Chapter 4 RANDOM VARIABLES

Experiments whose outcomes are numbers

EXAMPLE:
Select items at random from a batch of size N until the first defective item is found.
Record the number of non-defective items.

Sample Space: $\mathcal{S} = \{0, 1, 2, \ldots, N\}$

The result from the experiment becomes a **variable**; that is, a quantity taking different values on different occasions. Because the experiment involves selection at random, we call it a **random variable**.

Abbreviation: rv

Notation: capital letter, often X.

DEFINITION:

The set of possible values that a random variable X can take is called the **range** of X.

EQUIVALENCES

Unstructured Experiment	Random Variable
$\mathcal{E}$	X
Sample space	range of X
Outcome of $\mathcal{E}$	One possible value x for X
Event	Subset of range of X
Event A	$x \in$ subset of range of X e.g., $x = 3$ or $2 \leq x \leq 4$
Pr(A)	Pr($X = 3$), Pr($2 \leq X \leq 4$)

REMINDER:

The set of possible values that a random variable (rv) X can take is called the **range** of X.

DEFINITION:

A rv X is said to be **discrete** if its range consists of a **finite** or **countable** number of values.

Examples: based on tossing a coin repeatedly

No. of H in 1st 5 tosses: $\{0, 1, 2, \ldots., 5\}$

No. of T before first H: $\{0, 1, 2, \ldots.\}$

Note: although the definition of a discrete rv allows it to take any finite or countable set of values, the values are in practice almost always integers.

Probability Distributions

or 'How to describe the behaviour of a rv'

Suppose that the only values a random variable X can take are x_1, x_2, $\ldots, x_n$. That is, the range of X is the set of n values $x_1, x_2, \ldots x_n$.
Since we can list all possible values, this random variable X must be discrete.

Then the behaviour of X is completely described by giving the probabilities of all relevant events:

Event	Probability
$X = x_1$	$\Pr(X = x_1)$
$X = x_2$	$\Pr(X = x_2)$
$\vdots$	$\vdots$
$X = x_n$	$\Pr(X = x_n)$

In other words, we specify the function $\Pr(X = x)$ for all values x in the range of X.

DEFINITION:

The **Probability Function** of a discrete random variable X is the function $p(x)$ satisfying

$$p(x) \;=\; \mathrm{Pr}(X = x)$$

for all values x in the range of X.

Abbreviation: pf

Notation: $p(x)$ or $p_X(x)$. We use the $p_X(x)$ form when we need to make the identity of the rv clear.

Terminology: The pf is sometimes given the alternative name of probability mass function (pmf).

EXAMPLE:
Let the probability of a head on any toss of a particular coin be p. From independent successive tosses of the coin, we record the number X of tails before the first head appears.

Range of X : $\{0, 1, 2, \ldots\}$

$$\begin{aligned}
\Pr(X = 0) &= p \\
\Pr(X = 1) &= (1 - p)p \\
&\vdots \\
\Pr(X = x) &= (1 - p)^x p \\
&\vdots
\end{aligned}$$

The probability function for the random variable X gives a convenient summary of its behaviour; the pf $p_X(x)$ is given by:

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

X is said to have a **Geometric Distribution**.

Properties of a pf

If $p_X(x)$ is the pf of a rv X, then

- $p_X(x) \geq 0$, for all x in the range of X.
- $\sum p_X(x) = 1$, where the sum is taken over the range of X.

Informal 'definition' of a distribution:

The pf of a discrete rv describes how the total probability, 1, is split, or distributed, between the various possible values of X.

This 'split' or pattern is known as the **distribution** of the rv.

Note: The pf is not the only way of describing the distribution of a discrete rv. Any 1–1 function of the pf will do.

DEFINITION:

The **cumulative distribution function** of a rv X is the function $F_X(x)$ of x given by

$$F_X(x) \;=\; \Pr(X \leq x),$$

for all values x in the range of X.

Abbreviation: cdf

Terminology: The cdf is sometimes given the alternative name of distribution function.

Notation: $F(x)$ or $F_X(x)$. We use the $F_X(x)$ form when we need to make the identity of the rv clear.

Relationship with pf: For a **discrete** rv X,

$$F_X(x) = \sum_{y \leq x} p_X(y)$$

Example: If a rv has range $\{0, 1, 2, \ldots\}$,

$$F_X(3) = p_X(0) + p_X(1) + p_X(2) + p_X(3)$$

and

$$p_x(2) = F_X(2) - F_X(1).$$

EXAMPLE: **Discrete Uniform Distribution**

The rv X is equally likely to take each integer value in the range $1, 2, \ldots, n$.

Probability function:

$$p_X(x) = \begin{cases} \frac{1}{n}, & x = 1, 2, \ldots, n\,, \\ 0 & \text{elsewhere.} \end{cases}$$

Cumulative distribution function:

$$F_X(x) = \begin{cases} 0\,, & x < 1\,, \\[1ex] \frac{[x]}{n}, & 1 \leq x \leq n\,, \\[1ex] 1\,, & x \geq n\,, \end{cases}$$

where $[x]$ is the integer part of x.

Note: The cdf is defined for all values of x, not just the ones in the range of X.
For this distribution, the cdf is $\frac{1}{n}$ for all values of x in the range $1 \leq x < 2$, then jumps to $\frac{2}{n}$, and so on.

Properties of cdfs:

All cdfs

- are monotonic non-decreasing,
- satisfy $F_X(-\infty) = 0$,
- satisfy $F_X(\infty) = 1$.

Any function satisfying these conditions can be a cdf.

A function not satisfying these conditions cannot be a cdf.

For a discrete rv the cdf is always a **step** function.

Reminder: Properties of cdfs: **Any** function satisfying the following conditions can be a cdf:

- It is monotonic non-decreasing,

- It satisfies $F_X(-\infty) = 0$,
- It satisfies $F_X(\infty) = 1$.

DEFINITION: A random variable is said to be **continuous** if its cdf is a continuous function (see later).

This is an important case, which occurs frequently in practice.

EXAMPLE: The Exponential Distribution

Consider the rv Y with cdf

$$F_Y(y) = \begin{cases} 0\,, & y < 0\,, \\ 1 - e^{-y}\,, & y \geq 0\,. \end{cases}$$

This meets all the requirements above, and is not a step function.
The cdf is a continuous function.

Types of random variable

Most rvs are either discrete or continuous, **but**

- one can devise some complicated counter-examples, and
- there are practical examples of rvs which are partly discrete and partly continuous.

EXAMPLE: Cars pass a roadside point, the gaps (in time) between successive cars being exponentially distributed.
Someone arrives at the roadside and crosses as soon as the gap to the next car exceeds 10 seconds. The rv T is the delay before the person starts to cross the road.

The delay T may be zero or positive. The chance that $T = 0$ is positive; the cdf has a step at $t = 0$. But for $t > 0$ the cdf will be continuous.

Mean and Variance

The pf gives a complete description of the behaviour of a (discrete) random variable. In practice we often want a more concise description of its behaviour.

DEFINITION: The **mean** or **expectation** of a discrete rv X, $\mathsf{E}(X)$, is defined as

$$\mathsf{E}(X) = \sum_x x \, \mathsf{Pr}(X = x).$$

Note: Here (and later) the notation $\sum_x$ means the sum over all values x in the range of X.

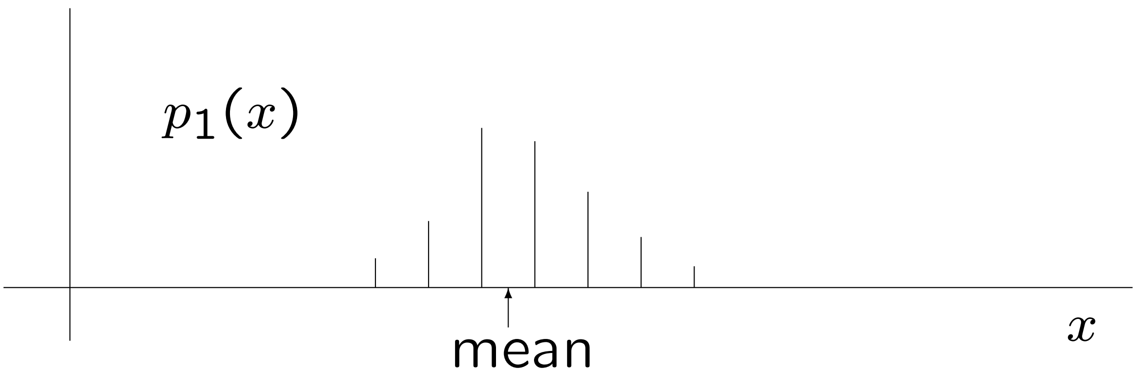
The expectation $\mathsf{E}(X)$ is a weighted average of these values. The weights always sum to 1.

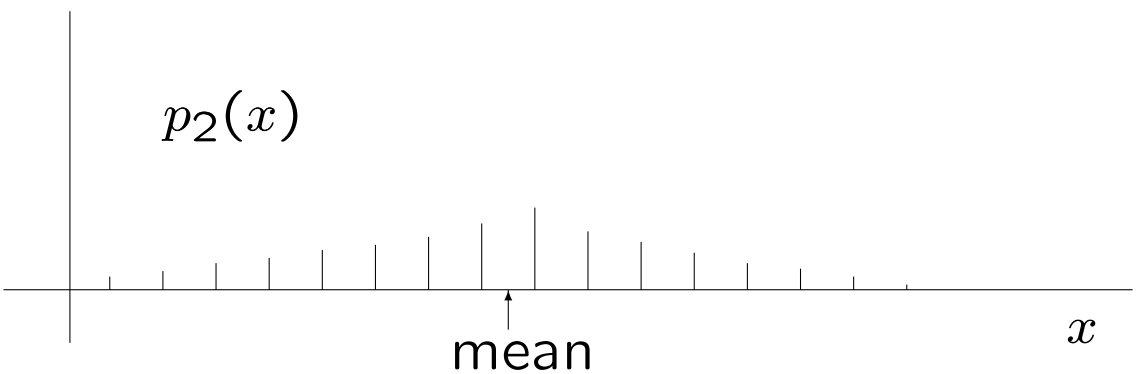
Extension: The concept of expectation can be generalised; we can define the expectation of any function of a rv. Thus we obtain, for a function $g(\cdot)$ of a discrete rv X,

$$\mathsf{E}\{g(X)\} = \sum_x g(x) \, \mathsf{Pr}(X = x) \, .$$

Measures of variability

Two rvs can have equal means but very different patterns of variability. Here is a sketch of the probability functions $p_1(x)$ and $p_2(x)$ of two rvs X_1 and X_2.





To distinguish between these, we need a measure of spread or **dispersion**.

Measures of dispersion

There are many possible measures. We look briefly at three plausible ones.

A. 'Mean difference': $\mathsf{E}\{X - \mathsf{E}(X)\}$.
Attractive superficially, but no use.

B. Mean absolute difference: $\mathsf{E}\{|X - \mathsf{E}(X)|\}$.
Hard to manipulate mathematically.

C. Variance: $\mathsf{E}\{X - \mathsf{E}(X)\}^2$.
The most frequently-used measure.

Notation for variance: $\mathsf{V}(X)$ or $\mathsf{Var}(X)$.

That is: $\quad \mathsf{V}(X) = \mathsf{Var}(X) = \mathsf{E}\{X - \mathsf{E}(X)\}^2$.

Summary and formula

The most important features of a distribution are its location and dispersion, measured by expectation and variance respectively.

Expectation: $\mathsf{E}(X) = \sum_x x \, \mathsf{Pr}(X = x) = \mu\,.$

Variance:

$$\begin{aligned}
\mathsf{Var}(X) &= \sum_x (x-\mu)^2 \, \mathsf{Pr}(X = x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2) \, \mathsf{Pr}(X = x) \\
&= \sum_x x^2 \, \mathsf{Pr}(X = x) - 2\mu \cdot \mu + \mu^2 \cdot 1 \\
&= \mathsf{E}(X^2) - \{\mathsf{E}(X)\}^2
\end{aligned}$$

Reminder: The notation $\sum_x$ means the sum over all values x in the range of X.

Notation: We often denote $\mathsf{E}(X)$ by μ, and $\mathsf{Var}(X)$ by σ^2.

EXAMPLE:

We find the mean and variance for the random variable X with pf as in the table:

x	1	2	3	4	5
$p(x) = \mathrm{Pr}(X = x)$	0.1	0.1	0.2	0.4	0.2

$\mathrm{E}(X) = \sum_x x \, \mathrm{Pr}(X = x)$, so

$$\mathrm{E}(X) = (1 \times 0.1) + (2 \times 0.1) + (3 \times 0.2) + (4 \times 0.4) + (5 \times 0.2) = 3.5.$$

$\mathrm{Var}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2$, and

$$\mathrm{E}(X^2) = (1^2 \times 0.1) + (2^2 \times 0.1) + (3^2 \times 0.2) + (4^2 \times 0.4) + (5^2 \times 0.2) = 13.7\,,$$

so $\mathrm{Var}(X) = 13.7 - (3.5)^2 = 1.45$.

Standard deviation of X: $\sqrt{1.45}$, or 1.20.

Notes

1. The concepts of expectation and variance apply equally to discrete and continuous random variables. The formulae given here relate to discrete rvs; formulae need (slight) adaptation for the continuous case.

2. Units: the mean is in the same units as X, the variance $\mathrm{Var}(X)$, defined as

$$\mathrm{Var}(X) = \mathrm{E}\{X - \mathrm{E}(X)\}^2$$

is in squared units.

A measure of dispersion in the same units as X is the **standard deviation (s.d.)**

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

CONTINUOUS RANDOM VARIABLES

Introduction

Reminder: a rv is said to be **continuous** if its cdf is a **continuous** function.

If the function $F_X(x) = \Pr(X \leq x)$ of x is continuous, what is $\Pr(X = x)$?

$$\begin{aligned} \Pr(X = x) &= \Pr(X \leq x) - \Pr(X < x) \\ &= 0, \quad \text{by continuity} \end{aligned}$$

A continuous random variable does not possess a probability function.

Probability cannot be assigned to individual values of x; instead, probability is assigned to intervals. [Strictly, half-open intervals]

Consider the events $\{X \leq a\}$ and $\{a < X \leq b\}$. These events are mutually exclusive, and

$$\{X \leq a\} \cup \{a < X \leq b\} = \{X \leq b\}.$$

So the addition law of probability (axiom **A3**) gives:

$$\begin{aligned} \Pr(X \leq b) &= \Pr(X \leq a) + \Pr(a < X \leq b), \\ \text{or } \Pr(a < X \leq b) &= \Pr(X \leq b) - \Pr(X \leq a) \\ &= F_X(b) - F_X(a). \end{aligned}$$

So, given the cdf for any continuous random variable X, we can calculate the probability that X lies in any interval.

Note: The probability $\Pr(X = a)$ that a continuous rv X is exactly a is 0. Because of this, we often do not distinguish between open, half-open and closed intervals for continous rvs.

Example: We gave earlier an example of a continuous cdf:

$$F_Y(y) = \begin{cases} 0\,, & y < 0\,, \\ 1 - e^{-y}\,, & y \geq 0\,. \end{cases}$$

This is the cdf of what is termed the exponential distribution with mean 1.

For the case of that distribution, we can find

$$\Pr(Y \leq 1) = F_Y(1) = 1 - e^{-1} = 0.6322$$

$$\begin{aligned} \Pr(2 \leq Y \leq 3) &= F_Y(3) - F_Y(2) \\ &= (1 - e^{-3}) - (1 - e^{-2}) = 0.0856 \end{aligned}$$

$$\begin{aligned} \Pr(Y \geq 2.5) &= F_Y(\infty) - F_Y(2.5) \\ &= 1 - (1 - e^{-2.5}) = 0.0821 \end{aligned}$$

Probability density function

If X is continuous, then Pr$(X = x) = 0$.

But what is the probability that 'X is close to some particular value x?'.
Consider Pr$(x < X \leq x + h)$, for small h.

Recall: $\dfrac{d\,F_X(x)}{dx} \simeq \dfrac{F_X(x+h) - F_X(x)}{h}$.

$$\text{So Pr}(x < X \leq x + h) \; = \; F_X(x+h) - F_X(x)$$
$$\simeq \; h\,\frac{d\,F_X(x)}{dx}.$$

DEFINITION: The derivative (w.r.t. x) of the cdf of a continous rv X is called the **probability density function** of X.

The probability density function is the limit of

$$\frac{\text{Pr}(x < X \leq x + h)}{h} \text{ as } h \to 0\,.$$

The probability density function

Alternative names: pdf,
density function,
density.

Notation for pdf: $f_X(x)$

Recall: The cdf of X is denoted by $F_X(x)$

Relationship: $$f_X(x) = \frac{d\,F_X(x)}{dx}$$

Care needed: Make sure f and F cannot be confused!

Interpretation

- When multiplied by a small number h, the pdf gives, approximately, the probability that X lies in a small interval, length h, close to x.
- If, for example, $f_X(4) = 2\,f_X(7)$, then X occurs near 4 twice as often as near 7.

Properties of probability density functions

Because the pdf of a rv X is the derivative of the cdf of X, it follows that

- $f_X(x) \geq 0, \quad \text{for all } x,$
- $\int_{-\infty}^{\infty} f_X(x)\, dx = 1,$
- $F_X(x) = \int_{-\infty}^{x} f_X(y) dy,$
- $\Pr(a < X \leq b) = \int_a^b f_X(x) dx.$

Mean and Variance

Reminder: for a discrete rv, the formulae for mean and variance are based on the probability function $\mathrm{Pr}(X = x)$. We need to adapt these formulae for use with continuous random variables.

DEFINITION:
For a continuous rv X with pdf $f_X(x)$, the expectation of a function $g(x)$ is defined as

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

Hence, for the **mean**:

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

Compare this with the equivalent definition for a discrete random variable:

$$\mathrm{E}(X) = \sum_x x\, \mathrm{Pr}(X = x)\,, \text{ or } \mathrm{E}(X) = \sum_x x p_X(x)\,.$$

For the **variance**, recall the definition.

$$\begin{aligned} \mathsf{Var}(X) &= \mathsf{E}[\{X - \mathsf{E}(X)\}^2] \\ \text{Hence } \mathsf{Var}(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\, dx \end{aligned}$$

As in the discrete case, the best way to **caclulate** a variance is by using the result:

$$\mathsf{Var}(X) = \mathsf{E}(X^2) - \{\mathsf{E}(X)\}^2 .$$

In practice, we therefore usually calculate

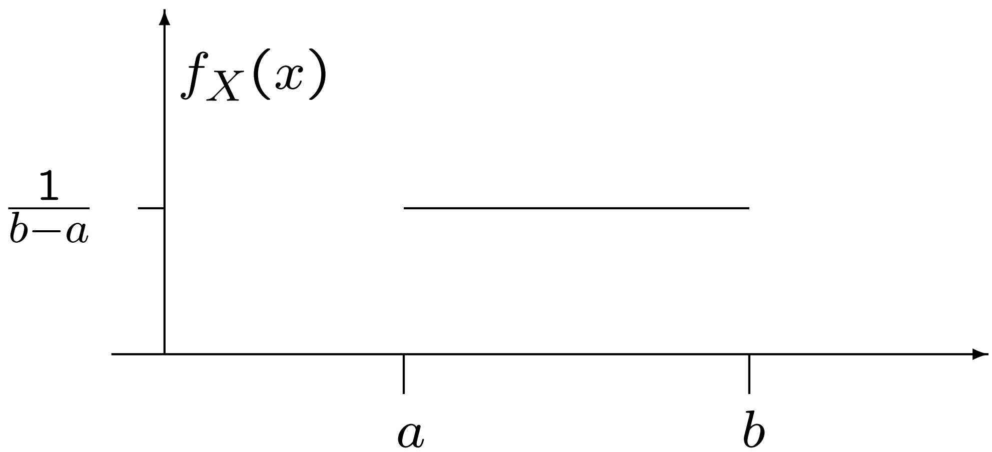
$$\mathsf{E}(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx$$

as a stepping stone on the way to obtaining $\mathsf{Var}(X)$.

The Uniform Distribution

Distribution of a rv which is equally likely to take any value in its range, say a to b ($b > a$).

The pdf is constant:



Because $f_X(x)$ is constant over $[a, b]$ and

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

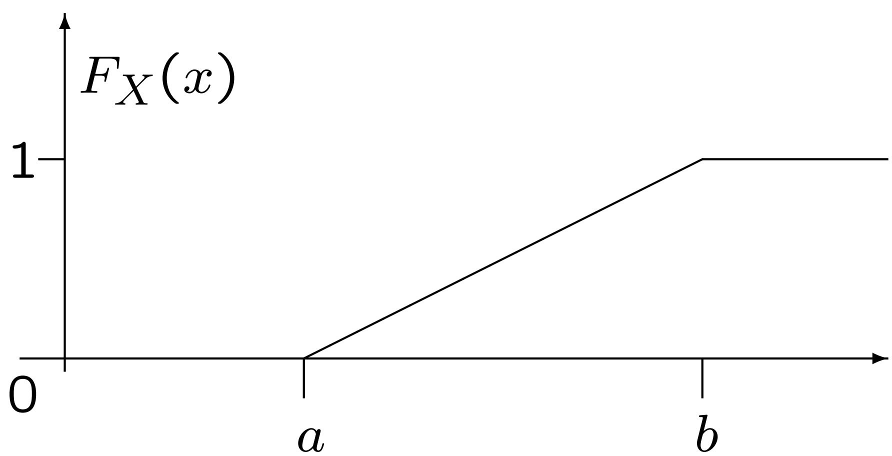
$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b, \\ 0 & \text{elsewhere.} \end{cases}$$

Uniform Distribution: cdf

For this distribution the cumulative distribution function (cdf) is

$$F_X(x) \;=\; \int_{-\infty}^{x} f_X(y)\,dy$$

$$= \begin{cases} 0\,, & x < a\,, \\ \frac{x-a}{b-a}\,, & a \le x \le b\,, \\ 1\,, & x > b\,. \end{cases}$$



Uniform Distribution: Mean and Variance

$$\begin{aligned} \mathsf{E}(X) = \mu &= \int_a^b x \frac{1}{b-a}\, dx \\ &= \tfrac{1}{2}(a+b). \end{aligned}$$

$$\begin{aligned} \mathsf{Var}(X) = \sigma^2 &= \mathsf{E}(X^2) - \mu^2 \\ &= \int_a^b x^2 \frac{1}{b-a}\, dx - \frac{(a+b)^2}{4} \\ &= \frac{1}{12}(b-a)^2. \end{aligned}$$

For example, if a random variable is uniformly distributed on the range (20,140), then $a = 20$ and $b = 140$, so the mean is 80. The variance is 1200, so the standard deviation is 34.64.

The exponential distribution

A continuous random variable X is said to have an exponential distribution if its range is $(0, \infty)$ and its pdf is proportional to $e^{-\lambda x}$, for some positive λ.

That is,

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

for some constant k. To evaluate k, we use the fact that all pdfs must integrate to 1.

Hence

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x)\,dx &= \int_0^{\infty} ke^{-\lambda x}\,dx \\ &= \frac{k}{\lambda}\left[-e^{-\lambda x}\right]_0^{\infty} \\ &= \frac{k}{\lambda} \end{aligned}$$

Since this must equal 1, $k = \lambda$.

Properties of the exponential distribution

The distribution has pdf

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

and its cdf is given by

$$\begin{aligned} F_X(x) &= \int_0^x \lambda e^{-\lambda y}\, dy \\ &= 1 - e^{-\lambda x}, \quad x > 0\,. \end{aligned}$$

Mean and Variance

$$\mathsf{E}(X) = \int_0^\infty x\, \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}.$$

For the variance, we use integration by parts to obtain

$$\mathsf{E}(X^2) = \int_0^\infty x^2\, \lambda e^{-\lambda x}\, dx = \frac{2}{\lambda^2}\,.$$

$$\begin{aligned} \text{Hence } \mathsf{Var}(X) &= \mathsf{E}(X^2) - \{\mathsf{E}(X)\}^2 \\ &= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}. \end{aligned}$$

The Normal Distribution

DEFINITION: A random variable X with probability density function

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

for all x, is said to have the Normal distribution with parameters μ and σ^2.

It can be shown that $\mathrm{E}(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$.

We write: $X \sim \mathrm{N}(\mu, \sigma^2)$.

Shape of the density function (pdf):

The pdf is symmetrical about $x = \mu$.
It has a single mode at $x = \mu$.
It has points of inflection at $x = \mu \pm \sigma$.
'A bell-shaped curve,' tails off rapidly.

Cumulative distribution function

If $X \sim \mathsf{N}(\mu, \sigma^2)$, the cdf of X is the integral:

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx.$$

This cannot be evaluated analytically.

Numerical integration is necessary: extensive tables are available.

The Standardised Normal Distribution

The Normal distribution with mean 0 and variance 1 is known as the standardised Normal distribution (SND). We usually denote a random variable with this distribution by Z. Hence

$$Z \sim \mathrm{N}(0, 1).$$

Special notation $\phi(z)$ is used for the pdf of $\mathrm{N}(0, 1)$. We write

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad -\infty < z < \infty.$$

The cdf of Z is denoted by $\Phi(z)$. We write

$$\begin{aligned} \Phi(z) &= \int_{-\infty}^{z} \phi(x)\, dx \\ &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx \end{aligned}$$

Tables of $\Phi(z)$ are available in statistical textbooks and computer programs.

Brief extract from a table of the SND

Z	$\Phi(z)$
0.0	0.5000
0.5	0.6915
1.0	0.8413
1.5	0.9332
2.0	0.9772

Tables in textbooks and elsewhere contain values of $\Phi(z)$ for $z =$ 0, 0.01, 0.02, and so on, up to $z = 4.0$ or further.

But the range of Z is $(-\infty, \infty)$, so we need values of $\Phi(z)$ for $z < 0$. To obtain these values we use the fact that the pdf of $\mathrm{N}(0, 1)$ is symmetrical about $z = 0$.
This means that

$$\Phi(z) = 1 - \Phi(-z).$$

This equation can be used to obtain $\Phi(z)$ for negative values of z.
For example, $\Phi(-1.5) = 1 - 0.9332 = 0.0668$.