Chap 2. Probability and Probability distributions

This chapter will discuss some basic notions of **probability** and some **probability distributions** for discrete and continuous data.

Probability

- Basic notions of probability
- **Classical** probability
- **Frequentist** probability
- Laws of probability
- **Independence**

Probability distributions

- **Discrete distributions**
  - The binomial distribution
  - The Poisson distribution
- **Continuous distributions**
  - The Normal distribution
  - The uniform distribution

2.1.1 Basic notions

Probability is the language we use to model uncertainty.

We all intuitively understand that few things in life are certain. There is usually an element of **uncertainty** or **randomness** around outcomes of our choices.

In engineering this uncertainty can mean the difference between life and death!

Hence an understanding of probability and how we might incorporate this into our decision making processes is important.

Definitions

We often use the letter P to represent a probability.

For example, $P(Rain)$ would be the probability that it rains.

An **Experiment** is an activity where we do not know for certain what will happen, but we can *observe* what happens.

An **Outcome** is one of the possible things that can happen.

The **Sample space** is the set of all possible outcomes.

An **Event** is a set of outcomes.

More stuff...

Probabilities are usually expressed in terms of **fractions**, **decimals** or **percentages**.

All probabilities are measured on a scale from zero to one.

- An **impossible** event has a probability of zero
- A **certain** event has a probability of one
- The collection of all possible outcomes – the **sample space** – has a probability of 1.

Two events are said to be **mutually exclusive** if both can not occur simultaneously. $P(E_1 \text{ and } E_2) = 0$.

Two events are said to be **independent** if the occurrence of one does not affect the probability of the second occurring.

$$P(E_1 \text{ and } E_2) = P(E_1) \times P(E_2).$$

For example, if you toss a coin and look out of the window, the events "get heads" and "it is raining" would be independent.

2.1.2 Classical Probability

This view is based on the concept of **equally likely events**.

If we toss a fair coin, we have two possible outcomes – **Heads** or **Tails**. Both outcomes are **equally likely**. Thus

$$P(Head) = \frac{1}{2} \quad \text{and} \quad P(Tail) = \frac{1}{2}$$

If we toss a fair coin twice, we have four possible outcomes – **HH**, **HT**, **TH** or **TT**. All outcomes are **equally likely**. Thus

$$P(HH) = \frac{1}{4},\ P(HT) = \frac{1}{4},\ P(TH) = \frac{1}{4} \text{ and } P(HH) = \frac{1}{4}$$

In this example, there is no reason to think that the outcome *Head* and the outcome *Tail* have different probabilities.

Classical probability

$$P(\text{Event}) = \frac{\text{Total number of outcomes in which event occurs}}{\text{Total number of possible outcomes}}$$

2.1.3 Frequentist probability

When the outcomes of an experiment are not equally likely, we can **conduct experiments** to give us an idea of how likely the different outcomes are.

By conducting experiments the probability of an event can easily be estimated using the following formula:

$$P(\text{Event}) = \frac{\text{Number of times an event occurs}}{\text{Total number of times experiment done}}.$$

The larger the experiment, the closer this probability is to the "**true**" probability.

Example The following data are the daily rainfall totals (in mm) for Kolkata, India:

41	**36**	**12**	**25**	**30**	**0**	**0**	**15**	**51**

$$P(\text{no more than 30 mm of rain}) = 6/9$$

$$P(\text{more than 25 mm of rain}) = 4/9.$$

Warning: the sample size is 9 which is rather small. The conclusions based on a small sample may be **misleading**!

2.1.4 Laws of probability

The probability of two **independent** events E_1 and E_2 both occurring is

$$P(E_1 \text{ and } E_2) = P(E_1) \times P(E_2).$$

This method of calculating probabilities extends to when there are *many* **independent** events:

$$P(E_1 \text{ and } E_2 \text{ and } \cdots \text{ and } E_n) = P(E_1) \times P(E_2) \times \cdots \times P(E_n).$$

Remarks: There is a more complicated rule for multiplying probabilities when the events are *not* independent.

Addition Law

The **addition law** describes the probability of any of two **or** more events occurring.

The addition law for two events E_1 and E_2 is

$$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2) - P(E_1 \text{ and } E_2).$$

This describes the probability of *either* event E_1 *or* event E_2 happening.

- If events E_1 and E_2 are **independent**, then

$$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2) - P(E_1) \times P(E_2).$$

- If events E_1 and E_2 are **mutually exclusive** then

$$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2).$$

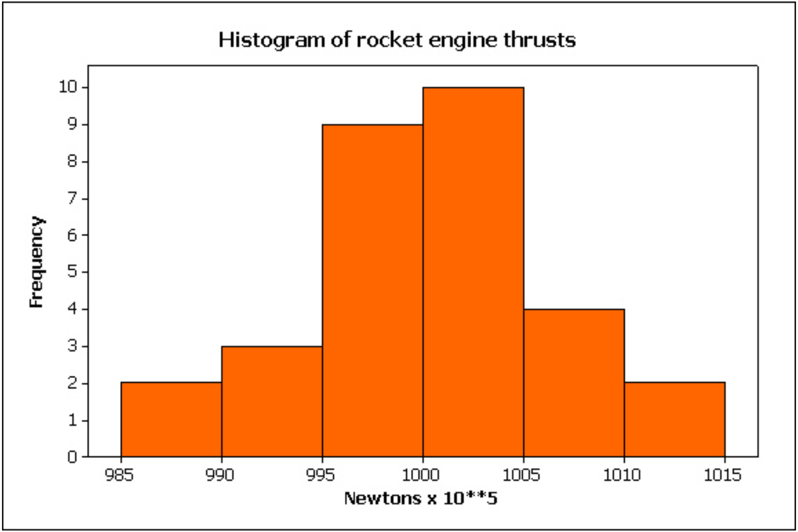
2.2 Probability distributions

Example: Rocket engine thrusts

The thrust of a rocket engine was measured at 10–minute intervals while being run at the same operating conditions. The following 30 observations were recorded (in Newtons$\times 10^5$).

999.1	1003.2	1002.1	999.2	989.7	1006.7	1012.3
996.4	1000.2	995.3	1008.7	993.4	998.1	997.9
1003.1	1002.6	1001.8	996.5	992.8	1006.5	1004.5
1000.3	1014.5	998.6	989.4	1002.9	999.3	994.7
1007.6	1000.9					

The following figure shows a histogram of these data, and their **probability distribution**.



X–the thrust of a rocket engine. Does X belong to any **standard** distribution (or model)?

2.2 Probability distributions

There are a number of '**standard**' probability distributions which data often adopt.

If we can learn to **recognise the situations** in which these 'standard' distributions occur, we can simplify the nature of the analysis which we perform on the data.

Two major subdivisions occur: **discrete distributions**, where we usually have counts, and **continuous distributions**, where values are from a continuous scale.

We will look at two discrete distributions (the **binomial** and **Poisson** distributions) and two continuous distributions (the **Normal** and **uniform** distributions).

2.2.1 Discrete distributions: binomial distribution

Bernoulli trial:

- There are only two possible outcomes, namely **success** and **failure**.
- $P($**success**$) = p$.
- Example: if we toss a coin, $P(\text{"}head\text{"}) = 0.5$.

Suppose the following statements hold:

- There are a fixed number of Bernoulli trials (n) (i.e. in each trial there are only two possible outcomes 'success' or 'failure');
- There is a constant probability of 'success', p;
- The outcome of each trial is independent of any other trial.

Then the total number of successes in n trials, X, follows a **binomial distribution**. We write $X \sim \text{Bin}(n, p)$.
Example: X – number of "heads" if we toss a coin three times.

If $X \sim \text{Bin}(n, p)$, then

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

where

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \qquad \text{and}$$

$$A! = A \times (A-1) \times (A-2) \ldots \times 3 \times 2 \times 1$$

If we assume a binomial distribution, then the following formulae give the **mean** and **variance**:

$$\begin{aligned} \text{mean} &= n \times p \\ \text{variance} &= n \times p \times (1 - p) \end{aligned}$$

Example: The probability that a fluorescent light has a life of over 500 hours is 0.9. Amongst a box of a dozen of such lights,

1. Find the probabilities that
   - exactly ten last for more than 500 hours;
   - at least ten last for more than 500 hours;
   - at most 2 last for less than 500 hours.
2. On average, how many lights in a box can last for more than 500 hours?

Poisson distribution

Suppose the following hold:

- There is no natural upper limit to the number of trials;
- Events occur independently, at a constant rate (λ);
- Two, or more, events cannot occur simultaneously.

Then the number of events, X, occurring with rate λ, has a **Poisson distribution**. We write

$$X \sim \text{Poi}(\lambda).$$

Poisson distribution

If $X \sim \text{Poi}(\lambda)$, then

$$P(X = x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}, \qquad x = 0, 1 \ldots$$

The **mean** and **variance** are :

$$\text{mean} = \lambda$$

$$\text{variance} = \lambda$$

Examples.

1. Vehicles pass a point on a busy road at an average rate of 210 per hour.
   - What is the average number passing in 2 minutes?
   - Find the probability that non pass in 2 minutes.
   - Find the probability that at least three vehicles pass in 2 minutes.
2. In the River Wear, a certain bacteria occurs at a rate of 5 per litre.
   - Find the probability of observing less than 3 bacteria in any one litre jar.
   - Find the probability of observing *more than* 2 in any one litre jar.
   - Find the probability of observing exactly 8 bacteria in a two -litre jar.

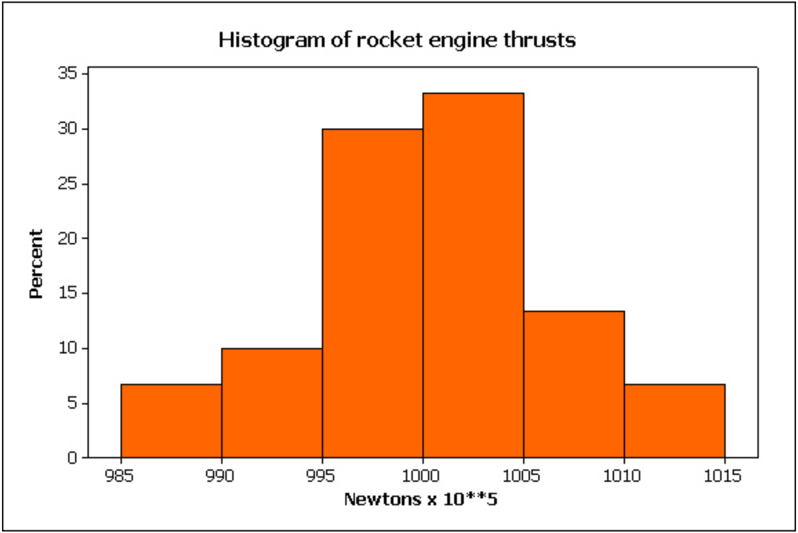
2.2.2 Continuous distributions

Example: Rocket engine thrusts

The thrust of a rocket engine was measured at 10–minute intervals while being run at the same operating conditions. The following 30 observations were recorded (in Newtons$\times 10^5$).

999.1	1003.2	1002.1	999.2	989.7	1006.7	1012.3
996.4	1000.2	995.3	1008.7	993.4	998.1	997.9
1003.1	1002.6	1001.8	996.5	992.8	1006.5	1004.5
1000.3	1014.5	998.6	989.4	1002.9	999.3	994.7
1007.6	1000.9					

The following slide shows a histogram of these data, and their **probability distribution**.


Histogram of rocket engine thrusts
Percent
35
30
25
20
15
10
5
0
985
990
995
1000
1005
1010
1015
Newtons x 10**5

A histogram gives an indication of the relative frequencies of different values. As n increases, it will tend to a smooth curve known as a **probability density function**.

The area under this curve between $[a, b]$ gives $P(a < X \leq b)$, i.e. the probability that X lies between a and b.

As with probability distributions in the discrete case, the **area under this curve must be equal to 1**.

The Normal distribution

The Normal distribution is without doubt the most widely–used statistical distribution in many practical applications:

- Normality arises naturally in many physical, biological and social measurement situations
- Normality is important in **Statistical inference** (see later)
- It has many guises:
  - Gaussian distribution
  - Laplacean distribution
  - "bell–shaped curve"

- The normal distribution is a **continuous distribution**
- It has **probability density function**, or **PDF**

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

- We write $X \sim N(\mu, \sigma^2)$
- The parameters μ and σ^2 are the **mean** and **variance** respectively

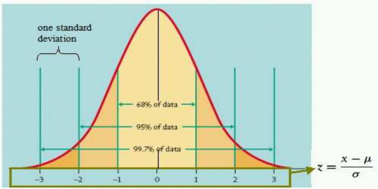
The expression above is the **general** form of the PDF for the normal distribution.

The *standard* Normal distribution

- The **standard** Normal distribution arises when $\mu = 0$ and $\sigma = 1$
- This gives the PDF

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\}$$

- **Statistical tables** give probabilities for the standard Normal distribution (see handout)



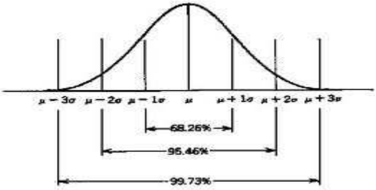
link

We can re–scale any normal distribution to the standard Normal! If $X \sim N(\mu, \sigma^2)$, then

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

The normal distribution has the following properties:

- Half of the population exceeds μ and half is less than μ;
- Approximately 2/3 of values lie within one standard deviation of the mean;
- Approximately 95% of values lie within two standard deviations of the mean;
- Almost all values lie within 3 three standard deviations of the mean.



link

Example Continue to consider the data on rocket thrusts. From our sample, the **mean thrust is 1000 Newtons**$\times 10^5$, with a **standard deviation of 6**. Assuming a Normal distribution for rocket thrusts, find the probability that:

(i) a randomly selected rocket has a thrust of less than 990 Newtons$\times 10^5$;

(ii) a randomly selected rocket has a thrust of more than 1005 Newtons$\times 10^5$;

(iii) the thrust of a rocket lies between 996 and 1002 Newtons$\times 10^5$.

Solution – part (i)

Let

$$X\text{: } \textbf{Rocket thrust} \text{ (Newtons} \times 10^5\text{)}$$

Then,

$$X \sim N(1000, 6^2)$$

We require $P(X < 990)$:

$$\begin{aligned} P(X < 990) &= P\left(Z < \frac{990 - 1000}{6}\right) \\ &= P\left(Z < -1.66667\right) \\ &= 1 - P(Z \leq 1.67) \\ &= 1 - 0.9525 = 0.0475. \end{aligned}$$

Solution – part (ii)

$$X \sim N(1000, 6^2)$$

We require $P(X > 1005)$:

$$\begin{aligned} P(X > 1005) &= 1 - P(X < 1005) \\ &= 1 - P\left(Z < \frac{1005 - 1000}{6}\right) \\ &= 1 - P(Z < 0.83) \\ &= 0.203. \end{aligned}$$

Solution – part (iii)

$$X \sim N(1000, 6^2)$$

We require $P(996 < X < 1002)$.

Actually, we have

$$P(996 < X < 1002) = P(X < 1002) - P(X < 996).$$

(ii) The uniform distribution

X is said to be a uniform random variable on a finite interval (a, b) if it takes any value in (a,b) with equal probability. We write $X \sim U(a, b)$ to mean that X can only take values in the inteval (a, b) and has PDF

$$f(x) = \frac{1}{b-a}, \quad \text{for} \quad a < x < b.$$

From the above PDF, we have

$$P(X < c) = \frac{c-a}{b-a} \quad \text{for any} \quad a < c < b.$$

If we assume a uniform distribution, then the following formulae give the **mean** and **variance**:

$$\begin{aligned} \text{mean} &= \frac{a+b}{2}; \\ \text{variance} &= \frac{(b-a)^2}{12}. \end{aligned}$$

(ii) The uniform distribution: Example

The amount of time, in minutes, that a person must wait for a bus is uniformly distributed between 0 and 15 minutes, inclusive.

1. What is the probability that a person waits fewer than 12.5 minutes?
2. On the average, how long must a person wait?
3. Ninety percent of the time, the time a person must wait falls below what value?