Victoria University of Wellington
School of Mathematics and Statistics
Te Kura Tatau

Probability Review
for OPRE354/COMP312
2017 Trimester 1/3

Lecturer Stefanka Chukova (course coordinator)
Cotton 535
Stefanka.Chukova@sms.vuw.ac.nz

"In probability theory we consider some underlying process/event that has some randomness or uncertainty modeled by random variables, and we aim to figure out what happens. In statistics we observe something that has happened, and try to figure out what underlying process/event would explain those observations."

Contents

Chapter 1

Probability

1.1 Introduction

Many physical phenomena can be described or predicted by a mathematical model. For example, the velocity v (m/s) of a body falling freely for t seconds in a vacuum is $v = gt$, where $g = 9.81$ m/s^2.

Ideally, repeated trials under identical conditions would produce the same result predicted by the model. This is the essential property of a **deterministic model**.

There are other phenomena in which the results seem to occur by chance. For example, the number of road deaths in a weekend, the outcome of a single game of roulette, the time between successive arrivals at a service counter. These events are either naturally unpredictable, or too complicated to describe by a deterministic model. The mathematical models which may be used to describe these situations of uncertainty are called **probability models** or **stochastic models** (from Greek *stochos* to guess) and, in this respect, Probability and Statistics might well be called the Mathematics of Uncertainty.

1.2 The Language of Probability

Concepts to do with chance, or unpredictable variation, have always been difficult to grasp. Historically they evolved from two processes, the analysis of games of chance (including insurance), and the analysis of sets of observations affected by errors (initially in astronomical data). From games of chance, a numerical measure of probability came into being. It was simply the ratio of the number of favourable outcomes to the total number of possible outcomes for a particular game. However, this measure assumes that each individual outcome is equally likely (equally probable). So it cannot be used as a definition of probability because the inclusion of the equally likely assumption would make the definition "circular". From the analyses of both games of chance and measurement errors, empirical evidence was also gathered indicating that relative frequencies appeared to stabilise when experiments were repeated many times under identical conditions. The notion of an empirical probability being equal to the proportion of occasions on which an event was observed then led to the idea of probability being some sort of limiting relative frequency. Unfortunately, it is not possible to conduct experiments indefinitely, nor is it possible to prove what such limits should be.

The concepts (if not quite the definition) of probability evolved as idealisations of issues such as these.

We start with the reasonable assumption that probability must be defined relative to a situation (called a *random experiment*) that has an uncertain outcome. We will assume that the set of all possible outcomes can be completely specified before such an experiment is undertaken.

Definition A *sample space* is a set of elements in one-one correspondence with the set of all possible outcomes of the situation of uncertainty being modelled. An element in a sample space is called a *sample point*.

The *sample space* S may be a listing of the set of all possible outcomes to the experiment.

Classical gambling games using coins, six-sided dice, and decks of 52 playing cards provide simple examples (mahjong would work well too, but would provide more complicated examples). For instance, most dice have six faces, each containing a unique number of dots (from 1 to 6). We can make up many probability problems with a die (the singular version of the word) or dice (more than one), for instance:

Example 1.1

(a) A single die is thrown and the result recorded.

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

(b) Two dice are thrown, and the result is recorded.

There are six possible outcomes for each die, so that, in all, there are $6 \times 6 = 36$ possible outcomes for the two together. Notice that we count a 1 on Die 1 and a 6 on Die 2 as distinct from 6 on Die 1 and 1 on Die 2.

The sample space is:

$$\begin{array}{rllllll} S = \{ & (1,1), & (1,2), & (1,3), & (1,4) & (1,5), & (1,6) \\ & (2,1), & (2,2), & \multicolumn{4}{l}{\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots} \\ & & \multicolumn{2}{l}{\dots\dots\dots\dots ,} & (6,4), & (6,5), & (6,6)\} \end{array}$$

(c) Two dice are thrown and the two results are added, and this total is recorded. This is the same as adding each pair of results in (b) above. The sample space for this experiment will have fewer elements than the sample space in (b) as we record only the possible totals, most of which can be obtained in more than one way (for instance, $1 + 3 = 3 + 1 = 2 + 2 = 4$).

The sample space is $S = \{2,\ 3,\ 4,\ \ldots,\ 10,\ 11,\ 12\}$ (it has 11 elements).

Example 1.2 An experiment involves tossing two coins and noting which sides are uppermost. Altogether there are 4 possible outcomes, depending on which way up each coin lands. A particular sample space is $S = \{$HH, HT, TH, TT$\}$, where the symbols H (heads) and T (tails) are used to identify the possible results for the first and second coins.

Alternatively, we might write $S = \{(1,\ 1),\ (1,\ 0),\ (0,\ 1),\ (0,\ 0)\}$ where "1" denotes a head and "0" denotes a tail. In either case S is a set of elements in one-one correspondence with the set of all possible outcomes.

Definition Any subset of the sample space S is called an *event*. An event A is said to *occur* if the observed outcome corresponds to an element of A.

Example 1.3 In Example 1.2 about tossing two coins, if A denotes the event of observing "at least one tail", then
$A = \{$HT, TH, TT$\}$.

In Example 1.1, if B denotes the event of an even number of spots on a single die, $B = \{2,\ 4,\ 6\}$; if C denotes the event of an equal number of spots on the two dice, then $C = \{(1,1),\ (2,2),\ (3,3),\ \ldots,\ (6,6)\}$; if D denotes the event of observing the sum of the spots on two dice to be greater than 6, then $D = \{7, 8, 9, 10, 11, 12\}$.

Definition An event defined by a subset consisting of a single element of S is called an *elementary event*.

The different *sample points* can be denoted by $s_1,\ s_2,\ \ldots,$ written $s_1 \in S$, $s_2 \in S\ \ldots$, and the elementary events would be $\{s_1\}$, $\{s_2\}$, ...

The elementary events of our coin-tossing sample space are $\{$HH$\}$, $\{$HT$\}$, $\{$TH$\}$ and $\{$TT$\}$.

The whole set S is sometimes called the *certain* event (because one of the outcomes of S must occur if the list of possible outcomes is complete).

The empty set $\emptyset$ is the *impossible* event.

A sample space S is said to be *finite* if it consists of a finite number of outcomes, like the coin-tossing example above.

Often there is no upper limit to the number of outcomes in S. The sample space is said to be *countably infinite* if the outcomes can be put into a one-to-one correspondence with the positive integers. For example, if a coin is tossed repeatedly until a head appears, then a possible sample space is $S = \{$H, TH, TTH, TTTH, ...$\}$. In principle, there is no 'last' element in this sample space. The element with n Ts is the $(n+1)$th element in the sequence, so it is clear that this infinite set is in one-to-one correspondence with the positive integers.

Definition If a sample space S is either finite or countably infinite, then it is called a *discrete sample space*.

In general, we say that a set is *countable* if it is finite or countably infinite. Otherwise it is *uncountable*. For example, the unit interval of real numbers, $\{x : 0 \leq x \leq 1\}$, is uncountable, as are all continuous intervals of real numbers.

We look now at the actual events on the sample space, and some relationships between events.

If we consider two events, one of two things must be true. Either the two events have at least one elementary event in common, or they have no events in common. To use our coin example again, we can define an event A, the event that there is at least one head. A is made up of the elementary events {HH}, {TH} and {HT}, so that we can write $A =$ {HH, HT, TH}. We can define the event B to be the event that both outcomes are the same. Then B is made of the elementary events {HH} and {TT}, or $B =$ {HH, TT}. We can see that events A and B have the elementary event {HH} in common.

But if we define event C to be the event that the two coins are different, then $C =$ {HT, TH}. Events A and C have the elementary events {HT} and {TH} in common, but events B and C have no elementary events in common. Using set notation, we can write $B \cap C = \emptyset$.

Definition Two events A and B are called *mutually exclusive* if $A \cap B = \emptyset$.

Example 1.4

(a) In the single die example, outcomes {1} and {2} are mutually exclusive.

(b) In the coin-tossing examples the outcomes (for a single coin) {H} and {T} are mutually exclusive.

The idea of mutually exclusive events can be extended to cover more than two events.

Definition Events A_1, A_2, A_3, ..., are said to be *mutually exclusive* if they are pairwise mutually exclusive. That is, if $A_i \cap A_j = \emptyset$ whenever $i \neq j$.

To discuss relationships among events, which are subsets of the sample space, the notation and methods of set theory are needed. The interpretation in terms of a random experiment gives this notation an intuitive content which is richer than the bare set theory content and for which special probability language has evolved.

Operation	Set Language	Probability Language
$A \subset B$	A is contained in B	A implies B
$A \cup B$	Union of A and B (points in *either A or B* (or both))	Either A or B occurs (or both)
$A \cap B$	Intersection of A and B (points in *both A and B*)	Both A and B occur
$\bar{A}$ (or A')	A complement (points in the sample space not included in A)	A does not occur

Example 1.5 If $A = \{2, 4, 6, 8\}$ and $B = \{4, 8\}$ then $B \subset A$.

Note that the same event A could equally validly have been defined in a number of other ways, for instance $A = \{ x : \ x = 2k$, where $k = 1, 2, 3, 4\}$ or $A = \{8, 4, 6, 2\}$ (order does not matter).

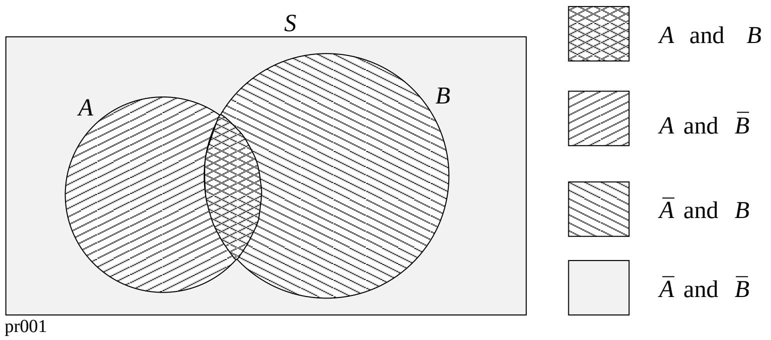
Note also that $2 \in A$ but $\{2\} \subset A$ — elements need not be sets, but subsets must be.

If the outcome of an event is a real number, we often use interval notation to represent the sets of interest. Remember that, if a and b are real numbers we can write

$$(a, b) \text{ for } \{ x : a < x < b\}$$
$$(a, b] \text{ for } \{ x : a < x \leq b\}$$
$$[a, b] \text{ for } \{ x : a \leq x \leq b\}$$

Relations between events can be easily represented by drawing a Venn diagram.

Example 1.6 $A \cup B = (A \cap \bar{B}) \cup (A \cap B) \cup (\bar{A} \cap B)$



The union of the three disjoint shaded areas (in the circles representing sets A and B) makes up the whole of $A \cup B$ and the union of all four disjoint shaded areas (including the whole of S that is outside the circles) makes up the whole of S.

We can also write

$$\begin{aligned} A \cup B &= \{x : x \in A \text{ or } x \in B\} \\ A \cap B &= \{x : x \in A \text{ and } x \in B\} \\ \bar{A} &= \{x : x \notin A,\ x \in S\} && \text{(the complement)} \\ A - B \text{ or } A \setminus B = A \cap \bar{B} &= \{x : x \in A,\ x \notin B\} && \text{(the difference)} \end{aligned}$$

1.3 Properties of Probability

Probability is a numerical measure of chance. Assigning probabilities to random events has been quite controversial. Two ways of assigning probability that we do not discuss in any depth are those of the frequentists (see also the first paragraph of Section 1.2) and subjectivists:

Frequentists regard probability as a limiting relative frequency. For example,

$$\mathrm{Prob}[\text{Heads on } \textit{one} \text{ toss of a coin}] = \lim_{n\to\infty} \frac{\text{Number of heads } (f)}{\text{Number of tosses } (n)}.$$

Tossing a coin *is* a repeatable experiment. But it is impossible to repeat an experiment indefinitely. And it is impossible to predict what such a limit should be.

However, it has been found experimentally that relative frequencies do seem to stabilise after many trials.

Subjectivists regard probability as a measure of “degree of belief”. A number between zero and one, it can be assigned in any situation of uncertainty without assuming repeatability.

Naturally, the results of a repeatable experiment would influence a subjectivist's “degree of belief”. (If 50 tosses of a coin produced 50 Heads, who could believe Heads and Tails were equi-likely?)

Definition Let A be an event, then the probability of A, denoted by $P(A)$, satisfies the following *axioms* (which we call P1, P2, P3 for future reference):

P1 $P(A) \geq 0$ for every A.

P2 $P(S) = 1$.

P3 If A_1, A_2, ... are a sequence of mutually exclusive events (i.e. no two have any sample points in common) then

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots .$$

Example 1.7 We may gamble on getting an even number on a single throw of a die. The winning event is a $\{2\}$, or a $\{4\}$, or a $\{6\}$, that is, the *union* of three elementary events. If we get a $\{2\}$, we can*not* get a $\{4\}$. Clearly, elementary events are mutually exclusive. If the probability of each of the 6 possible outcomes is $\frac{1}{6}$, then the probability of winning is $\frac{1}{6} + \frac{1}{6} + \frac{1}{6}$ (by P3), which is $\frac{1}{2}$.

Of these three axioms, P3 is the one with teeth in it, that carries the important consequences. If, for example, we can assign probabilities to all the elementary events of S, then since these are mutually exclusive, we can obtain probabilities for all other subsets of S.

The second axiom is a normalisation condition (in other words we specify that the total probability is 1) and the first states a preference for working with non-negative numbers (again, this agrees with our intuition).

Example 1.8 Suppose one of the questions on a questionnaire asks subjects to indicate into which age category they fall, where the six possible categories are 0–4, 5–14, 15–24, 25–44, 45–64, ≥ 65. The numbers in the categories are as follows:

Age	0–4	5–14	15–24	25–44	45–64	≥ 65
f	37	44	125	102	95	66

Suppose we call the event that a subject is aged 0–4 A_1, the event that a subject is aged 5–14 A_2, ... We can then extend the table to get:

Event	Age group	Frequency f_i	Relative Frequency p_i
A_1	0–4	37	$37/469 = 0.08$
A_2	5–14	44	$44/469 = 0.09$
A_3	15–24	125	$125/469 = 0.27$
A_4	25–44	102	$102/469 = 0.22$
A_5	45–64	95	$95/469 = 0.20$
A_6	≥ 65	66	$66/469 = 0.14$
S	Total:	469	1.00

P1, that $P(A_i) = p_i \geq 0$ for every A_i is obviously satisfied (see the rightmost column).

P2 that $P(S) = \sum_{i=1}^{6} p_i = 1$ is also satisfied (the entries in the p_i column sum to 1).

All our "events" are mutually exclusive, so can use P3 to establish the proportion of the sample that was, for instance, "young" (ages ≤ 24), or "probably in the work force" (ages 15–64):
$P(\text{"young"}) = P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) = 0.08 + 0.09 + 0.27 = 0.44.$
$P(\text{"working"}) = P(A_3 \cup A_4 \cup A_5) = P(A_3) + P(A_4) + P(A_5) = 0.27 + 0.22 + 0.20 = 0.69.$

The following immediate deductions from the axioms are important.

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

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

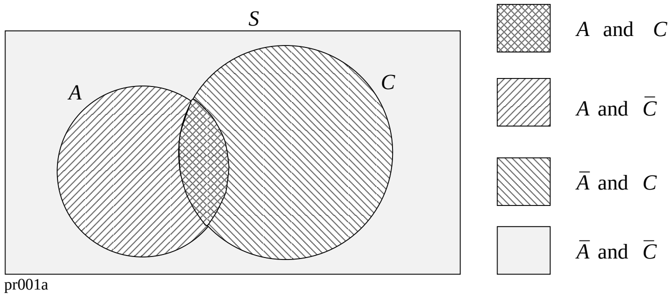
(iii) If $A \subseteq B$ then $P(A) \leq P(B).$

(iv) $P(A) = P(A \cap B) + P(A \cap \bar{B})$ This very useful result is known as the Law of Total Probability.

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

Example 1.9 Students at a university (like VUW) can register for one or more degrees (if they do a double degree). Let A be the event that a student is registered for an Arts degree and C be the event that they are registered for a Commerce degree. Suppose we know that 45% of the students are registered for an Arts degree and 50% are registered for a Commerce degree ($P(A) = 0.45$, $P(C) = 0.5$).

Problems such as this one, where amongst the probabilities of interest are $P(A)$, $P(C)$, $P(A \cup C)$ and $P(A \cap C)$ are usually best represented by a Venn diagram.



In order to solve any of the possible problems, we need at least 3 pieces of information. What pieces of information could we have? If we look at our Venn diagram, we see that the sample space is made up of 4 mutually exclusive events: $\bar{A} \cap \bar{C}$, $A \cap \bar{C}$, $\bar{A} \cap C$ and $A \cap C$.

Sometimes these events are of interest in their own right (the probability that a student is doing neither Arts nor Commerce; the probability a student is doing a double degree in Arts and Commerce), and sometimes it's events that are the union of two or more of the mutually exclusive events that are of interest (the probability of either Arts or Commerce or both; the probability of Arts (or both); the probability of Commerce (or both)).

In our example, we know $P(A) = 0.45$ and $P(C) = 0.5$. Let's suppose, in turn, that we're given all the possible different third pieces of information:

(a) Firstly, suppose $P(A \cap C) = 0.20$ (the probability of a student doing a double degree in Arts and Commerce). For practice, we now find all possible probabilities:

   (i) The probability a student is doing an Arts degree or a Commerce degree (or both) is
   $$\begin{aligned} P(A \cup C) &= P(A) + P(C) - P(A \cap C) \\ &= 0.45 + 0.5 - 0.2 = 0.75 \end{aligned}$$

   (ii) The probability a student is registered only for an Arts degree is:
   $$\begin{aligned} P(A \cap \bar{C}) &= P(A) - P(A \cap C) \\ &= 0.45 - 0.2 = 0.25 \end{aligned}$$

   (iii) The probability a student is registered only for a Commerce degree is:
   $$\begin{aligned} P(\bar{A} \cap C) &= P(C) - P(A \cap C) \\ &= 0.5 - 0.20 = 0.3 \end{aligned}$$

   (iv) The probability a student is registered for neither of these degrees is:
   $$\begin{aligned} P(\bar{A} \cap \bar{C}) &= 1 - P(A \cup C) \\ &= 1 - 0.75 = 0.25 \end{aligned}$$

   (v) The probability a student is registered for an Arts degree, or a Commerce degree, but not both, is:

$$\begin{aligned} P\big((A\cap\bar{C})\cup(\bar{A}\cap C)\big) &= P(A\cap\bar{C})+P(\bar{A}\cap C)\\ &= 0.25+0.3=0.55\\ \text{or} &= P(A\cup C)-P(A\cap C) \qquad \text{(the difference)}\\ &= 0.75-0.2=0.55 \end{aligned}$$

(b) Now suppose we were given $P(A\cup C)=0.75$. We again find all possible probabilities:

(i) $P(A\cap C)=P(A)+P(C)-P(A\cup C)$ (rearranging formula for $P(A\cup C)$)
$= 0.45+0.5-0.75=0.2$

(ii) $P(A\cap\bar{C})=P(A)-P(A\cap C)$ as before.

(iii) $P(\bar{A}\cap C)=P(C)-P(A\cap C)$ as before.

(iv) $P(\bar{A}\cap\bar{C})=1-P(A\cup C)$ as before.

(v) $P\big((A\cap\bar{C})\cup(\bar{A}\cap C)\big)=P(A\cup C)-P(A\cap C)$ as before.

(c) Next suppose we were given that $P(A\cap\bar{C})=0.25$ (Arts only).

(i)
$$\begin{aligned} P(A) &= P(A\cap C)+P(A\cap\bar{C}) \qquad \text{(Law of Total Prob.)}\\ \therefore\quad P(A\cap C) &= P(A)-P(A\cap\bar{C})\\ &= 0.45-0.25=0.2 \end{aligned}$$

(ii) . . . and the other probabilities can be calculated as before.

(d) Similarly, if we are given $P(\bar{A}\cap C)=0.3$, we can work out that

$$\begin{aligned} P(C) &= P(A\cap C)+P(\bar{A}\cap C) \qquad \text{(Law of Total Prob.)}\\ \therefore\quad P(A\cap C) &= P(C)-P(\bar{A}\cap C)\\ &= 0.5-0.3=0.2, \qquad \text{and again continue as before.} \end{aligned}$$

Thus, in a problem of this general type, given any three pieces of information about two events, we can find any other probability that can be expressed in terms of the two events. The principle can be extended to three or more events (it's just a bit messier as the number of basic mutually exclusive events is much larger).

It is all very well to have rules for handling probabilities, but how do we get the probabilities initially? When the experiment has only a finite number of outcomes, the essential task is to find the probabilities of each elementary event (outcome). The probabilities of other events can then be found by adding together the probabilities of the elementary events that they contain.

Example 1.10 **(Equally likely events)** In the simplest special case, all elementary events are equally likely, so that if there are N of them each has probability $\frac{1}{N}$. This links to the well known recipe

$$P(A)=\frac{\text{no. of outcomes in }A}{\text{total no. of outcomes}}=\frac{n(A)}{n(S)}= \text{ proportion of outcomes in } A.$$

1.4 Calculating Probabilities

1.4.1 Counting Problems

Even in the simplest case of equally likely outcomes, the problem of counting the number of outcomes "in favour" of a given event quickly gets complicated if there are many possible outcomes. The following two rules are useful for solving counting problems.

Rule 1: (Addition Rule) *If a given event can be broken down into a number of mutually exclusive sub-events, whose union fills out the whole of the original event, then the number of outcomes in favour of the original event is the sum of the numbers of outcomes in favour of each of the subevents.*

Rule 2: (Multiplication Rule) *If the outcomes leading to a certain event can be described in terms of stages which have to be undertaken in order, then the number of outcomes in the final event is the number of ways to produce the first stage times the number of ways to produce the second stage, etc.*

Many problems involving choosing objects "at random" can be answered more or less directly by appeal to these two rules, assuming that elementary outcomes are equally likely.

Example 1.11

(a) A motor car registration plate is characterised by a combination of two letters and four digits. How many distinct registration plates are possible? How many of these have all letters and numbers different?

(b) Assuming all choices are equally likely, what is the probability that a registration plate chosen at random has all numbers and letters different?

Answer

(a) $26 \times 26 \times 10 \times 10 \times 10 \times 10 = 6,760,000$ different possible registration plates.
$26 \times 25 \times 10 \times 9 \times 8 \times 7 = 3,276,000$ have all different entries.

(b) Prob (all different) $= \dfrac{3,276,000}{6,760,000} = 0.4846$

Example 1.12 How many tosses of 3 dice result in at least one 6?
Answer First we consider a subdivision into three subevents:

$$\begin{aligned} A_1 &= \text{exactly 1 six} \\ A_2 &= \text{exactly 2 sixes} \\ A_3 &= \text{exactly 3 sixes} \end{aligned}$$

We find the number of ways for each of these using rule 2 and then add them using rule 1.

Number of ways for A_3: $1 \times 1 \times 1 = 1$ way.

Number of ways for A_2: If the first two dice are sixes, then the third could be any number other than 6, so there are 1 × 1 × 5 = 5 possibilities. The same is true if the first and third are sixes, or the second and third are sixes, resulting in 15 possibilities altogether.

Number of ways for A_1: If the first dice is a six, the other two values can each be chosen in 5 ways, resulting in 1 × 5 × 5 = 25 possibilities. Again there are three cases, according to which dice is the six, so 75 altogether.

Hence the total number of ways is 1 + 15 + 75 = 91.

[Alternatively: Ans = (total no. of possibilities for 3 dice) - (those with no sixes) = 6^3 - 5^3 = 216 - 125 = 91.]

These results lead immediately to corresponding probabilities, assuming all outcomes are equally likely:
Prob (at least 1 six) = 1 − Prob(no sixes) = $1 - \left(\frac{5}{6}\right)^3 = \frac{91}{216}$.

1.4.2 Permutations and Combinations

Suppose we have a collection of n distinct objects (cards, balls in an urn, members of a population) and wish to select $k \leq n$ of these into a sample. How many ways are there of doing this?

1. *Ordered Sampling without replacement* The number of ordered samples of size k from a population of size n is
$$^{n}P_{k} = n(n-1)\cdots(n-k+1)$$
(the first object can be chosen in n ways, the second in $(n-1)$ ways, etc) $^{n}P_{k}$ is known as "the number of permutations of n objects taken k at a time".

   Note that if $k = n$, we have the number of different possible orderings (*permutations*) of the whole set, which is
$$^{n}P_{n} = n!$$
Remember that if $n! = n(n-1)\cdots 3 \cdot 2 \cdot 1$, and since $^{n}P_{k} = \frac{n!}{(n-k)!}$ it is quite consistent to define $1! = 1$ and $0! = 1$ (put $n = 1$ or $k = n$ in $^{n}P_{k}$).

2. *Unordered Sampling* Now suppose the order does not matter, that we are interested in <u>which</u> k objects are chosen but not in which order they are chosen. In 1. each set of k distinct objects will be counted in each of its $k!$ possible orderings. So if the number of distinct (unordered) samples is denoted by $^{n}C_{k}$ we should have
$$^{n}P_{k} = (k!)\,^{n}C_{k}$$
Hence
$$^{n}C_{k} = n(n-1)\cdots(n-k+1)/k! = \frac{n!}{k!(n-k)!}$$
$^{n}C_{k}$ is known as "the number of *combinations* of n objects taken k at a time".

Example 1.13 A drawer contains 5 knives, 4 forks and 3 spoons. Select three pieces of cutlery "at random" (all possibilities equally likely). What is the probability of selecting

(a) all knives

(b) exactly one knife

(c) one knife, one fork and one spoon?

Answer

(a) 3 knives $$\frac{\binom{5}{3}}{\binom{12}{3}} = \frac{5.4.3}{12.11.10} = \frac{1}{22}$$

(b) 1 knife $$\frac{\binom{5}{1}\binom{7}{2}}{\binom{12}{3}} = \frac{5}{1} \cdot \frac{7.6}{2.1} \cdot \frac{3.2.1}{12.11.10} = \frac{21}{44}$$

(c) 1 of each $$\frac{\binom{5}{1}\binom{4}{1}\binom{3}{1}}{\binom{12}{3}} = 5.4.3 \times \frac{3.2.1}{12.11.10} = \frac{3}{11}$$

Chapter 2

Discrete RV; Expectations

2.1 Definitions

Most of the quantities we might wish to study in a random experiment can be described by numbers (a few may better be described by *categories*, but even these can be given a numbered code e.g. red = 1, blue = 2, yellow = 3, green = 4). A numerical quantity which depends on the outcome of a *random experiment* is called a random variable. If we repeat the experiment, we will get different outcomes and hence different values of the random variable. Capital letters are often used to denote random variables, little letters to denote the possible values they may take (a number, or observed value).

Even in a simple experiment, like tossing a die once, one can define many random variables — e.g. the score X, the square of the score $Y = X^2$, etc. A *binary* random variable can have only two possible values (like parity); a *degenerate* random variable has only one possible value (i.e. it is a fixed number whatever the outcome).

Random variables are used as mathematical models for many numerical quantities observed in real life which have an uncertain outcome: cricket scores, air temperatures, student enrollment numbers etc. In this section we are concerned only with random variables which are *discrete* — the only possible values they can take are integers

Two important quantities associated with a discrete random variable are its *probability distribution* (also called its *probability (mass) function*) and its *expected value (expectation)*.

The *probability distribution* of the discrete random variable X is the set of numbers $p_n,\ n = 0,\ 1,\ 2,\ \ldots$ where

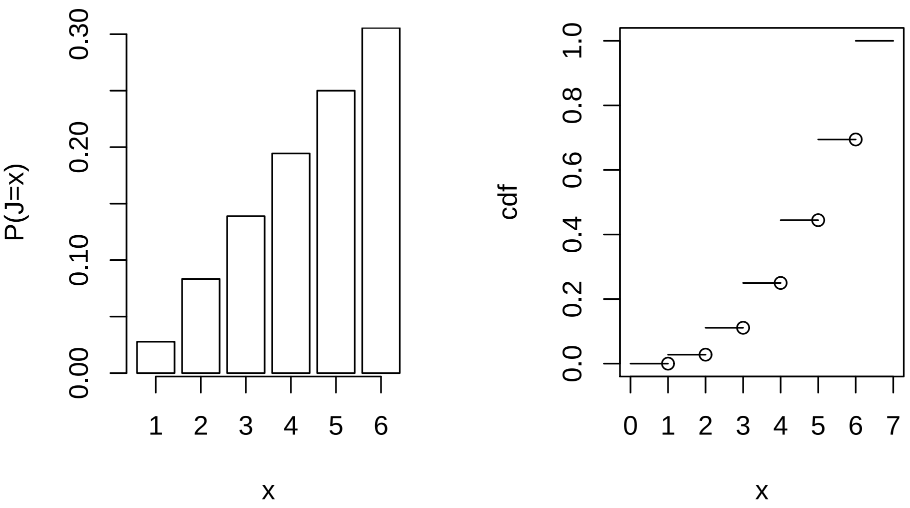
$$p_n = \mathrm{Prob}\{X = n\}$$

For example, if X is the score on a single toss of a die, the probability distribution has six non-zero values, $p_1,\ p_2,\ \ldots,\ p_6$ each of which is equal to $\frac{1}{6}$: thus $p_1 = p_2 = \cdots = p_6 = \frac{1}{6}$ (this is an example of a *discrete uniform distribution*).

Example 2.1 Suppose a single die is tossed twice. Find the probability distribution of the larger of the two scores, (say $J = \max(X_1,\ X_2)$).

Answer There are 36 possible outcomes ($S = \{(1,1),\ (1,2),\ \ldots,\ (6,5),\ (6,6)\}$). J can take any of the values 1 to 6. Because the dice are fair all outcomes are equally probable. By counting cases, we find $p_1 = P(J = 1) = \frac{1}{36}$, $p_2 = \frac{3}{36}$, $p_3 = \frac{5}{36}$, $p_4 = \frac{7}{36}$, $p_5 = \frac{9}{36}$, $p_6 = \frac{11}{36}$. (For example, only the outcomes (1,2), (2,1), and (2,2) contribute to p_2.)

A discrete distribution is best represented by a *bar graph* such as that illustrated below for the distribution of J.



Note that a bar graph differs from a histogram in that the columns are separated. This is to emphasise that the probabilities are attached to individual numbers (the integers), not spread across intervals.

The *cumulative distribution function* (c.d.f., or simply *distribution function*) usually denoted by $F(x)$, or $F_X(x)$ if the random variable needs to be noted, totals (adds) up successive values of the probability function, starting from the left. It therefore has the representation

$$F_X(x) = \mathrm{Prob}(X \leq x)$$

and for a discrete distribution is a step function with a step of height p_n at each integer n. (In the graph above, the open circles indicate that the right endpoint of each "step" is open, while the left endpoint is closed. So, for example, when $x = 1$, $F_X(x) = \frac{1}{36}$ and not 0.)

In the last example $\big(J = \max(X_1, X_2)\big)$, $F_X(3) = P(X \le 3) = p_1 + p_2 + p_3 = \frac{9}{36} = \frac{1}{4}$.

The cumulative distribution function (cdf) has the properties:

(i) $0 \le F(x) \le 1$;

(ii) $P(a < X \le b) = F(b) - F(a)$

(iii) If $a < b$, then $F(a) \le F(b)$ (from (ii)).

If X is discrete, we have introduced the cdf as the sum of p_i (the probability distribution). Working backwards, we see that for a discrete X, with possible values $S = \{x_1,\ x_2,\ \ldots,\ x_k,\ \ldots\}$, $p_k = F(x_k) - F(x_{k-1})$

Definition If X is a discrete random variable, then the *expectation* of X is given by

$$E(X) = x_1 p_1 + x_2 p_2 + \cdots$$

where $p_1,\ p_2,\ \ldots$, are the probabilities associated with $x_1,\ x_2,\ \ldots$, respectively.

To find the expectation the basic procedure is to form the weighted sum over all possible different outcomes

$$E(X) = \sum p_i x_i.$$

Example 2.2 Find the expected value of J in the last example. There are 36 points in the sample space, each with $p(s_i) = \frac{1}{36}$. Hence, recalling that there are three outcomes where $J = 2$, five outcomes where $J = 3$, etc, we have

$$\begin{aligned} E(J) &= \frac{1}{36}\{1 + (2+2+2) + (3+3+3+3+3) + \cdots\} \\ &= 1.\frac{1}{36} + 2.\frac{3}{36} + 3.\frac{5}{36} + 4.\frac{7}{36} + 5.\frac{9}{36} + 6.\frac{11}{36} \\ &= \frac{1+6+15+28+45+66}{36} = \frac{161}{36} = 4\tfrac{17}{36} \end{aligned}$$

A particularly important quantity is the *variance* of X, written Var$X$ or var(X), which is

$$\text{var}(X) = E\big[(X - E(X))^2\big].$$

If we expand the square and simplify, this can be shown to be equal to

$$\text{var}(X) = E(X^2) - (E(X))^2.$$

Example 2.3 Let X be the number of heads observed when a "fair" coin is to be tossed three times.

$$S = \{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH\}$$

$S_X = \{0, 1, 2, 3\}$, with four distinct elements.

$$p_0 = \frac{1}{8}, \quad p_1 = \frac{3}{8}, \quad p_2 = \frac{3}{8}, \quad p_3 = \frac{1}{8}$$

$\mu = \sum p_i x_i = \frac{1}{8} \times 0 + \frac{3}{8} \times 1 + \frac{3}{8} \times 2 + \frac{1}{8} \times 3$
$= \frac{12}{8} = 1.5$ the expected number of heads.

(In fact, this is a "binomial experiment" for which $\mu = np$, where $n = 3$ and $p = \frac{1}{2}$; see page 15.)

$$E(X^2) = \sum p_i x_i^2 = \frac{1}{8} \times 0^2 + \frac{3}{8} \times 1^2 + \frac{3}{8} \times 2^2 + \frac{1}{8} \times 3^2 = \frac{24}{8} = 3$$

Hence,

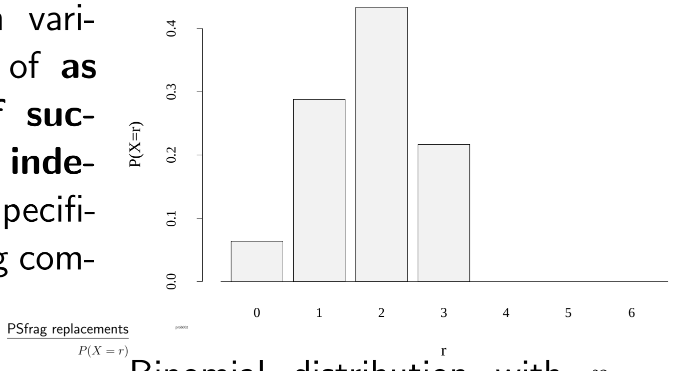
$$\sigma^2 = E(X^2) - \mu^2 = 3 - (1.5)^2 = 0.75 = \frac{3}{4}$$

(Again, for a "binomial experiment", $\sigma^2 = np(1-p) = 3 \times \frac{1}{2} \times \frac{1}{2}$.)

The importance of both expected values and probabilities lies in the so-called "laws of large numbers". These assert that in a long sequence of observations under identical conditions, the arithmetic mean of the observed values will be very close to the expected value of the random variable, and similarly the relative frequency with which a particular outcome is observed will be very close to its probability.

2.2 The Binomial Distribution

This distribution arises wherever we are dealing with a random variable which can be thought of **as recording the number of successes in a sequence of independent Yes/No trials**. Specifically we look for the following combination of circumstances.



Binomial distribution with $p = 0.6, \quad n = 3$.

1. *The experiment involves a sequence of independent, identical trials.*
2. *The outcome for each trial can be classified in just two ways, namely as "success" or "failure".*
3. *The probability p of a success is fixed for each trial.*
4. *The random variable X we are examining counts the number of successes in a fixed (predetermined) number of trials.*

Any set of trials with the properties *1.* and *2.* is called a set of *Bernoulli Trials.* We meet Bernoulli trials again later, when discussing geometric and negative binomial distributions.

Example 2.4 Tossing a coin and counting the number (X) of heads; drawing a random sample *with* replacement and counting the number (X) having some particular characteristic; the number of faulty products in a batch; the number of people with one of two possible views (like "yes" and "no" or "support" and "don't support") in a sample for an opinion poll; etc.

The distribution is specified by two quantities (usually called *parameters*), namely the fixed number of trials, n, and the probability of success in each trial, p. The abbreviation $\mathsf{B}(n,\, p)$ is used to denote the particular distribution corresponding to a given choice of n and p.

If X denotes the number of successes and $q = 1 - p$, then for $x = 0,\ 1,\ 2,\ \ldots,\ n$,

$$p_x = P(X = x) = \binom{n}{x} p^x q^{n-x}$$

It can be shown that

$$E(X) = np \qquad \mathsf{var}(X) = npq \qquad \sigma_X = \sqrt{npq}$$

The distribution is symmetrical if $p = q = \frac{1}{2}$, otherwise skewed towards the lower half of the range if $p < \frac{1}{2}$ and towards the upper half if $p > \frac{1}{2}$.

Example 2.5 If 80% of all students are in favour of lower fees, what is the probability that in a random sample of 10 students:

(a) there will be 7 in favour of lower fees;
(b) at least 7 will be in favour of lower fees?

Solution

Let X be the number in favour, then we can say $X \sim B(10, 0.8)$ as questioning each student can be regarded as a Bernoulli trial with $p = 0.8$. So

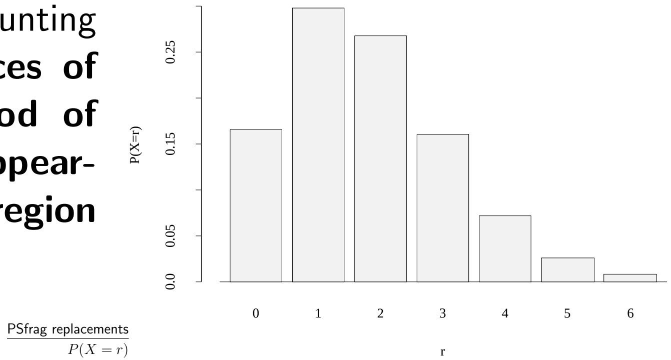
(a) $P(X = 7) = 0.2013.$
(b) $P(X \geq 7) = 0.8791$

Does 0.8791 agree with our intuition? If 80% are in favour, we'd expect around 8 out of the 10 to be in favour (i.e. the probability of "at least 7" to be high), so it does.

Most statistical computer packages can calculate binomial probabilities for almost any values of n and p, as can many calculators now.

2.3 The Poisson Distribution

This distribution is likely to be found whenever we are counting **the number of occurrences of some event over a period of time** , or **the number of appearances of some object in a region of space**.



Poisson distribution with $\mu = 1.8$.

1. *The experiment consists in counting the number of occurrences of a certain event (or object) in a fixed interval of time (or region of space).*
2. *These events occur one at a time and not simultaneously in groups.*
3. *The number of occurrences in any subinterval (subregion) is independent of the number of occurrences in any disjoint subinterval (subregion).*

The situations where the Poisson distribution arises differ from those where the binomial distribution arises because there is **no fixed set of trials**.

Example 2.6 The number of earthquakes with magnitude 5 or above in a given year; the number of telephone calls received in a given day; the number of fatal accidents over the Christmas period; the number of nuggets of gold in a mining claim; the number of blood cells in a given square in the field of view of a microscope; the number of galaxies in a given region of space, etc.

The Poisson distribution is characterised by a single parameter μ, which turns out to be both its mean and its variance, and the distribution is usually abbreviated to Pois(μ). The probabilities, which are defined for *all possible non-negative integers*, are given by

$$\boxed{p_x = \frac{\mu^x}{x!}e^{-\mu} \qquad x = 0,\ 1,\ 2,\ \ldots}$$

These probabilities decrease rapidly, so that all, save the first few, can be ignored in practice.

Sometimes r is used to denote the number of occurrences and λ to denote the parameter, in which case the probability density function would be written as

$$p_r = \frac{\lambda^r}{r!}e^{-\lambda} \qquad r = 0,\ 1,\ 2,\ \ldots$$

In these notes we will use μ to denote the average rate of occurrence in the interval of time or space **OF INTEREST**, and λ to denote the average rate **PER UNIT** time or space.

For the distribution we have

$$\boxed{\mu_X = \mu, \qquad \text{var}(X) = \mu, \qquad \sigma_X = \sqrt{\mu}}$$

Very commonly, the parameter is specified as an average rate, that is, the number of events per unit time (or per unit area, etc) and it is the length of time (or area, etc) which is of interest. Then μ is found by setting

$$\begin{aligned} \mu &= \text{average rate} \times \text{length of time} \\ &= \lambda t \end{aligned}$$

or in the case of areas

$$\begin{aligned} \mu &= \text{average density} \times \text{area} \\ &= \lambda A \end{aligned}$$

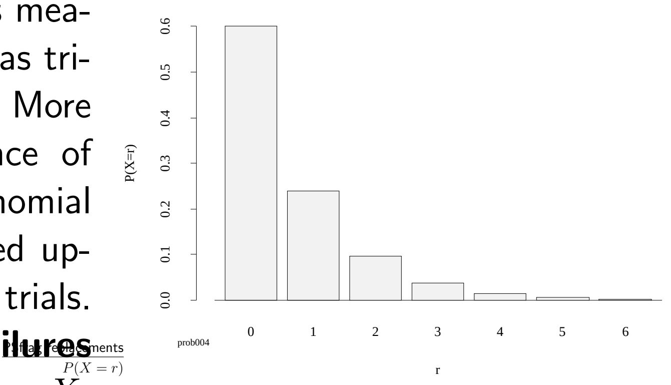
Probabilities can then be calculated from the formula given above.

Example 2.7 Suppose the average rate of large earthquakes is 2 per year. What is the probability of having 3 years without large earthquakes?

Answer Here $\mu = 2 \times 3 = 6$, so $p_0 = e^{-6} \approx 0.0025$.

2.4 The Geometric Distribution

The geometric distribution arises as a model for the distribution of **waiting times**, when time is measured in discrete units (such as trials) rather than continuously. More precisely, consider a sequence of Bernoulli trials, as for the binomial distribution, but with no fixed upper limit on the number of trials. We count the **number of failures before the first success**, say X. Then X has a geometric distribution. Note that X can take the values 0, 1, 2, ..., that is, it starts at zero.



Geometric distribution with $p = 0.6$, ($\mu = \frac{2}{3}$).

Note that sometimes the geometric distribution is defined as the number of trials Y *up to and including* the first success. Then Y can take the values 1, 2, . . . , starting at 1 and not at zero. In fact, $Y = X + 1$.

The probability of getting a sequence of trials starting with x failures in a row and then a success is $q^x p$, assuming the trials are independent and have a constant probability of success. Hence the geometric distribution has the form:

$$P(X = x) = p_x = q^x p, \qquad x = 0, 1, 2, \ldots$$

with

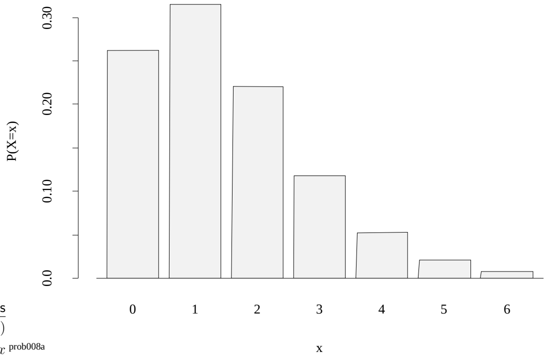
$$E(X) = \frac{q}{p}, \qquad \text{var}(X) = \frac{q}{p^2}$$

The characteristic feature of the distribution of X is that the terms decrease geometrically with x, each one being reduced by a further factor q from the previous one. The "tails" of the distribution (i.e. $P(X \geq x)$) also decay geometrically:

$$P(X \geq x) = q^x p + q^{x+1} p + \cdots = \frac{q^x p}{1-q} = q^x$$

2.5 The Negative Binomial Distribution

The negative binomial is a generalisation and extension of the geometric distribution. It arises in more general **waiting time situations, such as the number of trials before the 3rd success**, and is also used as a model for situations where events are often clustered together, such as the number of breakdowns in a computer in a given week, or the number of accidents on a given shift, or the number of plants in a given square metre of ground.



Negative Binomial distribution with $p = 0.8$, $r = 6$, $(\mu = 1.5)$.

The negative binomial distribution depends on two parameters, p and r. In terms of a waiting time situation, p is the probability of a "success", r is the number of "successes" that will "terminate" the series of trials, x is the number of "failures" before r "successes" are achieved (giving a total of $x+r$ "trials", the last of which is known to be a "success"). As usual when discussing a probability distribution based on Bernoulli trials, $q = 1 - p$.

The negative binomial distribution has probabilities of the form

$$\begin{aligned} P(X = x) = p_x &= \binom{x+r-1}{x} p^r q^x \\ &= \frac{(r+x-1)\cdots(r+1)r}{x!} p^r q^x, \qquad x = 0,\ 1,\ 2,\ \ldots \end{aligned}$$

The mean and variance are given by

$$E(X) = \sum x p_x = \frac{rq}{p} \qquad \text{var}(X) = \frac{rq}{p^2}$$

The geometric distribution corresponds to the special case $r = 1$.

Example 2.8 A patient taking a new drug, *Wonderdrug*, has a 20% probability of suffering side-effects. What is the probability that a doctor prescribing the drug finds that the fifth person starting to take the drug is the second to have side-effects?

Answer Let X be the number of people who do not have side-effects from *Wonderdrug*, then $r = 2$ (the second to have side-effects). $p = 0.2$ and $q = 0.8$ and we want $P(X = 3)$ (as there are 3 without side-effects and 1 with before the fifth person, who is the second to have side-effects).

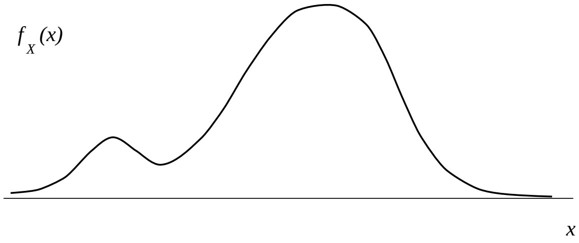
$$P(X=3)=p_3=\binom{x+r-1}{x}p^r q^x=\binom{4}{3}(0.2)^2(0.8)^3=0.08192$$

Chapter 4

Continuous Random Variables

4.1 Definitions and Examples

A continuous random variable X is one which can take on any value in a given range; e.g. heights and weights of people, magnitudes of earthquakes, levels of reservoirs, blood cholesterol levels, etc. More technically, it is a random variable with a continuous cumulative distribution function, i.e. without the steps characteristic of the discrete case.



pdf001

We cannot define probabilities for continuous random variables in the same way as for discrete random variables. For example if X is a Bin(n, p) random variable then X is discrete and takes on the values 0, 1, 2, ..., n. Hence we can easily write down the probabilities $P(X = 0),\ P(X = 1),\ \ldots,\ P(X = n)$. Suppose Y is a continuous random variable which can take on any value in [0,1] then we cannot even list all of the possible values for Y, i.e., Y has an uncountably infinite set of possible values. Hence we cannot produce a list of probabilities which describe the behaviour of Y. Instead it turns out that the best way to describe the behaviour of Y is to look at probabilities of the form $P(Y \leq y)$ rather than $P(Y = y)$.

Example 4.1 The error X in any weight measurement made by a certain set of electronic scales lies between 0 and 1 g. Suppose that X is a (continuous) random variable which is equally likely to take on any value between 0 and 1. Determine $P(X = x)$ and $P(X \le x)$ for $0 \le x \le 1$.

Answer $P(X = x) = 0$; $P(X \le x) = x$ for $0 \le x \le 1$.

Result For many continuous random variables X there exists a non-negative function $f_X(x)$ such that the area under the graph of $f_X(x)$ to the left of point x is equal to $P(X \le x)$, which is denoted by

$$P(X \le x) = \int_{-\infty}^{x} f_X(u)\,du \qquad (-\infty < x < \infty)$$

Definition The function $f_X(x)$ is called the *probability density function* (pdf) of X and

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(u)\,du \qquad (-\infty < x < \infty)$$

is called the *cumulative distribution function* (cdf) or *distribution function* of X.

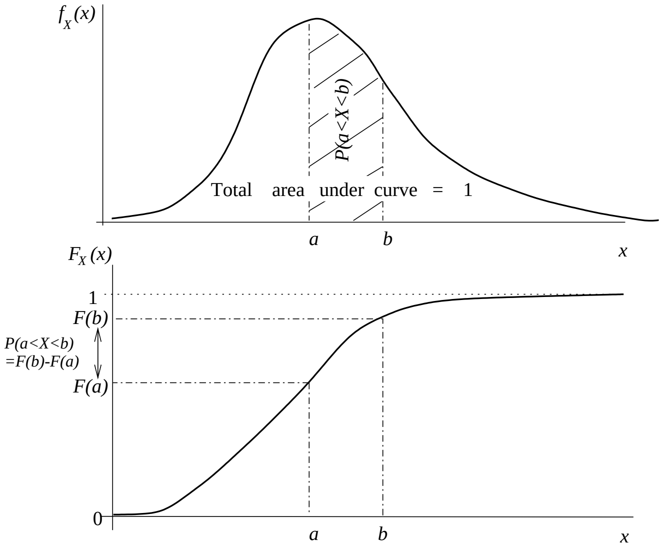
It follows from the above result that for any continuous r.v. X with probability density function (pdf) $f_X(x)$ and constants $a,\ b\ (a \le b)$

$$\begin{aligned} P(a < X < b) &= \int_a^b f_X(x)\,dx \\ &= F(b) - F(a) \end{aligned}$$

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

Moreover, wherever $F_X(x)$ is differentiable,

$$F'_X(x) = \frac{d}{dx}F_X(x) = f_X(x)$$



pdf002

Both the probability density function, $f_X(x)$, and the distribution function, $F_X(x)$, have simple interpretations. The distribution function is easiest since $F_X(x)$ is simply the probability $P(X \leq x)$ for any given value of x. Because of this property it is clear that the following results hold

- $0 \leq F_X(x) \leq 1$
- $F_X(x)$ is monotonic increasing, i.e. $F_X(a) \leq F_X(b)$ whenever $a \leq b$
- $\lim\limits_{x \to -\infty} F_X(x) = 0$ and $\lim\limits_{x \to \infty} F_X(x) = 1.$

The probability density function also gives probabilities but in a completely different way. The pdf $f_X(x)$ is a piecewise continuous function which gives probabilities as areas under the curve. Hence the probability $P(a \leq X \leq b)$ is given by the area under $f_X(x)$ between $x = a$ and $x = b$. Hence the following results apply to probability density functions

- $f_X(x) \geq 0$
- $\displaystyle\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ (since the probability that X lies in $(-\infty, \infty)$ must be 1!)

Note that $f_X(x)$ is NOT A PROBABILITY and so $f_X(x) > 1$ is possible.

Both the probability density function and the distribution function are ways of describing the probabilistic behaviour of a continuous random variable.

4.2 Expectations, Means and Variances (Optional/Advanced)

How do we define the expected value of a continuous random variable X? We can make use of our definition of expected value for discrete random variables to produce a sensible definition for continuous random variables. Suppose X is a continuous r.v. with pdf $f_X(x)$. Divide the x axis up into intervals of width dx (dx very small) and define the discrete r.v. Y by setting

$$Y = x \quad \text{if} \quad x \leq X < x + dx \qquad (-\infty < x < \infty).$$

Then since dx is very small

$$P(Y = x) = P(x \leq X < x + dx) \approx f_X(x)\,dx.$$

Moreover, from the theory of expectation for discrete r.v.s

$$\begin{aligned} E(Y) &= \sum x\, P(Y=x) \\ &\approx \sum x\, f_X(x)\, dx \ \overset{dx}{\underset{0}{\longrightarrow}} \ \int_{-\infty}^{\infty} x\, f_X(x)\, dx. \end{aligned}$$

Thus, as dx goes to 0, Y becomes X and so

$$E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx.$$

Definition For a continuous r.v. X with pdf $f_X(x)$ the *expected value* of X is

$$E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \mu_X.$$

Similarly, for any given function $g(x)$, the expected value of $g(X)$ is

$$E\big(g(X)\big) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx.$$

In particular, setting

$$g(X) = \big(X - E(X)\big)^2 = (X - \mu_X)^2$$

we obtain the *variance* of X (usually denoted by σ_X^2); i.e.

$$\sigma_X^2 = E\Big[\big(X - E(X)\big)^2\Big] = E\big[(X - \mu_X)^2\big] = \int_{-\infty}^{\infty} (x - \mu_X)^2\, f_X(x)\, dx.$$

Also

$$\sigma_X^2 = E(X^2) - \big(E(X)\big)^2 = E(X^2) - \mu_X^2.$$

Expected values can be interpreted as long run averages exactly as for discrete random variables. Hence we can take the expected value of any function of a random variable. However by far the most useful expected values are the mean and the variance given above.

Remember that expected values are in principle very easy to calculate. The definition above gives the formula for $E\big(g(X)\big)$ and this can be used no matter how complex the function $g(x)$ is. Evaluating the expected value is then simply a question of integration.

Example 4.2 X is a continuous random variable with distribution function given by

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x^3 & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

Check that you can reproduce the following results

1. $E(X) = \frac{3}{4}$
2. $E(X^2) = \frac{3}{5}$
3. $E(X^{-1}) = \frac{3}{2}$
4. $\text{Var}(X) = \frac{3}{80}$

Example 4.3 X is a continuous random variable with pdf given by

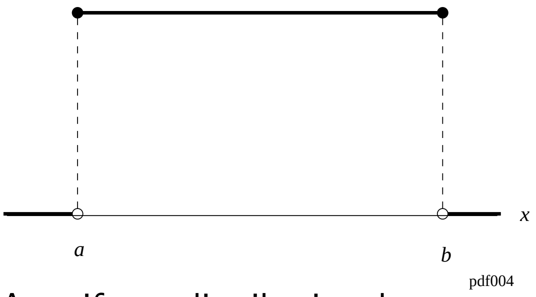
$$f_X(x) = \begin{cases} x^{-2} & x \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

Check that you can reproduce the following results:

1. $E(X)$ is infinite.
2. $E(X^{-1}) = \frac{1}{2}$
3. $E(X^{-2}) = \frac{1}{3}$

4.3 The Uniform Distribution

The uniform distribution describes situations where a continuous random variable lies in a fixed interval $[a, b]$ and is equally likely to lie in any subinterval of length d no matter where the subinterval lies in $[a, b]$. We can think of the probability as being uniformly smeared over the interval $[a, b]$. Hence this distribution is useful in situations where there is no evidence that any region is more likely than any other and the possible regions form a finite interval $[a, b]$. Alternatively you could use the uniform distribution to model **complete ignorance**.



A uniform distribution between a and b.

Characteristics The continuous r.v. X has a uniform distribution on the interval $[a, b]$ $(-\infty < a < b < \infty)$ if the pdf of X is

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & (a \le x \le b) \\ 0 & \text{(otherwise)} \end{cases}$$

The distribution function of X is given by

$$F_X(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$$

Result Let X be a continuous r.v. with a uniform distribution on $[a, b]$. Then

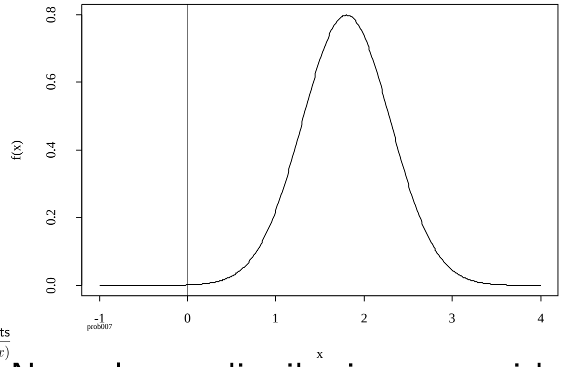
$$\mu_X = E(X) = \frac{a+b}{2}$$

$$\sigma_X = \text{ standard deviation of } X = \frac{b-a}{\sqrt{12}}$$

4.4 The Normal Distribution

By far the most important model for the pdf of a continuous r.v. is the normal distribution.

The basic reason that the normal distribution works well as a model for many different types of measurements generated in real experiments is that many measurements can be regarded as aggregates. Whenever responses tend to be sums or averages of independent quantities, the normal distribution quite likely will provide a reasonably good model for their relative frequency behaviour.



Normal distribution with $\mu = 1.8, \quad \sigma^2 = 0.25.$

Many naturally occurring measurements tend to have relative frequency distributions closely resembling the normal curve, probably because nature tends to "average out" the effects of the many variables that relate to a particular response. For example, heights of NZ women tend to have a distribution that shows many measurements clumped closely about a mean height, with relatively few very short or very tall women in the population. In other words, the relative frequency distribution is close to normal.

Example 4.4 Weight of powder in boxes of washing powder, height of NZ men, voltage in a power socket, cholesterol levels of smokers, aerobic fitness of students.

Characteristics A continuous r.v. X has a normal distribution with parameters μ and σ^2 $(-\infty < \mu < \infty,\, \sigma^2 > 0)$ if the pdf of X is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (-\infty < x < \infty)$$

Note: The following statements are all equivalent:

"X has a normal distribution with parameters μ and σ^2".
"X is a normal r.v. with parameters μ and σ^2".
"X is a $N(\mu,\, \sigma^2)$ random variable".
"X has a $N(\mu,\, \sigma^2)$ distribution".

"X is $N(\mu,\, \sigma^2)$".

Useful facts concerning the normal pdf are:-

- it is bell shaped and symmetric about μ;
- it can be shown that for $X \sim N(\mu,\, \sigma^2)$, 99.74% of the values of X lie in the interval $[\mu - 3\sigma, \mu + 3\sigma]$, 95.44% in the interval $[\mu - 2\sigma, \mu + 2\sigma]$, and 68.26% in the interval $[\mu - \sigma, \mu + \sigma]$.

Example 4.5 The mean length μ of mature karaka leaves is 151mm and the s.d., σ, is 15mm. Let X denote the length of a randomly chosen karaka leaf, and assume that the values of X are normally distributed.

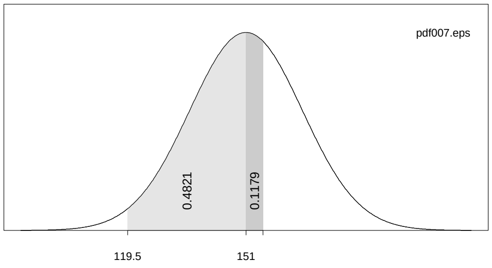
Find (a) $P(120 \leq X \leq 155)$ (b) $P(X > 185)$ (c) $P(X \leq 128)$.

(d) In a random sample of size $n = 500$ leaves, find the expected number of leaves with lengths between 120 and 155mm.

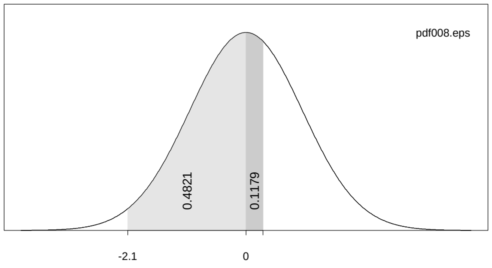
Solution

(a) Lengths recorded between 120mm and 155mm can have any value between 119.5mm and 155.5mm, if measured to the nearest mm. The lower and upper standard scores are

$$z_1 = \frac{X_1 - \mu}{\sigma} = \frac{119.5 - 151}{15} = -2.1 \qquad z_2 = \frac{X_2 - \mu}{\sigma} = \frac{155.5 - 151}{15} = 0.3$$



$X \sim N(151, 15^2)$

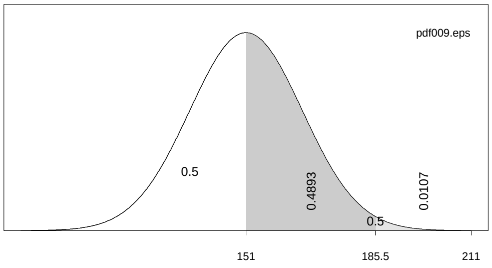


$Z \sim N(0, 1)$

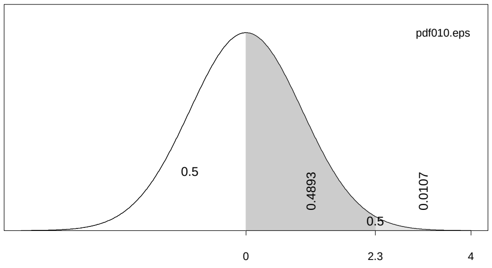
$$\begin{aligned} P(120 \leq X \leq 155) &= P(-2.1 < Z < 0.3) \\ &= P(-2.1 < Z < 0) + P(0 < Z < 0.3) \\ &= P(0 < Z < 2.1) + P(0 < Z < 0.3) \qquad \text{by symmetry} \\ &= 0.4821 + 0.1179 \qquad \text{(See Statistical Table for Normal distribution)} \\ &= 0.6000 \qquad \text{(or 60\%)} \end{aligned}$$

(b) To the nearest mm, lengths recorded as greater than 185 must actually be greater than 185.5.

$$z = \frac{X - \mu}{\sigma} = \frac{185.5 - 151}{15} = 2.3$$



$X \sim N(151, 15^2)$

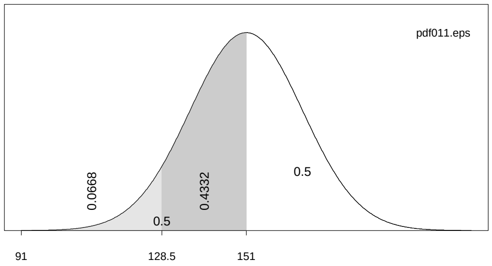


$Z \sim N(0, 1)$

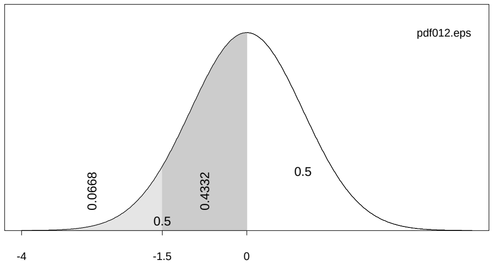
$$\begin{aligned} P(X > 185) &= P(Z > 2.3) \\ &= P(Z > 0) - P(0 < Z < 2.3) \\ &= 0.5000 - 0.4893 = 0.0107 \qquad \text{(or 1.07\%)} \end{aligned}$$

(c) To the nearest mm, lengths recorded as less than or equal to 128 must actually be less than 128.5.

$$z = \frac{X - \mu}{\sigma} = \frac{128.5 - 151}{15} = -1.5$$



$X \sim N(151, 15^2)$



$Z \sim N(0, 1)$

$$\begin{aligned} P(X \leq 128) &= P(Z < -1.5) \\ &= P(Z < 0) - P(-1.5 < Z < 0) \\ &= P(Z > 0) - P(0 < Z < 1.5) \qquad \text{(by symmetry)} \\ &= 0.5000 - 0.4332 = 0.0668 \qquad \text{(or 6.68\%)} \end{aligned}$$

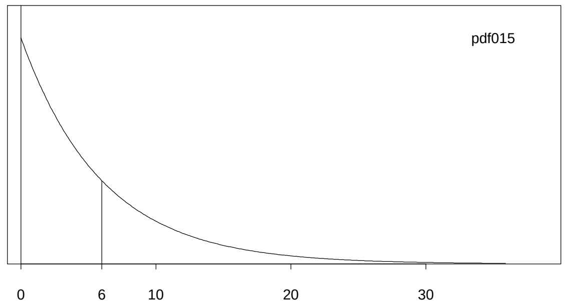
(d) In a random sample of size $n = 500$ leaves, we can treat $P(120 \leq X \leq 155)$ as a proportion, and the expected number would be 60% of 500, which is 300.

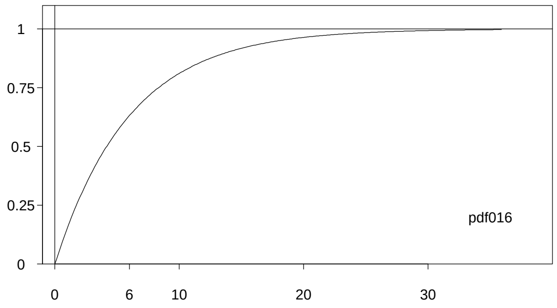
Alternatively, as a binomial experiment, a leaf is in the range [120, 155] or it is not (2 possible outcomes), $n = 500$, $p = 0.60$ and $\mu = np = 500 \times 0.60 = 300$, as before.

(And we would expect $np = 500 \times 0.0107 = 5.35$ to be longer than 185mm, and $np = 500 \times 0.0668 = 33.4$ to be shorter than 128mm.)

4.5 The Exponential Distribution

The exponential distribution models situations where the data takes on positive values only, is more likely to be near the origin and is increasingly less likely to be in an interval of fixed width as it moves further to the right. In fact many random variables in engineering and the sciences can be modelled appropriately as having exponential distributions. The exponential distribution can be thought of as a continuous analogue to the geometric distribution and is often used to model **waiting times**.





A pdf of an exponential distribution with mean of 6.   Corresponding cdf.

A common situation is when we are measuring times between certain events of interest. As long as the events occur fairly randomly then the time between them can be well approximated as an exponential random variable. In fact this can be proven analytically under certain assumptions.

Example 4.6 Time between vehicles passing a fixed point on a motorway, time between accidents on an airline, how long you have to wait for a lift in an office building.

Characteristics The continuous r.v. X has an exponential distribution with parameter θ ($\theta > 0$) if the pdf of X is

$$f_X(x) = \begin{cases} \theta\, e^{-\theta x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

The distribution function of X is given by

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\theta x} & x \geq 0 \end{cases}$$

Result Let the continuous r.v. X have an exponential distribution with parameter θ. Then

$$\mu_X = E(X) = \frac{1}{\theta}$$

$$\sigma_X = \text{standard deviation of } X = \sqrt{Var(X)} = \frac{1}{\theta}$$

Note For the exponential distribution, $\mu_X = \sigma_X$.

Example 4.7 The duration, X, in minutes of phone calls from company business phones is a continuous r.v. with pdf

$$f_X(x) = \begin{cases} \frac{1}{6} e^{-x/6} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

(a) What is the cdf, and show that this function is, in fact, a satisfactory cdf.
(b) Calculate the probability that a call will last between 3 and 6 minutes.
(c) Find the expected length of a call.
(d) Determine the variance of call length.

Solution

(a) $F(x) = 1 - e^{-x/6} \qquad x \geq 0$

We can confirm that this is a satisfactory cdf:

- $F(0) = 1 - 1 = 0$, that is, $P(X \leq 0) = 0$.
- $F(\infty) = 1 - 0 = 1$, that is, $P(R_X) = 1$. (Since $\lim_{x\to\infty} e^{-x/6} = 0$.)
- $F'(x) = f(x) \geq 0$ so $F(x)$ is an increasing function of x.

(b) $$\begin{aligned} P(3 < X \leq 6) &= F(6) - F(3) = (1 - e^{-6/6}) - (1 - e^{-3/6}) \\ &= 0.6065306 - 0.3678794 \\ &= 0.2386512 \end{aligned}$$

(c) The expected length of a call, using the result above, is $\mu = E(X) = \dfrac{1}{\theta} = \dfrac{1}{\frac{1}{6}} = 6$ minutes ($\theta = \frac{1}{6}$).

(d) The variance, using the result above, is var$(X) = \sigma^2 = \left(\dfrac{1}{\theta}\right)^2 = \left(\dfrac{1}{\frac{1}{6}}\right)^2 = 36.$