MODULE 1

PROBABILITY

LECTURE 2

Topics

1.2 AXIOMATIC APPROACH TO PROBABILITY AND PROPERTIES OF PROBABILITY MEASURE

In the following section we will discuss the modern approach to probability theory where we will not be concerned with how probabilities are assigned to suitably chosen subsets of Ω. Rather we will define the concept of probability for certain types of subsets Ω using a set of axioms that are consistent with properties (i)-(iii) of classical (or relative frequency) method. We will also study various properties of probability measures.

1.2 AXIOMATIC APPROACH TO PROBABILITY AND PROPERTIES OF PROBABILITY MEASURE

We begin this section with the following definitions.

Definition 2.1

(i) A set whose elements are themselves set is called a *class* of sets. A class of sets will be usually denoted by script letters $\mathcal{A}, \mathcal{B}, \mathcal{C}, \ldots$. For example $\mathcal{A} = \{\{1\}, \{1,3\}, \{2,5,6\}\}$;

(ii) Let $\mathcal{C}$ be a class of sets. A function $\mu: \mathcal{C} \rightarrow \mathbb{R}$ is called a *set function*. In other words, a real-valued function whose domain is a class of sets is called a set function. ■

As stated above, in many situations, it may not be possible to assign probabilities to all subsets of the sample space Ω such that properties (i)-(iii) of classical (or relative frequency) method are satisfied. Therefore one begins with assigning probabilities to members of an appropriately chosen class $\mathcal{C}$ of subsets of Ω (e.g., if $\Omega = \mathbb{R}$, then $\mathcal{C}$ may be class of all open intervals in $\mathbb{R}$; if Ω is a countable set, then $\mathcal{C}$ may be class of all singletons $\{\omega\}, \omega \in \Omega$). We call the members of $\mathcal{C}$ as *basic sets*. Starting from the basic sets in $\mathcal{C}$ assignment of probabilities is extended, in an intuitively justified manner, to as many subsets of Ω as possible keeping in mind that properties (i)-(iii) of classical (or relative frequency) method are not violated. Let us denote by $\mathcal{F}$ the class of sets for which the probability assignments can be finally done. We call the class $\mathcal{F}$ as *event space* and elements of $\mathcal{F}$ are called *events*. It will be reasonable to assume that $\mathcal{F}$ satisfies the following properties: (i) $\Omega \in \mathcal{F}$, (ii) $A \in \mathcal{F} \Longrightarrow A^C = \Omega - A \in \mathcal{F}$, and (iii) $A_i \in \mathcal{F}, i = 1,2, \ldots \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$. This leads to introduction of the following definition.

Definition 2.2

A sigma-field (σ-field) of subsets of Ω is a class $\mathcal{F}$ of subsets of Ω satisfying the following properties:

(i) $\Omega \in \mathcal{F}$;

(ii) $A \in \mathcal{F} \Rightarrow A^c = \Omega - A \in \mathcal{F}$ (closed under complements);

(iii) $A_i \in \mathcal{F}, i = 1, 2, \ldots \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ (closed under countably infinite unions). ■

Remark 2.1

(i) We expect the event space to be a σ-field;

(ii) Suppose that $\mathcal{F}$ is a σ-field of subsets of Ω. Then,

(a) $\phi \in \mathcal{F}$(since $\phi = \Omega^c$)

(b) $E_1, E_2, \ldots \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{\infty} E_i \in \mathcal{F}$ (since $\bigcap_{i=1}^{\infty} E_i = (\bigcup_{i=1}^{\infty} E_i^c)^c$);

(c) $E, F \in \mathcal{F} \Rightarrow E - F = E \cap F^c \in \mathcal{F}$ and $E \,\Delta F \stackrel{\text{def}}{=} (E - F) \cup (F - E) \in \mathcal{F}$;

(d) $E_1, E_2, \ldots, E_n \in \mathcal{F}$, for some $n \in \mathbb{N}, \Rightarrow \bigcup_{i=1}^{n} E_i \in \mathcal{F}$ and $\bigcap_{i=1}^{n} E_i \in \mathcal{F}$ (take $E_{n+1} = E_{n+2} = \cdots = \phi$ so that $\bigcup_{i=1}^{n} E_i = \bigcup_{i=1}^{\infty} E_i$ or $E_{n+1} = E_{n+2} = \cdots = \Omega$ so that $\bigcap_{i=1}^{n} E_i = \bigcap_{i=1}^{\infty} E_i$);

(e) although the power set of Ω $\left(\mathcal{P}(\Omega)\right)$ is a σ-field of subsets of Ω, in general, a σ-field may not contain all subsets of Ω. ■

Example 2.1

(i) $\mathcal{F} = \{\phi, \Omega\}$ is a sigma field, called the *trivial sigma-field*;

(ii) Suppose that $A \subseteq \Omega$. Then $\mathcal{F} = \{A, A^c, \phi, \Omega\}$ is a σ-field of subsets of Ω. It is the smallest sigma-field containing the set A;

(iii) Arbitrary intersection of σ-fields is a σ-field (see Problem 3 (i));

(iv) Let $\mathcal{C}$ be a class of subsets of Ω and let $\{F_\alpha : \alpha \in \Lambda\}$ be the collection of all σ-fields that contain $\mathcal{C}$. Then

$$\mathcal{F} = \bigcap_{\alpha \in \Lambda} \mathcal{F}_\alpha$$

is a σ-field and it is the smallest σ-field that contains class $\mathcal{C}$ (called the σ-field generated by $\mathcal{C}$ and is denoted by $\sigma(\mathcal{C})$) (see Problem 3 (iii));

(v) Let $\Omega = \mathbb{R}$ and let $\mathcal{J}$ be the class of all open intervals in $\mathbb{R}$. Then $\mathcal{B}_1 = \sigma(\mathcal{J})$ is called the *Borel σ-field* on $\mathbb{R}$. The Borel σ-field in $\mathbb{R}^k$ (denoted by $\mathcal{B}_k$) is the

σ-field generated by class of all open rectangles in $\mathbb{R}^k$. A set $B \in \mathcal{B}_k$ is called a Borel set in $\mathbb{R}^k$; here $\mathbb{R}^k = \{(x_1, \dots, x_k): -\infty < x_i < \infty,\ i = 1, \dots, k\}$ denotes the k-dimensional Euclidean space;

(vi) $\mathcal{B}_1$ contains all singletons and hence all countable subsets of $\mathbb{R}\left(\{a\} = \bigcap_{n=1}^{\infty}\left(a - \frac{1}{n}, a + \frac{1}{n}\right)\right)$. ■

Let $\mathcal{C}$ be an appropriately chosen class of basic subsets of Ω for which the probabilities can be assigned to begin with (e.g., if $\Omega = \mathbb{R}$ then $\mathcal{C}$ may be class of all open intervals in $\mathbb{R}$; if Ω is a countable set then $\mathcal{C}$ may be class of all singletons $\{\omega\}, \omega \in \Omega$). It turns out (a topic for an advanced course in probability theory) that, for an appropriately chosen class $\mathcal{C}$ of basic sets, the assignment of probabilities that is consistent with properties (i)-(iii) of classical (or relative frequency) method can be extended in an unique manner from $\mathcal{C}$ to $\sigma(\mathcal{C})$, the smallest σ-field containing the class $\mathcal{C}$. Therefore, generally the domain $\mathcal{F}$ of a probability measure is taken to be $\sigma(\mathcal{C})$, the σ-field generated by the class $\mathcal{C}$ of basic subsets of Ω. We have stated before that we will not care about how assignment of probabilities to various members of event space $\mathcal{F}$ (a σ-field of subsets of Ω) is done. Rather we will be interested in properties of probability measure defined on event space $\mathcal{F}$.

Let Ω be a sample space associated with a random experiment and let $\mathcal{F}$ be the event space (a σ-field of subsets of Ω). Recall that members of $\mathcal{F}$ are called events. Now we provide a mathematical definition of probability based on a set of axioms.

Definition 2.3

(i) Let $\mathcal{F}$ be a σ-field of subsets of Ω. A probability function (or a probability measure) is a set function P, defined on $\mathcal{F}$, satisfying the following three axioms:

(a) $P(E) \geq 0,\ \ \forall E \in \mathcal{F}$; $\qquad$ $(\text{Axiom 1: Non} - \text{negativity})$;

(b) If $E_1, E_2, \dots$ is a countably infinite collection of mutually exclusive events $\left(\text{i.e., } E_i \in \mathcal{F},\ i = 1, 2, \dots, E_i \cap E_j = \phi, i \neq j\right)$ then

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{1=1}^{\infty} P(E_i)\,; \qquad (\text{Axiom 2: Countably infinite additive})$$

(c) $P(\Omega) = 1$ $\qquad$ (Axiom 3: Probability of the sample space is 1).

(ii) The triplet $(\Omega, \mathcal{F}, P)$ is called a probability space. ■

Remark 2.2

(i) Note that if $E_1, E_2, \ldots$ is a countably infinite collection of sets in a σ-field $\mathcal{F}$ then $\bigcup_{i=1}^{\infty} E_i \in \mathcal{F}$ and, therefore, $P(\bigcup_{i=1}^{\infty} E_i)$ is well defined;

(ii) In any probability space $(\Omega, \mathcal{F}, P)$ we have $P(\Omega) = 1$ (or $P(\phi) = 0$; see Theorem 2.1 (i) proved later) but if $P(A) = 1$ (or $P(A) = 0$), for some $A \in \mathcal{F}$, then it does not mean that $A = \Omega$ (or $A = \phi$) (see Problem 14 (ii)).

(iii) In general not all subsets of Ω are events, i.e., not all subsets of Ω are elements of $\mathcal{F}$.

(iv) When Ω is countable it is possible to assign probabilities to all subsets of Ω using Axiom 2 provided we can assign probabilities to singleton subsets $\{x\}$ of Ω. To illustrate this let $\Omega = \{\omega_1, \omega_2, \ldots\}$ (or $\Omega = \{\omega_1, \ldots, \omega_n\}$, for some n $\in \mathbb{N}$) and let $P(\{\omega_i\}) = p_i,\ i = 1, 2, \ldots,$ so that $0 \le p_i \le 1,\ i = 1,2, \ldots$ (see Theorem 2.1 (iii) below) and $\sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} P(\{\omega_i\}) = P(\bigcup_{i=1}^{\infty}\{\omega_i\}) = P(\Omega) = 1$. Then, for any $A \subseteq \Omega$,

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

Thus in this case we may take $\mathcal{F} = P(\Omega)$, the power set of Ω. It is worth mentioning here that if Ω is countable and $\mathcal{C} = \big\{\{\omega\} :\ \omega \in \Omega\big\}$ (class of all singleton subsets of Ω) is the class of basic sets for which the assignment of the probabilities can be done, to begin with, then $\sigma(\mathcal{C}) = \mathcal{P}(\Omega)$ (see Problem 5 (ii)).

(v) Due to some inconsistency problems, assignment of probabilities for all subsets of Ω is not possible when Ω is continuum (e.g., if Ω contains an interval). ■

Theorem 2.1

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Then

(i) $P(\phi) = 0$;

(ii) $E_i \in \mathcal{F}, i = 1, 2, \ldots . n$, and $E_i \cap E_j = \phi,\ i \neq j \Rightarrow P(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i)$ (finite additivity);

(iii) $\forall E \in \mathcal{F}, 0 \le P(E) \le 1$ and $P(E^c) = 1 - P(E)$;

(iv) $E_1, E_2 \in \mathcal{F}$ and $E_1 \subseteq E_2 \Rightarrow P(E_2 - E_1) = P(E_2) - P(E_1)$ and $P(E_1) \le P(E_2)$ (monotonicity of probability measures);

(v) $E_1, E_2 \in \mathcal{F} \Rightarrow P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.

Proof.

(i) Let $E_1 = \Omega$ and $E_i = \phi,\ i = 2, 3, \ldots .$ Then $P(E_1) = 1$, (Axiom 3), $E_i \in \mathcal{F},\ i = 1, 2, \ldots$, $E_1 = \bigcup_{i=1}^{\infty} E_i$ and $E_i \cap E_j = \phi,\ i \neq j$. Therefore,

$$1 = P(E_1) = P\left(\bigcup_{i=1}^{\infty} E_i\right)$$

$$= \sum_{i=1}^{\infty} P(E_i) \qquad \text{(using Axiom 2)}$$

$$= 1 + \sum_{i=2}^{\infty} P(\phi)$$

$$\Rightarrow \sum_{i=2}^{\infty} P(\phi) = 0$$

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

(ii) Let $E_i = \phi,\ i = n+1,\ n+2, \dots$. Then $E_i \in \mathcal{F},\ i = 1, 2, \dots, E_i \cap E_j = \phi, i \neq j$ and $P(E_i) = 0,\ i = n+1, n+2, \dots$. Therefore,

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

$$= \sum_{i=1}^{\infty} P(E_i) \qquad \text{(using Axiom 2)}$$

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

(iii) Let $E \in \mathcal{F}$. Then $\Omega = E \cup E^c$ and $E \cap E^C = \phi$. Therefore

$$\begin{aligned} 1 &= P(\Omega) \\ &= P(E \cup E^c) \\ &= P(E) + P(E^c) \text{ (using (ii))} \\ &\Rightarrow P(E) \leq 1 \text{ and } P(E^c) = 1 - P(E) \quad (\text{since } P(E^c) \in [0,1]) \\ &\Rightarrow 0 \leq P(E) \leq 1 \text{ and } P(E^c) = 1 - P(E). \end{aligned}$$

(iv) Let $E_1, E_2 \in \mathcal{F}$ and let $E_1 \subseteq E_2$. Then $E_2 - E_1 \in \mathcal{F}, E_2 = E_1 \cup (E_2 - E_1)$ and $E_1 \cap (E_2 - E_1) = \phi$.

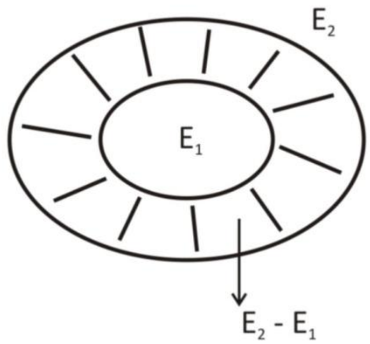

Figure 2.1

Therefore,

$$P(E_2) = P\big(E_1 \cup (E_2 - E_1)\big)$$

$$= P(E_1) + P(E_2 - E_1) \qquad \text{(using (ii))}$$

$$\Rightarrow P(E_2 - E_1) = P(E_2) - P(E_1).$$

As $P(E_2 - E_1) \geq 0$, it follows that $P(E_1) \leq P(E_2)$.

(v) Let $E_1, E_2 \in \mathcal{F}$. Then $E_2 - E_1 \in \mathcal{F}$, $E_1 \cap (E_2 - E_1) = \phi$ and $E_1 \cup E_2 = E_1 \cup (E_2 - E_1)$.

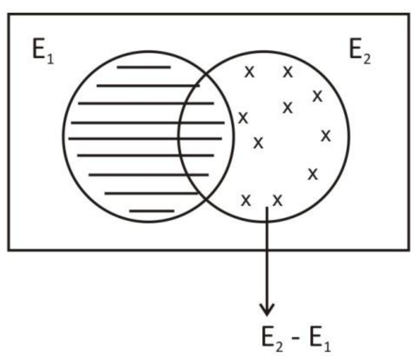

Figure 2.2

Therefore,

$$P(E_1 \cup E_2) = P\big(E_1 \cup (E_2 - E_1)\big)$$

$$= P(E_1) + P(E_2 - E_1) \qquad \text{(using (ii))} \qquad (2.1)$$

Also $(E_1 \cap E_2) \cap (E_2 - E_1) = \phi$ and $E_2 = (E_1 \cap E_2) \cup (E_2 - E_1)$. Therefore,

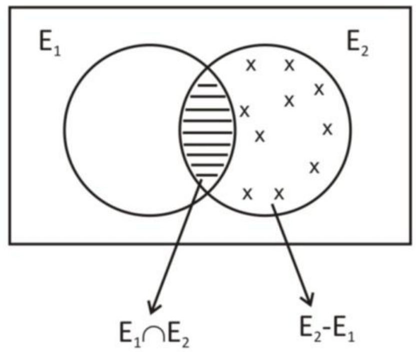

Figure 2.3

$$P(E_2) = P\big((E_1 \cap E_2) \cup (E_2 - E_1)\big)$$

$$= P(E_1 \cap E_2) + P(E_2 - E_1) \qquad \text{(using (ii))}$$

$$\Rightarrow P(E_2 - E_1) = P(E_2) - P(E_1 \cap E_2)\cdot \qquad (2.2)$$

Using (2.1) and (2.2), we get

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

1.2.1 Inclusion-Exclusion Formula

Theorem 2.2

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $E_1, E_2, \dots, E_n \in \mathcal{F}$ $(n \in \mathbb{N}, n \geq 2)$. Then

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{k=1}^{n} S_{k,n}\ ,$$

where $S_{1,n} = \sum_{i=1}^{n} P(E_i)$ and, for $k \in \{2, 3, \dots, n\}$,

$$S_{k,n} = (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P\big(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}\big).$$

Proof. We will use the principle of mathematical induction. Using Theorem 2.1 (v), we have

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

$$= S_{1,2} + S_{2,2},$$

where $S_{1,2} = P(E_1) + P(E_2)$ and $S_{2,2} = -P(E_1 \cap E_2)$. Thus the result is true for $n = 2$. Now suppose that the result is true for $n \in \{2, 3, \dots, m\}$ for some positive integer m ($\geq$ *2*. Then

$$P\left(\bigcup_{i=1}^{m+1} E_i\right) = P\left(\left(\bigcup_{i=1}^{m} E_i\right) \cup E_{m+1}\right)$$

$$= P\left(\bigcup_{i=1}^{m} E_i\right) + P(E_{m+1}) - P\left(\left(\bigcup_{i=1}^{m} E_i\right) \cap E_{m+1}\right) \quad \text{(using the result for } n = 2\text{)}$$

$$= P\left(\bigcup_{i=1}^{m} E_i\right) + P(E_{m+1}) - P\left(\bigcup_{i=1}^{m} (E_i \cap E_{m+1})\right)$$

$$= \sum_{i=1}^{m} S_{i,m} + P(E_{m+1}) - P\left(\bigcup_{i=1}^{m} (E_i \cap E_{m+1})\right) \quad \text{(using the result for } n = m\text{)} \quad (2.3)$$

Let $F_i = E_i \cap E_{m+1}, i = 1, \dots . m$. Then

$$P\left(\bigcup_{i=1}^{m} (E_i \cap E_{m+1})\right) = P\left(\bigcup_{i=1}^{m} F_i\right)$$

$$= \sum_{k=1}^{m} T_{k,m} \text{ (again using the result for } n = m\text{)}, \qquad (2.4)$$

where

$$T_{1,m} = \sum_{i=1}^{m} P(F_i) = \sum_{i=1}^{m} P(E_i \cap E_{m+1}) \text{ and, for } k \in \{2, 3, \cdots, m\},$$

$$T_{k,m} = (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq m} P\left(F_{i_1} \cap F_{i_2} \cap \cdots \cap F_{i_k}\right)$$

$$= (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq m} P\left(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k} \cap E_{m+1}\right).$$

Using (2.4) in (2.3), we get

$$P(\cup_{i=1}^{m+1} E_i) = \left(S_{1,m} + P(E_{m+1})\right) + \left(S_{2,m} - T_{1,m}\right) + \cdots + \left(S_{m,m} - T_{m-1,m}\right) - T_{m,m} \ .$$

Note that $S_{1,m} + P(E_{m+1}) = S_{1,m+1}$, $S_{k,m} - T_{k-1,m} = S_{k,m+1}, k = 2,3, \ldots, m$, and $T_{m,m} = -S_{m+1,m+1}$. Therefore,

$$P\left(\bigcup_{i=1}^{m+1} E_i\right) = S_{1,m+1} + \sum_{k=2}^{m+1} S_{k,m+1} = \sum_{k=1}^{m+1} S_{k,m+1} \, . \blacksquare$$

Remark 2.3

(i) Let $E_1, E_2, \ldots \in \mathcal{F}$. Then

$$P(E_1 \cup E_2 \cup E_3) = \underbrace{P(E_1) + P(E_2) + P(E_3)}_{S_{1,3}} \underbrace{-\left(P(E_1 \cap E_2) + P(E_1 \cap E_3) + P(E_2 \cap E_3)\right)}_{S_{2,3}} \underbrace{+P(E_1 \cap E_2 \cap E_3)}_{S_{3,3}}$$

$$= p_{1,3} - p_{2,3} + p_{3,3},$$

where $p_{1,3} = S_{1,3}$, $p_{2,3} = -S_{2,3}$ and $p_{3,3} = S_{3,3}$.

In general,

$$P(\cup_{i=1}^{n} E_i) = p_{1,n} - p_{2,n} + p_{3,n} \cdots + (-1)^{n-1} p_{n,n},$$

where

$$p_{i,n} = \begin{cases} S_{i,n}, & \text{if } i \text{ is odd} \\ -S_{i,n}, & \text{if } i \text{ is even} \end{cases}, \ i = 1, 2, \ldots n.$$

(ii) We have

$$1 \geq P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

$$\Rightarrow P(E_1 \cap E_2) \geq P(E_1) + P(E_2) - 1.$$

The above inequality is known as *Bonferroni's inequality*. $\blacksquare$