| | |
|---|---|
| Module: | Probability and Stochastic Modelling |
| Term: | Autumn 2024 |
| Lecturer: | Richard Pymar |

# 1 Basic concepts of Probability Theory and Univariate Distribution Theory

## 1.1 The Probability Space

We set up the mathematical framework on which we can construct random variables representing quantities associated with an experiment. There are three elements to this framework, known as the probability space: a sample space $\Omega$, a family of events $\mathcal{F}$, and a probability $\mathbb{P}$.

**Definition 1.1** (Sample Space) The set of all possible outcomes of an experiment is called the sample space and is denoted by $\Omega$.

In practice, $\Omega$ can just be any non-empty set. Members or elements of $\Omega$ are known as sample points, they denote elementary outcomes and are often denoted by $\omega$.

**Example 1.1** Suppose a coin is flipped. There are two possible outcomes: heads or tails. So the sample space $\Omega = \{H, T\}$.

**Example 1.2** Suppose a standard die is thrown. There are six possible outcomes. Thus $\Omega = \{1, 2, 3, 4, 5, 6\}$. We may be interested in the following events: (a) the outcome is an even number; (b) the outcome is even but does not exceed 3; (c) the outcome is not even; (d) the outcome is either even or does not exceed 3. These events correspond to different subsets of $\Omega$, respectively $\{2, 4, 6\}$, $\{2\}$, $\{1, 3, 5\}$ and $\{1, 2, 3, 4, 6\}$.

Examples 1.1 and 1.2 have finite sample spaces, the following examples have infinite sample spaces.

**Example 1.3** A coin is tossed repeatedly until a head occurs; we are interested in the number of tosses until this happens. The set of all outcomes is $\Omega = \{1, 2, 3, \cdots\}$. Here the sample space is countable.

**Example 1.4** Height of individuals in a class. The outcome should be positive.

$$\Omega = \{x : 0 < x < \infty\} \quad \text{or} \quad \{x : a < x < b\},$$

where real numbers $a$ and $b$ are properly chosen. Here the sample space is uncountable.

Now we explain the definition of countability of sets.

**Definition 1.2** (Countable) If a set $A$ has the same cardinality as $\mathbb{N}$ (the Natural numbers), then we say that $A$ is countable. In other words, a set is countable if there is a bijection from that set to $\mathbb{N}$.

An alternative way to define countable is: if there is a way to enumerate the elements of a set, then the set has the same cardinality as $\mathbb{N}$ and is called countable. An example of a set that is not countable is the real numbers $\mathbb{R}$. The rational numbers $\mathbb{Q}$ are countable since we can easily 'pair' them with the Natural numbers.

Table 1. The jargon of set theory and probability theory

| Typical notation | Set jargon | Probability jargon |
|---|---|---|
| $\varnothing$ | Empty set | Impossible event |
| $\Omega$ | Whole space | Certain event |
| $\omega$ | Member of $\Omega$ | Elementary event, outcome |
| $A$ | Subset of $\Omega$ | Event that some outcome in A occurs |
| $A^c$ $(\overline{A})$ | Complement of A | Event that no outcome in A occurs |
| $A \cap B$ | Intersection | Both A and B |
| $A \cup B$ | Union | Either A or B or Both |
| $A - B$ | Difference | A, but not B |
| $A \subseteq B$ | Inclusion | If A, then B |

There are several fundamental operations for constructing new events from given events (see Table 1).

**Example 1.5** (Example 1.2 revisited) Let

$$A = \{\text{outcome is an even number}\} = \{2, 4, 6\}$$

and

$$B = \{\text{outcome does not exceed 3}\} = \{1, 2, 3\}.$$

It follows that the answer to (b) is $A \cap B$ and (c) $A^c$ and (d) $A \cup B$.

The idea of construction can be applied to more than two events and infinite events.

For $n$ events $A_1, A_2, \cdots, A_n$, we have

$$\cup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n; \quad \cap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n.$$

For infinite countable events $A_1, A_2, \cdots$, we have

$$\cup_{i=1}^{\infty} A_i = \lim_{n\to\infty} \cup_{i=1}^{n} A_i; \quad \cap_{i=1}^{\infty} A_i = \lim_{n\to\infty} \cap_{i=1}^{n} A_i.$$

Properties of these operations are omitted here except De Morgan's Theorem:

$$(\cup_{i\in I} A_i)^c = \cap_{i\in I} A_i^c; \quad (\cap_{i\in I} A_i)^c = \cup_{i\in I} A_i^c,$$

where $I$ is a finite or countable set.

**Example 1.6** (Example 1.3 revisited) The event that the first head occurs upon completion of an even number of tosses is an infinite countable union of members of $\Omega$, that is,

$$A = \{2, 4, 6, \cdots\} = \{2\} \cup \{4\} \cup \{6\} \cup \cdots .$$

Not all the subsets of $\Omega$ are always of interest and when $\Omega$ is infinite, the collection of all the subsects of $\Omega$ is too large for probabilities to be assigned reasonably to all its members, so we will concentrate on a collection of some subsets of $\Omega$ which contains all the events we may be interested in.

**Definition 1.3** (Field) Let $\mathcal{F}$ be a collection of subsets of $\Omega$. If $\mathcal{F}$ satisfies the following conditions:

1. If $A, B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$;
2. If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$;
3. The empty set $\varnothing \in \mathcal{F}$.

Then $\mathcal{F}$ is called a field.

In the literature the term *algebra* is also used instead of *field*. This definition is fine when $\Omega$ is finite, but how to define a field when $\Omega$ is infinite?

**Definition 1.4** ($\sigma$-field) A collection $\mathcal{F}$ of subsets of $\Omega$ is called a $\sigma$-field if it satisfies the following conditions:

1. The set $\varnothing \in \mathcal{F}$;
2. If $A_1, A_2, \ldots \in \mathcal{F}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$;
3. If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.

At this point we may ask: "What is the difference between Definition 1.3 and Definition 1.4?". In fact the only difference is that when we consider a $\sigma$-field we are talking about a countable number of sets when operations such as union are considered. Hence we could have defined a $\sigma$-field in the following way: A class $\mathcal{F}$ of subsets of $\Omega$ is a $\sigma$-field if it is a field and if it is also closed under the formation of countable unions.

**Example 1.7**

1. The smallest $\sigma$-field associated with the sample space $\Omega$ is the collection $\mathcal{F} = \{\varnothing, \Omega\}$.
2. If A is any subset of $\Omega$ then $\mathcal{F} = \{\varnothing, A, A^c, \Omega\}$ is a $\sigma$-field.
3. The *power set* of $\Omega$, which is written $2^\Omega$ and contains all subsets of $\Omega$, is obviously a $\sigma$-field.

**Definition 1.5** (Events) The elements (or members) of a $\sigma$-field $\mathcal{F}$ are called events.

With any experiment, we may associate a pair $(\Omega, \mathcal{F})$, where $\Omega$ is the set of all possible outcomes and $\mathcal{F}$ is a $\sigma$-field of subsets $\Omega$ which contains all the events in whose occurrences we may be interested. Now we wish to assign probabilities to the occurrences of events. These probabilities should satisfy some properties: the probability that an event $A$ occurs is positive or zero for any event $A \in \mathcal{F}$; the probability of $\Omega$ is 1; the probability that the union of disjoint events occurs is equal to the sum of all the probabilities that each of the disjoint events occurs. Probability measure (or probability function) is defined by specifying these desirable properties.

**Definition 1.6** (Probability Measure) A probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a set function with the domain $\mathcal{F}$ satisfying:

1. $\mathbb{P}(A) \geq 0$ for every $A \in \mathcal{F}$,
2. $\mathbb{P}(\Omega) = 1$,
3. If $A_1, A_2, \ldots$ is a collection of disjoint members of $\mathcal{F}$, i.e, $A_i \cap A_j = \varnothing$ for $i \neq j$, then

$$\mathbb{P}\left(\cup_{i=1}^{\infty} A_i\right) = \Sigma_{i=1}^{\infty} \mathbb{P}(A_i).$$

**Definition 1.7** (Probability Space) The triple $(\Omega, \mathcal{F}, \mathbb{P})$ comprising of a set $\Omega$, a $\sigma$-field $\mathcal{F}$ of subsets of $\Omega$, and a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is called a probability space.

**Example 1.8** (Example 1.1 revisited) A coin, possibly biased, is tossed once. We can take $\Omega = \{H, T\}$ and $\mathcal{F} = \{\varnothing, H, T, \Omega\}$, and a possible measure $\mathbb{P} : \mathcal{F} \longrightarrow [0, 1]$ is given by

$$\mathbb{P}(\varnothing) = 0, \quad \mathbb{P}(H) = p, \quad \mathbb{P}(T) = 1 - p, \quad \mathbb{P}(\Omega) = 1,$$

where $p$ is a fixed real number in the interval $(0, 1)$. If $p = \frac{1}{2}$, then we say the coin is fair, or unbiased.

**Example 1.9** (Example 1.2 revisited) A die is thrown once. We can take $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = 2^{\Omega}$, and the probability measure $\mathbb{P}$ given by

$$\mathbb{P}(A) = \sum_{i \in A} p_i \quad \text{for any} \quad A \subseteq \Omega,$$

where $p_1, p_2, \cdots, p_6$ are specified numbers from the interval $[0, 1]$ having unit sum. The probability that $i$ turns up is $p_i$. The die is fair if $p_i = \frac{1}{6}$ for each $i$.

There are some important properties which are often required.

**Lemma 1.1**

1. $\mathbb{P}(\varnothing) = 0$,
2. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
3. if $A \subseteq B$ then $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B - A) \geq \mathbb{P}(A)$, in other words,
$$\mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A) \geq 0,$$
4. General law of addition: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$,
5. more generally, if $A_1, A_2, \cdots, A_n$ are events, then
$$\begin{aligned}\mathbb{P}(\bigcup_{i=1}^{n} A_i) = &\sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) \\ &+ \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots \\ &+ (-1)^{(n+1)} \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n),\end{aligned}$$
where, for example, $\sum_{i<j}$ sums over all unordered pairs $(i, j)$ with $i \neq j$.

**Definition 1.8** (Null, Almost Surely) An event $A$ is called null if $\mathbb{P}(A) = 0$. If $\mathbb{P}(A) = 1$ we say that $A$ occurs almost surely.

Note that the impossible event $\varnothing$ is null, but null events need not be impossible.

## 1.2 Conditional probability and statistical independence

**Example 1.10** (Example 1.2 revisited) Let $B = \{\text{outcome exceeds 3}\}$ and $A = \{\text{outcome is an even number}\}$. If B occurs, what is the probability of the occurrence of $A$?

$$\mathbb{P}(A|B) = \text{card}(\{4,6\})/\text{card}(\{4,5,6\}) = 2/3$$

while

$$\mathbb{P}(A) = \text{card}(\{2,4,6\})/\text{card}(\Omega) = 3/6 = 1/2,$$

where card(Z) is the number (cardinality) of elements in some set $Z$.

**Definition 1.9** (Conditional Probability) If $\mathbb{P}(B) > 0$ then the conditional probability that $A$ occurs given that $B$ occurs is defined to be

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

Remark: The definition above gives us another way to calculate the probability that the intersection of $A$ and $B$ occurs, which is

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

Here $\mathbb{P}(B) \geq 0$. This relation is called *General Law of Multiplication*.

When speaking of conditional probabilities we are conditioning on some given event $B$; that is , we assume that the experiment has resulted in some outcome in $B$. Thus $B$, in effect, then becomes our "new" sample space. One question that might be raised is : for a given event $B$ for which $\mathbb{P}(B) > 0$, is $\mathbb{P}(\cdot|B)$ a probability function having $\mathcal{F}$ as its domain? In other words, does $\mathbb{P}(\cdot|B)$ satisfy the three axioms? The answer is 'yes' because:

1. $\mathbb{P}(A|B) \geq 0$ for every $A \in \mathcal{F}$,
2. $\mathbb{P}(\Omega|B) = 1$,
3. if $A_1, A_2, \cdots$ , is a sequence of mutually exclusive events in $\mathcal{F}$, then

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B).$$

Hence, $\mathbb{P}(\cdot|B)$ for given $B$ satisfying $\mathbb{P}(B) > 0$ is a probability function, which justifies our calling it a conditional probability. $\mathbb{P}(\cdot|B)$ also enjoys the same properties as the unconditional probability.

**Lemma 1.2**

1. $\mathbb{P}(\varnothing|B) = 0$,
2. If $A$ is an event in $\mathcal{F}$, then $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$,

3. If $A_1$ and $A_2$ in $\mathcal{F}$ and $A_1 \subset A_2$, then $\mathbb{P}(A_1|B) \leq \mathbb{P}(A_2|B)$.

4. For every two events $A_1$ and $A_2$ in $\mathcal{F}$,

$$\mathbb{P}(A_1 \cup A_2|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B) - \mathbb{P}(A_1 \cap A_2|B).$$

There are a number of other useful formulae involving conditional probabilities that we will state as lemmas.

**Definition 1.10** (Partition) A family $B_1, B_2, \cdots, B_n$ of events is called a partition of the set $\Omega$ if

$$B_i \cap B_j = \varnothing \quad when \quad i \neq j, \quad and \quad \cup_{i=1}^{n} B_i = \Omega.$$

Each elementary event $\omega \in \Omega$ belongs to exactly one set in a partition of $\Omega$.

The next lemma is crucially important in probability theory.

**Lemma 1.3** (Total probabilities) For any events $A$ and $B$ such that $0 < \mathbb{P}(B) < 1$

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c).$$

More generally, let $B_1, B_2, \ldots, B_n$ be a partition of $\Omega$ s.t. $\mathbb{P}(B_i) > 0$ for all $i$. Then

$$\mathbb{P}(A) = \Sigma_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Similar result holds for a partition with infinite countable events $B_1, B_2, \cdots$.

Often it is easy to obtain an expression for the right hand side of these two equations, especially the first equation, by conditioning on the outcome of a particular binary event.

**Example 1.11** A student can either take a course in computer science, mathematics or chemistry. In computer science, the student will get an A grade with probability 1/3, in math with 1/4, in chemistry with 1/5. However the student is unaware of these probabilities and chooses a course at random with equally likely outcomes. What is the probability of the student getting an A grade?

Let $CS$ be computer science, $M$ be mathematics and $CH$ be chemistry and $A$ be the event of getting an A grade. Then it follows that

$$\mathbb{P}(CS) = \mathbb{P}(M) = \mathbb{P}(CH) = 1/3.$$

The conditional probabilities are

$$\mathbb{P}(A|CS) = 1/3, \quad \mathbb{P}(A|M) = 1/4, \quad \mathbb{P}(A|CH) = 1/5.$$

Hence,

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \cap CS) + \mathbb{P}(A \cap M) + \mathbb{P}(A \cap CH)\\ &= \mathbb{P}(A|CS)\mathbb{P}(CS) + \mathbb{P}(A|M)\mathbb{P}(M) + \mathbb{P}(A|CH)\mathbb{P}(CH)\\ &= (1/3)(1/3) + (1/4)(1/3) + (1/5)(1/3) = 47/180.\end{aligned}$$

Bayes' Theorem exploits the theorems of total probability in an inverse way. Suppose we have the same set-up as in lemma 3 and we ask the posterior question $\mathbb{P}(B_k|A)$, i.e, given the occurrence of A, what is the probability of $B_k$? This may be solved in the following way.

**Lemma 1.4** (Bayes' Theorem)

$$\begin{aligned}\mathbb{P}(B_k|A) &= \frac{\mathbb{P}(B_k \cap A)}{\mathbb{P}(A)}\\ &= \frac{\mathbb{P}(B_k \cap A)}{\sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}\\ \mathbb{P}(B_k|A) &= \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}\end{aligned}$$

Note that we can compute $\mathbb{P}(B_k|A)$ for $k = 1, \cdots, n$ in turn. The probabilities $\mathbb{P}(B)$ are known as *prior* probabilities and are sometimes denoted by $\pi_i(B_i)$ where $\sum_{i=1}^{n} \pi_i(B_i) = 1$ i.e, a proper probability distribution. The event $A$ is often called the *datum* or simply the *data*, so that we sometimes write $\mathbb{P}(B_k|A) = \mathbb{P}(B_k|data)$.

The concept of *prior* means before the conditioning event $A$ has occurred, i.e, before one has seen the data. So the concept of *posterior* means after the conditioning event $A$ has occurred, i.e, after one has seen the data.

Example 15 (example 14 revisited). Using the set up in Example 14, a student attains an A grade and tells you that that is so, but refuses to divulge the course pursued. Using the information available compute the posterior probabilities $\mathbb{P}(CS|A)$, $\mathbb{P}(M|A)$ and $\mathbb{P}(CH|A)$.

$$\begin{aligned}\mathbb{P}(M|A) &= \frac{\mathbb{P}(A|M)\mathbb{P}(M)}{\mathbb{P}(A|CS)\mathbb{P}(CS) + \mathbb{P}(A|M)\mathbb{P}(M) + \mathbb{P}(A|CH)\mathbb{P}(CH)}\\ &= \frac{(1/4)(1/3)}{(1/3)(1/3) + (1/4)(1/3) + (1/5)(1/3)}\\ &= 15/47\end{aligned}$$

Similarly, we get

$$\mathbb{P}(CS|A) = 20/47, \quad \mathbb{P}(CH|A) = 12/47.$$

Thus it is most likely that the student did computer science but the probability at $20/47$ is not convincingly high (for example it is not greater than $0.5$). Notice that we have a proper posterior distribution as one would expect, i.e, $\mathbb{P}(M|A) + \mathbb{P}(CS|A) + \mathbb{P}(CH|A) = 1$.

In general, the occurrence of some event $B$ changes the probability that another $A$ occurs, the original probability $\mathbb{P}(A)$ is replaced by $\mathbb{P}(A|B)$, i.e, $\mathbb{P}(A|B) \neq \mathbb{P}(A)$. If the probability remains unchanged, that is to say $\mathbb{P}(A|B) = \mathbb{P}(A)$, then we call $A$ and $B$ 'independent'. Further, noting the general law of multiplication, we have the following relation

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

This is the idea behind the definition of statistical independence.

**Definition 1.11** (Statistical Independence) Events $A$ and $B$ are called statistically independent if

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

More generally, a family $\{A_i | i \in I\}$ is called statistically independent if

$$\mathbb{P}(\cap_{i \in J} A_i) = \prod_{i \in J} \mathbb{P}(A_i)$$

for all finite subsets $J$ of $I$.

Note that independence does not mean $A \cap B = \varnothing$.

**Example 1.12** Choose a card from a pack of 52 playing cards, each being picked with probability $1/52$. Then the suit of the chosen card is independent of its rank. For example,

$$\mathbb{P}(\text{king}) = \frac{4}{52} = \frac{1}{13}, \qquad \mathbb{P}(\text{king}|\text{spade}) = \frac{1}{13}.$$

Alternatively

$$\mathbb{P}(\text{spade} \cap \text{king}) = \frac{1}{52} = \frac{1}{4}\frac{1}{13} = \mathbb{P}(\text{spade})\mathbb{P}(\text{king}).$$

**Example 1.13** (Knowledge changes probabilities in surprising ways) Consider 3 statements and in each determine the probability that both children are boys:
(a) I have 2 children, the elder is a boy,
(b) I have 2 children, at least one is a boy,
(c) I have 2 children, at least one is a boy born on a Thursday.
Since we have no further information we will assume outcomes are equally likely.
(a) Write BB, BG, GB, GG for the possible outcomes (first letter corresponds to elder child). Then from the definition of conditional prob-

ability we have

$$\mathbb{P}(\text{BB} \mid \text{BG} \cup \text{BB}) = \frac{\mathbb{P}(\text{BB})}{\mathbb{P}(\text{BG} \cup \text{BB})} = \frac{1/4}{1/2} = 1/2.$$

(b) Using the same notation we have

$$\mathbb{P}(\text{BB} \mid \text{BG} \cup \text{BB} \cup \text{GB}) = \frac{\mathbb{P}(\text{BB})}{\mathbb{P}(\text{BG} \cup \text{BB} \cup \text{GB})} = \frac{1/4}{3/4} = 1/3.$$

(c) Write TT for both children are boys born on a Thursday, TN for the elder is a boy born on a Thursday and the younger is a boy not born on a Thursday, etc. Notice that

$$\text{BB} = \text{TT} \cup \text{TN} \cup \text{NT} \cup \text{NN}.$$

Also $\mathbb{P}(\text{TT}) = (\frac{1}{2} \times \frac{1}{7})^2 = \frac{1}{14^2}$, $\mathbb{P}(\text{TN}) = (\frac{1}{2} \times \frac{1}{7}) \times (\frac{1}{2} \times \frac{6}{7}) = \frac{3}{98}$, and $\mathbb{P}(\text{TG}) = (\frac{1}{2} \times \frac{1}{7}) \times \frac{1}{2} = \frac{1}{28}$. So we have

$$\begin{aligned}&\mathbb{P}(\text{TT} \cup \text{TN} \cup \text{NT} \cup \text{NN} \mid \text{TT} \cup \text{TN} \cup \text{NT} \cup \text{TG} \cup \text{GT})\\ &= \frac{\mathbb{P}(\text{TT} \cup \text{TN} \cup \text{NT})}{\mathbb{P}(\text{TT} \cup \text{TN} \cup \text{NT} \cup \text{TG} \cup \text{GT})} = \frac{13}{27} \in \left(\frac{1}{3}, \frac{1}{2}\right).\end{aligned}$$

## 1.3 Random variables and the Distribution function

A fair coin is tossed once: $\Omega = \{H, T\}$. We may record the outcome $\omega$ in the following way:

$$X(\omega) = \begin{cases} 1 & if \quad \omega = H; \\ 0 & if \quad \omega = T. \end{cases}$$

Here $X$, called a random variable, is a function mapping the sample space $\Omega = \{H, T\}$ into a subset of the real numbers, in this case $\{0, 1\}$. More generally, after an experiment is done and the outcome $\omega \in \Omega$ is known, a random variable $X$ is a real valued function mapping $\omega \in \Omega$ into $\mathbb{R}$ or some subset of $\mathbb{R}$.

Suppose an experiment has a finite or countable sample space, i.e, $\Omega = \{\omega_1, \omega_2, \cdots\}$. A random variable $X$ associated with $\Omega$ is $X(\omega) = x_i$ for $\omega \in \Omega$. Furthermore, this real valued function $X$ should have an extra property called measurability, i.e, $\{\omega : X(\omega) = x_i\} \in \mathcal{F}$, that allows us to make probability statements about the random variable. In real life, it is more common that experiments have uncountable sample spaces: for example, height and weight of people; temperature in some area; survival time of a human being with some disease. For a random variable $X$ associated with uncountable sample space $\Omega$, we are more concerned with the events for which $X$ is located in some interval such as $X \leq b$, $X > a$ and $a < X \leq b$, where $a$ and $b$ are real numbers. We need to ensure that $X$ is measurable: noting that $\{X > a\} = (-\infty, +\infty) - \{X \leq a\}$ and $\{a < X \leq b\} = \{X \leq b\} - \{X \leq a\}$, it is sufficient to require that $\{\omega : X(\omega) \leq x\} \in \mathcal{F}$, where $x \in \mathbb{R}$.

**Definition 1.12** (Random variable) A random variable is a function $X : \Omega \longrightarrow \mathbb{R}$ with the property that $\{\omega \in \Omega : X(\omega) \le x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$. Such a function is said to be $\mathcal{F}$**-measurable**.

**Definition 1.13** (Distribution Function) The distribution function of a random variable $X$ is the function $F : \mathbb{R} \longrightarrow [0, 1]$ given by $F(x) = \mathbb{P}(X \le x) = \mathbb{P}(\{\omega : X(\omega) \le x\})$.

Now we state some properties that are satisfied by distribution functions.

**Lemma 1.5** A distribution function $F$ has the following properties:

1. $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$;
2. If $x < y$ then $F(x) \le F(y)$;
3. $F$ is right continuous, that is, $F(x + h) \to F(x)$ as $h \searrow 0$.

Note that the first condition is straightforward since we know that distribution functions must integrate to unity.

For the time being we can forget all about probability spaces and concentrate on random variables and their distribution functions. The distribution function $F(x)$ contains a great deal of information about $X$.

**Lemma 1.6** Let $F$ be the distribution function of $X$. Then

1. $\mathbb{P}(X > x) = 1 - F(x)$,
2. $\mathbb{P}(x < X \le y) = F(y) - F(x)$,
3. $\mathbb{P}(X = x) = F(x) - \lim_{y\nearrow x} F(y)$.

# Appendix A

## The Gamma Function

**Definition 1.14** The GAMMA FUNCTION, denoted by $\Gamma(\cdot)$, is given by

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$$

for $\alpha > 0$.

**Proposition 1.1** (Generalization of factorial function to non-integer values)

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha) \qquad \alpha > 0.$$

**Proof** Use integration by parts.

$$\begin{aligned}\Gamma(\alpha) &= \textstyle\int_0^\infty t^{\alpha-1}e^{-t}dt = \frac{t^\alpha}{\alpha}e^{-t}|_0^\infty - \int_0^\infty \frac{t^\alpha}{\alpha}(-e^{-t})dt \\ &= 0 + \textstyle\frac{1}{\alpha}\int_0^\infty t^{(\alpha+1)-1}e^{-t}dt = \frac{1}{\alpha}\Gamma(\alpha+1) \\ &\Rightarrow \Gamma(\alpha+1) = \alpha\Gamma(\alpha).\end{aligned}$$

(Note that $t^\alpha e^{-t} \longrightarrow 0$ as $t \longrightarrow \infty$ for $\alpha > 0$).

**Lemma 1.7** (More useful results)
(i) $\Gamma(1) = 1$.
(ii) $\Gamma(n+1) = n!$ for $n \in \mathbb{Z}^+$.
(iii) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

**Proof**
(i) $\Gamma(1) = \int_0^\infty e^{-t}dt = 1$.
(ii) Set $\alpha = n$ in earlier proposition. Then

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)\Gamma(n-1) = \ldots = n!\Gamma(1) = n!$$

(iii)

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty t^{-\frac{1}{2}}e^{-t}dt.$$

Put $x = \sqrt{2t}$. Then

$$\begin{aligned}\Gamma\left(\tfrac{1}{2}\right) &= \textstyle\int_0^\infty \frac{\sqrt{2}}{x}e^{-\frac{x^2}{2}}xdx = \sqrt{2}\int_0^\infty e^{-\frac{x^2}{2}}dx = \frac{\sqrt{2}}{2}\int_{-\infty}^\infty e^{-\frac{x^2}{2}}dx \\ &= \textstyle\frac{\sqrt{2}}{2}\sqrt{2\pi} \times \int_{-\infty}^\infty \underbrace{\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}}_{\text{p.d.f. of } N(0,1)} dx = \sqrt{\pi} \times 1 = \sqrt{\pi}.\end{aligned}$$

# Appendix B
## Some Standard Discrete Distributions

| **Distribution** | **Prob. Mass Func.** $f(k)$ | **Mean** $E[X]$ | **Variance** var$(X)$ | **p.g.f.** $G_X(z)$ |
|---|---|---|---|---|
| BERNOULLI($\theta$) $\theta \in (0,1)$ | $\theta^k(1-\theta)^{1-k}$ $k=0$ or $1$ | $\theta$ | $\theta(1-\theta)$ | $(1-\theta)+\theta z$ |
| BINOMIAL(n,$\theta$) $n \in \mathbb{Z}^+,\ \ \theta \in (0,1)$ | $\binom{n}{k}\theta^k(1-\theta)^{n-k}$ $k=0,1,\ldots,n$ | $n\theta$ | $n\theta(1-\theta)$ | $[(1-\theta)+\theta z]^n$ |
| POISSON($\lambda$) $\lambda > 0$ | $e^{-\lambda}\frac{\lambda^k}{k!}$ $k=0,1,\ldots$ | $\lambda$ | $\lambda$ | $e^{\lambda(z-1)}$ |
| GEOMETRIC($\theta$) $\theta \in (0,1)$ | $\theta(1-\theta)^{k-1}$ $k=1,2,\ldots$ | $\frac{1}{\theta}$ | $\frac{1-\theta}{\theta^2}$ | $\frac{\theta z}{1-(1-\theta)z}$ |
| SHIFTED GEOMETRIC($\theta$) $\theta \in (0,1)$ | $\theta(1-\theta)^k$ $k=0,1,2,\ldots$ | $\frac{1-\theta}{\theta}$ | $\frac{1-\theta}{\theta^2}$ | $\frac{\theta}{1-(1-\theta)z}$ |
| NEGATIVE BINOMIAL($n,\ \theta$) $n \in \mathbb{Z}^+,\ \ \theta \in (0,1)$ | $\binom{k-1}{n-1}\theta^n(1-\theta)^{k-n}$ $k=n,n+1,\ldots$ | $\frac{n}{\theta}$ | $\frac{n(1-\theta)}{\theta^2}$ | $\left[\frac{\theta z}{1-(1-\theta)z}\right]^n$ |
| DISCRETE UNIFORM($n$) $n \in \mathbb{Z}^+$ | $\frac{1}{n}$ $k=1,2,\ldots,n$ | $\frac{n+1}{2}$ | $\frac{n^2-1}{12}$ | $\frac{z(1-z^n)}{n(1-z)}$ |

# Appendix C

## Some Standard Continuous Distributions

| **Distribution** | **Prob. Density Func.** $f(x)$ | **Mean** $E[X]$ | **Variance** var$(X)$ | **m.g.f.** $M_X(t)$ |
|---|---|---|---|---|
| UNIFORM$(a,b)$ <br> $-\infty < a < b < \infty$ | $\frac{1}{b-a}$ <br> $x \in (a,b]$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ | $\frac{e^{bt}-e^{at}}{(b-a)t}$ |
| EXPONENTIAL <br> or EXP$(\beta)$ <br> $\beta > 0$ | $\beta e^{-\beta x}$ <br> $x > 0$ | $\frac{1}{\beta}$ | $\frac{1}{\beta^2}$ | $\frac{\beta}{\beta-t}$ |
| GAMMA$(\alpha,\beta)$ <br> $\alpha > 0,\ \beta > 0$ | $\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ <br> $0 < x < \infty$ | $\frac{\alpha}{\beta}$ | $\frac{\alpha}{\beta^2}$ | $\left(\frac{\beta}{\beta-t}\right)^\alpha$ |
| NORMAL <br> N$(\mu,\sigma^2)$ <br> $\mu \in \mathbb{R}$, $\sigma^2 > 0$ | $\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ <br> $-\infty < x < \infty$ | $\mu$ | $\sigma^2$ | $e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ |
| STANDARD NORMAL <br> N$(0,1)$ | $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ <br> $-\infty < x < \infty$ | $0$ | $1$ | $e^{\frac{t^2}{2}}$ |
| BETA$(\alpha,\beta)$ <br> $\alpha > 0,\ \beta > 0$ | $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ <br> $0 < x < 1$ | $\frac{\alpha}{\alpha+\beta}$ | $\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$ | No simple /useful form |
| CAUCHY$(\alpha,\beta)$ <br> $\alpha \in \mathbb{R},\ \beta > 0$ | $\frac{1}{\pi\beta\left(1+\frac{(x-\alpha)^2}{\beta^2}\right)}$ <br> $-\infty < x < \infty$ | Doesn't exist | Doesn't exist | Doesn't exist |

*Last updated: September 23, 2024*