PART I. PROBABILITY THEORY

Chapter 1. Probability Space

There are three approaches to the notion of *probability*:

- **classical**: intuitive, what most people are familiar with and think of when they hear the word "probability";
- **geometrical**: a natural extension of classical probability, for the case of infinite numbers of cases;
- **axiomatic**: rigorous, mathematical, enables proving probability formulas.

1 Experiments and Events

An **experiment** is any process or action whose outcome is not known (is random).
A **sample space**, denoted by S, is the set of all possible outcomes of an experiment.
Its elements are called **elementary events** (denoted by $e_i,\ i \in \mathbb{N}$).
An **event** is a collection of elementary events, i.e. it is a subset of S (events are denoted by capital letters, $A_i,\ i \in \mathbb{N}$).

Since events are defined as sets, we can employ set theory in describing them.

- two special events associated with every experiment:
  - the **impossible** event, denoted by $\emptyset$ ("never happens");
  - the **sure (certain)** event, denoted by S ("surely happens").
- for each event $A \subseteq S$, we define the event $\overline{A}$, the **complementary** event, to mean that $\overline{A}$ occurs if and only if A does not occur; $\overline{\overline{A}} = A$;

- we say that event A **implies (induces)** event B, $A \subseteq B$, if every element of A is also an element of B, or in other words, if the occurrence of A induces (implies) the occurrence of B; A and B are **equal (equivalent)**, $A = B$, if A implies B and B implies A;
- for any two events $A, B \subseteq S$, we define the following events:
  - **union** of A and B,

    $$A \cup B = \{e \in S \,|\, e \in A \text{ or } e \in B\},$$

    the event that occurs if either A or B or both occur;
  - **intersection** of A and B,

    $$A \cap B = \{e \in S \,|\, e \in A \text{ and } e \in B\},$$

    the event that occurs if both A and B occur;
  - **difference** of A and B,

    $$A \setminus B = \{e \in S \,|\, e \in A \text{ and } e \notin B\} = A \cap \overline{B},$$

    the event that occurs if A occurs and B does not;
  - **symmetric difference** of A and B,

    $$A \Delta B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B),$$

    the event that occurs if A or B occur, but not both.

The operations of union, intersection and symmetric difference are

- **commutative**:

$$A \cup B = B \cup A, \quad A \cap B = B \cap A, \quad A \Delta B = B \Delta A;$$

– **associative**:

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C),$$

$$(A \Delta B) \Delta C = A \Delta (B \Delta C);$$

– **distributive**:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C), \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C),$$

$$A \cap (B \Delta C) = (A \cap B) \Delta (A \cap C).$$

Definition 1.1.

- *Two events A and B are said to be* ***mutually exclusive (disjoint, incompatible)*** *if A and B cannot occur at the same time, i.e. $A \cap B = \emptyset$;*
- *Three or more events are mutually exclusive if* ***any two of them are****, i.e.*

$$A_i \cap A_j \ = \ \emptyset, \ \forall i \neq j;$$

- *A collection of events $\{A_i\}_{i \in I}$ from S is said to be* ***(collectively) exhaustive*** *if*

$$\bigcup_{i \in I} A_i = S;$$

- *A collection of events $\{A_i\}_{i \in I}$ from S is said to be a* ***partition*** *of S if the events are collectively exhaustive and mutually exclusive, i.e.*

$$\begin{aligned} \bigcup_{i \in I} A_i \ &= \ S \\ A_i \cap A_j \ &= \ \emptyset, \forall i, j \in I, \ i \neq j. \end{aligned}$$

Example 1.2. Consider the experiment of rolling a die. Then the sample space is

$$S = \{e_1, e_2, e_3, e_4, e_5, e_6\},$$

where the elementary events (outcomes) are $e_i,\ i = \overline{1,6}$, with e_i being the event that the face i shows on the die.

Consider the following events:
A: face 1 shows,
B: face 2 shows,
C: an even number shows,
D: a prime number shows,
E: a composite number shows.
Then we have

$$A = \{e_1\},\ B = \{e_2\},\ C = \{e_2, e_4, e_6\},\ D = \{e_2, e_3, e_5\},\ E = \{e_4, e_6\}.$$

We also have

$$B \subseteq C,\ A \cap B = \emptyset,\ A \cap D = \emptyset,\ A \cap E = \emptyset,\ D \cap E = \emptyset,$$

$$C \cap D = B,\ A \cup D \cup E = S.$$

So, for example, events $\{A, B\}$ and $\{A, D, E\}$ are mutually exclusive. In fact, these last three are also collectively exhaustive. Thus, events $\{A, D, E\}$ form a partition of S.

Proposition 1.3.
*For every collection of events $\{A_i\}_{i\in I}$, **De Morgan's laws** hold:*

a) $\overline{\bigcup\limits_{i\in I} A_i} = \bigcap\limits_{i\in I} \overline{A}_i,$

b) $\overline{\bigcap\limits_{i\in I} A_i} = \bigcup\limits_{i\in I} \overline{A}_i.$

2 Sigma Fields, Probability and Rules of Probability

Definition 2.1. *A collection $\mathcal{K}$ of events from S is said to be a* ***σ-field*** *(**σ-algebra**) over S if it satisfies the following conditions:*

(i) $\mathcal{K} \neq \emptyset$;

(ii) *if $A \in \mathcal{K}$, then $\overline{A} \in \mathcal{K}$;*

(iii) *if $A_n \in \mathcal{K}$ for all $n \in \mathbb{N}$, then $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{K}$.*

If $\mathcal{K}$ is a σ-field over the sample space S, then the pair $(S, \mathcal{K})$ is called a ***measurable space***.

Example 2.2. The power set $\mathcal{P}(S) = \{S' | S' \subseteq S\}$ is a σ-field over S.

Theorem 2.3. *Let $\mathcal{K}$ be a σ-field over S. Then the following properties hold:*

a) $\emptyset, S \in \mathcal{K}$.

b) *for all $A, B \in \mathcal{K}$, $A \cap B,\ A \setminus B,\ A \Delta B \in \mathcal{K}$.*

c) *if $A_n \in \mathcal{K}$, for all $n \in \mathbb{N}$, then $\bigcap\limits_{n=1}^{\infty} A_n \in \mathcal{K}$.*

Definition 2.4. *Let $\mathcal{K}$ be a σ-field over S. A mapping $P : \mathcal{K} \to \mathbb{R}$ is called* ***probability*** *if it satisfies the following conditions:*

(i) $P(S) = 1$;

(ii) $P(A) \geq 0$, *for all $A \in \mathcal{K}$;*

(iii) *for any sequence $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{K}$ of mutually exclusive events,*

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n), \tag{2.1}$$

(P is σ-additive).

The triplet $(S, \mathcal{K}, P)$ is called a ***probability space***.

Theorem 2.5. *Let $(S, \mathcal{K}, P)$ be a probability space, and let $A, B \in \mathcal{K}$. Then the following properties hold:*

a) $P(\overline{A}) = 1 - P(A)$ *and* $0 \le P(A) \le 1$.

b) $P(\emptyset) = 0$.

c) $P(A \setminus B) = P(A) - P(A \cap B)$.

d) *If $A \subseteq B$, then $P(A) \le P(B)$, i.e. P is monotonically increasing.*

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

Proof.
a) We have $A, \overline{A} \in \mathcal{K}$, $A \cup \overline{A} = S$ and $A, \overline{A}$ are mutually exclusive. Then

$$1 = P(S) = P(A \cup \overline{A}) \overset{(2.1)}{=} P(A) + P(\overline{A}),$$

i.e. $P(\overline{A}) = 1 - P(A)$.

Since $P(\overline{A}) \ge 0$, it follows that $P(A) \le 1$, so $0 \le P(A) \le 1$.

b) $P(\emptyset) = P(\overline{S}) = 1 - P(S) = 0$.

c) We have $A = (A \cap B) \cup (A \setminus B)$ and $A \cap B,\ A \setminus B$ are mutually exclusive. Thus

$$P(A) \overset{(2.1)}{=} P(A \cap B) + P(A \setminus B),$$

so $P(A \setminus B) = P(A) - P(A \cap B)$.

d) Since $A \subseteq B$, $A = A \cap B$. Then by c), we have

$$0 \leq P(B \setminus A) = P(B) - P(A),$$

which means $P(A) \leq P(B)$.

e) We have $A \cup B = A \cup (B \setminus (A \cap B))$ and $A, B \setminus (A \cap B)$ are mutually exclusive. Then using part c),

$$\begin{aligned} P(A \cup B) &\overset{(2.1)}{=} P(A) + P(B \setminus (A \cap B)) \\ &\overset{c)}{=} P(A) + P(B) - P(B \cap (A \cap B)) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

□

Part e) of Theorem 2.5 can be generalized to more than two events:

Theorem 2.6. *Let $(S, \mathcal{K}, P)$ be a probability space and $(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{K}$ a sequence of events. Then Poincaré's formula (the inclusion-exclusion principle) holds*

$$\begin{aligned} P(\bigcup_{i=1}^{n} A_i) &= \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) \\ &+ \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) \\ &+ \ldots + (-1)^{n-1} P(\bigcap_{i=1}^{n} A_i), \end{aligned} \tag{2.2}$$

for all $n \in \mathbb{N}$. As a consequence,

$$P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n), \tag{2.3}$$

i.e P is subadditive.

Example 2.7. Let us write formula (2.2) for three events $A, B, C \in \mathcal{K}$.

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

3 Classical Definition of Probability

Each event has an associated quantity which characterizes how likely its occurrence is; this is called the *probability* of the event. The classical definition of probability was given independently by B. Pascal and P. Fermat in the 17th century.

Definition 3.1. *Consider an experiment whose outcomes are finite and equally likely. Then the* ***probability of the occurrence of the event*** A *is given by*

$$P(A) = \frac{\text{number of favorable outcomes for the occurrence of } A}{\text{total number of possible outcomes of the experiment}} \stackrel{not}{=} \frac{N_f}{N_t}. \tag{3.1}$$

Remark 3.2. This approach can be used only when it is reasonable to assume that the possible outcomes of an experiment are equally likely (fair die, fair coin). Also, the two numbers have to be finite. When that is not the case, *geometrical probability* is used, when some continuous measure of a set is used (instead of the cardinality):

$$P(A) = \frac{\mu(A)}{\mu(S)}.$$

Remark 3.3. This notion is closely related to that of *relative frequency* of an event A: repeat an experiment a number of times N and count the number of times event A occurs, N_A. Then the relative frequency of the event A is

$$f_A = \frac{N_A}{N}.$$

Such a number is often used as an approximation to the probability of A. This is justified by the fact that

$$f_A \overset{N\to\infty}{\longrightarrow} P(A).$$

The relative frequency is used in computer simulations of random phenomena.

Example 3.4. Two dice are rolled. Find the probability of the events
A: a double appears;
B: the sum of the two numbers obtained is less than or equal to 5.

Solution We begin by computing the denominator in formula (3.1), because that number is common to both probabilities. The total number of possible outcomes is the number of elements of the sample space. The sample space is

$$S = \{e_{ij} \mid i, j = \overline{1, 6}\},$$

where e_{ij} (identified by the pair (i, j), for simplicity) represents the event that number i showed on the first die and number j on the second. Hence $N_t = 36$.

For event A, $N_f = 6$ (there are six doubles out of 36 possible outcomes), so

$$P(A) = \frac{1}{6}.$$

For event B, we count the number of favorable outcomes (i.e. the number of pairs (i, j) for which $i + j \le 5$). We have

$$\left.\begin{matrix} (1,1) & (1,2) & (1,3) & (1,4) \\ & (2,2) & (2,3) & \end{matrix}\right\} \; 6 \ \text{ outcomes}$$

By symmetry, we have $6 \times 2 = 12$, but two of the pairs were already symmetric, so

$N_f = 12 - 2 = 10$ cases. Thus

$$P(B) = \frac{5}{18}.$$

■