



Supporting Australian Mathematics Project

A guide for teachers - Years 11 and 12

Probability and statistics: Module 19

Discrete probability distributions



Education
Services
Australia



AMSI

AUSTRALIAN MATHEMATICAL
SCIENCES INSTITUTE

Discrete probability distributions - A guide for teachers (Years 11-12)

Professor Ian Gordon, University of Melbourne

Editor: Dr Jane Pitkethly, La Trobe University

Illustrations and web design: Catherine Tan, Michael Shaw

Full bibliographic details are available from Education Services Australia.

Published by Education Services Australia
PO Box 177
Carlton South Vic 3053
Australia

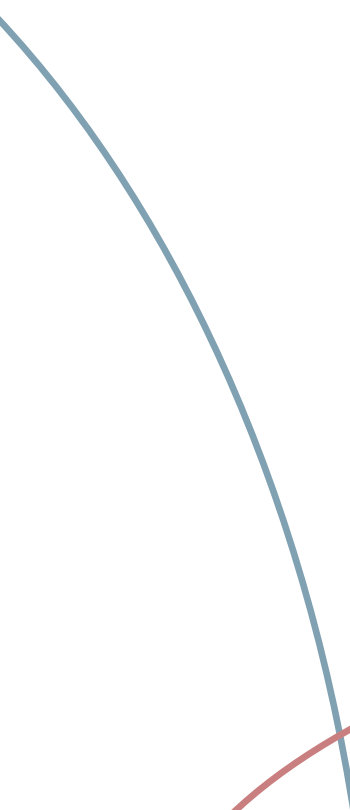
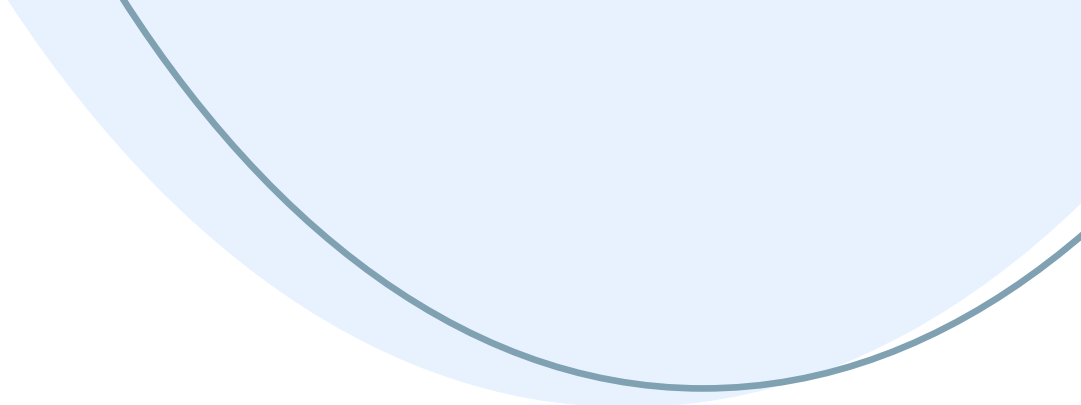
Tel: (03) 9207 9600
Fax: (03) 9910 9800
Email: info@esa.edu.au
Website: www.esa.edu.au

© 2013 Education Services Australia Ltd, except where indicated otherwise. You may copy, distribute and adapt this material free of charge for non-commercial educational purposes, provided you retain all copyright notices and acknowledgements.

This publication is funded by the Australian Government Department of Education, Employment and Workplace Relations.

Supporting Australian Mathematics Project

Australian Mathematical Sciences Institute
Building 161
The University of Melbourne
VIC 3010
Email: enquiries@amsi.org.au
Website: www.amsi.org.au



Assumed knowledge	4
Motivation	4
Content	5
Random variables	5
Discrete random variables: general ideas	7
Probability functions	9
Examples of discrete distributions	12
Mean of a discrete random variable	15
Variance of a discrete random variable	20
Answers to exercises	24

Discrete probability distributions

Assumed knowledge

The content of the module *Probability*, which formalises the ideas in the series of TIMES modules *Chance* (Years 1–10).

Motivation

This module introduces **discrete random variables**. A random variable can be either discrete or continuous. In this module, we cover the first type, and the module *Continuous probability distributions* covers the second.

The idea of a random variable builds on the fundamental ideas of probability. Students need to understand that random variables are conceptually different from the mathematical variables that they have met before. A random variable is linked to observations in the real world, where uncertainty is involved.

An informal — but important — understanding of a random variable is that it is a variable whose numerical value is determined by the outcome of a random procedure. In this module, we also see the more formal understanding, which is that a random variable is a function from the event space of a random procedure to the set of real numbers.

Random variables are central to the use of probability in practice. They are used to model random phenomena, which means that they are relevant to a wide range of human activity. In particular, they are used extensively in many branches of research, including agriculture, biology, ecology, economics, medicine, meteorology, physics, psychology and others. They provide a structure for making inferences about the world, when it is impossible to measure things comprehensively. They are used to model outcomes of processes that cannot be predicted in advance.

Random variables have **distributions**. In this module, we describe the essential properties of distributions of discrete random variables. Distributions can have many forms, but there are some special types of distributions that arise in many different practical

contexts. In this module, we discuss two of these special distributions: discrete uniform distributions and geometric distributions.

This module also covers the mean of a discrete random variable, which is a measure of central location, and the variance and standard deviation, which are measures of spread.

Content

Random variables

A **random variable** is a variable whose value is determined by the outcome of a **random procedure**. The concept of a random procedure was discussed in the module *Probability*. What makes the variable *random* is that — unlike the kind of variable we see in a quadratic equation — we cannot say what the observed value of the random variable is until we actually carry out the random procedure.

Example: Tetris

Consider the following example of a random procedure from the module *Probability*: During a game of Tetris, we observe a sequence of three consecutive pieces.

Each Tetris piece has one of seven possible shapes, which are labelled by the letters I, J, L, O, S, T and Z. So in this random procedure, we can observe a sequence such as JLL, ZOS, ZSZ, III and so on.

Based on this random procedure, we may define a number of random variables. For example:

- Define X to be the number of occurrences of ‘Z’ in a sequence of three pieces. Then X can take the value 0, 1, 2 or 3.
- Define Y to be the number of different shapes in a sequence of three pieces. Then Y can take the value 1, 2 or 3.

These are not the only random variables that could be defined in this context.

This example illustrates that a random variable takes a numerical value in a specific case, when the random procedure is carried out.

A convention of notation is that random variables are denoted by capital letters, usually near the end of the alphabet.

Example: Five people born in 1995

Consider another example of a random procedure from the module *Probability*: Five babies born in 1995 are followed up over their lives, and major health and milestone events are recorded.

This example is only vaguely described, and would be more tightly defined in practice. But we can see, again, that a number of random variables could be defined:

- Let U_i be the total number of times that individual i moves residence up to age 18. Then U_i can take values $0, 1, 2, \dots$
- Let V_i be the total number of mobile phones owned by individual i up to age 18. The possible values for V_i are $0, 1, 2, \dots$, which incidentally are the same as those for U_i .
- Let W be the average height of the five people at age 18. Then the value of W must be positive, but there is no obvious upper bound. The common practice in such cases is to say that the possible values are $W > 0$; we will assign extremely low probabilities to large values.
- Let T_i be the total time spent on Facebook by individual i up to age 18. Then T_i in this case is limited by the total time span being considered. If we measure T_i in years, then $0 \leq T_i \leq 18$; again, values anywhere near the logical maximum of 18 years will be assigned essentially zero probability.

The random variables given in the previous example are of two distinct types, which are handled in different ways:

- 1 A **discrete random variable** takes values confined to a range of separate or ‘discrete’ values. (More formally, a discrete random variable takes either a finite number of values or a countably infinite number of values.) In the example, the first two random variables U_i and V_i are counts: they can only take non-negative integer values. (A person cannot have moved residence 2.3 times by age 18, nor can a person have owned 9.6 mobile phones.) A random variable based on a count is an example of a discrete random variable.
- 2 A **continuous random variable** can take any value in an interval. In the example, the third and fourth random variables W and T_i are continuous random variables.

This module concerns discrete random variables. The module *Continuous probability distributions* deals with continuous random variables.

It is important to see that the definition of a random variable needs a specification of what is observed or recorded. In some situations, this is essentially implied; but it is always required implicitly or explicitly.

Example

A fair coin is spun vertically on a flat surface. (This example comes from an exercise in the module *Probability*.) Here are two related random variables:

- Let X be the number of heads showing when the coin comes to rest. Then X takes the value 0 if the coin finishes up ‘tails’, or 1 if the coin finishes up ‘heads’.
- Let Y be the time between the commencement of the spin and the coin coming to rest, measured in seconds.

Here X is discrete and Y is continuous.

A special case of a discrete random variable is one that can take only a finite number of values. We call this a **simple random variable**. Since X in the previous example can only take values 0 and 1, it is a simple random variable.

Discrete random variables: general ideas

We start with the classic example of rolling a fair six-sided die. This scenario has already been discussed extensively in the module *Probability*. You may wonder why introductory discussions of probability and random variables often start with examples involving dice, cards and coins. The reason is that these are concrete randomising devices with which we are familiar, and for which we are willing to entertain plausible assumptions about the idealised model.

As we have seen in the module *Probability*, for coin tossing in particular, this is not as straightforward as it seems. When considering the *actual* use of dice, cards and coins, the potential for other factors that will invalidate the model should always be kept in mind. Probability and statistics give us the framework to think about this rationally.

But for the initial example, we assume that the model is correct: the die is symmetric in every respect and is shaken well before it is rolled. The uppermost face is observed. The event space for this random procedure is $\mathcal{E} = \{1, 2, 3, 4, 5, 6\}$, and the possible outcomes are equiprobable.

Define the discrete random variable X to be the number on the uppermost face when the die comes to rest. In this case, there is a distinct value of the random variable for each distinct possible outcome of the random procedure. X can take the values 1, 2, 3, 4, 5, 6, and no other values. It takes discrete values, integers in this case — the die cannot end up with an outcome of π or 1.01 — so X is a discrete random variable. Since X can only take six possible values, it is a simple random variable.

The probability that a discrete random variable X takes the value x is denoted $\Pr(X = x)$. We read this as ‘the probability that X equals x ’, which means the probability that X takes the value x when we actually obtain an observation. For the die-rolling example,

$$\Pr(X = x) = \begin{cases} \frac{1}{6} & \text{if } x = 1, 2, 3, 4, 5, 6, \\ 0 & \text{otherwise,} \end{cases}$$

for each real number x . Often, for discrete random variables, it is sufficient to specify in some way the values with non-zero probability only; the values with zero probability are usually clear, or clearly implied.

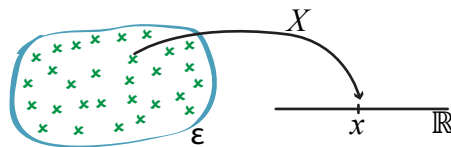
There are other simple random variables that can be defined for the random procedure of rolling a die. For example:

- Let Y be the number of even numbers appearing. Then Y takes value 1 if a 2, 4 or 6 is rolled, and Y takes value 0 otherwise.
- Let Z be the number of prime numbers appearing. Here Z takes value 1 if a 2, 3 or 5 is rolled, and takes value 0 otherwise.

These two examples are not terribly interesting; but they illustrate the important point that a single random procedure can accommodate several random variables. Note that “ $Y = 1$ ” and “ $Z = 0$ ” are **events**, in that they define subsets of the event space \mathcal{E} . The event “ $Y = 1$ ” is $\{2, 4, 6\}$. This is a crucial insight; it makes it feasible to obtain the probability distribution of a random variable.

This leads to the definition of a random variable from a formal perspective. Students accustomed to formal mathematical treatments of topics sometimes find the description of a random variable given so far somewhat elusive. A random variable can be defined formally in a way that strongly relates to mathematical topics that students have covered elsewhere, specifically, functions.

A random variable is a numerical-valued function that maps the event space \mathcal{E} to the set of real numbers. Students will not be familiar with a ‘variable’ that is a function. It is an important conceptual point, represented in the following diagram.



A random variable as a function from the event space \mathcal{E} to the set \mathbb{R} of real numbers.

From this diagram, we see that the random variable must take exactly one value for each element of the event space \mathcal{E} . So each possible outcome in the event space has a corresponding value for the random variable. As with functions generally, a number of possible outcomes in \mathcal{E} may have the same value of the random variable, and in practice this occurs frequently.

Probability functions

To work out the probability that a discrete random variable X takes a particular value x , we need to identify the event (the set of possible outcomes) that corresponds to “ $X = x$ ”.

In general, the function used to describe the probability distribution of a discrete random variable is called its **probability function** (abbreviated as **pf**). The probability function of X is the function $p_X: \mathbb{R} \rightarrow [0, 1]$ given by

$$p_X(x) = \Pr(X = x).$$

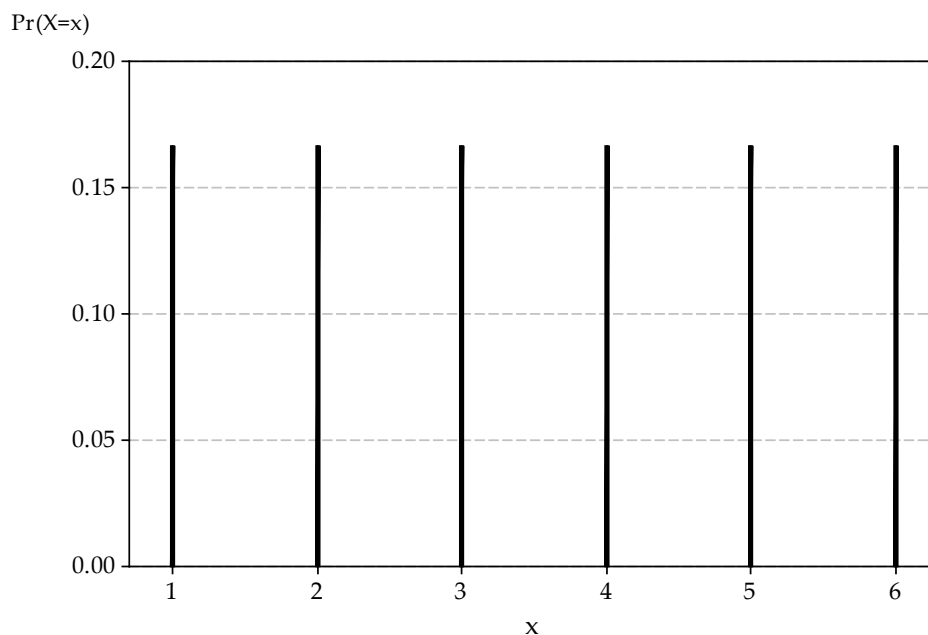
In general, the probability function $p_X(x)$ may be specified in a variety of ways. One way is to specify a numerical value for each possible value of X ; we have done that for the die-rolling example. In the die-rolling example, the random variable X can take exactly six values, and no others, and we assert that the probability that X takes any one of these values is the same, namely $\frac{1}{6}$.

As is the case generally for functions, the lower-case x here is merely the argument of the function. If we write $\Pr(X = y)$ it is essentially the same function, just as $f(x) = 2x^2 + 3x - 1$ and $f(y) = 2y^2 + 3y - 1$ are the same function. But it helps to associate the corresponding lower-case letter with the random variable we are considering.

Another way to specify the probability function is using a formula. We will see examples of this in the next section.

Less formally, the probability distribution may be represented using a graph, with a spike of height $p_X(x)$ at each possible value x of X . If there are too many possible values of X for this visual representation to work, we may choose to omit probabilities that are very close to zero; such values will typically be invisible on a graph anyway.

The following graph shows the probability function for the outcome of rolling a fair die.



The distribution of X , the number on the uppermost face when a fair die is rolled.

The importance of the notational convention to use capital letters for random variables now becomes apparent. A random variable X has a distribution. Formally, X is a function from the event space to the real line. An observed value of X , such as $x = 4$, is simply a number. Once we have rolled the die, we have the outcome. If we observed $x = 4$, this does not have a distribution. It is the observed value, and a number.

Quite often, there is more than one random variable being considered. That is the reason for writing $p_X(x)$ for the probability function of X , to distinguish it from $p_Y(y)$, say, the probability function of Y .

Since the probability function $p_X(x)$ is a probability, it must obey the axioms. Two important properties of a probability function follow from this:

- 1 $p_X(x) = \Pr(X = x) \geq 0$, for every real number x .
- 2 $\sum p_X(x) = 1$, where the sum is taken over all values x for which $p_X(x) > 0$.

In fact, these two properties characterise probability functions. If a function $f: \mathbb{R} \rightarrow \mathbb{R}$ satisfies these two properties, then we can view f as the probability function of a discrete random variable.

Exercise 1

Sneaky Sam has manufactured a deliberately biased six-sided die, with the following probability distribution for X , the number of spots on the uppermost face when the die is rolled.

x	1	2	3	4	5	6
$p_X(x)$	$\frac{1-\theta}{6}$	$\frac{1-\theta}{6}$	$\frac{1-\theta}{6}$	$\frac{1+\theta}{6}$	$\frac{1+\theta}{6}$	$\frac{1+\theta}{6}$

- What values of θ are feasible, in order for $p_X(x)$ to be a probability function?
- Find $\Pr(X \geq 4)$ in terms of θ .
- Find $\Pr(3 \leq X \leq 4)$.
- What is the probability of rolling an even number?

Random variables arising in practical settings of real importance are often much less readily dealt with than those arising in dice rolling. A common approach in such situations, which directly mirrors one of the approaches to probability, is to *estimate* the probability distribution of the random variable from data.

Example

Consider the random variable X defined to be the number of languages in which a randomly chosen Australian child attending school can hold an everyday conversation. Suppose we take a random sample of 1000 Australian school children and obtain the following data for the number of languages spoken.

x	1	2	3	4	5	6	Total
Frequency	663	226	66	22	19	4	1000

Then we obtain an estimate of $p_X(x)$, the pf of X , from the relative frequencies.

x	1	2	3	4	5	6
Estimate of $p_X(x)$	0.663	0.226	0.066	0.022	0.019	0.004

So we estimate that 22.6% of Australian school children can hold an everyday conversation in (exactly) two languages; equivalently, we estimate that $\Pr(X = 2) = 0.226$.

In this case, we cannot be sure that $x = 7$ (for example) is an impossible value. It may just be so rare that it did not crop up in the sample.

Examples of discrete distributions

Discrete uniform distribution

In the previous two sections, we considered the random variable defined as the number on the uppermost face when a fair die is rolled. This is an example of a random variable with a **discrete uniform distribution**. In general, for a positive integer m , let X be a discrete random variable with pf $p_X(x)$ given by

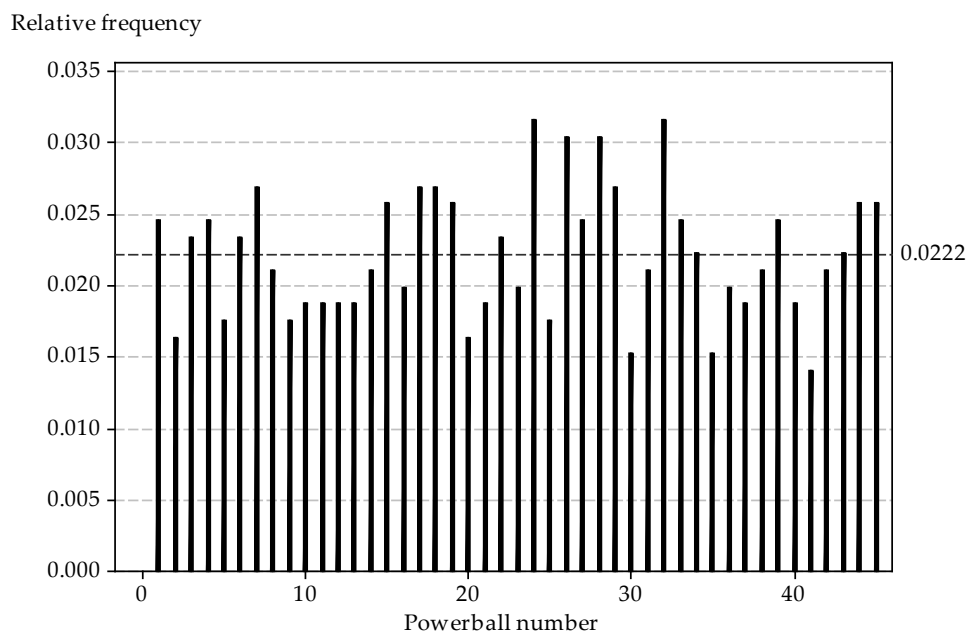
$$p_X(x) = \Pr(X = x) = \begin{cases} \frac{1}{m} & \text{if } x = 1, 2, \dots, m, \\ 0 & \text{otherwise.} \end{cases}$$

Then X has a discrete uniform distribution. This is a distribution that arises often in lotteries and games of chance.

We have seen this distribution in the Powerball example considered in the module *Probability*. In commercial lotteries, such as Powerball, it is a regulatory requirement that each outcome is equally likely. There are 45 possible Powerball numbers (1, 2, ..., 45). So if X is the Powerball drawn on a particular occasion, the pf $p_X(x)$ of X is given by

$$p_X(x) = \Pr(X = x) = \begin{cases} \frac{1}{45} & \text{if } x = 1, 2, \dots, 45, \\ 0 & \text{otherwise.} \end{cases}$$

If this model for the drawing of Powerball numbers is correct, we should expect that, over a large number of draws, the relative frequencies of the 45 possible numbers are all approximately equal to $\frac{1}{45} \approx 0.0222$. The following graph shows the relative frequencies observed in the 853 draws from May 1996 to September 2012.



Relative frequencies of Powerball numbers over 853 draws.
The model probability of 0.0222 is shown as a reference line.

Geometric distribution

The set of possible values for a discrete random variable may be infinite, as the following example shows.

Example

Fix a real number p with $0 < p < 1$. Now let X be the discrete random variable with probability function $p_X(x)$ given by

$$p_X(x) = \Pr(X = x) = \begin{cases} p(1-p)^x & \text{if } x = 0, 1, 2, \dots, \\ 0 & \text{otherwise.} \end{cases}$$

We will confirm that $p_X(x)$ has the two requisite properties of a probability function. First, we have $p_X(x) \geq 0$, since $p > 0$ and $1 - p > 0$. Second, we have

$$\begin{aligned} \sum p_X(x) &= p + p(1-p) + p(1-p)^2 + \dots \\ &= p(1 + (1-p) + (1-p)^2 + \dots) \\ &= p\left(\frac{1}{1-(1-p)}\right) && \text{(sum of an infinite geometric series)} \\ &= 1. \end{aligned}$$

We have already met one special distribution that is given a name (the discrete uniform distribution). The distribution in the previous example also has a name; it is called a **geometric distribution**.

Suppose that a sequence of independent ‘trials’ occur, and at each trial the probability of ‘success’ equals p . Define X to be the number of trials that occur before the first success is observed. Then X has a geometric distribution with parameter p .

We introduce here a symbol used throughout the modules on probability and statistics. If X has a geometric distribution with parameter p , we write $X \stackrel{d}{=} G(p)$. The symbol $\stackrel{d}{=}$ stands for ‘has the distribution’, meaning the distribution indicated immediately to the right of the symbol.

Note the use of the generic terms ‘trial’ and ‘success’. They are arbitrary, but they carry with them the idea of each observation involving a kind of test (i.e., trial) in which we ask the question: Which one of the two possibilities will be observed, a ‘success’ or a ‘failure’? In this sense, the words ‘success’ and ‘failure’ are just labels to keep track of the two possibilities for each trial.

Note that X can take the value 0, if a success is observed at the very first trial. Or it can take the value 1, if a failure is observed at the first trial and then a success at the second trial. And so on. What is the largest value that X can take? There is no upper limit, in theory. As the values of X increase, the probabilities become smaller and smaller. The sum of the probabilities equals one, as shown in the previous example.

Exercise 2

Recall that Tetris pieces have seven possible shapes, labelled I, J, L, O, S, T, Z. Assume that, at any stage of the game, all seven shapes are equally likely to be produced, independent of whatever pieces have been produced previously.

Consider the sequences of three consecutive pieces observed during a game of Tetris, such as JLL, ZOS, ZSZ, III and so on.

- a What is the probability that a sequence of three pieces does not contain a Z?
- b Hence, what is the probability that a sequence of three pieces has at least one Z?
- c What is the probability that a Z occurs in the first three-piece sequence observed?
- d What is the probability that a Z does not occur in the first three-piece sequence, but does occur in the second such sequence? (Here we are considering non-overlapping sequences of three pieces.)
- e What is the probability that a Z does not occur in the first x three-piece sequences, but does occur in the $(x + 1)$ st such sequence?
- f Hence, what is the distribution of X , the number of three-piece sequences observed before a three-piece sequence with a Z in it first appears? Write down its probability function $p_X(x)$.
- g Evaluate $p_X(x)$ for $x = 2, 3, 4, 10$.

Exercise 3

Julia and Tony play the hand game *Rock-paper-scissors*. Assume that, at each play, they make their choices with equal probability ($\frac{1}{3}$) for each of the three moves, independently of any previous play.

- a On any single play, what is the chance of a tie?
- b What is the chance that Julia wins at the first play?
- c Define X to be the number of plays before Julia wins for the first time. What is the distribution of X ?
- d Write down the probability function $p_X(x)$ of X .
- e Find $\Pr(X = 5)$.
- f Find $\Pr(X \geq 5)$.

Another important distribution arises in the context of a sequence of independent trials that each have the same probability of success p . This is the **binomial distribution**. There is an entire module devoted to it (*Binomial distribution*), so we do not consider it further here.

Mean of a discrete random variable

If you roll a fair die many times, what will be the average outcome? Imagine rolling it 6000 times. You would expect to roll about 1000 ones, 1000 twos, and so on: about 1000 occurrences of each possible outcome. What would be the average value of the outcomes obtained? Approximately, the average or mean would be

$$\frac{(1000 \times 1) + (1000 \times 2) + \cdots + (1000 \times 6)}{6000} = \frac{21\,000}{6000} = 3.5.$$

This can be thought of as the weighted average of the six possible values $1, 2, \dots, 6$, with weights given by the relative frequencies. Note that 3.5 is not a value that we can actually observe.

By analogy with data and relative frequencies, we can define the mean of a discrete random variable using probabilities from its distribution, as follows.

The **mean** μ_X of a discrete random variable X with probability function $p_X(x)$ is given by

$$\mu_X = \sum x p_X(x),$$

where the sum is taken over all values x for which $p_X(x) > 0$.

The mean can be regarded as a measure of ‘central location’ of a random variable. It is the weighted average of the values that X can take, with weights provided by the probability distribution.

The mean is also sometimes called the **expected value** or **expectation** of X and denoted by $E(X)$. These are both somewhat curious terms to use; it is important to understand that they refer to the long-run average. The mean is the value that we expect the long-run average to approach. It is not the value of X that we expect to observe.

Consider a random variable U that has the discrete uniform distribution with possible values $1, 2, \dots, m$. The mean is given by

$$\begin{aligned}\mu_U &= \sum_{x=1}^m \left(x \times \frac{1}{m} \right) \\ &= \frac{1}{m} \sum_{x=1}^m x \\ &= \frac{1}{m} \times \frac{m(m+1)}{2} \\ &= \frac{m+1}{2}.\end{aligned}$$

For example, the mean for the roll of a fair die is $\frac{6+1}{2} = 3.5$, as expected.

So in the long run, rolling a single die many times and obtaining the average of all the outcomes, we ‘expect’ the average to be close to 3.5, and the more rolls we carry out, the closer the average will be.

The use of the terms ‘expected value’ and ‘expectation’ is the reason for the notation $E(X)$, which also extends to functions of X .

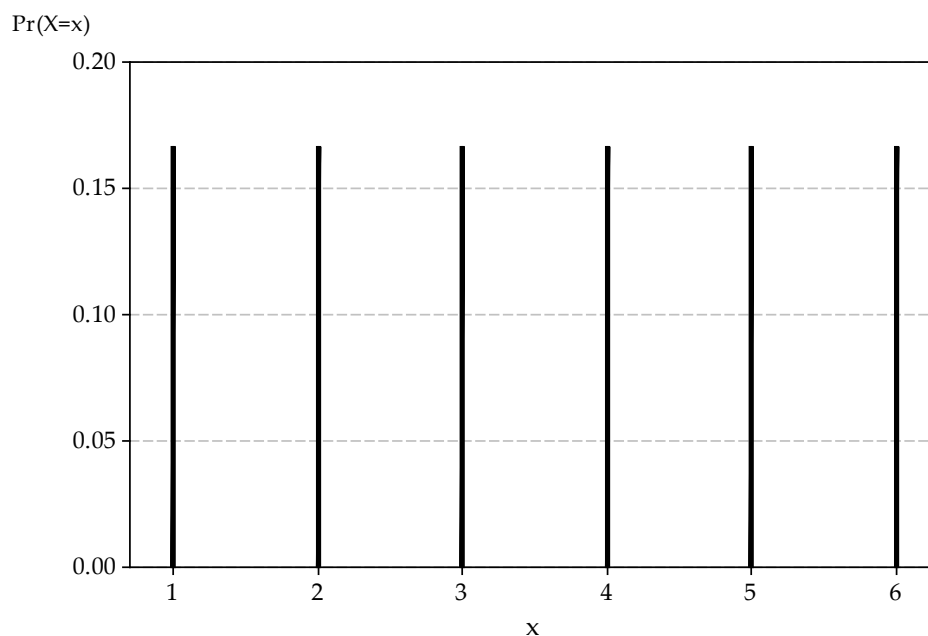
Exercise 4

Consider again the biased die made by Sneaky Sam. Recall that the distribution of X , the number of spots on the uppermost face when the die is rolled, is as follows.

x	1	2	3	4	5	6
$p_X(x)$	$\frac{1-\theta}{6}$	$\frac{1-\theta}{6}$	$\frac{1-\theta}{6}$	$\frac{1+\theta}{6}$	$\frac{1+\theta}{6}$	$\frac{1+\theta}{6}$

- Find μ_X , the mean of X .
- What is the largest possible value of μ_X ?

The following graph shows once again the probability function for the outcome of rolling a fair die. This distribution is symmetric, and the mean 3.5 is in the middle of the distribution; in fact, it is on the axis of symmetry.



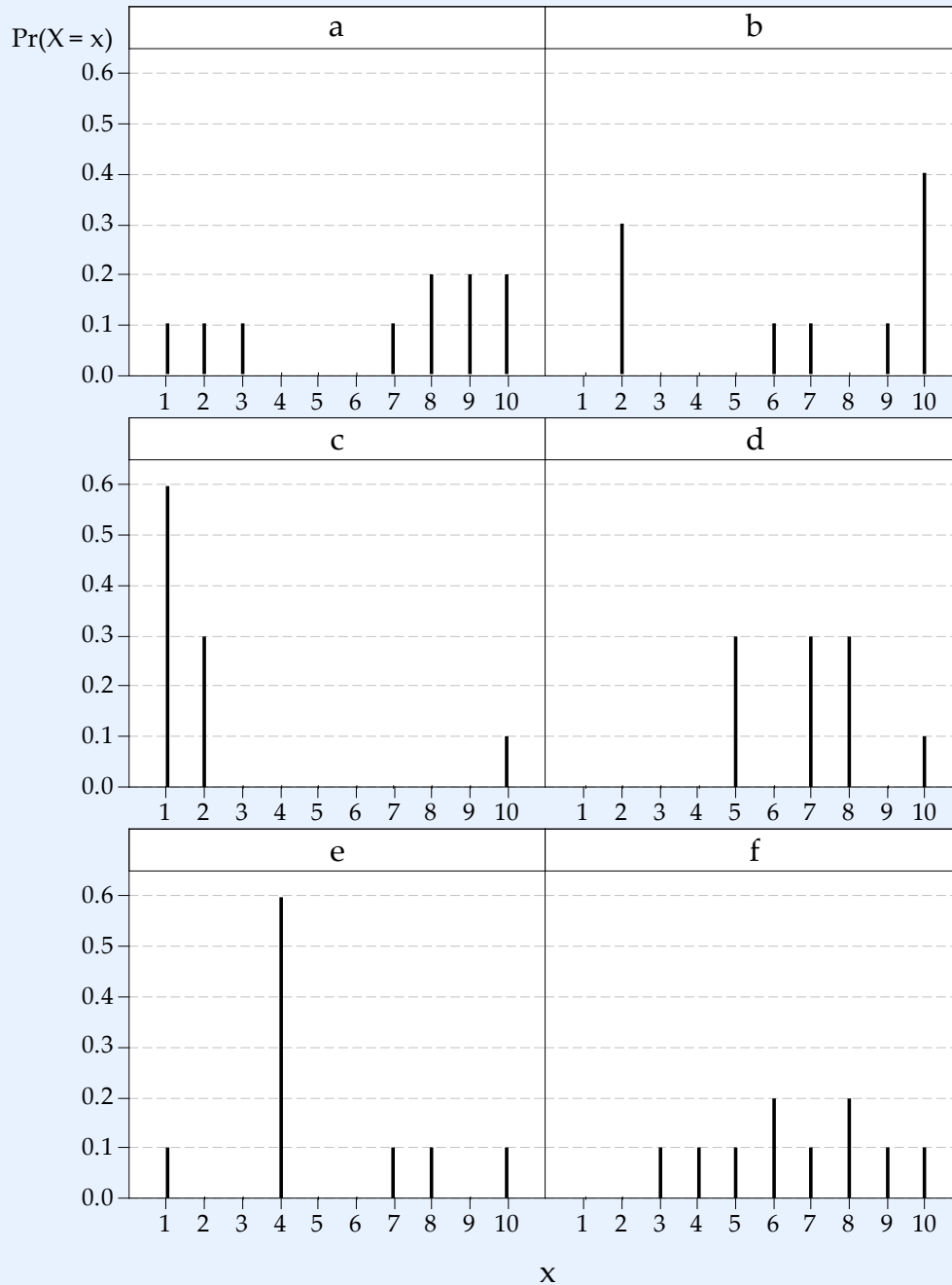
The distribution of X , the number on the uppermost face when a fair die is rolled.

We can give a general physical interpretation of the mean of a discrete random variable X with pf $p_X(x)$. Suppose we imagine that the x -axis is an infinite see-saw in each direction, and we place weights equal to $p_X(x)$ at each possible value x of X . Then the mean μ_X is at the point which will make the see-saw balance. In other words, it is at the centre of mass of the system.

If the distribution of a discrete random variable is represented graphically, then you should be able to guess the value of its mean, at least approximately, by using the 'centre of mass' idea. This is the topic of the next exercise.

Exercise 5

The distributions (labelled 'a' to 'f') of six different random variables are shown below.



For each distribution separately:

- a Confirm that the graph represents a probability function.
- b Guess the value of the mean of the corresponding random variable.
- c Calculate the value of the mean.

Suppose that X has a geometric distribution with parameter p , and therefore its probability function is

$$p_X(x) = p(1-p)^x, \quad x = 0, 1, 2, \dots$$

Recall that X is the number of trials before the first success in a sequence of independent trials, each with probability of success p . Do you expect there to be many trials before the first success, on average, or just a few?

A result which we state here without proof is that, for $X \stackrel{d}{=} G(p)$, we have

$$\mu_X = \frac{1-p}{p}.$$

If p is large (that is, close to 1), then successes are very likely and the wait before the first success is likely to be short; in this case, μ_X is small. On the other hand, if p is small (close to 0), then failures are very likely and the wait before the first success is likely to be long; in this case, μ_X is large.

Exercise 6

- One of the standard forms of commercial lottery selects six balls at random out of 45. What is the chance of winning first prize in such a lottery with a single entry?
- Suppose that someone buys a single entry in every draw. What is the distribution of the number of draws entered before the player wins first prize for the first time?
- What is the expected number of draws before winning first prize for the first time?
- Suppose the draws occur weekly. On average, how many years does the person have to wait before winning first prize for the first time?

We may wish to find the mean of a function of a random variable X , such as X^2 or $\log X$. For a discrete random variable X with pf $p_X(x)$, consider an arbitrary function of X , say $Y = g(X)$. Then the expectation of Y , that is, $E(Y) = \mu_Y$, is obtained as follows:

$$\begin{aligned} \mu_Y &= \sum y \Pr(Y = y) && \text{(summing over } y \text{ for which } \Pr(Y = y) > 0) \\ &= \sum g(x) \Pr(X = x) && \text{(summing over } x \text{ for which } \Pr(X = x) > 0) \\ &= \sum g(x) p_X(x). \end{aligned}$$

For the special case of a linear transformation $Y = aX + b$, we shall see that it follows that $\mu_Y = a\mu_X + b$. This is a very useful result; it says that, for a linear transformation, the mean of the transformed variable equals the transformation of the mean of the original variable. In particular, if $Y = aX$, then $\mu_Y = a\mu_X$, as you might expect. This applies to changes of units: for example, if the random variable X measures a time interval in days,

and we wish to consider the equivalent time in hours, then we can define $Y = 24X$ and we know that $\mu_Y = 24\mu_X$.

For a transformation $Y = g(X)$, it is *not true in general* that $\mu_Y = g(\mu_X)$. But in the special case of a linear transformation $Y = aX + b$, where $g(x) = ax + b$, we have

$$\begin{aligned}\mu_Y &= \sum g(x) p_X(x) \\ &= \sum (ax + b) p_X(x) \\ &= \sum ax p_X(x) + \sum b p_X(x) \\ &= a \sum x p_X(x) + b \sum p_X(x) \\ &= a\mu_X + b,\end{aligned}$$

as claimed.

Technical note. Every discrete random variable X with a finite set of possible values has a mean μ_X . But it is possible to construct examples where the mean does not exist. For example, consider the discrete random variable X with probability function $p_X(x) = \frac{1}{x}$, for $x = 2, 4, 8, 16, \dots$. Here $\sum p_X(x) = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1$, as required for a probability function. But $\mu_X = \sum x p_X(x) = \sum_{n=1}^{\infty} 1$ does not exist. Such complications are not considered in secondary school mathematics.

Variance of a discrete random variable

We have seen that the mean of a random variable X is a measure of the central location of the distribution of X . If we are summarising features of the distribution of X , it is clear that location is not the only relevant feature. The second most important feature is the spread of the distribution.

If values of X near its mean μ_X are very likely and values further away from μ_X have very small probability, then the distribution of X will be closely concentrated around μ_X . In this case, the spread of the distribution of X is small. On the other hand, if values of X some distance from its mean μ_X are likely, the spread of the distribution of X will be large.

These ideas lead to the most important measure of spread, the **variance**, and a closely related measure, the **standard deviation**.

Students have met the concepts of variance and standard deviation when summarising data. These were the sample variance and the sample standard deviation. The difference here is that we are referring to properties of the distribution of a random variable.

The variance of a random variable X is defined by

$$\text{var}(X) = E[(X - \mu)^2], \quad \text{where } \mu = E(X).$$

For a discrete random variable X , the variance of X is obtained as follows:

$$\text{var}(X) = \sum (x - \mu)^2 p_X(x),$$

where the sum is taken over all values of x for which $p_X(x) > 0$. So the variance of X is the weighted average of the squared deviations from the mean μ , where the weights are given by the probability function $p_X(x)$ of X .

The standard deviation of X is defined to be the square root of the variance of X . That is,

$$\text{sd}(X) = \sigma_X = \sqrt{\text{var}(X)}.$$

Because of this definition, the variance of X is often denoted by σ_X^2 .

In some ways, the standard deviation is the more tangible of the two measures, since it is in the same units as X . For example, if X is a random variable measuring lengths in metres, then the standard deviation is in metres (m), while the variance is in square metres (m^2).

Unlike the mean, there is no simple direct interpretation of the variance or standard deviation. The variance is analogous to the moment of inertia in physics, but that is not necessarily widely understood by students. What is important to understand is that, in relative terms:

- a small standard deviation (or variance) means that the distribution of the random variable is narrowly concentrated around the mean
- a large standard deviation (or variance) means that the distribution is spread out, with some chance of observing values at some distance from the mean.

Note that the variance cannot be negative, because it is an average of squared quantities. This is appropriate, as a negative spread for a distribution does not make sense. Hence, $\text{var}(X) \geq 0$ and $\text{sd}(X) \geq 0$ always.

Example

Consider the rolling of a fair six-sided die, with X the number on the uppermost face. We know that the pf of X is

$$p_X(x) = \frac{1}{6}, \quad x = 1, 2, 3, 4, 5, 6,$$

and that $\mu_X = 3.5$. The variance of X is given by

$$\begin{aligned} \text{var}(X) &= E[(X - \mu_X)^2] \\ &= \sum (x - \mu_X)^2 p_X(x) \\ &= \sum (x - \mu_X)^2 \frac{1}{6} \\ &= \frac{1}{6} \left((1 - 3.5)^2 + (2 - 3.5)^2 + (3 - 3.5)^2 + (4 - 3.5)^2 + (5 - 3.5)^2 + (6 - 3.5)^2 \right) \\ &= \frac{1}{6} \times 17.5 \\ &= \frac{35}{12} \approx 2.9167. \end{aligned}$$

Hence, the standard deviation of X is $\sigma_X = \sqrt{\frac{35}{12}} \approx 1.7078$.

Exercise 7

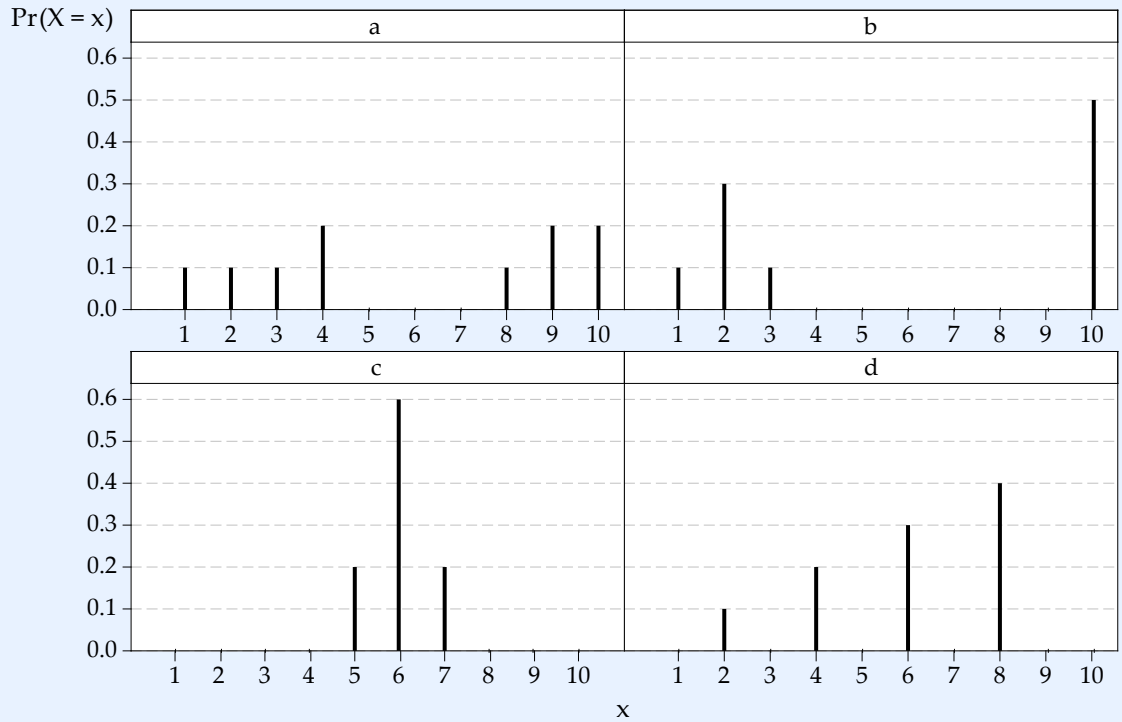
Consider again the example of the number of languages spoken by Australian school children. Define X to be the number of languages in which a randomly chosen Australian child attending school can hold an everyday conversation. Assume that the probability function of X , $p_X(x)$, is as shown in the following table.

x	1	2	3	4	5	6
$p_X(x)$	0.663	0.226	0.066	0.022	0.019	0.004

- What is the mean of X ?
- Find the variance and standard deviation of X .
- What are the units of the mean and standard deviation?

Exercise 8

Consider the following four discrete distributions.



- Confirm that each graph represents a probability function.
- Guess the value of the mean of each corresponding random variable.
- Calculate the value of the mean in each case.
- Guess the order of the variances, from largest to smallest.
- Find the variance and standard deviation in each case.

Answers to exercises

Exercise 1

- a We need $\sum p_X(x) = 1$, but this holds for any value of θ . We also need $p_X(x) \geq 0$. This holds if and only if $-1 \leq \theta \leq 1$.
- b $\Pr(X \geq 4) = 3 \times \left(\frac{1+\theta}{6}\right) = \frac{1}{2}(1 + \theta)$.
- c $\Pr(3 \leq X \leq 4) = \Pr(X = 3) + \Pr(X = 4) = \frac{1-\theta}{6} + \frac{1+\theta}{6} = \frac{1}{3}$.
- d $\Pr(X \text{ is even}) = \frac{1}{6}(3 + \theta)$.

Exercise 2

- a Assuming independence, $\left(\frac{6}{7}\right)^3 \approx 0.6297$.
- b $1 - \left(\frac{6}{7}\right)^3 \approx 0.3703$.
- c 0.3703.
- d $0.6297 \times 0.3703 \approx 0.2332$.
- e $0.6297^x \times 0.3703$.
- f X has a geometric distribution with parameter $p = 0.3703$. The probability function is given by $p_X(x) = 0.3703 \times 0.6297^x$, for $x = 0, 1, 2, \dots$
- g $p_X(2) = 0.1468$, $p_X(3) = 0.0925$, $p_X(4) = 0.0582$, $p_X(10) = 0.0036$.

Exercise 3

- a There are nine possible outcomes in a single play, three of which are ties. So the probability of a tie is $\frac{1}{3}$.
- b The chance of Julia winning on any single play, including on the first play, is $\frac{1}{3}$.
- c X has a geometric distribution with parameter $p = \frac{1}{3}$.
- d $p_X(x) = \frac{1}{3} \left(\frac{2}{3}\right)^x$, for $x = 0, 1, 2, \dots$
- e $p_X(5) = \frac{1}{3} \left(\frac{2}{3}\right)^5 \approx 0.0439$.
- f There are at least two ways to approach this problem. One method is to calculate:

$$\Pr(X \geq 5) = \sum_{x=5}^{\infty} p_X(x) = \sum_{x=5}^{\infty} \frac{1}{3} \left(\frac{2}{3}\right)^x = p_X(5) \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n = \frac{p_X(5)}{1 - \frac{2}{3}} \approx 0.1317,$$

using the formula for the sum of an infinite geometric series.

Alternatively, note that the event " $X \geq 5$ " (the first success is after at least five failures) occurs precisely when the first five trials are failures. The probability of failure in a single trial is $1 - p = \frac{2}{3}$ and the trials are independent, so $\Pr(X \geq 5) = \left(\frac{2}{3}\right)^5 \approx 0.1317$.

Exercise 4

- a $\mu_X = \sum x p_X(x) = 6\left(\frac{1-\theta}{6}\right) + 15\left(\frac{1+\theta}{6}\right) = \frac{1}{2}(7 + 3\theta)$.
- b From exercise 1, the largest value for θ is 1. So the largest mean is $\frac{1}{2}(7 + 3) = 5$.

Exercise 5

- a In each case, the probabilities are all clearly non-negative and the sum of the probabilities equals one, as required.
- b You should be able to get fairly close to the actual means (see part c) visually.
- c In each case, the mean is given by $\mu_X = \sum x p_X(x)$.

	a	b	c	d	e	f
μ_X	6.7	6.8	2.2	7.0	5.0	6.6

Exercise 6

- a The chance of a single entry winning first prize is equal to $\frac{1}{\binom{45}{6}} = \frac{1}{8\,145\,060}$.
- b The random variable X , the number of draws before a first prize is obtained, has a geometric distribution with parameter $p = \frac{1}{8\,145\,060}$.
- c The mean of X is given by $\frac{1-p}{p} = 8\,145\,059$.
- d About 157 000 years.

Exercise 7

- a $\mu_X = 1.52$.
- b $\text{var}(X) = 0.8216$, $\text{sd}(X) \approx 0.9064$.
- c The units are ‘number of languages’.

Exercise 8

- a For each graph, we have $p_X(x) \geq 0$ and $\sum p_X(x) = 1$.
- b You should be able to get close to the actual means (see part c) visually.
- c Each of the four means is equal to 6.
- d Guessing variances is much more difficult than guessing means. But for these four distributions, the *order* (largest to smallest) should be clear visually: $b > a > d > c$.

e

	a	b	c	d
$\text{var}(X)$	11.2	16.2	0.4	4.0
$\text{sd}(X)$	3.35	4.02	0.63	2.0

0 1 2 3 4 5 6 7 8 9 10 11 12