Mathematical Statistics

MAS 713

Chapter 2

Previous lecture

1. What is statistics ?
2. Descriptive Statistics.

Questions?

This lecture

2. Elements of Probability

- 2.1 Introduction
- 2.2 Events
- 2.3 Probabilities
- 2.4 Conditional Probabilities
- 2.5 Independent Events
- 2.6 Bayes' Formula

Additional reading : Chapter 1 in the textbook

The previous chapter described purely descriptive methods for a given sample of data

The subsequent chapters will describe **inferential methods**, that convert the information from random samples into information about the whole population which the sample has been drawn from

However, a sample only gives a partial and approximate picture of the population

$\leadsto$ drawing conclusions about the whole population, thus going beyond what we have effectively observed, necessarily involves some risk

$\leadsto$ it is important to quantify the amount of confidence or reliability in what we observe in the sample

It is important to keep in mind the crucial role played by random sampling

Without random sampling, statistics can only provide descriptive summaries of the data

With random sampling, the conclusions can be extended to the population, arguing that the randomness of the sample guarantees it to be representative of the population on average

"**Random**" is not synonymous with "chaotic" or "haphazard"

"Random" describes a situation in which an individual outcome is uncertain, but there is a regular distribution of outcomes in a large number of repetitions

Probability theory is the branch of mathematics concerned with analysis of random phenomena

$\rightsquigarrow$ probabilities and related concepts are the required tools to fill the gap between descriptive and inferential statistics.

Events

Random experiment

Definition

A **random experiment** (sometimes *chance experiment*) is any experiment whose exact outcome cannot be predicted with certainty

Random experiment

This definition includes the ‘usual’ introduction-to-probability random experiments...

Experiment 1 : toss a coin

Experiment 2 : roll a die

Experiment 3 : roll two dice

... as well as typical engineering experiments...

Experiment 4 : count the number of defective items produced on a given day

Experiment 5 : measure the current in a copper wire

Sample space

To model and analyse a random experiment, we must understand the set of all possible outcomes from the experiment

Definition

The set of all possible outcomes of a random experiment is called the **sample space** of the experiment. It is usually denoted S

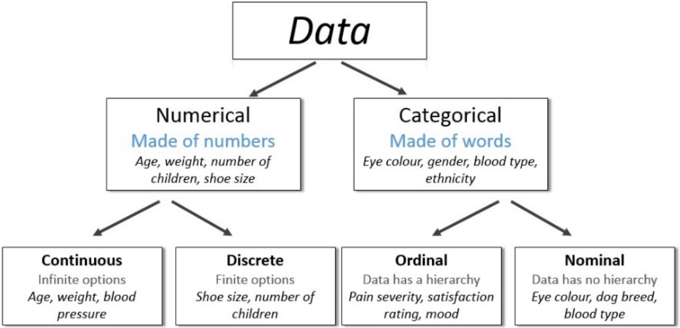
Sample space

Experiment 1 : $S = \{H, T\}$
Experiment 2 : $S = \{1, 2, 3, 4, 5, 6\}$
Experiment 3 : $S = \{(1, 1), (1, 2), (1, 3), \ldots, (6, 6)\}$
Experiment 4 : $S = \{0, 1, 2, \ldots n\}$ or $S = \{0, 1, 2, \ldots\}$
Experiment 5 : $S = [0, +\infty)$

Sample space

Each element of the sample space S, that is each possible outcome of the experiment, is a simple event, generically denoted ω

From the above examples, the distinction between discrete (finite or countable) and continuous sample spaces is clear



Often we are interested in a collection of related outcomes from a random experiment, that is by a subset of the sample space, which has a physical reality

Definition

An **event** E is a subset of the sample space of a random experiment

Examples of events :

Experiment 1 : $E_1 = \{H\}$ = "the coin shows up Heads"

Experiment 2 : $E_2 = \{2, 4, 6\}$ = "the die shows up an even number"

Experiment 3 : $E_3 = \{(1, 3), (2, 2), (3, 1)\}$ = "the sum of the dice is 4"

Experiment 4 : $E_4 = \{0, 1\}$ = "there is at most one defective item"

Experiment 5 : $E_5 = [1, 2]$ = "the current is between 1 and 2 A"

If the outcome of the experiment is contained in E,
then we say that E has occurred

The elements of interest are the events, which are sets

$\Rightarrow$ basic concepts of set theory will be useful

Set notation

- union $E_1 \cup E_2$ = event "either E_1 **or** E_2 occurs"
- intersection $E_1 \cap E_2$ = event "both E_1 **and** E_2 occur"
- complement E^c = event "E does **not** occur"

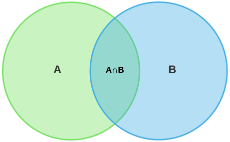
S is an event $\rightsquigarrow$ certain event $S^c \doteq \phi \rightsquigarrow$ impossible event

$E_1 \subseteq E_2 \rightsquigarrow E_1$ implies E_2

$E_1 \cap E_2 = \phi \rightsquigarrow$ mutually exclusive events

De Morgan's laws:
$$(E_1 \cup E_2)^c = E_1^c \cap E_2^c$$
$$(E_1 \cap E_2)^c = E_1^c \cup E_2^c$$

These relations are easily illustrated by means of Venn diagrams



Example

$$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$
$$E_1 = \{1, 3, 5, 7, 9\}$$
$$E_2 = \{1, 2, 3, 4, 5\}$$

Questions:

$$E_1 \cup E_2 = ?$$
$$E_1 \cap E_2 = ?$$

Probabilities

The axioms of probability theory

Intuitively, the probability $\mathbb{P}(E)$ of an event E is a real number which should measure

how likely E is to occur

Kolmogorov's probability axioms

The probability measure $\mathbb{P}(\cdot)$ satisfies :

i) $0 \leq \mathbb{P}(E) \leq 1$ for any event E

ii) $\mathbb{P}(S) = 1$

iii) for any (infinite) sequence of mutually exclusive events $E_1, E_2, \ldots$,

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

Useful implications of the axioms

- Any finite sequence of **mutually exclusive** events $E_1, E_2, \ldots, E_n$,

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

- $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$
- $\mathbb{P}(\phi) = 0$
- $E_1 \subseteq E_2 \Rightarrow \mathbb{P}(E_1) \leq \mathbb{P}(E_2)$ (increasing measure)
- $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_2)$,

Subadditivity of the probability measure $\mathbb{P}$

For **any** events $E_1, \ldots, E_n$

$$\mathbb{P}\Big(\bigcup_{i=1}^{n} E_i\Big) \leq \sum_{i=1}^{n} \mathbb{P}(E_i).$$

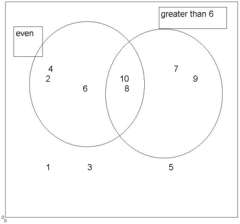
Additive Law of Probability

For **any** events $E_1, \ldots, E_n$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \mathbb{P}(E_i) - \sum_{i<j} \mathbb{P}(E_i \cap E_j) + \sum_{i<j<k} \mathbb{P}(E_i \cap E_j \cap E_k) + \ldots + (-1)^{n-1}\mathbb{P}\left(\bigcap_{i=1}^{n} E_i\right)$$

Application of **Additive Law of Probability**

One of the numbers from 1 to 10 is selected at random (call it N).
We want to find $\mathbb{P}$(even OR greater than 6).



$$\mathbb{P}(\text{ N even, OR N} > 6) = \mathbb{P}(N \text{ even}) + \mathbb{P}(N > 6) - \mathbb{P}(N \text{ even, AND } N > 6)$$

$$\mathbb{P}(N \text{ even}) = 5/10, \quad \mathbb{P}(N > 6) = 4/10, \quad \mathbb{P}(N \text{ even, AND } N > 6) = 2/10$$

$$\Longrightarrow$$

$$\mathbb{P}(\text{ N even, OR N} > 6) = \mathbb{P}(N \text{ even}) + \mathbb{P}(N > 6) - \mathbb{P}(N \text{ even, AND } N > 6)$$
$$= 5/10 + 4/10 - 2/10 = 7/10$$

Assigning probabilities

Note

The axioms state only the conditions an assignment of probabilities must satisfy, but they do not tell how to assign specific probabilities to events

$\leadsto$ different approaches can be used, the most widely held being the **frequentist approach**

Assigning probabilities

Frequentist definition of probability

If the experiment is repeated independently over and over again (infinitely many times), the proportion of time that event E occurs is its probability $\mathbb{P}(E)$

Suppose the experiment is repeated n times. Then, the probability of the event E is

$$\mathbb{P}(E) = \lim_{n\to\infty} \frac{\text{number of times } E \text{ occurs}}{n}$$

Assigning probabilities

Interpretation

probability $\simeq$ proportion of chances of occurrence of the event

It is straightforward to check that the so-defined probability measure satisfies the axioms

Of course, this definition remains theoretical, as assigning probabilities would require infinitely many repetitions of the experiment

Assigning probabilities

Besides, in many situations, the experiment cannot be faithfully replicated
(What is the probability that it will rain tomorrow?
What is the probability of finding oil in that region?)

$\rightsquigarrow$ essentially, assigning probabilities in practice relies on prior knowledge of the experimenter (belief and/or model)

A simple model is to assume that all the outcomes are equally likely, other more elaborated models define probability distributions.

A simple example

Experiment: flipping a coin $\rightsquigarrow S = \{H, T\}$
Does this experiment agree with the Kolmogorov axioms?

Kolmogorov's probability axioms

The probability measure $\mathbb{P}(\cdot)$ satisfies :

i) $0 \leq \mathbb{P}(E) \leq 1$ for any event E
ii) $\mathbb{P}(S) = 1$
iii) for any (infinite) sequence of mutually exclusive events $E_1, E_2, \ldots$,

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

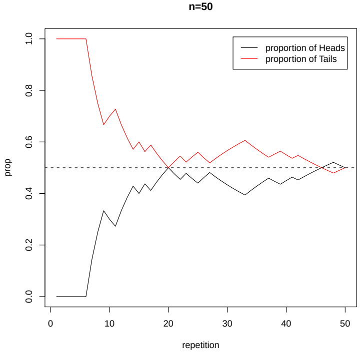
i) $0 \leq \mathbb{P}(H) \leq 1, 0 \leq \mathbb{P}(T) \leq 1$
ii) $\mathbb{P}(S) = \mathbb{P}(H \cup T) = 1$
iii) $\mathbb{P}(H \cup T) = \mathbb{P}(H) + \mathbb{P}(T)$

A simple example

The coin is tossed n times, we observe the proportion of H and T :

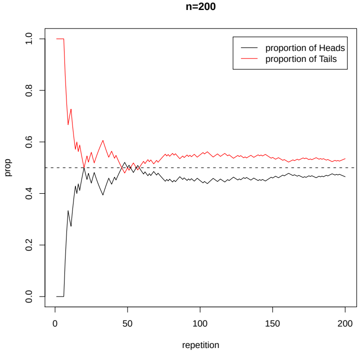
A simple example

The coin is tossed n times, we observe the proportion of H and T :



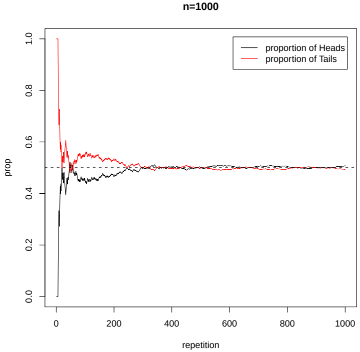
A simple example

The coin is tossed n times, we observe the proportion of H and T :



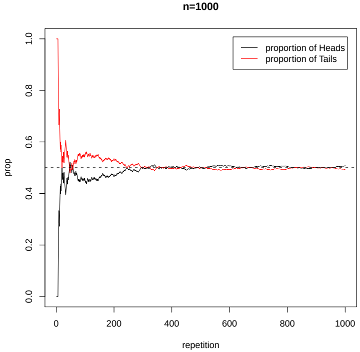
A simple example

The coin is tossed n times, we observe the proportion of H and T :



A simple example

The coin is tossed n times, we observe the proportion of H and T :



$$\rightsquigarrow \mathbb{P}(H) = \mathbb{P}(T) = \frac{1}{2}$$

(fair coin, in this case)

Assigning probabilities : equally likely outcomes

Assuming that all the outcomes of the experiment are equally likely provides important simplification

Suppose there are N possible outcomes $\{\omega_1, \omega_2, \dots, \omega_N\}$, equally likely to one another, $\mathbb{P}(\omega_k) = p$ for all k

Assigning probabilities : equally likely outcomes

Then, Axioms 2 and 3 impose $p + p + \ldots + p = Np = 1$, that is,

$$p = \frac{1}{N}$$

$\rightsquigarrow$ for an event E made up of k simple events, it follows from Axiom 3

$$\mathbb{P}(E) = \frac{k}{N} = \frac{\text{number of favorable cases}}{\text{number of total cases}}$$

$\rightsquigarrow$ “classical” definition of the probability

$\rightsquigarrow$ necessary to be able to effectively count the number of different ways that a given event can occur ($\rightsquigarrow$ combinatorics)

Basic combinatorics rules

Basic combinatorics rules

multiplication rule :
If an operation can be described as a sequence of k steps, and the number of ways of completing step i is n_i, then the total number of ways of completing the operation is

$$n_1 \times n_2 \times \ldots \times n_k$$

Example

A test consists of 6 mutiple-choice questions. Each question has 4 possible answers. There are

$$4 \times 4 \times 4 \times 4 \times 4 \times 4 = 4^6$$

ways to answer all 6 questions.

Basic combinatorics rules

permutations :
a permutation of the elements of a set is an ordered sequence of those elements. The number of different permutations of n elements is

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

Example

How many different ways can you arrange the letters X, Y, and Z?
We have 3 distinct objects so $n = 3$. Thus, the number of combinations is

$$3 \times 2 \times 1 = 6$$

Basic combinatorics rules

combinations :

a combination is a subset of elements selected from a larger set. The number of combinations of size r that can be selected from a set of n elements is

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

Example

How many different ways can you select 2 letters from the set of letters: X, Y, and Z?

We have 3 distinct objects so $n = 3$. And we want to arrange them in groups of 2, so $r = 2$. Thus, the number of combinations is

$$\frac{3!}{2!\,(3-2)!} = 3$$

Equally likely outcomes : example

Example 1

A computer system uses passwords that are 6 characters and each character is one of the 26 letters (a-z) or 10 integers (0-9). Uppercase letters are not used.
Let A the event that a password begins with a vowel (either a, e, i, o or u) and let B denote the event that a password ends with an even number (either 0, 2, 4, 6 or 8).

Suppose a hacker selects a password at random. What are the probabilities $\mathbb{P}(A)$, $\mathbb{P}(B)$, $\mathbb{P}(A \cap B)$ and $\mathbb{P}(A \cup B)$?

- All passwords are equally likely to be selected

$\Longrightarrow$ "classical" definition of probability

- Total number of cases $= 36^6 = 2,176,782,336$

- Let A the event that a password begins with a vowel (either a, e, i, o or u)
  Let B denote the event that a password ends with an even number (either 0, 2, 4, 6 or 8).

$$\mathbb{P}(A) = \frac{5}{36}, \qquad \mathbb{P}(B) = \frac{5}{36}$$

$$\mathbb{P}(A \cap B) = \frac{\text{number of favorable cases}}{\text{number of total cases}} = \frac{5 \times 36 \times 36 \times 36 \times 36 \times 5}{36^6} = \frac{5}{36} \times \frac{5}{36}$$

We use the Additive Law of Probability:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \frac{5}{36} + \frac{5}{36} - \frac{5}{36} \times \frac{5}{36}$$

Equally likely outcomes : example

Example 2 : the birthday problem

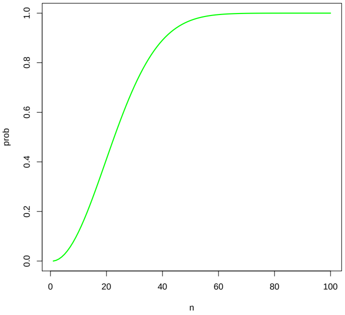
If n people are present in a room, what is the probability that no two of them celebrate their birthday on the same day of the year ?
How large need n to be so that this probability is less than 1/2 ?

Equally likely outcomes : example

We begin by calculating the probability that all birthdays are different :

$$\mathbb{P}(\text{all birthdays are different}) = \frac{\text{number of favorable cases}}{\text{number of total cases}}$$
$$= \frac{365 \times 364 \times 363 \times \cdots \times (365-n+1)}{365^n} = \frac{\binom{365}{n} n!}{365^n}$$

$$\Longrightarrow \mathbb{P}(\text{at least two have the same birthday}) = 1 - \frac{\binom{365}{n} n!}{365^n}$$



Conditional Probabilities

Conditional probabilities : definition

Sometimes probabilities need to be reevaluated as additional information becomes available

$\rightsquigarrow$ this gives rise to the notion of conditional probability

Definition

The **conditional probability** of E_1, conditional on E_2, is defined as

$$\mathbb{P}(E_1|E_2) = \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_2)} \qquad \text{(if } \mathbb{P}(E_2) > 0\text{)}$$

= probability of E_1, given that E_2 has occurred

$\rightsquigarrow$ as we know that E_2 has occurred, E_2 becomes the new sample space in the place of S

$\rightsquigarrow$ the probability of E_1 has to be calculated within E_2 and relatively to $\mathbb{P}(E_2)$

Conditional probabilities : properties

- $\mathbb{P}(E_1|E_2)$ = probability of E_1, given some extra information

  $\rightsquigarrow$ satisfies the axioms of probability

  e.g. $\mathbb{P}(S|E_2) = 1$, or $\mathbb{P}(E_1^c|E_2) = 1 - \mathbb{P}(E_1|E_2)$
- $\mathbb{P}(E_1|S) = \mathbb{P}(E_1)$
- $\mathbb{P}(E_1|E_1) = 1, \qquad \mathbb{P}(E_1|E_2) = 1$ if $E_2 \subseteq E_1$
- $\mathbb{P}(E_1|E_2) \times \mathbb{P}(E_2) = \mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_2|E_1) \times \mathbb{P}(E_1)$

Conditional probabilities : properties

Bayes' first rule

For any events E_1, E_2 with $\mathbb{P}(E_2) > 0$

$$\mathbb{P}(E_1|E_2) = \mathbb{P}(E_2|E_1) \times \frac{\mathbb{P}(E_1)}{\mathbb{P}(E_2)}$$

Conditional probabilities : properties

Multiplicative Law of Probability:

For **any** events $E_1, \ldots, E_n$ such that $\mathbb{P}(E_i) > 0$ for all $i = 1, \ldots, n$

$$\mathbb{P}\left(\bigcap_{i=1}^{n} E_i\right) = \mathbb{P}(E_1) \times \mathbb{P}(E_2|E_1) \times \mathbb{P}(E_3|E_1 \cap E_2) \times \ldots \times \mathbb{P}\left(E_n \,\middle|\, \bigcap_{i=1}^{n-1} E_i\right)$$

Example 3

A bin contains 5 defective, 10 partially defective and 25 acceptable transistors. Defective transistors immediately fail when put in use, while partially defective ones fail after a couple of hours of use. A transistor is chosen at random from the bin and put into use.

If it does not immediately fail, what is the probability it is acceptable?

Solution:

- Let E_1 be the event that transistor does not fail immediately.
- Let E_2 be the event that the selected transistor is acceptable.
- We want to calculate $\mathbb{P}(E_2|E_1)$

Using Bayes: $$\mathbb{P}(E_2|E_1) = \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_1)} = \frac{\mathbb{P}(E_1|E_2)\mathbb{P}(E_2)}{\mathbb{P}(E_1)} = \frac{1 \times 25/40}{35/40}.$$

Example 3

A bin contains 5 defective, 10 partially defective and 25 acceptable transistors. Defective transistors immediately fail when put in use, while partially defective ones fail after a couple of hours of use. A transistor is chosen at random from the bin and put into use.

If it does not immediately fail, what is the probability it is acceptable?

Solution:

- Let E_1 be the event that transistor does not fail immediately.
- Let E_2 be the event that the selected transistor is acceptable.
- We want to calculate $\mathbb{P}(E_2|E_1)$

Using Bayes: $$\mathbb{P}(E_2|E_1) = \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_1)} = \frac{\mathbb{P}(E_1|E_2)\,\mathbb{P}(E_2)}{\mathbb{P}(E_1)} = \frac{1 \times 25/40}{35/40}.$$

Example 4

A computer system has 3 users, each with a unique name and password. Due to a software error, the 3 passwords have been randomly permuted internally. Only the users lucky enough to have had their passwords unchanged in the permutation are able to continue using the system.

What is the probability that none of the three users has their password unchanged?

- Denote A =“none has their password unchanged”,
  E_i = “the ith user has their password unchanged” ($i = 1, 2, 3$).
- We are interested in finding

$$\mathbb{P}(A) = \mathbb{P}(E_1^c \cap E_2^c \cap E_3^c) = 1 - \mathbb{P}(E_1 \cup E_2 \cup E_3),$$

where

$$A^c = E_1 \cup E_2 \cup E_3 = \text{"at least one has their password unchanged"}.$$

Example 4

A computer system has 3 users, each with a unique name and password. Due to a software error, the 3 passwords have been randomly permuted internally. Only the users lucky enough to have had their passwords unchanged in the permutation are able to continue using the system.

What is the probability that none of the three users has their password unchanged?

- Denote $A =$"none has their password unchanged",
  $E_i =$ "the ith user has their password unchanged" ($i = 1, 2, 3$).
- We are interested in finding

$$\mathbb{P}(A) = \mathbb{P}(E_1^c \cap E_2^c \cap E_3^c) = 1 - \mathbb{P}(E_1 \cup E_2 \cup E_3),$$

where

$$A^c = E_1 \cup E_2 \cup E_3 = \text{"at least one has their password unchanged"}.$$

- By the **Additive Law of Probability**,

$$\begin{aligned}\mathbb{P}(E_1 \cup E_2 \cup E_3) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) \\ - \mathbb{P}(E_1 \cap E_2) - \mathbb{P}(E_1 \cap E_3) - \mathbb{P}(E_2 \cap E_3) \\ + \mathbb{P}(E_1 \cap E_2 \cap E_3).\end{aligned}$$

- Clearly, for $i = 1, 2, 3$

$$\mathbb{P}(E_i) = 1/3$$

(each user gets a password at random out of 3, including their own).

- By the **Additive Law of Probability**,

$$\begin{aligned}\mathbb{P}(E_1 \cup E_2 \cup E_3) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) \\ - \mathbb{P}(E_1 \cap E_2) - \mathbb{P}(E_1 \cap E_3) - \mathbb{P}(E_2 \cap E_3) \\ + \mathbb{P}(E_1 \cap E_2 \cap E_3).\end{aligned}$$

- Clearly, for $i = 1, 2, 3$

$$\mathbb{P}(E_i) = 1/3$$

(each user gets a password at random out of 3, including their own).

- From the **Multiplicative Law of Probability**,

$$\mathbb{P}(E_i \cap E_j) = \mathbb{P}(E_j | E_i) \times \mathbb{P}(E_i) \qquad \text{for any } i \neq j$$

Now, given E_i, that is knowing that the ith user has got their own password, there remain two passwords that the jth user may select, one of these two being their own. So

$$\mathbb{P}(E_j | E_i) = 1/2$$

and hence

$$\mathbb{P}(E_i \cap E_j) = 1/6.$$

- Likewise, given $E_1 \cap E_2$, that is knowing that the first two users have kept their own passwords, there is only one password left, the one of the third user, and

$$\mathbb{P}(E_3 | E_1 \cap E_2) = 1$$

so that (again Multiplicative Law of Probability)

$$\mathbb{P}(E_1 \cap E_2 \cap E_3) = \mathbb{P}(E_3 | E_1 \cap E_2) \times \mathbb{P}(E_2 | E_1) \times \mathbb{P}(E_1) = 1/6.$$

- Finally,

$$\mathbb{P}(E_1 \cup E_2 \cup E_3) = 3 \times 1/3 - 3 \times 1/6 + 1/6 = 2/3$$
$$\mathbb{P}(A) = 1 - \mathbb{P}(E_1 \cup E_2 \cup E_3) = 1/3.$$

- From the **Multiplicative Law of Probability**,

$$\mathbb{P}(E_i \cap E_j) = \mathbb{P}(E_j|E_i) \times \mathbb{P}(E_i) \qquad \text{for any } i \neq j$$

Now, given E_i, that is knowing that the ith user has got their own password, there remain two passwords that the jth user may select, one of these two being their own. So

$$\mathbb{P}(E_j|E_i) = 1/2$$

and hence

$$\mathbb{P}(E_i \cap E_j) = 1/6.$$

- Likewise, given $E_1 \cap E_2$, that is knowing that the first two users have kept their own passwords, there is only one password left, the one of the third user, and

$$\mathbb{P}(E_3|E_1 \cap E_2) = 1$$

so that (again Multiplicative Law of Probability)

$$\mathbb{P}(E_1 \cap E_2 \cap E_3) = \mathbb{P}(E_3|E_1 \cap E_2) \times \mathbb{P}(E_2|E_1) \times \mathbb{P}(E_1) = 1/6.$$

- Finally,

$$\mathbb{P}(E_1 \cup E_2 \cup E_3) = 3 \times 1/3 - 3 \times 1/6 + 1/6 = 2/3$$
$$\mathbb{P}(A) = 1 - \mathbb{P}(E_1 \cup E_2 \cup E_3) = 1/3.$$

- From the **Multiplicative Law of Probability**,

$$\mathbb{P}(E_i \cap E_j) = \mathbb{P}(E_j|E_i) \times \mathbb{P}(E_i) \qquad \text{for any } i \neq j$$

Now, given E_i, that is knowing that the ith user has got their own password, there remain two passwords that the jth user may select, one of these two being their own. So

$$\mathbb{P}(E_j|E_i) = 1/2$$

and hence

$$\mathbb{P}(E_i \cap E_j) = 1/6.$$

- Likewise, given $E_1 \cap E_2$, that is knowing that the first two users have kept their own passwords, there is only one password left, the one of the third user, and

$$\mathbb{P}(E_3|E_1 \cap E_2) = 1$$

so that (again Multiplicative Law of Probability)

$$\mathbb{P}(E_1 \cap E_2 \cap E_3) = \mathbb{P}(E_3|E_1 \cap E_2) \times \mathbb{P}(E_2|E_1) \times \mathbb{P}(E_1) = 1/6.$$

- Finally,

$$\mathbb{P}(E_1 \cup E_2 \cup E_3) = 3 \times 1/3 - 3 \times 1/6 + 1/6 = 2/3$$
$$\mathbb{P}(A) = 1 - \mathbb{P}(E_1 \cup E_2 \cup E_3) = 1/3.$$

Independent Events

Independence of two events

Definition

Two events E_1 and E_2 are said to be independent **if and only if**

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1) \times \mathbb{P}(E_2)$$

Note that independence implies

$$\mathbb{P}(E_1|E_2) = \mathbb{P}(E_1) \qquad \text{and} \qquad \mathbb{P}(E_2|E_1) = \mathbb{P}(E_2)$$

i.e. the probability of the occurrence of one of the event is unaffected by the occurrence or the nonoccurrence of the other

$\rightsquigarrow$ in agreement with our everyday usage of the word “independent” (“no link” between E_1 and E_2)

Caution : the ‘simplified’ multiplicative rule $\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1) \times \mathbb{P}(E_2)$ can only be used to assign a probability to $\mathbb{P}(E_1 \cap E_2)$ if E_1 and E_2 are independent, which can be known only from a fundamental understanding of the random experiment

Important property of independence

If E_1, E_2 are independent, then also:

- E_1^c, E_2 are independent
- E_1, E_2^c are independent
- E_1^c, E_2^c are independent

Proof: See Theorem 1.3.9 in the book.

Example

Example 5

We toss two fair dice, denote $E_1 =$"the sum of the dice is six", $E_2 =$"the sum of the dice is seven" and $F =$"the first die shows four".

Are E_1 and F independent? Are E_2 and F independent?

Recall that $S = \{(1,1),(1,2),(1,3),\dots,(6,5),(6,6)\}$ (there are thus 36 possible outcomes). See that, for the dice are fair,

$$E_1 = \{(1,5),(2,4),(3,3),(4,2),(5,1)\} \qquad \mathbb{P}(E_1) = 5/36$$

$$E_2 = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} \qquad \mathbb{P}(E_2) = 6/36$$

$$F = \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\} \qquad \mathbb{P}(F) = 6/36$$

$$E_1 \cap F = \{(4,2)\} \qquad \mathbb{P}(E_1 \cap F) = 1/36$$

$$E_2 \cap F = \{(4,3)\}, \qquad \mathbb{P}(E_2 \cap F) = 1/36$$

Hence, $\mathbb{P}(E_1 \cap F) \neq \mathbb{P}(E_1)\mathbb{P}(F)$ and $\mathbb{P}(E_2 \cap F) = \mathbb{P}(E_2)\mathbb{P}(F)$

$\rightsquigarrow$ E_2 and F are independent, but E_1 and F are not.

Example

Example 5

We toss two fair dice, denote E_1 ="the sum of the dice is six", E_2 ="the sum of the dice is seven" and F ="the first die shows four".

Are E_1 and F independent? Are E_2 and F independent?

Recall that $S = \{(1,1),(1,2),(1,3),\dots,(6,5),(6,6)\}$ (there are thus 36 possible outcomes). See that, for the dice are fair,

$$E_1 = \{(1,5),(2,4),(3,3),(4,2),(5,1)\} \qquad \mathbb{P}(E_1) = 5/36$$

$$E_2 = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} \qquad \mathbb{P}(E_2) = 6/36$$

$$F = \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\} \qquad \mathbb{P}(F) = 6/36$$

$$E_1 \cap F = \{(4,2)\} \qquad \mathbb{P}(E_1 \cap F) = 1/36$$

$$E_2 \cap F = \{(4,3)\}, \qquad \mathbb{P}(E_2 \cap F) = 1/36$$

Hence, $\mathbb{P}(E_1 \cap F) \neq \mathbb{P}(E_1)\mathbb{P}(F)$ and $\mathbb{P}(E_2 \cap F) = \mathbb{P}(E_2)\mathbb{P}(F)$

$\rightsquigarrow$ E_2 and F are independent, but E_1 and F are not.

Independence of more than two events

Definition

The events E_1, $E_2, \ldots, E_n$ are said to be independent **if and only if** for every subset $\{i_1, i_2, \ldots, i_r : r \leq n\}$ of $\{1, 2, \ldots, n\}$,

$$\mathbb{P}\left(\bigcap_{j=1}^{r} E_{i_j}\right) = \prod_{j=1}^{r} \mathbb{P}(E_{i_j})$$

For instance, E_1, E_2 and E_3 are independent iff

$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1) \times \mathbb{P}(E_2)$,

$\mathbb{P}(E_1 \cap E_3) = \mathbb{P}(E_1) \times \mathbb{P}(E_3)$,

$\mathbb{P}(E_2 \cap E_3) = \mathbb{P}(E_2) \times \mathbb{P}(E_3)$ and

$\mathbb{P}(E_1 \cap E_2 \cap E_3) = \mathbb{P}(E_1) \times \mathbb{P}(E_2) \times \mathbb{P}(E_3)$

Remark

Pairwise independent events need not be independent !

Example 6

Let a ball be drawn totally at random from an urn containing four balls numbered 1,2,3,4. Let $E = \{1, 2\}$, $F = \{1, 3\}$ and $G = \{1, 4\}$.

Because the ball is selected at random,
$\mathbb{P}(E) = \mathbb{P}(F) = \mathbb{P}(G) = 1/2$,
$\mathbb{P}(E \cap F) = \mathbb{P}(E \cap G) = \mathbb{P}(F \cap G) = \mathbb{P}(E \cap F \cap G) = \mathbb{P}(\{1\}) = 1/4$.

$\Longrightarrow$
$\mathbb{P}(E \cap F) = \mathbb{P}(E) \times \mathbb{P}(F)$, $\mathbb{P}(E \cap G) = \mathbb{P}(E) \times \mathbb{P}(G)$, $\mathbb{P}(F \cap G) = \mathbb{P}(F) \times \mathbb{P}(G)$,

But: $\mathbb{P}(E \cap F \cap G) \neq \mathbb{P}(E) \times \mathbb{P}(F) \times \mathbb{P}(G)$

The events E, F, G are pairwise independent, but not jointly independent

$\rightsquigarrow$ knowing that one event happened does not affect the probability of the others, but knowing that 2 events simultaneously happened does affect the probability of the third one

Example 7

Let a ball be drawn totally at random from an urn containing 8 balls numbered 1,2,3,...,8. Let $E = \{1,2,3,4\}$, $F = \{1,3,5,7\}$ and $G = \{1,4,6,8\}$.

It is clear that

- $\mathbb{P}(E) = \mathbb{P}(F) = \mathbb{P}(G) = 1/2$,
- $\mathbb{P}(E \cap F \cap G) = \mathbb{P}(\{1\}) = 1/8 = \mathbb{P}(E) \times \mathbb{P}(F) \times \mathbb{P}(G)$

But: $\mathbb{P}(F \cap G) = \mathbb{P}(\{1\}) = 1/8 \neq \mathbb{P}(F) \times \mathbb{P}(G)$

Hence, the events E, F, G are not independent, though

$$\mathbb{P}(E \cap F \cap G) = \mathbb{P}(E) \times \mathbb{P}(F) \times \mathbb{P}(G)$$

Example 8

An electric system composed of n separate components is said to be a parallel system if it functions when at least one of the components functions

1. If component i, independently of other components, functions with probability p_i, $i = 1, \ldots, n$, what is the probability the system functions?
2. What is the probability that component 1 is working, given that the system is functioning?

1) If component i, independently of other components, functions with probability p_i, $i = 1, \ldots, n$, what is the probability the system functions?

Answer:

$$\mathbb{P}(\text{the system functions}) = 1 - \mathbb{P}(\text{the system doesn't function})$$

$$\mathbb{P}(\text{the system doesn't function}) = \prod_{i=1}^{n}(1 - p_i)$$

$$\mathbb{P}(\text{the system functions}) = 1 - \prod_{i=1}^{n}(1 - p_i)$$

1) If component i, independently of other components, functions with probability p_i, $i = 1, \ldots, n$, what is the probability the system functions?

Answer:

$$\mathbb{P}(\text{the system functions}) = 1 - \mathbb{P}(\text{the system doesn't function})$$

$$\mathbb{P}(\text{the system doesn't function}) = \prod_{i=1}^{n}(1 - p_i)$$

$$\mathbb{P}(\text{the system functions}) = 1 - \prod_{i=1}^{n}(1 - p_i)$$

2) What is the probability that component 1 is working, given that the system is functioning?

Answer:

$$\mathbb{P}(\text{component 1 is working}|\text{the system functions})$$
$$= \frac{\mathbb{P}(\text{the system functions}|\text{component 1 is working})\mathbb{P}(\text{component 1 is working})}{\mathbb{P}(\text{the system functions})}$$
$$= \frac{1 \times p_1}{1 - \mathbb{P}(\text{the system doesn't function})}$$
$$= \frac{1 \times p_1}{1 - \prod_{i=1}^{n}(1 - p_i)}$$

2) What is the probability that component 1 is working, given that the system is functioning?

Answer:

$$\mathbb{P}(\text{component 1 is working}|\text{the system functions})$$
$$= \frac{\mathbb{P}(\text{the system functions}|\text{component 1 is working})\mathbb{P}(\text{component 1 is working})}{\mathbb{P}(\text{the system functions})}$$
$$= \frac{1 \times p_1}{1 - \mathbb{P}(\text{the system doesn't function})}$$
$$= \frac{1 \times p_1}{1 - \prod_{i=1}^{n}(1 - p_i)}$$

Example 9: falsely signalling a pollution problem

Many companies must monitor the effluent that is discharged from their plants in waterways. It is the law that some substances have water-quality limits that are below a limit L. The effluent is judged to satisfy the limit if every test specimen is below L.
Suppose the water does not contain the contaminant but that the variability in the chemical analysis still gives a 1% chance that a measurement on a test specimen will exceed L.

a) Find the probability that neither of two test specimens, both free of the contaminant, will fail to be in compliance

Answer: If the two samples are not taken too closely in time or space, we treat them as independent.
Denote E_i ($i = 1, 2$) the event "the sample i fails to be in compliance".
It follows

$$\mathbb{P}(E_1^c \cap E_2^c) = \mathbb{P}(E_1^c) \times \mathbb{P}(E_2^c) = 0.99 \times 0.99 = 0.9801$$

Example 9: falsely signalling a pollution problem

Many companies must monitor the effluent that is discharged from their plants in waterways. It is the law that some substances have water-quality limits that are below a limit L. The effluent is judged to satisfy the limit if every test specimen is below L.
Suppose the water does not contain the contaminant but that the variability in the chemical analysis still gives a 1% chance that a measurement on a test specimen will exceed L.

a) Find the probability that neither of two test specimens, both free of the contaminant, will fail to be in compliance

Answer: If the two samples are not taken too closely in time or space, we treat them as independent.
Denote E_i ($i = 1, 2$) the event "the sample i fails to be in compliance".
It follows

$$\mathbb{P}(E_1^c \cap E_2^c) = \mathbb{P}(E_1^c) \times \mathbb{P}(E_2^c) = 0.99 \times 0.99 = 0.9801$$

Example : falsely signalling a pollution problem

Many companies must monitor the effluent that is discharged from their plants in waterways. It is the law that some substances have water-quality limits that are below the a limit L. The effluent is judged to satisfy the limit if every test specimen is below L.
Suppose the water does not contain the contaminant but that the variability in the chemical analysis still gives a 1% chance that a measurement on a test specimen will exceed L.

b) If one test specimen is taken each week for two years (all free of the contaminant), find the probability that none of the test specimens will fail to be in compliance, and comment.

Answer: Treating the results for different weeks as independent,

$$\mathbb{P}\left(\bigcap_{i=1}^{104} E_i^c\right) = \prod_{i=1}^{104} \mathbb{P}(E_i^c) = 0.99^{104} = 0.35$$

$\rightsquigarrow$ even with excellent water quality, there is almost a two-thirds chance that at least once the water quality will be declared to fail to be in compliance with the law

Example : falsely signalling a pollution problem

Many companies must monitor the effluent that is discharged from their plants in waterways. It is the law that some substances have water-quality limits that are below the a limit L. The effluent is judged to satisfy the limit if every test specimen is below L.
Suppose the water does not contain the contaminant but that the variability in the chemical analysis still gives a 1% chance that a measurement on a test specimen will exceed L.

b) If one test specimen is taken each week for two years (all free of the contaminant), find the probability that none of the test specimens will fail to be in compliance, and comment.

Answer: Treating the results for different weeks as independent,

$$\mathbb{P}\left(\bigcap_{i=1}^{104} E_i^c\right) = \prod_{i=1}^{104} \mathbb{P}(E_i^c) = 0.99^{104} = 0.35$$

$\rightsquigarrow$ even with excellent water quality, there is almost a two-thirds chance that at least once the water quality will be declared to fail to be in compliance with the law

Example 10

The supervisor of a group of 20 construction workers wants to get the opinion of 2 of them (to be selected at random) about certain new safety regulations. If 12 workers favour the new regulations and the other 8 are against them, what is the probability that both of the workers chosen by the supervisor will be against the new regulations?

- Denote E_i ($i = 1, 2$) the event "the ith selected worker is against the new regulations". We desire $\mathbb{P}(E_1 \cap E_2)$
- **Notice**, E_1 and E_2 are not independent!
(whether first worker is against the regulations or not affects the proportion of workers against the regulations when the second one is selected)

So, $\mathbb{P}(E_1 \cap E_2) \neq \mathbb{P}(E_1)\mathbb{P}(E_2)$,

- But (by **multiplicative law of probability**)

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2|E_1) = \frac{8}{20}\frac{7}{19} = \frac{14}{95} \simeq 0.147$$

(if E_1 has occurred, then for the second selection it remains 19 workers including 7 who are against the new regulations)

Example 10

The supervisor of a group of 20 construction workers wants to get the opinion of 2 of them (to be selected at random) about certain new safety regulations. If 12 workers favour the new regulations and the other 8 are against them, what is the probability that both of the workers chosen by the supervisor will be against the new regulations?

- Denote E_i ($i = 1, 2$) the event "the ith selected worker is against the new regulations". We desire $\mathbb{P}(E_1 \cap E_2)$
- **Notice**, E_1 and E_2 are not independent!

(whether first worker is against the regulations or not affects the proportion of workers against the regulations when the second one is selected)

So, $\mathbb{P}(E_1 \cap E_2) \neq \mathbb{P}(E_1)\mathbb{P}(E_2)$,

- But (by **multiplicative law of probability**)

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2|E_1) = \frac{8}{20}\,\frac{7}{19} = \frac{14}{95} \simeq 0.147$$

(if E_1 has occurred, then for the second selection it remains 19 workers including 7 who are against the new regulations)

Bayes' Formula

Partition

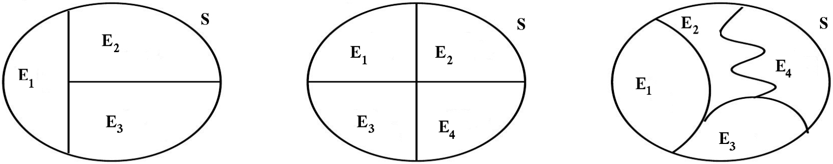
Definition

A sequence of events $E_1, E_2, \ldots, E_n$ such that

1. $S = \bigcup_{i=1}^{n} E_i$ and
2. $E_i \cap E_j = \phi$ for all $i \neq j$ (mutually exclusive),

is called a partition of S

Some examples:

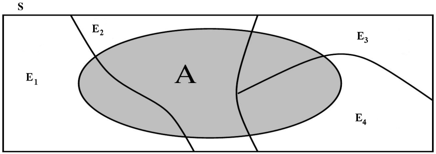


Simplest partition : $\{E, E^c\}$, for any event E

Law of Total Probability

From a partition $\{E_1, E_2, \ldots, E_n\}$, any event A can be written

$$A = (A \cap E_1) \cup (A \cap E_2) \cup \ldots \cup (A \cap E_n)$$



$$\implies \quad \mathbb{P}(A) = \mathbb{P}(A \cap E_1) + \mathbb{P}(A \cap E_2) + \ldots + \mathbb{P}(A \cap E_n)$$

Law of Total Probability

Law of Total Probability

Given a partition $\{E_1, E_2, \ldots, E_n\}$ of S such that $\mathbb{P}(E_i) > 0$ for all i, the probability of any event A can be written

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|E_i) \times \mathbb{P}(E_i)$$

- In particular, for any event A and any event E with $0 < \mathbb{P}(E) < 1$, we have

$$\mathbb{P}(A) = \mathbb{P}(A|E)\mathbb{P}(E) + \mathbb{P}(A|E^c)(1 - \mathbb{P}(E))$$

Bayes' second rule

Now, put the Law of Total Probability in Bayes' first rule and get

Bayes' second rule

Given a partition $\{E_1, E_2, \ldots, E_n\}$ of S such that $\mathbb{P}(E_i) > 0$ for all i, we have, for any event A such that $\mathbb{P}(A) > 0$,

$$\mathbb{P}(E_i|A) = \frac{\mathbb{P}(A|E_i)\mathbb{P}(E_i)}{\sum_{j=1}^{n} \mathbb{P}(A|E_j)\mathbb{P}(E_j)}$$

In particular :

$$\mathbb{P}(E|A) = \frac{\mathbb{P}(A|E)\mathbb{P}(E)}{\mathbb{P}(A|E)\mathbb{P}(E) + \mathbb{P}(A|E^c)(1 - \mathbb{P}(E))}$$

Example 11

Suppose a multiple choice test, with m multiple-choice alternatives for each question. A student knows the answer of a given question with probability p. If she does not know, she guesses.

Given that the student correctly answered a question, what is the probability that she effectively knew the answer?

Answer: Let $C =$"she answers the question correctly" and and $K =$"she knows the answer". Then, we desire $\mathbb{P}(K|C)$. We have

$$\mathbb{P}(K|C) = \mathbb{P}(C|K) \times \frac{\mathbb{P}(K)}{\mathbb{P}(C)}$$
$$= \frac{\mathbb{P}(C|K) \times \mathbb{P}(K)}{\mathbb{P}(C|K) \times \mathbb{P}(K) + \mathbb{P}(C|K^c) \times \mathbb{P}(K^c)}$$
$$= \frac{1 \times p}{1 \times p + (1/m) \times (1-p)}$$
$$= \frac{mp}{1 + (m-1)p}$$

Example 11

Suppose a multiple choice test, with m multiple-choice alternatives for each question. A student knows the answer of a given question with probability p. If she does not know, she guesses.

Given that the student correctly answered a question, what is the probability that she effectively knew the answer?

Answer: Let $C =$"she answers the question correctly" and and $K =$"she knows the answer". Then, we desire $\mathbb{P}(K|C)$. We have

$$\begin{aligned}\mathbb{P}(K|C) &= \mathbb{P}(C|K) \times \frac{\mathbb{P}(K)}{\mathbb{P}(C)} \\ &= \frac{\mathbb{P}(C|K) \times \mathbb{P}(K)}{\mathbb{P}(C|K) \times \mathbb{P}(K) + \mathbb{P}(C|K^c) \times \mathbb{P}(K^c)} \\ &= \frac{1 \times p}{1 \times p + (1/m) \times (1-p)} \\ &= \frac{mp}{1 + (m-1)p}\end{aligned}$$

Example 12: wireless communication system

Consider a wireless communication system which transmits binary symbols $S_0 = 0, S_1 = 1$, with transmission (i.e. sent) probabilities $\mathbb{P}(S_0) = 0.75, \mathbb{P}(S_1) = 0.25$

Due to the wireless channel random behaviour,
the transmitted symbol S_0 is received as S_1 with probability 0.25, and when the transmitted symbol is S_1, it is received as S_0 with prob. 0.3.

- If S_1 was received, what is the probability that S_1 was transmitted?

If S_1 was received, what is the probability that S_1 was transmitted?

Answer:

$\mathbb{P}(S_0) = 0.75, \quad \mathbb{P}(S_1) = 0.25$
$\mathbb{P}(S_1 \text{ received} | S_0 \text{ sent}) = 0.25$
$\mathbb{P}(S_0 \text{ received} | S_1 \text{ sent}) = 0.3$

Therefore,

$$\begin{aligned}
&\mathbb{P}(S_1 \text{ sent} | S_1 \text{ received}) \\
&= \frac{\mathbb{P}(S_1 \text{ received} | S_1 \text{ sent})\mathbb{P}(S_1 \text{ sent})}{\mathbb{P}(S_1 \text{ received} | S_1 \text{ sent})\mathbb{P}(S_1 \text{ sent}) + \mathbb{P}(S_1 \text{ received} | S_0 \text{ sent})\mathbb{P}(S_0 \text{ sent})} \\
&= \frac{0.7 \times 0.25}{0.7 \times 0.25 + 0.25 \times 0.75} = 0.48
\end{aligned}$$

Not a very good detection probability...

If S_1 was received, what is the probability that S_1 was transmitted?

Answer:

$\mathbb{P}(S_0) = 0.75, \quad \mathbb{P}(S_1) = 0.25$
$\mathbb{P}(S_1 \text{ received}|S_0 \text{ sent}) = 0.25$
$\mathbb{P}(S_0 \text{ received}|S_1 \text{ sent}) = 0.3$

Therefore,

$$\begin{aligned}
&\mathbb{P}(S_1 \text{ sent}|S_1 \text{ received}) \\
&= \frac{\mathbb{P}(S_1 \text{ received}|S_1 \text{ sent})\mathbb{P}(S_1 \text{ sent})}{\mathbb{P}(S_1 \text{ received}|S_1 \text{ sent})\mathbb{P}(S_1 \text{ sent}) + \mathbb{P}(S_1 \text{ received}|S_0 \text{ sent})\mathbb{P}(S_0 \text{ sent})} \\
&= \frac{0.7 \times 0.25}{0.7 \times 0.25 + 0.25 \times 0.75} = 0.48
\end{aligned}$$

Not a very good detection probability...

Objectives

Now you should be able to :

- understand and describe sample spaces and events for random experiments ☐
- interpret probabilities and use probabilities of outcomes to calculate probabilities of events ☐
- use permutations and combinations to count the number of outcomes in both an event and the sample space ☐
- calculate the probabilities of joint events such as unions and intersections from the probabilities of individual events ☐
- interpret and calculate conditional probabilities of events ☐
- determine whether events are independent and use independence to calculate probabilities ☐
- use Baye's rule(s) to calculate probabilities ☐

Put yourself to the test !

$\rightsquigarrow$ Q1.1 p.37, Q1.5 p.38, Q33.1 p.41, Q1.55 p.44