PROBABILITY (Chapter 2)

The notion of probability

What is "probability"?
In one sense it is a measure of belief regarding the occurrence of events.

$p > 0.5$ "more likely than not"
almost certain

Probability scale:

precise:	0	0.25	0.5	0.75	.99	1
		1/4, 25%	1/2, 50%			100%
vague:	impossible	'improbable'	as likely as not	'probable'		certain

p

Eg: A fair coin is tossed once. Then the probability of a head coming up is 1/2.

The last statement is actually an expression of the belief that if the coin were to be tossed many times (eg n = 1000) then H's would come up about half the time (eg y = 503). (We believe that as n increases indefinitely, the ratio y/n tends to 0.5 exactly. This is an example of the *law of large numbers*, which we'll look at in Ch 9.)

LLN

Thus pr is also a concept which involves the notion of (long-run) *relative frequency*.

We wish to be able to find pr's such as that of getting 2 H's on 5 tosses of a coin. In order to do this, we need a theory of probability.

The theory of probability

> In the most commonly accepted theory of probability:
> 1. *events* are expressed as *sets*
> 2. *probabilities* of events are expressed as *functions* of the corresponding sets.

Review of basic set theory

set = any collection of objects (elements)

eg: $A = \{0,1\}$ (finite)

$B = \{0,1,2,...\}$ (countably infinite) can be expressed as $\{e_1, e_2, ...\}$

$C = (0,1)$ (uncountably infinite).

(An *infinite* set is one which has an infinite number of elements.
A *countably infinite* set is an infinite set whose elements can be listed $e_1, e_2, e_3, \ldots$)

$\varnothing = \{\}$ = empty (or null) set = set with no elements.

S = universal set = set of all objects under consideration.

$\bar{A}$ = complement of A = set of all elements of S that are not in A.

($\bar{A} = \{e : e \in S, e \notin A\}$.) if.

NB: $\bar{A}$ may also be written A', $\sim A$, A^c or simply "not A".

A is a subset of B (written $A \subseteq B$ or $A \subset B$) if all elements of A are also in B.

(If $e \in A \Rightarrow e \in B$ then $A \subseteq B$.)

$A = B$ if $A \subseteq B$ and $B \subseteq A$ (equality).

$A \cap B$ = intersection of A and B = set of all elements common to A and B.

($A \cap B = \{e : e \in A, e \in B\}$, where the comma means "and".

So we may also write "A and B".)

AB is short for $A \cap B$.

$A \cup B$ = union of A and B = set of all elements in A or B (or both).

($A \cup B = \{e : e \in A$ or $e \in B$ or $e \in AB\}$. We may also write "A or B".)

$A - B$ = A minus B = set of all elements in A which are not in B.

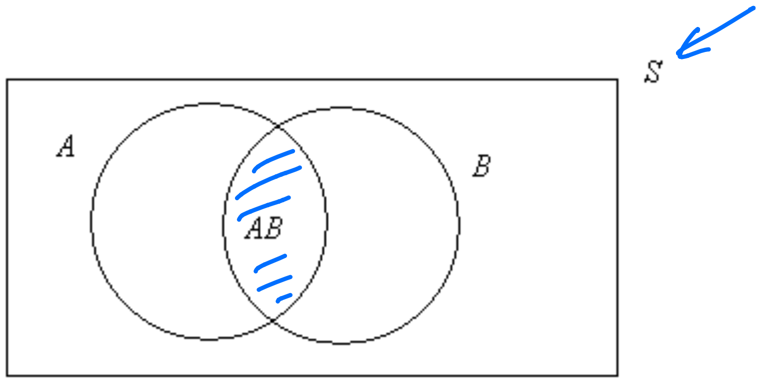
($A - B = A\bar{B} = \{e : e \in A, e \notin B\}$.)

NB: We will use the symbol $\subseteq$ rather than $\subset$ to indicate subsetting. The latter is potentially confusing because it looks like C. Also, it indicates 'proper' subsetting in some books, wherein $A \subset B$ if $A \subseteq B$ and $B - A \neq \varnothing$.

Two sets are disjoint (or mutually exclusive) if they have no elements in common. ME
(Thus A and B are disjoint if $AB = \varnothing$.)

Several sets are disjoint if no two of them have any elements in common.
(Thus, A, B and C are disjoint if $AB = AC = BC = \varnothing$.)

Venn diagrams are a useful tool when dealing with sets. For example:



Example 1 Suppose that: $S = \{1,2,3,4,5,6\}$
$A = \{2,6\}$, $B = \{1,2,3,6\}$, $C = \{3,5,6\}$.

$= S - A$

Then: $\overline{A} = \{1,3,4,5\}$, $A \subseteq B$, $A \not\subseteq C$,

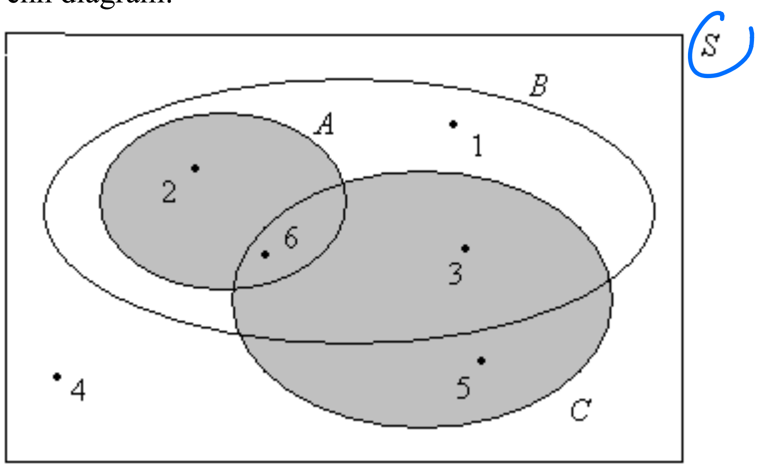
$AC = \{6\}$ (a singleton set, ie one that has only a single element)

$A \cup C = \{2,3,5,6\}$, $C - A = C\overline{A} = \{3,5\}$

A and C are not disjoint (since $AC \neq \varnothing$).

Therefore A, B and C are not disjoint .

Venn diagram:



$A \cup C$ Hence $\overline{A \cup C} = \{1,4\}$.

Q1: Find $D = A - \overline{BC}$
A1: $\{1,2,3,4,5\}$
Q2: Find $E = \overline{\overline{B} \cup C}$
A2: $\{1,2\}$

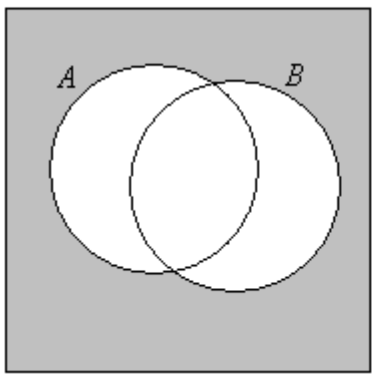
$1+2=2+1$

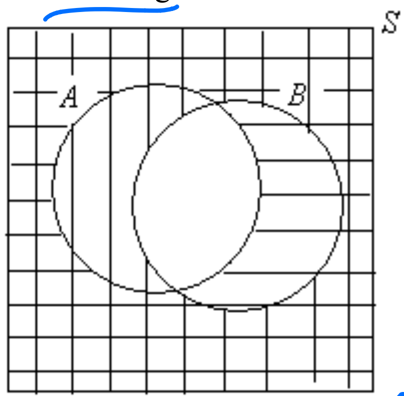
Algebra of sets

1.	$A\cup B = B\cup A$ $AB = BA$	(commutative laws)
2.	$A\cup(B\cup C) = (A\cup B)\cup C = A\cup B\cup C$ $A(BC) = (AB)C = ABC$	(associative laws)
3.	$A\cup(BC) = (A\cup B)(A\cup C)$ $A(B\cup C) = (AB)\cup(AC)$	(distributive laws)
4.	$\overline{A\cup B} = \bar{A}\bar{B}$ $= \bar{A}\cap\bar{B}$ $\overline{AB} = \bar{A}\cup\bar{B}$	(De Morgan's laws)

5. $A\cup\varnothing = A$, $A\cup A = A$, $AA = A$, $\bar{S} = \varnothing$, etc. $\bar{\varnothing} = S$ (basic identities)

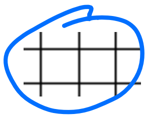
Illustration of De Morgan's 1st law via Venn diagrams:





white $A\cup B$

$\bar{A}$ $\bar{B}$

 $\overline{A\cup B}$ <--- same --->  $\bar{A}\bar{B}$

Note that the above is not a proper proof of De Morgan's 1st law.

The following is a proper proof (for interest only):

$e \in \overline{A \cup B} \Rightarrow e \notin A \cup B \Rightarrow e \notin A$ and $e \notin B$

taste of pure maths

$\Rightarrow e \in \bar{A}$ and $e \in \bar{B}$

$\Rightarrow e \in \bar{A}\bar{B}$.

Therefore $\overline{A \cup B} \subseteq \bar{A}\bar{B}$. (1)

Similarly, $e \in \bar{A}\bar{B} \Rightarrow e \in \overline{A \cup B}$ (the above argument can be reversed).

Therefore $\bar{A}\bar{B} \subseteq \overline{A \cup B}$. (2)

It follows from (1) and (2) that $\overline{A \cup B} = \bar{A}\bar{B}$.

The 2nd of De Morgan's laws can be proved by applying the 1st law as follows:

$\overline{\bar{A} \cup \bar{B}} = \bar{\bar{A}}\bar{\bar{B}} = AB$. $\therefore \bar{A} \cup \bar{B} = \overline{AB}$. QED

$\overline{A \cup B} = \bar{A}\bar{B}$

We will next show how events can be expressed as sets.

Some definitions relating to statistical experiments

experiment = process by which an observation is made:

this consists of 2 parts: what is *done* and what is *observed*

sample point = possible distinct observation (outcome) for the experiment

sample space = set of all sample points; we denote this by S

simple event = singleton set which contains a sample point

event = any subset of S

Example 2 Consider the experiment of rolling a die (what is *done*) and observing the number that comes up (what is *observed*).

For this experiment, the sample points are 1, 2, 3, 4, 5, 6.

The sample space is $S = \{1, 2, 3, 4, 5, 6\}$.

The simple events are $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5\}$, $\{6\}$.

Examples of events are:

$A = \{2, 4, 6\}$ = "An even number comes up"

$B = \{5\}$ = "5 comes up"

S = "Some number comes up"

$\varnothing$ = "No number comes up"

$C = A \cup B = \{2,4,5,6\}$ = "2,4,5 or 6 comes up"

= "1 or 3 does not come up", etc.

(i)

We now wish to express probabilities of events as functions of corresponding sets.

(ii)

Probability functions and the three axioms of probability

For a given experiment and associated sample space S, a *probability function P* is any real-valued function whose domain is the set of subsets of S (ie, $P:\{A:A\subseteq S\}\rightarrow\Re$) and which satisfies three conditions:

Axiom 1 $P(A)\geq 0$ for all $A\subseteq S$ (pr's can't be negative)

Axiom 2 $P(S)=1$ (*something* must happen)

Axiom 3 Suppose $A_1, A_2, \ldots$ is an infinite sequence of disjoint events. Then $P(A_1\cup A_2\cup\ldots)=P(A_1)+P(A_2)+\ldots$

These three conditions are known as *the three axioms of probability*. They do not completely specify P, but merely ensure that P is 'sensible'. It remains for P to be precisely defined in any given situation. Typically, P is defined by assigning 'reasonable' probabilities to each of the sample points (or simple events) in S.

Example 3 If the die in Example 2 is fair, then all of the possible outcomes 1, 2, 3, 4, 5, 6 are equally likely.
So it is reasonable to assign probability 1/6 to each one.

Thus we define the probability function P in this case by
$$P(\{1\})=P(\{2\})=\ldots=P(\{6\})=1/6.$$

Equivalently, we may write $P(\{k\})=1/6,\ k=1,\ldots,6$
or $P(\{k\})=1/6\ \forall\ k\in S$ (ie, for all k in S).

for all

It is conventional to leave out brackets. Thus we could also write $P(k)=1/6,\ k=1,\ldots,6$. (Leaving out brackets is technically incorrect, since $\{k\}\neq k$, but acceptable in practise and notationally very convenient.)

c

Some consequences of the three axioms

If $A_1, A_2, \ldots$ are disjoint then $P(A_1 \cup A_2 \cup \ldots)$
$= P(A_1) + P(A_2) + \ldots$

Theorem 1	$P(\varnothing) = 0$.

Pf: Apply Axiom 3 with $A_i = \varnothing$ for all i.

$\varnothing = \varnothing \cup \varnothing \cup \ldots$. Also $\varnothing\varnothing = \varnothing$ (ie $\varnothing$ and $\varnothing$ are disjoint). It follows that $\cancel{P(\varnothing)} = P(\varnothing \cup \varnothing \cup \ldots) = \cancel{P(\varnothing)} + P(\varnothing) + \ldots$. We now subtract $P(\varnothing)$ from both sides. Hence $0 = P(\varnothing) + P(\varnothing) + \ldots$ Therefore, $P(\varnothing) = 0$.   QED

Theorem 2	Axiom 3 also holds for *finite* sequences. Thus if $A_1, \ldots, A_n$ are disjoint events, then $P(A_1 \cup \ldots \cup A_n) = P(A_1) + \ldots + P(A_n)$.

Pf: Apply Axiom 3 and Thm 1, with $A_i = \varnothing$ for all $i = n+1, n+2, \ldots$

Theorem 3	$P(\bar{A}) = 1 - P(A)$.

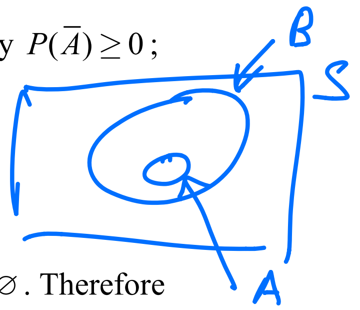
Pf: $1 = P(S)$ by Axiom 2

$= P(A \cup \bar{A})$ by the definition of complementation

$= P(A) + P(\bar{A})$ by Thm 2 with $n = 2$, since A and $\bar{A}$ are disjoint.

Theorem 4	$P(A) \le 1$.

Pf: This follows from Thm 3 and Axiom 1, whereby $P(\bar{A}) \ge 0$;
ie $P(A) = 1 - P(\bar{A}) \le 1 - 0 = 1$.



Another theorem: If $A \subseteq B$, then $P(A) \le P(B)$.

Pf: If $A \subseteq B$, then $B = A \cup (B\bar{A})$, where $A(B\bar{A}) = \varnothing$. Therefore

$P(B) = P(A \cup (B\bar{A}))$

$= P(A) + P(B\bar{A})$ by Theorem 2

$\ge P(A) + 0$ by Axiom 1, whereby $P(B\bar{A}) \ge 0$.

There are many other such results we could write down and prove. However, we now have enough theory to be able to apply our theory of probability to practical problems. The following is one of the two main basic strategies for computing probabilities.

1.

The sample-point method (5 steps)

1. Define the experiment (2 parts: what is *done* & what is *observed*)
2. List the sample points (or possible outcomes). $S = \ldots$
   (Equivalently, list the simple events or write down the sample space.)
3. Define a reasonable probability function P by assigning a pr to each sample point (ie, make sure the 3 axioms of pr are not violated). $A = \{\ldots\}$
4. Express the event of interest, say A, as a collection of sample points,
5. Compute $P(A)$ as the sum of the pr's associated with each of the sample points in A (ie, $P(A) = P(\{e_1,...,e_k\}) = P(\{e_1\}) + ... + P(\{e_k\})$).

Example 4 Find the pr that one head comes up on 2 tosses of a fair coin.
(Note: "One" here means "exactly one", not "at least one".)

1. The experiment consists of tossing a coin twice (1) and each time noting whether a head or tail comes up. (2)
2. Let HT denote heads on the 1st toss and tails on the 2nd,
   let HH denote 2 heads, etc.
   Then the sample points are HH, HT, TH, TT. So the simple events are
   $E_1 = \{HH\}$, $E_2 = \{\text{HT}\}$, $E_3 = \{\text{TH}\}$, $E_4 = \{\text{TT}\}$,
   and the sample space is $S = E_1 \cup E_2 \cup E_3 \cup E_4 = \{\text{HH, HT, TH, TT}\}$.
3. Assign pr 1/4 to each sample point.
   (Thus define the pr function by $P(E_i) = 1/4,\ i = 1,2,3,4$.)
4. Let A = "One head comes up".
   Then $A = \{\text{HT, TH}\} = E_2 \cup E_3$.
5. Therefore $P(A) = P(E_2) + P(E_3) = 1/4 + 1/4 = 1/2$.

In practice we often leave out details, such as brackets (mentioned earlier) and even whole steps. Thus the above solution may be simplified as follows:

2. S: HH, HT, TH, TT.
3. $P(E) = 1/4,\ E \in S$.
4 A = "One H" = {HT, TH}.
5. $P(A) = P(\text{HT}) + P(\text{TH}) = 1/4 + 1/4 = 1/2$.

Note that since all ~~the~~ sample points are equally likely, we may also think of $P(A)$ as n_A / n_S, where $n_A = 2$ is the number of sample points in A and $n_S = 4$ is the number of sample points in S. Thus $P(A) = 2/4 = 1/2$, as before.

Example 5 What's the pr of getting 2 heads on 3 tosses of a coin?

S = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}, $n_S = 8$.
$P(E) = 1/8$ for all E in S.
A = "2 H's" = {HHT, HTH, THH}, $n_A = 3$.
$P(A) = n_A / n_S = 3/8$.

Example 6 What's the pr of getting 2 heads on 5 tosses of a fair coin?

S = {HHHHH, HHHHT, HHHTH, ..., TTTTT}.

We see that listing all the sample points in this case is tedious and impractical.

What we need here are some tools for counting sample points easily, and that leads us to the next topic, *combinatorics*.