MATH 323: Probability

Fall 2021

Alia Sajjad & David A. Stephens

Department of Mathematics and Statistics, McGill University

Target Syllabus

Syllabus

1. The basics of probability.
    - Review of set theory notation.
    - Sample spaces and events.
    - The probability axioms and their consequences.
    - Probability spaces with equally likely outcomes.
    - Combinatorial probability.
    - Conditional probability and independence.
    - The Theorem of Total Probability.
    - Bayes Theorem.

Syllabus (cont.)

2. Random variables and probability distributions.
   - Random variables.
   - Univariate distributions: cdfs, pmfs and pdfs.
   - Moments: expectation and variance.
   - Moment generating functions (mgfs): derivation and uses.
   - Named distributions:
     - discrete uniform,
     - hypergeometric,
     - binomial,
     - Poisson,
     - continuous uniform,
     - gamma,
     - exponential,
     - chi-squared,
     - beta,
     - Normal.

Syllabus (cont.)

3. Probability calculation methods.
    - Transformations in one dimension.
    - Techniques for sums of random variables.

Syllabus (cont.)

4. Multivariate distributions.
   - Marginal cdfs and pdfs.
   - Conditional cdfs and pdfs.
   - Conditional expectation.
   - Independence of random variables.
   - Covariance and correlation.

5. Probability inequalities and theorems.
    - Markov's inequality.
    - Chebychev's inequality.
    - Definition of convergence in probability.
    - The Weak Law of Large Numbers.
    - The Central Limit Theorem and applications.

Pre-requisite knowledge

- MATH 140/141
  - Functions, limits
  - Basic calculus methods (differentiation/integration in 1 dimension)
  - Sequences and series
- MATH 133

Later in the course we will be introducing some basic methods concerning *multi-variable* calculus.

If you have already taken MATH 222 or MATH 240, or courses in analysis, that knowledge will be useful but not essential.

Introduction

Introduction

This course is concerned with developing mathematical concepts and techniques for modelling and analyzing situations involving **uncertainty**.

- *Uncertainty* corresponds to a lack of complete or perfect knowledge.
- Assessment of uncertainty in such real-life problems is a complex issue which requires a rigorous mathematical treatment.

This course will develop the probability framework in which questions of practical interest can be posed and resolved.

Introduction (cont.)

Uncertainty could be the result of

- incomplete observation of a system;
- unpredictable variation;
- simple lack of knowledge of the “state of nature”.

Introduction (cont.)

“State of nature”: some aspect of the real world.

- could be the *current* state that is imperfectly observed;
- could be the *future* state, that is, the result of an experiment yet to be carried out.

Example (Uncertain states of nature)

- the outcome of a single coin toss;
- the millionth digit of π;

  $$3.14159265358979323846264338327950288841971....$$

- the height of the building I am in;
- the number of people on campus now;
- the temperature at noon tomorrow.

Introduction (cont.)

Note

The millionth digit of π is a fixed number: however, unless we know its value, we are still in a state of uncertainty when asked to assess it, as we have a lack of perfect knowledge.

Example (Coin tossing)

If I take a coin and decide to carry out an experiment where I toss the coin once and see which face is turned upward when the coin comes to rest, then I can assess that there are two possible results of the toss:

The outcome is of the toss is uncertain *before* I toss the coin, and *after* I toss the coin *until I see the result*.

Example (Thumbtack tossing)

If I take a thumbtack and decide to carry out an experiment where I toss the thumbtack once, then I can assess that there are two possible results of the toss:

The outcome of the toss is uncertain *before* I toss the thumbtack, and *after* I toss the thumbtack *until I see the result*.

Chapter 1: The basics of probability

What is probability ?

By probability, we generally mean a numerical assessment of

the chance of a particular event occurring, given a particular set of circumstances.

What is probability ?

That is,

(i) the chance of a particular event occurring ...

What is probability ?

That is,

(i) the chance of a particular event occurring ...

(ii) ... given a particular set of circumstances.

What is probability ?

It has seemed to me that the theory of probabilities ought to serve as the basis for the study of all the sciences.

Adolphe Quetelet

The probable is what usually happens.

Aristotle

Probability is the very guide of life.

Cicero, *De Natura*, 5, 12

`https://www.stat.berkeley.edu/∼jpopen/probweb/quotes.html`

My thesis, paradoxically, and a little provocatively, but nonetheless genuinely, is simply this:

PROBABILITY DOES NOT EXIST

The abandonment of superstitious beliefs about the existence of the Phlogiston, the Cosmic Ether, Absolute Space and Time, ... or Fairies and Witches was an essential step along the road to scientific thinking.

What is probability ? (cont.)

Probability, too, if regarded as something endowed with some kind of objective existence, is no less a misleading misconception, an illusory attempt to exteriorize or materialize our true probabilistic beliefs.

B. de Finetti, *Theory of Probability*, 1970.

What is probability ? (cont.)

That is,

- probability is not some *objectively* defined quantity – it has a *subjective* interpretation.
- this subjective element may be commonly agreed upon in certain circumstances, but in general is always present.
- recall
  - coin example,
  - thumbtack example.

Constructing the mathematical framework

For a given ‘experiment’, we

- identify the possible outcomes;
- assign numerical values to collections of possible values to represent the chance that they will coincide with the actual outcome;
- lay out the rules for how the numerical values can be manipulated.

Basic concepts in set theory

A set S is a collection of individual *elements*, such as s.

We write

$$s \in S$$

to mean "s is one of the elements of S".

- $S = \{0, 1\}$; $0 \in S, 1 \in S$.
- $S = \{\text{'dog'}, \text{'cat'}, \text{'mouse'}, \text{'rat'}\}$; $\text{'rat'} \in S$
- $S = [0, 1]$ (the *closed* interval from zero to one); then

$$0.3428 \in S$$

S may be

- *finite* if it contains a finite number of elements;
- *countable* if it contains a countably infinite number of elements;
- *uncountable* if it contains an uncountably infinite number of elements.

A is a *subset* of S if it contains some, all, or none of the elements of S

$$\text{some: } A \subset S \qquad \text{some or all: } A \subseteq S.$$

That is, A is a subset of S if every element of A is also an element of S

$$s \in A \implies s \in S.$$

- $S = \{0, 1\}$; $A = \{0\}$, or $A = \{1\}$, or $A = \{0, 1\}$.
- $S = \{\text{'dog'}, \text{'cat'}, \text{'mouse'}, \text{'rat'}\}$; $A = \{\text{'cat'}, \text{'mouse'}\}$
- $S = [0, 1]$; $A = (0.25, 0.3)$.

Special case: the *empty* set, $\emptyset$, is a subset that contains no elements.

Basic concepts in set theory (cont.)

Note

In probability, it is necessary to think of *sets of subsets* of S.
For example:

- $S = \{1, 2, 3, 4\}$.
- Consider all the subsets of S:

$$\text{One element:}\{1\}, \{2\}, \{3\}, \{4\}$$

$$\text{Two elements:}\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}$$

$$\text{Three elements:}\{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}$$

$$\text{Four elements:}\{1, 2, 3, 4\}.$$

- Add to this list the empty set (zero elements), $\emptyset$.

Thus there is a collection of 16 (that is, 2^4) subsets of S.

Basic concepts in set theory (cont.)

Note

This is a bit trickier if S is an interval ...

Set operations

To manipulate sets, we use three basic operations:

- *intersection* $\cap$
- *union* $\cup$
- *complement* $'$

Consider two sets A and B which are subsets of a set S.

Set operations (cont.)

- intersection: the *intersection* of two sets A and B is the collection of elements that are elements of *both* A and B.

$$s \in A \cap B \qquad \Longleftrightarrow \qquad s \in A \text{ and } s \in B.$$

We have for any $A, B \subseteq S$ that

$$A \cap \emptyset = \emptyset$$

$$A \cap S = A$$

$$A \cap B \subseteq A$$

$$A \cap B \subseteq B.$$

Set operations (cont.)

- union: the *union* of two sets A and B is the set of distinct elements that are either in A, or in B, or in both A and B.

$$s \in A \cup B \quad \iff \quad s \in A \text{ or } s \in B \ (\text{or } s \in A \cap B).$$

We have for any $A, B \subseteq S$ that

$$A \cup \emptyset = A$$

$$A \cup S = S$$

$$A \subseteq A \cup B$$

$$B \subseteq A \cup B$$

Set operations (cont.)

- complement: the *complement* of set A, a subset of S, is the collection of elements of S that are not elements of A.

$$s \in A' \quad \iff \quad s \in S \text{ but } s \notin A.$$

We have that

$$A \cap A' = \emptyset \quad \text{and} \quad A \cup A' = S.$$

We have for any $A \subseteq S$ that

$$(A')' = A.$$

Set operations (cont.)

Note

The notations $\overline{A}$ and A^c may also be used.

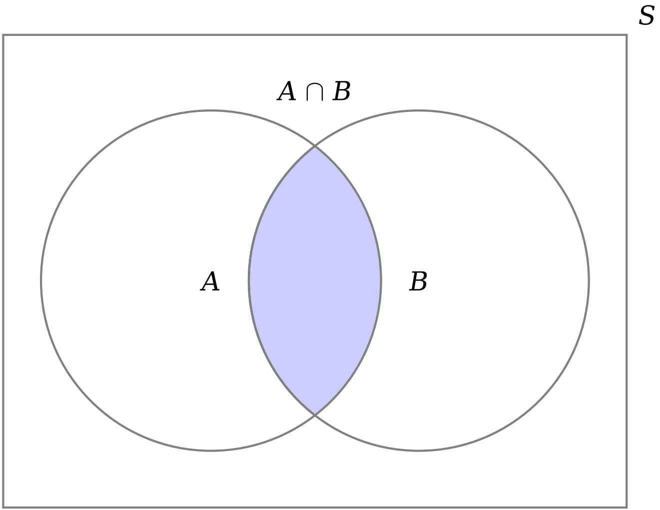
The textbook uses

$$\overline{A} \qquad \overline{(A \cap B)}$$

etc., but I find this leads to confusion.

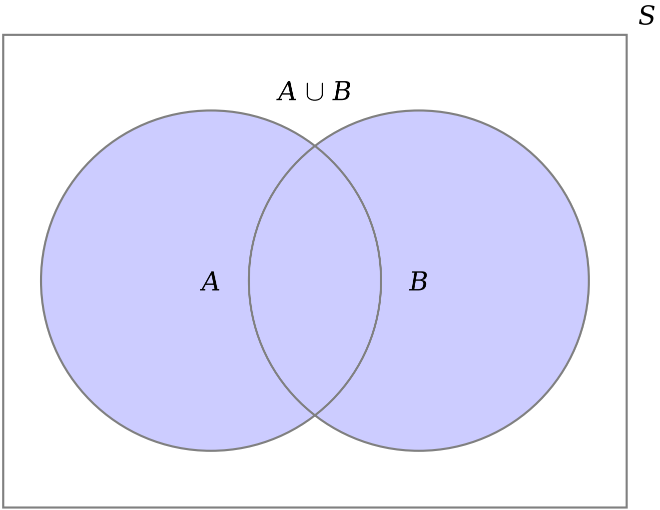
Set operations: in pictures

Intersection:



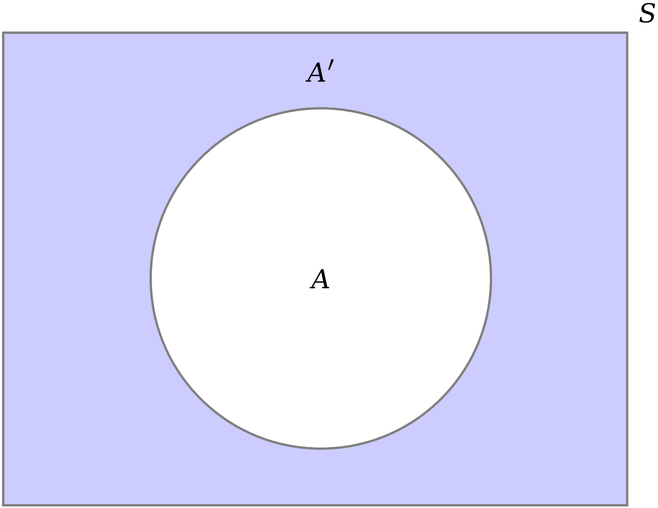
Set operations: in pictures (cont.)

Union:



Set operations: in pictures (cont.)

Complement:



Example

Example (Finite set)

- $S = \{1, 2, 3, \ldots, 9, 10\}$
- $A = \{2, 4, 6\}$
- $B = \{1, 2, 5, 7, 9\}$

Then

- $A \cap B = \{2\}$
- $A \cup B = \{1, 2, 4, 5, 6, 7, 9\}$
- $A' = \{1, 3, 5, 7, 8, 9, 10\}$

Set operations: extensions

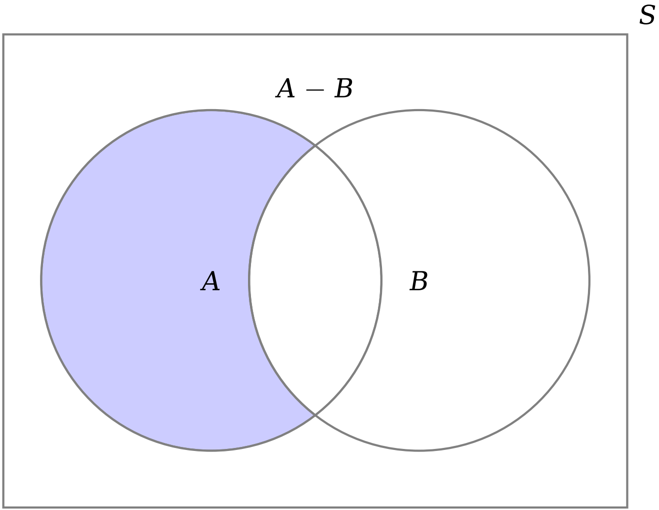
- Set difference: $A - B$ (or $A \backslash B$)

$$s \in A - B \iff s \in A \text{ and } s \in B'$$

that is

$$A - B \equiv A \cap B'.$$

Set operations: extensions (cont.)



Set operations: extensions (cont.)

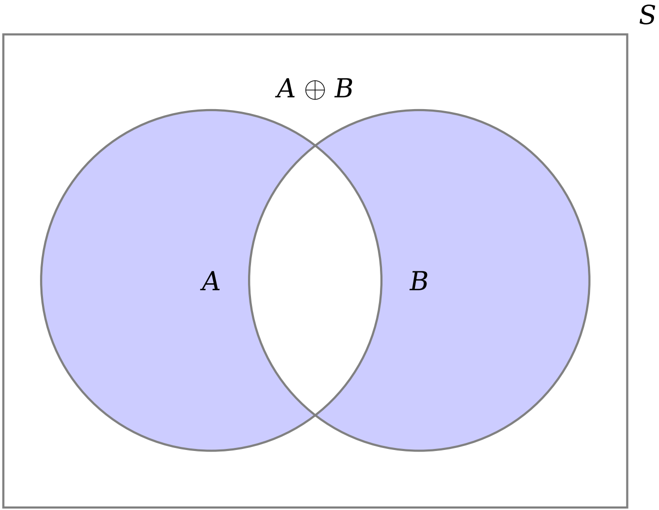
- A or B but not both: $A \oplus B$

$$s \in A \oplus B \iff s \in A \text{ or } s \in B, \text{ but } s \notin A \cap B$$

aka

$$(A \cup B) - (A \cap B)$$

Set operations: extensions (cont.)



Combining set operations

Both intersection and union operators are *binary* operators, that is, take *two* arguments:

$$A \cap B \qquad A \cup B.$$

We have immediately from the definitions that

$$A \cap B = B \cap A$$

and that

$$A \cup B = B \cup A$$

(that is, *the order does not matter*).

Combining set operations (cont.)

However, A and B are arbitrary, so suppose there is a third set $C \subseteq S$, and consider

$$(A \cap B) \cap C$$

as $A \cap B$ is a set in its own right. Then

$$\begin{aligned} s \in (A \cap B) \cap C \quad &\iff \quad s \in (A \cap B) \text{ and } s \in C \\ &\iff \quad s \in A \text{ and } s \in B \text{ and } s \in C. \end{aligned}$$

Combining set operations (cont.)

That is

$$(A \cap B) \cap C = A \cap B \cap C$$

and by the same logic

$$(A \cap B) \cap C = A \cap (B \cap C)$$

and so on.

Combining set operations (cont.)

This extends to the case of any finite collection of sets

$$A_1, A_2, \ldots, A_K$$

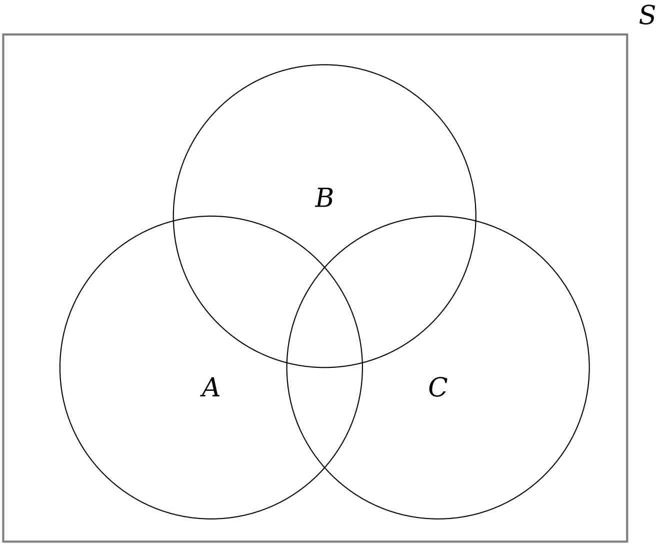
and we may write

$$A_1 \cap A_2 \cap \ldots \cap A_K \equiv \bigcap_{k=1}^{K} A_k$$

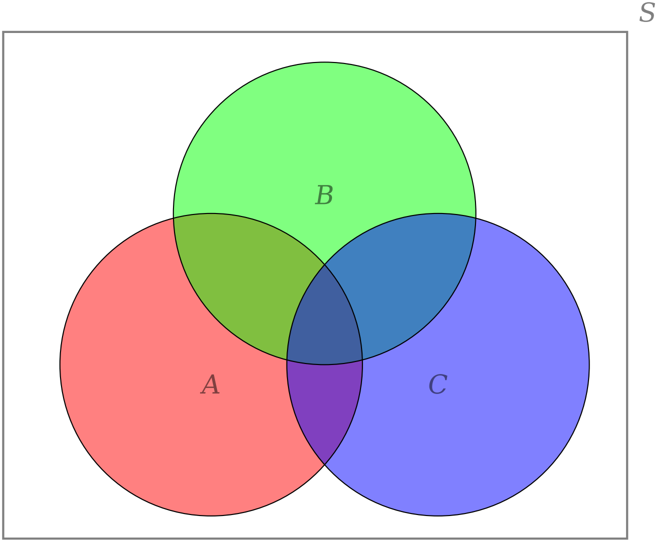
where

$$s \in \bigcap_{k=1}^{K} A_k \qquad \Longleftrightarrow \qquad s \in A_k \text{ for all } k$$

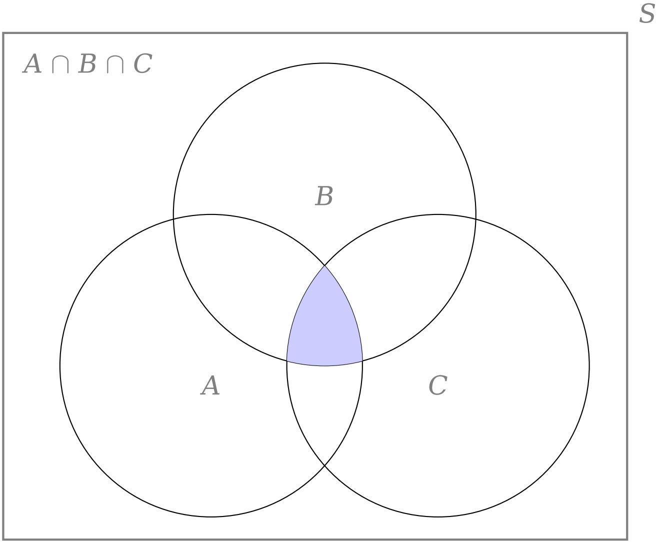
Combining set operations (cont.)



Combining set operations (cont.)



Combining set operations (cont.)



Combining set operations (cont.)

The same logic holds for the union operator:

$$(A \cup B) \cup C = A \cup B \cup C = A \cup (B \cup C)$$

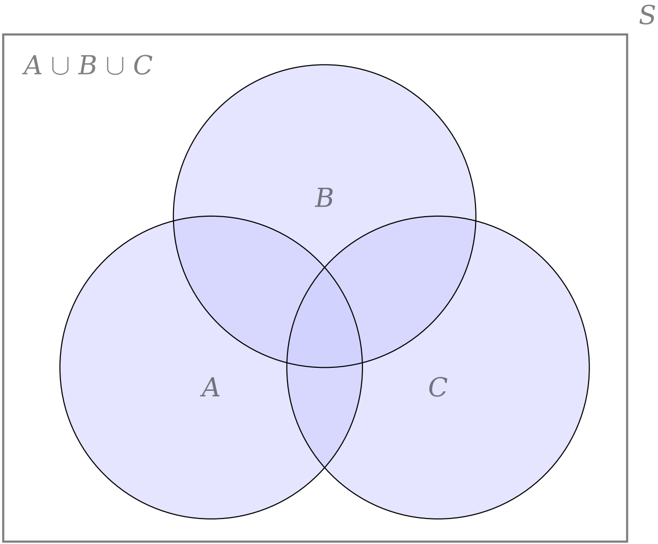
and

$$A_1 \cup A_2 \cup \ldots \cup A_K \equiv \bigcup_{k=1}^{K} A_k$$

where

$$s \in \bigcup_{k=1}^{K} A_k \qquad \Longleftrightarrow \qquad s \in A_k \text{ for at least one } k$$

Combining set operations (cont.)



Combining set operations (cont.)

Note

The intersection and union operations can be extended to work with a countably infinite number of sets

$$A_1, A_2, \ldots, A_k, \ldots$$

and we can consider

$$\bigcap_{k=1}^{\infty} A_k \quad : \quad s \in \bigcap_{k=1}^{\infty} A_k \iff s \in A_k \text{ for all } k$$

$$\bigcup_{k=1}^{\infty} A_k \quad : \quad s \in \bigcup_{k=1}^{\infty} A_k \iff s \in A_k \text{ for at least one } k$$

Combining set operations (cont.)

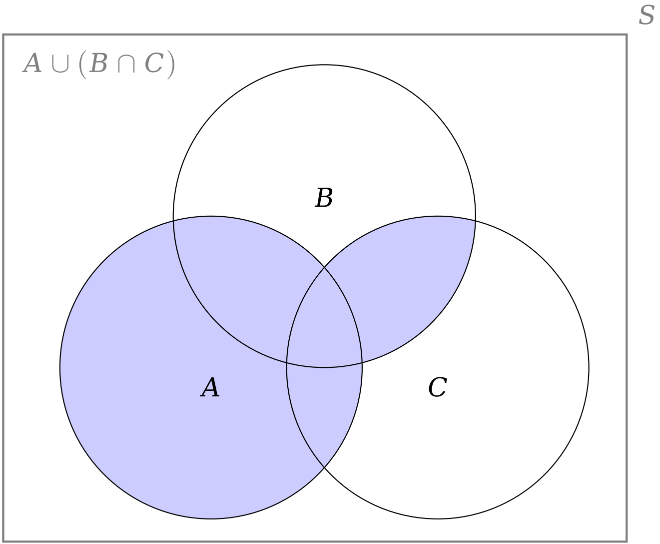
Consider now

$$A \cup (B \cap C)$$

To be an element in this set, you need to be

- an element of A or
- an element of $B \cap C$ or
- an element of both A and $B \cap C$.

Combining set operations (cont.)



Combining set operations (cont.)

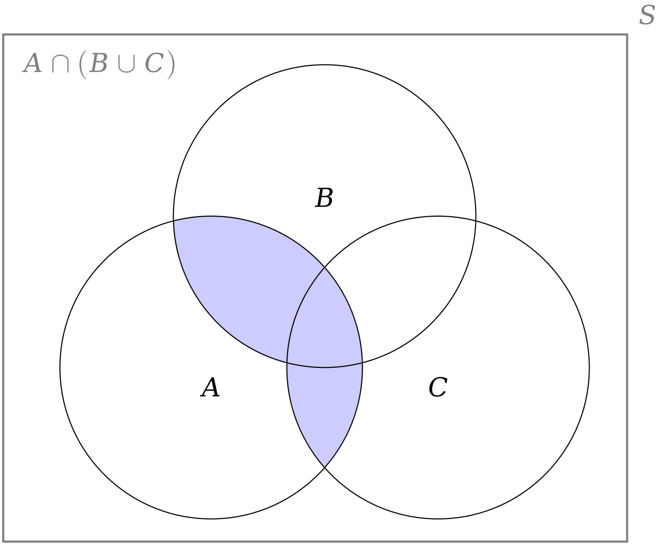
We can therefore deduce that

$$A \cup (B \cap C) \equiv (A \cup B) \cap (A \cup C).$$

Similarly

$$A \cap (B \cup C) \equiv (A \cap B) \cup (A \cap C).$$

Combining set operations (cont.)



Partitions

We call $A_1, A_2, \ldots, A_K$ a *partition* of S if these sets are

- pairwise *disjoint* (or *mutually exclusive*):

$$A_j \cap A_k = \emptyset \quad \text{for all } j \neq k$$

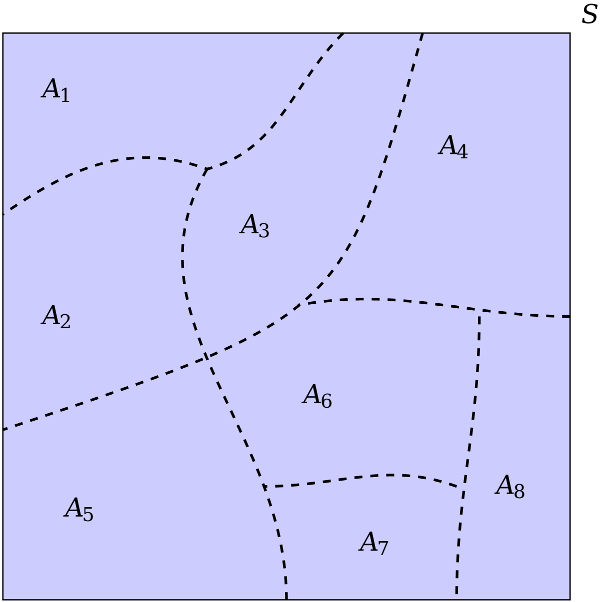
- exhaustive:

$$\bigcup_{k=1}^{K} A_k = S.$$

(that is, the As cover the whole of S, but do not overlap).

For every $s \in S$, s is an element of precisely one of the A_k.

Partitions (cont.)

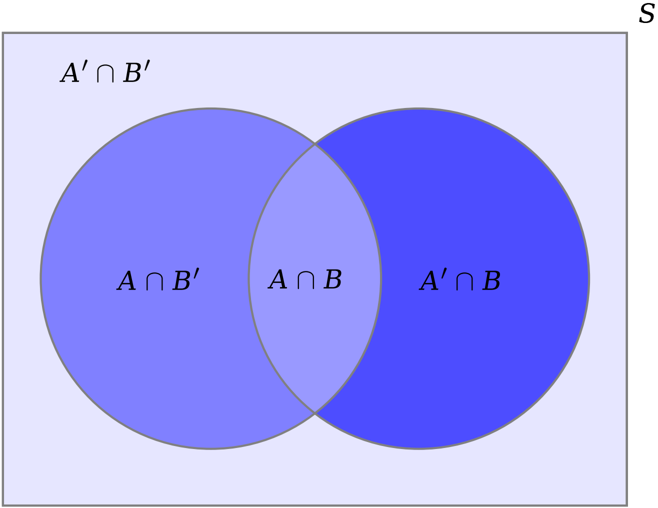


Some simple theorems

For any two subsets A and B of S, we have the partition of S via the four sets

$$A \cap B \qquad A \cap B' \qquad A' \cap B \qquad A' \cap B'$$

Some simple theorems (cont.)



Some simple theorems (cont.)

This picture implies that

$$(A \cup B)' = A' \cap B'.$$

We can show this as follows: let

$$A_1 = A \cup B \qquad\qquad A_2 = A' \cap B'.$$

We need to show that

$$A_1' = A_2$$

that is,

(i) $A_1 \cap A_2 = \emptyset$;

(ii) $A_1 \cup A_2 = S$.

Some simple theorems (cont.)

(i) Disjoint:

$$\begin{aligned}
A_1 \cap A_2 &= (A \cup B) \cap A_2 \\
&= (A \cap A_2) \cup (B \cap A_2) \\
&= (A \cap A' \cap B') \cup (B \cap A' \cap B') \\
&= \emptyset \cup \emptyset \\
&= \emptyset
\end{aligned}$$

(ii) Exhaustive:

$$
\begin{aligned}
A_1 \cup A_2 &= A_1 \cup (A' \cap B') \\
&= (A_1 \cup A') \cap (A_1 \cup B') \\
&= (A \cup B \cup A') \cap (A \cup B \cup B') \\
&= S \cap S \\
&= S
\end{aligned}
$$

Some simple theorems (cont.)

Similarly

$$A' \cup B' = (A \cap B)'$$

which is equivalent to saying

$$(A' \cup B')' = A \cap B.$$

Some simple theorems (cont.)

The first result

$$(A \cup B)' = A' \cap B'.$$

holds for arbitrary sets A and B. In particular it holds for the sets

$$C = A' \qquad\qquad D = B'$$

that is we have

$$(C \cup D)' = C' \cap D'.$$

Some simple theorems (cont.)

But

$$C' = (A')' = A \qquad\qquad D' = (B')' = B$$

Therefore

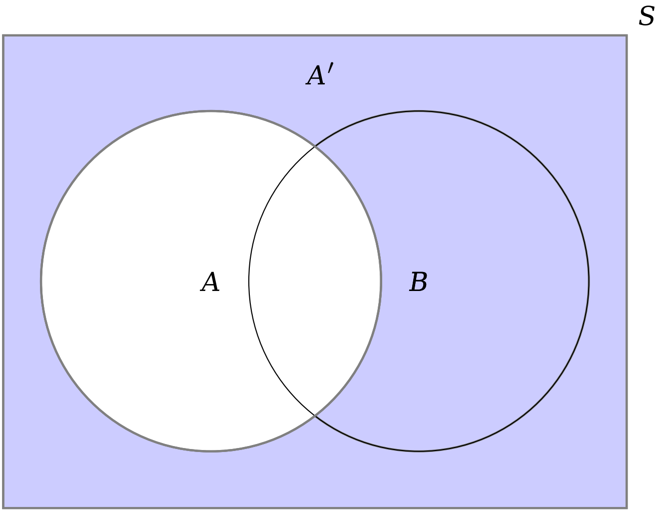
$$(A' \cup B')' = A \cap B$$

or equivalently

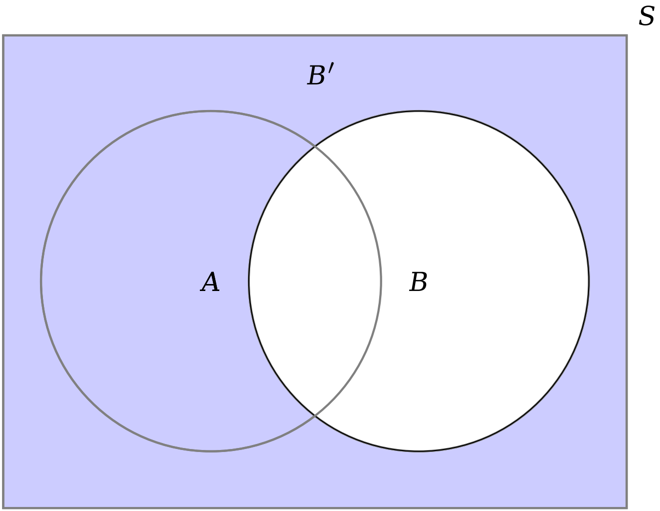
$$A' \cup B' = (A \cap B)'$$

as required.

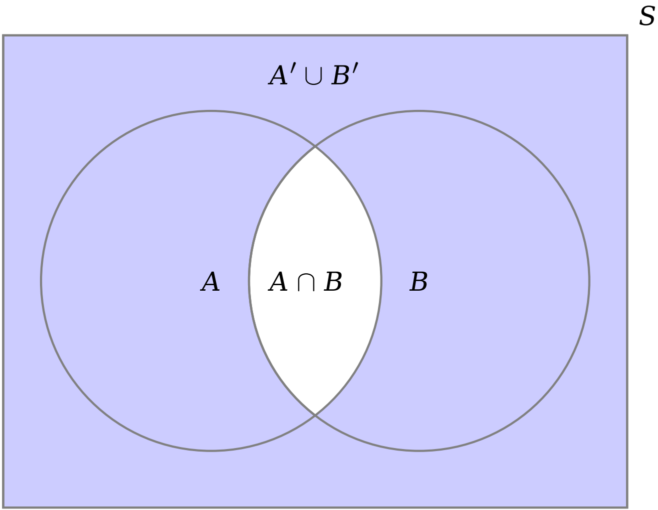
Some simple theorems (cont.)



Some simple theorems (cont.)



Some simple theorems (cont.)



Some simple theorems (cont.)

These two results

$$A' \cap B' = (A \cup B)'$$

$$A' \cup B' = (A \cap B)'$$

are sometimes known as de Morgan's Laws.

Sample spaces and Events

We now utilize the set theory formulation and notation in the probability context.

Recall the earlier informal definition:

> *By probability, we generally mean the chance of a particular event occurring, given a particular set of circumstances. The probability of an event is generally expressed as a quantitative measurement.*

We need to carefully define what an 'event' is, and what constitutes a 'particular set of circumstances'.

Sample spaces and Events (cont.)

We will consider the general setting of an *experiment*:

- this can be interpreted as any setting in which an uncertain consequence is to arise;
- could involve observing an outcome, taking a measurement etc.

1. **Consider the possible outcomes of the experiment:** Make a 'list' of the outcomes that can arise, and denote the corresponding set by S.
   - $S = \{0, 1\}$;
   - $S = \{\text{'head'}, \text{'tail'}\} \equiv \{\mathrm{H}, \mathrm{T}\}$;
   - $S = \{\text{'cured'}, \text{'not cured'}\}$;
   - $S = \{\text{'Arts'}, \text{'Engineering'}, \text{'Medicine'}, \text{'Science'}\}$;
   - $S = \{1, 2, 3, 4, 5, 6\}$;
   - $S = \mathbb{R}^+$.

   The set S is termed the *sample space* of the experiment. The individual elements of S are termed *sample points* (or *sample outcomes*).

Constructing the framework (cont.)

2. **Events**: An *event* A is a collection of sample outcomes. That is, A is a subset of S,

$$A \subseteq S.$$

   For example

   - $A = \{0\}$;
   - $A = \{\text{'tail'}\} \equiv \{\mathrm{T}\}$;
   - $A = \{\text{'cured'}\}$;
   - $A = \{\text{'Arts'}, \text{'Engineering'}\}$;
   - $A = \{1, 3, 5\}$;
   - $A = [2, 3)$.

   The individual sample outcomes are termed *simple* (or *elementary*) events, and may be denoted

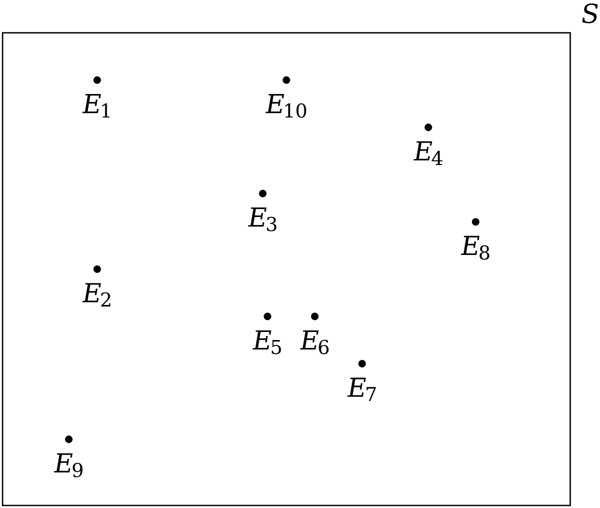
$$E_1, E_2, \ldots, E_K, \ldots.$$

3. **Terminology**: We say that event A *occurs* if the actual outcome, s, is an element of A.

   For two events A and B

   - $A \cap B$ occurs if and only if A occurs and B occurs, that is
   $$s \in A \cap B.$$
   - $A \cup B$ occurs if A occurs or if B occurs, or if both A and B occur, that is
   $$s \in A \cup B$$
   - if A occurs, then A' does not occur.

Constructing the framework (cont.)

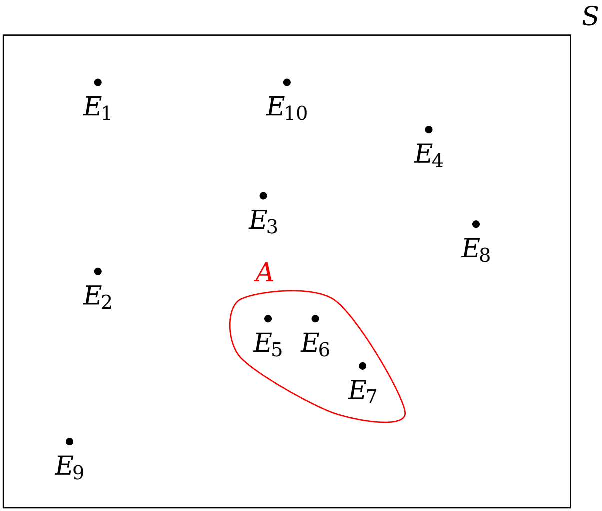


Constructing the framework (cont.)

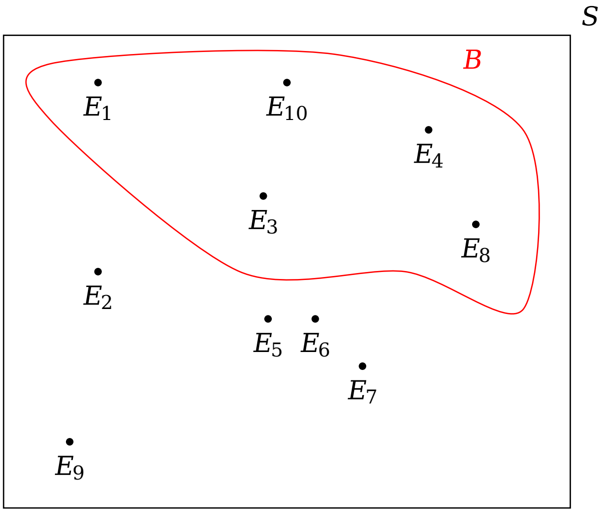
In this case

$$S = \bigcup_{k=1}^{10} E_k .$$

Constructing the framework (cont.)



Constructing the framework (cont.)



Constructing the framework (cont.)

In this case

$$A = E_5 \cup E_6 \cup E_7$$

and

$$B = E_1 \cup E_3 \cup E_4 \cup E_8 \cup E_{10}.$$

Constructing the framework (cont.)

Note

- The event S is termed the certain event;
- The event $\emptyset$ is termed the impossible event.

Mathematical definition of probability

Probability is a means of assigning a quantitative measure of uncertainty to events in a sample space.

Formally, the *probability function*, $P(.)$ is a function that assigns numerical values to events.

$$P : \mathcal{A} \longrightarrow \mathbb{R}$$

$$A \longmapsto p$$

that is

$$P(A) = p$$

that is, the probability assigned to event A (a set) is p (a numerical value).

Mathematical definition of probability (cont.)

Note

(i) The function $P(.)$ is a set function (that is, it takes a set as its argument).

(ii) $\mathcal{A}$ is a "set of subsets of S"; we pick a subset $A \in \mathcal{A}$ and consider its probability.

(iii) $\mathcal{A}$ has certain nice properties that ensure that the probability function can operate successfully.

Mathematical definition of probability (cont.)

This definition is too general:

- what properties does $P(.)$ have ?
- how does $P(.)$ assign numerical values; that is, how do we compute

  $$P(A)$$

  for a given event A in sample space S ?

The probability axioms

Suppose S is a sample space for an experiment, and A is an event, a subset of S. Then we assign $P(A)$, the probability of event A, so that the following axioms hold:

(I) $P(A) \geq 0$.

(II) $P(S) = 1$.

(III) If $A_1, A_2, \ldots$ form a (countable) sequence of events such that

$$A_j \cap A_k = \emptyset \qquad \text{for all } j \neq k$$

then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

that is, we say that $P(.)$ is *countably additive*.

The probability axioms (cont.)

Note

Axiom (III) immediately implies that $P(.)$ is *finitely additive*, that is, for all n, $1 \leq n < \infty$, if $A_1, A_2, \ldots, A_n$ form a sequence of events such that $A_j \cap A_k = \emptyset$ for all $j \neq k$, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

To see this, fix n and define $A_i = \emptyset$ for $i > n$.

The probability axioms (cont.)

Example (Partitions)

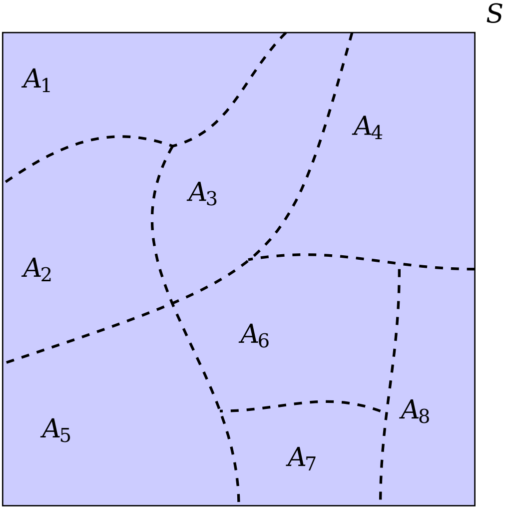
If $A_1, A_2, \ldots, A_n$ form a partition of S, and

$$P(A_i) = p_i \qquad i = 2, \ldots, n$$

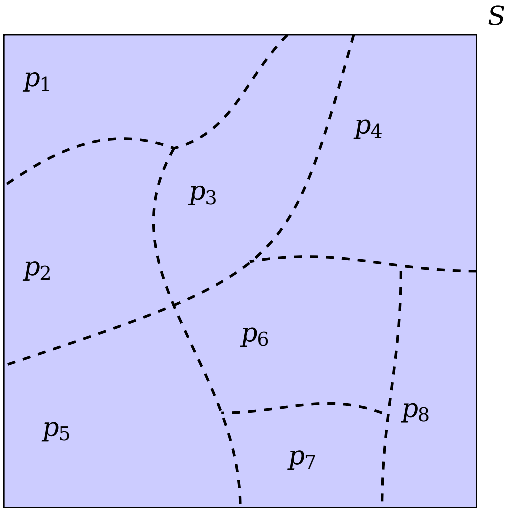
say, then

$$\sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{n} p_i = 1.$$

The probability axioms (cont.)



The probability axioms (cont.)



The probability axioms (cont.)

Some immediate corollaries of the axioms:

(i) For any A, $P(A') = 1 - P(A)$.

- We have that $S = A \cup A'$. By Axiom (III) we have that

$$P(S) = P(A) + P(A').$$

- By Axiom (II), $P(S) = 1$, so therefore

$$1 = P(A) + P(A') \qquad \therefore \qquad P(A') = 1 - P(A).$$

The probability axioms (cont.)

(ii) $P(\emptyset) = 0$.

- Apply the result from point (i) to the set $A \equiv S$.
- Note that by Axiom (II), $P(S) = 1$.

The probability axioms (cont.)

(iii) For any A, $P(A) \leq 1$.

- By Axiom (III) and point (i) we have

$$1 = P(S) = P(A) + P(A') \geq P(A)$$

as $P(A') \geq 0$ by Axiom (I).

The probability axioms (cont.)

(iv) For any two events A and B, if $A \subseteq B$, then

$$P(A) \leq P(B)$$

- We may write in this case

$$B = A \cup (A' \cap B)$$

  and, as the two events on the right hand side are mutually exclusive, by Axiom (III)

$$P(B) = P(A) + P(A' \cap B) \geq P(A)$$

  as $P(A' \cap B) \geq 0$ by Axiom (I).

The probability axioms (cont.)

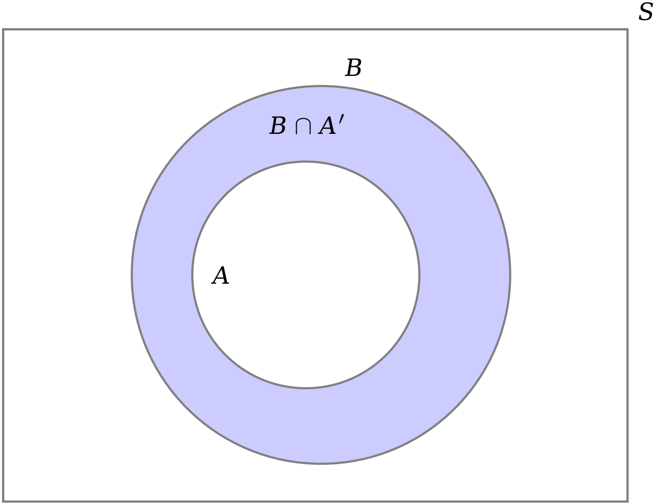
Therefore, for example,

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

that is

$$P(A \cap B) \leq \min\{P(A), P(B)\}.$$

The probability axioms (cont.)

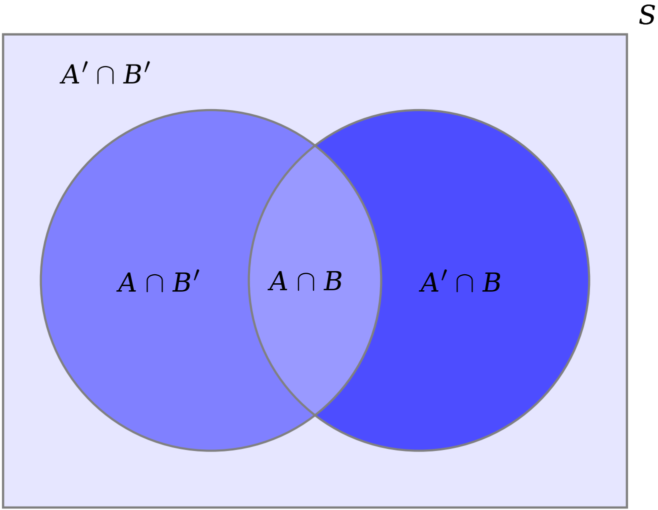


The probability axioms (cont.)

(v) **General Addition Rule:** For two arbitrary events A and B,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

The probability axioms (cont.)



The probability axioms (cont.)

- We have that

$$(A \cup B) = A \cup (A' \cap B)$$

so that by Axiom (III)

$$P(A \cup B) = P(A) + P(A' \cap B).$$

But

$$B = (A \cap B) \cup (A' \cap B)$$

so that

$$P(B) = P(A \cap B) + P(A' \cap B)$$

and therefore

$$P(A' \cap B) = P(B) - P(A \cap B)$$

The probability axioms (cont.)

Substituting back in, we have

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

as required.

- We may deduce from this that

$$P(A \cup B) \leq P(A) + P(B)$$

as $P(A \cap B) \geq 0$.

The probability axioms (cont.)

In general, for events $A_1, A_2, \ldots, A_n$, we can construct a formula for

$$P\left(\bigcup_{i=1}^{n} A_i\right)$$

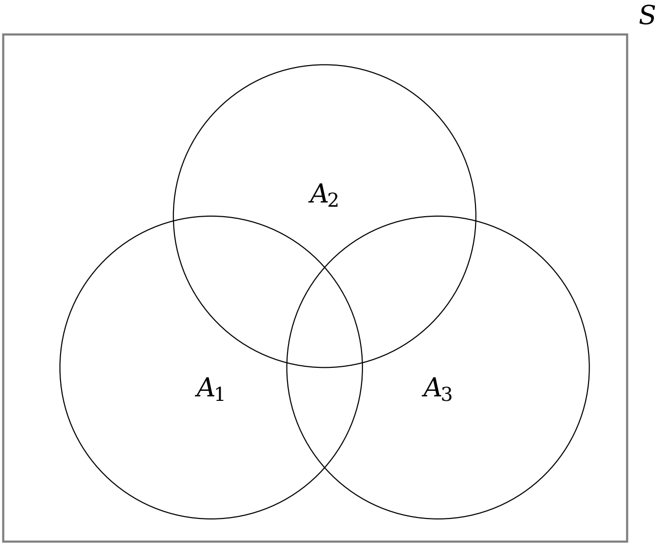
using inductive arguments.

The probability axioms (cont.)

For example, with $n = 3$, we have that

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) = \;& P(A_1) + P(A_2) + P(A_3) \\ & - P(A_1 \cap A_2) \\ & - P(A_1 \cap A_3) \\ & - P(A_2 \cap A_3) \\ & + P(A_1 \cap A_2 \cap A_3). \end{aligned}$$

The probability axioms (cont.)



A more straightforward general result is Boole's Inequality

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P\left(A_i\right).$$

Proof of this result follows by a simple inductive argument.

Probability tables

It is sometimes helpful to lay out probabilities in table form:

	A	A'	Total
B	$p_{A\cap B}$	$p_{A'\cap B}$	p_B
B'	$p_{A\cap B'}$	$p_{A'\cap B'}$	$p_{B'}$
Total	p_A	$p_{A'}$	p_S

that is, for example, $p_{A\cap B} = P(A \cap B)$, and

$$A = (A \cap B) \cup (A \cap B')$$

so that

$$p_A = P(A) = P(A \cap B) + P(A \cap B') = p_{A\cap B} + p_{A\cap B'}$$

and so on.

Example

Suppose

$$P(A) = \frac{1}{3} \qquad P(B) = \frac{1}{2} \qquad P(A \cup B) = \frac{3}{4}.$$

	A	A'	Total
B			$\frac{1}{2}$
B'			$\frac{1}{2}$
Total	$\frac{1}{3}$	$\frac{2}{3}$	1

Probability tables (cont.)

Example

We have

$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &= \frac{1}{3} + \frac{1}{2} - \frac{3}{4} \\ &= \frac{1}{12} \end{aligned}$$

Probability tables (cont.)

Example

Hence

	A	A'	Total
B	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{1}{2}$
B'	$\frac{3}{12}$	$\frac{3}{12}$	$\frac{1}{2}$
Total	$\frac{1}{3}$	$\frac{2}{3}$	1

Probability tables (cont.)

Not all configurations of entries in the table are valid: we need to ensure that

$$0 \leq p_* \leq 1$$

for all entries, with

$$p_A = p_{A \cap B} + p_{A \cap B'}$$

and

$$p_B = p_{A \cap B} + p_{A' \cap B}$$

Specifying probabilities

We have defined how the probability function must behave *mathematically*.

We now consider three ways via which we could specify the *numerical* probability of an event.

1. **Equally likely sample outcomes:** Suppose that sample space S is finite, with N sample outcomes in total that are considered to be equally likely to occur. Then for the elementary events $E_1, E_2, \ldots, E_N$, we have

$$P(E_i) = \frac{1}{N} \qquad i = 1, \ldots, N$$

Specifying probabilities (cont.)

Every event $A \subseteq S$ can be expressed

$$A = \bigcup_{i=1}^{n} E_{i_A}$$

for some $n \leq N$, and collection $E_{1_A}, \ldots, E_{n_A}$, with indices

$$i_A \in \{1, \ldots, n\}.$$

Then

$$P(A) = \frac{n}{N} = \frac{\text{Number of sample outcomes in } A}{\text{Number of sample outcomes in } S}.$$

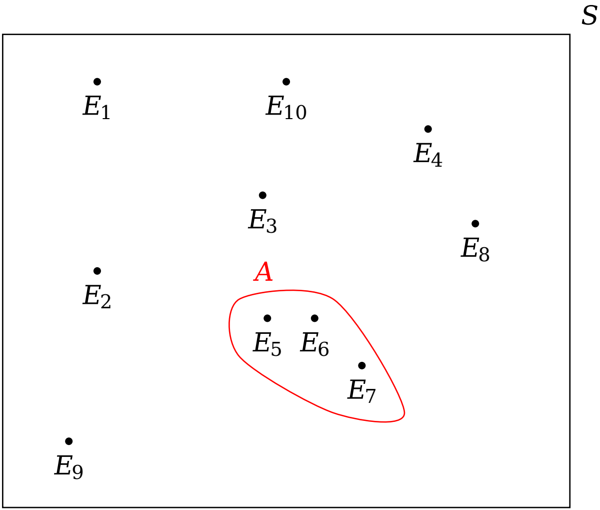
Specifying probabilities (cont.)

“**Equally likely outcomes**”

This is a strong (and subjective) assumption, but if it holds, then it leads to a straightforward calculation.

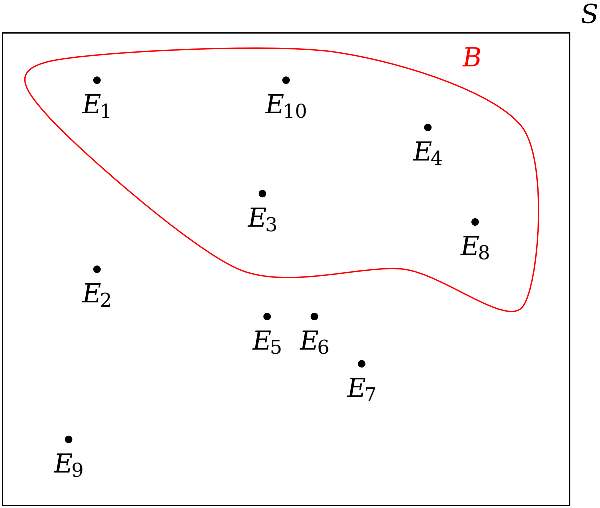
Suppose we have $N = 10$:

$$P(A) = \frac{3}{10}\colon\ 1_A = 5, 2_A = 6, 3_A = 7.$$



Specifying probabilities (cont.)

$$P(B) = \frac{5}{10} = \frac{1}{2}\colon\ 1_B = 1, 2_B = 3, 3_B = 4, 4_B = 8, 5_B = 10.$$



Specifying probabilities (cont.)

We usually apply the logic of "equally likely outcomes" in certain mechanical settings where we can appeal to some form of symmetry argument.

(a) **Coins:** It is common to assume the existence of a 'fair' coin, and an experiment involving a single flip. Here the sample space is

$$S = \{\text{Head}, \text{Tail}\}$$

with elementary events $E_1 = \{\text{Head}\}$ and $E_2 = \{\text{Tail}\}$, and it is usual to assume that

$$P(E_1) = P(E_2) = \frac{1}{2}.$$

Specifying probabilities (cont.)

(b) **Dice:** Consider a single roll of a 'fair' die, with the outcome being the upward face after the roll is complete. Here the sample space is

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

with elementary events $E_i = \{i\}$ for $i = 1, 2, 3, 4, 5, 6$, and it is usual to assume that

$$P(E_i) = \frac{1}{6}.$$

Let A be the event that the outcome of the roll is an even number. Then

$$A = E_2 \cup E_4 \cup E_6$$

and

$$P(A) = \frac{3}{6} = \frac{1}{2}.$$

Specifying probabilities (cont.)

(c) **Cards:** A standard deck of cards contains 52 cards, comprising four suits (Hearts, Clubs, Diamonds, Spades) with thirteen cards in each suit, with cards denominated

$$2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, King, Ace.$$

Thus each card has a suit and a denomination. An experiment involves selecting a card from the deck after it has been well shuffled.

There are 52 elementary outcomes, and

$$P(E_i) = \frac{1}{52} \quad i = 1, \ldots, 52.$$

If A is the event *Ace* is selected, then

$$P(A) = \frac{\text{Total number of } Aces}{\text{Total number of cards}} = \frac{4}{52} = \frac{1}{13}.$$

Specifying probabilities (cont.)

In a second experiment, five cards are to be selected from the deck *without replacement*. The elementary outcomes correspond to all sequences of five cards that could be obtained. All such sequences are equally likely.

Let A be the event that the five cards contain three cards of one denomination, and two cards of another denomination.

What is $P(A)$?

Need some rules to help in counting the elementary outcomes.

Specifying probabilities (cont.)

The concept of "equally likely outcomes" can be extended to the uncountable sample space case:

- pick a point from the interval $(0, 1)$ with each point equally likely; if A is the event

  "point picked is in the interval $(0, 0.25)$"

  then we can set $P(A) = 0.25$.

Specifying probabilities (cont.)

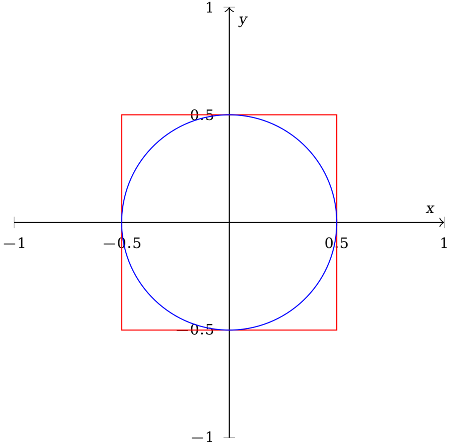
- a point is picked internal to the square centered at $(0, 0)$ with side length one, with each point equally likely. Let A be the event that

  "point lies in the circle centered at $(0, 0)$ with radius $\frac{1}{2}$"

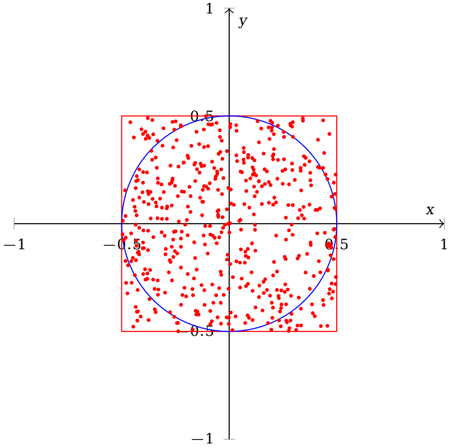
  Then

$$P(A) = \frac{\text{Area of circle}}{\text{Area of Square}} = \frac{\pi}{4} = 0.7853982.$$

Specifying probabilities (cont.)

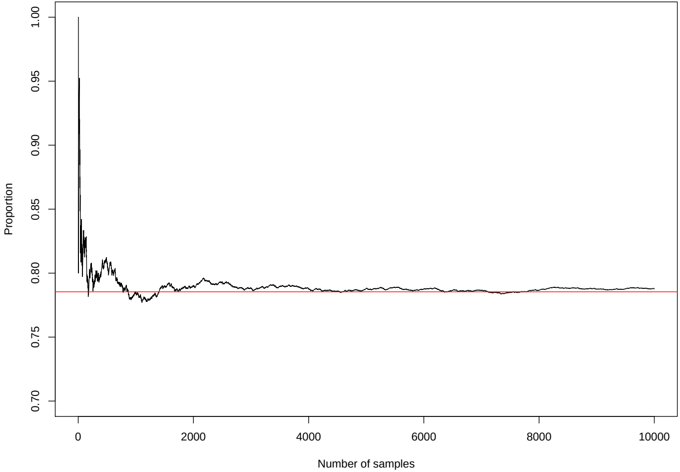


Specifying probabilities (cont.)



Specifying probabilities (cont.)

A simulation: 10000 points picked from the square.



Specifying probabilities (cont.)

Example (Thumbtack tossing)

- $S = \{\text{Up}, \text{Down}\}$.
- $A = \{\text{Up}\}$.

What is $P(A)$?

Specifying probabilities (cont.)

2. **Relative frequencies:** Suppose S is the sample space, and A is the event of interest. Consider an *infinite* sequence of repeats of the experiment under identical conditions.

   Then we may define $P(A)$ by considering the relative frequency with which event A occurs in the sequence.

Specifying probabilities (cont.)

- Consider a finite sequence of N repeat experiments.
- Let n be the number of times (out of N) that A occurs.
- Define

$$P(A) = \lim_{N \longrightarrow \infty} \frac{n}{N}.$$

The *frequentist* definition of probability; it generalizes the “equally likely outcomes” version of probability.

Specifying probabilities (cont.)

The frequentist definition would cover the thumbtack example:

Is it always possible to consider an infinite sequence of repeats ?

3. **Subjective assessment:** For a given experiment with sample space S, the probability of event A is
   - a numerical representation of your own personal

     degree of belief

     that the actual outcome lies in A;
   - you are *rational* and *coherent* (that is, internally consistent in your assessment).

   This generalizes the "equally likely outcomes" and "frequentist" versions of probability.

- Especially useful for ‘one-off’ experiments.
- Assessment of ‘odds’ on an event can be helpful:

$$\text{Odds}: \quad \frac{P(A)}{P(A')} = \frac{P(A)}{1 - P(A)}$$

eg Odds = 10:1, then $P(A) = 10/11$ etc.

Multiplication principle: A sequence of k operations, in which operation i can result in n_i possible outcomes, can result in

$$n_1 \times n_2 \times \cdots \times n_k$$

possible sequences of outcomes.

Rules for counting outcomes (cont.)

Example (Two dice)

Two dice are rolled: there are $6 \times 6 = 36$ possible outcomes.

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Rules for counting outcomes (cont.)

Example

We are to pick $k = 5$ numbers sequentially from the set

$$\{1, 2, 3, \ldots, 100\},$$

where every number is available on every pick. The number of ways of doing that is

$$100 \times 100 \times 100 \times 100 \times 100 = 100^5.$$

Rules for counting outcomes (cont.)

Example

We are to pick $k = 11$ players sequentially from a squad of 25 players. The number of ways of doing that is

$$25 \times 24 \times 23 \times \cdots \times 15.$$

as players cannot be picked twice.

Selection principles: When selecting repeatedly from a finite set $\{1, 2, \ldots, N\}$ we may select

- with replacement, that is,
  - each successive selection can be one of the original set, irrespective of previous selections,

  or

- without replacement, that is,
  - the set is depleted by each successive selection,

Rules for counting outcomes (cont.)

Ordering: When examining the result of a sequence of selections, it may be required that

- order is important, that is

  13456   is considered distinct from   54163

  or

- order is unimportant, that is

  13456   is considered identical to   54163

Rules for counting outcomes (cont.)

Example (Two dice)

Two dice are rolled: there are $(7 \times 6)/2 = 21$ possible unordered outcomes.

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	-	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	-	-	(3,3)	(3,4)	(3,5)	(3,6)
4	-	-	-	(4,4)	(4,5)	(4,6)
5	-	-	-	-	(5,5)	(5,6)
6	-	-	-	-	-	(6,6)

Distinguishable items: It may be that the objects being selected are

- distinguishable, that is, individually uniquely labelled.
  - eg lottery balls.
- indistinguishable, that is, labelled according to a type, but not labelled individually.
  - eg a bag containing 7 red balls, 2 green balls, and 3 yellow balls.

An ordered arrangement of r distinct objects is called a *permutation*.

The number of ways of ordering n distinct objects taken r at a time is denoted P_r^n, and by the multiplication rule we have that

$$P_r^n = n \times (n-1) \times (n-2) \times \cdots (n-r+2) \times (n-r+1) = \frac{n!}{(n-r)!}$$

Multinomial coefficients

The number of ways of partitioning n distinct objects into k disjoint subsets of sizes

$$n_1, n_2, \ldots, n_k$$

where

$$\sum_{i=1}^{k} n_i = n$$

is

$$N = \binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! \times n_2! \ldots \times n_k!}.$$

Multinomial coefficients (cont.)

To see this, consider first selecting the n objects in order; there are

$$P_n^n = n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$$

ways of doing this. Then designate

- objects selected 1 to n_1 as Subset 1,
- objects selected $n_1 + 1$ to $n_1 + n_2$ as Subset 2,
- objects selected $n_1 + n_2 + 1$ to $n_1 + n_2 + n_3$ as Subset 3,
- $\ldots$
- objects selected $n_1 + n_2 + \cdots + n_{k-1} + 1$ to $n_1 + n_2 + \cdots + n_k$ as Subset k.

Multinomial coefficients (cont.)

Then note that the specific ordered selection that achieves the partition is only one of several that yield the same partition; there are

- $n_1!$ ways of permuting Subset 1,
- $n_2!$ ways of permuting Subset 2,
- $n_3!$ ways of permuting Subset 3,
- $\ldots$
- $n_k!$ ways of permuting Subset k,

that yield the same partition.

Multinomial coefficients (cont.)

Therefore, we must have that

$$P_n^n = n! = N \times (n_1! \times n_2! \ldots \times n_k!)$$

and hence

$$N = \frac{n!}{n_1! \times n_2! \times \ldots \times n_k!}$$

Multinomial coefficients (cont.)

Example ($n = 8$, $k = 3$, $(n_1, n_2, n_3) = (4, 2, 2)$)

- One specific ordered selection:

$$4, 6, 2, 1, 8, 5, 3, 7$$

  – there are $8!$ ways of obtaining such a selection.
- Partition into subsets with (4,2,2) elements:

$$(4, 6, 2, 1), (8, 5), (3, 7)$$

Multinomial coefficients (cont.)

Example ($n = 8$, $k = 3$, $(n_1, n_2, n_3) = (4, 2, 2)$)

- Consider all within-subset permutations:

$$(4, 6, 2, 1), (8, 5), (3, 7)$$

$$(4, 6, 1, 2), (8, 5), (3, 7)$$

$$(4, 1, 2, 6), (8, 5), (3, 7)$$

$$\vdots$$

Multinomial coefficients (cont.)

Example ($n = 8$, $k = 3$, $(n_1, n_2, n_3) = (4, 2, 2)$)

- There are

$$\frac{8!}{4! \times 2! \times 2!} = \frac{40320}{24 \times 2 \times 2} = 420$$

possible distinct partitions;

$$(4, 6, 2, 1), (8, 5), (3, 7)$$

is regarded as identical to

$$(1, 2, 4, 6), (5, 8), (3, 7).$$

Multinomial coefficients (cont.)

Example (Department committees)

A University Department comprises 40 faculty members. Two committees of 12 faculty members, and two committees of 8 faculty members, are needed.

The number of distinct committee configurations that can be formed is

$$N = \binom{40}{12, 12, 8, 8} = \frac{40!}{12! \times 12! \times 8! \times 8!} = 1785474512.$$

Multinomial coefficients (cont.)

Note: Binomial coefficients

Recall the binomial theorem:

$$(a+b)^n = \sum_{j=0}^{n} \binom{n}{j} a^j b^{n-j}$$

where

$$\binom{n}{j} = \frac{n!}{j!(n-j)!} \equiv \binom{n}{j,\, n-j}$$

Multinomial coefficients (cont.)

Note: Binomial coefficients

Here we are solving a partitioning problem; partition n elements into one subset of size j and one subset of size $n - j$.

We can use this result and the multiplication principle to derive the general multinomial coefficient result.

Multinomial coefficients (cont.)

Note: Binomial coefficients

For a collection of n elements:

1. Partition into two subsets of n_1 and $n - n_1$ elements;
2. Take the $n - n_1$ elements and partition them into two subsets of n_2 and $n - n_1 - n_2$ elements;
3. Repeat until the final partition step; at this stage we have a set of $n - n_1 - \cdots - n_{k-2}$ elements, which we partition into a subset of n_{k-1} elements and a subset of

$$n_k = n - n_1 - \cdots - n_{k-2} - n_{k-1}$$

elements.

Multinomial coefficients (cont.)

Note: Binomial coefficients

By the multiplication rule, the number of ways this sequence of partitioning steps can be carried out is

$$\binom{n}{n_1} \times \binom{n-n_1}{n_2} \times \cdots \times \binom{n-n_1-\cdots-n_{k-2}}{n_{k-1}}$$

that is

$$\frac{n!}{n_1!(n-n_1)!} \times \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \times \cdots \times \frac{(n-n_1-\cdots-n_{k-2})!}{n_{k-1}!n_k!}$$

Multinomial coefficients (cont.)

Note: Binomial coefficients

Cancelling terms top and bottom in successive factors, we see that this equals

$$\frac{n!}{n_1 \times n_2! \times \ldots \times n_{k-1}! \times n_k!}$$

as required.

The number of *combinations* of n objects taken r at a time is the number of subsets of size r that can be formed.

This number is denoted

$$C_r^n = \binom{n}{r}$$

where

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

i.e. we "choose r from n".

Combinations (cont.)

We first consider sequential selection of r objects: the number of possible selections (without replacement) is

$$n \times (n-1) \times \ldots \times (n-r+1) = \frac{n!}{(n-r)!} = P_r^n$$

leaving $(n-r)$ objects unselected.

We then remember that the order of selected objects is not important in identifying a combination, so therefore we must have that

$$P_r^n = r! \times C_r^n.$$

as there are $r!$ equivalent combinations that yield the same permutation. The result follows.

Binary sequences

$$101000100101010$$

arise in many probability settings; if we take '1' to indicate inclusion and '0' to indicate exclusion, then we can identify combinations with binary sequences.

- the number of binary sequences of length n containing r 1s is
$$\binom{n}{r}$$

Combinatorial probability

The above results are used to compute probabilities in the case of equally likely outcomes.

- S: complete list of possible sequences of selections.
- A: sequences having property of interest.
- We have

$$P(A) = \frac{\text{Number of elements in } A}{\text{Number of elements in } S} = \frac{n_A}{n_S}$$

Example (Cards)

Five cards are selected without replacement from a standard deck. What is the probability they are all Hearts ?

- Number of elements in S:

$$n_S = \binom{52}{5}$$

- Number of elements in A:

$$n_A = \binom{13}{5}$$

Combinatorial probability (cont.)

Example (Cards)

Therefore

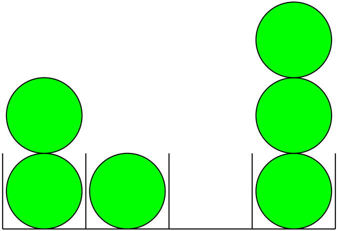
$$P(A) = \frac{\binom{13}{5}}{\binom{52}{5}} = 0.0004951981.$$

Other combinatorial problems

- Poker hands;
- Hypergeometric selection;
- Occupancy (or allocation) problems – allocate r objects to n boxes and identify
  - the occupancy pattern;
  - the occupancy of a specific box.

Other combinatorial problems (cont.)

Allocate $r = 6$ indistinguishable balls to $n = 4$ boxes: how many distinct allocation patterns are there ?



Conditional probability

We now consider probability assessments in the presence of

partial knowledge.

In the case of ‘ordinary’ probability, we have (by assumption) that the outcome **must** be an element in the set S, and proceed to assess the probability that the outcome is an element of the set $A \subseteq S$.

That is, the only ‘certain’ knowledge we have is that the outcome, s, is in S.

Conditional probability (cont.)

Now suppose we have the 'partial' knowledge that, in fact,

$$s \in B \subseteq S$$

for some B;

- that is, we know that event B occurs.

Conditional probability (cont.)

How does this change our probability assessment concerning the event A ?

- in light of the information that B occurs, what do we now think about the probability that A occurs also ?

Conditional probability (cont.)

First, we are considering the event that *both* A and B occur, that is

$$s \in A \cap B$$

so $P(A \cap B)$ must play a role in the calculation.

Secondly, with the knowledge that event B occurs, we restrict the parts of the sample space that should be considered; we are *certain* that the sample outcome must lie in B.

Conditional probability (cont.)

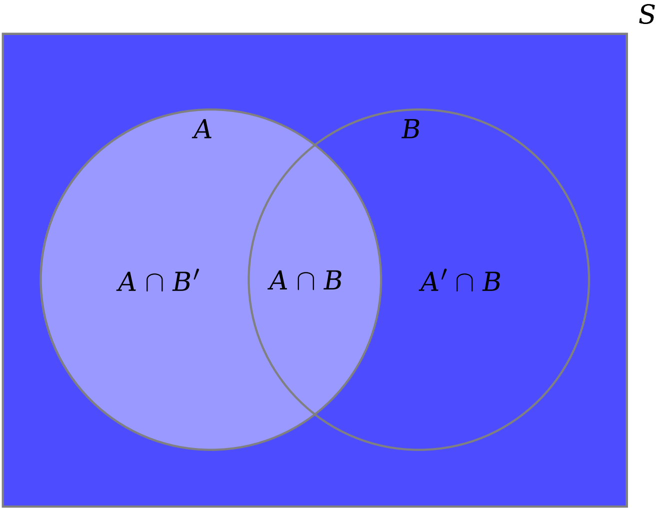
For two events A and B, the *conditional probability* of A *given* B is denoted $P(A|B)$, and is defined by

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

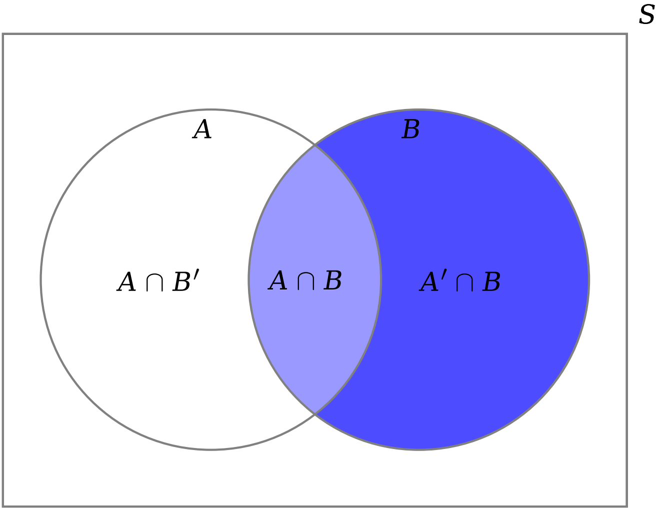
Note that we consider this *only* in cases where $P(B) > 0$.

Conditional probability (cont.)

Ordinary Probability: $P(A)$.



Conditional Probability: $P(A|B) = P(A \cap B)/P(B)$.



Example

Experiment: roll a fair die, record the score.

- $S = \{1, 2, 3, 4, 5, 6\}$, all outcomes equally likely.
- $A = \{1, 3, 5\}$ (score is odd).
- $B = \{4, 5, 6\}$ (score is more than 3).
- $A \cap B = \{5\}$.

$$P(A) = \frac{3}{6} = \frac{1}{2} \qquad P(B) = \frac{3}{6} = \frac{1}{2} \qquad P(A \cap B) = \frac{1}{6}$$

so

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

Conditional probability (cont.)

Example

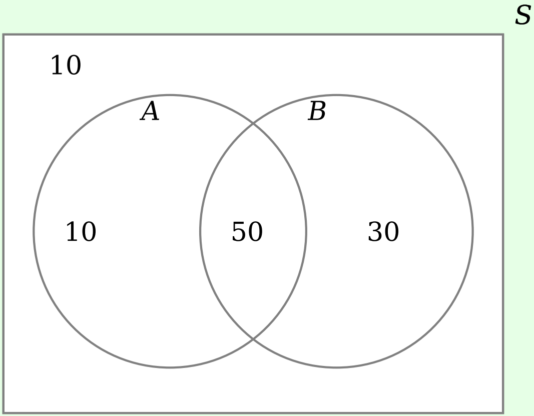
In a class of 100 students:

- 60 are from Faculty of Science, 40 are from Faculty of Arts.
- 80 are in a Major program, 20 are in another program.
- 50 of the Science students are in a Major program.

Conditional probability (cont.)

Example

A – Faculty of Science; B – Major program.



Conditional probability (cont.)

Example

A student is selected from the class, with all students equally likely to be selected.

- What is the probability that the selected student is from Science, $P(A)$?:

$$P(A) = \frac{60}{100} = \frac{3}{5}$$

Conditional probability (cont.)

Example

- What is the probability that the selected student is from Science and in a Major program, $P(A \cap B)$?:

$$P(A \cap B) = \frac{50}{100} = \frac{1}{2}$$

Conditional probability (cont.)

Example

- If the selected student is **known** to be in a Major program, what is the probability that the student is from Science, $P(A|B)$?:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{50/100}{80/100} = \frac{50}{80} = \frac{5}{8}.$$

Conditional probability (cont.)

Note: some consequences

- If $B \equiv S$, then

$$P(A|S) = \frac{P(A \cap S)}{P(S)} = \frac{P(A)}{1} = P(A).$$

- If $B \equiv A$, then

$$P(A|A) = \frac{P(A \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1.$$

- We have

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \leq \frac{P(B)}{P(B)} = 1.$$

Conditional probability (cont.)

Direct from the definition, we have that

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

(recall $P(B) > 0$).

Conditional probability (cont.)

Note

It is important to understand the distinction between

$$P(A \cap B) \qquad \text{and} \qquad P(A|B).$$

- $P(A \cap B)$ records the chance of A **and** B occurring relative to S. A and B are treated symmetrically in the calculation.
- $P(A|B)$ records the chance of A **and** B occurring relative to B. A and B are **not** treated symmetrically; from the definition, we see that in general

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

Note

From the above calculation, we see that we could have

$$P(A|B) \leq P(A)$$

or

$$P(A|B) \geq P(A)$$

or

$$P(A|B) = P(A).$$

That is, certain knowledge that B occurs could decrease, increase or leave unchanged the probability that A occurs.

Conditional probability (cont.)

Example

A box of 500 light bulbs is purchased from each of two factories (labelled 1 and 2).

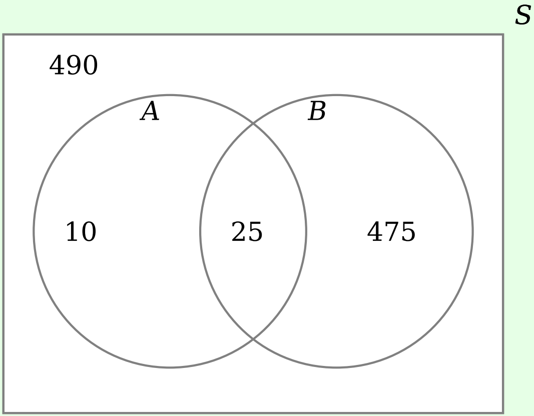
- In the box from Factory 1, there are 25 defective bulbs.
- In the box from Factory 2, there are 10 defective bulbs.

The bulbs are unpacked and placed in storage.

Conditional probability (cont.)

Example

A – Defective; B – Factory 1.



Conditional probability (cont.)

Example

At the time of the next installation, a bulb is selected from the store.

- What is the probability that the bulb is defective, $P(A)$?:

$$P(A) = \frac{35}{1000} = 0.035.$$

Conditional probability (cont.)

Example

- What is the probability that the bulb came from Factory 1, $P(B)$?:

$$P(B) = \frac{500}{1000} = 0.5.$$

Conditional probability (cont.)

Example

- If the selected bulb **came from Factory 1**, what is the probability it is **defective** $P(A|B)$?:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{25/1000}{500/1000} = \frac{25}{500} = 0.05.$$

Conditional probability (cont.)

Example

- If the selected bulb is **defective**, what is the probability it came from Factory 1, $P(B|A)$?:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{25/1000}{35/1000} = \frac{25}{35} = \frac{5}{7}.$$

Conditional probability (cont.)

Example

Experiment: We measure the failure time of an electrical component.

- $S = R^+$.
- $A = [10, \infty)$ (fails after 10 hours or more).
- $B = [5, \infty)$ (fails after 5 hours or more).

Suppose we assess

$$P(A) = 0.25 \qquad P(B) = 0.4.$$

Now here $A \subset B \subset S$, so therefore $P(A \cap B) \equiv P(A)$, and

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{0.25}{0.4} = 0.625.$$

The conditional probability function satisfies the Probability Axioms: suppose B is an event in S such that $P(B) > 0$.

(I) Non-negativity:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0.$$

Conditional probability (cont.)

(II) For $P(S|B)$ we have

$$P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Conditional probability (cont.)

(III) Countable additivity: if $A_1, A_2, \ldots$ form a (countable) sequence of events such that $A_j \cap A_k = \emptyset$ for all $j \neq k$.

Conditional probability (cont.)

Then

$$
\begin{aligned}
P\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) &= \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right)}{P(B)} \\
&= \frac{P\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{P(B)} \\
&= \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} \\
&= \sum_{i=1}^{\infty} P(A_i | B).
\end{aligned}
$$

Conditional probability (cont.)

$$
\begin{aligned}
P\left(A_1 \cup A_2 \middle| B\right) &= \frac{P\left((A_1 \cup A_2) \cap B\right)}{P(B)} \\
&= \frac{P\left((A_1 \cap B) \cup (A_2 \cap B)\right)}{P(B)} \\
&= \frac{P(A_1 \cap B) + P(A_1 \cap B)}{P(B)} \\
&= P(A_1|B) + P(A_2|B).
\end{aligned}
$$

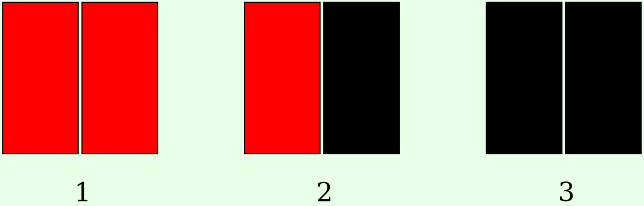
as

$$A_1, A_2 \text{ disjoint} \quad \Longrightarrow \quad (A_1 \cap B), (A_2 \cap B) \text{ disjoint.}$$

Conditional probability (cont.)

Example (Three cards)

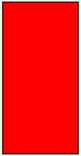
Three two sided cards:



One of the three cards is selected, with all cards being equally likely to be chosen.

Example (Three cards)

One side is displayed:



What is the probability that the other side of this card is red ?

Conditional probability (cont.)

Example (Three cards)

Let

- $S = \{R_1, R_2, R, B, B_1, B_2\}$ be the possible exposed sides. The outcomes in S are equally likely.
- A – Card 1 is selected.

  $$A = \{R_1, R_2\}$$

  and $P(A) = 2/6$.
- B – A red side is exposed.

  $$B = \{R_1, R_2, R\}$$

  and $P(B) = 3/6$.

Conditional probability (cont.)

Example (Three cards)

Then

$$A \cap B = \{R_1, R_2\} \equiv A$$

and hence

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{2/6}{3/6} = \frac{2}{3}.$$

Example (Drug testing)

A drug testing authority tests 100000 athletes to assess whether they are using performance enhancing drugs (PEDs). The drug test is not perfect as it produces

- *False positives:* declares an athlete to be using PEDs, when in reality they are not.
- *False negatives:* declares an athlete not to be using PEDs, when in reality they are.

Conditional probability (cont.)

Example (Drug testing)

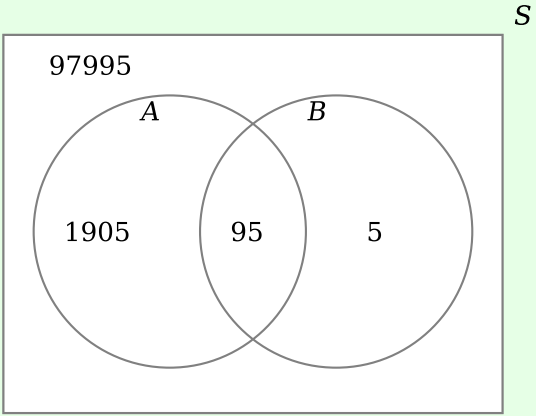
Suppose that after detailed investigation it is discovered that

- 2000 athletes gave positive test results.
- 100 athletes were using PEDs.
- 95 athletes who were using PEDs tested positive.

Conditional probability (cont.)

Example (Drug testing)

A – Drug test is positive; B – Athlete is using PED.



Conditional probability (cont.)

Example (Drug testing)

However, in the original analysis, an athlete is selected at random from the 100000, and is observed to test positive.

What is the probability that they actually were using PEDs ?

Conditional probability (cont.)

Example (Drug testing)

The required conditional probability is

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{95}{2000} = 0.0475.$$

Note that is very different from

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{95}{100} = 0.95.$$

Two events A and B in sample space S are *independent* if

$$P(A|B) = P(A);$$

equivalently, they are independent if

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

Note that if $P(A|B) = P(A)$, then

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

Independence (cont.)

Note

Independence is not a property of the events A and B themselves, it is a property of the probabilities assigned to them.

Independence (cont.)

Note

Independence is **not** the same as mutual exclusivity.

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

$$\text{Mutually exclusive: } P(A \cap B) = 0.$$

Example

A fair coin is tossed twice, with the outcomes of the two tosses independent. We have

$$S = \{HH, HT, TH, TT\}$$

with all four outcomes equally likely.

(a) If the first toss result is H, what is the probability that the second toss is also H ?

(b) If one result is H, what is the probability that the other is H ?

Independence (cont.)

Example

Let

- $A = \{HH\}$ (both heads).
- $B = \{HH, HT, TH\}$ (at least one head).
- $C = \{HH, HT\}$ (first result is head).

(a) We want $P(A|C)$:

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(A)}{P(C)} = \frac{1/4}{2/4} = \frac{1}{2}.$$

(b) We want $P(A|B)$:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Independence for multiple events

Independence as defined above is a statement concerning two events.

What if we have more than two events ?

- Events A_1, A_2, A_3 in sample space S.
- We can consider independence pairwise:

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$

$$P(A_1 \cap A_3) = P(A_1)P(A_3)$$

$$P(A_2 \cap A_3) = P(A_2)P(A_3)$$

Independence for multiple events (cont.)

- What about

$$P(A_1 \cap A_2 \cap A_3);$$

  Can we deduce

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)?$$

- In general, no.

Independence for multiple events (cont.)

Example (Two dice)

Suppose we roll two dice with outcomes independent.

- A_1 – first roll outcome is odd.
- A_2 – second roll outcome is odd.
- A_3 – total score is odd.

We have

$$P(A_1) = P(A_2) = P(A_1|A_3) = P(A_2|A_3) = \frac{1}{2}$$

– can compute this by identifying all 36 pairs of scores, which are all equally likely, and counting the relevant sample outcomes.

Thus A_1, A_2 and A_3 are *pairwise* independent.

Independence for multiple events (cont.)

Example (Two dice)

However,

$$P(A_1 \cap A_2 \cap A_3) = 0$$

and

$$P(A_1 | A_2 \cap A_3) = 0$$

etc.

Independence for multiple events (cont.)

Mutual Independence: Events $A_1, A_2, \ldots, A_K$ are *mutually independent* if

$$P\left(\bigcap_{k\,\in\,\mathcal{I}} A_k\right) = \prod_{k\,\in\,\mathcal{I}} P(A_k)$$

for **all subsets** $\mathcal{I}$ of $\{1, 2, \ldots, K\}$.

For example, if $K = 3$, we require that

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$

$$P(A_1 \cap A_3) = P(A_1)P(A_3)$$

$$P(A_2 \cap A_3) = P(A_2)P(A_3)$$

$$\underline{\text{and}}\ P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3).$$

Conditional independence: Consider events A_1, A_2 and B in sample space S, with $P(B) > 0$. Then A_1 and A_2 are *conditionally independent* given B if

$$P(A_1 | A_2 \cap B) = P(A_1 | B)$$

or equivalently

$$P(A_1 \cap A_2 | B) = P(A_1 | B) P(A_2 | B).$$

General Multiplication Rule

For events $A_1, A_2, \ldots, A_K$, we have the general result that

$$P(A_1 \cap A_2 \cap \cdots \cap A_K) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\ldots$$
$$\ldots P(A_K|A_1 \cap A_2 \cap \cdots \cap A_{K-1})$$

This follows by the recursive calculation

$$\begin{aligned}
P(A_1 \cap A_2 \cap \cdots \cap A_K) &= P(A_1)P(A_2 \cap \cdots \cap A_K|A_1) \\
&= P(A_1)P(A_2|A_1) \\
&\qquad P(A_3 \cap \cdots \cap A_K|A_1 \cap A_2) \\
&= \ldots
\end{aligned}$$

Also known as the *Chain Rule* for probabilities.

General Multiplication Rule (cont.)

If the events are mutually independent, then we have that

$$P(A_1 \cap A_2 \cap \cdots \cap A_K) = \prod_{k=1}^{K} P(A_k)$$

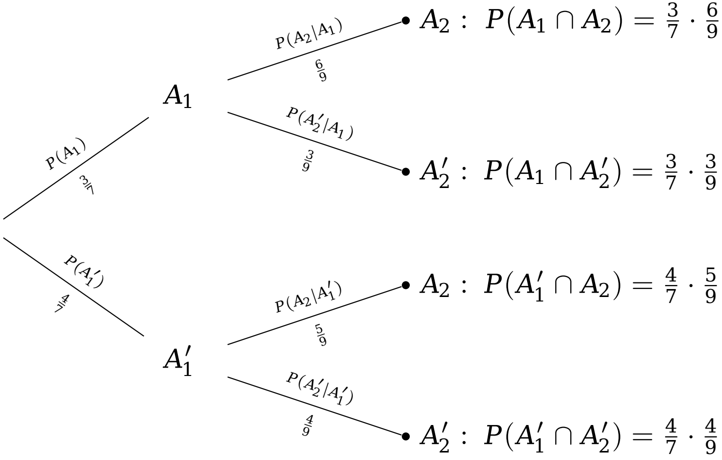
Probability Trees

Probability trees are simple ways to display joint probabilities of multiple events. They comprise

- *Junctions*: correspond to the multiple events
- *Branches*: correspond to the sequence of (conditional) choices of events, given the previous choices.

We multiply along the branches to get the joint probabilities.

Probability Trees (cont.)



Probability Trees (cont.)

Note that each junction has two possible branches coming from it, corresponding to

$$A_k \quad \text{and} \quad A_k'$$

respectively.

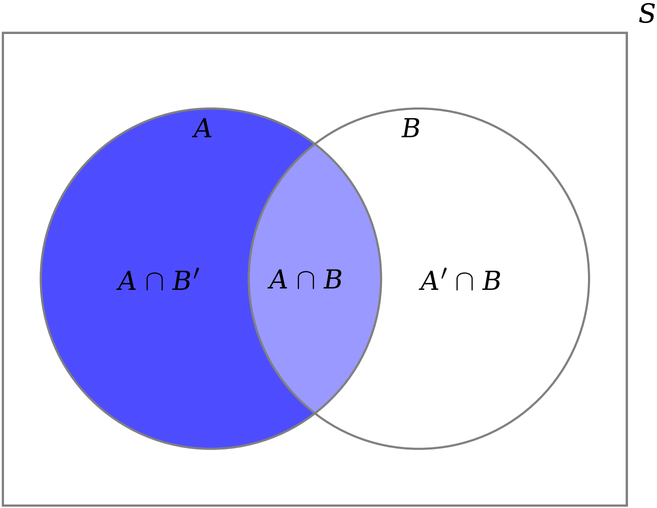
Such a tree extends to as many events as we need.

The Theorem of Total Probability

For two events A and B in sample space S, we have the partition of A as

$$A = (A \cap B) \cup (A \cap B').$$

The Theorem of Total Probability (cont.)



The Theorem of Total Probability (cont.)

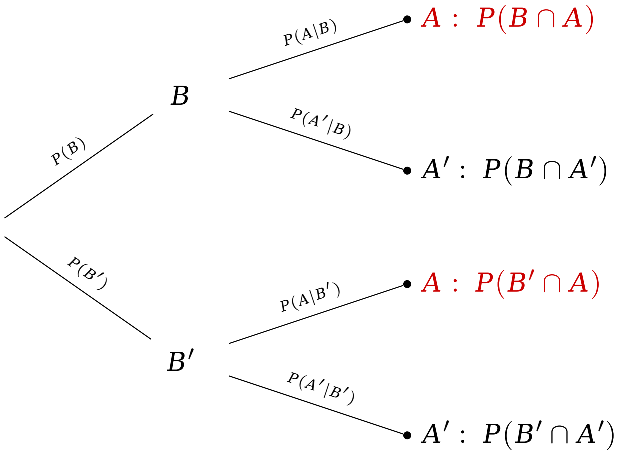
Therefore

$$P(A) = P(A \cap B) + P(A \cap B')$$

and using the multiplication rule, we may rewrite this as

$$P(A) = P(A|B)P(B) + P(A|B')P(B').$$

The Theorem of Total Probability (cont.)



The Theorem of Total Probability (cont.)

That is, there are two ways to get to A:

- via B;
- via B'.

To compute $P(A)$ we add up all the probabilities on paths that end up at A.

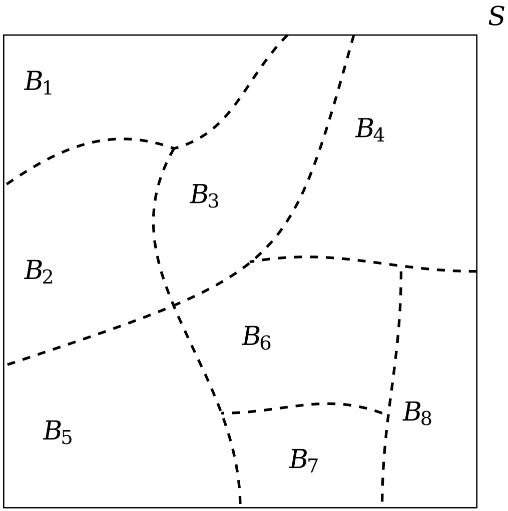
Note that B and B' together form a partition of S.

Now suppose we have a partition into n subsets

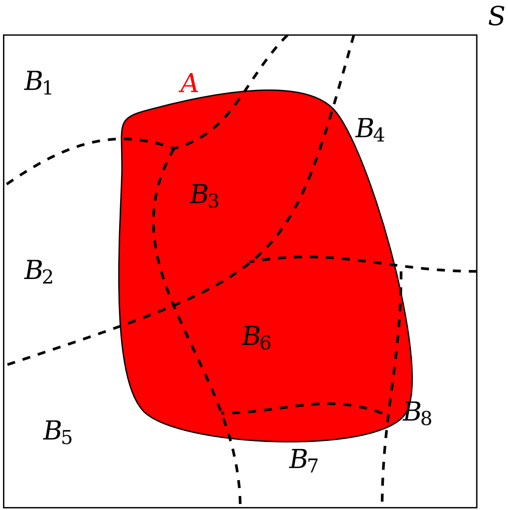
$$B_1, B_2, \ldots, B_n.$$

Again, these events also partition A.

The Theorem of Total Probability (cont.)



The Theorem of Total Probability (cont.)



The Theorem of Total Probability (cont.)

That is, we have that

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n)$$

and

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

The Theorem of Total Probability (cont.)

Using the definition of conditional probability, we therefore have

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

that is

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

This result is known as the *Theorem of Total Probability*.

The Theorem of Total Probability (cont.)

Notes

- The formula assumes that $P(B_i) > 0$ for $i = 1, \ldots, n$.
- It might be that

  $$P(A \cap B_i) = P(A|B_i) = 0.$$

  for some i.
- This formula is a mathematical encapsulation of the probability tree when it is used to compute $P(A)$.

The Theorem of Total Probability (cont.)

Example (Three bags)

Suppose that an experiment involves selecting a ball from one of three bags. Let

- Bag 1: 4 red and 4 white balls.
- Bag 2: 1 red and 10 white balls.
- Bag 3: 7 red and 11 white balls.

A bag is selected (with all bags equally likely), and then a ball is selected from that bag (with all balls equally likely).

What is the probability that the ball selected is red ?

Example (Three bags)

Let

- S: all possible selections of balls.
- B_1: bag 1 selected.
- B_2: bag 2 selected.
- B_2: bag 3 selected.
- A: ball selected is red.

The Theorem of Total Probability (cont.)

Example (Three bags)

We have that

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) \\ &= \frac{4}{8} \times \frac{1}{3} + \frac{1}{11} \times \frac{1}{3} + \frac{7}{18} \times \frac{1}{3} \\ &= \frac{97}{297} \end{aligned}$$

The Theorem of Total Probability (cont.)

Example (Three bags)

Note that this is **not** equal to

$$\frac{\text{Total number of red balls}}{\text{Total number of balls}}$$

that is

$$\frac{4+1+7}{8+11+18} = \frac{12}{37}$$

as the red balls are not equally likely to be selected – this is only true conditional on the bag selection.

Bayes Theorem

The second great theorem of probability is *Bayes Theorem*.

- Named after Reverend Thomas Bayes (1701 – 1761)

  `https://en.wikipedia.org/wiki/Thomas_Bayes`

- Could be written *Bayes's Theorem*.
- It is not really a theorem.

Bayes Theorem: For two events A and B in sample space S, with $P(A) > 0$ and $P(B) > 0$,

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

If $0 < P(B) < 1$, we may write by the Theorem of Total Probability.

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

Bayes Theorem (cont.)

"Proof": By the definition of conditional probability

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

Then as $P(A) > 0$ and $P(B) > 0$ we can cross multiply and write

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

If $0 < P(B) < 1$, then we can legitimately write

$$P(A) = P(A|B)P(B) + P(A|B')P(B')$$

and substitute this expression in the denominator.

Bayes Theorem (cont.)

General version: Suppose that $B_1, B_2, \ldots, B_n$ form a partition of S, with $P(B_j) > 0$ for $j = 1, \ldots, n$. Suppose that A is an event in S with $P(A) > 0$.

Then for $i = 1, \ldots, n$

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}.$$

Bayes Theorem (cont.)

Probability tree interpretation: *Given* that we end up at A, what is the probability that we got there via branch B_i ?

Bayes Theorem (cont.)

Example (Three bags)

Suppose that an experiment involves selecting a ball from one of three bags. Let

- Bag 1: 4 red and 4 white balls.
- Bag 2: 1 red and 10 white balls.
- Bag 3: 7 red and 11 white balls.

A bag is selected (with all bags equally likely), and then a ball is selected from that bag (with all balls equally likely).

The ball selected is red. What is the probability it came from Bag 2 ?

Bayes Theorem (cont.)

Example (Three bags)

Let

- S: all possible selections of balls.
- B_1: bag 1 selected.
- B_2: bag 2 selected.
- B_2: bag 3 selected.
- A: ball selected is red.

Bayes Theorem (cont.)

Example (Three bags)

Recall that

$$
\begin{aligned}
P(A) &= P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) \\
&= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) \\
&= \frac{4}{8} \times \frac{1}{3} + \frac{1}{11} \times \frac{1}{3} + \frac{7}{18} \times \frac{1}{3} \\
&= \frac{97}{297}
\end{aligned}
$$

Example (Three bags)

Then

$$
\begin{aligned}
P(B_2|A) &= \frac{P(A|B_2)P(B_2)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)} \\
&= \frac{\frac{1}{11} \times \frac{1}{3}}{\frac{4}{8} \times \frac{1}{3} + \frac{1}{11} \times \frac{1}{3} + \frac{7}{18} \times \frac{1}{3}} \\
&= \frac{1/33}{97/297} \\
&= \frac{9}{97} \doteq 0.09278.
\end{aligned}
$$

Bayes Theorem (cont.)

Example (Three bags)

We can also compute that

$$P(B_1|A) = \frac{4/24}{97/297} = \frac{297}{582} \doteq 0.51031$$

and

$$P(B_3|A) = \frac{21/54}{97/297} = \frac{1746}{2079} \doteq 0.39691.$$

Note

We have that

$$P(B_1|A) + P(B_2|A) + P(B_3|A) = 1$$

but note that

$$P(A|B_1) + P(A|B_2) + P(A|B_3) \neq 1.$$

In the first formula, we are conditioning on A everywhere; in the second, we have different conditioning sets.

Using Bayes Theorem

Bayes theorem is often used to make probability statements concerning an event B that has not been observed, given an event A that has been observed.

Example (Medical screening)

A health authority tests individuals from a population to assess whether they are sufferers from some disease. The screening test is not perfect as it produces

- *False positives:* declares someone to be a sufferer, when in reality they are a non-sufferer.
- *False negatives:* declares someone to be a non-sufferer, when in reality they are a sufferer.

Using Bayes Theorem (cont.)

Example (Medical screening)

Suppose that we denote by

- A – Screening test is positive;
- B – Person is actually a sufferer.

Using Bayes Theorem (cont.)

Example (Medical screening)

Suppose that

- $P(B) = p$;
- $P(A|B) = 1 - \alpha$ (true positive rate);
- $P(A|B') = \beta$ (false positive rate).

for probabilities p, α, β.

Using Bayes Theorem (cont.)

Example (Medical screening)

We have for the rate of positive tests

$$P(A) = P(A|B)P(B) + P(A|B')P(B') = (1-\alpha)p + \beta(1-p)$$

by the Theorem of Total Probability, and by Bayes theorem

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

$$= \frac{(1-\alpha)p}{(1-\alpha)p + \beta(1-p)}.$$

Example (Medical screening)

Similarly, for the false negative rate, we have

$$P(B|A') = \frac{P(A'|B)P(B)}{P(A'|B)P(B) + P(A'|B')P(B')}$$
$$= \frac{\alpha p}{\alpha p + (1-\beta)(1-p)}.$$

Using Bayes Theorem (cont.)

Example (Medical screening)

We have

$$P(A|B) = (1 - \alpha) \qquad P(B|A) = (1 - \alpha)\frac{p}{(1 - \alpha)p + \beta(1 - p)}$$

and these two values are potentially very different.

Using Bayes Theorem (cont.)

Example (Medical screening)

That is, in general,

$$P(\text{“Spots”}|\text{“Measles”}) \neq P(\text{“Measles”}|\text{“Spots”})$$

Using Bayes Theorem (cont.)

Note

This phenomenon is sometimes known as the *Prosecutor's Fallacy*:

$$P(\text{“Evidence”}|\text{“Guilt”}) \neq P(\text{“Guilt”}|\text{“Evidence”})$$

Bayes Theorem and Odds

Recall that the odds on event B is defined by

$$\frac{P(B)}{P(B')} = \frac{P(B)}{1 - P(B)}$$

The *conditional odds* given event A (with $P(A) > 0$) is defined by

$$\frac{P(B|A)}{P(B'|A)} = \frac{P(B|A)}{1 - P(B|A)}$$

Bayes Theorem and Odds (cont.)

We have that

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

that is, in light of knowledge that B occurs, the odds change by a factor

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

Bayes Theorem with multiple events

If A_1 and A_2 are two events in a sample space S so that

$$P(A_1 \cap A_2) > 0$$

and there is a partition of S via $B_1, \ldots, B_n$, with $P(B_i) > 0$, then

$$P(B_i | A_1 \cap A_2) = \frac{P(A_1 \cap A_2 | B_i) P(B_i)}{\sum_{j=1}^{n} P(A_1 \cap A_2 | B_j) P(B_j)}$$

- we use the previous version of Bayes Theorem with

$$A \equiv A_1 \cap A_2.$$

- this extends to conditioning events $A_1, A_2, \ldots, A_n$.

Bayes Theorem with multiple events (cont.)

Note that we have by earlier results

$$P(A_1 \cap A_2 | B_i) = P(A_1 | B_i) P(A_2 | A_1 \cap B_i)$$

that is, the chain rule applies to conditional probabilities also.