CHAPTER 1: Introduction to Probability Theory

1.2 Sample Space and Events

- A sample space $S$ is the set of all possible outcomes of an experiment.
- An event $E$ is a subset of the sample space $S$.
- The union of two events is the event consisting of outcomes that are in either of the two events. $E \cup F = \{x : x \in E \text{ or } x \in F\}$.
  
  Also, $\bigcup_{i=1}^{\infty} E_i = \{x \in E_1 \text{ or } x \in E_2 \text{ or } x \in E_3 \text{ or } \ldots\}$

- The intersection of two events is the event consisting of outcomes that are in both events. $E \cap F = EF = \{x \in E \text{ and } x \in F\}$.
  
  Also, $\bigcap_{i=1}^{\infty} E_i = \{x \in E_i \text{ for } i = 1, 2, \ldots\}$

- If $EF = \emptyset$, then $E$ and $F$ are mutually exclusive events (or disjoint events).

- The complement of event $E$ consists of outcomes that are in the sample space $S$ but are not in $E$. $E^c = E' = \overline{E} = \{x \in S \text{ and } x \notin E\}$.

- If outcome $x$ occurs in an experiment and $x \in E$, then we say $E$ occurs.

1.3 Probabilities Defined on Events

The probability of an event $E$, denoted $P(E)$, is a number between 0 and 1 inclusive that measures the likelihood of the occurrence of $E$. The basic 3 axioms (developed by Kolmogorov) for probability theory are the following:

(i) $0 \leq P(E) \leq 1$.
(ii) $P(S) = 1$ where $S$ denotes the sample space for the experiment.
(iii) For a sequence of mutually exclusive events $E_1, E_2, \ldots$,

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

The probability of an event $E$ can be interpreted as

1. the long-term relative frequency of $E$ in identical repeated experiments
2. the proportion of outcomes that are in $E$
   
   (when all outcomes are equally likely and finite in number)
3. a subjective assignment of likelihood based on experience, judgement, or intuition

Some probability rules derived from the basic axioms:

1. $P(E^c) = 1 - P(E)$
2. $P(E \cup F) = P(E) + P(F) - P(EF)$
3. $P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i} P(E_i) - \sum_{i<j} P(E_iE_j) + \sum_{i<j<k} P(E_iE_jE_k) - \sum_{i<j<k<l} P(E_iE_jE_kE_l) + \ldots + (-1)^{n+1} P(E_1E_2 \cdots E_n)$
1.4 Conditional Probabilities

The conditional probability that event $E$ occurs given that event $F$ has occurred is denoted $P(E|F)$ and is defined by

$$P(E|F) = \frac{P(EF)}{P(F)}.$$ 

Hence, the following multiplication rule holds:

$$P(EF) = P(E|F) \cdot P(F)$$

1.5 Independent Events

Two events $E$ and $F$ are (pairwise) independent if and only if $P(EF) = P(E)P(F)$. Equivalently, $E$ and $F$ are independent iff $P(E|F) = P(E)$.

Three events $E$, $F$, and $G$ are (mutually or jointly) independent if and only if $P(EF) = P(E)P(F)$, $P(EG) = P(E)P(G)$, $P(FG) = P(F)P(G)$, and $P(EGF) = P(E)P(F)P(G)$.

$N$ events are (mutually or jointly) independent if and only if the probability of the joint occurrence of each nonempty subset of the events equals the product of the probabilities of the events in the subset.

A sequence of experiments consists of independent trials if

(i) each experiment results in a “success” or “failure”, and

(ii) every subset of “success” events consists of independent events.

1.6 Bayes' Formula

Events $F_1$, $F_2$, . . . , $F_n$ are said to partition a sample space $S$ if they are mutually exclusive and their union equals $S$. In such a case, we have $S = \bigcup_{i=1}^{n} F_i$

and, for any event $E$,

$$E = \bigcup_{i=1}^{n} EF_i.$$ 

Then $P(E) = \sum_{i=1}^{n} P(EF_i) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$ and, subsequently,

$$P(F_j|E) = \frac{P(EF_j)}{P(E)} = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)} \quad \text{(Bayes' formula).}$$
Chapter 1 Supplementary Notes

• If \( E \subseteq F \), then \( P(E) \leq P(F) \).

• \( P(E_1E_2\cdots E_n) = P(E_1)P(E_2\mid E_1)P(E_3\mid E_1E_2)\cdots P(E_n\mid E_1E_2\cdots E_{n-1}) \)

• \( P(A \cup B) = 1 - P(A^cB^c) \)

• \( P(AB) = 1 - P(A^c \cup B^c) \)

• **Multiplicative counting principle.** If a task has \( k \) steps and there are \( n_i \) ways to perform step \( i \), then the number ways to do the task is \( \prod_{i=1}^{k} n_i \).

• The number of **permutations** of \( n \) items taken \( r \) at a time is given by

\[
nP_r = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}.
\]

• The number of **combinations** of \( n \) items taken \( r \) at a time is given by

\[
nC_r = \binom{n}{r} = \frac{n!}{(n-r)!r!}.
\]

• **Inclusion/Exclusion Principle.** Let \( |A| \) denote the cardinality of set \( A \).

For sets \( A_1, A_2, \ldots, A_n \), we have

\[
| \bigcup_{i=1}^{n} A_i | = \sum_{i=1}^{n} |A_i| - \sum_{i<j} |A_iA_j| + \sum_{i<j<k} |A_iA_jA_k| - \sum_{i<j<k<l} |A_iA_jA_kA_l| + \cdots
\]

\[
+ (-1)^{n+1} |A_1A_2\cdots A_n|.
\]

• **Boole’s Inequality.** \( | \bigcup_{i=1}^{n} A_i | \leq \sum_{i=1}^{n} |A_i| \)

• **Bonferroni Inequality.** \( P\left( \bigcup_{i=1}^{n} A_i \right) \leq \sum_{i=1}^{n} P(A_i) \).

Proof. Let \( B_1 = A_1, B_2 = A_2A_1^c, B_3 = A_3A_2^cA_1^c, B_4 = A_4A_3^cA_2^cA_1^c, \ldots, \)

\( B_n = A_nA_{n-1}A_{n-2}^c\cdots A_1^c. \) Since the \( B_i \)'s are mutually exclusive, \( B_i \subseteq A_i \) for all \( i \), and

\( \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i \), we have

\[
P\left( \bigcup_{i=1}^{n} A_i \right) = P\left( \bigcup_{i=1}^{n} B_i \right)
\]

\[
= \sum_{i=1}^{n} P(B_i) \quad \text{[since the } B_i \text{ are mutually exclusive]}
\]

\[
\leq \sum_{i=1}^{n} P(A_i) \quad \text{[since } B_i \subseteq A_i \text{ for all } i].
\]