Statistics

MATH2089

THE UNIVERSITY OF NEW SOUTH WALES

Semester 1, 2018 – Lecture 3

This lecture

3. Elements of Probability

Additional reading: Sections 5.1, 5.2 and 5.3 in the textbook

3. Elements of Probability

Introduction

The previous chapter (Chapter 2) described purely descriptive methods for a given sample of data.

The subsequent chapters (Chapters 6-12) will describe **inferential methods**, that convert information from random samples into information about the whole population from which the sample has been drawn.

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

$\Rightarrow$ drawing conclusions about the whole population, thus going beyond what we have observed, inherently involves some risk

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

Introduction

- It is important to keep in mind the crucial role played by random sampling (Lecture 1)
- Without random sampling, statistics can only provide descriptive summaries of the observed 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 to be interpreted as "chaotic" or "haphazard". It 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.

$\Rightarrow$ Probability theory (Chapters 3-5) is a necessary link between descriptive and inferential statistics

Random experiment

Definition

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

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

... and obviously the "random sampling" experiment

Experiment 6: select a random sample of size n from a population

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.

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: $S = [0, +\infty)$
Experiment 6: $S = \{\text{sets of } n \text{ individuals out of the population}\}$

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.

Events

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

Events

The elements of interest are the events, which are (sub)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" $(= E')$

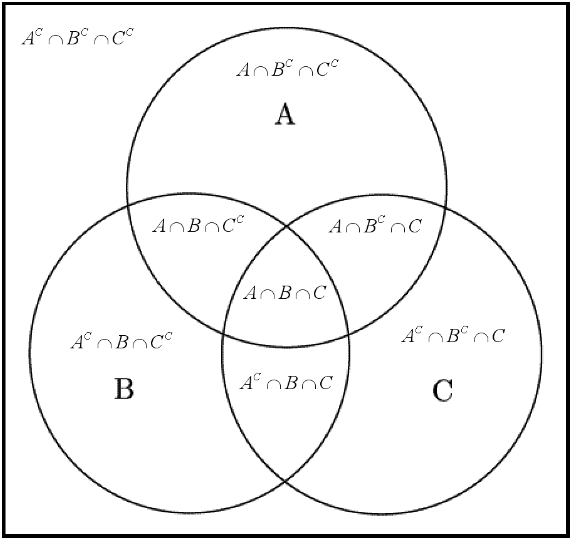
$E_1 \subseteq E_2 \Rightarrow E_1$ implies E_2

$E_1 \cap E_2 = \phi \Rightarrow$ mutually exclusive events (they cannot occur together)

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 can be clearly illustrated by means of Venn diagrams.

Venn diagrams



The axioms of probability theory

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

how likely E is to occur

Firm mathematical footing $\Rightarrow$ Kolmogorov's axioms (1933)

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

- For any finite sequence of mutually exclusive events $E_1, E_2, \dots, E_n$,

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

- $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$
- $\mathbb{P}(\phi) = 0$
- $E_1 \subseteq E_2 \Rightarrow \mathbb{P}(E_1) \leq \mathbb{P}(E_2)$ (increasing measure)
- $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_2)$, and by induction:
- Additive Law of Probability (or inclusion/exclusion principle)

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

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.

Experiment 1 (ctd.)

- experiment: flipping a coin $\Rightarrow S = \{H, T\}$
- Axioms: $\begin{cases} i) & 0 \leq \mathbb{P}(H) \leq 1, 0 \leq \mathbb{P}(T) \leq 1 \\ ii) & \mathbb{P}(S) = \mathbb{P}(H \cup T) = 1 \\ iii) & \mathbb{P}(H \cup T) = \mathbb{P}(H) + \mathbb{P}(T) \end{cases}$

$\Rightarrow$ The axioms only state that $\mathbb{P}(H)$ and $\mathbb{P}(T)$ are two non-negative numbers such that $\mathbb{P}(H) + \mathbb{P}(T) = 1$, nothing more!

The exact values of $\mathbb{P}(H)$ and $\mathbb{P}(T)$ depend on the coin itself (fair, biased, fake).

Assigning probabilities

To effectively assign probabilities to events, different approaches can be used, the most widely held being the **frequentist approach**.

Frequentist definition of probability

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

Let n be the number of repetitions of the experiment. Then, the probability of the event E is

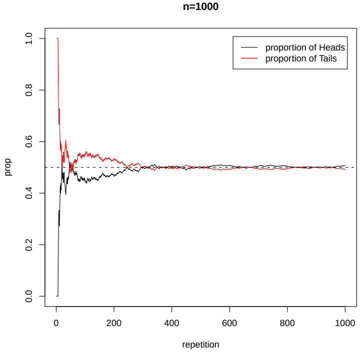
$$\mathbb{P}(E) = \lim_{n\to\infty} \frac{\text{number of times } E \text{ occurs}}{n}$$

Assigning probabilities: a simple example

Experiment 1 (ctd.)

- experiment: flipping a coin $\Rightarrow S = \{H, T\}$
- Axioms: $\begin{cases} i) & 0 \leq \mathbb{P}(H) \leq 1, 0 \leq \mathbb{P}(T) \leq 1 \\ ii) & \mathbb{P}(S) = \mathbb{P}(H \cup T) = 1 \\ iii) & \mathbb{P}(H \cup T) = \mathbb{P}(H) + \mathbb{P}(T) \end{cases}$

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



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

(fair coin, in this case)

Assigning probabilities

Interpretation

probability $\simeq$ proportion of occurrences of the event

- It is straightforward to check that the so-defined ‘frequentist’ probability measure satisfies the axioms
- Of course, this definition remains theoretical, as assigning probabilities would require infinitely many repetitions of the experiment
- 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?)

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

- A simple model assumes that all the outcomes are equally likely, other more elaborated models define probability distributions

($\rightsquigarrow$ Chapter 5)

Assigning probabilities: equally likely outcomes

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

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

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

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

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

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

$\Rightarrow$ "Classical" definition of probability

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

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

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

- **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

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

- **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

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

Equally likely outcomes: example

Example

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 $\mathbb{P}(A \cup B)$?

Equally likely outcomes: example

Example: the birthday problem

If n people are present in a room, what is the probability that at least two of them celebrate their birthday on the same day of the year ? How large need n to be so that this probability is more than 1/2 ?

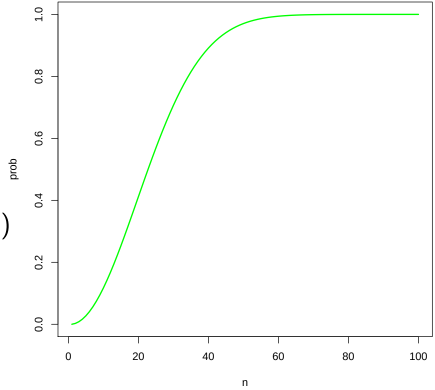
We have:

$$\mathbb{P}(\text{all birthdays are different}) = \frac{\binom{365}{n}\, n!}{365^n},$$

so that

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

$\Rightarrow$ Prob $> 1/2 \iff n > 23$



Conditional probabilities: definition

Sometimes probabilities need to be re-evaluated as additional information becomes available

$\Rightarrow$ this gives rise to the concept 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

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

$\Rightarrow$ The probability of E_1 has to be calculated within E_2 and relative to $\mathbb{P}(E_2)$

Conditional probabilities: properties

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

  $\rightarrow$ 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 $E_2 \subseteq E_1$
- $\mathbb{P}(E_1|E_2) \times \mathbb{P}(E_2) = \mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_2|E_1) \times \mathbb{P}(E_1)$

  $\Rightarrow$ Bayes' first rule: if $\mathbb{P}(E_1) > 0$ and $\mathbb{P}(E_2) > 0$,

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

- Multiplicative Law of Probability:

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

Example

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?

Example

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 kept their original password?

Denote $A =$"no user kept their original password", and $E_i =$ "the ith user has the same password" ($i = 1, 2, 3$). See that

$$A^c = E_1 \cup E_2 \cup E_3,$$

for $A^c =$ at least one user has kept their original password. 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

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

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,

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

and

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

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 non-occurrence of the other

$\rightarrow$ in agreement with 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.

Example

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 $S = \{(1,1),(1,2),(1,3),\ldots,(6,5),(6,6)\}$ (there are thus 36 possible outcomes).

$$
\begin{aligned}
E_1 &= \{(1,5),(2,4),(3,3),(4,2),(5,1)\} & \mathbb{P}(E_1) &= 5/36 \\
E_2 &= \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} & \mathbb{P}(E_2) &= 6/36 \\
F &= \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\} & \mathbb{P}(F) &= 6/36 \\
E_1 \cap F &= \{(4,2)\} & \mathbb{P}(E_1 \cap F) &= 1/36 \\
E_2 \cap F &= \{(4,3)\}, & \mathbb{P}(E_2 \cap F) &= 1/36
\end{aligned}
$$

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

$\Rightarrow$ 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 iff 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 jointly independent !

Example

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, $\mathbb{P}(E) = \mathbb{P}(F) = \mathbb{P}(G) = 1/2$, and

$$\mathbb{P}(E \cap F) = \mathbb{P}(E \cap G) = \mathbb{P}(F \cap G) = \mathbb{P}(E \cap F \cap G) = \mathbb{P}(\{1\}) = 1/4.$$

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

$$\text{but } \mathbb{P}(E \cap F \cap G) \neq \mathbb{P}(E) \times \mathbb{P}(F) \times \mathbb{P}(G)$$

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

$\Rightarrow$ 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

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 $\mathbb{P}(E) = \mathbb{P}(F) = \mathbb{P}(G) = 1/2$, and

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

but

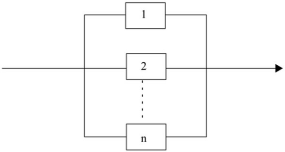
$$\mathbb{P}(F \cap G) = \mathbb{P}(\{1\}) = 1/8 \neq \mathbb{P}(F) \times \mathbb{P}(G)$$

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

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. For such a system, if component i, **independently** of other components, functions with probability p_i, $i = 1, \ldots, n$, what is the probability the system functions?



Define the events W = the system functions and W_i = component i functions

Then, $\mathbb{P}(W^c) = \mathbb{P}(W_1^c \cap W_2^c \cap \ldots \cap W_n^c) = \prod_{i=1}^{n} \mathbb{P}(W_i^c) = \prod_{i=1}^{n}(1 - p_i)$, hence

$$\mathbb{P}(W) = 1 - \prod_{i=1}^{n}(1 - p_i)$$

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 some 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

If the two samples are not taken too closely in time or space, we can 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 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.

Treating the results for different weeks as independent,

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

→ 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

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

However, E_1 and E_2 are not independent! (whether the 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 the 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)

Partition

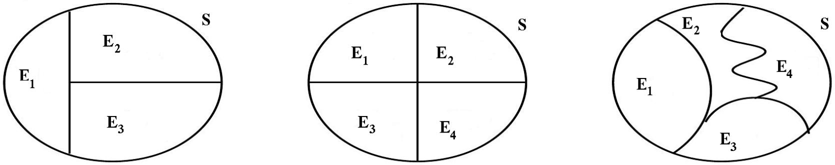
Definition

A sequence of events E_1, E_2, ..., 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:

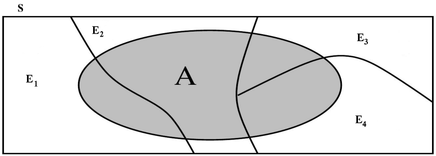


Simplest partition is $\{E, E^c\}$, for any event E

Law of Total Probability

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

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



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

Law of Total Probability

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

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

Bayes' second rule

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

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

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

A new medical procedure has been shown to be effective in the early detection of an illness and a medical screening of the population is proposed. The probability that the test correctly identifies someone with the illness as positive is 0.99, and the probability that someone without the illness is correctly identified by the test is 0.95. The incidence of the illness in the general population is 0.0001. You take the test, and the result is positive. What is the probability that you have the disease?

Example

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?

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 □

Recommended exercises

→ Q1, Q2 p.197, Q8, Q9 p.203, Q12 p.209, Q17 p.210, Q19 p.210, Q20 p.210, Q59 p.238, Q61 p.238, Q73 p.240 (2nd edition)

→ Q1, Q2 p.200, Q8, Q9 p.207, Q13 p.213, Q18 p.214, Q20 p.214, Q21 p.214, Q61 p.242, Q63 p.243, Q76 p.244 (3rd edition)