4 Permutations and Combinations

In this section, we introduce some symbols for counting **combinations**.

• The symbol $n!$ represents the product of all integers from 1 to n. In other words, it means that

$$\boxed{n! = n\,(n-1)\,(n-2)\,(n-3) \times 3 \times 2 \times 1.}$$

For simplicity of discussion , we define

$$\boxed{0! = 1.}$$

• It is easy to see that that given a set of n distinct objects, there are $n!$ ways to arrange them in an **order** of sequence.

Because there are n possible objects to be places at Position 1. After placing an object in Position 1, there are $n-1$ possible objects to be placed at Position 2. After it has been placed, there are $n-2$ objects to be placed at Position 3. The argument can continue until the last object being placed at Position n. Therefore we have $n!$ different ways.

• Suppose we are going to choose only r ($r \leq n$) objects from the set of n objects, how many distinct sequences are there?

The answer of the problem is called the number of **permutation** and is denoted by nPr.

Since there are n possible objects to be places at Position 1. After placing an object in Position 1, there are $n-1$ possible objects to be placed at Position 2. After it has been placed, there are $n-2$ objects to be placed at Position 3. The argument can continue until the rth object being placed at Position r. Finally we have

$$\boxed{nPr = \frac{n!}{(n-r)!} = n \times (n-1) \times \cdots \times (n-r+1).}$$

• One may also interest in the number of ways that r elements can be selected from $n(r \leq n)$ elements **independent** of the order. The answer of the question is called the number of **combination** and is denoted by nCr.

Since there are nPr permutations of length r, if the order is ignored then we have to divide nPr by $r!$, because there are $r!$ permutation of a sequence of r distinction objects.

In fact, it is given by the following formula:

$$\boxed{nCr = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}.}$$

Exercise 1. Show that

$$\binom{n}{r} = \binom{n}{n-r}.$$

Exercise 2. Show that

$$r! \times nCr = nPr.$$

Example 1. Three balls are randomly selected from a box which contains four balls with numbers $1, 2, 3$ and 4. How many different possible combinations are there?

There are a total of four balls and we are going to select three of them. Hence, $n = 4$ and $r = 3$. Applying the combination formula, one gets

$$\binom{4}{3} = \frac{4!}{3!\,(4-3)!} = \frac{24}{6 \times 1} = 4.$$

In fact, we can list out all the possible combinations as follows.

$$\{123, 124, 134, 234\}.$$

We note that in the above example, in counting the combinations, the followings are equivalent and will count as one combination only

$$\{123, 132, 213, 231, 312, 321\}.$$

Exercise 3. Let $A = \{1, 2, 3, 4, 5, 6\}$, how many distinct subsets of A are of size two? Recall that $\{1, 2\} \subset A$ and $\{1, 2\} = \{2, 1\}$.

4.1 Stirling's Formula*

This subsection is optional and it aims at providing an approximation for $n!$.

For large value of n, it can be shown that

$$\boxed{n! \approx \sqrt{2n\pi}\left(\frac{n}{e}\right)^n = \sqrt{2\pi e}\left(\frac{n}{e}\right)^{n+\frac{1}{2}}}$$

which is called the **Stirling formula**.

- Thus if n is large then we have

$$2nCn = \binom{2n}{n} = \frac{(2n)!}{n!n!} \approx \frac{\sqrt{4n\pi}\left(\frac{2n}{e}\right)^{2n}}{\sqrt{2n\pi}\left(\frac{n}{e}\right)^n \cdot \sqrt{2n\pi}\left(\frac{n}{e}\right)^n}.$$

Hence we have

$$2nCn \approx \frac{2^{2n}}{\sqrt{\pi n}}.$$

Figure 1: St Augustine and Monica by Ary Scheffer (1846). Taken from Wikipedia.

5 Some History of Probability*

This section is optional and it aims to introduce some story of probability.

With the advent of Christianity, the concept of **random events** developed by philosophers was rejected in the early time. According to **St. Augustine** (354-430), nothing occurred by chance, everything being minutely controlled by the will of God. If events appear to occur at **random**, then it is because of our ignorance and not in the nature of events. One should only seek for the will of God instead of looking at **patterns** of behavior in aggregates of events. [1]

[1] Poker faces: the life and work of professional card players by David M. Hayano, UCP Press, 1982.

The amazing contents and applications of probability theory owes its origin to a question on gambling (**game**).

The question was raised by **Chevalier de Mere** (1607-1684) on his problem of throwing a die. He had a title Chevalier (Knight) and educated at Mere. The problem was solved by **Pascal**. [2]

5.1 Problem 1

De Mere made considerable money over the years in betting **double odds** on rolling at **least one** "**6**" in **4** throws of a fair die (six faces).

He then thought that the same should occur for betting on at **least one double-six in 24 throws of two fair dice** (This is their ancient believes). It turned out that it did not work well.

Why? In 1654, he challenged his friends Pascal and Fermat for the reasons.

[2] Religion reformation (1517-1648) and Enlightenment age (1637-1789).

 

Figure 2: Pascal (1623-1662) (Left). Fermat (1601-1665) (Right). Taken from Wikipedia.

The probability of getting no "6" in four **independent** throws of a fair die:

$$\boxed{(5/6) \times (5/6) \times (5/6) \times (5/6) = \frac{625}{1296}.}$$

Therefore the probability of having **at least one** "6" in 4 throws will be equal to

$$\boxed{1 - \frac{625}{1296} = \frac{671}{1296} = 0.5177 > 0.5000.}$$

This explained why de Mere got a good amount of money on double odds on his bet.

This is not a fair game, the player has advantage over the house.

The probability of getting no double "6" in the throw of two fair dice is

$$\boxed{1 - (\frac{1}{6} \times \frac{1}{6}) = \frac{35}{36}.}$$

The probability of getting no double "6" in "24" **independent**[3] throws is

$$(\frac{35}{36})^{24}.$$

Therefore the probability of having at least one double 6 in 24 throws is equal to

$$\boxed{1 - (\frac{35}{36})^{24} = 0.4914 < 0.5.}$$

This explained why de Mere did not get a good amount of money on double odds on this bet.

Again this is not a fair game, the house has advantage over the player.

[3]We shall define and discuss this concept later

5.2 Problem 2

• Two Players A and B are playing a series of games which requires to score 5 points (games) in order to win. In each game there is no draw.

At the moment that Player A is leading 4 points to 3 points, the game was interrupted and cannot continue. How should the players divide the stakes on the unfinished game?

8th Point	9th Point	Final Winner
A Wins	A Wins	Player A
A Wins	B Wins	Player A
B Wins	A Wins	Player A
B Wins	B Wins	Player B

Assume all the 4 outcomes are equal likely then the stake should be divided by the ratio 1:3 (B:A).

Exercise 4. Suggest another method to divide the stake.

5.3 Further Development

C. Huygens (a teacher of Leibniz), learned of the incident. Later in 1657, he published the first book on probability; entitled "De Ratiociniis in Ludo Aleae", it was a study on problems related to gambling. Probability soon became popular, and developed rapidly during the 18th century and the major contributors were **J. Bernoulli** (1654-1705) and **A. de Moivre** (1667-1754).

In 1812 **P. Laplace** (1749-1827) introduced new ideas and mathematical techniques in his book, Theorie Analytique des Probabilites. He applied probabilistic ideas to many scientific and practical problems such as mathematical statistics, actuarial mathematics and statistical mechanics etc. Many workers then contributed to the theory, e.g. **Chebyshev**, **Markov** and **Kolmogorov**. [4]

[4] Taken from Calculus, Volume II by Tom M. Apostol (2nd edition, John Wiley & Sons, 1969)

One of the main difficulties in "mathematicalizing" probability was its definition. The search for a generally acceptable definition took nearly 300 years.

It was finally resolved in the 20th century by treating probability theory on an **axiomatic basis**. In 1933 a monograph by a Russian mathematician **A. Kolmogorov** gave an axiomatic approach that forms the basis for the modern theory.

Kolmogorov's monograph is available in English translation:

Foundations of Probability Theory, Chelsea, New York, 1950.

Since then the ideas have been refined somewhat and probability theory is now part of a more general discipline known as **measure theory**.

6 Samples Spaces and Events

This section discusses the elementary concepts of **probability**. In our daily activities, we always experience the sense of uncertainty. Take for a simple example, when we toss a fair coin, we may have a Head (H) or a Tail (T). One can never tell exactly the outcome. If you throw a fair dice you may get $1, 2, 3, 4, 5$ or 6 dots. In fact, each of the above actions is called an **experiment** and each of the possible outcomes of the experiments is called an **event**.

An experiment is defined to be any process which generates well defined **outcomes**. This means that on any single repetition of the experiment, **one and only one** of the possible experimental outcomes will occur.

A **sample space** is defined as the set of all possible experimental outcomes. Any particular outcome is referred to as a **sample point** and is called an **element** of the sample space. Using the language of set, the sample space of tossing a coin is $\{H, T\}$ and the sample space of throwing a dice is $\{1, 2, 3, 4, 5, 6\}$.

Here let us introduce the concept of an *event*. An event is a collection of one or more of the outcomes of an experiment. An event that includes one and only one of the (final) outcomes for an experiment is called a **simple event** and is usually denoted by E_i. A **compound event** is a collection of **more than one outcome** for an experiment. With the concepts of "outcome" and "event", we are going to define the "probability of an event".

Probability is a numerical measure of the **likelihood** that a specific event will occur. Let E_i be a simple event and A be a compound event. The following are two important properties of probability. The probability of an event always lies in the range from zero to one, i.e.

$$0 \leq P(E_i) \leq 1 \quad \text{for all } i, \quad \text{and} \quad 0 \leq P(A) \leq 1.$$

Furthermore, the sum of the probabilities of all simple events (or final outcomes) for an experiment, denoted by $\sum P(E_i)$ is always equal to one, i.e.

$$\boxed{P(E_1) + P(E_2) + \ldots + P(E_n) = \sum_{i=1}^{n} P(E_i) = 1.}$$

7 Basic Probability Rules

Strictly speaking, probability is a **set function** P that assigns to each event A in the **sample space** S a number $P(A)$, called **the probability of the event** A, such that the following properties (**Axioms**)[5] are satisfied:

(i) $0 \leq P(A) \leq 1$,

(ii) $P(S) = 1$,

(iii) If $A_1, A_2, A_3, \ldots$ are events such that $A_i \cap A_j = \phi$ for $i \neq j$, then

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots.$$

Property (iii) implies that for any **finite collection** of events $A_1, \ldots, A_n$, we have

$$P(A_1 \cup A_2 \cup A_3 \ldots \cup A_n) = P(A_1) + P(A_2) + P(A_3) + \ldots + P(A_n)$$

provided that the events are **mutually exclusive**, i.e.,

$$\boxed{A_i \cap A_j = \phi \text{ if } i \neq j.}$$

The following propositions give some important properties of the probability set function.

[5] (Taken from Wikipedia) An axiom is a premise or starting point of reasoning. As classically conceived, an axiom is a premise so evident as to be accepted as true without controversy.

The first proposition tells that the probability sum of an event and its complement A' must be one.

Proposition 1. *For any event A,* $\boxed{P(A) = 1 - P(A').}$

Proof. Since $S = A \cup A'$ and $A \cap A' = \phi$, it follows that

$$1 = P(S) = P(A \cup A') = P(A) + P(A').$$

Hence $P(A) = 1 - P(A')$. □

Proposition 2. *If ϕ is the empty set, then* $\boxed{P(\phi) = 0.}$

Proof. Since $A = A \cup \phi$ and $A \cap \phi = \phi$, we have

$$P(A) = P(A \cup \phi) = P(A) + P(\phi).$$

Hence $P(\phi) = 0$. □

Proposition 3. *If events A and B are such that $A \subseteq B$, then $P(A) \leq P(B)$.*

Proof. We note that (i) $B = A \cup (B \cap A')$[6] and (ii) A and $B \cap A'$ are **mutually exclusive** because $A \cap (B \cap A') = \phi$. Therefore we have

$$P(B) = P(A) + P(B \cap A') \geq P(A)$$

because $P(B \cap A') \geq 0$. □

[6]You may verify this by using the Venn diagram. Or since for any sets X, Y, Z, we have $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$. Let $X = A, Y = B, Z = A'$, we have $R.H.S. = (A \cup B) \cap (A \cup A') = (A \cup B) \cap S = B \cap S = B$.

Proposition 4. *If A and B are any two events, then*

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B).}$$

Proof. The event $A \cup B$ can be represented as a union of **disjoint sets**, that is,

$$A \cup B = A \cup (A' \cap B).^{7}$$

Hence

$$P(A \cup B) = P(A) + P(A' \cap B). \tag{7.1}$$

Similarly, we can express B in terms of union of disjoint sets, that is,

$$B = (A \cap B) \cup (A' \cap B).^{8}$$

Thus

$$P(A' \cap B) = P(B) - P(A \cap B). \tag{7.2}$$

Combining Equations (7.1) and (7.2), we get the result. □

Exercise 5. If A, B, C are any three events, show that

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

---

[7] You may verify this by using the Venn diagram. Or since for any sets X, Y, Z, we have $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$. Let $X = A, Y = A', Z = B$, we have $R.H.S. = (A \cup A') \cap (A \cup B) = S \cap (A \cup B) = A \cup B$.

[8] You may verify this by using the Venn diagram. Or since for any sets X, Y, Z, we have $X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$. Let $X = B, Y = A, Z = A'$, we have $R.H.S. = (B \cap A) \cup (B \cap A') = (A \cup A') \cap B = S \cap B = B$.

8 Conditional Probability, Independent Events and Bayes' Theorem

In this section, we are going to introduce the concepts of **conditional probability**.

• In some situations, when extra information about an event is known then the probability of the occurrence of the event will be different.

• When a dice is thrown and you are asked to guess the result. Then the probability that you can get a correct answer is of course $1/6$.

But if you are told that the number of dots got is an even number, then you know that the result should be $2, 4$ or 6 and $1, 3$ and 5 should be excluded.

Hence your chance of getting a correct answer increases from $1/6$ to $1/3$.

If A and B are two events, then the conditional probability of A is written as $P(A|B)$ (read as "the probability of A given B has already occurred") and defined as

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}.}$$

Proposition 5. *Let A, B be two events of a sample space S. Then*

(i) $0 \leq P(B|A) \leq 1$.

(ii) $P(S|A) = 1$.

(iii) $P(B_1 \cup \ldots \cup B_n|A) = P(B_1|A) + \ldots + P(B_n|A)$ *if* $B_i \cap B_j = \phi$ *for* $i \neq j$.

Proof. **(i)** It is easy to see that the following inequalities hold.

$$0 \leq P(B|A) = \frac{P(A \cap B)}{P(A)} \leq \frac{P(A)}{P(A)} = 1.$$

(ii) By definition we have the following

$$P(S|A) = \frac{P(S \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1.$$

(iii) We note that

$$P(B_1 \cup B_2 \cup \ldots \cup B_n|A) = \frac{P((\cup_{i=1}^{n} B_i) \cap A)}{P(A)} = \frac{P(\cup_{i=1}^{n}(B_i \cap A))}{P(A)}.^{9}$$

Since $(A \cap B_i) \cap (A \cap B_j) = A \cap B_i \cap B_j = A \cap \phi = \phi$, we have

$$P(B_1 \cup B_2 \cup \ldots \cup B_n|A) = \frac{P(B_1 \cap A)}{P(A)} + \frac{P(B_2 \cap A)}{P(A)} + \ldots + \frac{P(B_n \cap A)}{P(A)}.$$

By definition, we have

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} \quad \text{for } i = 1, 2, \ldots, n$$

and therefore

$$P(B_1 \cup B_2 \cup \ldots \cup B_n|A) = P(B_1|A) + P(B_2|A) + \ldots + P(B_n|A).$$

□

[9] Recall that $A \cap (B_1 \cup B_2 \cup \cdots \cup B_n) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n)$.

Proposition 6. *(The Bayes' Theorem) Suppose that* $B_1, B_2, \ldots, B_n$ *are* n *mutually exclusive events and the union of these events is the entire sample space* S*. Then*

$$\boxed{P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(B_1)P(A|B_1) + \ldots + P(B_n)P(A|B_n)}.}$$

Proof. Since

$$A = A \cap S = A \cap (\cup_{i=1}^{n} B_i)$$

and

$$A \cap (\cup_{i=1}^{n} B_i) = \cup_{i=1}^{n} (A \cap B_i),$$

we have

$$P(A) = P(\cup_{i=1}^{n}(A \cap B_i)) =^{10} \sum_{i=1}^{n} P(A \cap B_i) =^{11} \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

Hence

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}.$$

□

[10]Since $(A \cap B_i) \cap (A \cap B_j) = \phi$ for $i \neq j$, we may apply the third axiom.
[11]By definition, we have $P(A \cap B_i) = P(A|B_i)P(B_i)$.

Finally let us define the concept of **independence** of two events.

Definition 1. Two events A and B are **independent** if

$$\boxed{P(A \cap B) = P(A)P(B).}$$

Remark 1. If A and B are independent then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

and

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = P(B).$$

This means that the outcome of event A will have no effect on the outcome of event B and vice versa.

Example 2. In a manufacturing plant, there are three machines, B_1, B_2 and B_3, producing 50%, 30% and 20%, respectively, of the products. It is known from the past experience that 1%, 2% and 3% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective? If a product was chosen randomly and was found to be defective, what is the probability that it was made by machine B_3?

This is an application of probability model to quality control and machine maintenance. It will cover the first two learning outcomes. [12]

Let us define the following events.

A is the event that the product is defective;
B_1 is the event that the product was made by machine B_1;
B_2 is the event that the product was made by machine B_2;
B_3 is the event that the product was made by machine B_3.

[12] (1) Demonstrate knowledge and understanding of the essential engineering mathematics as well as their relationship to the engineering problems in general. (2) Model an engineering problem into a mathematical form or a mathematical model, which can be an algebraic equation, a differential equation, a graph, or some other mathematical expression.

We have

$$P(B_1) = 0.5, \quad P(B_2) = 0.30, \quad P(B_3) = 0.20$$

and

$$P(A|B_1) = 0.01, \quad P(A|B_2) = 0.02, \quad P(A|B_3) = 0.03$$

Therefore we have

$$\begin{aligned} P(A) &= P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) \\ &= (0.5)(0.01) + (0.3)(0.02) + (0.2)(0.03) \\ &= 0.005 + 0.006 + 0.006 \\ &= 0.017. \end{aligned}$$

Using Bayes' theorem, we write

$$\begin{aligned} P(B_3|A) &= \frac{P(B_3)P(A|B_3)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)} \\ &= \frac{(0.2)(0.03)}{(0.5)(0.01) + (0.3)(0.02) + (0.2)(0.03)} \\ &= \frac{0.006}{0.017} \\ &= \frac{6}{17}. \end{aligned}$$

9 Discrete and Continuous Random Variables

Many problems have "**uncertainty**" in nature. An event is uncertain means that it is not deterministic, you can never know the result before it happens. For example, when we toss a fair coin, the result can be a head or a tail. We can never know the result until the coin is tossed. The concept of "probability distribution" is a useful tool for describing and modeling these uncertain events.

• Probability also helps us in solving many other practical problems. In this section, we are going to introduce the concept of a **random variable** and **probability distribution** or **probability density function** (p.d.f.). They are useful tools to describe uncertain events.

• There are two types of probability distributions: the **discrete** type and the **continuous** type. For a probability distribution, no matter it is discrete or continuous, in this course, we will pay a lot of attention to its **mean** and its **variance**.

The mean of a probability distribution describes the average behavior of the uncertainty.

The variance measures how the "**uncertainty**" deviate from the mean.

A **random variable** is a variable whose value is determined by the outcome of a random experiment. In other words, its value is not deterministic. There are two types of random variables, the **discrete random variable** and the **continuous random variable**.

A random variable that assumes **countable values** (for example, the sample space $S \subseteq \mathbb{Z}$) is called a discrete random variable.

A random variable that can assume any value contained in one or more intervals is called a **continuous random variable**.

Table 1: The credit cards

Number of cards	Owned frequency	Relative frequency (Probabilities)
0	50	$50/1000 = 0.05$
1	550	$550/1000 = 0.55$
2	250	$250/1000 = 0.25$
3	100	$100/1000 = 0.10$
4	50	$50/1000 = 0.05$
	$N = 1000$	Sum $= 1.00$

Example 3. Table 1 gives the frequency and relative frequency distributions of the **credit cards** owned by all 1000 families living in a certain area. Suppose one family out of the 1000 is randomly selected. The act of randomly selecting a family is called a **random experiment**.

Let X denotes the number of credit cards owned by the selected family. Then X can be any one of the five possible values $(0, 1, 2, 3$ and $4)$ listed in the first column of the Table 1.

The value of X depends on which family is selected. Thus, this value depends on the **outcome** of a **random experiment**.

The **probability distribution** or **probability density function** (p.d.f.) of a discrete **random variable** (r.v.) lists all the possible values that the random variable can assume and their corresponding probabilities.

We note that the probability distribution of a **discrete random variable** has the following two characteristics:

(i) $0 \leq p(x) \leq 1$ for each x and

(ii) $\sum_x p(x) = 1$.

Moreover, the function

$$\boxed{F(t) = \sum_{x=-\infty}^{t} p(x)}$$

is called the **cumulative probability distribution** of X. Moreover,

$$\boxed{P(X \geq a) = \sum_{x=a}^{\infty} p(x) \quad \text{and} \quad P(X \leq b) = \sum_{x=-\infty}^{b} p(x).}$$

Table 2: The probability distribution (incomplete) of errors

Number of errors	0	1	2	3	4	5	more than 5
Probability	0.01	0.05	0.04	0.20	0.30	0.30	0.10

Example 4. Table 2 lists the probability distribution of the number of reading errors of a hard disk per 10^{10} bits of data based on the past historical data. Find the probability that the number reading errors for the hard disk per 10^{10} bits of data is (i) one to two (ii) more than three.
Let x denotes the number of errors for this machine during a given week. We can calculate the required probabilities as follows.
(i) The probability of one to two reading errors is given by the sum of the probabilities of 1 and 2 reading errors.

$$P(\text{1 to 2 reading errors}) \;=\; P(1 \le X \le 2) = P(X = 1) + P(X = 2) = 0.09.$$

(ii) The probability of more than three reading errors is obtained by adding the probabilities of $4, 5$ and more than 5 reading errors.

$$\begin{aligned} P(\text{more than three errors}) &= P(X > 3) \\ &= P(X = 4) + P(X = 5) + P(X > 5) \\ &= 0.30 + 0.30 + 0.10 = 0.70. \end{aligned}$$

In the continuous case, the random variables can assume values on a continuous scale. A random variable X is said to be a continuous random variable if there exists a function $p(x)$, the probability density function satisfies
(i) $p(x) \geq 0$ and;
(ii) $\int_{-\infty}^{\infty} p(x)dx = 1$. (summation is replaced by integration)
Moreover, the function

$$\boxed{F(t) = \int_{-\infty}^{t} p(x)dx}$$

is called the **cumulative probability distribution** of X.

One can compare this with the probability distribution in the discrete case. The probability $P(a \leq X \leq b)$ is defined as

$$\boxed{P(a \leq X \leq b) = \int_{a}^{b} p(x)dx.}$$

Similar to the discrete case we have

$$\boxed{P(X \geq a) = \int_{a}^{\infty} p(x)dx \quad \text{and} \quad P(X \leq b) = \int_{-\infty}^{b} p(x)dx.}$$

Example 5. Consider the function

$$p(x) = \begin{cases} 3x^2 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Determine if it is a probability distribution. If "yes" find $P(X \geq 0.5)$. [13]

We note that $p(x) \geq 0$ for any $x \in [0, 1]$. Secondly, we see that

$$\int_{-\infty}^{\infty} p(x)dx = \int_{0}^{1} 3x^2 dx = 1.$$

Therefore $p(x)$ is a probability distribution. Now

$$P(X \geq 0.5) = \int_{0.5}^{\infty} p(x)dx = \int_{0.5}^{1} 3x^2 dx = 1 - 0.5^3 = \frac{7}{8}.$$

[13] Different from the case of discrete probability distribution, here we have $P(X \geq 0.5) = P(X > 0.5)$.

Example 6. The **normal distribution** has the form

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.}$$

Given $\int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} dx = \sqrt{2\pi}$, show that $f(x)$ is a probability density function.

We note that $f(x)$ is non-negative and we need to show that [14]

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

We consider the substitution

$$y = \frac{x-\mu}{\sigma} \quad \text{then} \quad \frac{dy}{dx} = \frac{1}{\sigma}.$$

The upper and lower limits for y are ∞ and $-\infty$ respectively. Then

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-y^2}{2}} \sigma dy = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy = 1.$$

This is an example of solving a problem in normal distribution model by a mathematical technique. It will cover the third learning outcome. [15]

[14] An exercise in Tutorial 1

[15] (3) Solve the model by selecting and applying a suitable mathematical method, skill or technique learned.

Exercise 6. Suppose the demand of certain new product follows the **uniform distribution** on $[a, b]$ (where $a < b$).

The probability density function takes the form:

$$f(x) = \begin{cases} K & a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases}$$

Here K is a positive constant to be determined.

(a) Show that if $K = \frac{1}{b-a}$ then $f(x)$ is a probability density function.
(b) Find the probability that the demand of a new product lies in $[a, (b + 2a)/3]$.

10 The Expectation of a Random Variable

Let X be a random variable with probability distribution $p(x)$.

The mean or the expected value is defined as

$$\boxed{\mu = E(X) = \sum_x xp(x) \quad \text{if } x \text{ is discrete,}}$$

and

$$\boxed{\mu = E(X) = \int_{-\infty}^{\infty} xp(x)dx \quad \text{if } x \text{ is continuous.}}$$

The meaning of $E(x)$ is the **long-run average** of the random variable X.

Example 7. Find the mean number of heads obtained in tossing a fair coin twice.

We multiply each value of x by its probability and add these products. This sum gives the mean of the probability distribution of X where X is the number of heads obtained in tossing the coin twice.

$p(X=0)=(\frac{1}{2})^2=0.25,\ p(X=1)=(\frac{1}{2}\times\frac{1}{2})+(\frac{1}{2}\times\frac{1}{2})=0.50,\ p(X=2)=(\frac{1}{2})^2=0.25.$

The mean, denoted by μ is

$$E(X)=\mu=\sum_{x=0}^{2} xp(x)=0\times 0.25+1\times 0.5+2\times 0.25=1.$$

When the experiment of tossing the fair coin twice is performed many times, then in certain occasion we will observe no head; for sometimes we will observe 1 head; and for sometimes we will observe 2 heads.

The mean number of heads obtained in the experiments will tend to 1.

Let us look at an example of continuous random variable.

Example 8. Let X be the random variable that denotes the lifetime in hours of a certain light bulb. We assume it follows the **exponential distribution:** $f(t) = \lambda e^{-\lambda t}$ for $t \geq 0$ with $\lambda = 1/1000$. Then the probability density function is given by

$$f(x) = \begin{cases} \frac{1}{1000} e^{-\frac{x}{1000}} & x \geq 0 \\ 0 & \text{elsewhere.} \end{cases}$$

Find the expected lifetime of this type of light bulb.

By definition and apply integration by parts,

$$\begin{aligned} \mu &= E(X) = \int_0^\infty \frac{x}{1000} e^{-x/1000} dx \\ &= -\int_0^\infty x d(e^{-x/1000}) \\ &= \boxed{-xe^{-x/1000}\Big|_0^\infty} + \int_0^\infty e^{-x/1000} dx = 1000. \end{aligned}$$

[16]

[16] $\lim_{x \to \infty} \frac{x}{e^{x/1000}} = 0.$

The concept of **expectation** $E(X)$ can be generalized to the case of general expectation of functions $E(g(X))$.

Let x be a random variable with probability distribution $p(x)$.

The mean or expected value of the random variable $g(x)$ is defined as

$$\boxed{\mu_{g(X)} = E\left[g(X)\right] = \sum_x g(x)p(x) \quad \text{if } X \text{ is discrete,}}$$

and

$$\boxed{\mu_{g(X)} = E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)p(x)dx \quad \text{if } X \text{ is continuous.}}$$

We remark that if X and Y are two independent randoms variables then we have

$$\boxed{E(X+Y) = E(X) + E(Y)}^{17}$$

and

$$\boxed{E(XY) = E(X)E(Y).}^{18}$$

---

[17] If X follows $p(x)$ and Y follows $q(y)$ then $E(X+Y) = \sum_x \sum_y (x+y)p(x)q(y) = \sum_x \sum_y xp(x)q(y) + yp(x)q(y) = \sum_x \sum_y xp(x)q(y) + \sum_x \sum_y yp(x)q(y) = \sum_x xp(x) + \sum_x p(x)E(Y) = E(X) + E(Y)$.

[18] If X follows $p(x)$ and Y follows $q(y)$ then $E(XY) = \sum_x \sum_y xyp(x)q(y) = \sum_x xp(x) \sum_y yq(y) = E(X)E(Y)$.

Table 3: The probability distribution

x	0	1	2	3	4	5
$p(x)$	0.1	0.1	0.2	0.2	0.3	0.1

Example 9. Consider a 24-hour car park with five parking space. Let X be the random variable representing the number of cars parking per hour and it has the probability distribution given in Table 3. Suppose the parking fee is \$20 each car per hour and \$5 per hour is paid to the operator of the car park. Find the expected earnings of the car park (per hour). The profit function per hour is

$$\boxed{g(X) = 20X - 5}$$

which represents the amount of money in dollars, paid to the car park manager. Thus $E(g(X))$ is the expected earnings per hour.

$$\begin{aligned} E\left[g(X)\right] &= E\left[20X - 5\right] = \sum_x (20x - 5)\, p(x) \\ &= (2 \times 0 - 5)\frac{1}{10} + (20 \times 1 - 5)\frac{1}{10} + (20 \times 2 - 5)\frac{1}{5} \\ &\quad + (20 \times 3 - 5)\frac{1}{5} + (20 \times 4 - 5)\frac{3}{10} + (20 \times 5 - 5)\frac{1}{10} = 51. \end{aligned}$$

We see that the average daily earnings of the car park is $24 \times \$51 = \$1224$.

Example 10. Let X be the random variable representing the duration of an distant call in minutes. Here X is a random variable follows the following probability density function

$$f(x) = \begin{cases} a\sqrt{x} & 0 \le x \le 100 \\ 0 & \text{elsewhere,} \end{cases}$$

where a is a constant. Suppose the cost per minute of a call is \$5 and there is a \$10 service charge. Find the expected cost $(5X + 10)$.

We first find the constant a. Since

$$1 = \int_0^{100} a\sqrt{x}dx = \frac{2ax^{3/2}}{3}\bigg|_0^{100} = \frac{2000a}{3}.$$

Thus $a = 3/2000$.

$$E(5X + 10) = \int_0^{100} 3(5x + 10)\frac{\sqrt{x}}{2000}dx = 310.$$

The above two examples demonstrate an interrelation among mathematical theory, result and a operation management problem and therefore the fourth learning outcome. [19]

[19] (4) Have a general grasp on the interrelation among mathematical theory, result and the engineering problem.

Example 11. Let X be a discrete random variable taking values in $\{0, 1, 2, \ldots, \}$ and having probability distribution $p_i = P(x = i)$. We observe that

$$\begin{aligned} E(X) &= 0p_0 + 1p_1 + 2p_2 + 3p_3 + 4p_4 + \ldots + \\ &= (p_1 + p_2 + p_3 + p_4 + \ldots) + (p_2 + p_3 + p_4 + \ldots) + (p_3 + p_4 + \ldots) + \ldots + \\ &= G(1) + G(2) + G(3) + \ldots \end{aligned}$$

where

$$\boxed{G(x) = 1 - F(x-1) = 1 - \sum_{i=0}^{x-1} p_i}$$

and $F(x)$ is the cumulative probability distribution. Thus we have

$$\boxed{E(X) = \sum_{i=0}^{\infty} (1 - F(i)).}$$

11 The Variance of a Random Variable

Let X be a random variable with probability distribution $p(x)$ and mean μ. The **variance** of X, denoted by $\mathrm{Var}(X)$, is defined as

$$\boxed{\sigma^2 = E\left[(X-\mu)^2\right] = \sum_x (x-\mu)^2\, p(x) \quad \text{if } X \text{ is discrete}}$$

and

$$\boxed{\sigma^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx \quad \text{if } X \text{ is continuous.}}$$

The variance is a measure of how "**dispersive**" the random variable from the mean is.

Suppose that the variance is very close to zero, then the probability of getting a data drawn from the distribution close to the mean μ is very large.

The square root of variance is called the **standard deviation**.

Table 4: The probability distribution

x	0	1	2	3	4	5
$p(x)$	0.1	0.1	0.2	0.2	0.3	0.1

Table 5: The values of $(x-\mu)^2 p(X=x)$

x	0	1	2	3	4	5
$(x-\mu)^2 p(x)$	0.784	0.324	0.128	0.008	0.432	0.484

Let us calculate the variance of the distribution in the car park example.

Example 12. We recall that the density function of the car park example is given in Table 4. To find $Var(X)$, we first find $E(X)$.

$$\mu = E(X) = 0 \cdot \frac{1}{10} + 1 \cdot \frac{1}{10} + 2 \cdot \frac{1}{5} + 3 \cdot \frac{1}{5} + 4 \cdot \frac{3}{10} + 5 \cdot \frac{1}{10} = 2.8.$$

Table 5 gives $(x-\mu)^2 P(X=x)$ for $x = 0, 1, 2, 3, 4, 5$. $Var(X)$ is then given by

$$0.784 + 0.324 + 0.128 + 0.008 + 0.432 + 0.484 = 2.52.$$

In the following, we are going to introduce a useful formula for relating variance and mean and therefore computation of variance.

Proposition 7. *The variance of a random x is given by*

$$\boxed{\sigma^2 = E\left(X^2\right) - \mu^2 = E(X^2) - E(X)^2.}$$

Proof. Here we prove the formula for discrete random variable case. For the case of continuous variable the proof is similar. In the discrete case, one can write

$$\begin{aligned}\sigma^2 &= \sum_x (x-\mu)^2 f(x) = \sum_x \left(x^2 - 2x\mu + \mu^2\right) f(x) \\ &= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x).\end{aligned}$$

Since by definition

$$\mu = E(X) = \sum_x x f(x) \quad \text{and} \quad \sum_x f(x) = 1$$

for any discrete probability distribution, it follows that

$$\sigma^2 = \sum_x x^2 f(x) - \mu^2 = E\left(X^2\right) - \mu^2.$$

□

In the following we relate the variance of a stock price with its risk.

Example 13. Consider two stocks (A) Telecom and (B) Bank of East Asia. Suppose the yearly return in billions of these two stocks A and B are given by

$$p_A(x) = \frac{3}{x^4}, \quad 1 \leq x$$

and

$$p_B(x) = 0.5e^{-0.5x}, \quad 0 \leq x$$

respectively.

One can show $\mu_B = 2$ and $\mu_A = 1.5$. Furthermore, we have

$$E(X_B^2) = 8 \quad \text{and} \quad E(X_A^2) = 3.$$

Hence we get

$$\sigma_B^2 = 8 - 2^2 = 4 \quad \text{and} \quad \sigma_A^2 = 3 - 1.5^2 = 0.75.$$

Thus we conclude the followings:
(i) The expected return of stock A is less than stock B.
(ii) The variance of return of stock A is less than stock B.
Stock B is more risky than stock A though that it has a higher expected return.

Similar to the case of Mean, the concept of Variance can be further generalized as follows.

Let X be a random variable with probability distribution $p(x)$. The variance of the random variable $g(x)$ is

$$\boxed{\sigma^2_{g(X)} = E\left[\left(g(X) - \mu_{g(X)}\right)^2\right] = \sum_x \left(g(x) - \mu_{g(x)}\right)^2 p(x)}$$

if X is discrete and

$$\boxed{\sigma^2_{g(X)} = E\left[\left(g(X) - \mu_{g(X)}\right)^2\right] = \int_{-\infty}^{\infty} \left(g(x) - \mu_{g(x)}\right)^2 p(x)dx}$$

if X is continuous.

12 Means and Variances of Linear Combinations of Random Variables

In this section, we will discuss several formula for the calculation of mean and variance. The formula are presented as propositions and proved for the case of discrete random variable. For the continuous case, the proofs are similar so we leave them as exercises.

Very often, we know the value $E(X)$ for a distribution $p(x)$. But we want to find for instance $E(aX + b)$ where a and b are constants. Of course we can find it by the definition that

$$\boxed{E(aX + b) = \sum_x (ax + b)p(x).}$$

Is there any shortcut if we know $\mu = E(X)$?

Yes. The following proposition provides an answer.

Proposition 8. *If a and b are* **constants**, *then*

$$\boxed{E(aX+b) = aE(X)+b.}$$

Proof.

$$\begin{aligned} E(aX+b) &= \sum_x (ax+b)p(x) \\ &= \sum_x axp(x) + \sum_x bp(x) \\ &= a\sum_x xp(x) + b\sum_x p(x) \\ &= aE(X)+b \end{aligned}$$

because

$$E(X) = \sum_x xp(x) \quad \text{and} \quad \sum_x p(x) = 1.$$

□

How about the case for $\mathrm{Var}(aX+b)$?

In fact, we have similar result as follows.

Proposition 9. *If a and b are constants, then*

$$\boxed{\sigma^2_{aX+b} = a^2\sigma^2.}$$

Proof. By definition,

$$\sigma^2_{ax+b} = E\left[((aX+b) - \mu_{aX+b})^2\right].$$

Since

$$\mu_{aX+b} = E\left(aX+b\right) = a\mu + b.$$

we have,

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

□

Let us apply the above propositions to solving the following problem.

Example 14. Let x be a random variable whose probability density function is the **uniform distribution** $p(x)$:

$$p(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find $E(2X+1)$ and $Var(2X+1)$. First of all, we have

$$E(X) = \int_0^1 x dx = \frac{1}{2}$$

and

$$E(X^2) = \int_0^1 x^2 dx = \frac{1}{3}.$$

Thus

$$Var(X) = E(X^2) - E(X)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Using the above propositions we have

$$E(2X+1) = 2E(X) + 1 = 1 + 1 = 2$$

and

$$Var(2X+1) = 2^2\text{Var}(X) = \frac{4}{12} = \frac{1}{3}.$$

Proposition 10. *If* X_1 *and* X_2 *are two independent random variables having the same probability density function whose variance is* σ^2 *then*

$$\boxed{\sigma^2_{X_1+X_2} = 2\sigma^2.}$$

Proof. By definition,

$$\sigma^2_{X_1+X_2} = E\left[\left((X_1 + X_2) - \mu_{X_1+X_2}\right)^2\right].$$

Since

$$\mu_{X_1+X_2} = E\left(X_1 + X_2\right) = \mu + \mu.$$

we have

$$\begin{aligned} \sigma^2_{X_1+X_2} &= E\left[\left((X_1 + X_2) - (\mu + \mu)\right)^2\right] \\ &= E\left[(X_1 - \mu)^2\right] + E\left[(X_2 - \mu)^2\right] + 2E\left[(X_1 - \mu)\left(X_2 - \mu\right)\right] \\ &= E\left[(X_1 - \mu)^2\right] + E\left[(X_2 - \mu)^2\right] + 0 = 2\sigma^2. \end{aligned}$$

□

Remark 2. We recall that if X and Y are independent random variables then $E(XY) = E(X)E(Y)$. Thus $E(X_1 - \mu)(X_2 - \mu) = E(X_1X_2) + \mu^2 - \mu(E(X_1) + E(X_2)) = E(X_1)E(X_2) + \mu^2 - 2\mu^2 = 2\mu^2 - 2\mu^2 = 0$.

Exercise 7. Let $X_1, X_2, \ldots, X_n$ be n independent random variables sharing the same probability distribution with mean μ and variance σ^2.

Let

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

What are the values of

$$E(\bar{X}) \quad \text{and} \quad Var(\bar{X})?$$

What will happen when $n \to \infty$?

Please interpret the result.

12.1 Results in Large Number of Independent Observations*

This section is optional and we state without proof of an important result.

Proposition 11. *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identical distributed random variables having mean μ and finite variance σ^2. Then we have*

$$\boxed{P\left(\lim_{n\to\infty} \frac{X_1 + X_2 + \ldots + X_n}{n} = \mu\right) = 1.}$$

This is the famous **Strong Law of Large Numbers***.*

Summary

1. Combination
2. Permutation
3. Sample space
4. Random variable
5. Conditional probability and independent events
6. Probability density function (p.d.f.)
7. Cumulative probability distribution
8. Uniform distribution
9. Exponential distribution
10. Normal distribution
11. Expected value $E(X)$
12. Variance $Var(X)$
13. $Var(X) = E(X^2) - E(X)^2$