Probability

Jan-April 2014

Chennai Mathematical Institute

Professor Nandini Kannan

Chapter 1

Axioms of Probability Theory

Probability theory provides the mathematical foundation to quantify the chance of occurrence of events. The origins of modern Probability may be traced to games of chance around the 17th century. The two French mathematicians Pascal and Fermat discovered many of the rules for computing probabilities. The rigorous and formal development of the theory of probability was presented by the Russian mathematician Andrei Kolmogorov in 1933 in his book *Foundations of the Theory of Probability*.
To introduce the mathematical foundations of probability, we need to review some basic set theory concepts.

1.1 Experiments, Sample Space and Events

An **experiment** is a "repeated" process that generates observations or measurements. The outcomes of the experiment are assumed to be random. We sometimes refer to this as a ***random experiment***.

Examples:

1. Tossing a coin: the observations are Head and Tail
2. Rolling a die: the observations are 1 through six.
3. Classifying an email: the observations are "Spam" or "Good".
4. In a survey, individuals are classified according to income (low, medium, high) and gender (male, female): the observations are all possible combinations of the two categorical variables along with the response of interest.
5. Recording the survival times for patients suffering from cancer: the measurements are the actual lifetimes of the patients, which could be any positive real number.

Definition: The **Sample Space** $\mathcal{S}$ is the set of all possible outcomes of a given experiment. Each outcome in a sample space is called an **element** or a **sample point**.

Examples:

1. Experiment: Tossing a coin

   $\mathcal{S}$ = {H,T}

2. Experiment: Rolling a die

   $\mathcal{S}$ = {1,2,. . . ,6}

3. Experiment: Rolling a die twice

   $$\mathcal{S} = \left\{ \begin{array}{llll} (1,1) & . & . & (1,6) \\ (2,1) & . & . & (2,6) \\ . & . & . & . \\ . & . & . & . \\ (6,1) & . & . & (6,6) \end{array} \right\}$$

4. Experiment:Recording the number of cases of malaria in Chennai

   $\mathcal{S} = \{0, 1, \ldots\}$.

5. Experiment: Recording Profits for a company

   $\mathcal{S} = \{(-\infty, \infty)\}$.

Observe that in the first three examples, we can list all possible outcomes of the experiment. Such sample spaces are called **finite** and **discrete**.

In example 4, the sample space is **infinite** but there is an ordering of the possible outcomes, i.e. all whole numbers. Such a sample space is **discrete, countable**.

Example 5 has a sample space which is infinite and an interval. Such sample spaces are usually referred to as **continuous** sample spaces.

Definition: An **event** is any outcome or collection of outcomes of a given experiment. (In other words, an event is a subset of the sample space.)
Events are denoted by capital letters like A, B, E etc.

For each of the above examples, we can define several events.

1. Experiment: Rolling a die
$\mathcal{S}$ = {1,2,. . . ,6}
Event A: observing an odd number
A = {1,3,5}
Event B: observing a number that is a multiple of 5
B = {5}

2. Experiment: Rolling a die twice

$$\mathcal{S} = \left\{ \begin{array}{llll} (1,1) & . & . & (1,6) \\ (2,1) & . & . & (2,6) \\ . & . & . & . \\ . & . & . & . \\ (6,1) & . & . & (6,6) \end{array} \right\}$$

Event E: sum of the two rolls is 7.
E = { (1,6), (2,5), (3,4), (4,3), (5,2), (6,1)}

3. Experiment: Recording profits for a company
$\mathcal{S} = \{(-\infty, \infty)\}$.
Event C: Profits above 10 crores.
C = $\{(10, \infty)\}$

Definition: An event which has only one outcome in it is called **simple**, otherwise it is known as a **composite** event.

1.1.1 Algebra of Sets

Let **A** and **B** be two events defined on the same sample space.

Definition: The **Union** of **A** and **B**, denoted by $A \cup B$ is the event consisting of all outcomes which are in A **or** in B **or** in both.

Definition: The **intersection** of A and B, denoted by $A \cap B$ is the event consisting of all outcomes that are in both A **and** in B.

Definition: Event A is said to be a **subset** of B if every outcome in A is also an outcome in B. We write this as $A \subset B$.

Definition: If $A \subset B$ and $B \subset A$, then the two sets have the same outcomes and are said to be **equal**. We write $A = B$.

Definition: Two events are said to be **mutually exclusive** or **disjoint** if they cannot occur simultaneously. When two events are mutually exclusive, their intersection is the **empty** or **null** set ϕ.

Definition: The **complement** of an event A, denoted by A^c or $\bar{A}$ is the set of all outcomes in the sample space that do not belong to A.

Examples

1. Rolling a die.

   A: observing an odd number = {1,3,5}

B: observing an even number = {2,4,6}

C: observing a multiple of 3 = {3,6}

We have

$A \cup B$ = {1,2,3,4,5,6}

$A \cap C$ = {3}

$B \cup C$ = {2,3,4,6}

$A \cap B = \phi$, i.e. A and B are mutually exclusive.

2. Consider the sample space $\mathcal{S} = \mathcal{R}$. Define $A = \{x : 0 \leq x \leq 1\}, B = \{x : 0 < x < 1\}, C = [2,3] = \{x : 2 \leq x \leq 3\}, D = \mathcal{R}^+$.

   We have

   $B \subseteq A$.

   $A \cap B = (0,1)$

   $A \cup C = [0,1] \cup [2,3]$

   $A \cap C = \phi$.

1.1.2 Properties of Set Operations

Let $\mathcal{S}$ be the sample space and A, B, C events defined on $\mathcal{S}$.

Theorem 1.1.1. *1. Commutative Property*

$$\begin{aligned} A \cup B &= B \cup A \\ A \cap B &= B \cap A. \end{aligned}$$

2. Associative Property

$$\begin{aligned} A \cup (B \cup C) &= (A \cup B) \cup C \\ A \cap (B \cap C) &= (A \cap B) \cap C. \end{aligned}$$

3. Distributive Property

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C). \end{aligned}$$

4. DeMorgan's Law

$$\begin{aligned} (A \cup B)^c &= A^c \cap B^c \\ (A \cap B)^c &= A^c \cup B^c. \end{aligned}$$

Proof of 4, Part 1

We need to show $L = (A \cup B)^c$ and $R = A^c \cap B^c$ are equal.
Let x be an outcome in L. We have the following:

$$x \in (A \cup B)^c$$
$$\Rightarrow x \notin A \cup B$$
$$\Rightarrow x \notin A \;\; and \;\; x \notin B$$
$$\Rightarrow x \in A^c \;\; and \;\; x \in B^c$$
$$\Rightarrow x \in A^c \cap B^c.$$

This proves that

$$L \subseteq R. \tag{1.1}$$

To show the reverse, let $x \in R$. Then

$$x \in A^c \cap B^c$$
$$\Rightarrow x \in A^c \;\; and \;\; x \in B^c$$
$$\Rightarrow x \notin A \;\; and \;\; x \notin B$$
$$\Rightarrow x \notin A \cup B$$
$$\Rightarrow x \in (A \cup B)^c$$

This proves

$$R \subseteq L. \tag{1.2}$$

From (1.1) and (1,2), we have

$$L = R.$$

■

The definitions of unions and intersections can be extended to infinite and uncountable collections of events.

Definition: Let $\{A_i\}$ be a collection of events. Then

$$\bigcup_{i=1}^{\infty} A_i = \{x : x \in A_i \;\; \text{for some} \;\; i\}.$$

$$\bigcap_{i=1}^{\infty} A_i = \{x : x \in A_i \;\; \text{for all} \;\; i\}.$$

If Γ is an index set then, we have

$$\bigcup_{a \in \Gamma} A_a = \{x : x \in A_a \ \text{ for some } \ a\}.$$

$$\bigcap_{a \in \Gamma} A_a = \{x : x \in A_a \ \text{ for all } \ a\}.$$

Example: Let $\mathcal{S} = \mathcal{R}$. Define $A_i = \left(1 - \frac{1}{i}, 1\right]$. Then

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

Definition: Let $\{A_i\}$ be a collection of events. The events are said to be ***pairwise disjoint*** if

$$A_i \cap A_j = \phi \qquad \forall i \neq j.$$

Definition: Let $\{A_i\}$ be a collection of pairwise disjoint events such that

$$\bigcup_{i=1}^{\infty} A_i = \mathcal{S}.$$

Then the sets $\{A_i\}$ are said to form a ***partition*** of $\mathcal{S}$.

1.2 Counting Rules

Fundamental Theorem of Counting: Consider an experiment that consists of k distinct steps. Let n_i = number of possible outcomes for the $i-$th step. Then the total number of outcomes for the combined experiment is

$$n_1 \times n_2 \times \ldots \times n_k.$$

This is also called the **Multiplication Rule**.

Example: Alphonse Bertillon, viewed as the founder of Anthropometry developed a system of identifying individuals (criminals) based on anthropometric measurements. Individuals were identified by measurements of the head and body, markings (tattoos, scars), and personality characteristics that were assumed to be unchanged during the individuals's adult life. The body measurements were broken down into three intervals: small, medium, and large.
The measurements allowed for records to be sorted, first with respect to height, then arm-span, upper body height, head length etc. A person's Bertillon configuration was an

ordered sequence of 11 letters, say

$$ssmmlslssms,$$

where a letter indicated the individual's size relative to a particular measurement. How many different Bertillon configurations are possible?
Solution: $3^{11} = 177,147$.

The Bertillon system was used in Europe before the advent of modern fingerprinting.

1.2.1 Permutations

Definition: A **permutation** is an ordered arrangement of objects.

Theorem 1.2.1. *Consider n distinct objects. The number of permutations of r objects selected from the group of n objects (repetitions not allowed) is denoted by the symbol P_r^n, where*

$$P_r^n = n\ (n-1)\ldots(n-r+1) = \frac{n!}{(n-r)!}. \tag{1.3}$$

Proof: Use the Multiplication Rule.
■

Corollary 1.2.2. *The number of permutations of the entire set of n objects is P_n^n, where*

$$P_n^n = n!. \tag{1.4}$$

Example: At a family reunion, a group of 4 families, each with 8 members, are lined up for a photograph. In how many ways can the group be arranged if members of a family must stay together?

Solution:
$4!(8!)^4$

■

Example: A three-digit number is to be formed from the digits 1 through 7, with no digit being used more than once. How many such numbers are less than 289?

Solution:
60

■

1.2.2 Permutations when the objects are not distinct

Theorem 1.2.3. *The number of permutations of n objects of which n_1 are of one kind, n_2 are of a second kind, ..., n_r are of a $r-th$ kind is*

$$\frac{n!}{n_1!n_2!\ldots n_r!}, \tag{1.5}$$

where $n_1 + \ldots + n_r = n$.

Proof: Let P denote the total number of such arrangements. For any one of these P arrangements, we have $n_1!n_2!\ldots n_r!$ ways of arranging the similar objects (if they were truly different). Therefore

$$P \times n_1!n_2!\ldots n_r! = n!$$

which implies

$$P = \frac{n!}{n_1!n_2!\ldots n_r!}.$$

■

Example: A chess tournament has 10 competitors of which 4 are Russian, 3 are from the US, 2 from the UK, and 1 from Germany. If the tournament results list just the nationalities of the players in the order in which they placed, how many outcomes are possible?

Solution: There are

$$\frac{10!}{4!3!2!1!} = 12,600 \qquad \text{possible outcomes.}$$

■

Theorem 1.2.4. *The number of ways of partitioning a set of n objects into r cells with n_1 elements in the first cell, n_2 elements in the second cell, ..., n_r elements in the $r-th$ cell is*

$$\left(\begin{array}{c} n \\ n_1, n_2, \ldots, n_r \end{array} \right) = \frac{n!}{n_1!n_2!\ldots n_r!},$$

where $n_1 + \ldots + n_r = n$.

The terms $\left(\begin{array}{c} n \\ n_1, n_2, \ldots, n_r \end{array} \right)$ are referred to as multinomial coefficients.

■

Example: The security company used by CMI employs 10 security guards. If the company policy is to have 5 of the guards on duty at the hostel, 2 of the officers working full time at the main entrance, and 3 of the officers on reserve at the main office, how many different divisions of the 10 guards into the 3 groups are possible?

Solution: There are

$$\frac{10!}{5!2!3!} = 2520 \quad \text{divisions.}$$

■

1.2.3 Combinations

In many problems, we are interested in the number of ways of selecting r objects from n without regard to order. These selections are called **combinations**.

Theorem 1.2.5. *The number of combinations of n distinct objects taken r at a time is*

$$C_r^n = \begin{pmatrix} n \\ r \end{pmatrix} = \frac{n!}{r!(n-r)!}$$

■

If the selection is done with replacement, the formulae above need to be modified.

Example: A box contains n marbles numbered $1, 2, \ldots$. A marble is drawn at random, the number noted, and the marble returned to the box. If r marbles are drawn, the sample space consist of all r-tuples $(x_1, \ldots, x_r)$, where x_i is the outcome of the $i-$th draw. The total number of outcomes is

$$n^r.$$

■

1.3 Probability

Once we have defined an experiment, obtained the associated sample space $\mathcal{S}$, and defined events of interest, we would like to assign probabilities to these events.
The **Probability** associated with an event A, denoted by **P(A)** is a quantification of our belief that the experiment will yield those particular outcomes.
How do we assign probabilities to all events (and combinations of events) in a sample space in a mathematically consistent fashion?
To answer that question, let us start with some simple intuitive examples.

Example: Consider the simple experiment of tossing a coin. We know the sample space is $\mathcal{S} = \{H, T\}$. There are only two events of interest: the two possible outcomes. These are simple events.
We wish to find the probability of the event that a Head occurs. We may start with the **classical** approach to assigning probability
We say that the chance of the coin turning up heads is 50 %. The reason for this statement is that we assume the coin is fair, there are only two possible outcomes (Head and Tail), and both are equally likely to happen. Thus classical probability assumes a structure to the process.

This idea may be extended to the case in which the experiment may result in one of N equally likely outcomes. Let $\{E_1, \ldots, E_N\}$ denote the N outcomes (simple events). We can assign a probability of $1/N$ to each E_i.
The probability of any event A may be defined as

$$P(A) = \frac{\text{number of outcomes favourable to A}}{\text{number of outcomes in the sample space}}$$

Empirical Probability: Consider again the problem of tossing a coin. Suppose we have no information on whether the coin is fair or not. To find the chance of obtaining a head, we could toss the coin a large number of times, observe the number of times a head turned up, and use the proportion of heads observed as an approximation to the chance of observing a head.

This is also called the relative frequency approach to probability. The empirical proportion should get "closer" to the true proportion as the number of replications increase.

For finite sample spaces, we can assign probabilities to all the individual outcomes. This assignment of probabilities (using either the classical or the relative frequency approach) satisfies the following:

- The probability of any event lies between 0 and 1.
- The sum of the probabilities of all possible outcomes equals 1.

- Probability of an event A may be obtained by adding the probabilities of the simple events that make up A.

This approach is adequate for many of the standard examples involving games of chance (poker hands, roll of dice etc).

The theory of discrete sample spaces provides the foundation for the development of many interesting topics in mathematical probability. However, the theory is not adequate to provide a rigorous treatment for two kinds of problems:

1. Experiments involving an infinitely repeated operation such as infinite sequence of tosses of a coin.

2. Problems involving an "infinitely fine" operation, such as the random drawing of a point from a segment.

In order to develop a mathematically complete description of probability, we need to introduce some ideas from Measure Theory.

1.4 A Trip to Measure Land

Let Ω be an abstract space, viz, a nonempty set of elements called points and denoted generically by ω.
In probability theory, Ω represents the sample space, and points refer to outcomes. A subset of Ω is an event.
We will assume Ω is an arbitrary nonempty space,

Definition: A collection (or class) $\mathcal{F}$ of subsets of Ω is called an ***Algebra*** or a ***Field*** if it satisfies the following three properties:

1. $\phi \in \mathcal{F}$.

2. If $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$.

3. If $A_1, A_2 \in \mathcal{F} \Rightarrow A_1 \cup A_2 \in \mathcal{F}$

A field is said to be **closed** under the formation of complements and finite unions.

As a consequence of the definition, we have the following additional properties:

$1'$ $\Omega \in \mathcal{F}$

$2'$ If $A_1, A_2 \in \mathcal{F} \Rightarrow A_1 \cap A_2 \in \mathcal{F}$

Fill in the Proof!

■

Example: Let $\Omega = \{1, 2, 3, \ldots\}$. Let $\mathcal{C}$ be the class of subsets C of Ω such that either C contains a finite number of points or C^c contains a finite number of points. Show that $\mathcal{C}$ is a field.

Definition: The ***trivial field*** is $\mathcal{F} = \{\Phi, \Omega\}$.

Definition: The set of all possible subsets of Ω is called the ***power set*** or ***power class***, and is denoted by $\mathcal{P}(\mathcal{S})$. The power set is necessarily a field.

Definition: A collection (or class) $\mathcal{F}$ of subsets of Ω is called a $\sigma-$ ***Algebra*** or a $\sigma-$ ***Field*** if it satisfies the following three properties:

1. $\phi \in \mathcal{F}$.
2. If $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$.
3. If $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_i A_i \in \mathcal{F}$

A $\sigma-$ field is said to be **closed** under the formation of complements and countable unions. Every $\sigma-$ field is a field but the converse is not true.

Example: Let $\Omega = \{1, 2, 3, \ldots\}$. Let $\mathcal{C}$ be the class of subsets C of Ω such that either C contains a finite number of points or C^c contains a finite number of points. We have shown that $\mathcal{C}$ is a field. Let

$$A_i = \{2i\}.$$

Each $A_i \in \mathcal{C}$. Let

$$C = \bigcup_i A_i = \{2, 4, 6, \ldots\}.$$

We have

$$C^c = \{1, 3, 5, \ldots\},$$

which is also a set containing infinitely many points. Therefore $C \notin \mathcal{C}$, which implies $\mathcal{C}$ is **NOT** a $\sigma-$ field. ■

Example: Consider the class $\mathcal{F}$ of all intervals of the form $(a, b) : a, b \in \mathcal{R}; a < b$ and the set ϕ.
We have

$$(a,b) \cap (c,d) = \begin{cases} \phi, & \text{if } a < b < c \text{ or } c < d < a < b; \\ (c,b), & \text{if } a < c < b < d; \\ (a,d), & \text{if } c < a < d < b; \\ (c,d), & \text{if } a < c < d < b; \\ (a,b), & \text{if } c < a < b < d. \end{cases}$$

So $\mathcal{F}$ is closed under finite intersection. It is **not** closed under complementation or unions.

$$(a, b)^c = (-\infty, a] \cup [b, \infty) \notin \mathcal{F}.$$

Further $(a, b) \cup (c, d)$ is not an interval if $a < b < c < d$ or $c < d < a < b$. ■

A field is closed under the finite set-theoretic operations and a σ-field is closed under the countable ones. In general, we start with a small class $\mathcal{A}$. However, we may find that constructions involving finite and countable operations may lead to sets outside this initial class (See previous example).
We therefore expand to a class of sets that

1. Contains $\mathcal{A}$ and
2. is a σ-field.

We also want this class to be as small as possible. The next result allows us to construct this "small" class.

Theorem 1.4.1. *The intersection of an arbitrary number of σ-fields is also a σ-field.*

Proof: Use the definition.

Given any class $\mathcal{A}$, the minimal σ-field containing $\mathcal{A}$ is denoted by $\sigma(\mathcal{A})$. It is the intersection of all the σ-fields containing $\mathcal{A}$. It is also called the σ-field generated by $\mathcal{A}$. It is minimal in the sense that it is contained in every σ-field that contains $\mathcal{A}$.

Example: The Borel Field Let $\Omega = \mathcal{R}$. Consider the class $\mathcal{A}$ of all intervals of the form $(-\infty, x), x \in \mathcal{R}$. This class is closed under finite intersections, but not under complementation nor under countable intersections (Need to show this!).
Let $\mathcal{B} = \sigma(\mathcal{A})$ be the minimal σ-field containing $\mathcal{A}$. Clearly $\mathcal{B}$ contains intervals of the form $[x, \infty)$ which are complements of sets in $\mathcal{A}$. It also contains intervals of the form

$$(-\infty, a] = \bigcap_{n=1}^{\infty} \left(-\infty, a + \frac{1}{n}\right).$$

$$(a, \infty) = (-\infty, a]^c.$$

$$(a, b) = (-\infty, b) \cap (a, \infty), a < b.$$

It also includes intervals of the form $(a, b], [a, b)$.
$\mathcal{B}$ is called the ***Borel field*** of subsets of $\mathcal{R}$. The sets of $\mathcal{B}$ are called Borel sets. $\mathcal{B}$ contains all the subsets of $\mathcal{R}$ encountered in normal probability. It is large enough for all practical purposes. It does NOT, however, contain all subsets of $\mathcal{R}$

1.5 The Probability Function

Kolmogorov's Axioms of Probability

Definition: Let $\mathcal{S}$ be a sample space, and let $\mathcal{B}$ be a σ-field $\mathcal{B}$ on $\mathcal{S}$. A **probability function** is a set function with domain $\mathcal{B}$ that satisfies the following three axioms:

A1. $P(A) \geq 0 \qquad \forall A \in \mathcal{B}$.

A2. $P(\mathcal{S}) = 1$.

A3. If $A_1, A_2, ... \in \mathcal{B}$, A_i's are pairwise disjoint, then

$$P\left(\bigcup_1^\infty A_i\right) = \sum_1^\infty P(A_i).$$

A3 is known as the **countable additivity** property.

Remarks:

- These axioms simply state the conditions $P(.)$ must satisfy, but do not provide any guidelines as to how $P(.)$ must be selected. For any sample space, many different probability functions can be defined.
- Sets in the σ-field $\mathcal{B}$ are the only sets for which $P(.)$ is defined. $\mathcal{B}$ cannot always be taken to be the power class because, sometimes, one cannot define a probability function consistent with the above axioms.

Definition: If P is a probability function on a σ-field $\mathcal{B}$ in $\mathcal{S}$, then $(\mathcal{S}, \mathcal{B}, P)$ is called a **probability space**.

Example: Consider a pointer free to spin about the centre of a circle. Each point on the circumference is a possible outcome of the experiment.

$$\mathcal{S} = \{x : 0 < x < 2\pi r\},$$

where r is the radius of the circle.
Events of interest are those in which the pointer stops at a point belonging to a specified arc.
The probability could correspond to the area of an arc.
■

Example: Let $\mathcal{S} = \mathcal{R}^+$, $\mathcal{B}$ the Borel σ-field on $\mathcal{S}$. For each interval $I \subseteq \mathcal{S}$, define

$$P(I) = \int_I e^{-x} dx.$$

Clearly

1. $P(I) \geq 0$ for all intervals I.
2. $P(\mathcal{S}) = \int\limits_0^\infty e^{-x} dx = 1$.
3. $P(\cup_i A_i) = \int\limits_{\cup_i A_i} e^{-x} dx = \sum_i \int\limits_{A_i} e^{-x} dx$ if $A_i's$ are disjoint intervals.

■

1.5.1 Properties of the Probability Function

Theorem 1.5.1.

$$P(\phi) = 0.$$

Proof:

$$\begin{aligned}
&\mathcal{S} = \phi \cup \mathcal{S}; \qquad \phi \cap \mathcal{S} = \phi \\
&\Rightarrow P(\mathcal{S}) = P(\mathcal{S}) + P(\phi) \qquad \text{(Property A3)} \\
&\Rightarrow 1 = 1 + P(\phi) \qquad \text{(Property A2)} \\
&\Rightarrow P(\phi) = 0.
\end{aligned}$$

■

Theorem 1.5.2. *P is finitely additive.*

Proof: Fill in the steps.

■

Theorem 1.5.3. *If $A \in \mathcal{B}$, then*

$$P(A^c) = 1 - P(A).$$

Proof:

$$A \cup A^c = \mathcal{S}; \qquad A \cap A^c = \phi$$

$$
\begin{aligned}
&\Rightarrow P(A) + P(A^c) = P(\mathcal{S}) \qquad \text{(Property A3)}\\
&\Rightarrow P(A) + P(A^c) = 1 \qquad \text{(Property A2)}\\
&\Rightarrow P(A^c) = 1 - P(A).
\end{aligned}
$$

■

Theorem 1.5.4. *If $A, B \in \mathcal{B}$, then*

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

<u>**Proof:**</u>

$$
\begin{aligned}
&(A \cap B) \cup (A \cap B^c) = A; \qquad (A \cap B) \cap (A \cap B^c) = \phi\\
&\Rightarrow P(A \cap B) + P(A \cap B^c) = P(A). \qquad \text{(Property A3)}
\end{aligned}
$$

■

Theorem 1.5.5. *Monotonicity Property. If $A, B \in \mathcal{B}$ and $A \subset B$, then*

$$P(A) \leq P(B)$$

<u>**Proof:**</u> We may write

$$
\begin{aligned}
&B = A \cup (B \cap A^c); \qquad A \cap (B \cap A^c) = \phi\\
&\Rightarrow P(B) = P(A) + P(B \cap A^c) \qquad \text{(Property A3)}\\
&\Rightarrow P(B) \geq P(A) \qquad \text{since} \;\; P(B \cap A^c) \geq 0.
\end{aligned}
$$

■

Theorem 1.5.6. *For every $A \in \mathcal{B}$,*

$$0 \leq P(A) \leq 1.$$

<u>**Proof:**</u> Axiom A1 states that $P(A) \geq 0$. We have

$$
\begin{aligned}
&A \subseteq \mathcal{S}\\
&\Rightarrow P(A) \leq P(\mathcal{S}) \qquad \text{(Theorem 1.5.5)}\\
&\Rightarrow P(A) \leq 1. \qquad \text{(Property A2)}
\end{aligned}
$$

■

Theorem 1.5.7. *Additive Rule. If $A, B \in \mathcal{B}$, then*

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

<u>Proof:</u> We may write

$$\begin{aligned}
&A \cup B = A \cup (B \cap A^c); \qquad A \cap (B \cap A^c) = \phi \\
&\Rightarrow P(A \cup B) = P(A) + P(B \cap A^c) \qquad \text{(Property A3)} \\
&\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B). \qquad \text{(Theorem 1.5.4)}
\end{aligned}$$

■

We may extend the previous result.

Theorem 1.5.8. *The Inclusion-Exclusion Formula. Let $A_1, A_2, \ldots, A_n \in \mathcal{B}$. Then*

$$\begin{aligned}
P\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P\left(A_i \bigcap A_j\right) \\
&+ \sum_{i<j<k} P\left(A_i \bigcap A_j \bigcap A_k\right) + \ldots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} A_i\right).
\end{aligned}$$

<u>Proof:</u> The general proof is obtained by induction.
■

Bonferroni's Inequality provides a bound for the probability of the intersection of two events. It is used in the construction of simultaneous confidence intervals. The result is a simple consequence of the additive rule.

Theorem 1.5.9. *Bonferroni's Inequality. If $A, B \in \mathcal{B}$, then*

$$P(A \cap B) \geq P(A) + P(B) - 1.$$

<u>Proof:</u> From the additive rule, we have

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

Since $P(A \cup B) \leq 1$, we have

$$P(A \cap B) \geq P(A) + P(B) - 1.$$

■

Theorem 1.5.10. *Law of Total Probability . If $A \in \mathcal{B}$, and $\{C_i\}$ is a partition of $\mathcal{S}$, such* *that* $C_i \in \mathcal{B} \ \ \forall \ \ i$, *then*

$$P(A) = \sum_{i=1}^{\infty} P(A \cap C_i).$$

<u>Proof:</u> Since $C_1, C_2, \ldots$ form a partition, we have

$$\bigcup_{i=1}^{\infty} C_i = \mathcal{S} \qquad C_i \cap C_j = \phi, i \neq j.$$

Further

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

Therefore

$$P(A) = P\left(\bigcup_{i=1}^{\infty} (A \cap C_i) \right) = \sum_{i=1}^{\infty} P(A \cap C_i).$$

■

Theorem 1.5.11. *Boole's Inequality. Let* $A_1, A_2, \ldots, A_n \in \mathcal{B}$. *Then*

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

<u>Proof:</u> Define

$$\begin{aligned}
B_1 &= A_1 \\
B_2 &= A_2 - A_1 = A_2 \bigcap A_1^c \\
&\vdots \\
B_k &= A_k - \left(\bigcup_{j=1}^{k-1} A_j \right).
\end{aligned}$$

Clearly $B_i \in \mathcal{B}$ and $B_i \bigcap B_j = \phi, i \neq j$. We also have

$$B_i \subseteq A_i \qquad \forall i.$$

This implies, from Theorem 1.4.6,

$$P(B_k) \leq P(A_k).$$

We also see that

$$\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i.$$

Therefore

$$\begin{aligned}
&P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{n} B_i\right) = \sum_{i=1}^{n} P(B_i) \qquad \text{(Property A3)}\\
&\Rightarrow P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(B_i) \le \sum_{i=1}^{n} P(A_i).
\end{aligned}$$

■

Theorem 1.5.12. *Generalized Bonferroni's Inequality. If $\{A_i\}$ is a sequence of sets in $\mathcal{B}$, then*

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

Proof: Let us apply Boole's inequality to the A_i^c's. We have

$$P\left(\bigcup_{i=1}^{n} A_i^c\right) \le \sum_{i=1}^{n} P(A_i^c).$$

We know that

$$\bigcup_{i=1}^{n} A_i^c = \left(\bigcap_{i=1}^{n} A_i\right)^c.$$

Therefore

$$\begin{aligned}
P\left(\bigcup_{i=1}^{n} A_i^c\right) &= P\left[\left(\bigcap_{i=1}^{n} A_i\right)^c\right]\\
&= 1 - P\left[\bigcap_{i=1}^{n} A_i\right]\\
&\le \sum_{i=1}^{n} [1 - P(A_i)]\\
&= n - \sum_{i=1}^{n} P(A_i).
\end{aligned}$$

Rearranging terms, we get

$$P\left(\bigcap_{i=1}^{n} A_i\right) \ge \sum_{i=1}^{n} P(A)_i - (n-1)$$

■

Definition: A sequence of sets $\{A_n\}$ is said to be **non-decreasing** if $A_n \subseteq A_{n+1}$ for each

n. Since

$$\bigcup_{k=1}^{n} A_k = A_n,$$

$\bigcup_{k=1}^{\infty} A_k = A$ is called the **limit** of the sequence and we write $A_n \uparrow A$

Definition: If $A_n \supseteq A_{n+1}$ for each n, the sequence is said to be **non-increasing**. Then $\bigcap_{k=1}^{n} A_k = A_n$ and $\bigcap_{k=1}^{\infty} A_k = A$ is called the **limit** of the sequence. We write $A_n \downarrow A$

Example: Let $\mathcal{S} = (-\infty, \infty)$ and $A_n = \{\omega : 0 < \omega < 1 - \frac{1}{n}\}$

Then $A_n \uparrow A$, where $A = \{\omega : 0 < \omega < 1\}$.
$B_n = \{\omega : 0 < \omega < 1 + \frac{1}{n}\}$

Then $B_n \downarrow B$, where $B = \{\omega : 0 < \omega \leq 1\}$.
■

Theorem 1.5.13. *Axiom of Continuity.*

1. *Let $\{A_n\}$ be an **increasing** sequence of sets in $\mathcal{B}$. Then*

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n).$$

   This is referred to as continuity from below.

2. *Let $\{A_n\}$ be a **decreasing** sequence of sets in $\mathcal{B}$ with $\lim A_n = A \in \mathcal{B}$. Then*

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n).$$

   This is referred to as continuity from above.

Proof: Part 1.
Let $B_1 = A_1, \quad B_i = A_i - A_{i-1}, i = 2, 3, \ldots$.
Clearly $B_i's$ are disjoint. Therefore, we have

$$\begin{aligned} P(\lim_{n\to\infty} A_n) &= P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right) \\ &= \sum_{n=1}^{\infty} P(B_n) \\ &= \lim_{n\to\infty} \sum_{j=1}^{n} P(B_j) \end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to\infty}\{P(A_1)+\sum_{j=2}^{n}[P(A_j)-P(A_{j-1})]\} \\
&= \lim_{n\to\infty} P(A_n).
\end{aligned}
$$

Part 2. If $\{A_n\}$ is decreasing, then $\{B_n\} = \{A_n^c\}$ is increasing. From Part 1, we have

$$P(B) = P(\lim_{n\to\infty} B_n) = \lim_{n\to\infty} P(B_n). \tag{1.6}$$

Clearly,

$$\lim_{n\to\infty} P(B_n) \;=\; 1 - \lim_{n\to\infty} P(A_n). \tag{1.7}$$

Using (1.3) and (1.4), we have

$$
\begin{aligned}
1 - \lim_{n\to\infty} P(A_n) &= \lim_{n\to\infty} P(B_n) \\
&= P(\lim_{n\to\infty} B_n) \\
&= P\left(\bigcup_{n=1}^{\infty} B_n\right) = P\left(\bigcup_{n=1}^{\infty} A_n^c\right) \\
&= P\left[\left(\bigcap_{n=1}^{\infty} A_n\right)^c\right] \\
&= 1 - P\left(\bigcap_{n=1}^{\infty} A_n\right) \\
&= 1 - P(\lim_{n\to\infty} A_n).
\end{aligned}
$$

This shows that

$$P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n).$$

■

Corollary 1.5.14. *Axiom of Continuity (Special Case).*
*If $\{A_n\}$ is an infinite sequence of **decreasing** sets with $A_n \downarrow \phi$, then*

$$P(A_n) \to 0.$$

Proof: Direct application of previous result.
■

Corollary 1.5.15. *In the presence of finite additivity, this special axiom of continuity implies the axiom of countable additivity.*

Proof: Let $\{A_n\}$ be a sequence of pairwise disjoint sets in $\mathcal{B}$. We have

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{n} A_i\right) + P\left(\bigcup_{i=n+1}^{\infty} A_i\right).$$

Finite additivity implies

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{n} P(A_i) + P\left(\bigcup_{i=n+1}^{\infty} A_i\right).$$

$\bigcup_{i=n+1}^{\infty} A_i \downarrow \phi$. Using the special axiom of continuity, we have

$$P\left(\bigcup_{i=n+1}^{\infty} A_i\right) \to 0 \qquad as \;\; n \to \infty.$$

Therefore

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

■

Theorem 1.5.16. *Countable Subadditvity.*
Let $\{A_n\}$ be a sequence of sets in $\mathcal{B}$. Then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n).$$

Proof: Let $B_n = \bigcup_{i=1}^{n} A_i$. Clearly (!), $\{B_n\}$ is an increasing sequence with limit $A = \bigcup_{n=1}^{\infty} A_n$. Further, $B_n = B_{n-1} \cup A_n$. Using Boole's inequality, we have

$$P(B_n) \leq P(B_{n-1}) + P(A_n),$$

which is equivalent to

$$P(B_n) - P(B_{n-1}) \leq P(A_n).$$

Applying Theorem 1.4.11 to the $\{B_n\}$ sequence, we have

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} B_n\right) \\ &= \lim_{n\to\infty} P(B_n) \\ &= \lim_{n\to\infty} \left\{ P(B_1) + \sum_{j=2}^{n} [P(B_j) - P(B_{j-1})] \right\} \\ &\leq \lim_{n\to\infty} \sum_{j=1}^{n} P(A_j) \end{aligned}$$

$$= \sum_{n=1}^{\infty} P(A_n).$$

■

1.6 Examples

Example: Consider an experiment wherein n marbles are to be placed in n cells. If the marbles are randomly placed, find the probability that each cell will be occupied.

$$\frac{n!}{n^r}.$$

■

Example: Birthday Problem. Consider a group of r individuals. What is the probability that at least two share the same birthday?

Consider the sample space consisting of the birthdays of these r individuals. Ignoring leap years, each individual might have a birthday on any one of the 365 days of the year. The number of outcomes in the sample space is

$$365^r.$$

Let A be the event that all r birthdays are different. Then A contains

$$P_r^{365} = 365 \times 364 \times \ldots \times (365 - r + 1)$$

outcomes. Assuming all sequences of birthdays are equally likely, we have

$$P(A) = \frac{P_r^{365}}{365^r}.$$

The probability that at least two share the same birthday is $P(A^c)$.
For $r = 23$, the probability exceeds 0.5.

■

Example: Consider the random distribution of r marbles in n cells. Let

$$A_k = \text{event that a specified call has exactly } k \text{ marbles} \quad k = 0, \ldots, r.$$

The number of ways k marbles can be selected from the r is

$$\binom{r}{k}.$$

The remaining $r-k$ marbles can be distributed in any of the remaining $n-1$ cells in

$$(n-1)^{r-k}$$

ways. Therefore

$$P(A_k) = \binom{r}{k} \frac{(n-1)^{r-k}}{n^r} = \binom{r}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{r-k}.$$

■

Example: Poker hands. Find the probability of (a) A = a full house, (b) B=one pair, and (c) C=a straight.
A full house consists of three cards of one denomination and two of another.

$$P(A) = \frac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}}.$$

$$P(B) = \frac{\binom{13}{1}\binom{4}{2}\binom{12}{3}\binom{4}{1}\binom{4}{1}\binom{4}{1}}{\binom{52}{5}}.$$

$$P(C) = \frac{10[4^5 - 4]}{\binom{52}{5}}.$$

Example: Occupancy Problems

If we place r distinguishable marbles into n cells, the number of possible outcomes is n^r.
Suppose the r marbles are indistinguishable. The events of interest are the number of marbles in the cells as opposed to the ordered arrangements.
The sample space may be described be a vector of the form $(x_1, \ldots, x_n)$, where x_i is the number in the i-th cell.
The problem reduces to finding the number of distinct non-negative integer valued vectors $(x_1, \ldots, x_n)$ such that

$$x_1 + \ldots + x_n = r.$$

First, consider the number of positive integer valued solutions.
Imagine, we have r indistinguishable objects lined up, and that we want to divide them into n nonempty groups. We can select $n-1$ of $r-1$ spaces between the adjacent objects as our dividing points. For example, if $r = 8$ and $n = 3$, we can choose the 2 divisors as below:

$$\square \ \square \ \square \| \square \ \square \ \square \| \square \ \square$$

Here $x_1 = 3, x_2 = 3, x_3 = 2$. There are

$$\binom{r-1}{n-1}$$

possible selections.
To obtain the number of nonnegative (as opposed) to positive solutions, we note that the number of nonnegative solutions of

$$x_1 + \ldots + x_n = r$$

is the same as the number of positive solutions of

$$y_1 + \ldots + y_n = n + r,$$

letting $y_i = x_i + 1$. Therefore the number of distinct nonnegative solutions (integer valued) is given by

$$\binom{n+r-1}{n-1} = \binom{n+r-1}{r}.$$

Another solution for this problem is obtained by treating the objects as $\square$'s and the cells by $\|$'s. The n cells can be represented by the n spaces between $n+1$ bars. The two ends are fixed, but the remaining $(n-1)$ bars and the r objects may appear in an arbitrary order.
There are

$$\binom{n+r-1}{r}$$

ways of selecting r places out of $n+r-1$ objects.

Example: Out of $(2n+1)$ tickets consecutively numbered, 3 are drawn at random. Find the probability that the numbers form an arithmetic progression.

The number of ways of choosing 3 tickets is

$$\binom{2n+1}{3}.$$

The common difference for an AP can take values $1, 2, \ldots, n$. If $d = 1$, the first term a can range from $1, 2, \ldots, 2n-1$ [$2n-1$ possible values.]
If $d = 2$, there are $2n-3$ possible AP's.
If $d = n-1$, there are 3 AP's $1, n, 2n-1; 2, n+1, 2n; 3, n+2, 2n+1$.
If $d = n$, there is only one AP.
The total number of AP's is

$$(2n-1) + (2n-3) + \ldots + 1 = \frac{n}{2}[2 + (2n-1)] = n^2.$$

The probability that the numbers form an AP is

$$\frac{n^2}{\binom{2n+1}{3}}.$$