}(A_i).\end{aligned}$$

This brings us to the *Law of Total Probability.*

If $A_1, A_2, \ldots$ are disjoint and exhaustive events then, for any event H,

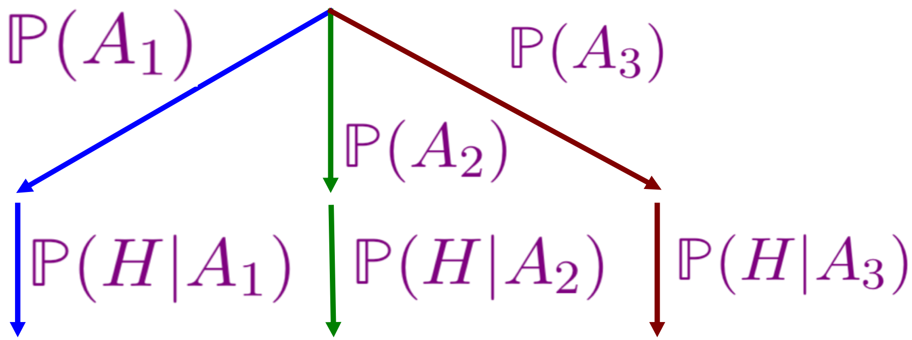
$$\mathbb{P}(H) = \sum_i \mathbb{P}(H|A_i)\mathbb{P}(A_i).$$

A Cause-Effect Interpretation of Law of Total Prob.

We can interpret the event H as the effect/result of some process, and interpret the partition $\{A_i\}$ as all possible different causes of the effect.

The law of total probability can be interpreted as reaching the effect through any of the possible causes.

Multiply along the branches and add the results to get $\mathbb{P}(H)$:



Example

Suppose a test for HIV is 90% effective in the sense that if a person is HIV positive then the test has a 90% chance of saying that they are HIV positive. If they are not positive there is still a 5% chance that the test says that they are. Assume that in a large population, there are 0.01% of people who are HIV positive.

What is the probability that a randomly selected person from the population will have a positive HIV test result?

Solution

Define:

$$\begin{aligned} A_1 &: \text{ the person is HIV positive} \\ A_2 &: \text{ the person is HIV negative} \\ H &: \text{ HIV test for the person is positive.} \end{aligned}$$

By the assumption,

$$\mathbb{P}(A_1) = 0.0001, \quad \mathbb{P}(A_2) = 0.9999$$
$$\mathbb{P}(H|A_1) = 0.9, \quad \mathbb{P}(H|A_2) = 0.05.$$

Therefore,

$$\mathbb{P}(H) = \mathbb{P}(A_1)\mathbb{P}(H|A_1) + \mathbb{P}(A_2)\mathbb{P}(H|A_2) \approx 0.05$$

A different question: If a person receives a positive HIV test result, what is the probability that the person is actually HIV positive?

Bayes' Formula (Ghahramani 3.4)

Let $A_1, A_2, \ldots$ be a set of disjoint and exhaustive events.

Then for an event H,

$$\mathbb{P}(A_i|H) = \frac{\mathbb{P}(H|A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(H|A_j)\mathbb{P}(A_j)}.$$

The Bayes formula is interpreted as:

- given the effect, find a particular cause.

Proof

From the Multiplication Formula

$$\mathbb{P}(A_i|H) = \frac{\mathbb{P}(A_i \cap H)}{\mathbb{P}(H)} = \frac{\mathbb{P}(H|A_i)\mathbb{P}(A_i)}{\mathbb{P}(H)}.$$

Substitution of

$$\mathbb{P}(H) = \sum_j \mathbb{P}(H|A_j)\mathbb{P}(A_j)$$

from the Law of Total Probability gives the result.

HIV example solution:

HIV example (conclusion)

- Under these numbers a person is unlikely to be HIV positive even if the test says that they are. Such phenomena are well known in the epidemiological literature. Tests for rare diseases have to be very accurate.

Example (Multiple Choice Exams)

Consider a multiple choice exam that has m choices of answer for each question. Assume that the probability that a student knows the correct answer to a question is p. A student that doesn't know the correct answer marks an answer at random. Suppose that the answer marked to a particular question was correct. What is the probability that the student was guessing?

Solution

Here we have the disjoint and exhaustive events

- A_1 – the student knew the correct answer
- A_2 – the student was guessing

and the observed event H – the correct answer was marked.

The conditional probabilities are

$$\mathbb{P}(H|A_1) = 1, \quad \mathbb{P}(H|A_2) = 1/m.$$

We want to find $\mathbb{P}(A_2|H)$:

$$\begin{aligned}
\mathbb{P}(A_2|H) &= \frac{\mathbb{P}(H|A_2)\mathbb{P}(A_2)}{\mathbb{P}(H|A_1)\mathbb{P}(A_1) + \mathbb{P}(H|A_2)\mathbb{P}(A_2)} \\
&= \frac{1/m\,(1-p)}{p + 1/m\,(1-p)} \\
&= \frac{1-p}{mp+1-p}\,.
\end{aligned}$$