SM-1402/CU-0304 Basic Statistics

Chapter 2 (Probability)

Dr. Haziq Jamil

Semester 1, 2020/21

In a room containing (n) individuals, what is the probability that at least two of them share the same birthday?

By the end of this probability lecture we will have the necessary tools to answer this question.

Probability

- The probability of an event is a *measure* of the likelihood that it will happen/occur.

- Probabilities are measured on a numerical scale from 0 to 1.

- Probabilities can be expressed as percentages, fractions or decimals.

- **Important**

  - A probabability of 0 indicates that the event is **impossible**.
  - Conversely, a probability of 1 indicates that the event is **certain** to happen.
  - All other events have probability between 0 and 1.
  - The higher the value, the more likely it is to occur, and vice-versa.

Probability

Examples

- The probability of a fair coin landing 'heads' when tossed is 50% or 1/2 or 0.5.

- There is a one in four chance of cutting a pack of cards at a diamond; the probability is 1/4 or 0.25 or 25%.

- The weather forecaster may say that there is a 70% chance of rain tomorrow.

Probability relates to our *uncertainty* about the world. The concept of *randomness* is also inherintly linked to probabilities.

Probability

Notation

Any statistical experiment or trial has a number of possible outcomes.

- The set of all possible outcomes is called the **possibility space** S. This is also known as the **sample space**.

- An event A of the experiment is defined to be the subset of S.

Example: A six-sided die is thrown and the outcome is recorded.

- The sample space is $S = \{1, 2, 3, 4, 5, 6\}$.
- An event A is any subset of S. Let A be the event that 'the score is less than 3'. Then,
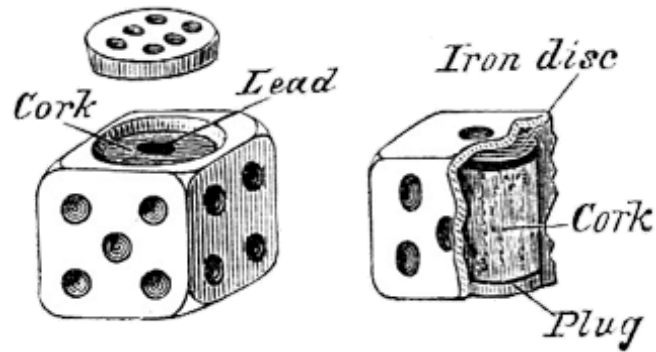
$$A = \{1, 2\}$$

Probability

A fair six-sided die

Probability

An unfair six-sided die

Probability

Calculating probabilities

Example: Two six-sided dice are thrown and the outcome for both dice are recorded.

- There are 36 possible outcomes: $\{1,1\}, \{1,2\}, \{2,1\}, \{1,3\}$, etc.

- Define A to be the event 'the sum of the two scores is 6. These outcomes are

  - $\{1,5\}, \{2,4\}, \{3,3\}, \{4,2\}$, and $\{5,1\}$
  - There are five such outcomes. We denote this by the notation $n(\cdot)$, as in $n(A) = 5$.

- Let $P(A)$ denote the probability of A. Then,

$$P(A) = \frac{n(A)}{n(S)} = \frac{5}{36}.$$

Probability
Possibility space diagram

Complementary events

- The complementary event A' denotes 'the event A does not occur'.

- $n(A') = n(S) - n(A)$.

$$P(A') = \frac{n(S) - n(A)}{n(S)}$$
$$= 1 - \frac{n(A)}{n(S)}$$
$$= 1 - P(A)$$

- Therefore, $P(A) + P(A') = 1$.

Complementary events

Example: A group of 20 university students contain eight who are in their first year of study. A student is picked at random to represent the group at a meeting. Find the probability that the student is not in the first year of study.

- Let A denote the event that 'a student *is* in the first year of study'.

- Therefore, A' is the event that 'a student *is not* in the first year of study'.

- $P(A) = 8/20 = 0.4$.

- So $P(A') = 1 - 0.4 = 0.6$.

Venn diagrams

Suppose A and B are two events associated with the same experiment. Consider the outcomes described below

Union

- The set that contains the outcomes that are in A or B or both is called the *union* of A and B, and is written $A \cup B$.
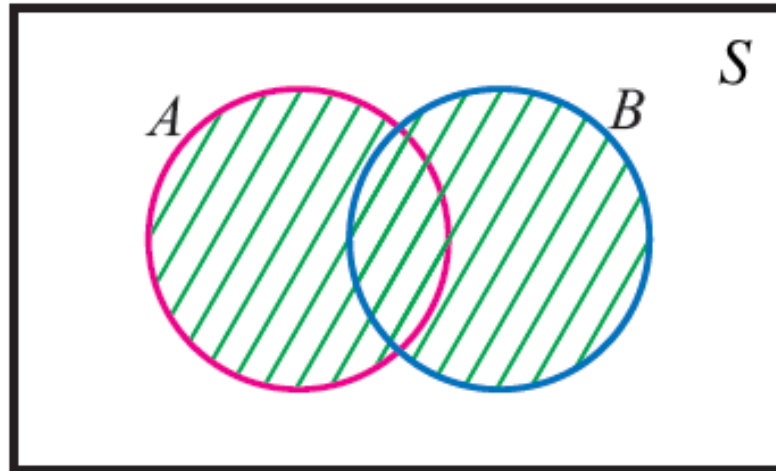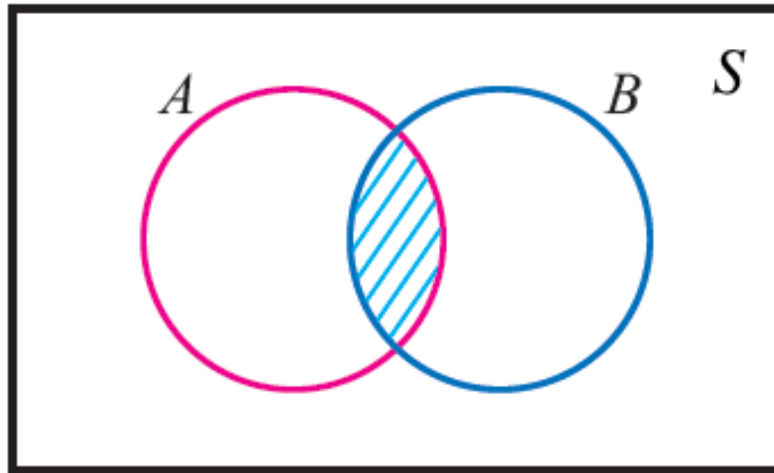
Venn diagrams

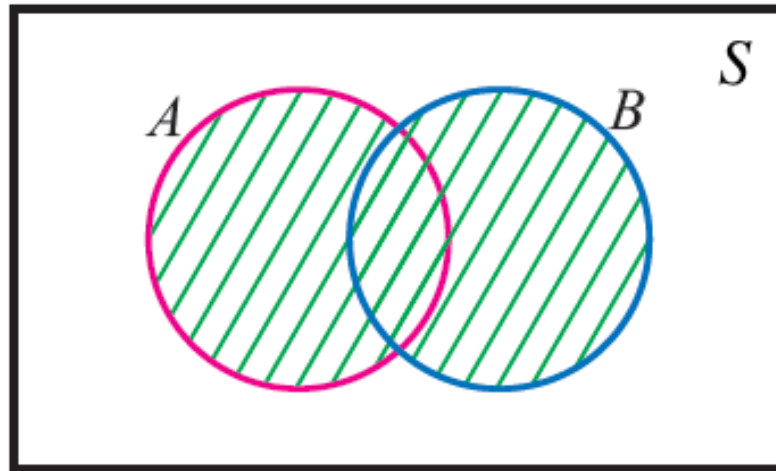Suppose A and B are two events associated with the same experiment. Consider the outcomes described below

Intersection

- The set that contains the outcomes that are in both A and B is called the *intersection* of A and B, and is written $A \cap B$.

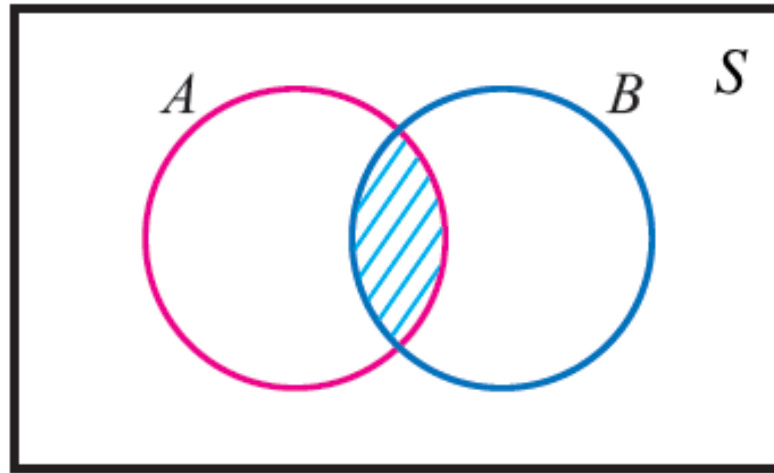# Probability results for combined events

Result 1

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

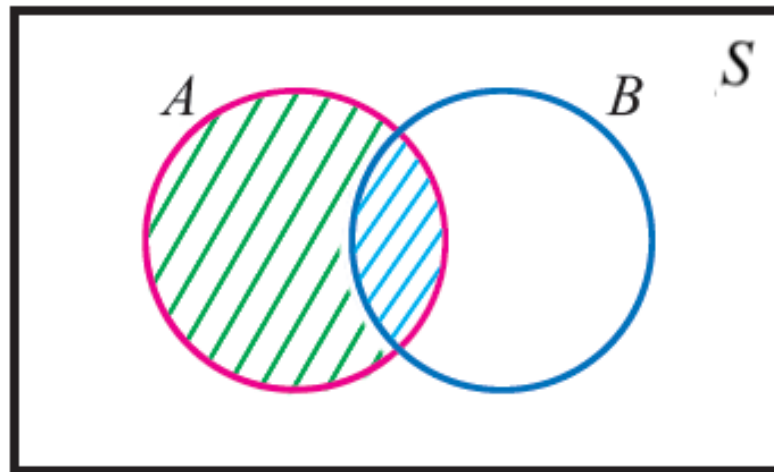# Probability results for combined events

Result 2

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

Probability results for combined events

Result 3

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

Probability results for combined events

Result 4
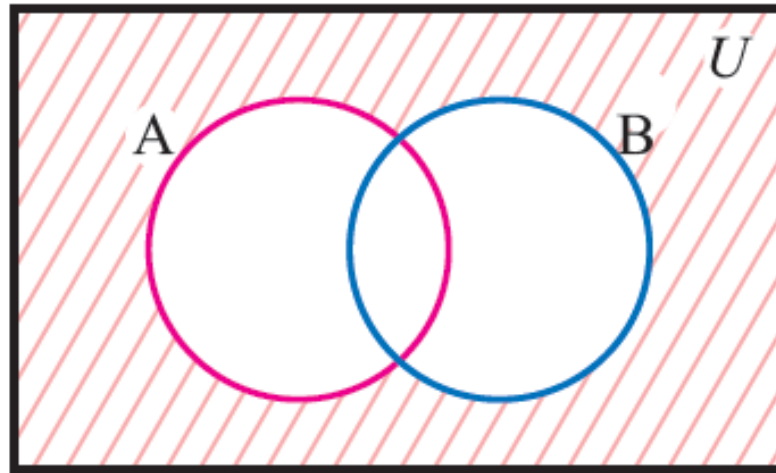
Probability of 'neither A nor B'

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

Examples

Example 2.3

In a class of 20 children, 4 of the 9 boys and 3 of 11 girls are in the athletics team. A person from the class is chosen to be in the 'egg and spoon' race on Sports Day. Find the probability that the person chosen is

(a) in the athletics team,

(b) female,

(c) A female member of the athletics team,

(d) A female or in the athletics team.

Examples

Solution 2.3

Let A be the event that they are in the athletics team, and let F be the event that they are female.

- $n(S) = 20$.
- $n(A) = 4$ boys $+3$ girls $= 7$ altogether.
- $n(F) = 11$.
- $n(A \cap F) = 3$.

(a) $P(A) = n(A)/n(S) = 7/20 = 0.35$.

(b) $P(F) = n(F)/n(S) = 11/20 = 0.55$.

(c) $P(A \cap F) = n(A \cap F)/n(S) = 3/20 = 0.15$.

(d) $P(A \cup F) = P(A) + P(F) - P(A \cap F) = 0.75$.

Examples

Example 2.6

Events A and B are such that $P(A) = 0.3$, $P(B) = 0.4$, and $P(A \cap B) = 0.1$. Find (a) $P(A \cap B')$ and (b) $P(A' \cap B')$.

Hint: $P(A) = P(A \cap B') + P(A \cap B)$ and $P(A' \cap B') = 1 - P(A \cup B)$.

Solution 2.6

(a)

$$
\begin{aligned}
P(A \cap B') + P(A \cap B) &= P(A) \\
\Rightarrow P(A \cap B') &= P(A) - P(A \cap B) \\
&= 0.3 - 0.1 \\
&= 0.2
\end{aligned}
$$

Examples

Solution 2.6

(b) Recall that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, so

$$P(A \cup B) = 0.3 + 0.4 - 0.1 = 0.6$$

Therefore,

$$\begin{aligned} P(A' \cap B') &= 1 - P(A \cup B) \\ &= 1 - 0.6 \\ &= 0.4 \end{aligned}$$

Mutually exclusive events

Consider events A and B of the same experiment. The two events are said to be **mutually exclusive** if they cannot occur at the same time.

Examples:

- Throwing a single six-sided die and getting both a '3' and a '5' simultaneously.
- Flipping a coin and obtaining both 'heads' and 'tails' simultaneously.
- A car turning left and turning right simultaneously.

Examples of non-mutually exclusive events:

- Drawing an even-numbered card and a red card from a deck of playing cards.
- Throwing a single six-sided die and getting both an odd score and a score greater than 3.

Mutually exclusive events

Since A and B are mutually exclusive, $A \cap B$ is an impossible event. Therefore, $P(A \cap B) = 0$.
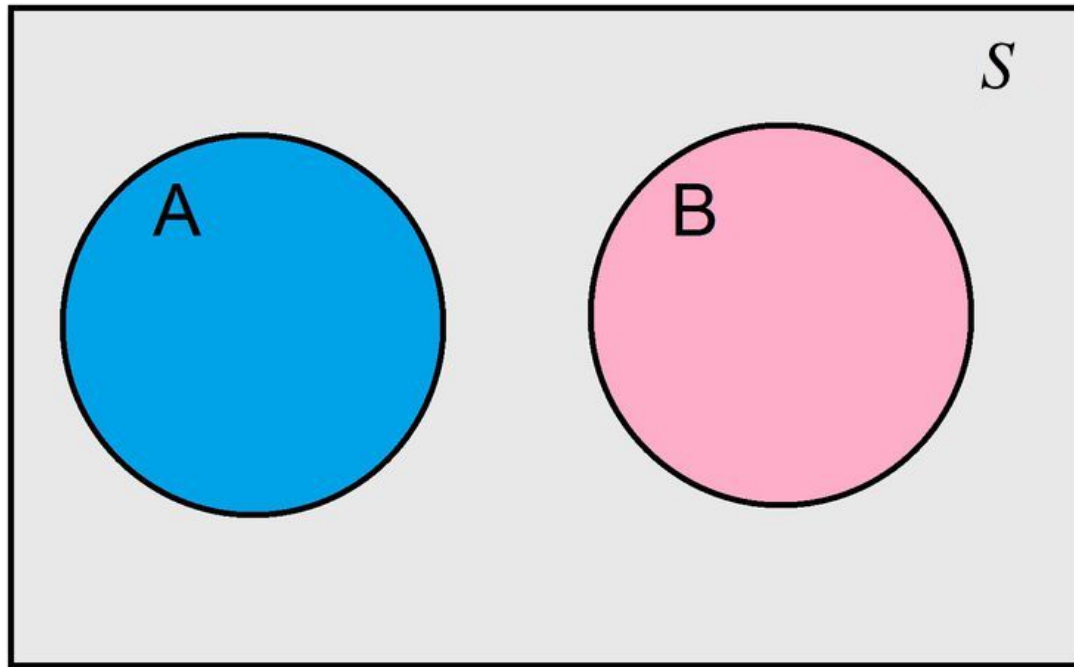
There is no overlap of A and B!

Mutually exclusive events

For mutually exlusive events,

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

Extending this result to n mutually exclusive events $A_1, A_2 \ldots, A_n$,

- $P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n)$

Exhaustive events

If two events A and B are such that between them they make up the whole of the possibility space, then A and B are said to be **exhaustive events**, and

- $P(A \cup B) = 1$ (for exhaustive events A and B)
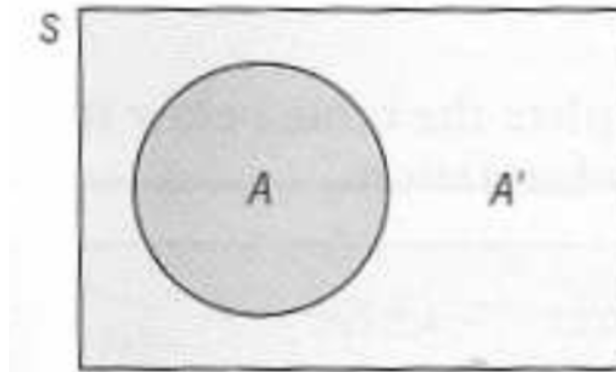
Example:

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

Exhaustive events

Special case

Consider an event A and its complementary event A'.

Then,

- $P(A \cap A') = 0$
- $P(A \cup A') = P(A) + P(A') = 1$
- Any event A and its complementary A' are both mutually exclusive and exhaustive.

Exhaustive events

Special case

We can extend this to n exhasutive events $A_1, A_2 \ldots, A_n$:

$P(A_1) + \cdots + P(A_n) = 1$ if and only if the n events are both mutually exclusive and exhaustive.

Conditional probability

If A and B are two events, not necessarily from the same experiment, then the **conditional probability** that A occurs, given that B has already occured, is written $P(A|B)$.

The conditional probability is defined as

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

This can be rearranged to give $P(A \cap B) = P(A|B)P(B)$.

Also, by definition, $P(B \cap A) = P(B|A)P(A)$. Thus,

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

Independent events

If either of the events A and B can occur without being affected by the other, then the two events are said to be **independent**. Mathematically,

- $P(A|B) = P(A)$ if and only if A and B are independent.
- It is also true that $P(B|A) = P(B)$.

Remember that using the conditional probability definition, we have

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

So for *independent events* A and B, the result becomes

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

Independent events

Three conditions for A and B to be independent:

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

Any of these three conditions may be used as a test of independence.

The independence results can be extended to n independent events $A_1, A_2 \ldots, A_n$:

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \times \cdots \times P(A_n)$$

Examples

Example 2.8: When a die was thrown the score was an odd number. What is the probability that it was a prime number?

Let A be the event 'score is prime', and B be the event 'the score is odd'. We are interested in $P(A|B)$.

Note that

- $n(B) = 3$, so $P(B) = 3/6 = 1/2$.
- $n(A \cap B) = n(\{3, 5\}) = 2$, and thus $P(A \cap B) = 2/6 = 1/3$.

Therefore,

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

Examples

Example 2.10: A fair six-sided die is thrown twice. Find the probability that two fives are thrown.

Firstly, realise that the two throws are *independent* of each other. The result of the first throw does not affect the result of the second throw.

Let A_1 be 'score of 5 on first throw' and A_2 be 'score of 5 on second throw'. We are interested in the probability of A_1 *and* A_2 occurring together.

$$P(A_1 \cap A_2) = P(A_1)P(A_2) = 1/6 \times 1/6 = 1/36.$$

Probability trees

A useful way of tackling many probability problems is to draw a probability tree. The method is illustrated in the following example.

Example 2.17: In a certain selection of flower seeds, 2/3 have been treated to improve germination and 1/3 have been left untreated. The seeds which have been treated have a probability of germination of 0.8, whereas the untreated seeds have a probability of germination of 0.5.

(a) Find the probability that a seed, selected at random, will germinate (the seeds were sown and given time to germinate).

(b) Find the probability that a seed selected at random had been treated, given that it had germinated.
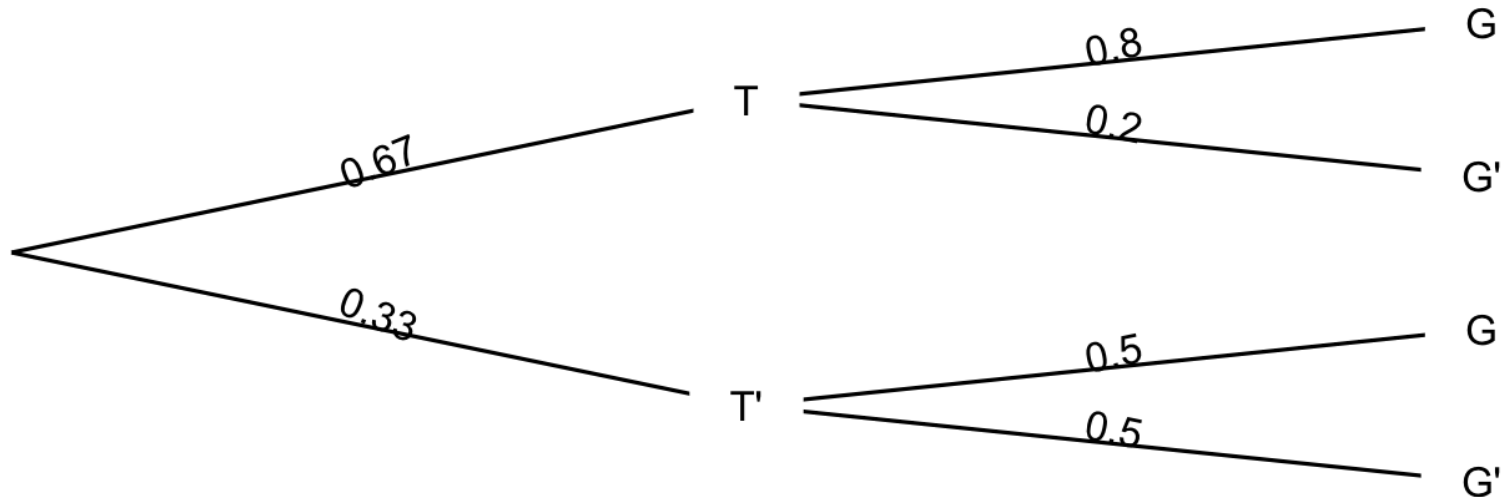
Probability trees

This is what we know...

- Let T be 'seed is treated'. So $P(T) = 2/3$ and $P(T') = 1/3$.
- Let G be 'seed germinates'. So $P(G|T) = 0.8$ and $P(G|T') = 0.5$

We'll draw a probability tree to determine

(a) $P(G)$

(b) $P(T|G)$

Probability trees

$$P(G) = P(G \cap T) + P(G \cap T')$$
$$= P(G|T)P(T) + P(G|T')P(T')$$
$$= 2/3 \times 0.8 + 1/3 \times 0.5 = 0.7$$
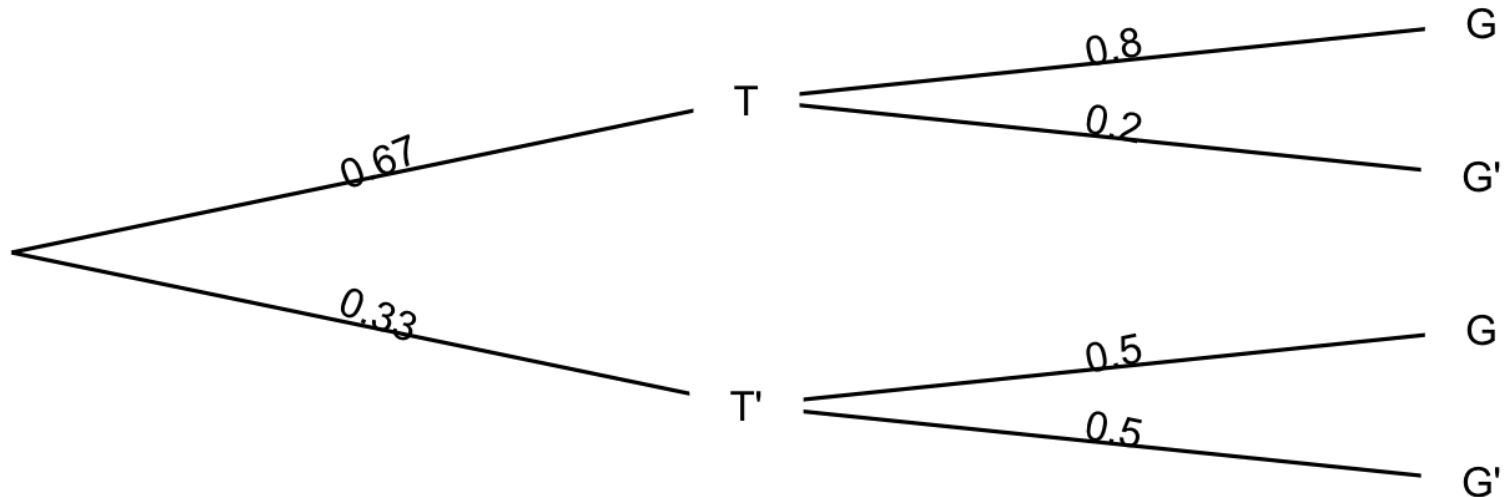
Probability trees

$$P(T|G) = \frac{P(G \cap T)}{P(G)} = \frac{P(G|T)P(T)}{P(G)}$$

$$= \frac{0.8 \times 2/3}{0.7} = 0.762 \text{ (3d.p.)}$$

Law of Total Probability

In the previous probability tree example, we wanted to find out the **marginal probability** $P(G)$, which we calculated using other conditional/joint probabilities involving the event G.

Consider a set of exhaustive events $\{B, A_1, \ldots, A_n\}$ whose union makes up the entire sample space S. The law of total probability states that

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Bayes' Theorem

Thomas Bayes (1701-1761) was an English statistician, philosopher and Presbyterian minister

Bayes' Theorem

Bayes' theorem connects the two probabilities $P(A|B)$ and $P(B|A)$ through the conditional probability rule.

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

This can be extended as well, by replacing the marginal probability $P(B)$ in the formula with the result of the *law of total probability*.

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

The message here is that you can calculate $P(A|B)$ if you have knowledge regarding the "reverse" conditionals $P(B|A)$.

Problems involving an 'at least' situation

Example 2.21 (a): Find the probability of obtaining at least one six when five dice are thrown.

The events that we are interested in are

- One die scores a 6;
- Two dice score 6
- Three dice score 6
- ...and so on.

It is easier to think of the complement of the event 'at least one six is thrown', which is

- 'NO die scores a 6' or 'All 5 die do not score 6'.

Problems involving an 'at least' situation

Example 2.21 (a): Find the probability of obtaining at least one six when five dice are thrown.

Let A be the event 'at least one six is thrown'. Then,

$$P(A') = 1 - P(A)$$
$$= 1 - \left(\frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \right)$$
$$= 0.598 \text{ (3 d.p)}$$

Arrangements

In order to calculate the number of possible outcomes in a possibility space or an event, the following results are often used.

Result 1

The number of ways of arranging n unlike objects in a line is $n!$

$$n! = n \times (n-1) \times (n-2) \times \cdots 3 \times 2 \times 1$$

The ' $!$ ' symbol is pronounced 'factorial'. Note that $0! = 1$ by definition.

Result 2

The number of ways of arranging in a line n objects, of which p are alike, is $\frac{n!}{p!}$

Arrangements

Example: Consider the letters A, B, C, D. How many possible arrangements are there from these four letters?

Imagine that you have four boxes to fill with these four letters (withour replacement)

- 1st box: Any of the **4** letters can be chosen.
- 2nd box: The remaining **3** letters can be chosen.
- 3rd box: The remaining **2** letters can be chosen.
- 4th box: No choice to be made here.
- Therefore, total number of arrangements = $4! = 24$.

```
ABCD  ABDC  ACBC  ACDB  ADCB  ADBC
BCDA  BCAD  BDAC  BDCA  BACD  BADC
CDBA  CDAB  CABD  CADB  CBAD  CBDA
DABC  DACB  DBCA  DBAC  DCAB  DCBA
```

Arrangements

Example: Consider the letters A, A, A, D. How many possible arrangements are there from these four letters?

As before, we imagine four boxes to fill with these four letters. However, we notice that three of the letters are the same, which makes some of the arrangements indistinguishable from one another.

AAAD AADA ADAA DAAA

Using the formula, we have

$$\frac{4!}{3!} = \frac{4 \times 3 \times 2 \times 1}{3 \times 2 \times 1}$$
$$= 4$$

Arrangements

Example: If a four-digit number is formed from the digits 1, 2, 3, and 5, and repetitions are not allowed, find the probability that the four-digit number is divisible by 5.

- In total, there are $4! = 24$ possible arrangements.
- However, for a number to be divisible by 5, it must end in 5.
- So the first 3 digits can be arranged in any order using the remaining 3 digits.
- There are $3! = 6$ possible ways to do this.
- Thus, the probability is $6/24 = 1/4$.

Permutations

Consider the number of ways of placing the letters A, B, C, D, E, F, G in three empty spaces.

- The first space can be filled in **7** ways
- The second space can be filled in **6** ways
- The third space can be filled in **5** ways
- Therefore there are $7 \times 6 \times 5 = 210$ ways in total.

This is the number of **permutations** of three objects taken from seven, and is written $^7P_3$.

Note that the order is important: ABC is different from BAC, although they contain the same three letters.

Permutations

The number of permutations, or ordered arrangements, of r objects taken from n unlike objects, is written $^{n}P_{r}$, where

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

Combinations

If instead we viewed all arrangements with the same letters as identical, i.e. treat ABC, ACB, CAB, BAC, BCA, CBA all the same, then we are only interested in **combinations** of letters.

For each choice of three letters, there are $3!$ possible arrangements.

Denote the number of combinations of these 3 objects from 7 objects as $^{7}C_3$. Then, we have the relationship

$$^{7}P_3 = 3! \times ^{7}C_3$$

Which then implies

$$^{7}C_3 = \frac{^{7}P_3}{3!} = \frac{7!}{3!(7-3)!} = 35$$

Combinations

A combination of n objects taken r at a time is a choice of r out of n objects. This is given by the formula

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

Examples

Example 2.27: In how many ways can a hand of four cards be dealt from an ordinary pack of 52 playing cards?

Hint: Is this a permutation or a combination problem?

It is a combination problem: The order in which the cards are dealt is not important.

The answer is $^{52}C_4 = 270,725$ possible combinations.

Examples

Example 2.28: Four letters are chosen randomly from the word RANDOMLY. Find the probability that all four letters chosen are consonants.

Hint: Is this a permutation or a combination problem?

- What is the size of the sample space? $n(S) = {}^8C_4 = 70$.
- Let E represent 'all four letters are consonants'.
- What is the size of the event space? $n(E) = {}^6C_4 = 15$.
- Thus $P(E) = n(E)/n(S) = 15/70 = 0.214$ (2 d.p.).

The Birthday Problem

In a room containing n individuals, what is the probability that at least two of them share the same birthday?

First, some assumptions.

- There are only 365 days in a year.
- Each day is equally likely to be a birthday.
- Birthdays are independent of each other.

Hint: 'at least' means look at the event complement.

- Let E be 'at least two people share same birthday'

- Therefore E' is the event 'no one shares the same birthday'

- What is the size of the sample space? $n(S) = 365^n$, since each person has a birthday on one of the 365 days. For two people it is 365×365 (recall the possibility diagram for the six-sided dice). By extension for n people it is 365^n.

- What is the size of the complement of the event space? We are interested in choosing n different days of the year (from 365). Birthdays are unique to the individuals, so order matters (it is a permutation problem). Therefore, $n(E') = ^{365}P_n$

- The probability of the event complement is therefore

$$P(E') = n(E')/n(S) = \frac{365!}{(365-n)!365^n}$$

- Alternatively, we can use conditional probabilities.

- Let A_i be the event that person i has a unique birthday in the year. We are interested in

$$P(A_1 \cap A_2 \cap \cdots \cap A_n)$$

- Using the definition of conditional probabilities,

$$
\begin{aligned}
P(A_1 \cap A_2 &\cap \cdots \cap A_n) \\
&= P(A_1)P(A_2 \cap \cdots \cap A_n | A_1) \\
&= P(A_1)P(A_2 | A_1)P(A_3 \cap \cdots \cap A_n | A_1, A_2) \\
&= \vdots \\
&= P(A_1)P(A_2 | A_1) \cdots P(A_n | A_1, \ldots, A_{n-1})
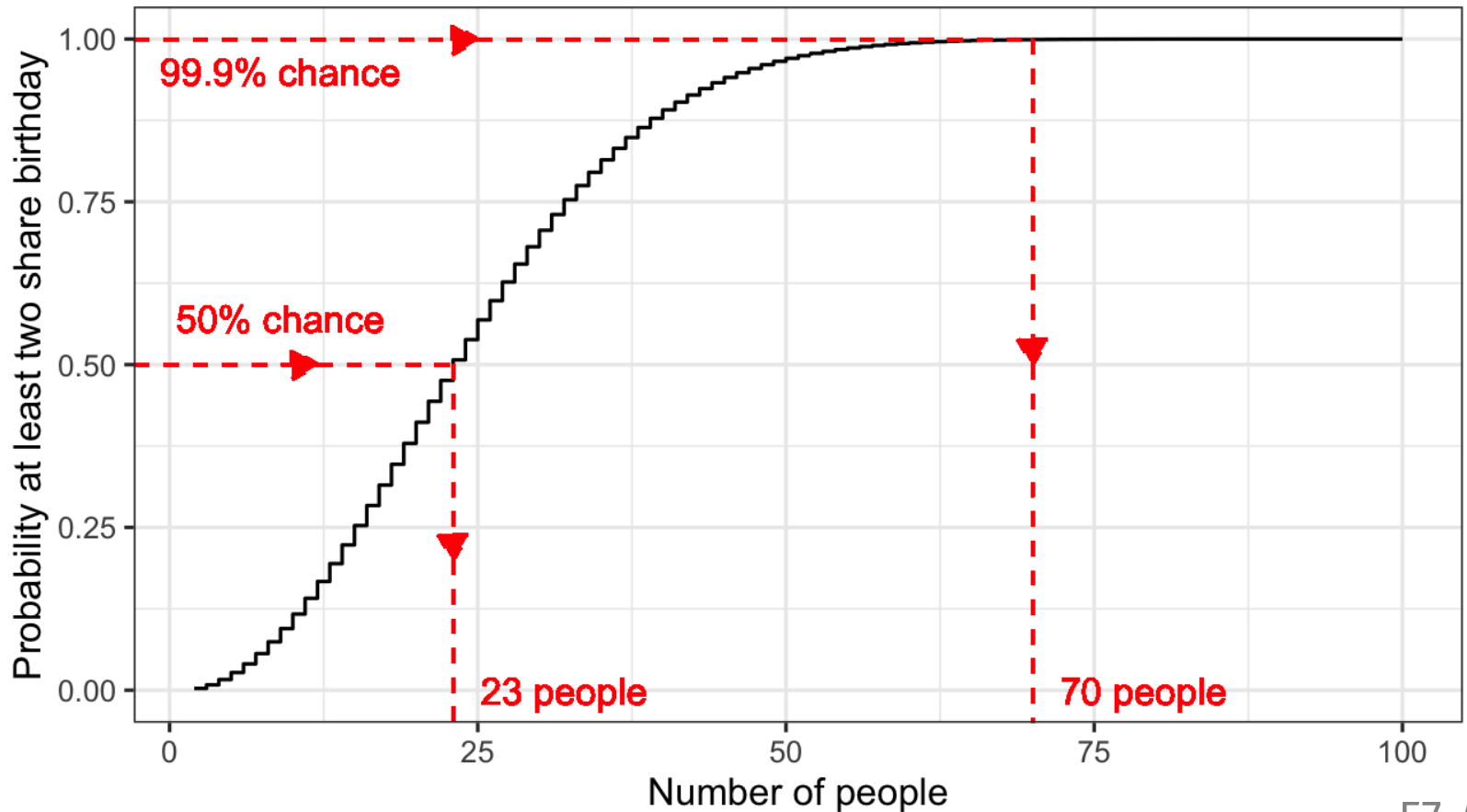\end{aligned}
$$

- Taking each of these probabilities in turn,

  - $P(A_1) = 365/365 = 1$

  - $P(A_2|A_1) = 364/365$ since one birthday has already been taken, now only 364 days of the year remains

  - $P(A_3|A_1, A_2) = 363/365$ by a similar argument

  - $P(A_n|A_1, \ldots, A_{n-1}) = (365 - n + 1)/365$ in general (following the pattern)
- Therefore

$$
\begin{aligned}
P(E') &= P(A_1 \cap A_2 \cap \cdots \cap A_n) \\
&= \frac{365}{365} \cdot \frac{364}{365} \cdots \frac{365 - n + 1}{365} \\
&= \frac{365!}{(365 - n)!365^n}
\end{aligned}
$$

Therefore the probability of interest is

$$P(E) = 1 - P(E') = 1 - \frac{365!}{(365-n)!\,365^n}$$

- Remark: The probability to calculate is better evaluated using logarithms due to the very large numbers resulting from the factorials.

$$\frac{365!}{(365-n)!365^n} = \exp\left[\log\left(\frac{365!}{(365-n)!365^n}\right)\right]$$
$$= \exp\left[\log 365! - \log(365-n)! - n\log 365\right]$$

END