1 Probability Theory

1.1 Set Theory

Definition 1.1 *The set, S, of all possible outcomes of a particular experiment is called the sample space for the experiment.*

Example 1.2 *Tossing a coin twice*

$$S = \{HH, HT, TH, TT\}.$$

Example 1.3 *Choosing a real number in the interval [0,1]*

$$S = \{x : 0 \leq x \leq 1\}.$$

Sample spaces are either countable or uncountable.

Definition 1.4 *An event is any collection of possible outcomes of an experiment, that is, any subset of S (including S itself).*

Example 1.5 *Events in Example 1.2 are:*

$$\begin{array}{l}
\{HH\}, \{HT\}, \{TH\}, \{TT\} \\
\{HH, HT\}, \{HH, TH\}, \{HH, TT\}, \{HT, TH\}, \{HT, TT\}, \{TH, TT\} \\
\{HH, HT, TH\}, \{HH, HT, TT\}, \{HH, TH, TT\}, \{HT, TH, TT\} \\
\{HH, HT, TH, TT\}
\end{array}$$

Union: The union of A and B, written $A \cup B$, is the set of elements that belong to either A or B or both:

$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$

Intersection: The intersection of A and B, written $A \cap B$, is the set of elements that belong to both A and B:

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

Complementation: The complement of A, written A' (A^c), is the set of all elements that are not in A:

$$A' = \{x : x \notin A\}.$$

Example 1.6 *In Example 1.2, the event of observing at least one head is $A = \{HH, TH, HT\}$ and the event of observing two heads is $B = \{HH\}$. From these events, we have*

$$A \cup B = \{HH, TH, HT\};\ A \cap B = \{HH\};\ A' = \{TT\}.$$

Theorem 1.7 *For any three events A, B and C defined in a sample space S,*
a. Commutativity: $A \cup B = B \cup A$ *and* $A \cap B = B \cap A$;
b. Associativity: $A \cup (B \cup C) = (A \cup B) \cup C$ *and* $A \cap (B \cap C) = (A \cap B) \cap C$;
c. Distributive laws: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ *and* $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
d. De Morgan's law: $(A \cup B)' = A' \cap B'$ *and* $(A \cap B)' = A' \cup B'$.

The operations of union and intersection can be extended to infinite collections of sets.

$$\begin{aligned} \cup_{i=1}^{\infty} A_i &= \{x \in S : x \in A_i \text{ for some } i\} \\ \cap_{i=1}^{\infty} A_i &= \{x \in S : x \in A_i \text{ for all } i\} \end{aligned}$$

or

$$\begin{aligned} \cup_{\alpha \in \Gamma} A_\alpha &= \{x \in S : x \in A_\alpha \text{ for some } \alpha\} \\ \cap_{\alpha \in \Gamma} A_i &= \{x \in S : x \in A_\alpha \text{ for all } \alpha\}, \end{aligned}$$

where Γ is an index set possibly uncountable.

Example 1.8 *Let $S = (0, 1]$ and define $A_i = [(1/i), 1]$. Then,*

$$\begin{aligned} \cup_{i=1}^{\infty} A_i &= \cup_{i=1}^{\infty} [(1/i), 1] = \{x \in (0, 1]\}; \\ \cap_{i=1}^{\infty} A_i &= \cap_{i=1}^{\infty} [(1/i), 1] = \{x \in [1, 1]\} = \{1\}. \end{aligned}$$

Definition 1.9 *Two events A and B are disjoint (or mutually exclusive) if $A \cap B = 0$. The events $A_1, A_2, \ldots$ are pairwise disjoint (or mutually exclusive) if $A_i \cap A_j = 0$ for all i, j.*

Definition 1.10 *If $A_1, A_2, \ldots$ are pairwise disjoint and $\cup_{i=1}^{\infty} A_i = S$, then the collection $A_1, A_2, \ldots$ forms a partition of S.*

1.2 Basics of Probability Theory

Probabilities are defined by a function that satisfy some axioms. We are not concerned with the interpretation of probabilities.

Interpretations of the probabilities are quite another matter. The "frequency of occurrence" of an event is one example of a particular interpretation of probability. Another possible interpretation is a subjective one, where rather than thinking of probability as frequency, we can think of it as a belief in the chance of an event occurring.

1.2.1 Axiomatic Foundations

For each event A in the sample space S we want to associate with A a number between zero and one that will be called the probability of A, denoted by $P(A)$. It would seem natural to define the domain of P as all subsets of S. We define $P(A)$ as the probability that A occurs.

Definition 1.11 *A collection of subsets of S is called a sigma algebra (or Borel field), denoted by $\mathcal{B}$, if it satisfies the following three properties:*
a. $\emptyset \in \mathcal{B}$.
b. If $A \in \mathcal{B}$, *then* $A^c \in \mathcal{B}$.
c. If $A_1, A_2, ... \in \mathcal{B}$, *then* $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$.

According to this definition,

$$S \in B;\ A_1^c, A_2^c, ... \in \mathcal{B};\ \cup_{i=1}^{\infty} A_i^c \in \mathcal{B} \text{ and } (\cup_{i=1}^{\infty} A_i^c)^c = \cap_{i=1}^{\infty} A_i \in \mathcal{B}.$$

Example 1.12 *In Example 1.2, the largest sigma-field is the set containing*

$$\begin{aligned}&\{HH\}, \{HT\}, \{TH\}, \{TT\}\\ &\{HH, HT\}, \{HH, TH\}, \{HH, TT\}, \{HT, TH\}, \{HT, TT\}, \{TH, TT\}\\ &\{HH, HT, TH\}, \{HH, HT, TT\}, \{HH, TH, TT\}, \{HT, TH, TT\}\\ &\{HH, HT, TH, TT\}, \emptyset\end{aligned}$$

Definition 1.13 *Given a sample space S and an associated sigma algebra $\mathcal{B}$, a probability function is a function P with domain $\mathcal{B}$ that satisfies*
a. $P(A) \geq 0$ *for all* $A \in \mathcal{B}$.
b. $P(S) = 1$.
c. If $A_1,\ A_2, ... \in \mathcal{B}$ *are pairwise disjoint, then* $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

The three properties given in this definition are usually referred to as the Axioms of Probability (or the Kolmogorov Axioms). Any function $P(\cdot)$ that satisfies the Axioms of Probability is called a probability function.

Example 1.14 *Consider the simple experiment of tossing a fair coin, so* $S = \{H, T\}$. *By a "fair" coin we mean* $P(\{H\}) = P(\{T\})$. *Since* $S = \{H\} \cup \{T\}$, $P(\{H\} \cup \{T\}) = 1$. *Also,* $\{H\}$ *and* $\{T\}$ *are disjoint, so* $P(\{H\} \cup \{T\}) = P(\{H\}) + P(\{T\}) = 1$. *Thus,* $P(\{H\}) = P(\{T\}) = 1/2$.

The following theorem gives a common method of defining a legitimate probability function.

Theorem 1.15 *Let* $S = \{s_1, ..., s_n\}$ *be a finite set. Let $\mathcal{B}$ be any sigma algebra of subsets of S. Let* $p_1, ..., p_n$ *be nonnegative numbers that sum to 1. For any* $A \in \mathcal{B}$, *define* $P(A)$ *by*

$$P(A) = \sum_{\{i: s_i \in A\}} p_i.$$

Then, P is a probability function on $\mathcal{B}$.

Proof. We need to check whether the function $P(\cdot)$ satisfies the Kolmogorov Axioms. For any $A \in \mathcal{B}$, $P(A) \geq 0$. Thus, Axiom 1 is satisfied. Axiom 2 is also true, because $P(S) = \sum_{\{i:s_i \in S\}} p_i = 1$. Let $A_1, ..., A_k$ be pairwise disjoint sets. Then,

$$P(\cup_{i=1}^k A_i) = \sum_{\{j:s_j \in \cup_{i=1}^k A_i\}} p_j = \sum_{i=1}^k \sum_{\{j:s_j \in A_i\}} p_j = \sum_{i=1}^k P(A_i),$$

where the second equality is true because A_i are disjoint. Thus, Axiom 3 is satisfied. ■

Example 1.16 *Let* $S = \{\{HH\}, \{HT\}, \{TH\}, \{TT\}\}$ *and* $A = \{\{HH\}, \{HT\}, \{TH\}\}$. *Then,* $P(A) = 3/4$.

1.2.2 The Calculus of Probabilities

Theorem 1.17 *If P is a probability function and A is any set in $\mathcal{B}$, then*
a. $P(\emptyset) = 0$.
b. $P(A) \leq 1$.
c. $P(A^c) = 1 - P(A)$.

Proof. The sets $A$and A^c form a partition of the sample space, that is, $S = A \cup A^c$. Therefore,

$$P(A \cup A^c) = P(S) = 1 \tag{1}$$

by Axiom 2. Since A and A^c are disjoint, Axiom 3 yields

$$P(A \cup A^c) = P(A) + P(A^c). \tag{2}$$

Combining (1) and (2) gives c. Since $P(A^c) \geq 0$, b follows from c. Relation a follows from

$$P(S) = P(S \cup \emptyset) = P(S) + P(\emptyset) = 1.$$

■

Theorem 1.18 *If P is a probability function and A and B are any sets in $\mathcal{B}$, then*
a. $P(B \cap A^c) = P(B) - P(A \cap B)$.
b. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
c. If $A \subset B$, *then* $P(A) \leq P(B)$.

Proof. For any sets A and B,

$$B = \{B \cap A\} \cup \{B \cap A^c\}$$

and the two sets on the RHS of this equation are disjoint. Thus,

$$P(B) = P(B \cap A) + P(B \cap A^c),$$

proving part a. Using the identity relation

$$A \cup B = A \cup \{B \cap A^c\}$$

and the fact that A and $\{B \cap A^c\}$ are disjoint, we have

$$P(A \cup B) = P(A) + P(B \cap A^c) = P(A) + P(B) - P(A \cap B),$$

where the second equality employs part a.
If $A \subset B$, then $A \cap B = A$. Therefore, part a gives

$$0 \leq P(B \cap A^c) = P(B) - P(A),$$

establishing part c. ■

Bonferroni's inequality: Part b gives Bonferroni's inequality $P(A \cap B) \geq P(A) + P(B) - 1$.

Theorem 1.19 *If P is a probability function, then*
a. $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ *for any partition $C_1, C_2, \ldots$ of S.*
b. (Boole's inequality) $P\left(\cup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$ *for any sets $A_1, A_2, \ldots$*

Proof. Since $C_1, C_2, \ldots$ form a partition of S, we have

$$A = A \cap S = A \cap \left(\cup_{i=1}^{\infty} C_i\right) = \cup_{i=1}^{\infty}(A \cap C_i).$$

Since $A \cap C_i$ are disjoints, $P(A) = P(\cup_{i=1}^{\infty}(A \cap C_i)) = \sum_{i=1}^{\infty} P(A \cap C_i)$, establishing part a. Define

$$A_1^* = A_1, \ A_i^* = A_i \backslash \left(\cup_{j=1}^{i-1} A_j\right), \ i = 2, 3, \ldots,$$

where $A \backslash B = A \cap B^c$. Then, $\cup_{i=1}^{\infty} A_i^* = \cup_{i=1}^{\infty} A_i$, A_i^* are disjoint and $A_i^* \subset A_i$. Thus,

$$P\left(\cup_{i=1}^{\infty} A_i\right) = P\left(\cup_{i=1}^{\infty} A_i^*\right) = \sum_{i=1}^{\infty} P(A_i^*) \leq \sum_{i=1}^{\infty} P(A_i).$$

■

A general version of Bonferroni's inequality can be derived form Boole's inequality. Boole's inequality gives

$$P\left(\cup_{i=1}^{n} A_i^c\right) \leq \sum_{i=1}^{n} P(A_i^c),$$

which is equivalent to

$$1 - P\left(\cap_{i=1}^{n} A_i\right) \leq n - \sum_{i=1}^{n} P(A_i).$$

This becomes on rearrangement

$$P\left(\cap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1).$$

1.2.3 Counting

Example 1.20 *(Lottery) For a number of years the New York state lottery operated according to the following scheme. From the numbers 1, 2, ... , 44, a person may pick any six for her ticket. The winning number is then decided by randomly selecting six numbers from the forty-four. To be able to calculate the probability of winning we first must count how many different groups of six numbers can be chosen from the forty-four.*

Example 1.21 *(Tournament) In a single-elimination tournament, such as the U.S. Open tennis tournament, players advance only if they win (in contrast to double-elimination or round-robin tournaments). If we have 16 entrants, we might be interested in the number of paths a particular player can take to victory, where a path is taken to mean a sequence of opponents.*

Theorem 1.22 *(Fundamental theorem of counting) If a job consists of k separate tasks, the $i-$th of which can be done in n_i ways, $i = 1, ..., k$, then the entire job can be done in $n_1 \times n_2 \times ... \times n_k$ ways.*

Proof. If suffices to prove this for $k = 2$. The first task can be done in n_1 ways, and for each of these ways we have n_2 choices for the second task. Thus, we can do the job in $n_1 \times n_2$ ways. ■

Example 1.23 *(Lottery) 1. Ordered, without replacement*

$$44 \times 43 \times 42 \times 41 \times 40 \times 39.$$

2. Ordered, with replacement

$$44 \times 44 \times 44 \times 44 \times 44 \times 44.$$

3. Unordered, without replacement

$$\frac{44 \times 43 \times 42 \times 41 \times 40 \times 39}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = \begin{pmatrix} 44 \\ 6 \end{pmatrix} = \frac{44!}{6!38!}.$$

4. Unordered, with replacement

$$\frac{49!}{6!43!}$$

Consider the problem of choosing r numbers from the set $\{a_1, a_2, ..., a_n\}$ with replacements. Chosen samples are $(a_{i_1}, ..., a_{i_r})$ with each $a_{i_k} \in \{a_1, ..., a_n\}$. Let r_k be the number of occurrences of a_k. Then, $r_1 + ... + r_n = r$. Thus, we must count the number of nonnegative integer solutions $(r_1, ..., r_n)$ of the equation $r_1 + ... + r_n = r$.

Let $n = 3$ and $r = 4$. Typical, chosen samples are:

$\| \quad * \quad \|$	$\| \quad * \quad * \quad * \quad \|$	$\|$		$\| \quad * \quad \|$	$* \quad \|$	$* \quad * \quad \|$	
$r_1 = 1$	$r_2 = 0$	$r_3 = 3$	.	$r_1 = 1$	$r_2 = 1$	$r_1 = 2$	.
(sample $a_1a_3a_3a_3$)				(sample $a_1a_2a_3a_3$)			

The number of chosen samples are $\begin{pmatrix} \#\text{ of } * + \#\text{ of } \mid \\ 4 \end{pmatrix} = \begin{pmatrix} n+r-1 \\ r \end{pmatrix}$ or $\begin{pmatrix} \#\text{ of } * + \#\text{ of } \mid \\ 2 \end{pmatrix} = \begin{pmatrix} n+r-1 \\ n-1 \end{pmatrix}$.

Table: Number of possible arrangements of size r from n objects

	Without replacement	With replacement
Ordered	$\frac{n!}{(n-r)!}$	n^r
Unordered	$\begin{pmatrix} n \\ r \end{pmatrix}$	$\begin{pmatrix} n+r-1 \\ r \end{pmatrix} = \begin{pmatrix} n+r-1 \\ n-1 \end{pmatrix}$

1.2.4 Enumerating Outcomes

Suppose that $S = \{s_1, ..., s_N\}$ and that $P(s_i) = 1/N$ for all i. Then, we have for any event A

$$P(A) = \sum_{s_i \in A} P(\{s_i\}) = \sum_{s_i \in A} \frac{1}{N} = \frac{\#\ of\ elements\ in\ A}{\#\ of\ elements\ in\ S}.$$

Example 1.24 *Consider choosing a five-card poker hand from a standard deck of 52 playing cards. Obviously, we are sampling without replacement from the deck. For this example we use the unordered outcomes, so the sample space consists of all the five-card hands (outcomes) that can be chosen from the 52-card deck. There are* $\begin{pmatrix} 52 \\ 5 \end{pmatrix}$*=2,598,960 possible hands. If the deck is well shuffled and the cards are randomly dealt, it is reasonable to assign probability 1/2,598,960 to each possible hand. What is the probability of having four aces? How many different hands are there with four aces? If we specify that four of the cards are aces, then there are 48 different ways of specifying the fifth card. Thus,*

$$P(four\ aces) = \frac{48}{2,598,960} < \frac{1}{50,000}.$$

What is the probability of having four of a kind (e.g., all queens, all aces, etc.)? There are 13 ways to specify which denomination there will be four of. After we specify these four cards, there are 48 ways of specifying the fifth. Thus, the total number of hands with four of a kind is 13×48 and

$$P(four\ of\ a\ kind) = \frac{13 \times 48}{2,598,960}.$$

1.3 Conditional Probability and Independence

Example 1.25 *Four cards are dealt from the top of a well-shuffled deck. What is the probability that they are the four aces? We can calculate this probability by the methods of the previous section. The number of distinct groups of four cards is* $\binom{52}{4} = 270,725$. *the probability of being dealt all four aces is* $1/270,725$. *We can also calculate this probability by an "updating" argument, as follows. The probability that the first card is an ace is 4/52. Given that the first card is an ace, the probability that the second card is an ace is 3/51 . Continuing this argument, we get the desired probability as*

$$\frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49} = \frac{1}{270,725.}$$

Definition 1.26 *If A and B are events in S, and $P(B) > 0$, then the conditional probability of A given B, written $P(A \mid B)$, is*

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

Example 1.27 *Four cards will again be dealt from a well-shuffled deck, and we now calculate*

$$P(4 \text{ aces in } 4 \text{ cards} \mid i \text{ aces in } i \text{ cards}), \; i = 1, 2, 3.$$

According to the definition of conditional probability, we have

$$\begin{aligned} P(4 \text{ aces in } 4 \text{ cards} \quad & \mid \quad i \text{ aces in } i \text{ cards}) \\ &= \frac{P(\{4 \text{ aces in } 4 \text{ cards}\} \cap \{i \text{ aces in } i \text{ cards}\})}{P(\{i \text{ aces in } i \text{ cards}\})} \\ &= \frac{P(\{4 \text{ aces in } 4 \text{ cards}\})}{P(\{i \text{ aces in } i \text{ cards}\})} \\ &= \frac{1/\binom{52}{4}}{\binom{4}{i}/\binom{52}{i}}. \end{aligned}$$

Theorem 1.28 *(Bayes rule) Let $A_1, A_2, ...$ be a partition of S, and let B any set. Then, for each $i = 1, 2, ...,$*

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B \mid A_j)P(A_j)}.$$

Definition 1.29 *Two events, A and B, are statistically independent if*

$$P(A \cap B) = P(A)P(B).$$

Many gambling games provide models of independent events. The spins of a roulette wheel and the tosses of a pair of dice are both series of independent events.

Example 1.30 *(Rolling a die)*

$$\begin{aligned} P(\textit{at least 1 six in 4 rolls}) &= 1 - P(\textit{no six in 4 rolls}) \\ &= 1 - \Pi_{i=1}^{4} P(\textit{no six on roll } i) \\ &= 1 - (\frac{5}{6})^4 = 0.518. \end{aligned}$$

The second equality holds due to independence of the rolls.

Theorem 1.31 *If A and B are independent events, then the following pairs are also independent:*
a. A and B^c,
b. A^c and B,
c. A^c and B^c.

Proof. Theorem 1.18 gives

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) \\ &= P(A)P(B^c), \end{aligned}$$

proving part a. The rest are left for exercises. ■

Example 1.32 *(Letters) Let the sample space S consist of the 3! permutations of the letters a, b, and c along with the three triples of each letter. Thus,*

$$\begin{aligned} S &= \{aaa, bbb, ccc, \\ &\quad abc, bac, cba, \\ &\quad acb, bac, cab\}. \end{aligned}$$

Each element of S has probability 1/9. Define

$$A_i = \{i - \textit{th place in the triple is occupied by } a\}.$$

Then,

$$P(A_i) = \frac{1}{3},\ i = 1, 2, 3$$

and

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P((A_2 \cap A_3) = \frac{1}{9},$$

so the A_i are pairwise independent. But

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{9} \neq P(A_1)P(A_2)P(A_3).$$

Here, the events are pairwise independent, but they are not indeed independent.

Definition 1.33 *A collection of events $A_1, ..., A_n$ are mutually independent if for any subcollection $A_{i_1}, ..., A_{i_k}$, we have*

$$P(\cap_{j=1}^k A_{i_j}) = \Pi_{j=1}^k P(A_{i_j}).$$

Example 1.34 *(Three coin tosses - I) Consider the experiment of tossing a coin three times. The sample space for this experiment has eight points, namely,*

$$\{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}.$$

Let H_i, $i = 1, 2, 3$, denote the event that the $i-th$ toss is a head. For example,

$$H_1 = \{HHH, HHT, HTH, HTT\}.$$

We find that

$$P(H_1) = P(H_2) = P(H_3) = \frac{1}{2}.$$

For this experiment,

$$P(H_1 \cap H_2 \cap H_3) = P(\{HHH\}) = \frac{1}{8} = P(H_1)P(H_2)P(H_3)$$

and

$$P(H_1 \cap H_2) = P(\{HHH, HHT\}) = \frac{2}{8} = P(H_1)P(H_2).$$

The latter relation is also true for the other two pairs. Thus, H_1, H_2 and H_3 are mutually independent.

1.4 Random Variables

Definition 1.35 *A random variable is a function from a sample space S into the real numbers.*

Example 1.36

Experiment	*Random variable*
Toss two dice	*$X =$sum of the numbers*
Toss a coin 25 times	*$X =$number of heads in 25 tosses*

Suppose we have a sample space

$$S = \{s_1, ..., s_n\}$$

with a probability function P and we define a random variable X with range $\mathcal{X} = \{x_1, ..., x_m\}$. We can define a probability function P_X on X as

$$P_X(X = x_i) = P(\{s_j \in S : X(s_j) = x_i).$$

The function P_X, is an induced probability function on $\mathcal{X}$, defined in terms of the original function P.

A note on notation: Random variables will always be denoted with uppercase letters and the realized values of the variable (or its range) will be denoted by the corresponding lowercase letters. Thus, the random variable X can take the value x.

Example 1.37 *(Three coin tosses) Define the random variable X to be the number of heads obtained in the three tosses. A complete enumeration of the value of X for each point in the sample space is*

s	HHH	HHT	HTH	THH	TTH	THT	HTT	TTT
$X(s)$	3	2	2	2	1	1	1	0

The induced probability function is given by

x	*0*	*1*	*2*	*3*
$P_X(X = x)$	*1/8*	*3/8*	*3/8*	*1/8*

If X is uncountable, we define the induced probability function, P_X, similarly. For any set $A \subset \mathcal{X}$,

$$P_X(X \in A) = P(\{s \in S : X(s) \in A).$$

1.5 Distribution Functions

Definition 1.38 *The cumulative distribution function or cdf of a random variable X, denoted by $F_X(x)$, is defined by*

$$F_X(x) = P_X(X \leq x), \ for \ all \ x.$$

Example 1.39 *(Three coin tosses) Define the random variable X to be the number of heads obtained in the three tosses. The cdf of X is*

$$F_X(x) = \begin{cases} 0 & if \ -\infty < x < 0 \\ 1/8 & if \ 0 \leq x < 1 \\ 1/2 & if \ 1 \leq x < 2 \\ 7/8 & if \ 2 \leq x < 3 \\ 1 & if \ 3 \leq x < \infty \end{cases}$$

The function F_X is continuous when a point is approached from the right. That is,

$$\lim_{x \to x_0, x \geq x_0} F_X(x) = F_X(x_0).$$

The property of right-continuity is a consequence of the definition of the cdf. In contrast, if we had defined $F_X(x) = P_X(X < x)$ (note strict inequality), F_X would then be left-continuous. The size of the jump at any point x is equal to $P(X = x)$.

Theorem 1.40 *The function $F(x)$ is a cdf if and only if the following three conditions hold:*
a. $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to \infty} F(x) = 1$.
b. $F(x)$ is a nondecreasing function of x.
c. $F(x)$ is right-continuous; that is, for every number x_0, $\lim_{x \to x_0, x \geq x_0} F(x) = F(x_0)$.

Example 1.41 *(Continuous cdf) An example of a continuous cdf is the function*

$$F(x) = \frac{1}{1 + e^{-x}}.$$

For this function we have

$$a. \lim_{x\to-\infty} F(x) = 0 \text{ and } \lim_{x\to\infty} F(x) = 1.$$

$$b. \frac{d}{dx}F(x) = \frac{e^{-x}}{(1 + e^{-x})^2} > 0,$$

showing that F is increasing. The function F is continuous. Thus, F is a cdf.

Definition 1.42 *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.*

F_X completely determines the probability distribution of a random variable X. This is true if $P(X \in A)$ is defined only for events A in $\mathcal{B}^1$, the smallest sigma algebra containing all the intervals of real numbers of the form (a, b), $[a, b)$, $(a, b]$, and $[a, b]$.

Definition 1.43 *The random variables X and Y are identically distributed if, for every set $A \in \mathcal{B}^1$, $P(X \in A) = P(Y \in A)$.*

Theorem 1.44 *The following two statements are equivalent:*
a. The random variables X and Y are identically distributed.
b. $F_X(x) = F_Y(x)$ for every x.

Proof. Showing that a implies b is easy. But showing the converse is more difficult. See Chung (1974). ■

1.6 Density and Mass Functions

Definition 1.45 *The probability mass function (pmf) of a discrete random variable X is given by*

$$f_X(x) = P(X = x) \text{ for all } x.$$

Example 1.46 *(Geometric probabilities) Suppose we do an experiment that consists of tossing a coin until a head appears. Let p = probability of a head on any given toss, and define a random variable X = number of tosses required to get a head. Then, for any x = 1,2, ...,*

$$f_X(x) = P(X = x) = (1 - p)^{x-1}p.$$

This is called the geometric distribution. The cdf of this distribution is given by

$$P(X \le b) = \sum_{k=1}^{b} f_X(k) = F_X(b).$$

Definition 1.47 *The probability density function or pdf, $f_X(x)$, of a continuous random variable X is the function that satisfies*

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx \text{ for all } x.$$

If X is a continuous random variable,

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b),$$

because $P(X = x) = 0$.

Example 1.48 *For the logistic distribution*

$$F(x) = \frac{1}{1 + e^{-x}}$$

and

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{e^{-x}}{(1 + e^{-x})^2}.$$

Theorem 1.49 *A function $f_X(x)$ is a pdf (or pmf) of a random variable X if and only if*
a. $f_X(x) \geq 0$ for all x.
b. $\sum_x f_X(x) = 1$ or $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

Proof. The necessity part follows from the definition of pdf. To prove the sufficiency part, let $F_X(x) = \int_{-\infty}^{x} f_X(x)dx$ and use Theorem 1.40. ■