

## 2. AXIOMS OF PROBABILITY

**Definition 2.1.** In an experiment, each outcome is called a **sample point** and the set of all possible outcomes is defined to be the **sample space** and denoted by  $S$ . Any subset of  $S$  is called an **event**.

*Example 2.1.* Consider the experiment of flipping a coin four times. Assume that the head (H) and the tail (T) are the only outcomes of each flip. Then, the sample space is

$$S = \{(a, b, c, d) | a, b, c, d \in \{H, T\}\}.$$

If  $E$  be the event that exactly one head appears, then

$$E = \{(H, T, T, T), (T, H, T, T), (T, T, H, T), (T, T, T, H)\}.$$

*Example 2.2.* Consider the game of flipping four coins at one time. If only the numbers of heads and tails are concerned, then the sample space is

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

Let  $E, F$  be two events.

- The event that  $E$  or  $F$  happens is  $E \cup F$  (the union of  $E$  and  $F$ ).
- The event that both  $E$  and  $F$  happen is  $E \cap F$  (or written by  $EF$ , the intersection of  $E$  and  $F$ ).
- The event that  $F$  happens but  $E$  does not happen is  $F \setminus E = \{x \in F | x \notin E\}$  (the complement of  $E$  in  $F$ ). If  $F = S$ , write  $E^c$  for  $S \setminus E$  and simply call it the complement of  $E$ .
- $E$  and  $F$  are said to be **mutually exclusive** if  $EF = \emptyset$ , the empty set. That is,  $E$  and  $F$  are mutually disjoint.

Laws of the above operations.

- (Commutative laws)  $E \cup F = F \cup E$  and  $EF = FE$ .
- (Associative laws)  $(E \cup F) \cup G = E \cup (F \cup G)$  and  $(EF)G = E(FG)$ .
- (Distributive laws)  $E(F \cup G) = (EF) \cup (EG)$  and  $E \cup (FG) = (E \cup F)(E \cup G)$ .
- (DeMorgan's laws)  $(\bigcup_{n=1}^{\infty} E_n)^c = \bigcap_{n=1}^{\infty} E_n^c$  and  $(\bigcap_{n=1}^{\infty} E_n)^c = \bigcup_{n=1}^{\infty} E_n^c$ .

**Definition 2.2.** Given a sample space  $S$ , a probability  $P$  is a function defined on a structured class ( $\sigma$ -algebra) of subsets of  $S$  which satisfies

- (Axiom 1)  $P(E) \in [0, 1]$  for all  $E \subset S$ .
- (Axiom 2)  $P(S) = 1$ .
- (Axiom 3) If  $E_1, E_2, \dots$  are mutually exclusive events, then

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

By setting  $E_1 = S$  and  $E_n = \emptyset$  for  $n > 1$ , we have

$$P(S) = P(S) + P(\emptyset) + P(\emptyset) + \dots.$$

This implies  $P(\emptyset) = 0$ . Using this fact and Axiom 3, if  $E_1, \dots, E_n$  are mutually exclusive, then

$$P(E_1 \cup \dots \cup E_n) = P(E_1) + \dots + P(E_n).$$

*Remark 2.1.* If  $E$  consists of exactly one sample point, say  $s$ , we briefly write  $P(s)$  for  $P(E) = P(\{s\})$ .

*Example 2.3* (Sample spaces with equally likely outcomes). An experiment with a finite sample space has equally likely outcomes if the probability  $P$  satisfies  $P(s) = 1/|S|$  for all  $s \in S$ . In this case,  $P(E) = |E|/|S|$  for all  $E \subset S$ .

*Example 2.4.* Consider an urn with 5 white balls and 6 black balls. For convenience, we let  $W_1, \dots, W_5$  be the white balls and  $B_1, \dots, B_6$  be the black balls. Let  $S_1$  be the sample space when 3 balls are randomly drawn and  $S_2$  be the sample space when 3 balls are randomly drawn one by one without replacement. Let  $P_1$  and  $P_2$  be the probability on  $S_1$  and  $S_2$ . Consider the event that one ball is white and two balls are black. Let  $E_1, E_2$  be respectively the events in  $S_1$  and  $S_2$ . Then,

$$E_1 = \{\{W_i, B_j, B_k\} | 1 \leq i \leq 5, 1 \leq j \neq k \leq 6\}$$

and

$$E_2 = \{(W_i, B_j, B_k), (B_j, W_i, B_k), (B_j, B_k, W_i) | 1 \leq i \leq 5, 1 \leq j \neq k \leq 6\}.$$

Observe that  $|S_1| = \binom{11}{3} = 165$ ,  $|S_2| = 11 \times 10 \times 9 = 990$ ,  $|E_1| = \binom{5}{1} \binom{6}{2} = 75$  and  $|E_2| = 3 \times (6 \cdot 5 \cdot 5) = 450$ . This implies

$$P_1(E_1) = \frac{75}{165} = \frac{5}{11}, \quad P_2(E_2) = \frac{450}{990} = \frac{5}{11}.$$

*Example 2.5.* Suppose that  $n$  red balls, say  $r_1, \dots, r_n$ , and  $m$  blue balls, say  $b_1, \dots, b_m$ , are arranged in a line. Let  $S$  be the sample space consisting of all the arrangements of balls in a linear order and  $P$  be a probability on  $S$  with equally likely outcomes. Let  $S'$  be the set of all linear arrangements of  $n$  letters “ $r$ ” and  $m$  letters “ $b$ ” and let  $f : S \rightarrow S'$  be the mapping by erasing the subindex. For instance, if  $n = 3$  and  $m = 2$ , then

$$f(r_1, b_2, b_1, r_3, r_2) = (r, b, b, r, r).$$

Define  $P'$  as a probability on  $S'$  by setting

$$P'(s') = P(f^{-1}(\{s'\})) \quad \forall s' \in S',$$

where  $f^{-1}(\{s'\}) = \{s \in S | f(s) = s'\}$ . For  $s' \in S'$ ,  $|f^{-1}(\{s'\})| = n!m!$  and this implies  $P'(s') = n!m!/(n+m)!$ .

*Example 2.6.* A poker hand consists of 5 cards. Regardless of the suits, if the values of cards are distinct and consecutive (including  $\{A, 2, 3, 4, 5\}$  and  $\{10, J, Q, K, A\}$ ), then the poker hand is called a straight. Find the probability that randomly selected five cards form a straight poker hand.

**Solution:** Let  $S$  be the sample space consisting of all possible poker hands. Clearly,  $|S| = \binom{52}{5}$ . Note that there are 10 types of straights, which are  $\{A, 2, 3, 4, 5\}, \dots, \{10, J, Q, K, A\}$ . In a deck of 52 cards, each value has 4 different suits, which are spades, hearts, diamonds and clubs. This implies that there are  $10 \cdot 4^5$  different straights and the probability of a straight poker hand is equal to  $10 \cdot 4^5 / \binom{52}{5} \approx 0.004$ .

*Remark 2.2.* How to determine the value of  $P(E)$  for any event  $E$ ? It is intuitive to consider the ratio  $n(E)/n$ , where  $n(E)$  denotes the number of times  $E$  occurs if the experiment is repeatedly performed under the “same” condition for  $n$  times. Clearly,  $n(E)/n$  refers to the frequency that  $E$  occurs. If the ratio converges, then it is natural to assign the limit to  $P(E)$ . To realize this idea, one has to ensure that the ratio converges with common limit for any sequence of experiments and this is related to the strong law of large numbers.

**Lemma 2.1.** *Let  $E$  and  $F$  be two events.*

- (1)  $P(E^c) = 1 - P(E)$ .
- (2) If  $E \subset F$ , then  $P(E) \leq P(F)$ .
- (3)  $P(E \cup F) = P(E) + P(F) - P(EF)$ .

*Proof.* If  $E \subset F$ , then  $F = E \cup (F \setminus E)$  and this implies  $P(F) = P(E) + P(F \setminus E) \geq P(E)$ . As  $S$  is a mutually disjoint union of  $E$  and  $E^c$ , one has  $1 = P(S) = P(E) + P(E^c)$ . For (3), note that  $E \cup F = E \cup (F \setminus E)$  and  $F = (EF) \cup (F \setminus E)$ . By the additivity of  $P$ , these facts imply

$$P(E \cup F) = P(E) + P(F \setminus E), \quad P(EF) + P(F \setminus E) = P(F).$$

Adding up both equalities leads to the desired identity.  $\square$

**Theorem 2.2** (The inclusion-exclusion formula). *Let  $E_1, \dots, E_n$  be events and  $E = E_1 \cup E_2 \cup \dots \cup E_n$ . Then,*

$$\begin{aligned} P(E) &= \sum_{m=1}^n (-1)^{m+1} \sum_{\substack{i_1, \dots, i_m \in \mathbb{N} \\ 1 \leq i_1 < \dots < i_m \leq n}} P(E_{i_1} \cdots E_{i_m}) \\ &= \sum_{i=1}^n P(E_i) - \sum_{1 \leq i < j \leq n} P(E_i E_j) + \sum_{1 \leq i < j < k \leq n} P(E_i E_j E_k) + \dots + (-1)^{n+1} P(E_1 \cdots E_n). \end{aligned}$$

*Remark 2.3.* Note that

$$E = \bigcup_{i=1}^n E_i = E_1 \cup (E_1^c E_2) \cup (E_1^c E_2^c E_3) \cup \dots \cup (E_1^c \cdots E_{n-1}^c E_n).$$

By setting  $B_1 = S$  and  $B_i = E_1^c \cdots E_{i-1}^c$ , one may rewrite  $E = \bigcup_{i=1}^n (B_i E_i)$ . For  $i < j$ ,  $(B_i E_i)(B_j E_j) \subset E_i B_j = \emptyset$ . This implies

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

Next, for  $i \geq 2$ ,

$$P(B_i E_i) = P(E_i) - P(B_i^c E_i) = P(E_i) - P\left(E_i \bigcup_{j=1}^{i-1} E_j\right).$$

As a result, we obtain

$$P(E) = \sum_{i=1}^n P(E_i) - \sum_{i=2}^n P\left(\bigcup_{j=1}^{i-1} (E_i E_j)\right) \geq \sum_{i=1}^n P(E_i) - \sum_{1 \leq j < i \leq n} P(E_i E_j).$$

Inductively, one can show that if  $a_m = \sum_{1 \leq i_1 < \dots < i_m \leq n} P(E_{i_1} \cdots E_{i_m})$ , then

$$P(E) \leq a_1, \quad P(E) \geq a_1 - a_2, \quad P(E) \leq a_1 - a_2 + a_3, \quad \dots$$

*Example 2.7* (The matching problem). At a party,  $N$  men throw their hats to the center of the room and mix up. Then, each man selects a hat randomly. What is the probability that no one selects his own hat?

**Solution:** Let  $E_i$  be the event that the  $i$ th man selects his own hat and  $E = E_1 \cup \dots \cup E_N$ . Note that, for any  $1 \leq k \leq N$  and  $1 \leq i_1 < i_2 < \dots < i_k \leq N$ ,

$$P(E_{i_1} \cdots E_{i_k}) = \frac{(N-k)!}{N!}.$$

This implies

$$P(E) = \sum_{k=1}^N (-1)^{k+1} \binom{N}{k} \times \frac{(N-k)!}{N!} = \sum_{k=1}^N \frac{(-1)^{k+1}}{k!}.$$

As  $N \rightarrow \infty$ , we have

$$P(\{\text{no one selects his own hat}\}) = 1 - \sum_{k=1}^N \frac{(-1)^{k+1}}{k!} = \sum_{k=0}^N \frac{(-1)^k}{k!} \rightarrow \frac{1}{e} \approx 0.37.$$

In fact, the remainder estimate of alternating series gives

$$|P(\{\text{no one selects his own hat}\}) - 1/e| \leq \frac{1}{N!}.$$

**Proposition 2.3.** *Let  $P$  be a probability on  $S$  and let  $(E_n)_{n=1}^{\infty}$  and  $(F_n)_{n=1}^{\infty}$  be events satisfying  $E_n \subset E_{n+1}$  and  $F_n \supset F_{n+1}$ . Then,*

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n \rightarrow \infty} P(E_n), \quad P\left(\bigcap_{n=1}^{\infty} F_n\right) = \lim_{n \rightarrow \infty} P(F_n).$$

*Proof.* By DeMorgan's laws, the second identity is a corollary of the first one. To prove the first identity, set  $G_1 = E_1$  and  $G_n = E_n \setminus E_{n-1}$  for  $n > 1$ . Note that  $G_n$ 's are mutually exclusive,  $E_n = G_1 \cup \dots \cup G_n$  and  $\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} G_n$ . By Axiom 3 of probabilities, this implies

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(G_n) = \lim_{n \rightarrow \infty} \sum_{k=1}^n P(G_k) = \lim_{n \rightarrow \infty} P(E_n).$$

□

*Example 2.8.* Consider an experiment performed between 11 : 59 and 12 : 00. At 1 minute to 12, we put  $N$  balls indexed from 1 to  $N$  into an urn and then randomly withdraw one. At 1/2 minute to 12, we put another  $N$  balls indexed from  $N + 1$  to  $2N$  into the urn and then randomly withdraw one. At 1/4 minute to 12, we do the same thing as before and then keep going on. Let  $E$  be the event that ball 1 is in the urn at 12 and  $E_n$  is the event that ball 1 is in the urn at  $2^{1-n}$  minute to 12. It is clear that  $E_n \supset E_{n+1}$  and  $E = \bigcap_{n=1}^{\infty} E_n$ . If  $P$  is the probability on the outcomes of the above experiment, then

$$P(E) = \lim_{n \rightarrow \infty} P(E_n), \quad P(E_n) = \frac{(N-1)(2N-2)(3N-3) \cdots (nN-n)}{N(2N-1)(3N-2) \cdots (nN-n+1)}.$$

Note that

$$\frac{1}{P(E_n)} = \prod_{k=1}^n \left(1 + \frac{1}{k(N-1)}\right) > \sum_{k=1}^n \frac{1}{k(N-1)} = \frac{1}{N-1} \sum_{k=1}^n \frac{1}{k} \rightarrow \infty$$

as  $n \rightarrow \infty$ . This implies  $P(E) = 0$ . In a similar argument, one can prove that, for  $m \geq 1$ , if  $F_m$  is the event that ball  $m$  is in the urn at 12, then  $P(F_m) = 0$ . As a consequence, we obtain

$$P(\{\text{The urn is nonempty}\}) = P\left(\bigcup_{m=1}^{\infty} F_m\right) \leq \sum_{m=1}^{\infty} P(F_m) = 0.$$

Thus,  $P(\{\text{The urn is empty}\}) = 1$ .