Chapter 1

Elements of Probability Distribution Theory

1.1 Introductory Definitions

Statistics gives us methods to make inference about a population based on a random sample representing this population. For example, in clinical trials a new drug is applied to a group of patients who suffer from a disease, but we draw conclusions about the drug’s effect on any person suffering from that disease. In agricultural experiments we may be comparing new varieties of wheat and observe a sample of yield or resistance to some infections. These samples will represent a population of all possible values of the observed variables. Randomness is an important element of such experiments and here we start with some fundamental definitions of probability theory.

Definition 1.1. The set, $\Omega$, of all possible outcomes of a particular experiment is called the sample space of the experiment.

Example 1.1. Dice and reaction time.

(a) The experiment consists of throwing a fair die and observing the number of dots on top face. Then, the sample space is

$$\Omega = \{1, 2, 3, 4, 5, 6\}, \text{ which is countable.}$$

(b) The experiment involves observation of reaction time to a certain stimulus
CHAPTER 1. ELEMENTS OF PROBABILITY DISTRIBUTION THEORY

of randomly selected patients. Then, the sample space is

\[ \Omega = (0, \infty) \]

which is uncountable.

Definition 1.2. An event is any collection of possible outcomes of an experiment, that is, any subset of \( \Omega \).

We will use the following notation:

- \( \emptyset \) - impossible event (empty set),
- \( A \cup B \) - union of two events (sets),
- \( A \cap B \) - intersection of two events (sets),
- \( A^c \) - complement of a set \( A \) in sample space \( \Omega \),
- \( \bigcup_{i=1}^{\infty} A_i \) - union of an infinite number of events,
- \( \bigcap_{i=1}^{\infty} A_i \) - intersection of an infinite number of events.

Exercise 1.1. Recall the following properties of set operations: commutativity, associativity, distributive laws, DeMorgan’s laws.

Definition 1.3. Two events \( A \) and \( B \) are disjoint if \( A \cap B = \emptyset \). The events \( A_1, A_2, \ldots \) are pairwise disjoint (mutually exclusive) if \( A_i \cap B_j = \emptyset \) for all \( i \neq j \).

Definition 1.4. If \( A_1, A_2, \ldots \) are pairwise disjoint and \( \bigcup_{i=1}^{\infty} A_i = \Omega \), then the collection \( A_1, A_2, \ldots \) forms a partition of \( \Omega \).

Partitions let us to divide a sample space into non-overlapping pieces. For example, intervals \( A_i = [i, i+1) \) for \( i = 0, 1, 2, \ldots \) form a partition of \( \Omega = [0, \infty) \).
1.2 Probability Function

Random events have various chances to occur. These chances are measured by probability functions. Here we formally introduce such functions.

**Definition 1.5.** We call $\mathcal{A}$ a $\sigma$-algebra defined on $\Omega$ if it is a collection of subsets of the set $\Omega$ such that

1. $\emptyset \in \mathcal{A}$,
2. if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$,
3. if $A_1, A_2, \ldots \in \mathcal{A}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. $\square$

The pair $(\Omega, \mathcal{A})$ is called the measurable space. Property (1) states that an empty set is always in the $\sigma$-algebra. Then, by property (2), $\Omega \in \mathcal{A}$ as well. Properties (2) and (3) together with DeMorgan’s Laws assure that also

$$\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}.$$ 

That is, a $\sigma$-algebra is closed under both union and intersection of its elements.

**Exercise 1.2.** Show that a $\sigma$-algebra is closed under intersection of its elements.

**Example 1.2.** The smallest and largest $\sigma$-algebras:

(a) **The Smallest $\sigma$-algebra** containing an event $A \subset \Omega$ is

$$\mathcal{A} = \{\emptyset, A, A^c, \Omega\}.$$ 

(b) **The Power Set** of $\Omega$ is the $\sigma$-algebra with the largest number of elements, that is, it is the set of all subsets of $\Omega$,

$$\mathcal{A} = \{A : A \subseteq \Omega\}.$$ 

$\square$
Example 1.3. Given the sample space as in (a) of Example 1.1, the $\sigma$-algebra which is a power set consists of all events showing a specific number of dots on the top face of a tossed die, any of two numbers of dots, any of three and so on; also an empty set:

$$ A = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \ldots, \{1, 2, 3, 4, 5, 6\}, \{\emptyset\} \}. $$

On the other hand, in (b) of Example 1.1, an analogous $\sigma$-algebra contains all intervals of the form

$$ [a, b], \ (a, b], \ (a, b), \ (a, b) ] $$

where $a, b \in \mathbb{R}_+ \ a \leq b$.

Definition 1.6. Given a measurable space $(\Omega, A)$, $P$ is a probability function if it is a measure such that

1. $P : A \rightarrow [0, 1]$,
2. $P(\Omega) = 1$
3. If $A_1, A_2, \ldots$ are pairwise disjoint then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

The triple $(\Omega, A, P)$ is called a probability space.

Example 1.4. For a fair die, we may assume that the probability of getting number $i$ on its face is

$$ P(\{i\}) = \frac{1}{6}. $$

Then the triple $(\Omega, A, P)$, where $\Omega$ and $A$ are defined as in Examples 1.1 and 1.3 is a probability space.

Some fundamental properties of the probability function are given in the following theorem.

Theorem 1.1. For a given probability space $(\Omega, A, P)$ and any events $A$ and $B$ in $A$ the following properties hold:

(a) $P(\emptyset) = 0$;
(b) $P(A) \leq 1$;
1.2. PROBABILITY FUNCTION

(c) \( P(A^c) = 1 - P(A) \);

(d) \( P(B \cap A^c) = P(B) - P(A \cap B) \);

(e) \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \);

(f) If \( A \subset B \) then \( P(A) \leq P(B) \).

□

Two other useful properties are given in the following theorem.

**Theorem 1.2.** If \( P \) is a probability function, then

(a) \( P(A) = \sum_{i=1}^{\infty} P(A \cap C_i) \) for any partition \( C_1, C_2, \ldots \).

(b) \( P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i) \) for any events \( A_1, A_2, \ldots \). [Boole’s Inequality]

□

**Exercise 1.3.** Prove Theorem 1.2.

The following theorem establishes the connection between limits of probabilities and sequences of events.

**Theorem 1.3.** For a given probability space \( (\Omega, A, P) \) the following hold:

(a) If \( \{A_1, A_2, A_3, \ldots\} \) is an increasing sequence of events, i.e., \( A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots \), then

\[
\lim_{n \to \infty} P(A_n) = P(\bigcup_{i=1}^{\infty} A_i).
\]

(b) If \( \{A_1, A_2, A_3, \ldots\} \) is a decreasing sequence of events, i.e., \( A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots \), then

\[
\lim_{n \to \infty} P(A_n) = P(\bigcap_{i=1}^{\infty} A_i).
\]

□
1.3 Random Variables

Definition 1.7. A random variable $X$ is a function that assigns one and only one numerical value to each outcome of an experiment, that is

$$X : \Omega \to X \subseteq \mathbb{R}.$$ 

Given a probability space $(\Omega, A, P)$ a random variable $X$ can be viewed as follows:

(a) if $\Omega$ is a countable set, i.e., $\Omega = \{\omega_1, \omega_2, \ldots\}$, then we will observe $X = x_i$ iff the outcome of the random experiment is an $\omega_j \in \Omega$ such that $X(\omega_j) = x_i$.

Hence,

$$P(X = x_i) = P\left(\{\omega_j \in \Omega : X(\omega_j) = x_i\}\right).$$

(b) If $\Omega$ is uncountable, then analogously we can write

$$P(X \in B) = P\left(\{\omega \in \Omega : X(\omega) \in B\}\right).$$

Note that $P(X = x_i)$ ($P(X \in B)$) is a probability function on $X$, defined in terms of the probability function on $\Omega$.

We will denote rvs by capital letters: $X$, $Y$ or $Z$ and their values by small letters: $x$, $y$ or $z$ respectively.

There are two types of rvs: discrete and continuous. Random variables that take a countable number of values are called discrete. Random variables that take values from an interval of real numbers are called continuous.

Example 1.5. In an efficacy preclinical trial a drug candidate is tested on mice. The observed response can either be “efficacious” or “non-ef ficacious”. Here we have

$$\Omega = \{\omega_1, \omega_2\},$$

where $\omega_1$ is an outcome of the experiment meaning efficacious response to the drug, $\omega_2$ is an outcome meaning non-ef ficacious response. Assume that $P(\omega_1) = p$ and $P(\omega_2) = q = 1 - p$. 

It is natural to define a random variable $X$ as follows:

$$X(\omega_1) = 1, X(\omega_2) = 0.$$  

Then

$$P(X = 1) = P\left(\{\omega_j \in \Omega : X(\omega_j) = 1\}\right) = P(\omega_1) = p$$

and

$$P(X = 0) = P\left(\{\omega_j \in \Omega : X(\omega_j) = 0\}\right) = P(\omega_0) = q.$$  

Here $\mathcal{X} = \{0, 1\}$.  

**Example 1.6.** In the same experiment we might observe a continuous outcome, for example, time to a specific reaction. Then

$$\Omega = (0, \infty).$$

Many random variables could be of interest in this case. For example, it could be a continuous rv, such as

$$Y : \Omega \to \mathcal{Y} \subseteq \mathbb{R},$$

where

$$Y(\omega) = \ln \omega.$$  

Here $\mathcal{Y} = (-\infty, \infty)$. Then, for example,

$$P(Y \in (-1, 1)) = P(\{\omega \in \Omega : Y(\omega) \in (-1, 1)\}) = P(\omega \in (e^{-1}, e)).$$

On the other hand, if we are interested in just two events: either that the time is less than a pre-specified value, say $t^*$, or that it is more than this value, than we categorize the outcome and the sample space is

$$\Omega = \{\omega_1, \omega_2\},$$

where $\omega_1$ means that we observed the time to the reaction shorter than or equal to $t^*$, $\omega_2$ means that the time was longer than $t^*$. Then we can define a discrete random variable as in Example 1.5.  

\[\square\]
1.4 Distribution Functions

**Definition 1.8.** The probability of the event \( X \leq x \) expressed as a function of \( x \in \mathbb{R} \):

\[
F_X(x) = P(X \leq x)
\]

is called the **cumulative distribution function** (cdf) of the rv \( X \).

\[\square\]

**Example 1.7.** The cdf of the rv defined in Example 1.5 can be written as

\[
F_X(x) = \begin{cases} 
0, & \text{for } x \in (-\infty, 0); \\
q, & \text{for } x \in [0, 1); \\
q + p = 1, & \text{for } x \in [1, \infty).
\end{cases}
\]

\[\square\]

Properties of cumulative distribution functions are given in the following theorem.

**Theorem 1.4.** The function \( F_X(x) \) is a cdf iff the following conditions hold:

(i) The function is nondecreasing;

(ii) \( \lim_{x \to -\infty} F_X(x) = 0 \);

(iii) \( \lim_{x \to \infty} F_X(x) = 1 \);

(iv) \( F_X(x) \) is right-continuous.

**Proof.** Note that a cdf can be equivalently written as

\[
F_X(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) = P(A_x),
\]

where \( A_x = \{\omega \in \Omega : X(\omega) \leq x\} \).

(i) For any \( x_i < x_j \) we have \( A_{x_i} \subseteq A_{x_j} \). Thus by 1.1 part (f), we have

\[
P(A_{x_i}) \leq P(A_{x_j}) \quad \text{and so } F_X(x_i) \leq F_X(x_j).
\]
(ii) Let \( \{x_n : n = 1, 2, \ldots\} \) be a sequence of decreasing real numbers such that \( x_n \to -\infty \) as \( n \to \infty \). Then, for \( x_n \geq x_{n+1} \) we have \( A_{x_n} \supseteq A_{x_{n+1}} \) and 
\[
\bigcap_{n=1}^{\infty} A_{x_n} = \emptyset.
\]
Hence, by Theorem 1.3 (b), we get 
\[
\lim_{n \to \infty} F_X(x_n) = \lim_{n \to \infty} P(A_{x_n}) = P\left( \bigcap_{n=1}^{\infty} A_{x_n} \right) = P(\emptyset) = 0.
\]
Since the above holds for any sequence \( \{x_n\} \) such that \( \{x_n\} \to -\infty \), we conclude that \( \lim_{x \to -\infty} F_X(x) = 0 \).

(iii) Can be proved by similar reasoning as in (ii).

(iv) A function \( g : \mathbb{R} \to \mathbb{R} \) is right continuous if \( \lim_{\delta \to 0^+} g(x + \delta) = g(x) \).

Let \( x_n \) be a decreasing sequence such that \( x_n \to x \) as \( n \to \infty \). Then, by definition, \( A_x \subseteq A_{x_n} \) for all \( n \) and \( A_x \) is the largest set for which it is true. This gives 
\[
\bigcap_{n=1}^{\infty} A_{x_n} = A_x
\]
and by Theorem 1.3 (b), we get 
\[
\lim_{n \to \infty} F_X(x_n) = \lim_{n \to \infty} P(A_{x_n}) = P\left( \bigcap_{n=1}^{\infty} A_{x_n} \right) = P(A_x) = F_X(x).
\]
Since the above holds for any sequence \( \{x_n\} \) such that \( \{x_n\} \to x \), we conclude that \( \lim_{\delta \to 0^+} F_X(x + \delta) = F_X(x) \) for all \( x \) and so \( F_X \) is right continuous.

\[\square\]

Example 1.8. We will show that 
\[
F_X(x) = \frac{1}{1 + e^{-x}}
\]
is the cdf of a rv \( X \). It is enough to show that the function meets the requirements of Theorem 1.4. The derivative of \( F(x) \) is 
\[
F'_X(x) = \frac{e^{-x}}{(1 + e^{-x})^2} > 0,
\]

showing that \( F(x) \) is increasing.

\[
\begin{align*}
\lim_{x \to -\infty} F_X(x) &= 0 \quad \text{since} \quad \lim_{x \to -\infty} e^{-x} = \infty; \\
\lim_{x \to \infty} F_X(x) &= 1 \quad \text{since} \quad \lim_{x \to \infty} e^{-x} = 0.
\end{align*}
\]

Also, it is a continuous function, not only right-continuous.

Now we can define discrete and continuous rvs more formally.

**Definition 1.9.** A random variable \( X \) is continuous if \( F_X(x) \) is a continuous function of \( x \). A random variable \( X \) is discrete if \( F_X(x) \) is a step function of \( x \).

### 1.5 Density and Mass Functions

#### 1.5.1 Discrete Random Variables

Values of a discrete rv are elements of a countable set \( \{x_1, x_2, \ldots \} \). We associate a number \( p_X(x_i) = P(X = x_i) \) with each value \( x_i, \ i = 1, 2, \ldots \), such that:

1. \( p_X(x_i) \geq 0 \) for all \( i \);
2. \( \sum_{i=1}^{\infty} p_X(x_i) = 1 \).

Note that

\[
F_X(x_i) = P(X \leq x_i) = \sum_{x \leq x_i} p_X(x), \tag{1.1}
\]

\[
p_X(x_i) = F_X(x_i) - F_X(x_{i-1}). \tag{1.2}
\]

The function \( p_X \) is called the **probability mass function** (pmf) of the random variable \( X \), and the collection of pairs

\[
\{(x_i, p_X(x_i)), \ i = 1, 2, \ldots \} \tag{1.3}
\]

is called the **probability distribution** of \( X \). The distribution is usually presented in either tabular, graphical or mathematical form.
Example 1.9. Consider Example 1.5, but now, we have $n$ mice and we observe efficacy or no efficacy for each mouse independently. We are interested in the number of mice which respond positively to the applied drug candidate. If $X_i$ is a random variable as defined in Example 1.5 for each mouse, and we may assume that the probability of a positive response is the same for all mice, then we may create a new random variable $X$ as the sum of all $X_i$, that is,

$$X = \sum_{i=1}^{n} X_i.$$  

$X$ denotes $k$ successes in $n$ independent trials and it has a binomial distribution, which we denote by

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

where $p$ is the probability of success. The pmf of a binomially distributed rv $X$ with parameters $n$ and $p$ is

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, 2, \ldots, n,$$

where $n$ is a positive integer and $0 \leq p \leq 1$.

Exercise 1.4. For $n = 8$ and the probability of success $p = 0.4$ obtain mathematical, tabular and graphical form (pmf and cdf) of $X \sim \text{Bin}(n, p)$. □

1.5.2 Continuous Random Variables

Values of a continuous rv are elements of an uncountable set, for example a real interval. A cdf of a continuous rv is a continuous, nondecreasing, differentiable function. An interesting difference from a discrete rv is that for a $\delta > 0$

$$P(X = x) = lim_{\delta \to 0} (F_X(x + \delta) - F_X(x)) = 0.$$

For a continuous random variable $P(X \leq x)$ is

$$F_X(x) = \int_{-\infty}^{x} f_X(t) dt,$$  \hspace{1cm} (1.4)

where

$$f_X(x) = \frac{d}{dx} F_X(x)$$  \hspace{1cm} (1.5)
14 CHAPTER 1. ELEMENTS OF PROBABILITY DISTRIBUTION THEORY

is the probability density function (pdf).

Similarly to the properties of the probability distribution of a discrete rv we have the following properties of the density function:

1. \( f_X(x) \geq 0 \) for all \( x \in X \);
2. \( \int_X f_X(x) dx = 1 \).

Probability of an event that \( X \in (-\infty, a) \), is expressed as an integral

\[
P(-\infty < X < a) = \int_{-\infty}^{a} f_X(x) dx = F_X(a)
\]  

(1.6)

or for a bounded interval \((b, c)\) as

\[
P(b < X < c) = \int_{b}^{c} f_X(x) dx = F_X(c) - F_X(b).
\]

(1.7)

**Exercise 1.5.** A certain river floods every year. Suppose that the low-water mark is set at 1 and a high-water mark \( X \) has distribution function

\[
F_X(x) = \begin{cases} 
0, & \text{for } x < 1; \\
1 - \frac{1}{x^2}, & \text{for } x \geq 1.
\end{cases}
\]

1. Verify that \( F_X(x) \) is a cdf.
2. Find the pdf of \( X \).
3. Calculate the probability that the high-water mark is between 3 and 4.