# Chapter 2

# Random Variables

## 2.1 Random variables and probability distributions

### The definition of a random variable

#### Example 1

The FTSE 100 index indicates the total value of the 100 most highly capitalised UK companies listed on the London Stock Exchange. Suppose we decide to record the number of these shares that fall in value tomorrow. The outcome of this experiment is intrinsically numerical, and a suitable sample space is $S = \{0, 1, 2, \ldots, 100\}$.

#### Example 2

We might ask a University of Glasgow student to respond to the statement, "MyCampus is working well for recording student data", by choosing one of the options: Strongly disagree, Disagree, Neutral, Agree, Strongly agree. A suitable sample space is $S =$ {'Strongly disagree', 'Disagree', 'Neutral', 'Agree', 'Strongly agree'}.

The outcome of this experiment is not intrinsically numerical but we might choose to represent the outcomes by the following numerical codes $\{-2, -1, 0, 1, 2\}$.

#### Example 3

Suppose we record a thermograph, a continuous trace of the ambient temperature, for 24 hours. The outcome of this experiment is a continuous, time-dependent function. Suppose that we read off the maximum temperature recorded during the day. Though the outcome of this experiment is not simply a number, we have now derived a single number of considerable interest from it.

#### Example 4

Some knock-out football competitions have experimented with the following scheme for deciding the outcome of a match. Normally, a football match lasts 90 minutes. If one team has scored more goals than the opposition at the end of 90 minutes, then the game ends and that team is the outright winner. If the teams are drawing after 90 minutes, then the match goes into extra playing time. The first team to score a goal wins the match (which is stopped immediately). If no goal is scored in 30 minutes of extra time, then play is stopped and the match is decided on a penalty shoot-out. The playing time (minutes) of a match played under these rules is a number in the range $[90, 120]$.

## Definition

A **random variable** is a function, $X : S \to \mathbb{R}$, which associates a single numerical value $X(s)$ with every outcome $s$ in the sample space $S$.

When the outcomes in $S$ are intrinsically numerical, the identity function often defines the random variable of particular interest.

Every time the experiment is conducted, one and only one outcome $s \in S$ is observed so one and only one value, $X(s)$, of the random variable is observed. This is called a realisation of the random variable. It is conventional to represent a random variable by a capital letter (such as $X$), and a general realisation of it by the corresponding lower-case letter (for example, $x$).

## Definition

The range space, $R_X$, of the random variable $X$ is the set of all possible realisations of $X$. $R_X = S$ if the sample space is intrinsically numerical and the identity function defines the random variable of interest.

### Example 1 (continued)

Let the random variable $X$ be the number of the 100 leading shares that fall in value tomorrow. Then $X$ is the identity function:

$$X(s) = s \qquad \text{for every outcome } s \in S$$

and $R_X = S = \{0, 1, 2, \ldots, 100\}$.

### Example 2 (continued)

The random variable $Y$ can be defined as follows: $Y(\text{'Strongly disagree'}) = -2$, $Y(\text{'Disagree'}) = -1$, $Y(\text{'Neutral'}) = 0$, $Y(\text{'Agree'}) = 1$, $Y(\text{'Strongly agree'}) = 2$ and $R_Y = \{-2, -1, 0, 1, 2\}$.

### Example 3 (continued)

The random variable $Z$ is defined by:

$$Z = \max[g(t); 0 \leqslant t \leqslant 24],$$

where $g(t)$ is the temperature (°C) recorded at time $t$ and $R_Z = [-273.15, \infty)$.

The outcome of the underlying experiment completely determines which realisation of a random variable is observed. This allows us to speak of probabilities associated with $X$, for example $P(X = x)$ or $P(X \leqslant x)$, as long as we recognise that such probabilities are derived from probabilities associated with the original sample space. For example, $P(X \leqslant x)$ is the probability of the equivalent event $\{s \in S : X(s) \leqslant x\}$.

**Example 5**

A psychologist who is conducting learning experiments with rats has prepared a simple maze with just two narrow paths through it. When the rat is introduced into the maze, it must go down one of the paths. One leads to food, the other does not. Each rat is introduced to the same maze three times, and the outcome on each trial is noted.

A suitable sample space for this experiment is

$$S = \{SSS, SSF, SFS, SFF, FSS, FSF, FFS, FFF\}$$

where, for example, $SSF$ means that the rat succeeds in reaching the food on the first two attempts, but fails on the third. Assuming that a rat does not learn from its experience (good or bad) on any of its attempts at the maze, then each of these eight outcomes is equally likely with probability $\frac{1}{8} = 0.125$.

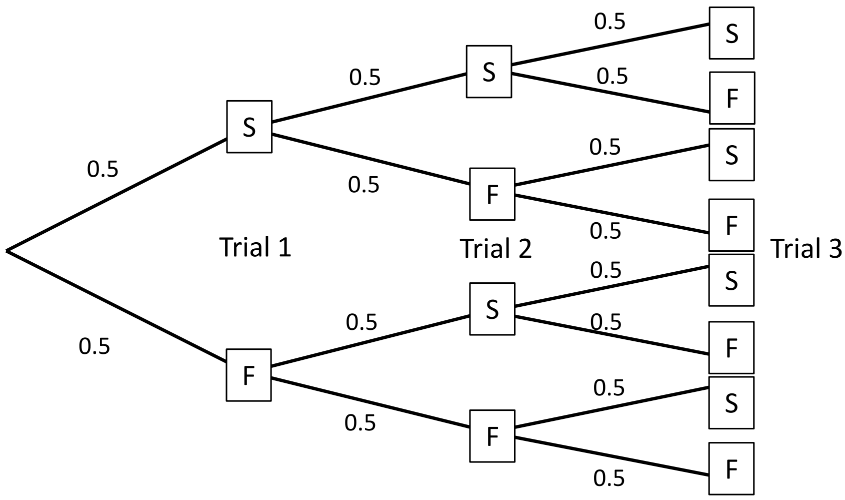


Figure 2.1: Representation as a tree of the rat's choices.

Let the random variable $X$ be the number of times the rat succeeds in reaching the food. Then, $R_X = \{0, 1, 2, 3\}$ and

$$P(X = 0) =$$
$$P(X = 1) =$$
$$P(X = 2) =$$
$$P(X = 3) =$$

This model does not seem very realistic. We would expect a rat to learn where the food is in the course of the experiment. Suppose that the psychologist believes instead that a rat has probability 0.5 of reaching the food on its first attempt, but that, once the rat succeeds in reaching food, it will always succeed in reaching food on subsequent trials. Until the rat first reaches food, its chance of reaching food on a subsequent trial is unchanged at 0.5.

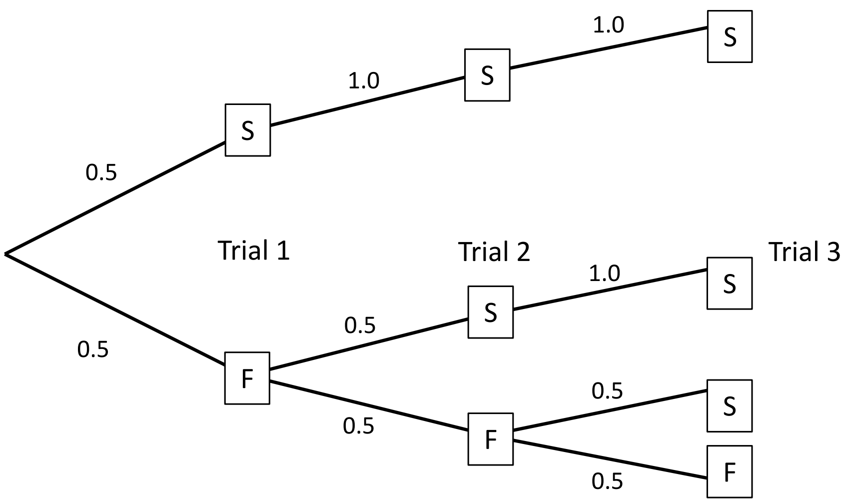


Figure 2.2: Representation as a tree of the rat's more realistic choices.

In this case,

$$P(X = 0) =$$
$$P(X = 1) =$$
$$P(X = 2) =$$
$$P(X = 3) =$$

Notice that, in both models, the four probabilities associated with the random variable $X$ add up to 1. This has to be true, since the events $\{s \in S : X(s) = 0\}$, $\{s \in S : X(s) = 1\}$, $\{s \in S : X(s) = 2\}$ and $\{s \in S : X(s) = 3\}$ must partition the sample space, $S$.

## The distribution function and percentiles

Once we have identified a random variable, $X$, that records the information from the experiment that is of interest to us, we can restrict our attention to probabilities associated with $X$. For reasons we will explore over the next couple of lectures, probabilities of the form $P(X \leqslant x)$ are generally of most interest to us theoretically.

### Definition

The **distribution function** (d.f.), sometimes known as the **cumulative distribution function** (c.d.f.), of a random variable, $X$, is defined as

$$F_X(x) = P(X \leqslant x) \qquad x \in \mathbb{R}.$$

This function is defined for every real value, $x$, whether or not $x \in R_X$. Any valid distribution function $F_X$ has the following properties:

(i) $F_X(-\infty) = 0$ and $F_X(\infty) = 1$;

(ii) whenever $a \leqslant b$, then $F_X(a) \leqslant F_X(b)$, i.e., $F_X$ is non-decreasing on $\mathbb{R}$;

(iii) for all $x \in \mathbb{R}$, $\lim_{h \to 0^+} F_X(x + h) = F_X(x)$, i.e., $F_X$ is continuous on the right.

The distribution function incorporates all the probability information we need about X, since all other probabilities associated with X can be recovered from it. For example,

$$P(a < X \leqslant b) = P(X \leqslant b) - P(X \leqslant a) = F(b) - F(a).$$

**Example 5 (continued)**

We can construct the distribution function as follows:

$$\begin{array}{ll}
x < 0 & F_X(x) = P(X \leqslant x) = 0 \quad \text{(since this event is impossible)} \\
x = 0 & F_X(0) = P(X \leqslant 0) = P(X = 0) = 0.125 \\
0 < x < 1 & F_X(x) = P(X \leqslant x) = P(X = 0) = 0.125 \\
x = 1 & F_X(1) = P(X \leqslant 1) = P(X = 0) + P(X = 1) = 0.25 \\
\vdots & \quad \vdots \\
x = 3 & F_X(3) = P(X \leqslant 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1 \\
x > 3 & F_X(x) = P(X \leqslant x) = P(X \leqslant 3) = 1 \quad \text{(since this event is certain)}
\end{array}$$

**Definition**

A random variable that has a finite or countable range space is known as a **discrete random variable**.

The random variable in Example 5 is discrete. Every discrete random variable has a distribution function that consists of a series of discrete steps at the values $x \in R_X$. Such a distribution function is right continuous but not left continuous and therefore not continuous.

**Example 6**

Many simulation studies begin by drawing a "random number", X, between 0 and 1 (inclusive). One way of thinking about this experiment is to say that the probability that X lies in any interval $(a, b] \subseteq (0, 1]$ is equal to the width of the interval $(a, b]$, i.e.,

$$P(a < X \leqslant b) = b - a.$$

We can construct the distribution function of X as follows:

$$\begin{array}{ll}
x \leqslant 0 & F_X(x) = 0 \quad \text{(an impossible event)} \\
0 < x \leqslant 1 & F_X(x) = P(X \leqslant x) = P(0 < X \leqslant x) = x - 0 = x \\
x \geqslant 1 & F_X(x) = P(X \leqslant x) = 1 \quad \text{(a certain event)}
\end{array}$$

This random variable has a range space that is uncountably infinite, so it is not discrete. Its distribution function is not only right continuous but continuous everywhere in $\mathbb{R}$. Such random variables are said to be **continuous**.

**Example 4 (continued)**

Extensive research shows that there is probability 0.7 of a match being won within the 90 minutes of normal play. There is also probability 0.1 that a match will go to a penalty shoot out at the end of 120 minutes of play. The length of a match that is decided by a goal scored between 90 and 120 minutes of play is random (in the sense of Example 6) over that range.

The distribution function of the random variable, $X$ = the length of play in a randomly-selected match determined on this basis, is given by the following formula:

$$F_X(x) = \begin{cases} 0, & x < 90 \\ 0.7 + \frac{0.2(x-90)}{(120-90)}, & 90 \leqslant x < 120 \\ 1, & 120 \leqslant x \end{cases}.$$

Since $R_X = [90, 120]$, then $X$ is not a discrete random variable. At first sight, it appears to be continuous but note that $F_X(x)$ is not left continuous (i.e., not continuous) at $x = 90$ or $x = 120$; so $X$ is not a continuous random variable either.

Notice that $F_X(x)$ can be expressed as the weighted sum of a discrete distribution function, $F_d(x)$, and a continuous distribution function, $F_c(x)$, since $F_X(x) = 0.8F_d(x) + 0.2F_c(x)$, where

$$F_d(x) = \begin{cases} 0, & x < 90 \\ \frac{7}{8}, & 90 \leqslant x < 120 \\ 1, & 120 \leqslant x \end{cases}, \qquad F_c(x) = \begin{cases} 0, & x < 90 \\ \frac{x-90}{30}, & 90 \leqslant x < 120 \\ 1, & 120 \leqslant x \end{cases}.$$

$X$ in this example is a **mixed random variable**. In general, a random variable is called mixed if its distribution function can be written in the form

$$F_X(x) = \lambda F_d(x) + (1-\lambda)F_c(x),$$

where $0 < \lambda < 1$, $F_d(x)$ is a discrete distribution function and $F_c(x)$ is a continuous distribution function.

**Definition**

Let $X$ be a random variable with distribution function $F_X$. For $0 < \alpha < 1$, the $100\alpha$th **percentile** of $X$ (or, equivalently, of its probability distribution) is the smallest value of $x$ for which $F_X(x) \geqslant \alpha$. In particular, the **median** is the 50th percentile of a distribution.

**Example 5 (continued)**

The plot of the distribution function shows that $F_X(2) = 0.5$ while $F_X(x) < 0.5$ for $x < 2$. Therefore, the median is 2.

**Example 6 (continued)**

Since $X$ is continuous, we can find the median by solving the equation $F_X(x) = 0.5$, i.e., $x = 0.5$. Therefore, the median is 0.5.

**Example 4 (continued)**

There is no value of $x$ for which $F_X(x) = 0.5$. At $x = 90$, $F_X(x)$ jumps from 0 to 0.7. Therefore $x = 90$ is the smallest value for which $F_X(x) \geqslant 0.5$. So the median is 90.

## 2.2 Some properties of random variables

### The probability mass function

The distribution function does not usually give the most informative summary of the probability information about a random variable. Different random variables can have distribution functions that look alike and this reduces the usefulness of the distribution function as a way of distinguishing between different distributions.

#### Definition

Let $X$ be a discrete random variable. The probability mass function (p.m.f.) is defined as

$$p_X(x) = P(X = x) \qquad x \in \mathbb{R}.$$

Now $P(X = x) = 0$ for $x \notin R_X$. So, when discussing a particular p.m.f., it is usual to restrict attention to the list of values $\{(x, p_X(x)), x \in R_X\}$.

There are two general properties of all valid probability mass functions:

(a) $0 \leqslant p_X(x) \leqslant 1$, for all $x \in \mathbb{R}$. This follows because $p_X(x)$ is a probability.

(b)

$$\sum_{x \in R_X} p_X(x) = 1.$$

This follows because the events $\{s : X(s) = x\}$, for $x \in R_X$, must partition $S$.

#### Example 7

To test the hardness of a batch of plastic, a rod made from the plastic is hit repeatedly with a known force until it breaks. Let the random variable $X$ be the number of hits required until the rod breaks. If the plastic has been manufactured as specified, then there is probability 0.5 that the rod breaks on any given hit. It is assumed that the outcomes of different hits are independent.

The range space of $X$ is $R_X = \{1, 2, 3, \dots\}$. Also

$$\begin{aligned}
p_X(1) &= P(X = 1) = P(\text{rod breaks on 1st hit}) = 0.5 \\
p_X(2) &= P(X = 2) = P(\text{rod does not break on 1st hit but breaks on 2nd hit}) \\
&= P(\text{rod does not break on 1st hit}) \times P(\text{rod breaks on 2nd hit}) \\
&= 0.5 \times 0.5 = 0.25 \\
p_X(3) &= P(X = 3) \\
&= P(\text{rod does not break on 1st hit}) \times P(\text{rod does not break on 2nd hit}) \\
&\quad \times P(\text{rod breaks on 3rd hit}) \\
&= 0.5 \times 0.5 \times 0.5 = 0.125 \\
p_X(x) &= P(X = x) = \left(\tfrac{1}{2}\right)^x \qquad x = 1, 2, 3, \dots
\end{aligned}$$

Now

$$\sum_{x \in R_X} p_X(x) = \sum_{x=1}^{\infty} p_X(x) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots .$$

This is an example of the sum to infinity of a **Geometric Series**, which can be written in the general form:

$$T = a + ar + ar^2 + ar^3 + \cdots .$$

Each term of this series is found from the previous one by multiplying it by the constant $r$. When $-1 < r < 1$, the sum to infinity of the series is:

$$T = \frac{a}{1-r}.$$

When $r \geqslant 1$ or $r \leqslant -1$, the sum $T$ does not converge to any finite limit. The sum of the first $k$ terms of a Geometric Series is:

$$T_k = a + ar + \cdots + ar^{k-1} = \frac{a(1-r^k)}{1-r}.$$

**Example 7 (continued)**

$\sum_{x \in R_X} p_X(x)$ is the sum to infinity of a Geometric Series with $a = \frac{1}{2}$ and $r = \frac{1}{2}$, which is

$$T = \frac{0.5}{1-0.5} = 1,$$

as required for a valid p.m.f.. The probability that 3 or fewer blows are required to break a rod is

$$\begin{aligned} P(X \leqslant 3) &= P(X=1) + P(X=2) + P(X=3) \\ &= p_X(1) + p_X(2) + p_X(3) \\ &= 0.5 + 0.25 + 0.125 = 0.875 \end{aligned}$$

In general, for $x \in R_X$,

$$\begin{aligned} P(X \leqslant x) &= \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^x} \\ &= \frac{1}{2}\left(1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^{x-1}}\right) \\ &= \tfrac{1}{2}\frac{1-(\frac{1}{2})^x}{1-\frac{1}{2}} = 1 - \left(\frac{1}{2}\right)^x . \end{aligned}$$

Here, the distribution function has been derived from the probability mass function. Knowing one of these functions always allows us to derive the other.

## The probability density function

Suppose that $X$ is a continuous random variable. Then $P(X = x) = 0$ for all $x$.

### Proof

For any $x \in \mathbb{R}$, define the decreasing sequence of events:

$$E_n = \left[x - \frac{1}{n}, x + \frac{1}{n}\right] \qquad n = 1, 2, 3, \ldots,$$

so that $\lim_{n\to\infty} E_n = \{x\}$. Then:

$$\begin{aligned}
\lim_{n\to\infty} P(E_n) &= \lim_{n\to\infty} P\left\{x - \frac{1}{n} < X \leqslant x + \frac{1}{n}\right\} \\
&= \lim_{n\to\infty} \left\{F_X\left(x + \frac{1}{n}\right) - F_X\left(x - \frac{1}{n}\right)\right\} \\
&= \lim_{n\to\infty} \left\{F_X\left(x + \frac{1}{n}\right)\right\} - \lim_{n\to\infty} \left\{F_X\left(x - \frac{1}{n}\right)\right\} \\
&= F_X(x) - F_X(x) = 0.
\end{aligned}$$

This uses the fact that $F$ is continuous, i.e., left continuous as well as right continuous.

A general result states that, for decreasing sequences of events $P\left(\lim_{n\to\infty} E_n\right) = \lim_{n\to\infty} P(E_n)$. So, $P(X = x) = 0$.

For a continuous random variable, then, the concept of a probability mass function is meaningless. A function that does usefully summarise the probability information about $X$ is the probability density function, defined below.

### Definition

Let $X$ be a continuous random variable. The **probability density function** (p.d.f.) is defined as

$$f_X(x) = \frac{d}{dx} F_X(x) \qquad x \in \mathbb{R}.$$

Since integration is anti-differentiation, it follows from this definition that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt \qquad x \in \mathbb{R}.$$

Notice that $f_X(x) = 0$ for all $x \notin R_X$.

The probability that a continuous random variable $X$ lies in the interval $(a, b)$ can be found by integrating its probability density function over the appropriate range (Figure 2.3):

$$\begin{aligned}
P(a < X \leqslant b) &= P(X \leqslant b) - P(X \leqslant a) \\
&= F_X(b) - F_X(a) \\
&= \int_{-\infty}^{b} f_X(t)\, dt - \int_{-\infty}^{a} f_X(t)\, dt \\
&= \int_{a}^{b} f_X(t)\, dt.
\end{aligned}$$

Since $X$ is continuous, $P(X = a) = P(X = b) = 0$ and therefore

$$P(a < X < b) = P(a < X \leqslant b) = P(a \leqslant X < b) = P(a \leqslant X \leqslant b) = F_X(b) - F_X(a).$$

Since $P(X = x) = 0$ for all $x$, the p.d.f. of a continuous random variable cannot represent probabilities in the same way as the p.m.f. of a discrete random variable. There is a connection, though, which can be seen from the following argument.

Let $x$ be any real number and let $\delta$ be a small positive value. Then,

$$P\left(x - \frac{\delta}{2} < X \leqslant x + \frac{\delta}{2}\right) = \int_{x-\frac{\delta}{2}}^{x+\frac{\delta}{2}} f_X(t)\, dt \approx \delta \times f_X(x).$$

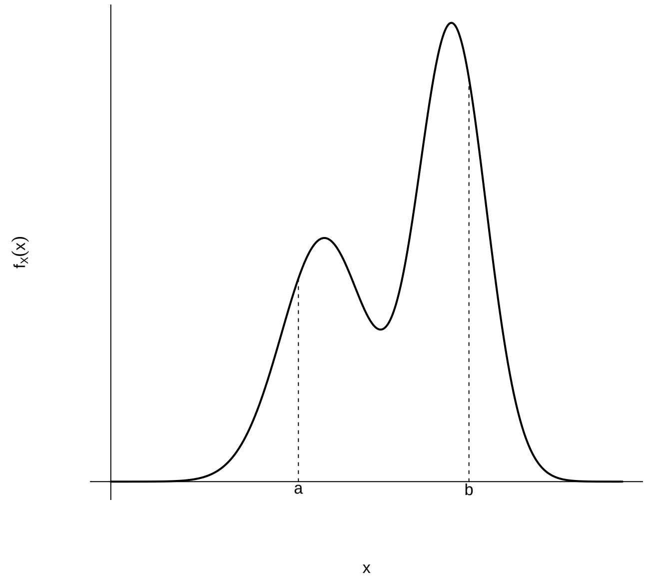


Figure 2.3: Integrating a p.d.f.

So $f_X(x)$ is proportional to the probability that $X$ lies in a small interval (of fixed width) centred on $x$. It is for this reason that the probability density function is preferred to the distribution function as a way of summarising the probabilities associated with $X$.

The following two properties are shared by all valid probability density functions.

(a) $0 \leqslant f_X(x)$, for all $x \in \mathbb{R}$. This follows because $f_X$ is the derivative of the non-decreasing function $F_X$.

(b) $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$. This follows because $F_X(-\infty) = 0$ and $F_X(\infty) = 1$ by definition.

**Example 6 (continued)**

The continuous random variable $X$ has distribution function

$$F_X(x) = \begin{cases} 0, & x \leqslant 0 \\ x, & 0 < x \leqslant 1 \\ 1, & 1 < x \end{cases}.$$

Differentiating once with respect to $x$, we obtain the following p.d.f.:

$$f_X(x) = \begin{cases} 0, & x \leqslant 0 \\ 1, & 0 < x \leqslant 1 \\ 0, & 1 < x \end{cases}.$$

This special distribution is known as the **Uniform distribution** on the range 0 to 1. We could write $X \sim U(0, 1)$.

**Definition**

The continuous random variable $X$ is said to have the Uniform distribution on the interval $[a, b]$, written $X \sim U(a, b)$, if $X$ has the following distribution function and probability density function:

$$F_X(x) = \begin{cases} 0, & x \leqslant a \\ \frac{x-a}{b-a}, & a < x \leqslant b \\ 1, & b < x \end{cases}, \qquad f_X(x) = \begin{cases} 0, & x \leqslant a \\ \frac{1}{b-a}, & a < x \leqslant b \\ 0, & b < x \end{cases}.$$

## 2.3 Moments

### Definition

The expected value of a random variable, X, if it exists, is defined as:

$$E(X) = \sum_{x \in R_X} x p_X(x)$$

when X is a discrete random variable and

$$E(X) = \int_{x \in R_X} x f_X(x)\, dx$$

when X is a continuous random variable. The expected value is only well-defined when (respectively)

$$\sum_{x \in R_X} |x| p_X(x) \qquad \text{or} \qquad \int_{x \in R_X} |x| f_X(x)\, dx$$

converges to a finite limit ("absolute convergence").

The interpretation of E(X) is as follows. If the underlying random experiment is conducted on a large number of occasions, and the value of X is recorded each time, then in the limit, as the number of trials tends to infinity, the mean value of X is E(X).

In general, if X has a symmetric distribution then E(X) occurs at the centre of symmetry, assuming that E(X) exists at all.

### Example 5 (continued)

In this example, X is a discrete random variable, $R_X = \{0, 1, 2, 3\}$ and $p_X(0) = 0.125$, $p_X(1) = 0.125$, $p_X(2) = 0.25$, $p_X(3) = 0.5$. So

$$E(X) =$$

This example shows that, when X is a discrete random variable, E(X) does not need to be a possible realisation of X.

### Example 8

Suppose that the continuous random variable $X \sim U(a, b)$. Then,

$$E(X) = \int_a^b \frac{1}{b-a} x\, dx = \left[\frac{1}{2(b-a)} x^2\right]_a^b = \frac{1}{2}\frac{b^2-a^2}{b-a} = \frac{a+b}{2}.$$

Notice that this is a symmetric distribution on the interval $[a, b]$ so E(X) is, as anticipated, at the centre of the interval. The median of the distribution, $\xi_{50}$, is found as follows:

$$F(\xi_{50}) = 0.5 \implies \frac{\xi_{50} - a}{b-a} = 0.5 \implies \xi_{50} = \frac{a+b}{2}.$$

Since the distribution is symmetric, the median is equal to the expected value; it, too, is found at the centre of the interval $[a, b]$.

If $g(X)$ is a function of the random variable X, then it is itself a random variable. The expected value of $g(X)$ is defined as follows:

$$E[g(X)] = \sum_{x \in R_X} g(x) p_X(x), \qquad \text{when } g(X) \text{ is a discrete random variable, or}$$

$$E[g(X)] = \int_{x \in R_X} g(x) f_X(x)\, dx, \qquad \text{when } g(X) \text{ is a continuous random variable.}$$

This assumes that the required sum or integral is absolutely convergent.

**Example 9**

Let X be a discrete random variable with finite expected value $E(X)$. Let $a$ and $b$ be real constants. Then $aX + b$ has finite expected value

$$E[aX + b] = \sum_{x \in R_X} (ax + b) p_X(x) = a \sum_{x \in R_X} x p_X(x) + b \sum_{x \in R_X} p_X(x) = aE(X) + b.$$

Although we have only proved this result for the case when X is a discrete random variable, it is true also for the continuous case (and the proof is easily adapted).

**Definitions**

In general, if it exists, the kth moment of the random variable X around a value c is $E[(X - c)^k]$.

The kth moment (or **raw moment** or **crude moment**) is usually taken to mean $\mu_k \equiv E(X^k)$. So the first moment is the expected value of X.

The kth **central moment** is $E[(X - \mu)^k]$, where $\mu = E(X)$.

The second central moment $E[(X - \mu)^2]$ is also called the **variance** of X, denoted $\text{Var}(X)$. It is a measure of the spread of the distribution of X. A distribution with a higher variance has a distribution that is more spread out than one with a lower variance.

The variance is the average squared distance of X from its expected value. Since it is an expected value (or long-term average) of a non-negative quantity, then the variance itself must be non-negative, i.e., $\text{Var}(X) \geqslant 0$.

**Example 9 (continued)**

Let X be a discrete random variable with finite variance $\text{Var}(X)$. Let $a$ and $b$ be real constants. Then $aX + b$ has finite variance

$$\begin{aligned}
\text{Var}(aX + b) &= E\left([(aX + b) - E(aX + b)]^2\right) \\
&= E\left([(aX + b) - (aE(X) + b)]^2\right) \\
&= E\left(a^2[X - E(X)]^2\right) \\
&= a^2 \text{Var}(X).
\end{aligned}$$

Although we have only proved this result for the case when X is a discrete random variable, it is true also for the continuous case (and the proof is easily adapted).

In general, $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$. This is proved as follows for the continuous case:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\big[(X - \mathrm{E}[X])^2\big) \\
&= \int_{x \in R_X} [x - \mathrm{E}(X)]^2 f_X(x)\, dx \\
&= \int_{x \in R_X} x^2 f_X(x)\, dx - 2\mathrm{E}(X) \int_{x \in R_X} x f_X(x)\, dx + [\mathrm{E}(X)]^2 \int_{x \in R_X} f_X(x)\, dx \\
&= \mathrm{E}[X^2] - 2\mathrm{E}[X]\mathrm{E}[X] + [\mathrm{E}(X)]^2 \\
&= \mathrm{E}[X^2] - [\mathrm{E}(X)]^2
\end{aligned}$$

**Example 8 (continued)**

Here, $X \sim U(a, b)$. We have already shown that $\mathrm{E}(X) = \frac{a+b}{2}$.

$$\begin{aligned}
\mathrm{E}(X^2) &= \int_a^b \frac{1}{b-a} x^2\, dx = \left[\frac{1}{3(b-a)} x^3\right]_a^b = \frac{1}{3}\frac{b^3 - a^3}{b-a} = \frac{1}{3}(a^2 + ab + b^2) \\
\mathrm{Var}(X) &= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 \\
&= \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 \\
&= \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12} \\
&= \frac{(b-a)^2}{12}.
\end{aligned}$$

The wider the interval $[a, b]$ on which $X$ is defined, the higher is the variance of $X$ (indicating that the distribution of $X$ is more spread out).

Notice that $\mathrm{Var}(X)$ is measured in the square of the units in which $X$ itself is measured. For this reason, an alternative measure of spread is sometimes preferred that is measured in the same units as $X$. This measure, known as the **standard deviation** of $X$, is defined by:

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}.$$

## 2.4 The Binomial and related distributions

Here are brief descriptions of random variables associated with four experiments. Though they arise in very different contexts, they share a number of crucial features.

(a) A fair coin is tossed 1,000 times; $X$ is the number of heads recorded.

(b) A fair die is rolled 100 times; $X$ is the number of times a score of 6 is recorded.

(c) 20 electronic components produced consecutively on a production line are tested; $X$ is the number of them that are found to be satisfactory.

(d) $X$ is the number of girls in a family of four children (none of whom are twins, triplets, etc.).

In each case, a basic experiment is replicated a number of times (tossing a coin, rolling a die, testing a component, recording the sex of a child). The sequences of trials described above share four common features.

(1) Each trial has two possible outcomes, which are conventionally described as 'success' and 'failure'. In the above examples, a 'success' might be obtaining (a) heads, (b) a score of 6, (c) a satisfactory result to the test, (d) a girl.

(2) The outcomes of the trials are independent.

(3) In each trial, the probability of a success has the same value, $\theta$ (where $0 \leqslant \theta \leqslant 1$ and $\theta$ is often unknown in advance of the experiment).

(4) The number of trials is fixed in advance ($n = 1000$, 100, 20, 4).

Trials with the features (1) to (3) are called **Bernoulli trials**. When $X$ is the number of successes in $n$ Bernoulli trials, each with success probability $\theta$, then $X$ is said to be a Binomial random variable. This is written $X \sim \mathrm{Bi}(n, \theta)$. So, in the examples above, the random variables have the following distributions:

(a) $\mathrm{Bi}(1000, 0.5)$

(b) $\mathrm{Bi}(100, \frac{1}{6})$

(c) $\mathrm{Bi}(20, \theta)$, where $\theta$ is the probability that a randomly-selected component is satisfactory

(d) $\mathrm{Bi}(4, \theta)$ where $\theta$ is the probability that a randomly-selected child in a family of 4 children is a girl [does $\theta = 0.5$?]

Suppose that $X \sim \mathrm{Bi}(n, \theta)$. Then $R_X = \{0, 1, \ldots, n\}$, since there are $n$ trials, each of which may result in a success or a failure. For any given $x \in R_X$, the following is a typical sequence of results with successes on $x$ given trials:

$$\underbrace{SS\cdots S}_{x}\underbrace{FF\cdots F}_{n-x}$$

Since each success occurs with probability $\theta$, each failure occurs with probability $1-\theta$, and the trials are independent, it follows that the probability of this given sequence of results is:

$$\theta\theta\cdots\theta(1-\theta)(1-\theta)\cdots(1-\theta) = \theta^x(1-\theta)^{n-x}.$$

The number of different sequences of trials with exactly $x$ successes is:

$$\binom{n}{x} = {}^{n}C_{x} = \frac{n!}{x!(n-x)!}.$$

Each of these sequences has the same probability. So, $X$ has probability mass function:

$$p_X(x) = P(X = x) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \qquad (x = 0, 1, \ldots, n).$$

This is a valid probability mass function, since:

(a) $0 \leqslant p_X(x) \leqslant 1$ for all $x$

(b)

$$\begin{aligned}\sum_{x=0}^{n} p_X(x) &= \sum_{x=0}^{n}\binom{n}{x}\theta^x(1-\theta)^{n-x}\\ &= [\theta + (1-\theta)]^n \qquad \text{[Binomial theorem]}\\ &= 1.\end{aligned}$$

### Example 10

In the course of a certain printing process, 20 different colours are inked on to the same sheet of perspex. Unless all 20 colours are inked satisfactorily, the printed item has to be scrapped. The colours are inked on consecutively and independently, 2.5% of items being inked unsatisfactorily with each colour. Overall, what is the percentage of items printed by this process that has to be scrapped?

### Solution

### Example 11

Suppose that $X \sim \text{Bi}(10, 0.1)$. Write out the probability mass function of $X$.

### Solution

In this case it is easy to obtain values of the probability mass function directly from the definition. We will, however, use the following recursive formula, which is particularly useful when carrying out hand calculations or programming a calculator or computer when $n$ is large or $\theta$ is small.

Suppose that $X \sim \text{Bi}(n, \theta)$. Then, for $x = 0, 1, \ldots, n-1$,

$$\begin{aligned} p_X(x+1) &= \binom{n}{x+1}\theta^{x+1}(1-\theta)^{n-(x+1)} \\ &= \frac{n!}{(x+1)!(n-x-1)!}\theta^{x+1}(1-\theta)^{n-x-1} \\ &= \frac{n!}{x!(n-x)!}\theta^{x}(1-\theta)^{n-x}\frac{n-x}{x+1}\frac{\theta}{1-\theta} \\ &= p_X(x)\frac{n-x}{x+1}\frac{\theta}{1-\theta}. \end{aligned}$$

For Example 11, we can work out the probability mass function of X as follows:

$$\begin{aligned} p_x(0) &= \\ p_x(1) &= \\ p_x(2) &= \\ p_x(3) &= \\ p_x(4) &= \\ p_x(5) &= \\ &\text{etc.} \end{aligned}$$

Suppose that $X \sim \text{Bi}(n, \theta)$. Then X has expected value:

$$\begin{aligned} E(X) &= \sum_{x=0}^{n} x p_X(x) \\ &= \sum_{x=1}^{n} x\frac{n!}{x!(n-x)!}\theta^{x}(1-\theta)^{n-x} \\ &= n\theta\sum_{x=1}^{n}\frac{(n-1)!}{(x-1)!(n-x)!}\theta^{x-1}(1-\theta)^{n-x} \\ &= n\theta\sum_{x=1}^{n}\binom{n-1}{x-1}\theta^{x-1}(1-\theta)^{n-x} \\ &= n\theta\sum_{y=0}^{n-1}\binom{n-1}{y}\theta^{y}(1-\theta)^{n-x} \qquad \text{[where } y = x-1\text{]} \\ &= n\theta[\theta + (1-\theta)]^{n-1} \qquad \text{[Binomial theorem]} \\ &= n\theta. \end{aligned}$$

It can also be shown that $\text{Var}(X) = n\theta(1-\theta)$.

Here are variations of the experiments (a) – (c) listed at the start of this section.

(a′) A fair coin is tossed repeatedly, until it lands tails up for the first time.

(b′) A fair die is rolled repeatedly, until the first score of 1, 2, 3, 4 or 5 is recorded.

(c′) Components off a production line are tested, in turn, until the first defective item is found.

The common features of these experiments can be expressed in the following way.

(1) There is a *potentially infinite* series of Bernoulli trials, each with success probability $\theta$ ($0 \leqslant \theta \leqslant 1$).

(2) The trials continue until the first failure occurs. Here, a failure might be: (a′) tails, (b′) a score $< 6$, (c′) a defective component.

Let the random variable $X$ be the total number of trials required until the first failure occurs. Then $X$ is a **Geometric** random variable with parameter $\theta$, written $X \sim \text{Geo}(\theta)$.

In the examples above: (a′) $X \sim \text{Geo}(0.5)$; (b′) $X \sim \text{Geo}(\frac{1}{6})$; (c′) $X \sim \text{Geo}(\theta)$, where $\theta$ is the probability that a randomly-selected component is satisfactory.

It is sometimes natural to think of a $\text{Geo}(\theta)$ random variable as the number of Bernoulli trials till the first success, when $P(\text{success}) = 1 - \theta$ [or $P(\text{failure}) = \theta$].

Suppose that $X \sim \text{Geo}(\theta)$. Then the range space of $X$ is $R_X = \{1, 2, \ldots\}$. For any $x \in R_X$,

$$\begin{aligned} P(X = x) &= P\,(\text{success on 1st trial}) \\ &\quad \times \cdots \times P\,(\text{success on } (x-1)\text{th trial}) \\ &\quad \times P\,(\text{failure on } x\text{th trial}) \\ &= \theta^{x-1}(1-\theta). \end{aligned}$$

So $X$ has the following probability mass function:

$$p_X(x) = P(X = x) = \theta^{x-1}(1-\theta) \qquad (x = 1, 2, \ldots).$$

The terms in this probability mass function form a *Geometric Series*; the first term is $1 - \theta$ and the constant ratio between consecutive terms is $\theta$, where $0 \leqslant \theta \leqslant 1$.

This is a valid probability mass function since:

(a) $0 \leqslant p_X(x) \leqslant 1$, for all $x$.

(b)

$$\begin{aligned} \sum_{x=1}^{\infty} p_X(x) &= \sum_{x=1}^{\infty} \theta^{x-1}(1-\theta) \\ &= (1-\theta)(1 + \theta + \theta^2 + \cdots) \\ &= (1-\theta) \times \frac{1}{1-\theta} \\ &= 1. \end{aligned}$$

It can easily be shown that

$$E(X) = \frac{1}{1-\theta} \qquad \text{and} \qquad \text{Var}(X) = \frac{\theta}{(1-\theta)^2}.$$

### Example 12

In a certain quiz show, contestants are asked a series of multiple choice questions, where each question has 4 different possible answers of which only one is correct. The first contestant to play is chosen at random from among all the contestants. This contestant continues to play (i.e., answer questions) until he or she first get an answer wrong, at which point a new contestant is chosen to play. Assume that the first contestant guesses the answer to every question asked. On average, how many questions will this contestant attempt before the next contestant gets to play?

**Solution**

## 2.5 Functions of a random variable

### Example 13

A certain equity (share) is currently priced at £$S$. We are interested in forecasting its price, £$Y$, at a time $T$ years in the future. One financial model divides the period of $T$ years into $n$ non-overlapping sub-periods, each of length $T/n$ years. It is assumed that, in each of these periods, the price of the equity will either rise by £$a$ (with probability $\theta$) or fall by £$a$ (with probability $1-\theta$), where $a \leqslant S/n$. It is further assumed that what happens to the price of the equity in one sub-period is independent of what happens to it in any other sub-period.

Let the random variable $X$ be the number of time periods in which the share price increases. Then $X \sim \mathrm{Bi}(n,\theta)$, so that

$$P(X=x) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \qquad (x = 0, 1, \ldots, n),$$

and

$$E[X] = n\theta \qquad \text{and} \qquad \mathrm{Var}(X) = n\theta(1-\theta).$$

Now $Y$ is also a discrete random variable and we can derive its probability mass function from our knowledge of the probability mass function of $X$. Note first of all that $X = x$ if and only if the share price increases in exactly $x$ periods and, therefore, falls in the other $n-x$ periods. This means that, when $X = x$, $Y$ takes the value:

$$y = S + ax - a(n-x) = 2ax + S - an.$$

In terms of the random variables, we can therefore write:

$$Y = 2aX + S - an.$$

$Y$ has range space $R_Y = \{S - an, S - a(n-2), S - a(n-4), \ldots, S + an\}$. For $y \in R_Y$,

$$P(Y=y) = P(2aX + S - an = y) = P\left(X = \frac{y + an - S}{2a}\right) = p_X\left(\frac{y + an - S}{2a}\right).$$

In order to determine $E(Y)$ or $\mathrm{Var}(Y)$ directly, we would need to sum series related to this probability distribution function, which could be quite challenging. However, we can exploit the linear relationship between $X$ and $Y$ to write immediately:

$$\begin{aligned} E(Y) &= E(2aX + S - an) = 2aE(X) + S - an = 2an\theta + S - an = S + an(2\theta - 1) \\ \mathrm{Var}(Y) &= \mathrm{Var}(2aX + S - an) = 4a^2\mathrm{Var}(X) = 4a^2 n\theta(1-\theta). \end{aligned}$$

In general, when $X$ is a discrete random variable and a new random variable $Y$ is defined by $Y = h(X)$, then $Y$ must also be a discrete random variable. The probability mass function of Y is found using $P(Y = y) = P(\{x : h(x) = y\})$. $E(Y)$ and $Var(Y)$ may, as usual, be found directly from this probability distribution or indirectly as functions of $E(X)$ and $Var(X)$.

When X is a continuous random variable and $Y = h(X)$ then Y might be either discrete or continuous depending on the function h. When Y is defined by a continuous function of X, then Y is a continuous random variable whose probability density function can be derived using the distribution function of X.

### Example 14

Suppose that the continuous random variable X has the **Exponential** distribution with parameter $\theta > 0$ written $X \sim \text{Expo}(\theta)$. This means that X has the probability density function

$$f_X(x) = \theta e^{-\theta x} \qquad x > 0.$$

Let $Y = kX$, where $k > 0$ is a real constant. We can find the probability density function of Y, and therefore identify its distribution, as follows.

Figure 2.4 shows a diagram of this general procedure for finding the probability density function of the continuous random variable Y, when Y can be expressed as a function of another continuous random variable X whose distribution function we already know. In certain circumstances, it is possible to short-circuit this procedure

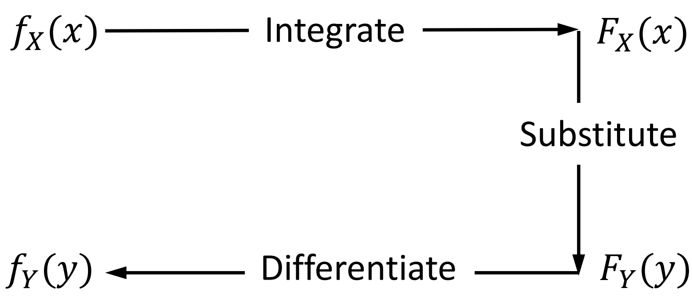


Figure 2.4: Finding the p.d.f. of a function of a continuous random variable.

and go from $f_X(x)$ to $f_Y(y)$ without having to work out the distribution functions explicitly. The two most important situations like this are described in Propositions 2.1 and 2.2 below.

### Proposition 2.1

Suppose that $X$ is a continuous random variable with range space $R_X = (a, b)$ and probability density function $f_X(x)$, $x \in R_X$. Define the random variable $Y$ by $Y = h(X)$, where $h$ is either a strictly increasing or a strictly decreasing, differentiable function on $R_X$. Then the range space of $Y$ is $R_Y = \big(\min[h(a), h(b)], \max[h(a), h(b)]\big)$ and the p.d.f. of $Y$ is

$$f_Y(y) = f_X\big(h^{-1}(y)\big)\left|\frac{dx}{dy}\right| \qquad y \in R_Y.$$

### Proof

We prove the case where $h(x)$ is a strictly increasing, differentiable function. In this case, $R_Y = \{y : h(a) < y < h(b)\}$. For any $y$ in this range,

$$F_Y(y) = P(Y \leqslant y) = P\big(h(X) \leqslant y\big) = P\big(X \leqslant h^{-1}(y)\big) = F_X\big(h^{-1}(y)\big).$$

So,

$$f_Y(y) = \frac{d}{dy}F_X\big(h^{-1}(y)\big) = \frac{d}{dx}F_X\big(h^{-1}(y)\big)\frac{dx}{dy} = f_X\big(h^{-1}(y)\big)\frac{dx}{dy}.$$

The proof for the case where $h(X)$ is a strictly decreasing function is left as an exercise.

### Example 14 (continued)

In this example, $h(x) = kx$ is a strictly increasing function since $k > 0$. Clearly, $h^{-1}(y) = y/k$.

For any $y \in R_Y = [0, \infty)$,

$$f_Y(y) = f_X\big(h^{-1}(y)\big)\frac{dx}{dy} = \theta e^{-\theta y/k}\frac{1}{k} = \frac{\theta}{k}e^{-y(\theta/k)}.$$

### Proposition 2.2

Suppose that the continuous random variable $X$ has the range space $R_X = (-\infty, \infty)$ and the probability density function $f_X(x)$ on $R_X$. Define the random variable $Y = X^2$. Then $Y$ has probability density function

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})] \qquad y > 0.$$

### Proof

Since $Y$ is not a strictly increasing function of $X$ on the whole of $R_X$, Proposition 2.1 does not apply. We proceed using the general method. $R_Y = [0, \infty)$. For any $y \in R_Y$,

$$F_Y(y) = P(Y \leqslant y) = P(X^2 \leqslant y) = P(-\sqrt{y} \leqslant X \leqslant \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Differentiating with respect to $y$ gives the probability density function of $Y$:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{2\sqrt{y}}f_X(\sqrt{y}) - \frac{-1}{2\sqrt{y}}f_X(-\sqrt{y}) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})].$$

### Example 15

Suppose that X has the **Standard Normal** distribution:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Let $Y = X^2$. The p.d.f. of X is symmetric about $x = 0$. So, applying Proposition 2.2,

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})] = \frac{1}{2\sqrt{y}}[2f_X(\sqrt{y})] = \frac{1}{\sqrt{2\pi}\sqrt{y}} \exp\left(-\tfrac{1}{2}y\right) \qquad y > 0.$$

This is an important distribution in its own right, called the **chi-squared distribution** with one degree of freedom, written $Y \sim \chi^2_1$.

To end this discussion of functions of a random variable, we will discuss the relationship between one particular continuous distribution, the Exponential, and one discrete distribution, the Poisson. The discrete random variable, X, is said to have the **Poisson** distribution with expected value $\lambda > 0$, written $X \sim \text{Poi}(\lambda)$, if it has probability mass function

$$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad (x = 0, 1, 2, \dots).$$

### Example 16

Suppose that jobs arrive at a computer server according to a **Poisson Process**. This means that the number of jobs that arrive at the server in any time period of length $t$ seconds, $X_t$ say, is a Poisson random variable with expected value $t\lambda$, for some $\lambda > 0$. The numbers of jobs that arrive in non-overlapping time intervals are independent random variables.

An inter-arrival time is the length of time that elapses between consecutive jobs arriving at the server. Suppose that one job arrives and let the continuous random variable Y be the length of time (minutes) that elapses until the next job arrives. Clearly, Y has range space $R_Y = [0, \infty)$. The distribution function of Y is:

When events occur at an average rate of $\lambda$ events per unit time in a Poisson Process, the number of events in a time period of $t$ minutes is a $\text{Poi}(t\lambda)$ random variable, while the time that elapses between consecutive arrivals is an $\text{Expo}(\lambda)$ random variable.

When X is a continuous random variable, it can often be difficult or even impossible to determine the probability density function of $Y = h(X)$. The following results are often used to find approximations to the expected value and variance of Y in these circumstances. Warning: these results can give very misleading answers.

### Proposition 2.3

Let $X$ be a continuous random variable with finite mean $\mu_X$ and finite standard deviation $\sigma_X$. Let $Y = h(X)$, where $h$ is a continuous function whose first and second derivatives exist at all points $x \in R_X$. Then, approximately,

$$E(Y) \approx h(\mu_X) + \frac{\sigma_x^2}{2} h''(\mu_X).$$

### Proof

Expand $h$ as a Taylor series around $\mu_X$:

$$Y = h(X) = h(\mu_X) + (X - \mu_X)h'(\mu_X) + \tfrac{1}{2}(X - \mu_X)^2 h''(\mu_X) + \cdots .$$

Taking expected values of both sides:

$$\begin{aligned} E(Y) &= h(\mu_X) + E(X - \mu_X)h'(\mu_X) + \tfrac{1}{2}E(X - \mu_X)^2 h''(\mu_X) + \cdots \\ &= h(\mu_X) + 0 \times h'(\mu_X) + \tfrac{1}{2}\sigma_X^2 h''(\mu_X) + \cdots \end{aligned}$$

So $E(Y) \approx h(\mu_X) + \frac{1}{2}\sigma_x^2 h''(\mu_X)$.

### Proposition 2.4

Let $X$ be a continuous random variable with finite mean $\mu_X$ and finite standard deviation $\sigma_X$. Let $Y = h(X)$, where $h$ is a continuous function whose first and second derivatives exist at all points $x \in R_X$. Then, approximately,

$$\mathrm{Var}(Y) \approx [h'(\mu_X)]^2 \sigma_X^2.$$

### Proof

Again expanding $h$ as a Taylor series around $\mu_X$, but this time taking just the first two terms:

$$Y = h(X) \approx h(\mu_X) + (X - \mu_X)h'(\mu_X).$$

Since $\mu_X$, $h(\mu_X)$ and $h'(\mu_X)$ are all constants, taking the variance of both sides gives

$$\mathrm{Var}(Y) \approx [h'(\mu_X)]^2 \mathrm{Var}(X) = [h'(\mu_X)]^2 \sigma_X^2.$$

**Example 15 (continued)**

## 2.6 The Normal distribution

### Definition

Suppose that the continuous random variable, $X$, can take any real value and has the following p.d.f. for real numbers $\mu$ and $\sigma$ (where $\sigma > 0$):

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right].$$

Then $X$ is said to have a **Normal** distribution, with parameters $\mu$ and $\sigma^2$, written $X \sim N(\mu, \sigma^2)$.

Many different kinds of continuous random variables have been found to follow Normal distributions, for example:

- some anthropometric measurements, such as the height of a fully-grown man or woman from a particular ethnic group;
- scientific measurements that are subject to experimental error, for example the measured distance from the Earth to a given star;
- dimensions of manufactured objects, for example the weight of a (nominal) $\frac{1}{4}$ lb hamburger.

This is a valid p.d.f. because:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] > 0 \qquad \text{for any real } x$$

and

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \\
&= 2\int_{\mu}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \\
&= 2\int_{0}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}z^2\right] \sigma\, dz \qquad \text{where } z = (x-\mu)/\sigma \\
&= 2\int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-u\right) \frac{1}{\sqrt{2}\sqrt{u}}\, du \qquad \text{where } u = \tfrac{1}{2}z^2 \\
&= \frac{1}{\sqrt{\pi}} \int_0^{\infty} \frac{1}{\sqrt{u}} \exp(-u)\, du \\
&= \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right) \\
&= 1.
\end{aligned}$$

In a similar way, it is possible to find the central moments of $X$. For $k = 1, 2, 3, \ldots$

$$\begin{aligned}
E[(X-\mu)^k] &= \int_{-\infty}^{\infty} (x-\mu)^k \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \\
&= \int_{-\infty}^{\mu} (x-\mu)^k \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \\
&\qquad + \int_{\mu}^{\infty} (x-\mu)^k \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx.
\end{aligned}$$

When $k$ is an odd number, this integral is 0. In particular, this means that

$$E(X-\mu) = 0 \implies E(X) = \mu.$$

When $k$ is an even number,

$$\begin{aligned}
E[(X-\mu)^k] &= 2\int_{\mu}^{\infty} (x-\mu)^k \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \\
&= 2\sigma^{k+1} \int_0^{\infty} z^k \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}z^2\right] dz \qquad \text{where } z = (x-\mu)/\sigma \\
&= 2\sigma^k \int_0^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-u) \frac{2^{k/2}u^{k/2}}{\sqrt{2}\sqrt{u}}\, du \qquad \text{where } u = \tfrac{1}{2}z^2 \\
&= \frac{2^{k/2}\sigma^k}{\sqrt{\pi}} \int_0^{\infty} u^{(k-1)/2} \exp(-u)\, du \\
&= \frac{2^{k/2}\sigma^k}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right).
\end{aligned}$$

At $k = 2$,

$$\mathrm{Var}(X) = E[(X-\mu)^2] = \frac{2\sigma^2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right) = \frac{2\sigma^2}{\sqrt{\pi}}\frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \sigma^2.$$

The two parameters of the Normal distribution allow for a very wide range of probability distribution functions. The expected value, $\mu$, can be any real number. The standard deviation, $\sigma$, can be any positive real number. The larger $\sigma$ is, the flatter/broader the probability density function is.

As before, we want to be able to calculate probabilities of the form $P(a < X < b)$, which means evaluating integrals of the form:

$$\int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx.$$

This cannot be done algebraically, or in closed form, but only using numerical methods of integration.

Tables of the distribution function of one very important Normal distribution are widely available. This is the $N(0,1)$ distribution, i.e., the Normal distribution with expected value 0 and variance 1, which is known as the **Standard Normal** distribution. Table 2.1 lists these values. This function is so important in practice that we reserve a special notation for it, the capital Greek letter phi ($\Phi$). If $Z \sim N(0,1)$, then:

$$P(Z \leqslant z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx.$$

### Example 17

Suppose that $Z$ has the Standard Normal distribution, $Z \sim N(0,1)$. Then, using Table 2.1, we can easily find probabilities like the following.

$$\begin{aligned} P(Z \leqslant 0) &= \\ P(Z > 1.96) &= \\ P(2 < Z < 3) &= \end{aligned}$$

Table 2.1 does not list $\Phi(z)$ for $z < 0$. When $Z \sim N(0,1)$, then $f_Z(z)$ is symmetric: $f_Z(-z) = f_Z(z)$. Hence, $\Phi(-z) = P(Z \leqslant -z) = P(Z > z) = 1 - \Phi(z)$.

### Example 17 (continued)

This result allows us to use the table to find probabilities like the following:

$$\begin{aligned} P(Z < -1) &= \\ P(-2 < Z < -1) &= \\ P(-1.645 < Z < 1.645) &= \end{aligned}$$

The following result allows us to work out similar probabilities for other Normal distributions too.

### Proposition 2.5

If $X \sim N(\mu, \sigma^2)$, for any real $\mu$ and positive $\sigma$, then

$$Z \equiv \frac{X-\mu}{\sigma} \sim N(0,1).$$

Table 2.1: Standard normal distribution function. When $Z$ is a $N(0, 1)$ random variable, this table gives $\Phi(z) = P(Z \leqslant z)$ for values of $z$ from 0.00 to 3.67 in steps of 0.01. When $z < 0$, $\Phi(z)$ can be found from this table using the relationship $\Phi(-z) = 1 - \Phi(z)$.

| $z$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9979 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9983 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 | 0.9990 |
| 3.1 | 0.9990 | 0.9991 | 0.9991 | 0.9991 | 0.9992 | 0.9992 | 0.9992 | 0.9992 | 0.9993 | 0.9993 |
| 3.2 | 0.9993 | 0.9993 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9995 | 0.9995 | 0.9995 |
| 3.3 | 0.9995 | 0.9995 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9997 |
| 3.4 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9998 |
| 3.5 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 |
| 3.6 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9999 | | |

**Proof**

$z = h(x) = \frac{x-\mu}{\sigma}$, where $h$ is a strictly increasing function. So we can use Proposition 2.1 to find the p.d.f. of $Z$:

$$\begin{aligned}
x &= h^{-1}(z) = \mu + \sigma z \\
\frac{dx}{dz} &= \sigma \\
f_Z(z) = f_X\left(h^{-1}(z)\right)\left|\frac{dx}{dz}\right| &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left[\frac{\mu+\sigma z-\mu}{\sigma}\right]^2\right)\sigma = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right).
\end{aligned}$$

**Example 18**

Suppose that $X \sim N(10, 9)$. Then

$$Z \equiv \frac{X-10}{3} \sim N(0,1).$$

We can use Table 2.1 to obtain probabilities like the following:

$$\begin{aligned}
P(X \geqslant 13) &= \\
P(X < 0) &= \\
P(0 \leqslant X \leqslant 10) &=
\end{aligned}$$

**Example 19**

(a) The nominal weight of a box of a particular breakfast cereal is 1 kg. The actual weight (g) is a $N(1100, 100^2)$ random variable. What proportion of all boxes of this cereal weigh at least 1 kg?

(b) Suppose now that the manufacturing process can be changed so that the weight of a box of cereal is a $N(\mu, 50^2)$ random variable. To what value may $\mu$ be reduced while ensuring that 95% of all boxes of this cereal weigh at least 1 kg?

### Solution

### Definition

A **log-normal** distribution is the probability distribution of a random variable whose natural logarithm is Normally distributed. If $X$ is a random variable with a Normal distribution, then $Y = \exp(X)$ has a log-normal distribution. Alternatively, if $Y$ is log-normally distributed, then $X = \log_e(Y)$ is Normally distributed.

A random variable might have a log-normal distribution if it is the product of many independent random variables, each of which must take positive values.

### Example 20

Suppose that $S_0$ is the current price of a stock (equity), and let $S_t$ denote the price of the same stock at time $t$ ($> 0$) years in the future. It is often assumed that

$$\log_e\left(\frac{S_t}{S_0}\right) \sim N\left(\left[\mu - \frac{\sigma^2}{2}\right] t, \sigma^2 t\right).$$

In this standard model, which is known as **Geometric Brownian Motion**, $\mu$ is called the expected rate of return and $\sigma$ is the volatility rate. Suppose that a certain stock has a current price of \$64, expected return of 5% p.a. and volatility of 12% p.a.. What is the probability that the stock price will be less than \$60 in 4 months' time?

### Solution

We require to find:

$$P(S_{4/12} < 60) = P\left(\log_e \frac{S_{4/12}}{S_0} < \log_e \frac{60}{64} = -0.0645\right).$$

Now

$$\log_e \frac{S_{4/12}}{S_0} \sim N\left(\left[0.05 - \frac{0.12^2}{2}\right] \times \frac{4}{12}, 0.12^2 \times \frac{4}{12}\right).$$

That is,

$$\log_e \frac{S_{4/12}}{S_0} \sim N(0.0143, 0.0048).$$

So,

$$
\begin{aligned}
P\left(\log_e \frac{S_{4/12}}{S_0} < -0.0645\right) &= P\left(Z < \frac{-0.0645 - 0.0143}{\sqrt{0.0048}} = -1.14\right) \\
&= 1 - \Phi(1.14) \\
&= 1 - 0.8729 \\
&= 0.1271.
\end{aligned}
$$

## 2.7 Generating functions

### The characteristic function

Generating functions provide alternative ways of characterising the probability distribution of a random variable, rather than the distribution function, which are useful mainly for theoretical purposes. We will briefly discuss three such generating functions at this point, but will return later to make use of one of them in Chapter 4.

### Definition

Let $X$ be any random variable. The **characteristic function** (c.f.) of $X$ is defined by

$$\phi_X(t) = E(e^{Xti}) \qquad t \in (-\infty, \infty),$$

where $i$ denotes the complex number, $\sqrt{-1}$.

In the particular case where $X$ is a continuous random variable, the characteristic function is essentially the Fourier transform of the probability density function:

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{xti} f_X(x)\, dx.$$

Whatever the distribution of $X$, its characteristic function is guaranteed to exist and it uniquely identifies the distribution (in the sense that the distribution function can be uniquely recovered from the characteristic function). However, because finding and using the characteristic function requires complex integration, we will not refer to it further in this course.

### The moment-generating function

### Definition

When it exists, the **moment-generating function** (m.g.f.) of $X$ is defined by

$$M_X(t) = E(e^{Xt}) = \begin{cases} \displaystyle\sum_{x \in R_X} e^{xt} p_X(x), & \text{when } X \text{ is discrete;} \\ \displaystyle\int_{R_X} e^{xt} f_X(x)\, dx, & \text{when } X \text{ is continuous.} \end{cases}$$

Notice that $M_X(t)$ is a function of (the dummy variable) $t$. As usual for expected values, the moment-generating function exists only for those values of $t$ at which the relevant sum or integral is absolutely convergent. It is not guaranteed that $M_X(t)$ is defined for all real values, $t$, or even that it is defined at all.

Using the Maclaurin expansion of $e^{xt}$, it is possible to write (under certain regularity conditions):

$$\begin{aligned} M_X(t) &= \int_{R_X} e^{xt} f_X(x)\, dx \\ &= \int_{R_X} \left(1 + xt + \frac{(xt)^2}{2!} + \frac{(xt)^3}{3!} + \cdots\right) f_X(x)\, dx \\ &= \int_{R_X} f_X(x)\, dx + t\int_{R_X} x f_X(x)\, dx + \frac{t^2}{2!}\int_{R_X} x^2 f_X(x)\, dx + \cdots \\ &= 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \cdots . \end{aligned}$$

So

$$M_X'(t) = E(X) + tE(X^2) + \frac{t^2}{2!}E(X^3) + \dots$$

and

$$M_X''(t) = E(X^2) + tE(X^3) + \frac{t^2}{2!}E(X^4) + \dots .$$

In general, the $r$th derivative ($r = 1, 2, 3, \dots$) is

$$M_X^{(r)}(t) = \frac{d^r}{dt^r}M_X(t) = E(X^r) + tE(X^{r+1}) + \frac{t^2}{2!}E(X^{r+2}) + \cdots .$$

In particular,

$$M_X^{(r)}(0) = E(X^r).$$

So, a general method for calculating the $r$th moment of $X$ is to find the moment-generating function, differentiate it $r$ times with respect to $t$ and then set $t = 0$ in the derivative.

**Example 21: the moment-generating function of the Binomial distribution**

Suppose that $X \sim Bi(n, \theta)$. Then, the m.g.f. of $X$ can be found as follows.

$$\begin{aligned}
M_X(t) &= E(e^{Xt}) \\
&= \sum_{x=0}^{n} e^{xt}\binom{n}{x}\theta^x(1-\theta)^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x}(\theta e^t)^x(1-\theta)^{n-x} \\
&= (\theta e^t + 1 - \theta)^n \qquad \text{by the Binomial theorem.}
\end{aligned}$$

So, in this case, $M_X(t)$ is defined for every real value, $t$. Now,

$$\begin{aligned}
M_X'(t) &= n(\theta e^t + 1 - \theta)^{n-1}\theta e^t \\
M_X''(t) &= n(n-1)(\theta e^t + 1 - \theta)^{n-2}\theta e^t\theta e^t + n(\theta e^t + 1 - \theta)^{n-1}\theta e^t.
\end{aligned}$$

Therefore,

$$\begin{aligned}
E(X) = M_X'(0) &= n(\theta e^0 + 1 - \theta)^{n-1}\theta e^0 = n\theta \\
E(X^2) = M_X''(0) &= n(n-1)(\theta e^0 + 1 - \theta)^{n-2}\theta e^0\theta e^0 + n(\theta e^0 + 1 - \theta)^{n-1}\theta e^0 \\
&= n(n-1)\theta^2 + n\theta.
\end{aligned}$$

Then

$$Var(X) = E(X^2) - [E(X)]^2 = n(n-1)\theta^2 + n\theta - (n\theta)^2 = n\theta(1-\theta).$$

**Example 22: the moment-generating function of the Normal distribution**

When $X \sim N(\mu, \sigma^2)$, its moment-generating function can be found as follows:

$$M_X(t) = E(e^{Xt}) = \int_{-\infty}^{\infty} e^{xt}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx.$$

The exponent (power of e) in this integral is:

$$\begin{aligned}
xt - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 &= -\frac{(x-\mu)^2 - 2xt\sigma^2}{2\sigma^2} = -\frac{x^2 - 2(\mu + t\sigma^2)x + \mu^2}{2\sigma^2} \\
&= -\frac{[x-(\mu+t\sigma^2)]^2 - 2t\mu\sigma^2 - t^2\sigma^4}{2\sigma^2} \qquad \text{(completing square in } x\text{)} \\
&= -\frac{[x-(\mu+t\sigma^2)]^2}{2\sigma^2} + \mu t + \tfrac{1}{2}t^2\sigma^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
M_X(t) &= \exp\left(\mu t + \tfrac{1}{2}t^2\sigma^2\right)\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\left(\frac{x-(\mu+t\sigma^2)}{\sigma}\right)^2\right]dx \\
&= \exp\left(\mu t + \tfrac{1}{2}t^2\sigma^2\right).
\end{aligned}$$

The last step follows because the integrand is the p.d.f. of a $N(\mu + t\sigma^2, \sigma^2)$ random variable and hence must integrate to one. Then,

$$M_X'(t) = (\mu + t\sigma^2)\exp\left(\mu t + \tfrac{1}{2}t^2\sigma^2\right),$$

so that

$$E(X) = M_X'(0) = \mu$$

and

$$M_X''(t) = (\mu + t\sigma^2)^2\exp\left(\mu t + \tfrac{1}{2}t^2\sigma^2\right) + \sigma^2\exp\left(\mu t + \tfrac{1}{2}t^2\sigma^2\right),$$

so that

$$E(X^2) = M_X''(0) = \mu^2 + \sigma^2$$

and

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \sigma^2.$$

It might already appear from these two examples that the moment-generating function is useful since, in many cases, it allows us to obtain expected values and variances more conveniently than through direct summation or integration. The moment-generating function, however, is most useful in theoretical work because it characterises a probability distribution uniquely. We state without proof the following important property.

### Proposition 2.6: uniqueness property of moment-generating functions

Let $X$ and $Y$ be random variables. If the moment-generating functions $M_X(t)$ and $M_Y(t)$ exist and are equal for all values of $t$ in an interval $(-h, h)$, for some $h > 0$, then the distribution functions $F_X$ and $F_Y$ are the same.

This uniqueness property of moment-generating functions will be used extensively to derive theoretical results in later parts of the course.

### Proposition 2.7

Suppose that $X$ is a random variable with moment-generating function $M_X(t)$. Define $Y$ by $Y = aX + b$, for some constants $a$ and $b$. Then, $Y$ has moment-generating function

$$M_Y(t) = e^{bt}M_X(at).$$

**Proof**

$$M_Y(t) = E(e^{Yt}) = E(e^{(aX+b)t}) = E(e^{bt}e^{Xat}) = e^{bt}E(e^{Xat}) = e^{bt}M_X(at).$$

## Example 22 (continued)

Suppose that $X \sim N(\mu, \sigma^2)$, for $\sigma > 0$. Define the random variable $Y$ by $Y = aX + b$. Then, using Proposition 2.7, $Y$ has moment-generating function

$$M_Y(t) = e^{bt}M_X(at) = e^{bt}\exp\left(at\mu + \tfrac{1}{2}(at)^2\sigma^2\right) = \exp\left((a\mu + b)t + \tfrac{1}{2}t^2(a\sigma)^2\right).$$

This is the moment-generating function of the Normal distribution with expected value $a\mu + b$ and variance $(a\sigma)^2$. The Uniqueness Property (Proposition 2.6) therefore tells us that $Y \sim N(a\mu + b, [a\sigma]^2)$.

## The probability-generating function

## Definition

Let $X$ be a discrete random variable whose range space is a subset of the non-negative integers. Then the **probability-generating** function of $X$ is defined by:

$$G_X(t) = E(t^X) = \sum_{x=0}^{\infty} t^x p_X(x).$$

This power series converges absolutely at least for all $t$ with $|t| \leqslant 1$, so the probability-generating function always exists at least in that interval.

## Example 23: the probability-generating function of the Poisson distribution

Suppose that the random variable $X$ has the Poi($\lambda$) distribution. Then the probability-generating function of $X$ is found as follows:

Taking derivatives of the probability-generating function, $G_X(t)$, gives:

$$\begin{aligned}
G_X(t) &= E(t^X) = p_X(0) + tp_X(1) + t^2p_X(2) + t^3p_X(3) + \cdots \\
G'_X(t) &= p_X(1) + 2tp_X(2) + 3t^2p_X(3) + \cdots \\
G''_X(t) &= 2p_X(2) + 6tp_X(3) + \cdots
\end{aligned}$$

So:

$$\begin{aligned}
G_X^{(r)}(t) &= r!p_X(r) + (r+1)!tp_X(r+1) + \cdots \\
\implies p_X(r) &= \frac{G_X^{(r)}(0)}{r!}.
\end{aligned}$$

Also:

$$\begin{aligned} G'_X(1) &= p_X(1) + 2p_X(2) + 3p_X(3) + \cdots \\ &= 0p_X(0) + 1p_X(1) + 2p_X(2) + 3p_X(3) + \cdots \\ &= E(X). \end{aligned}$$

Again,

$$\begin{aligned} G''_X(1) &= 2p_X(2) + 6p_X(3) + \cdots \\ &= 0 \times (-1) \times p_X(0) + 1 \times 0 \times p_X(1) + 2 \times 1 \times p_X(2) + 3 \times 2 \times p_X(3) + \cdots \\ &= E[X(X-1)], \end{aligned}$$

which means that

$$E(X^2) = E[X(X-1)] + E[X] = G''_X(1) + G'_X(1)$$

and

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = G''_X(1) + G'_X(1) - [G'_X(1)]^2.$$

**Example 23 (continued)**