

Chapter 3

Basic Concepts of Probability

The theory of probability provides the mathematical tools necessary to study and analyze uncertain phenomena that occur in nature. It establishes a formal framework to understand and predict the outcome of a random experiment. It can be used to model complex systems and characterize stochastic processes. This is instrumental in designing efficient solutions to many engineering problems. Two components define a probabilistic model, a sample space and a probability law.

3.1 Sample Spaces and Events

In the context of probability, an *experiment* is a random occurrence that produces one of several *outcomes*. The set of all possible outcomes is called the *sample space* of the experiment, and it is denoted by Ω . An *admissible* subset of the sample space is called an *event*.

Example 17. *The rolling of a die forms a common experiment. A sample space Ω corresponding to this experiment is given by the six faces of a die. The set of prime numbers less than or equal to six, namely $\{2, 3, 5\}$, is one of many possible events. The actual number observed after rolling the die is the outcome of the experiment.*

There is essentially no restriction on what constitutes an experiment. The flipping of a coin, the flipping of n coins, and the tossing of an infinite sequence of coins are all random experiments. Also, two similar experiments may have

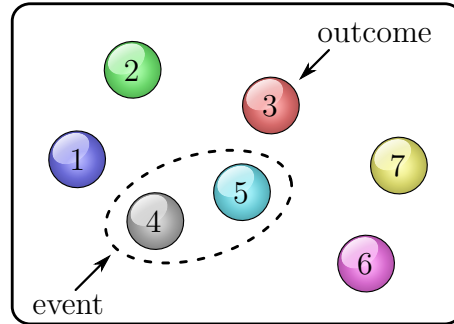


Figure 3.1: A sample space contains all the possible outcomes; an admissible subset of the sample space is called an event.

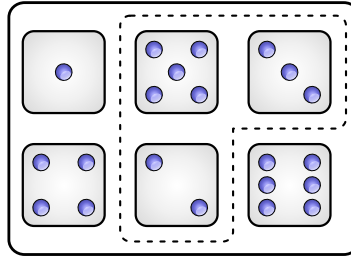


Figure 3.2: A possible sample space for the rolling of a die is $\Omega = \{1, 2, \dots, 6\}$, and the subset $\{2, 3, 5\}$ forms a specific event.

different sample spaces. A sample space Ω for observing the number of heads in n tosses of a coin is $\{0, 1, \dots, n\}$; however, when describing the complete history of the n coin tosses, the sample space becomes much larger with 2^n distinct sequences of heads and tails. Ultimately, the choice of a particular sample space depends on the properties one wishes to analyze. Yet some rules must be followed in selecting a sample space.

1. The elements of a sample space should be *distinct* and *mutually exclusive*. This ensures that the outcome of an experiment is unique.
2. A sample space must be *collectively exhaustive*. That is, every possible outcome of the experiment must be accounted for in the sample space.

In general, a sample space should be precise enough to distinguish between all outcomes of interest, while avoiding frivolous details.

Example 18. Consider the space composed of the odd integers located between one and six, the even integers contained between one and six, and the prime numbers less than or equal to six. This collection cannot be a sample space for the rolling of a die because its elements are not mutually exclusive. In particular, the numbers three and five are both odd and prime, while the number two is prime and even. This violates the uniqueness criterion.

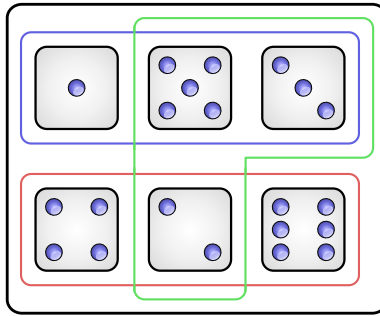


Figure 3.3: This collection of objects cannot be a sample space as the three proposed outcomes (even, odd and prime) are not mutually exclusive.

Alternatively, the elements of the space composed of the odd numbers between one and six, and the even numbers between one and six, are distinct and mutually exclusive; an integer cannot be simultaneously odd and even. Moreover, this space is collectively exhaustive because every integer is either odd or even. This latter description forms a possible sample space for the rolling of a die.

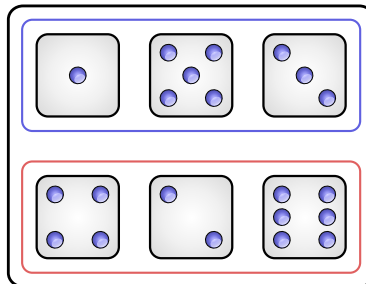


Figure 3.4: A candidate sample space for the rolling of a die is composed of two objects, the odd numbers and the even numbers between one and six.

3.2 Probability Laws

A *probability law* specifies the likelihood of events related to an experiment. Formally, a probability law assigns to every event A a number $\Pr(A)$, called the *probability of event A* , such that the following axioms are satisfied.

1. (**Nonnegativity**) $\Pr(A) \geq 0$, for every event A .
2. (**Normalization**) The probability of the sample space Ω is equal to one,

$$\Pr(\Omega) = 1.$$

3. (**Countable Additivity**) If A and B are disjoint events, $A \cap B = \emptyset$, then the probability of their union satisfies

$$\Pr(A \cup B) = \Pr(A) + \Pr(B).$$

More generally, if A_1, A_2, \dots is a sequence of disjoint events and $\bigcup_{k=1}^{\infty} A_k$ is itself an admissible event then

$$\Pr\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \Pr(A_k).$$

A number of important properties can be deduced from the three axioms of probability. We prove two such properties below. The first statement describes the relation between inclusion and probabilities.

Proposition 1. *If $A \subset B$, then $\Pr(A) \leq \Pr(B)$.*

Proof. Since $A \subset B$, we have $B = A \cup (B - A)$. Noting that A and $B - A$ are disjoint sets, we get

$$\Pr(B) = \Pr(A) + \Pr(B - A) \geq \Pr(A),$$

where the inequality follows from the nonnegativity of probability laws. \square

Our second result specifies the probability of the union of two events that are not necessarily mutually exclusive.

Proposition 2. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.

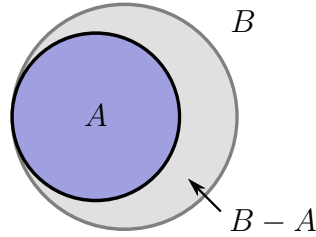


Figure 3.5: If event A is a subset of event B , then the probability of A is less than or equal to the probability of B .

Proof. Using the third axiom of probability on the disjoint sets A and $(A \cup B) - A$, we can write

$$\Pr(A \cup B) = \Pr(A) + \Pr((A \cup B) - A) = \Pr(A) + \Pr(B - A).$$

Similarly, applying the third axiom to $A \cap B$ and $B - (A \cap B)$, we obtain

$$\Pr(B) = \Pr(A \cap B) + \Pr(B - (A \cap B)) = \Pr(A \cap B) + \Pr(B - A).$$

Combining these two equations yields the desired result. \square

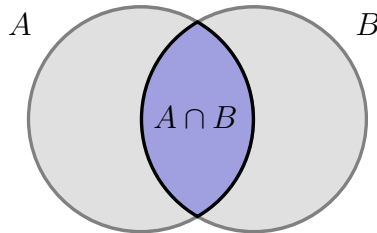


Figure 3.6: The probability of the union of A and B is equal to the probability of A plus the probability of B minus the probability of their intersection.

The statement of Proposition 2 can be extended to finite unions of events. Specifically, we can write

$$\Pr\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{\mathbb{I} \subset \{1, \dots, n\}, |\mathbb{I}|=k} \Pr\left(\bigcap_{i \in \mathbb{I}} A_i\right)$$

where the rightmost sum runs over all subsets \mathbb{I} of $\{1, \dots, n\}$ that contain exactly k elements. This more encompassing result is known as the *inclusion-exclusion principle*.

We can use Proposition 2 recursively to derive a bound on the probabilities of unions. This theorem, which is sometimes called the *Boole inequality*, asserts that the probability of at least one event occurring is no greater than the sum of the probabilities of the individual events.

Theorem 1 (Union Bound). *Let A_1, A_2, \dots, A_n be a collection of events, then*

$$\Pr\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n \Pr(A_k). \quad (3.1)$$

Proof. We show this result using induction. First, we note that the claim is trivially true for $n = 1$. As an inductive hypothesis, assume that (3.1) holds for some $n \geq 1$. Then, we have

$$\begin{aligned} \Pr\left(\bigcup_{k=1}^{n+1} A_k\right) &= \Pr\left(A_{n+1} \cup \left(\bigcup_{k=1}^n A_k\right)\right) \\ &= \Pr(A_{n+1}) + \Pr\left(\bigcup_{k=1}^n A_k\right) - \Pr\left(A_{n+1} \cap \left(\bigcup_{k=1}^n A_k\right)\right) \\ &\leq \Pr(A_{n+1}) + \Pr\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^{n+1} \Pr(A_k). \end{aligned}$$

Therefore, by the principle of mathematical induction, (3.1) is valid for all positive integers. \square

The union bound is often employed in situations where finding the joint probability of multiple rare events is difficult, but computing the probabilities of the individual components is straightforward.

Example 19. *An urn contains 990 blue balls and 10 red balls. Five people each pick a ball at random, without replacement. We wish to compute the probability that at least one person picks a red ball. Let B_k denote the event that person k draws a red ball. We note that the probability of interest can be written as $\Pr\left(\bigcup_{k=1}^5 B_k\right)$.*

First, let us approximate this probability using the union bound. Clearly, the probability that the first person picks a red ball, $\Pr(B_1)$, is equal to $1/100$. As we will see later, it turns out that $\Pr(B_k) = 1/100$ for $k = 1, \dots, 5$. Applying

(3.1), we get a bound on the probability that at least one person picks a red ball,

$$\Pr\left(\bigcup_{k=1}^5 B_k\right) \leq \sum_{k=1}^5 \Pr(B_k) = \sum_{k=1}^5 \frac{1}{100} = \frac{1}{20}.$$

We can also compute this probability exactly. The probability that no red balls are selected is given by $\binom{990}{5}/\binom{1000}{5}$. Hence, the probability that a least one person draws a red ball becomes

$$\Pr\left(\bigcup_{k=1}^5 B_k\right) = 1 - \frac{\binom{990}{5}}{\binom{1000}{5}} \approx 0.0491.$$

As a second application, consider the problem where the five people each draw two balls from the urn, without replacement. This time, we wish to approximate the probability that at least one person gets two red balls. Using the same steps as before, we get

$$\Pr\left(\bigcup_{k=1}^5 C_k\right) \leq \sum_{k=1}^5 \Pr(C_k) = \sum_{k=1}^5 \frac{\binom{10}{2}}{\binom{1000}{2}} = \frac{1}{2220},$$

where C_k represents the event that person k draws two red balls. In this latter scenario, computing the exact probability is much more challenging.

3.2.1 Finite Sample Spaces

If a sample space Ω contains a finite number of elements, then a probability law on Ω is completely determined by the probabilities of its individual outcomes. Denote a sample space containing n elements by $\Omega = \{s_1, s_2, \dots, s_n\}$. Any event in Ω is of the form $A = \{s_i \in \Omega | i \in \mathbb{I}\}$, where \mathbb{I} is a subset of the integers one through n . The probability of event A is therefore given by the third axiom of probability,

$$\Pr(A) = \Pr(\{s_i \in \Omega | i \in \mathbb{I}\}) = \sum_{i \in \mathbb{I}} \Pr(s_i).$$

We emphasize that, by definition, distinct outcomes are always disjoint events.

If in addition the elements of Ω are equally likely with

$$\Pr(s_1) = \Pr(s_2) = \dots = \Pr(s_n) = \frac{1}{n},$$

then the probability of an event A becomes

$$\Pr(A) = \frac{|A|}{n} \quad (3.2)$$

where $|A|$ denotes the number of elements in A .

Example 20. *The rolling of a fair die is an experiment with a finite number of equally likely outcomes. The probability of observing any of the faces labeled one through six is therefore equal to $1/6$. The probability of any event can easily be computed by counting the number of distinct outcomes included in the event. For instance, the probability of rolling a prime number is*

$$\Pr(\{2, 3, 5\}) = \Pr(2) + \Pr(3) + \Pr(5) = \frac{3}{6}.$$

3.2.2 Countably Infinite Models

Consider a sample space that consists of a countably infinite number of elements, $\Omega = \{s_1, s_2, \dots\}$. Again, a probability law on Ω is specified by the probabilities of individual outcomes. An event in Ω can be written as $A = \{s_j \in \Omega | j \in \mathbb{J}\}$, where \mathbb{J} is a subset of the positive integers. Using the third axiom of probability, $\Pr(A)$ can be written as

$$\Pr(A) = \Pr(\{s_j \in \Omega | j \in \mathbb{J}\}) = \sum_{j \in \mathbb{J}} \Pr(s_j).$$

The possibly infinite sum $\sum_{j \in \mathbb{J}} \Pr(s_j)$ always converges since the summands are nonnegative and the sum is bounded above by one; it is consequently well defined.



Figure 3.7: A countable set is a collection of elements with the same cardinality as some subset of the natural numbers.

Example 21. *Suppose that a fair coin is tossed repetitively until heads is observed. The number of coin tosses is recorded as the outcome of this experiment. A natural sample space for this experiment is $\Omega = \{1, 2, \dots\}$, a countably infinite set.*

The probability of observing heads on the first trial is $1/2$, and the probability of observing heads for the first time on trial k is 2^{-k} . The probability of the entire sample space is therefore equal to

$$\Pr(\Omega) = \sum_{k=1}^{\infty} \Pr(k) = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1,$$

as expected. Similarly, the probability of the number of coin tosses being even can be computed as

$$\Pr(\{2, 4, 6, \dots\}) = \sum_{k=1}^{\infty} \Pr(2k) = \sum_{k=1}^{\infty} \frac{1}{2^{2k}} = \frac{1}{4} \frac{1}{(1 - \frac{1}{4})} = \frac{1}{3}.$$

3.2.3 Uncountably Infinite Models

Probabilistic models with uncountably infinite sample spaces differ from the finite and countable cases in that a probability law may not necessarily be specified by the probabilities of single-element outcomes. This difficulty arises from the large number of elements contained in the sample space when the latter is uncountable. Many subsets of Ω do not have a finite or countable representation, and as such the third axiom of probability cannot be applied to relate the probabilities of these events to the probabilities of individual outcomes. Despite these apparent difficulties, probabilistic models with uncountably infinite sample spaces are quite useful in practice. To develop an understanding of uncountable probabilistic models, we consider the unit interval $[0, 1]$.



Figure 3.8: The unit interval $[0, 1]$, composed of all real numbers between zero and one, is an example of an uncountable set.

Suppose that an element is chosen at random from this interval, with uniform weighting. By the first axiom of probability, the probability that this element belongs to the interval $[0, 1]$ is given by $\Pr([0, 1]) = 1$. Furthermore, if two intervals have the same length, the probabilities of the outcome falling

in either interval should be identical. For instance, it is natural to anticipate that $\Pr((0, 0.25)) = \Pr((0.75, 1))$.

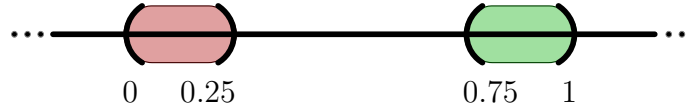


Figure 3.9: If the outcome of an experiment is uniformly distributed over $[0, 1]$, then two subintervals of equal lengths should have the same probabilities.

In an extension of the previous observation, we take the probability of an open interval (a, b) where $0 \leq a < b \leq 1$ to equal

$$\Pr((a, b)) = b - a. \quad (3.3)$$

Using the third axiom of probability, it is then possible to find the probability of a finite or countable union of disjoint open intervals.

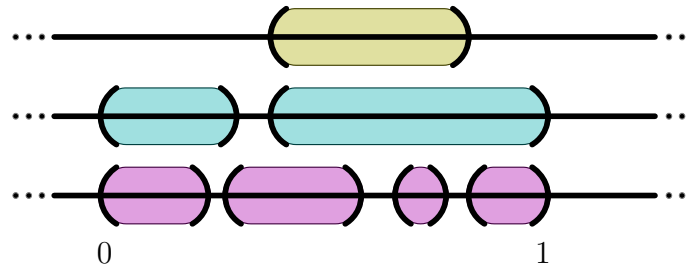


Figure 3.10: The probabilities of events that are formed through the union of disjoint intervals can be computed in a straightforward manner.

Specifically, for constants $0 \leq a_1 < b_1 < a_2 < b_2 < \dots \leq 1$, we get

$$\Pr\left(\bigcup_{k=1}^{\infty} (a_k, b_k)\right) = \sum_{k=1}^{\infty} (b_k - a_k).$$

The probabilities of more complex events can be obtained by applying additional elementary set operations. However, it suffices to say for now that specifying the probability of the outcome falling in (a, b) for every possible open interval is enough to define a probability law on Ω . In the example at hand, (3.3) completely determines the probability law on $[0, 1]$.

Note that we can give an alternative means of computing the probability of an interval. Again, consider the open interval (a, b) where $0 \leq a < b \leq 1$. The probability of the outcome falling in this interval is equal to

$$\Pr((a, b)) = b - a = \int_a^b dx = \int_{(a,b)} dx.$$

Moreover, for $0 \leq a_1 < b_1 < a_2 < b_2 < \dots \leq 1$, we can write

$$\Pr\left(\bigcup_{k=1}^{\infty} (a_k, b_k)\right) = \sum_{k=1}^{\infty} (b_k - a_k) = \int_{\bigcup_{k=1}^{\infty} (a_k, b_k)} dx.$$

For this carefully crafted example, it appears that the probability of an admissible event A is given by the integral

$$\Pr(A) = \int_A dx.$$

This is indeed accurate for the current scenario. In fact, the class of admissible events for this experiment is simply the collection of all sets for which the integral $\int_A dx$ can be computed. In other words, if a number is chosen at random from $[0, 1]$, then the probability of this number falling in set $A \subset [0, 1]$ is

$$\Pr(A) = \int_A dx.$$

This method of computing probabilities can be extended to more complicated problems. In these notes, we will see many probabilistic models with uncountably infinite sample spaces. The mathematical tools required to handle such models will be treated alongside.

Example 22. *Suppose that a participant at a game-show is required to spin the wheel of serendipity, a perfect circle with unit radius. When subjected to a vigorous spin, the wheel is equally likely to stop anywhere along its perimeter. A sampling space for this experiment is the collection of all angles from 0 to 2π , an uncountable set. The probability of Ω is invariably equal to one, $\Pr([0, 2\pi)) = 1$.*

The probability that the wheel stops in the first quadrant is given by

$$\Pr\left(\left[0, \frac{\pi}{2}\right)\right) = \int_0^{\frac{\pi}{2}} \frac{1}{2\pi} d\theta = \frac{1}{4}.$$

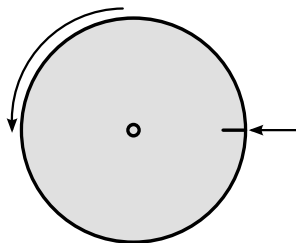


Figure 3.11: The wheel of serendipity forms an example of a random experiment for which the sample space is uncountable.

More generally, the probability that the wheel stops in an interval (a, b) where $0 \leq a \leq b < 2\pi$ can be written as

$$\Pr((a, b)) = \frac{b - a}{2\pi}.$$

If $B \subset [0, 2\pi)$ is a set representing all winning outcomes, then the probability of success at the wheel becomes

$$\Pr(B) = \int_B \frac{1}{2\pi} d\theta.$$

3.2.4 Probability and Measure Theory*

A thorough treatment of probability involves advanced mathematical concepts, especially when it comes to infinite sample spaces. The basis of our intuition for the infinite is the set of *natural numbers*,

$$\mathbb{N} = \{1, 2, \dots\}.$$

Two sets are said to have the same *cardinality* if their elements can be put in one-to-one correspondence. A set with the same cardinality as a subset of the natural numbers is said to be *countable*. That is, the elements of a countable set can always be listed in sequence, s_1, s_2, \dots ; although the order may have nothing to do with any relation between the elements. The integers and the rational numbers are examples of countably infinite sets. It may be surprising at first to learn that there exist uncountable sets. To escape beyond the countable, one needs set theoretic tools such as *power sets*. The set of real numbers is uncountably infinite; it cannot be put in one-to-one correspondence

with the natural numbers. A typical progression in analysis consists of using the finite to gain intuition about the countably infinite, and then to employ the countably infinite to get at the uncountable.

It is tempting to try to assign probabilities to every subset of a sample space Ω . However, for uncountably infinite sample spaces, this leads to serious difficulties that cannot be resolved. In general, it is necessary to work with special subclasses of the class of all subsets of a sample space Ω . The collections of the appropriate kinds are called fields and σ -fields, and they are studied in *measure theory*. This leads to measure-theoretic probability, and to its unified treatment of the discrete and the continuous.

Fortunately, it is possible to develop a working understanding of probability without worrying excessively about these issues. At some point in your academic career, you may wish to study analysis and measure theory more carefully and in greater details. However, it is not our current purpose to initiate the rigorous treatment of these topics.

Further Reading

1. Ross, S., *A First Course in Probability*, 7th edition, Pearson Prentice Hall, 2006: Chapter 2.
2. Bertsekas, D. P., and Tsitsiklis, J. N., *Introduction to Probability*, Athena Scientific, 2002: Section 1.2.
3. Miller, S. L., and Childers, D. G., *Probability and Random Processes with Applications to Signal Processing and Communications*, 2004: Sections 2.1–2.3.
4. Gubner, J. A., *Probability and Random Processes for Electrical and Computer Engineers*, Cambridge, 2006: Sections 1.1, 1.3–1.4.

