Probability and Statistics

Kristel Van Steen, PhD[2]

Montefiore Institute - Systems and Modeling

GIGA - Bioinformatics

ULg

kristel.vansteen@ulg.ac.be

CHAPTER 1: PROBABILITY THEORY

1 Introduction

2 Different flavors of probability

2.1 Classical or a priori probability

2.2 A posteriori or frequency probability

3 Axiomatic probability theory

3.1 Set theory

3.2 Sample space and event

3.3. Redefining probability: an axiomatic definition

3.4 Modeling experiments using finite sample spaces

3.5 Independence and conditional probability

4 Useful results

4.1 Total probability

4.2 Bayes' theorem

5 Reliability of a system

6 Supplementary section

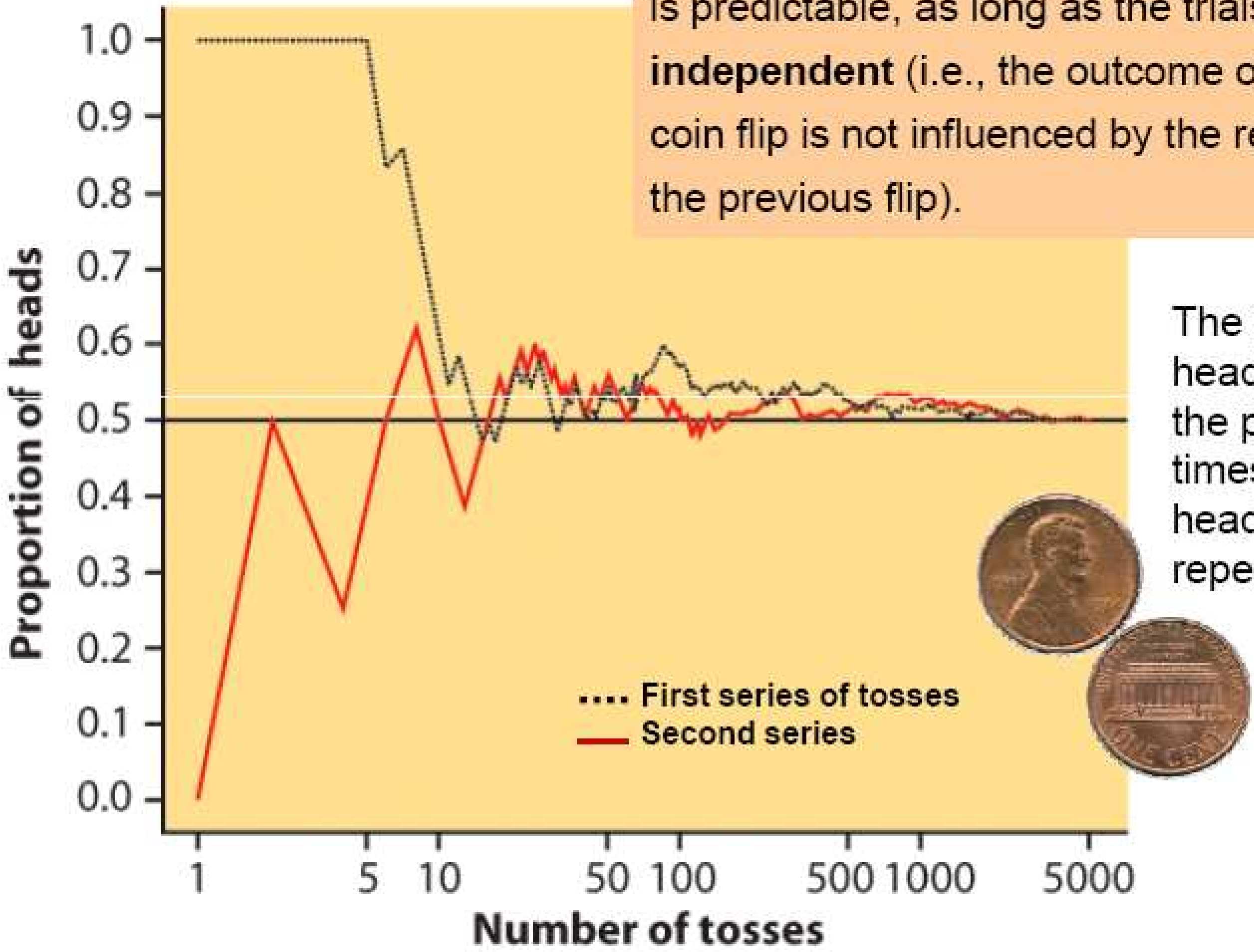
Combinatorics

1 Introduction

- One of the fundamental tools of statistics is *probability*, which had its formal beginnings with games of chance in the 17th century.
- Even though the outcome of a particular trial (like tossing a coin or spinning a roulette wheel) may be uncertain, there is a *predictable* long-term outcome

Coin toss

The result of any single coin toss is random. But the result over many tosses is predictable, as long as the trials are **independent** (i.e., the outcome of a new coin flip is not influenced by the result of the previous flip).

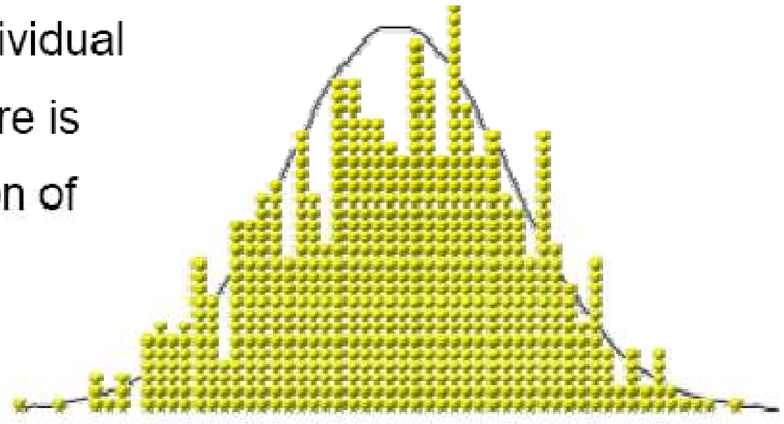


The probability of heads is 0.5 = the proportion of times you get heads in many repeated trials.

The gambling industry relies on probability distributions to calculate the odds of winning. The rewards are then fixed precisely so that, **on average**, players lose and the house wins.

The industry is very tough on so called “cheaters” because their probability to win exceeds that of the house. Remember that it is a business, and therefore it has to be profitable.

A phenomenon is **random** if individual outcomes are uncertain, but there is nonetheless a regular distribution of outcomes in a large number of repetitions.



The **probability** of any outcome of a random phenomenon can be defined as the proportion of times the outcome would occur in a very long series of repetitions.

- Similar type of uncertainty and long-term regularity often occurs in experimental science

2 Different flavors of probability

2.1 Classical or a priori probability

- The classical definition of probability is prompted by the close association between the theory of probability of the early ages and games of chance.

  Classical probability: If a random experiment can result in n mutually exclusive and equally likely outcomes and if n_A of these outcomes have an attribute A, then the probability of A is the fraction n_A/ n.

- In this context

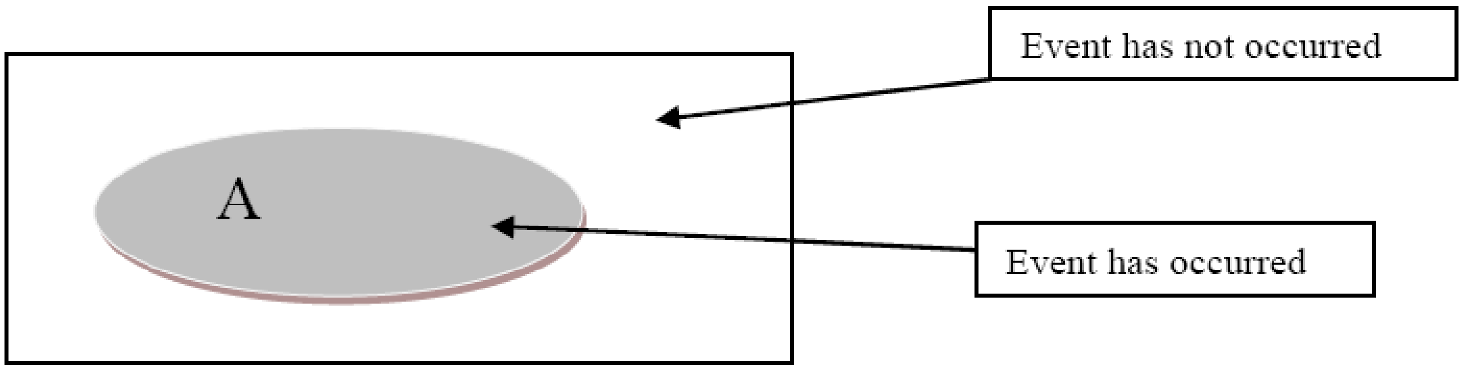
  Event: a possible outcome or set of possible outcomes of an experiment or observation. Typically denoted by a capital letter (e.g., A = result of coin toss)

- Also

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(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(*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 models

Probability models describe, mathematically, the outcome of random processes. They consist of two parts:

1) ***S* = Sample Space:** This is a set, or list, of all possible outcomes of a random process. An **event** is a subset of the sample space.

2) A **probability** for each possible event in the sample space *S*.

Example: Probability Model for a Coin Toss:

S = {Head, Tail}

Probability of heads = 0.5

Probability of tails = 0.5

- The probability of 0.5 only holds when the *trials* are independent

Two events are **independent** if the probability that one event occurs on any given trial of an experiment is not affected or changed by the occurrence of the other event.

When are trials not independent?

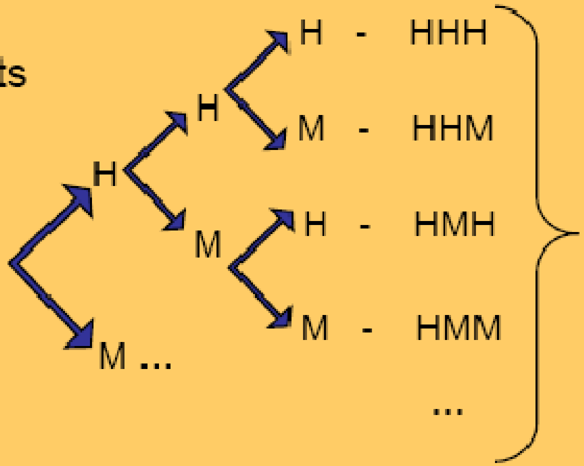
Imagine that these coins were spread out so that half were heads up and half were tails up. Close your eyes and pick one. The probability of it being heads is 0.5. However, if you don't put it back in the pile, the probability of picking up another coin that is heads up is now less than 0.5.

The trials are independent only when you put the coin back each time. It is called **sampling with replacement.**

Sample spaces

It's the question that determines the sample space.

A. A basketball player shoots three free throws. What are the possible sequences of hits (H) and misses (M)?



S = { HHH, HHM, HMH, HMM, MHH, MHM, MMH, MMM }

Note: 8 elements, 2^3

B. A basketball player shoots three free throws. What is the number of baskets made?

S = { 0, 1, 2, 3 }

C. A nutrition researcher feeds a new diet to a young male white rat. What are the possible outcomes of weight gain (in grams)?

$S = [0, \infty[$ = (all numbers ≥ 0)

Rules of Probability

Coin Toss Example:
S = {Head, Tail}
Probability of heads = 0.5
Probability of tails = 0.5

1) Probabilities range from 0 (*no chance of the event*) to 1 (*the event has to happen*). For any event A, $0 \leq P(A) \leq 1$	Probability of getting a Head = 0.5 We write this as: *P*(Head) = 0.5 *P*(neither Head nor Tail) = 0 *P*(getting either a Head or a Tail) = 1

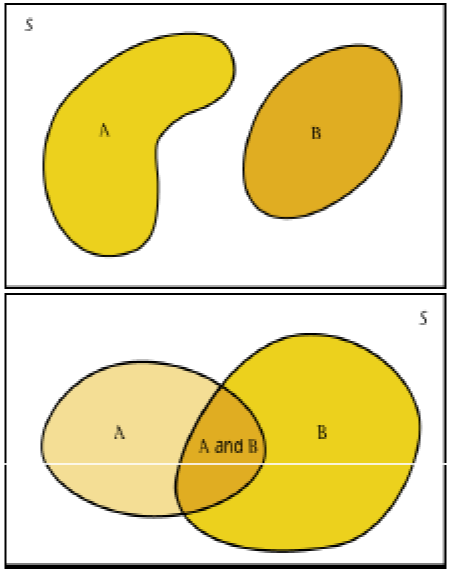
2) Because some outcome must occur on every trial, the sum of the probabilities for all possible outcomes (the sample space) must be exactly 1. *P*(sample space) = 1	Coin toss: S = {Head, Tail} *P*(head) + *P*(tail) = 0.5 + 0.5 =1 → *P*(sample space) = 1

3) Two events A and B are **disjoint** if they have no outcomes in common and can never happen together**.** The probability that A <u>or</u> B occurs is then the sum of their individual probabilities.

P(A or B) = "*P*(A U B)" = *P*(A) + *P*(B)

This is the **addition rule for disjoint events.**

Venn diagrams:
A and B disjoint



A and B <u>not</u> disjoint

Example: If you flip two coins, and the first flip does not affect the second flip:

S = {HH, HT, TH, TT}. The probability of each of these events is 1/4, or 0.25.

The probability that you obtain "only heads <u>or</u> only tails" is:

P(HH <u>or</u> TT) = *P*(HH) + *P*(TT) = 0.25 + 0.25 = 0.50

Coin Toss Example:
S = {Head, Tail}
Probability of heads = 0.5
Probability of tails = 0.5

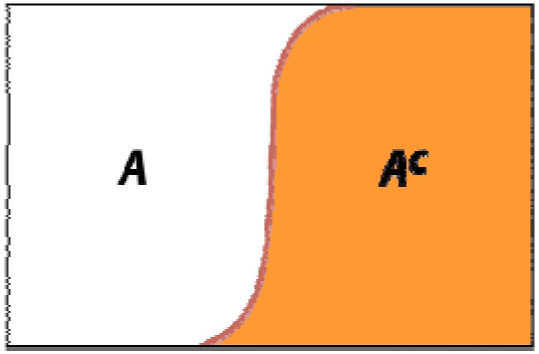
4) The complement of any event A is the event that A does not occur, written as A^c.

The complement rule states that the probability of an event not occurring is 1 minus the probability that is does occur.

$P(\text{not A}) = P(A^c) = 1 - P(A)$

Tail^c = not Tail = Head

$P(\text{Tail}^c) = 1 - P(\text{Head}) = 0.5$



Venn diagram:
Sample space made up of an event A and its complementary A^c, i.e., everything that is not A.

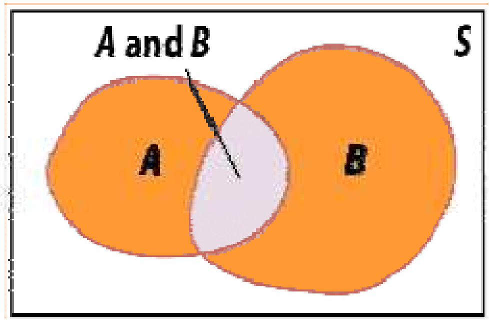
Coin Toss Example:
S = {Head, Tail}
Probability of heads = 0.5
Probability of tails = 0.5

5) Two events A and B are independent if knowing that one occurs does not change the probability that the other occurs.

If A and B are independent, P(A and B) = P(A)P(B)

This is the **multiplication rule for independent events.**

Two consecutive coin tosses:
P(first Tail and second Tail) = P(first Tail) * P(second Tail) = 0.5 * 0.5 = 0.25



Venn diagram:
Event A and event B. The intersection represents the event {A and B} and outcomes common to both A and B.

- For instance, a couple wants 3 children. What is the arrangement of boys (B) and girls (G)?

Genetics tell us that the probability that a baby is a boy or a girl is the same, 0.5.

Sample space: {BBB, BBG, BGB, GBB, GGB, GBG, BGG, GGG}

➔ All eight outcomes in the sample space are **equally likely**.
The probability of each is thus 1/8.

➔ Each birth is independent of the next, so we can use the **multiplication rule**.

Example: $P(BBB) = P(B)^* P(B)^* P(B) = (1/2)^*(1/2)^*(1/2) = 1/8$

- A couple wants 3 children. What are the number of girls (G) they could have?

The same genetic laws apply. We can use the probabilities above and the **addition rule for disjoint events** to calculate the probabilities for *X*.

Sample space: {0, 1, 2, 3}

➔ $P(X = 0) = P(BBB) = 1/8$

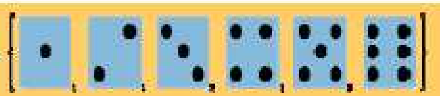
➔ $P(X = 1) = P(BBG \text{ or } BGB \text{ or } GBB) = P(BBG) + P(BGB) + P(GBB) = 3/8$

…

Value of X	0	1	2	3
Probability	1/8	3/8	3/8	1/8

Probabilities: finite number of outcomes

Finite sample spaces deal with **discrete data** — data that can only take on a limited number of values. These values are often integers or whole numbers.

Throwing a die: $S =$ 

$S = \{1, 2, 3, 4, 5, 6\}$

The individual outcomes of a random phenomenon are always disjoint.

➔ The probability of any event is the sum of the probabilities of the outcomes making up the event (addition rule).

M&M candies

If you draw an M&M candy at random from a bag, the candy will have one of six colors. The probability of drawing each color depends on the proportions manufactured, as described here:

Color	Brown	Red	Yellow	Green	Orange	Blue
Probability	0.3	0.2	0.2	0.1	0.1	?

What is the probability that an M&M chosen at random is blue?

S = {brown, red, yellow, green, orange, blue}

$P(S) = P(\text{brown}) + P(\text{red}) + P(\text{yellow}) + P(\text{green}) + P(\text{orange}) + P(\text{blue}) = 1$

$P(\text{blue}) = 1 - [P(\text{brown}) + P(\text{red}) + P(\text{yellow}) + P(\text{green}) + P(\text{orange})]$

$= 1 - [0.3 + 0.2 + 0.2 + 0.1 + 0.1] = 0.1$

What is the probability that a random M&M is either red, yellow, or orange?

$P(\text{red or yellow or orange}) = P(\text{red}) + P(\text{yellow}) + P(\text{orange})$

$= 0.2 + 0.2 + 0.1 = 0.5$

Probabilities: equally likely outcomes

We can assign probabilities either:

- **empirically** → from our knowledge of numerous similar past events
  - Mendel discovered the probabilities of inheritance of a given trait from experiments on peas without knowing about genes or DNA.

- **or theoretically** → from our understanding of the phenomenon and symmetries in the problem
  - A 6-sided fair die: each side has the same chance of turning up
  - Genetic laws of inheritance based on meiosis process

If a random phenomenon has *k* equally likely possible outcomes, then each individual outcome has probability 1/*k*.

And, for any event A:

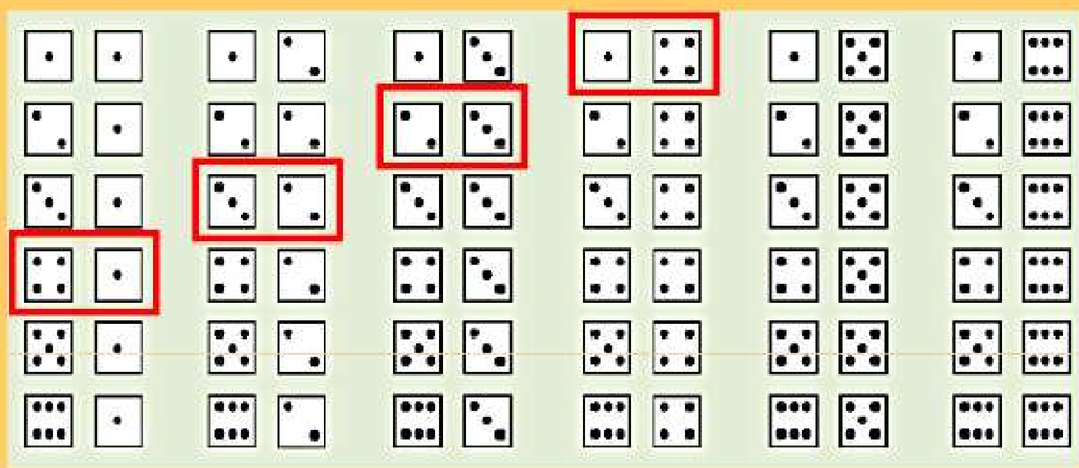
$$P(\mathrm{A}) = \frac{\text{count of outcomes in A}}{\text{count of outcomes in S}}$$

Dice

You toss two dice. What is the probability of the outcomes summing to 5?

This <u>is</u> *S*:

{(1,1), (1,2), (1,3), ……etc.}



There are 36 possible outcomes in *S*, all equally likely (given fair dice).

Thus, the probability of any one of them is 1/36.

P(the roll of two dice sums to 5) =

P(1,4) + *P*(2,3) + *P*(3,2) + *P*(4,1) = 4 / 36 = 0.111

A priori probabilities

- The probabilities determined by the classical definition are called a priori probabilities.
- Results can be derived by pure deductive reasoning. Nothing is said about how one can determine whether or not a particular coin is true
- The fact that we shall deal with ideal objects in developing a theory of probability will not trouble us because that is a common requirement of mathematical systems
    - E.g., geometry deals with conceptually perfect circles, lines with zero width, and so forth, but it is a useful branch of knowledge, which can be applied to diverse practical problems

2.2 A posteriori or frequency probability

Limitations of the classical definition

- Limitation 1: The definition of probability must be modified somehow when the total number of possible outcomes is infinite

  - What is the probability that an integer drawn at random from the positive integers be even? Start with the first 2*N* integers…

    - Natural ordering: 1,2,3,4,5,6,…
    - Ordering 1,3,2; 5,7,4; 9,11,6;…
    - The natural numbers can be ordered that the ratio will oscillate and never approach any definite value as *N* increases.

- Limitation 2: Suppose that we toss a coin known to be biased in favor of heads (it is bent so that a head is more likely to appear than a tail).

    - What is the probability of a head?
    - The classical definition leaves us completely helpless…

A posteriori probabilities

We assume that a series of observations (or experiments) can be made under quite uniform conditions:

- An observation of a random experiment is made
- Then the experiment is repeated under similar conditions, and another observation is taken
- This is repeated many times, and while conditions are similar each time, there is an uncontrollable variation which is haphazard or random so that the observations are individually unpredictable.

- In many cases the observations will fall into certain classes wherein the relative frequencies are quite stable.
- This suggests that we postulate a number *p*, called the probability of the event, and approximate *p* by the relative frequency with which the repeated observations satisfy the event

Applet Probabilities

3 Axiomatic probability theory

3.1 Set theory

Introduction

We begin with a wide collection of objects:

- Each object in our collection is called a *point* or *element* ω
- The collection is large enough so that it includes all the points under consideration
- The totality of these points is called the *space, universe, universal set* Ω
- We will call it the **space**, anticipating that it will become the sampling space when we speak of probabilities. Note that $\omega \in \Omega$.
- A set is a collection of objects; for the sequel we assume that sets consist of points in the space Ω

Definitions

Definition **Subset** If every element of a set A is also an element of a set B, then A is defined to be a *subset* of B, and we shall write $A \subset B$ or $B \supset A$; read "A is contained in B" or "B contains A." ////

Definition **Equivalent sets** Two sets A and B are defined to be *equivalent*, or *equal*, if $A \subset B$ and $B \subset A$. This will be indicated by writing $A = B$. ////

Definition **Empty set** If a set A contains no points, it will be called the *null set*, or *empty set*, and denoted by ϕ. ////

Definition **Complement** The *complement* of a set A with respect to the space Ω, denoted by $\bar{A}$, A^c, or $\Omega - A$, is the set of all points that are in Ω but not in A. ////

Definition **Union** Let A and B be any two subsets of Ω; then the set that consists of all points that are in A or B or both is defined to be the *union* of A and B and written $A \cup B$. ////

Definition **Intersection** Let A and B be any two subsets of Ω; then the set that consists of all points that are in both A and B is defined to be the *intersection* of A and B and is written $A \cap B$ or AB. ////

Definition **Set difference** Let A and B be any two subsets of Ω. The set of all points in A that are not in B will be denoted by $A - B$ and is defined as *set difference*. ////

EXAMPLE Let $\Omega = \{(x, y): 0 \le x \le 1 \text{ and } 0 \le y \le 1\}$, which is read the collection of all points (x, y) for which $0 \le x \le 1$ and $0 \le y \le 1$. Define the following sets:

$$A_1 = \{(x, y): 0 \le x \le 1; 0 \le y \le \tfrac{1}{2}\},$$

$$A_2 = \{(x, y): 0 \le x \le \tfrac{1}{2}; 0 \le y \le 1\},$$

$$A_3 = \{(x, y): 0 \le x \le y \le 1\},$$

$$A_4 = \{(x, y): 0 \le x \le \tfrac{1}{2}; 0 \le y \le \tfrac{1}{2}\}.$$

(We shall adhere to the practice initiated here of using braces to embrace the points of a set.)
The set relations below follow.

$$A_4 \subset A_1; \qquad A_4 \subset A_2; \qquad A_1 \cap A_2 = A_1A_2 = A_4;$$

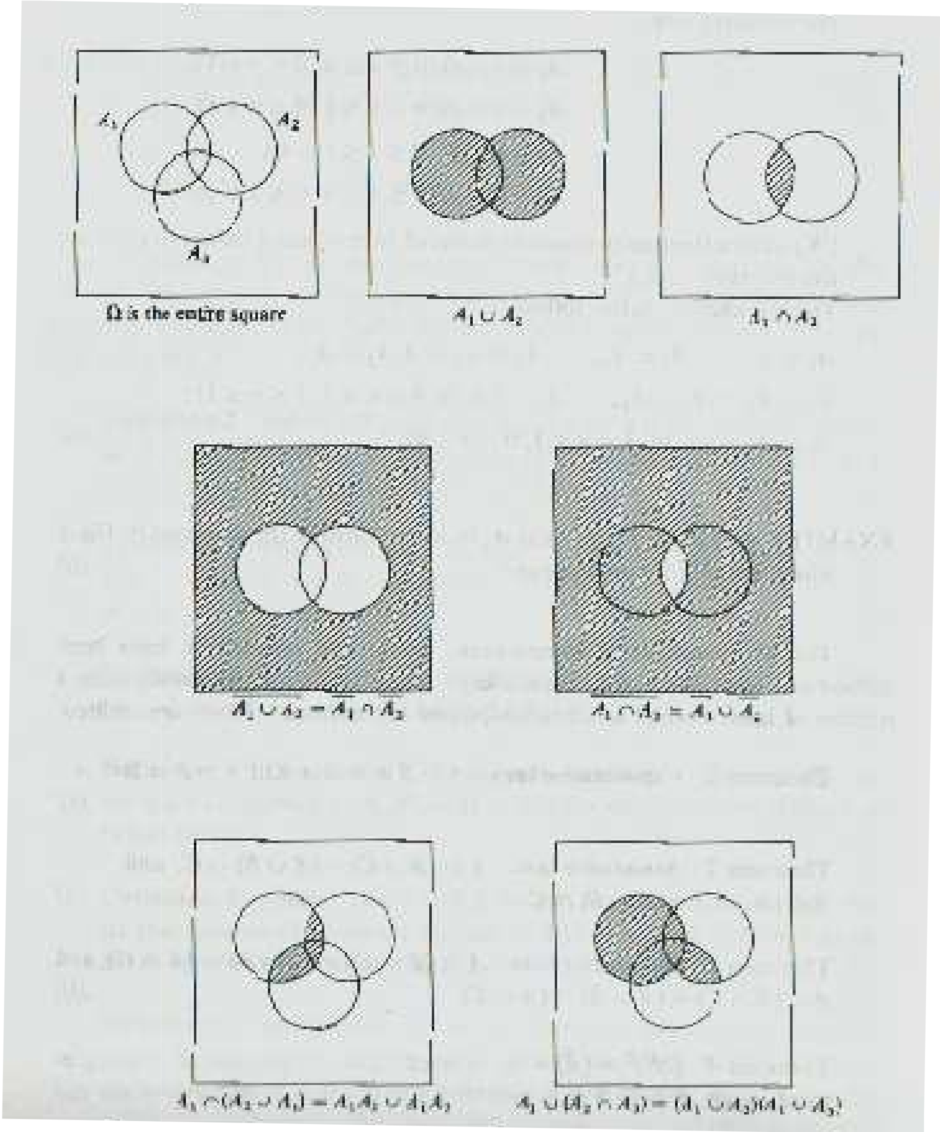
$$A_2 \cup A_3 = A_4 \cup A_3; \qquad \bar{A}_1 = \{(x, y): 0 \le x \le 1; \tfrac{1}{2} < y \le 1\};$$

$$A_1 - A_4 = \{(x, y): \tfrac{1}{2} < x \le 1; 0 \le y \le \tfrac{1}{2}\}.$$

////

Definition **Union and intersection of sets** Let Λ be an index set and $\{A_\lambda : \lambda \in \Lambda\} = \{A_\lambda\}$, a collection of subsets of Ω indexed by Λ. The set of points that consists of all points that belong to A_λ for at least one λ is called the *union* of the sets $\{A_\lambda\}$ and is denoted by $\bigcup_{\lambda \in \Lambda} A_\lambda$. The set of points that consists of all points that belong to A_λ for every λ is called the *intersection* of the sets $\{A_\lambda\}$ and is denoted by $\bigcap_{\lambda \in \Lambda} A_\lambda$. If Λ is empty, then define $\bigcup_{\lambda \in \Lambda} A_\lambda = \phi$ and $\bigcap_{\lambda \in \Lambda} A_\lambda = \Omega$. ////

Definition **Disjoint or mutually exclusive** Subsets A and B of Ω are defined to be *mutually exclusive* or *disjoint* if $A \cap B = \phi$. Subsets $A_1, A_2, \ldots$ are defined to be *mutually exclusive* if $A_i A_j = \phi$ for every $i \neq j$. ////


Ω is the entire square
A1 ∪ A2
A1 ∩ A2

Probability laws

Theorem 1 **Commutative laws** $A \cup B = B \cup A$ and $A \cap B = B \cap A$. ////

Theorem 2 **Associative laws** $A \cup (B \cup C) = (A \cup B) \cup C$, and $A \cap (B \cap C) = (A \cap B) \cap C$. ////

Theorem 3 **Distributive laws** $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$. ////

Theorem 4 $(A^C)^C = \overline{(\overline{A})} = A$; in words, the complement of A complement equals A. ////

Theorem 5 $A\Omega = A$; $A \cup \Omega = \Omega$; $A\phi = \phi$; and $A \cup \phi = A$. ////

Theorem 6 $A\overline{A} = \phi$; $A \cup \overline{A} = \Omega$; $A \cap A = A$; and $A \cup A = A$. ////

Theorem 7 $\overline{(A \cup B)} = \bar{A} \cap \bar{B}$, and $\overline{(A \cap B)} = \bar{A} \cup \bar{B}$. These are known as *De Morgan's laws.* ////

Although we will feel free to use any of the above laws, it might be instructive to give a proof of one of them just to illustrate the technique. For example, let us show that $\overline{(A \cup B)} = \bar{A} \cap \bar{B}$. By definition, two sets are equal if each is contained in the other. We first show that $\overline{(A \cup B)} \subset \bar{A} \cap \bar{B}$ by proving that if $\omega \in \overline{A \cup B}$, then $\omega \in \bar{A} \cap \bar{B}$. Now $\omega \in \overline{(A \cup B)}$ implies $\omega \notin A \cup B$, which implies that $\omega \notin A$ and $\omega \notin B$, which in turn implies that $\omega \in \bar{A}$ and $\omega \in \bar{B}$; that is, $\omega \in \bar{A} \cap \bar{B}$. We next show that $\bar{A} \cap \bar{B} \subset \overline{(A \cup B)}$. Let $\omega \in \bar{A} \cap \bar{B}$, which means ω belongs to both $\bar{A}$ and $\bar{B}$. Then $\omega \notin A \cup B$ for if it did, ω must belong to at least one of A or B, contradicting that ω belongs to both $\bar{A}$ and $\bar{B}$; however, $\omega \notin A \cup B$ means $\omega \in \overline{(A \cup B)}$, completing the proof.

Theorem 8 $A - B = A\bar{B}$. ////

Theorem 9 De Morgan's theorem Let Λ be an index set and $\{A_\lambda\}$ a collection of subsets of Ω indexed by Λ. Then,

(i) $\overline{\bigcup_{\lambda \in \Lambda} A_\lambda} = \bigcap_{\lambda \in \Lambda} \bar{A}_\lambda$.

(ii) $\overline{\bigcap_{\lambda \in \Lambda} A_\lambda} = \bigcup_{\lambda \in \Lambda} \bar{A}_\lambda$. ////

Theorem 10 If A and B are subsets of Ω, then (i) $A = AB \cup A\bar{B}$, and (ii) $AB \cap A\bar{B} = \phi$.

PROOF (i) $A = A \cap \Omega = A \cap (B \cup \bar{B}) = AB \cup A\bar{B}$. (ii) $AB \cap A\bar{B} = A\,AB\bar{B} = A\phi = \phi$. ////

Theorem 11 If $A \subset B$, then $AB = A$, and $A \cup B = B$.

PROOF Left as an exercise. ////

Rules of probability revisited

using set representations

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(Event happens) = 1 - P(Event doesn't happen)

or

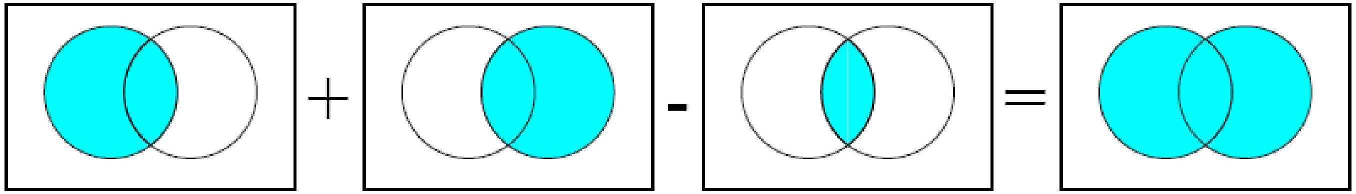
$$P(A) = 1 - P(A^c)$$
$$P(A^c) = 1 - P(A)$$

E.g. when throwing a fair die, P(not 6) = 1 – 1/6 = 5/6.

2. Addition Rule

For any two events A and B:

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



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

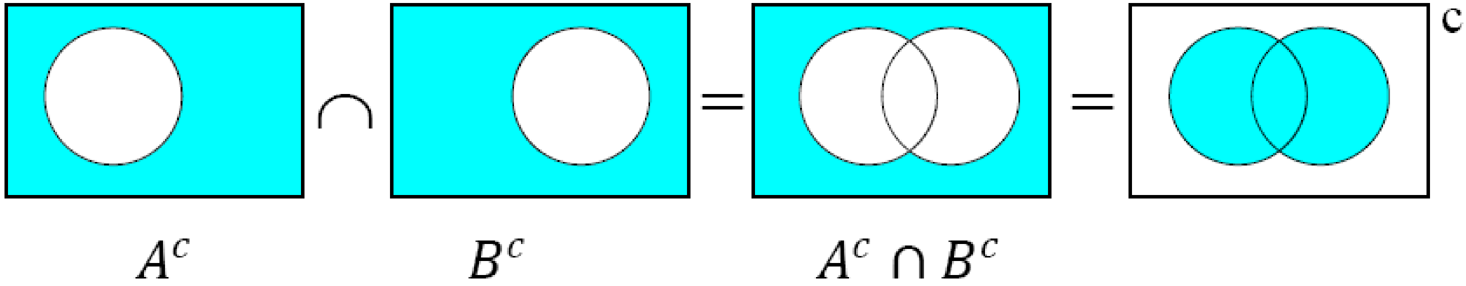
E.g. Throwing a fair die, let events be

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

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

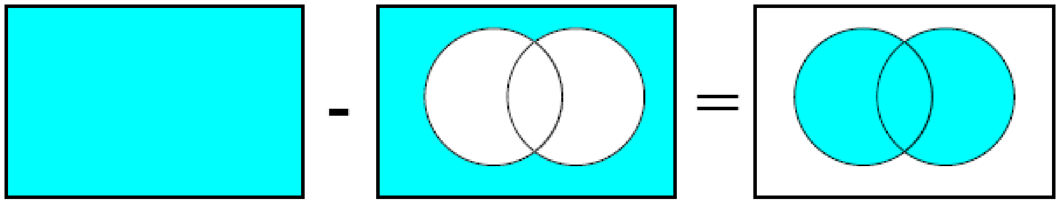
This is consistent, since $P(A \cup B) = P(\{1,3,5,6\}) = \frac{4}{6} = \frac{2}{3}$

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 = get an odd number, B = 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:

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

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 $A_1, A_2, \ldots, A_k$ are mutually exclusive (i.e. at most one of them can happen in a single experiment) then

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

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

3.2 Sample space and event

- When talking about probability models, we have in mind a conceptual experiment, whose possible outcomes we would like to study by assessing the probability of certain outcomes or collection of outcomes.
- Two important concepts to assess these probabilities:

    - **Sample space**: The sample space denoted by Ω is the collection or totality of all possible outcomes of a conceptual experiment

    - **Event and event space**: An event is a subset of the sample space. The class of all events associated with a given experiment is defined to be the event space (usually denoted by a script Latin letter, such as $\mathscr{A}$)

- Note:
  - An event is always a subset of the sample space, but for sufficiently large sample spaces not all subsets will be events
  - The class of all subsets of the sample space will not necessarily correspond to the event space
  - If the sample space consists of only a finite number of points, then the corresponding event space will be the class of all subsets of the sample space

EXAMPLE The experiment is the tossing of a single die (a regular six-sided polyhedron or cube marked on each face with one to six spots) and noting which face is up. Now the die can land with any one of the six faces up; so there are six possible outcomes of the experiment:

$$\Omega = \{ ⚀,\ ⚁,\ ⚂,\ ⚃,\ ⚄,\ ⚅ \}.$$

Let A = {even number of spots up}. A is an event; it is a subset of Ω. $A = \{⚁, ⚃, ⚅\}$. Let A_i = {i spots up}; $i = 1, 2, \ldots, 6$. Each A_i is an elementary event. For this experiment the sample space is finite; hence the event space is all subsets of Ω. There are $2^6 = 64$ events, of which only 6 are elementary, in $\mathscr{A}$ (including both ϕ and Ω).

////

EXAMPLE Select a light bulb, and record the time in hours that it burns before burning out. Any nonnegative number is a conceivable outcome of this experiment; so $\Omega = \{x: x \geq 0\}$. For this sample space not all

subsets of Ω are events; however, any subset that can be exhibited will be an event. For example, let

$$A = \{\text{bulb burns for at least } k \text{ hours but burns out before } m \text{ hours}\}$$
$$= \{x: k \leq x < m\};$$

then A is an event for any $0 \leq k < m$. ////

- Clearly, our definitions of event and event space are not entirely satisfactory.
- We said that if the event space is sufficiently large (whatever this means), not all subsets of the sample space are events. But which subsets would be seen as event and which not remains to be resolved
- Rather than developing the necessary mathematics to precisely define which subsets of Ω constitute our event space, we will state some properties of it that seem reasonable to require:

(i) $\Omega \in \mathscr{A}$.

(ii) If $A \in \mathscr{A}$, then $\bar{A} \in \mathscr{A}$.

(iii) If A_1 and $A_2 \in \mathscr{A}$, then $A_1 \cup A_2 \in \mathscr{A}$.

- Any collection of events with properties (i) to (iii) is called a Boolean algebra, or just algebra, of events. Note: collection of all subsets of Ω necessarily satisfies the above properties.
- Several results follow (see next slides)
- Assuming that we impose the event space to be an algebra, and redefining probability in the next section, will allow us to explain why an event space cannot always be taken to be the collection of all subsets of Ω.

Theorem $\phi \in \mathscr{A}$.

PROOF By property (i) $\Omega \in \mathscr{A}$; by (ii) $\overline{\Omega} \in \mathscr{A}$; but $\overline{\Omega} = \phi$; so $\phi \in \mathscr{A}$. ////

Theorem If A_1 and $A_2 \in \mathscr{A}$, then $A_1 \cap A_2 \in \mathscr{A}$.

PROOF $\overline{A}_1$ and $\overline{A}_2 \in \mathscr{A}$; hence $\overline{A}_1 \cup \overline{A}_2$, and $\overline{(\overline{A}_1 \cup \overline{A}_2)} \in \mathscr{A}$, but $\overline{(\overline{A}_1 \cup \overline{A}_2)} = \overline{\overline{A}}_1 \cap \overline{\overline{A}}_2 = A_1 \cap A_2$ by De Morgan's law. ////

Theorem If $A_1, A_2, \ldots, A_n \in \mathscr{A}$, then $\bigcup_{i=1}^{n} A_i$ and $\bigcap_{i=1}^{n} A_i \in \mathscr{A}$.

PROOF Follows by induction. ////

3.3. Redefining probability: an axiomatic definition

The general concept of a function

The definition of a function The following terminology is frequently used to describe a function: A *function*, say $f(\cdot)$, is a rule (law, formula, recipe) that associates each point in one set of points with one and only one point in another set of points. The first collection of points, say A, is called the *domain*, and the second collection, say B, the *counterdomain*.

Definition **Function** A *function*, say $f(\cdot)$, with domain A and counterdomain B, is a collection of ordered pairs, say (a, b), satisfying (i) $a \in A$ and $b \in B$; (ii) each $a \in A$ occurs as the first element of some ordered pair in the collection (each $b \in B$ is not necessarily the second element of some ordered pair); and (iii) no two (distinct) ordered pairs in the collection have the same first element. ////

If $(a, b) \in f(\cdot)$, we write $b = f(a)$ (read "b equals f of a") and call $f(a)$ the *value* of $f(\cdot)$ at a. For any $a \in A$, $f(a)$ is an element of B; whereas $f(\cdot)$ is a set of ordered pairs. The set of all values of $f(\cdot)$ is called the *range* of $f(\cdot)$; i.e., the range of $f(\cdot) = \{b \in B: b = f(a) \text{ for some } a \in A\}$ and is always a subset of the counterdomain B but is not necessarily equal to it. $f(a)$ is also called the *image* of a under $f(\cdot)$, and a is called the *preimage* of $f(a)$.

EXAMPLE Let $f_1(\cdot)$ and $f_2(\cdot)$ be the two functions, having the real line for their domain and counterdomain, defined by

$$f_1(\cdot) = \{(x, y): y = x^3 + x + 1, -\infty < x < \infty\}$$

and

$$f_2(\cdot) = \{(x, y): y = x^2, -\infty < x < \infty\}.$$

The range of $f_1(\cdot)$ is the counterdomain, the whole real line, but the range of $f_2(\cdot)$ is all nonnegative real numbers, not the same as the counterdomain. ////

Indicator functions

Definition **Indicator function** Let Ω be any space with points ω and A any subset of Ω. The *indicator function* of A, denoted by $I_A(\cdot)$, is the function with domain Ω and counterdomain equal to the set consisting of the two real numbers 0 and 1 defined by

$$I_A(\omega) = \begin{cases} 1 & \text{if} \quad \omega \in A \\ 0 & \text{if} \quad \omega \notin A. \end{cases}$$

$I_A(\cdot)$ clearly "indicates" the set A. ////

Properties of Indicator Functions Let Ω be any space and $\mathscr{A}$ any collection of subsets of Ω:

(i) $I_A(\omega) = 1 - I_{\bar{A}}(\omega)$ for every $A \in \mathscr{A}$.

(ii) $I_{A_1 A_2 \cdots A_n}(\omega) = I_{A_1}(\omega) \cdot I_{A_2}(\omega) \cdots I_{A_n}(\omega)$ for $A_1, \ldots, A_n \in \mathscr{A}$.

(iii) $I_{A_1 \cup A_2 \cup \cdots \cup A_n}(\omega) = \max [I_{A_1}(\omega), I_{A_2}(\omega), \ldots, I_{A_n}(\omega)]$ for $A_1, \ldots, A_n \in \mathscr{A}$.

(iv) $I_A^2(\omega) = I_A(\omega)$ for every $A \in \mathscr{A}$.

e.g.,

$$I_{\{[0,\,1)\}}(x) = I_{[0,\,1)}(x) = \begin{cases} 1 & \text{if} \quad 0 \le x < 1 \\ 0 & \text{otherwise,} \end{cases}$$

and if I^+ denotes the set of positive integers,

$$I_{I^+}(x) = \begin{cases} 1 & \text{if } x \text{ is some positive integer} \\ 0 & \text{otherwise.} \end{cases}$$

Probability functions

Definition Probability function A *probability function* $P[\cdot]$ is a set function with domain $\mathscr{A}$ (an algebra of events)* and counterdomain the interval [0, 1] which satisfies the following axioms:

(i) $P[A] \geq 0$ for every $A \in \mathscr{A}$.

(ii) $P[\Omega] = 1$.

(iii) If $A_1, A_2, \ldots$ is a sequence of mutually exclusive events in $\mathscr{A}$

(that is, $A_i \cap A_j = \phi$ for $i \neq j; i, j = 1, 2, \ldots$) and if $A_1 \cup A_2 \cup \cdots = \bigcup_{i=1}^{\infty} A_i \in \mathscr{A}$, then $P\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} P[A_i]$. ////

*In defining a probability function, many authors assume that the domain of the set function is a sigma-algebra rather than just an algebra. For an algebra $\mathscr{A}$, we had the property

$$\text{if } A_1 \text{ and } A_2 \in \mathscr{A}, \text{ then } A_1 \cup A_2 \in \mathscr{A}.$$

A sigma-algebra differs from an algebra in that the above property is replaced by

$$\text{if } A_1, A_2, \ldots, A_n, \ldots \in \mathscr{A}, \text{ then } \bigcup_{n=1}^{\infty} A_n \in \mathscr{A}.$$

It can be shown that a sigma-algebra is an algebra, but not necessarily conversely. If the domain of the probability function is taken to be a sigma-algebra then axiom (iii) can be simplified to

if $A_1, A_2, \ldots$ is a sequence of mutually exclusive events in $\mathscr{A}$, $P\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} P[A_i]$.

A fundamental theorem of probability theory, called the *extension theorem*, states that if a probability function is defined on an algebra (as we have done) then it can be extended to a sigma-algebra.

- Note:
  - o The axioms for a probability function are clearly motivated by the definitions of classical and frequency probability.
  - o The axiomatic definition is a mathematical one, telling us which set of functions can be called probability functions
  - o However, the axiomatic definition does not tell us what value the probability function P[.] assigns to a given event.
  - o We will have to *model our random experiment* in some way in order to obtain values for the probability of events

Probability space

Definition **Probability space** A *probability space* is the triplet $(\Omega, \mathscr{A}, P[\cdot])$, where Ω is a sample space, $\mathscr{A}$ is a collection (assumed to be an algebra) of events (each a subset of Ω), and $P[\cdot]$ is a probability function with domain $\mathscr{A}$. ////

Properties of $P[\cdot]$ For each of the following theorems, assume that Ω and $\mathscr{A}$ (an algebra of events) are given and $P[\cdot]$ is a probability function having domain $\mathscr{A}$.

Theorem 15 $P[\phi] = 0$.

PROOF Take $A_1 = \phi$, $A_2 = \phi$, $A_3 = \phi$, ...; then by axiom (iii)

$$P[\phi] = P\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} P[A_i] = \sum_{i=1}^{\infty} P[\phi],$$

which can hold only if $P[\phi] = 0$. ////

Theorem 16 If $A_1, \ldots, A_n$ are mutually exclusive events in $\mathscr{A}$, then

$$P[A_1 \cup \cdots \cup A_n] = \sum_{i=1}^{n} P[A_i].$$

Theorem 17 If A is an event in $\mathscr{A}$, then

$$P[\bar{A}] = 1 - P[A].$$

PROOF $A \cup \bar{A} = \Omega$, and $A \cap \bar{A} = \phi$; so

$$P[\Omega] = P[A \cup \bar{A}] = P[A] + P[\bar{A}].$$

But $P[\Omega] = 1$ by axiom (ii); the result follows. ////

Theorem 18 If A and $B \in \mathscr{A}$, then $P[A] = P[AB] + P[A\bar{B}]$, and $P[A - B] = P[A\bar{B}] = P[A] - P[AB]$.

PROOF $A = AB \cup A\bar{B}$, and $AB \cap A\bar{B} = \phi$; so $P[A] = P[AB] + P[A\bar{B}]$. ////

Theorem 19 For every two events A and $B \in \mathscr{A}$, $P[A \cup B] = P[A] + P[B] - P[AB]$. More generally, for events $A_1, A_2, \ldots, A_n \in \mathscr{A}$

$$P[A_1 \cup A_2 \cup \cdots \cup A_n] = \sum_{j=1}^{n} P[A_j] - \sum_{i<j}\sum P[A_i A_j]$$

$$+ \sum_{i<j<k}\sum\sum P[A_i A_j A_k] - \cdots + (-1)^{n+1} P[A_1 A_2 \ldots A_n].$$

PROOF $A \cup B = A \cup \bar{A}B$, and $A \cap \bar{A}B = \phi$; so

$$P[A \cup B] = P[A] + P[\bar{A}B]$$

$$= P[A] + P[B] - P[AB].$$

The more general statement is proved by mathematical induction.

////

Theorem 20 If A and $B \in \mathscr{A}$ and $A \subset B$, then $P[A] \leq P[B]$.

Theorem 21 Boole's inequality If $A_1, A_2, \ldots, A_n \in \mathscr{A}$, then

$$P[A_1 \cup A_2 \cup \cdots \cup A_n] \leq P[A_1] + P[A_2] + \cdots + P[A_n].$$

PROOF $P[A_1 \cup A_2] = P[A_1] + P[A_2] - P[A_1 A_2] \leq P[A_1] + P[A_2]$.
The proof is completed using mathematical induction. ////

3.4 Modeling experiments using finite sample spaces

Finite samples spaces with equally likely points

- For certain random experiments, there is a finite number of outcomes N, and it is often realistic to assume that the probability of each outcome is $1/N$
- The classical definition of probability is generally adequate for these problems
- We show how the axiomatic definition is applicable as well

Let $\omega_1, \omega_2, \ldots, \omega_N$ be the N sample points in a finite space Ω. Suppose that the set function $P[\cdot]$ with domain the collection of all subsets of Ω satisfies the following conditions:

(i) $P[\{\omega_1\}] = P[\{\omega_2\}] = \cdots = P[\{\omega_N\}]$.

(ii) If A is any subset of Ω which contains $N(A)$ sample points [has size $N(A)$], then $P[A] = N(A)/N$.

Then it is readily checked that the set function $P[\cdot]$ satisfies the three axioms and hence is a probability function.

Definition **Equally likely probability function** The probability function $P[\cdot]$ satisfying conditions (i) and (ii) above is defined to be an *equally likely probability function.* ////

- Given that a random experiment can be realistically modeled by assuming equally likely sample points, the only problem left in determining the value of the probability of event A is to find $N(\Omega) = N$ and $N(A)$. Strictly speaking this is just a problem of counting—count the number of points in A and the number of points in Ω.

- If $N(A)$ and $N(\Omega)$ are large for a given random experiment with a finite number of equally likely outcomes, the counting itself can become a difficult problem. Such counting can often be facilitated by use of certain combinatorial formulas.

(see supplementary section to chapter)

Finite samples spaces withOUT equally likely points

For finite sample spaces without equally likely sample points, things are not quite as simple, but we can completely define the values of $P[A]$ for each of the $2^{N(\Omega)}$ events A by specifying the value of $P[\cdot]$ for each of the $N = N(\Omega)$ elementary events. Let $\Omega = \{\omega_1, \ldots, \omega_N\}$, and assume $p_j = P[\{\omega_j\}]$ for $j = 1, \ldots, N$. Since

$$1 = P[\Omega] = P\left[\bigcup_{j=1}^{N} \{\omega_j\}\right] = \sum_{j=1}^{N} P[\{\omega_j\}],$$

$$\sum_{j=1}^{N} p_j = 1.$$

For any event A, define $P[A] = \Sigma p_j$, where the summation is over those ω_j belonging to A. It can be shown that $P[\cdot]$ so defined satisfies the three axioms and hence is a probability function.

EXAMPLE Consider an experiment that has N outcomes, say $\omega_1, \omega_2, \ldots, \omega_N$, where it is known that outcome ω_{j+1} is twice as likely as outcome ω_j, where $j = 1, \ldots, N-1$; that is, $p_{j+1} = 2p_j$, where $p_i = P[\{\omega_i\}]$. Find $P[A_k]$, where $A_k = \{\omega_1, \omega_2, \ldots, \omega_k\}$. Since

$$\sum_{j=1}^{N} p_j = \sum_{j=1}^{N} 2^{j-1} p_1 = p_1(1 + 2 + 2^2 + \cdots + 2^{N-1}) = p_1(2^N - 1) = 1,$$

$$p_1 = \frac{1}{2^N - 1}$$

and

$$p_j = 2^{j-1}/(2^N - 1);$$

hence

$$P[A_k] = \sum_{j=1}^{k} p_j = \sum_{j=1}^{k} 2^{j-1}/(2^N - 1) = \frac{2^k - 1}{2^N - 1}. \qquad ////$$

3.5 Independence and conditional probability

Independence of events If $P[A|B]$ does not depend on event B, that is, $P[A|B] = P[A]$, then it would seem natural to say that event A is independent of event B. This is given in the following definition.

Definition **Independent events** For a given probability space $(\Omega, \mathscr{A}, P[\cdot])$, let A and B be two events in $\mathscr{A}$. Events A and B are defined to be *independent* if and only if any one of the following conditions is satisfied:

(i) $P[AB] = P[A]P[B]$.

(ii) $P[A|B] = P[A]$ if $P[B] > 0$.

(iii) $P[B|A] = P[B]$ if $P[A] > 0$. ////

Remark Some authors use "statistically independent," or "stochastically independent," instead of "independent." ////

Something to think about …

The property of independence of two events A and B and the property that A and B are mutually exclusive are distinct, though related, properties. For example, two mutually exclusive events A and B are independent if and only if $P[A]P[B] = 0$, which is true if and only if either A or B has zero probability. Or if $P[A] \neq 0$ and $P[B] \neq 0$, then A and B independent implies that they are not mutually exclusive, and A and B mutually exclusive implies that they are not independent. Independence of A and B implies independence of other events as well.

Definition : Independence of several events For a given probability space $(\Omega, \mathscr{A}, P[\cdot])$, let $A_1, A_2, \ldots, A_n$ be n events in $\mathscr{A}$. Events $A_1, A_2, \ldots, A_n$ are defined to be *independent* if and only if

$$P[A_i A_j] = P[A_i]P[A_j] \qquad \text{for } i \neq j$$

$$P[A_i A_j A_k] = P[A_i]P[A_j]P[A_k] \qquad \text{for } i \neq j, j \neq k, i \neq k$$

$$\vdots$$

$$P\left[\bigcap_{i=1}^{n} A_i\right] = \prod_{i=1}^{n} P[A_i].$$

////

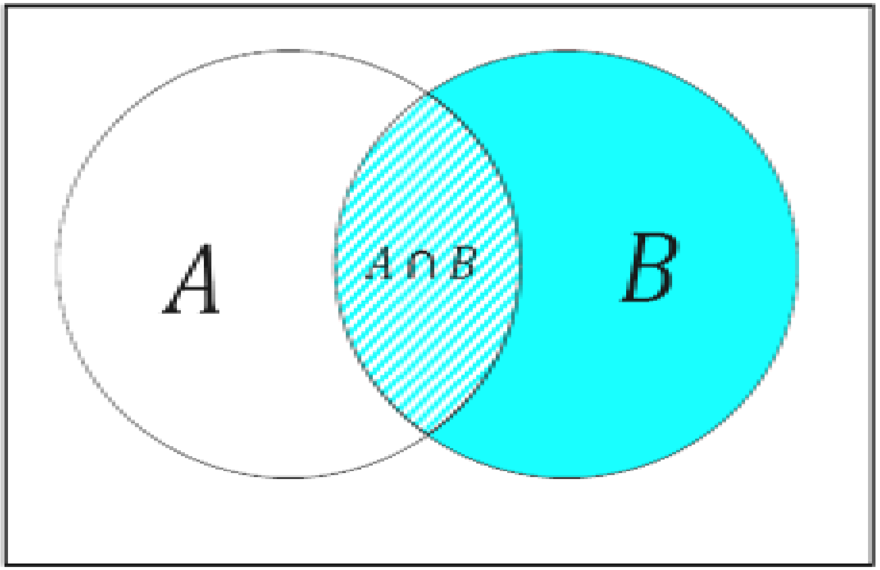
Conditional probability: P(A|B) means the probability of A given that B has happened or is true.
e.g. P(result of coin toss is heads | the coin is fair) =1/2
P(Tomorrow is Tuesday | it is Monday) = 1
P(card is a heart | it is a red suit) = 1/2

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 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:



Multiplication Rule

We can re-arrange the definition of the conditional probability

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

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 hearts is

$$P(\text{first is a heart and second is a heart})$$
$$= P(\text{first is a heart}) \times P(\text{second is a heart} \mid \text{first is a heart})$$

$$= \frac{13}{52} \times \frac{12}{51} = \frac{1}{4} \times \frac{12}{51} = \frac{3}{51}$$

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
$P(H_1 \text{ and } T_2) = P(H_1)P(T_2) = ½ \times ½ = ¼$.

E.g. A die is thrown 3 times. The probability of getting the first six on the last throw is
P(not 6)P(not 6)P(6) = 5/6 x *5/6* x *1/6 = 25/216 = 0.116..*

Properties of $P[\cdot\,|B]$ Assume that the probability space $(\Omega, \mathscr{A}, P[\cdot])$ is given, and let $B \in \mathscr{A}$ satisfy $P[B] > 0$.

Theorem 22 $P[\phi\,|\,B] = 0.$ ////

Theorem 23 If $A_1, \ldots, A_n$ are mutually exclusive events in $\mathscr{A}$, then

$$P[A_1 \cup \cdots \cup A_n | B] = \sum_{i=1}^{n} P[A_i | B]. \qquad ////$$

Theorem 24 If A is an event in $\mathscr{A}$, then

$$P[\bar{A} | B] = 1 - P[A | B]. \qquad ////$$

Theorem 25 If A_1 and $A_2 \in \mathscr{A}$, then

$$P[A_1 | B] = P[A_1 A_2 | B] + P[A_1 \bar{A}_2 | B]. \qquad ////$$

Theorem 26 For every two events A_1 and $A_2 \in \mathscr{A}$,

$$P[A_1 \cup A_2 | B] = P[A_1 | B] + P[A_2 | B] - P[A_1 A_2 | B]. \qquad ////$$

Theorem 27 If A_1 and $A_2 \in \mathscr{A}$ and $A_1 \subset A_2$, then

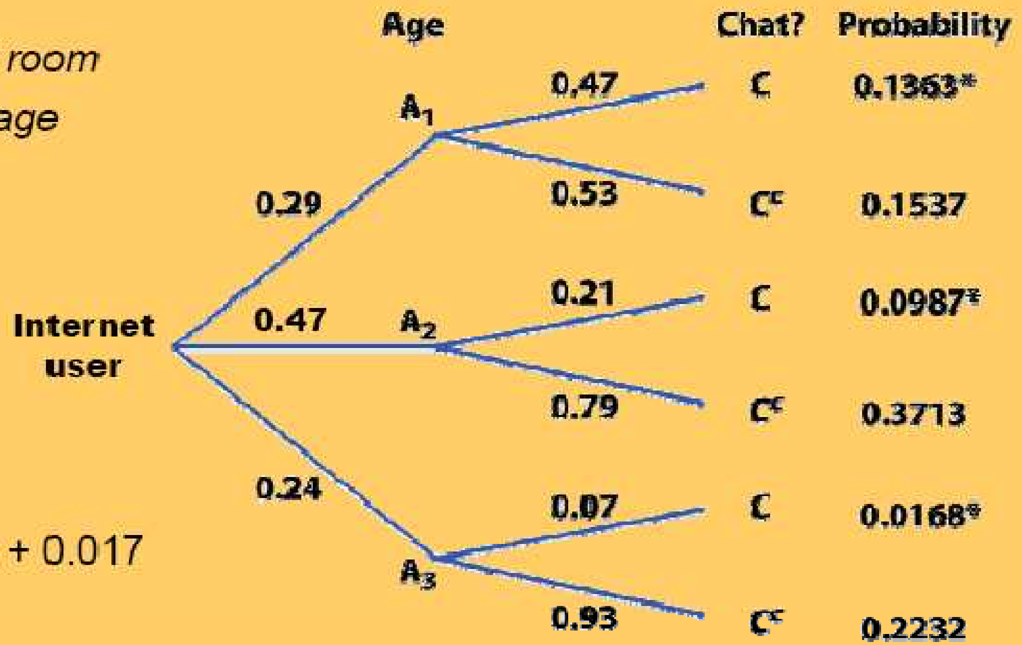
$$P[A_1 | B] \leq P[A_2 | B]. \qquad ////$$

Theorem 28 If $A_1, A_2, \ldots, A_n \in \mathscr{A}$, then

$$P[A_1 \cup A_2 \cup \cdots \cup A_n | B] \leq \sum_{j=1}^{n} P[A_i | B]. \qquad ////$$

Conditional probabilities can get complex, and it is often a good strategy to build a **probability tree** that represents all possible outcomes graphically and assigns conditional probabilities to subsets of events.

Tree diagram for chat room habits for three adult age groups



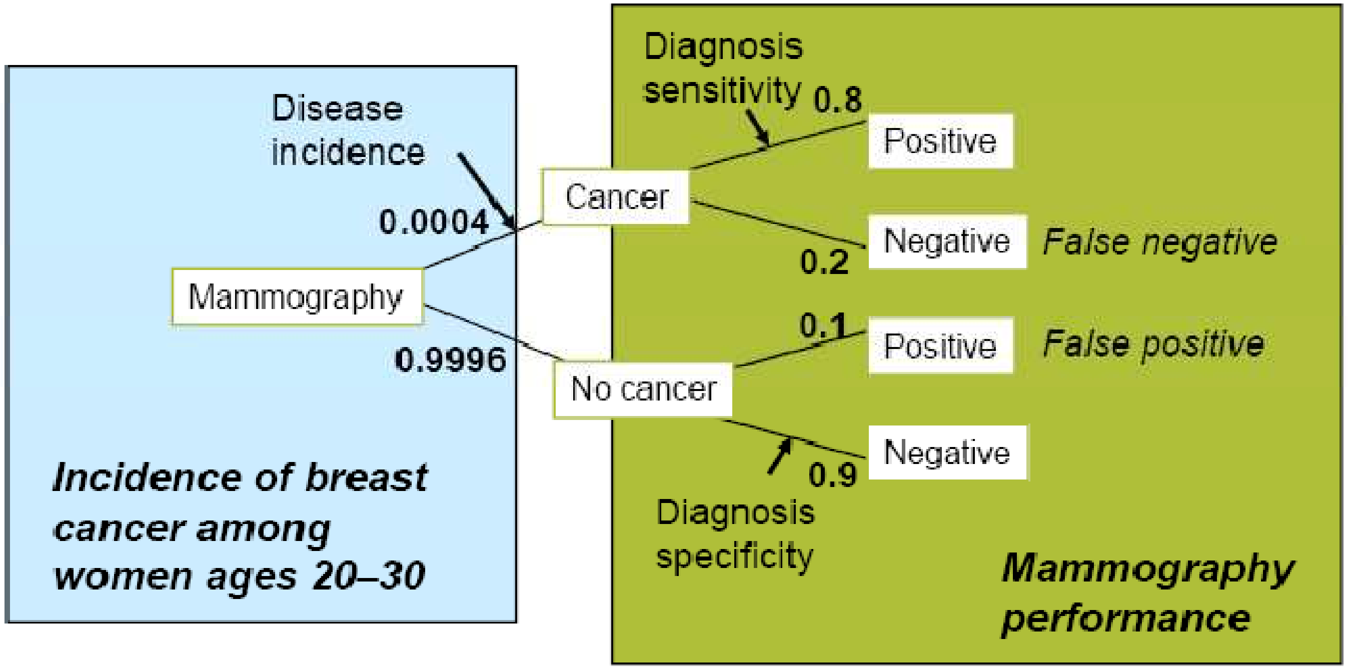
P(chatting) $= 0.136 + 0.099 + 0.017$

$= 0.252$

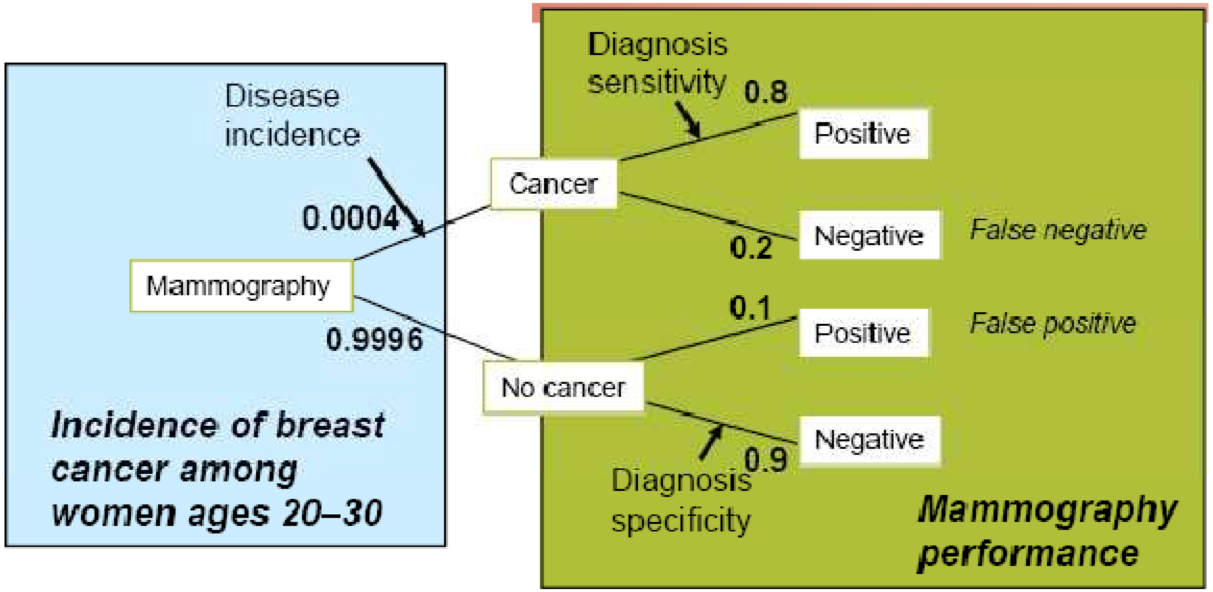
About 25% of all adult Internet users visit chat rooms.

Breast cancer screening

If a woman in her 20s gets screened for breast cancer and receives a positive test result, what is the probability that she does have breast cancer?



She could either have a positive test and have breast cancer or have a positive test but not have cancer (false positive).



Possible outcomes given the positive diagnosis: positive test and breast cancer or positive test but no cancer (false positive).

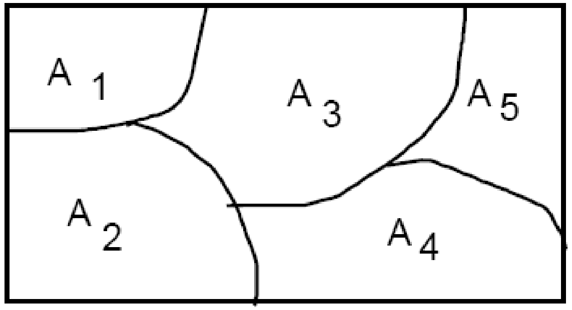
$$P(cancer \mid pos) = \frac{P(cancer\ and\ pos)}{P(cancer\ and\ pos) + P(nocancer\ and\ pos)}$$

$$= \frac{0.0004 * 0.8}{0.0004 * 0.8 + 0.9996 * 0.1} \approx 0.3\%$$

This value is called the positive predictive value, or *PV+*. It is an important piece of information but, unfortunately, is rarely communicated to patients.

4. Extremely useful results

Total Probability



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

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

Proof: This follows since

$$\begin{aligned}
\mathrm{P}(B) &= \mathrm{P}(B \mid A_1)\mathrm{P}(A_1) + \mathrm{P}(B \mid A_2)\mathrm{P}(A_2) + \ldots + \mathrm{P}(B \mid A_k)\mathrm{P}(A_k) \\
&= \mathrm{P}(B \cap A_1) + \mathrm{P}(B \cap A_2) + \ldots + \mathrm{P}(B \cap A_k) \\
&= \mathrm{P}(B \cap A_1 \text{ or } B \cap A_2 \text{ or.. + or } B \cap A_k) \\
&= \mathrm{P}(B \cap (A_1 \text{ or } A_2 \text{ or } A_k)) \\
&= \mathrm{P}(B)
\end{aligned}$$

Corollary For a given probability space $(\Omega, \mathscr{A}, P[\cdot])$ let $B \in \mathscr{A}$ satisfy $0 < P[B] < 1$; then for every $A \in \mathscr{A}$

$$P[A] = P[A \mid B]P[B] + P[A \mid \bar{B}]P[\bar{B}]. \qquad ////$$

Remark The theorem of total probabilities remains true when n is infinitely large

Bayes' Theorem

The multiplication rule gives $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$.
Bayes' theorem follows by diving 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):

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

Corollary For a given probability space $(\Omega, \mathscr{A}, P[\cdot])$ let A and $B \in \mathscr{A}$ satisfy $P[A] > 0$ and $0 < P[B] < 1$; then

$$P[B|A] = \frac{P[A|B]P[B]}{P[A|B]P[B] + P[A|\bar{B}]P[\bar{B}]}. \qquad ////$$

Remark Bayes' theorem remains true when n is infinitely large

***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

A 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(heads)=1/2. So the 8th toss of a fair coin is still P(tails)= 1/2.

Let B = coin is biased, F= coin is fair ($F = B^c$) and 7H be seeing seven heads in first seven tosses. Know $P(7H|\text{B}) = 1, P(7H|F) = \left(\frac{1}{2}\right)^7 = \frac{1}{128}, P(F) = P(B) = \frac{1}{2}$ 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 at P(A|B,C) means 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.

Bayes' rule for multiple events

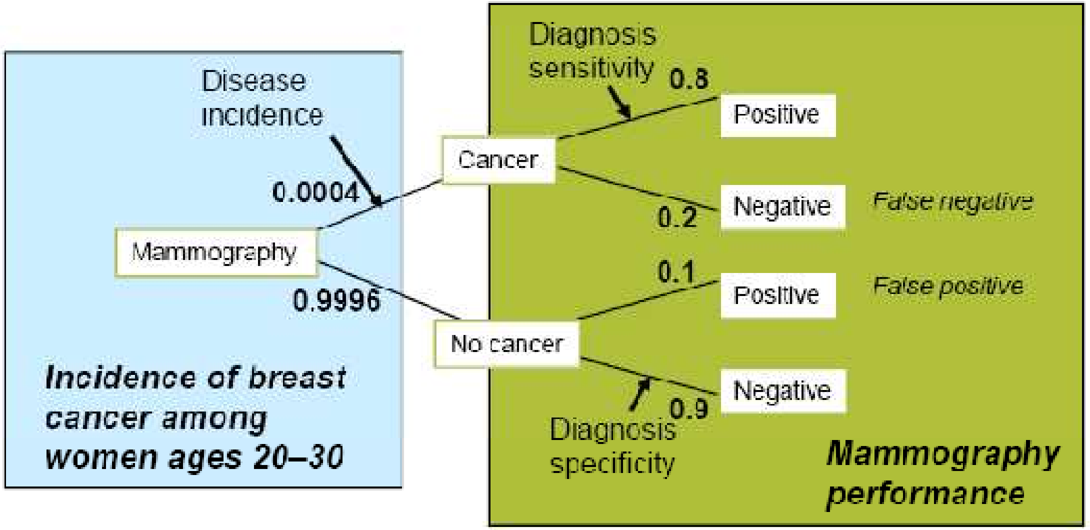
* If a sample space is decomposed in k disjoint events, A_1, A_2, … , A_k
— none with a null probability but $P(A_1) + P(A_2) + … + P(A_k) = 1$,
* And if C is any other event such that $P(C)$ is not 0 or **1**, then:

$$P(A_i \mid C) = \frac{P(C \mid A_i)P(A_i)}{P(C \mid A_1)P(A_1) + P(C \mid A_2)P(A_2) + \cdots + P(A_k)P(C \mid A_k)}$$

However, it is often intuitively much easier to work out answers with a probability tree than with these lengthy formulas.

Example: Breast cancer screening

If a woman in her 20s gets screened for breast cancer and receives a positive test result, what is the probability that she does have breast cancer?



This time, we use Bayes's rule: $P(A_i \mid C) = \dfrac{P(C \mid A_i)P(A_i)}{P(C \mid A_1)P(A_1) + P(C \mid A_2)P(A_2) + \cdots + P(A_k)P(C \mid A_k)}$

A1 is cancer, A2 is no cancer, *C* is a positive test result.

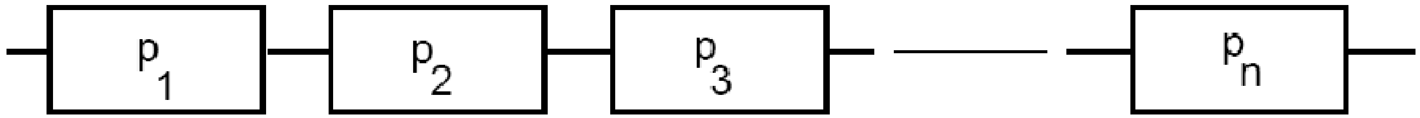
$$P(cancer \mid pos) = \frac{P(pos \mid cancer)P(cancer)}{P(pos \mid cancer)P(cancer) + P(pos \mid nocancer)P(nocancer)}$$

$$= \frac{0.8 * 0.0004}{0.8 * 0.0004 + 0.1 * 0.9996} \approx 0.3\%$$

5. 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).

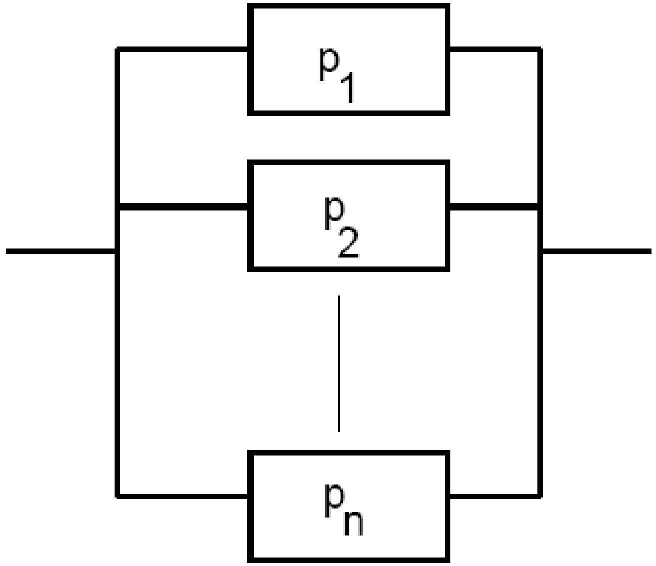
i.e. P(System does not fail) =

P(Element 1 doesn't fail <u>and</u> Element 2 doesn't fail <u>and</u> ... <u>and</u> 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]

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

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)

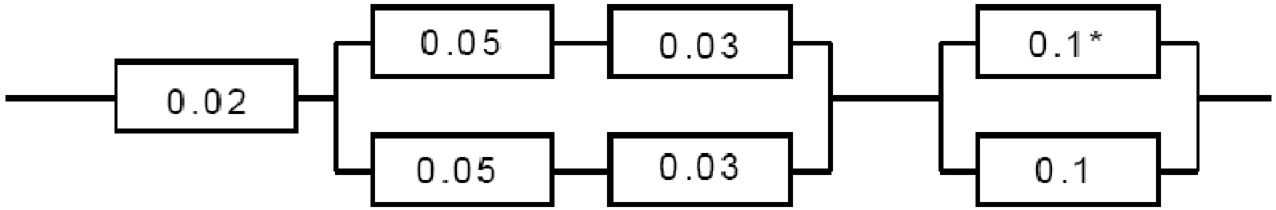
= P(Element 1 fails)P(Element 2 fails) ... P(Element n fails)

[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.



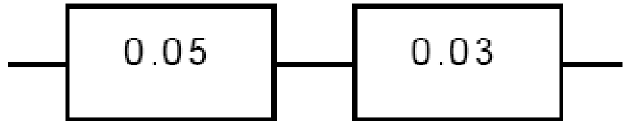
(a) What is the probability that the system does not fail in the next year?

(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?

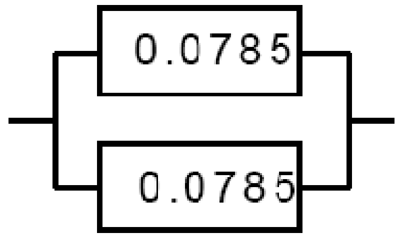
Solution

Subsystem 1:



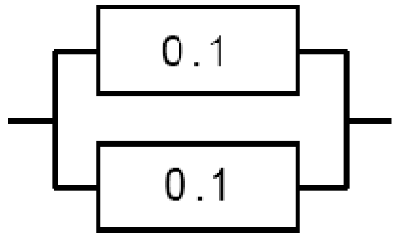
P(Subsystem 1 doesn't fail) = (1 - 0.05)(1 - 0.03) = 0.9215
P(Subsystem 1 fails) = 0.0785

Subsystem 2: (two units of subsystem 1)



P(Subsystem 2 fails) = 0.0785 x 0.0785 = 0.006162

Subsystem 3:
P(Subsystem 3 fails) = 0.1 x 0.1 = 0.01



System (summarised):

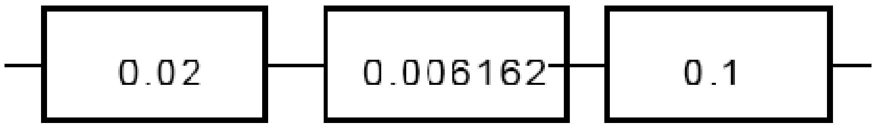
P(System doesn't fail) =

(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 and C).
Now, P(C) = 0.1.
Also, P(B | C) = P(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) x (1 - 6.162x10^{-3})(1 - 0.1) = 0.8766

$\therefore$ P(B and C) = P(B | C) P(C) = 0.8766 x 0.1 = 0.08766

Supplementary section

Combinatorics

Permutations - ways of ordering k items: k!

Factorials: for a positive integer k, $k! = k(k-1)(k-2) \ldots 2.1$

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 $_nC_k$ 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

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

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! x 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 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

$$\frac{n!}{k!\,(n-k)!} = \frac{n(n-1)(n-2)\ldots(n-k+1)(n-k)(n-k-1)..1}{k!\,(n-k)(n-k-1)\ldots 1}$$
$$= \frac{n(n-1)..(n-k+1)}{k(k-1)(k-2)..1}$$

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”.