2 Probability and Probability distributions

This chapter will give a brief review on some basic notions of ***probability*** and introduce 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 Probability

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. **For example:**

- We will ask someone whether or not they have used our product.
- We will observe the temperature at midday tomorrow.
- We will toss a coin and observe whether it shows "heads" or "tails".

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

The ***Sample space*** is the set of all possible outcomes. For example, it could be the set of all shoe sizes.

An ***Event*** is a set of outcomes. For example "the shoe size of the next customer is less than 9" is an event. It is made up of all of the outcomes where the shoe size is less than 9.

Probabilities are usually expressed in terms of ***fractions***, ***decimals*** or ***percentages***. Therefore we could express the probability of it raining today as

$$P(Rain) = \frac{1}{20} = 0.05 = 5\%.$$

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.
- An ***evens*** event has a probability of 0.5.
- Can you imagine where about on this scale a ***likely*** event will lie? Or an ***extremely unlikely event***?

The collection of all possible outcomes – the ***sample space*** – has a probability of 1. For example: suppose an event has only two outcomes – *success* or *failure*, then

$$P(success \text{ or } failure) = 1.$$

Another example: Suppose we have a fair six–sided die, then

$$P(1 \text{ or } 2 \text{ or } 3 \text{ or } 4 \text{ or } 5 \text{ or } 6) = 1.$$

Two events are said to be ***mutually exclusive*** if both can not occur simultaneously. In the example above, the outcomes *success* and *failure* are mutually exclusive.

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

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

The underlying idea behind this view of probability is ***symmetry***.

In this example, there is no reason to think that the outcome *Head* and the outcome *Tail* have different probabilities. Since there are two outcomes and one of them must occur, both outcomes must have probability 1/2.

Another commonly used example is ***rolling dice***. There are **six** possible outcomes – (1, 2, 3, 4, 5, 6) – if the die is fair, each of them should have an equal chance of occurring. With this in mind, we have the following ***Probability distribution*** table:

Outcome	1	2	3	4	5	6
Probability	*1/6*	*1/6*	*1/6*	*1/6*	*1/6*	*1/6*

Using the following formula:

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

we find the probabilities of these events:

$$P(\text{Even Number}) = 1/2.$$

$$P(\text{Odd Number}) = 1/2.$$

$$P(\text{multiple of three}) = 1/3.$$

Example. A washing basket contains two ***green*** socks, three ***yellow*** socks, an ***orange*** sock and a ***purple*** sock.

We can obtain the following probabilities:

$$P(\textbf{\textit{green}}\text{ sock}) = 2/7,$$

and

$$P(\textbf{orange}\text{ if a }\textbf{\textit{yellow}}\ \textit{has already been removed}) = 1/6.$$

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.

Examples

- *Probability of producing a defective item in a manufacturing process*: We could monitor the process over a long period of time and the probability of a defective could be measured by the proportion of defectives in our sample.
- *Imagine we believed a coin was* ***unfair***: Toss the coin a large number of times and see how many heads you obtain, and express $P(Head)$ as a proportion.

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

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

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

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

We have the following (frequentist) probabilities:

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

and

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

The larger the experiment, the closer this probability is to the "***true***" probability. The frequentist view of probability regards probability as the long run relative frequency (or proportion). In the defects example, the "true" probability of getting a defective item is the proportion obtained in a very large experiment (strictly an ***infinitely*** long sequence of trials).

2.1.4 Laws of probability

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

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

For example, the probability of throwing a six followed by another six on two rolls of a die is calculated as follows: The outcomes of the two rolls of the die are independent. Let E_1 denote a six on the first roll and E_2 a six on the second roll.

$$P(\text{two sixes}) = P(E_1 \text{ and } E_2) = P(E_1) \times P(E_2) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}.$$

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.

Example: A mugging in California

An elderly woman was assaulted and robbed in an alley in San Pedro, California. A witness saw a blonde woman with a pony–tail running out of the alley and get into a yellow car driven by a black male with a beard and a moustache.

A couple answering that description were arrested nearby and brought to trial. The prosecutor calculated:

$$P(\textsf{blonde}) = \frac{1}{3}, \qquad P(\textsf{pony–tail}) = \frac{1}{10},$$

$$P(\textsf{beard}) = \frac{1}{10}, \qquad P(\textsf{moustache}) = \frac{1}{4},$$

$$P(\textsf{yellow car}) = \frac{1}{10}, \qquad P(\textsf{black male with white female}) = \frac{1}{1000},$$

so that

$$P(\textsf{coincidence}) = \frac{1}{3} \times \frac{1}{10} \times \frac{1}{10} \times \frac{1}{4} \times \frac{1}{10} \times \frac{1}{1000} = \textsf{1 in 12 million.}$$

Not surprisingly, the verdict was guilty. This evidence was challenged on appeal, however, and the verdict reversed. ***Why***? The events might be 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.

Example. 50% of families in a certain city subscribe to the morning newspaper, 65% subscribe to the afternoon newspaper, and 30% of the families subscribe to both newspapers.

What proportion of families subscribe to at least one newspaper?

$$\begin{aligned} P(E_1 \text{ or } E_2) &= P(E_1) + P(E_2) - P(E_1 \text{ and } E_2) \\ &= 0.5 + 0.65 - 0.3 = 0.85. \end{aligned}$$

A more basic version of the rule works where events are ***mutually exclusive***.

If events E_1 and E_2 are mutually exclusive then

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

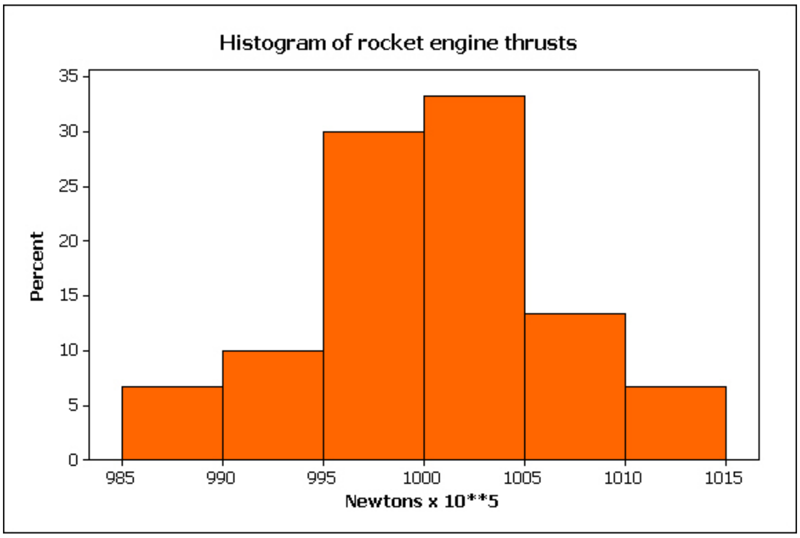
This simplification occurs because when two events are mutually exclusive they cannot happen together and so $P(E_1 \text{ and } E_2) = 0$.

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***.



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

(i) Binomial Distribution

Bernoulli trial: There are only two possible outcomes, namely *success* and failure.

$$P(success) = p.$$

For example: if we toss a coin, $P(\text{'head'}) = 0.5$.

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 $X \sim \mathsf{Bin}(n, p)$, and

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

where

$$\begin{aligned} \binom{n}{x} &= \frac{n!}{x!(n-x)!} \qquad \text{and} \\ A! &= A \times (A-1) \times (A-2) \ldots \times 3 \times 2 \times 1. \end{aligned}$$

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
Suppose that 95% of bathing beaches pass E.U. hygiene regulations. In a random sample of 12 beaches, what is the probability that more than than 9 beaches pass the regulations?

Let

X: ***number of beaches which pass***.

Since we have a fixed number of trials (12), there are two possible outcomes for each trial, and we have a constant probability of 'success' (95% =0.95), we can say that

$$X \;\sim\; \mathsf{Bin}(12, 0.95).$$

We need to calculate

$$\begin{aligned} P(X > 9) &= P(X \geq 10) \\ &= P(X = 10) + P(X = 11) + P(X = 12). \end{aligned}$$

Now

$$\begin{aligned} P(X=10) &= \binom{12}{10} \times 0.95^{10} \times 0.05^{2} \\ &= 66 \times 0.59874 \times 0.0025 \\ &= 0.09879. \end{aligned}$$

Similarly,

$$\begin{aligned} P(X=11) &= \binom{12}{11} \times 0.95^{11} \times 0.05^{1} = 0.3414, \text{ and} \\ P(X=12) &= \binom{12}{12} \times 0.95^{12} \times 0.05^{0} = 0.5404. \end{aligned}$$

Thus,

$$P(X>9) = P(X=10)+P(X=11)+P(X=12) = 0.9805.$$

Conclusion: the possibility that more than 9 beaches can pass the regulations is about 98%.

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
   (a) exactly ten last for more than 500 hours;
   (b) at least ten last for more than 500 hours;
   (c) 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?

(ii)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 $X \sim \mathsf{Poi}(\lambda)$, and

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

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

$$\begin{aligned} \mathsf{mean} &= \lambda; \\ \mathsf{variance} &= \lambda. \end{aligned}$$

Example. 'Jonah' Jones has sailed on tankers during his long sea–going career. Assuming the average incident rate for tankers is 0.231 per ship per year, calculate the probability of more than 1 incident in any year of Jones' career.

Let

X: ***number of incidents (per year)***.

Since we have a constant rate for the number of incidents per year – $\lambda = 0.231$ – and there is no upper limit to the number of incidents which may occur in any given year, we can say that

$$X \sim \mathsf{Poi}(0.231).$$

We need

$$\begin{aligned} P(X > 1) &= P(X = 2) + P(X = 3) + \ldots \\ &= 1 - P(X \leq 1) \\ &= 1 - \{P(X = 0) + P(X = 1)\} \\ &= 1 - \left\{ \frac{e^{-0.231} \times (0.231)^0}{0!} + \frac{e^{-0.231} \times (0.231)^1}{1!} \right\} \\ &= 1 - \{0.7939 + 0.1834\} = 2.3\%. \end{aligned}$$

More examples.

1. Vehicles pass a point on a busy road at an average rate of 420 per hour.

(a) What is the averge number passing in 2 minutes?

(b) Find the probability that non pass in 2 minutes.

(c) Find the probability that at least two vehicles pass in 2 minutes.

2. In the River Wear, a certain bacteria occurs at a rate of 5 per litre.

(a) Find the probability of observing less than 3 bacteria in any one litre jar.

(b) Find the probability of observing *more than* 2 in any one litre jar.

(c) Find the probability of observing exactly 8 bacteria in a two -litre jar.

2.2.2 Continuous distributions

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***.

(i) 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***.
- 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***

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

- 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

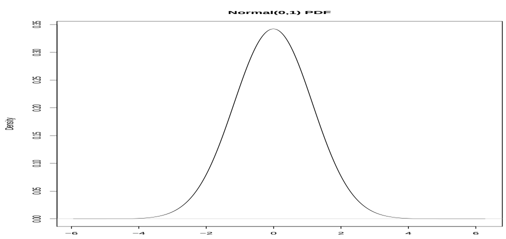
$$f(z) \quad = \quad \frac{1}{\sqrt{2\pi}} \mathsf{exp} \left\{ -\frac{1}{2} z^2 \right\}.$$

- Statistical tables give probabilities for the standard Normal distribution.
- However, 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 three standard deviations of the mean.



Example 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$\times 10^5$.

Let

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

Since we are assuming that thrusts follow a Normal distribution, we have:

$$X \sim N(1000,\ 36).$$

(i) We require $P(X < 990)$. Now we don't have tables for this Normal distribution, but we *do* have tables for the ***standard Normal distribution***.

$$Z = \frac{X - 1000}{6} \sim N(0, 1).$$

So,

$$\begin{aligned} P(X < 990) &= P\left(Z < \frac{990 - 1000}{6}\right) \\ &= P\left(Z < -1.66667\right) \\ &= 0.0446. \end{aligned}$$

(ii) We require $P(X > 1005)$:

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

(iii) We require $P(996 < X < 1002)$. Can you see that this is the same as $P(X < 1002) - P(X < 996)$? Try this one yourself...

More examples

1. The actual diameter (in millimetres) of a rivet with nominal diameter 10 mm is a $N(10, 0.01)$ random variable. To be usable, a rivet must have a diameter in the range 9.8 to 10.2 mm. What proportion of rivets are usable?
2. Yearly peak flows (in $\mathsf{m}^3\mathsf{s}^{-1}$) at a location in Tynedale are assumed to have a Normal distribution with a mean of 4.1 and a standard deviation of 0.9.
   (a) Find the probability that, in any given year, a peak flow of less than 3.2 will be observed.
   (b) Find the probability that, in any given year, the peak flow will lie between 3.8 and 4.2.

(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 interval (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***:

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

Examples.

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

(a) What is the probability that a person waits fewer than 12.5 minutes?

(b) On the average, how long must a person wait?

(c) Ninety percent of the time, the time a person must wait falls below what value?

2. A point D is chosen on the line AB, whose midpoint is C and whose length is b. If X, the distance from D to A, is a random variable having the uniform density $U(0, b)$, what is the probability that AD, BD and AC will form a triangle?