Chapter 1

Axioms of Probability

1.1 Sample Space and Events

Definition. The set of all possible outcomes of an experiment is called the *sample space* of the experiment and is denoted by Ω.

Example. If the experiment consists of tossing two dice, then the sample space is

$$\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}.$$

Definition. Let Ω be a sample space of an experiment. A subset Σ of the power set of Ω is called a *σ-algebra* if the following conditions hold.

(a) $\Omega \in \Sigma$.

(b) For all $E \in \Sigma$, $\Omega \setminus E \in \Sigma$.

(c) If $E_1, E_2, \ldots$ is a sequence of elements in Σ, then

$$\bigcup_{i=1}^{\infty} E_i \in \Sigma.$$

Definition. A pair (Ω, Σ) where Σ is a σ-algebra in Ω is called a *measurable space*.

Definition. In a measurable space (Ω, Σ), each element of Σ is called an *event* of Ω (in Σ).

Remark. An event of Ω is a subset of Ω.

Definition. Two events E and F are *mutually exclusive* if $E \cap F = \varnothing$.

Definition. Let (Ω, Σ) be a measurable space. A function $P : \Sigma \to \mathbb{R}$ is called a *probability function* and (Ω, Σ, P) is a *probability space* if the following conditions hold.

(a) For all $E \in \Sigma$, $P(E) \geq 0$.

(b) $P(\Omega) = 1$.

(c) If $E_1, E_2, \ldots$ is a sequence of events that are pairwise mutually exclusive, then

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

Theorem 1.1. Let (Ω, Σ, P) be a probability space. Let $E, F \in \Sigma$. Then $P(F \setminus E) = P(F) - P(E \cap F)$.

Proof. Since $E \cap F$ and $F \setminus E$ are mutually exclusive, we have

$$P(F) = P((E \cap F) \cup (F \setminus E)) = P(E \cap F) + P(F \setminus E).$$

Thus, $P(F \setminus E) = P(F) - P(E \cap F)$. $\square$

Corollary. $P(\Omega \setminus E) = 1 - P(E)$ holds for any event E, implying $P(\varnothing) = 0$.

Corollary. If $E \subseteq F$, then $P(E) \leq P(F)$ because $P(F) - P(E) = P(F \setminus E) \geq 0$.

Theorem 1.2 (Inclusive-exclusive Principle)**.** Let (Ω, Σ, P) be a probability space. If $E_1, \ldots, E_n \in \Sigma$, then

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{1 \leq i_1 < \cdots < i_r \leq n} P(E_{i_1} \cap \cdots \cap E_{i_r}).$$

Proof. The proof is by induction on n. The theorem holds for $n = 0$ and $n = 1$ trivially. For $n = 2$, since $E_1 \cap E_2$ and $E_1 \setminus E_2$ are mutually exclusive, we have

$$P(E_1) = P((E_1 \cap E_2) \cup (E_1 \setminus E_2)) = P(E_1 \cap E_2) + P(E_1 \setminus E_2).$$

Thus, since $E_1 \setminus E_2$ and E_2 are mutually exclusive, we have

$$\begin{aligned} P(E_1 \cup E_2) &= P((E_1 \setminus E_2) \cup E_2) \\ &= P(E_1 \setminus E_2) + P(E_2) \\ &= P(E_1) - P(E_1 \cap E_2) + P(E_2). \end{aligned}$$

Now suppose that the theorem holds for some $n \geq 2$, and we prove that the theorem is true for $n + 1$. Since $E_1 \cup \cdots \cup E_n$ and E_{n+1} are mutually exclusive, we have

$$P(E_1 \cup \cdots \cup E_n \cup E_{n+1}) = P(E_1 \cup \cdots \cup E_n) + P(E_{n+1}) - P((E_1 \cup \cdots \cup E_n) \cap E_{n+1}),$$

where the first term can be written as

$$P(E_1 \cup \cdots \cup E_n) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{1 \leq i_1 < \cdots < i_r \leq n} P(E_{i_1} \cap \cdots \cap E_{i_r}),$$

and the last term can be written as

$$\begin{aligned} &P((E_1 \cup \cdots \cup E_n) \cap E_{n+1}) \\ &= P((E_1 \cap E_{n+1}) \cup \cdots \cup (E_k \cap E_{n+1})) \\ &= \sum_{s=1}^{n} (-1)^{s+1} \sum_{1 \leq i_1 \cdots \leq i_s \leq n} P((E_{i_1} \cap E_{n+1}) \cap \cdots \cap (E_{i_s} \cap E_{n+1})) \\ &= \sum_{s=1}^{n} (-1)^{s+1} \sum_{1 \leq i_1 \cdots \leq i_s \leq n} P(E_{i_1} \cap \cdots \cap E_{i_s} \cap E_{n+1}) \\ &= -\sum_{r=2}^{n+1} (-1)^{r+1} \sum_{\substack{1 \leq i_1 \cdots \leq i_{r-1} \leq n \\ i_r = n+1}} P(E_{i_1} \cap \cdots \cap E_{i_{r-1}} \cap E_{i_r}). \end{aligned}$$

Now we consider r, which is the number of sets in each intersection. The second term is actually the case with $r = 1$, and the last term consists of the cases with $r \geq 2$. Thus,

$$P(E_{n+1}) - P((E_1 \cup \cdots E_n) \cap E_{n+1}) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{\substack{1 \leq i_1 \cdots \leq i_{r-1} \leq n \\ i_r = n+1}} P(E_{i_1} \cap \cdots \cap E_{i_r}).$$

Furthermore, note that the first term consists of the case where E_{n+1} does not appear in the intersection, while the difference above consists of the case where E_{n+1} appears in the intersection. Thus, by summing up all terms, we have

$$P(E_1 \cup \cdots \cup E_n \cup E_{n+1}) = \sum_{r=1}^{n+1} (-1)^{r+1} \sum_{1 \leq i_1 \leq \cdots \leq i_r \leq n+1} P(E_{i_1} \cap \cdots \cap E_{i_r}),$$

which completes the proof. $\square$

Example. For any three events E_1, E_2, E_3, we have $P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_2 \cap E_3) - P(E_3 \cap E_1) + P(E_1 \cap E_2 \cap E_3)$.

1.2 Sample Spaces with Equally Likely Outcomes

Theorem 1.3. Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be a finite sample space and let P be a probability function such that $P(\{\omega_i\}) = P(\{\omega_j\})$ for $i, j \in \{1, \ldots, n\}$. Then for each event $E \subseteq \Omega$ with $|E| = k$, we have

$$P(E) = \frac{k}{n}.$$

Proof. Let p denote the probability of each elementary event $\{\omega_i\}$ for all $i \in \{1, \ldots, n\}$. Then we have

$$1 = P(\Omega) = P\left(\bigcup_{i=1}^{n} \{\omega_i\}\right) = \sum_{i=1}^{n} P(\{\omega_i\}) = np.$$

Thus,

$$p = \frac{1}{n}.$$

Let $E = \{\omega_{i_1}, \ldots, \omega_{i_k}\}$. Then

$$P(E) = P\left(\bigcup_{r=1}^{k} \{\omega_{i_r}\}\right) = \sum_{r=1}^{k} P(\{\omega_{i_r}\}) = \frac{k}{n}. \qquad \square$$

Chapter 2

Conditional Probability and Independence

2.1 Conditional Probability

Definition. Let (Ω, Σ, P) be a probability space. If E is an event with $P(E) > 0$, then define

$$P(F \mid E) = \frac{P(E \cap F)}{P(E)}$$

for any event F.

Theorem 2.1. Let (Ω, Σ, P) be a probability space. If E is an event with $P(E) > 0$, then the function $P_E : \Sigma \to \mathbb{R}$ is a probability function if

$$P_E(F) = P(F \mid E)$$

for any event F.

Proof. For events E and F,

$$P_E(F) = \frac{P(E \cap F)}{P(E)} \geq 0.$$

Moreover,

$$P_E(\Omega) = \frac{P(E \cap \Omega)}{P(E)} = \frac{P(E)}{P(E)} = 1.$$

If $F_1, F_2, \ldots$ is a sequence of events that are piecewise mutually exclusive, then

$$P_E\left(\bigcup_{i=1}^{\infty} F_i\right) = \frac{P\Big(E \cap \bigcup_{i=1}^{\infty} F_i\Big)}{P(E)} = \frac{P\Big(\bigcup_{i=1}^{\infty} (E \cap F_i)\Big)}{P(E)} = \sum_{i=1}^{\infty} \frac{P(E \cap F_i)}{P(E)} = \sum_{i=1}^{\infty} P_E(F_i).$$

Thus, P_E is a probability function. $\square$

2.2 Bayes' Formula

Definition. A *partition* of Ω is a family of nonempty events such that each element in Ω is in exactly one of these events.

Theorem 2.2. Let $E_1, \ldots, E_n$ form a partition of Ω such that $P(E_i) > 0$ for each $i \in \{1, \ldots, n\}$. Then for any event F,

$$P(F) = \sum_{i=1}^{n} P(F \mid E_i) P(E_i).$$

Proof. Since

$$F = F \cap \Omega = F \cap \bigcup_{i=1}^{n} E_i = \bigcup_{i=1}^{n} (F \cap E_i),$$

it follows that

$$P(F) = P\left(\bigcup_{i=1}^{n} (F \cap E_i)\right) = \sum_{i=1}^{n} P(F \cap E_i) = \sum_{i=1}^{n} P(F \mid E_i) P(E_i). \qquad \square$$

Theorem 2.3 (Bayes' Formula)**.** Let $E_1, \ldots, E_n$ form a partition of Ω such that $P(E_j) > 0$ for each $j \in \{1, \ldots, n\}$. Then for any event F with $P(F) > 0$, for any $i \in \{1, \ldots, n\}$, we have

$$P(E_i \mid F) = \frac{P(F \mid E_i) P(E_i)}{\sum_{j=1}^{n} P(F \mid E_j) P(E_j)}.$$

Proof. By Theorem 2.2, we have

$$P(E_i \mid F) = \frac{P(E_i \cap F)}{P(F)} = \frac{P(F \mid E_i) P(E_i)}{\sum_{j=1}^{n} P(F \mid E_j) P(E_j)}. \qquad \square$$

2.3 Independence

Definition. Let $E_1, \ldots, E_n$ be events in a probability space (Ω, Σ, P).

- $E_1, \ldots, E_n$ are *independent* if

  $$P\left(\bigcap_{i \in I} E_i\right) = \prod_{i \in I} P(E_i)$$

  holds for any nonempty subset I of $\{1, \ldots, n\}$.

- $E_1, \ldots, E_n$ are *dependent* if they are not independent.

Definition. Let $E_1, \ldots, E_n$ and F be events in a probability space (Ω, Σ, P), where $P(F) > 0$. Then $E_1, \ldots, E_n$ are *independent given event* F if

$$P\left(\bigcap_{i \in I} E_i \;\middle|\; F\right) = \prod_{i \in I} P(E_i \mid F)$$

holds for any nonempty subset I of $\{1, \ldots, n\}$.