Probability

INTRODUCTION

Probability is the study of random or nondeterministic experiments. If a die is tossed in the air, then it is certain that the die will come down, but it is not certain that, say, a 6 will appear. However, suppose we repeat this experiment of tossing a die; let s be the number of successes, i.e. the number of times a 6 appears, and let n be the number of tosses. Then it has been empirically observed that the ratio $f = s/n$, called the *relative frequency*, becomes stable in the long run, i.e. approaches a limit. This stability is the basis of probability theory.

In probability theory, we define a mathematical model of the above phenomenon by assigning "probabilities" (or: the limit values of the relative frequencies) to the "events" connected with an experiment. Naturally, the reliability of our mathematical model for a given experiment depends upon the closeness of the assigned probabilities to the actual relative frequency. This then gives rise to problems of testing and reliability which form the subject matter of statistics.

Historically, probability theory began with the study of games of chance, such as roulette and cards. The probability p of an event A was defined as follows: if A can occur in s ways out of a total of n equally likely ways, then

$$p = P(A) = \frac{s}{n}$$

For example, in tossing a die an even number can occur in 3 ways out of 6 "equally likely" ways; hence $p = \frac{3}{6} = \frac{1}{2}$. This classical definition of probability is essentially circular since the idea of "equally likely" is the same as that of "with equal probability" which has not been defined. The modern treatment of probability theory is purely axiomatic. This means that the probabilities of our events can be perfectly arbitrary, except that they must satisfy certain axioms listed below. The classical theory will correspond to the special case of so-called *equiprobable spaces.*

SAMPLE SPACE AND EVENTS

The set S of all possible outcomes of some given experiment is called the *sample space*. A particular outcome, i.e. an element in S, is called a *sample point* or *sample*. An *event* A is a set of outcomes or, in other words, a subset of the sample space S. The event $\{a\}$ consisting of a single sample $a \in S$ is called an *elementary event*. The empty set $\emptyset$ and S itself are events; $\emptyset$ is sometimes called the *impossible* event, and S the *certain* or *sure* event.

We can combine events to form new events using the various set operations:

(i) $A \cup B$ is the event that occurs iff A occurs *or* B occurs (or both);

(ii) $A \cap B$ is the event that occurs iff A occurs *and* B occurs;

(iii) A^c, the complement of A, is the event that occurs iff A does *not* occur.

Two events A and B are called *mutually exclusive* if they are disjoint, i.e. if $A \cap B = \emptyset$. In other words, A and B are mutually exclusive if they cannot occur simultaneously.

Example 1: Experiment: Toss a die and observe the number that appears on top. Then the sample space consists of the six possible numbers:

$$S = \{1, 2, 3, 4, 5, 6\}$$

Let A be the event that an even number occurs, B that an odd number occurs and C that a prime number occurs:

$$A = \{2, 4, 6\}, \quad B = \{1, 3, 5\}, \quad C = \{2, 3, 5\}$$

Then:

$A \cup C = \{2, 3, 4, 5, 6\}$ is the event that an even or a prime number occurs;

$B \cap C = \{3, 5\}$ is the event that an odd prime number occurs;

$C^c = \{1, 4, 6\}$ is the event that a prime number does not occur.

Note that A and B are mutually exclusive: $A \cap B = \emptyset$; in other words, an even number and an odd number cannot occur simultaneously.

Example 2: Experiment: Toss a coin 3 times and observe the sequence of heads (H) and tails (T) that appears. The sample space S consists of eight elements:

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Let A be the event that two or more heads appear consecutively, and B that all the tosses are the same:

$$A = \{HHH, HHT, THH\} \quad \text{and} \quad B = \{HHH, TTT\}$$

Then $A \cap B = \{HHH\}$ is the elementary event in which only heads appear. The event that 5 heads appear is the empty set $\emptyset$.

Example 3 : Experiment: Toss a coin until a head appears and then count the number of times the coin was tossed. The sample space of this experiment is $S = \{1, 2, 3, \ldots, \infty\}$. Here ∞ refers to the case when a head never appears and so the coin is tossed an infinite number of times. This is an example of a sample space which is *countably infinite*.

AXIOMS OF PROBABILITY

Let S be a sample space, let $\mathcal{E}$ be the class of events, and let P be a real-valued function defined on $\mathcal{E}$. Then P is called a *probability function*, and $P(A)$ is called the *probability* of the event A if the following axioms hold:

[$\mathbf{P}_1$] For every event A, $0 \leq P(A) \leq 1$.

[$\mathbf{P}_2$] $P(S) = 1$.

[$\mathbf{P}_3$] If A and B are mutually exclusive events, then

$$P(A \cup B) = P(A) + P(B)$$

[$\mathbf{P}_4$] If $A_1, A_2, \ldots$ is a sequence of mutually exclusive events, then

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$$

The following remarks concerning the axioms [$\mathbf{P}_3$] and [$\mathbf{P}_4$] are in order. First of all, using [$\mathbf{P}_3$] and mathematical induction we can prove that for any mutually exclusive events $A_1, A_2, \ldots, A_n$,

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n) \qquad (*)$$

We emphasize that [$\mathbf{P}_4$] does not follow from [$\mathbf{P}_3$] even though (*) holds for every positive integer n. However, if the sample space S is finite, then clearly the axiom [$\mathbf{P}_4$] is superfluous.

We now prove a number of theorems which follow directly from our axioms.

Theorem 1: If $\emptyset$ is the empty set, then $P(\emptyset) = 0$.

Proof: Let A be any set; then A and $\emptyset$ are disjoint and $A \cup \emptyset = A$. By [$\mathbf{P}_3$],

$$P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset)$$

Subtracting $P(A)$ from both sides gives our result.

Theorem 2: If A^c is the complement of an event A, then $P(A^c) = 1 - P(A)$.

Proof: The sample space S can be decomposed into the mutually exclusive events A and A^c; that is, $S = A \cup A^c$. By [$\mathbf{P}_2$] and [$\mathbf{P}_3$] we obtain

$$1 = P(S) = P(A \cup A^c) = P(A) + P(A^c)$$

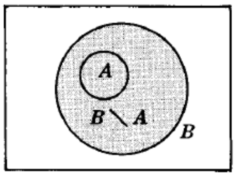
from which our result follows.

Theorem 3: If $A \subset B$, then $P(A) \leq P(B)$.

Proof. If $A \subset B$, then B can be decomposed into the mutually exclusive events A and $B \setminus A$ (as illustrated on the right). Thus

$$P(B) = P(A) + P(B \setminus A)$$

The result now follows from the fact that $P(B \setminus A) \geq 0$.



B is shaded.

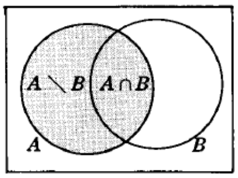
Theorem 4: If A and B are any two events, then

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

Proof. Now A can be decomposed into the mutually exclusive events $A \setminus B$ and $A \cap B$; that is, $A = (A \setminus B) \cup (A \cap B)$. Thus by [$\mathbf{P}_3$],

$$P(A) = P(A \setminus B) + P(A \cap B)$$

from which our result follows.



A is shaded.

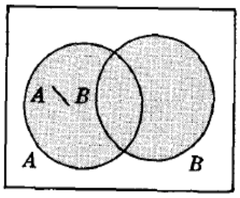
Theorem 5: If A and B are any two events, then

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

Proof. Note that $A \cup B$ can be decomposed into the mutually exclusive events $A \setminus B$ and B; that is, $A \cup B = (A \setminus B) \cup B$. Thus by [$\mathbf{P}_3$] and Theorem 3.4,

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

which is the desired result.



$A \cup B$ is shaded.

Applying the above theorem twice (Problem 3.23) we obtain

Corollary 6: For any events A, B and C,

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

FINITE PROBABILITY SPACES

Let S be a finite sample space; say, $S = \{a_1, a_2, \ldots, a_n\}$. A finite probability space is obtained by assigning to each point $a_i \in S$ a real number p_i, called the *probability* of a_i, satisfying the following properties:

(i) each p_i is nonnegative, $p_i \geq 0$

(ii) the sum of the p_i is one, $p_1 + p_2 + \cdots + p_n = 1$.

The *probability* $P(A)$ of any event A, is then defined to be the sum of the probabilities of the points in A. For notational convenience we write $P(a_i)$ for $P(\{a_i\})$.

Example 5: Let three coins be tossed and the number of heads observed; then the sample space is $S = \{0, 1, 2, 3\}$. We obtain a probability space by the following assignment

$$P(0) = \tfrac{1}{8}, \quad P(1) = \tfrac{3}{8}, \quad P(2) = \tfrac{3}{8} \quad \text{and} \quad P(3) = \tfrac{1}{8}$$

since each probability is nonnegative and the sum of the probabilities is 1. Let A be the event that at least one head appears and let B be the event that all heads or all tails appear:

$$A = \{1, 2, 3\} \quad \text{and} \quad B = \{0, 3\}$$

Then, by definition,

$$P(A) = P(1) + P(2) + P(3) = \tfrac{3}{8} + \tfrac{3}{8} + \tfrac{1}{8} = \tfrac{7}{8}$$

and $$P(B) = P(0) + P(3) = \tfrac{1}{8} + \tfrac{1}{8} = \tfrac{1}{4}$$

Example 6: Three horses A, B and C are in a race; A is twice as likely to win as B and B is twice as likely to win as C. What are their respective probabilities of winning, i.e. $P(A)$, $P(B)$ and $P(C)$?

Let $P(C) = p$; since B is twice as likely to win as C, $P(B) = 2p$; and since A is twice as likely to win as B, $P(A) = 2P(B) = 2(2p) = 4p$. Now the sum of the probabilities must be 1; hence

$$p + 2p + 4p = 1 \quad \text{or} \quad 7p = 1 \quad \text{or} \quad p = \tfrac{1}{7}$$

Accordingly,

$$P(A) = 4p = \tfrac{4}{7}, \quad P(B) = 2p = \tfrac{2}{7}, \quad P(C) = p = \tfrac{1}{7}$$

Question: What is the probability that B or C wins, i.e. $P(\{B, C\})$? By definition

$$P(\{B, C\}) = P(B) + P(C) = \tfrac{2}{7} + \tfrac{1}{7} = \tfrac{3}{7}$$

FINITE EQUIPROBABLE SPACES

Frequently, the physical characteristics of an experiment suggest that the various outcomes of the sample space be assigned equal probabilities. Such a finite probability space S, where each sample point has the same probability, will be called an *equiprobable* or *uniform space*. In particular, if S contains n points then the probability of each point is $1/n$. Furthermore, if an event A contains r points then its probability is $r \cdot \frac{1}{n} = \frac{r}{n}$. In other words,

$$P(A) = \frac{\text{number of elements in } A}{\text{number of elements in } S}$$

or

$$P(A) = \frac{\text{number of ways that the event } A \text{ can occur}}{\text{number of ways that the sample space } S \text{ can occur}}$$

examples

Let a die be weighted so that the probability of a number appearing when the die is tossed is proportional to the given number (e.g. 6 has twice the probability of appearing as 3). Let $A = \{\text{even number}\}$, $B = \{\text{prime number}\}$, $C = \{\text{odd number}\}$.

(i) Describe the probability space, i.e. find the probability of each sample point.

(ii) Find $P(A)$, $P(B)$ and $P(C)$.

(iii) Find the probability that: (*a*) an even or prime number occurs; (*b*) an odd prime number occurs; (*c*) A but not B occurs.

(i) Let $P(1) = p$. Then $P(2) = 2p$, $P(3) = 3p$, $P(4) = 4p$, $P(5) = 5p$ and $P(6) = 6p$. Since the sum of the probabilities must be one, we obtain $p + 2p + 3p + 4p + 5p + 6p = 1$ or $p = 1/21$. Thus

$$P(1) = \tfrac{1}{21}, \quad P(2) = \tfrac{2}{21}, \quad P(3) = \tfrac{1}{7}, \quad P(4) = \tfrac{4}{21}, \quad P(5) = \tfrac{5}{21}, \quad P(6) = \tfrac{2}{7}$$

(ii) $P(A) = P(\{2, 4, 6\}) = \tfrac{4}{7}$, $\quad P(B) = P(\{2, 3, 5\}) = \tfrac{10}{21}$, $\quad P(C) = P(\{1, 3, 5\}) = \tfrac{3}{7}$.

(iii) (*a*) The event that an even or prime number occurs is $A \cup B = \{2, 4, 6, 3, 5\}$, or that 1 does not occur. Thus $P(A \cup B) = 1 - P(1) = \tfrac{20}{21}$.

(*b*) The event that an odd prime number occurs is $B \cap C = \{3, 5\}$. Thus $P(B \cap C) = P(\{3, 5\}) = \tfrac{8}{21}$.

INFINITE SAMPLE SPACES

Now suppose S is a countably infinite sample space; say $S = \{a_1, a_2, \ldots\}$. As in the finite case, we obtain a probability space by assigning to each $a_i \in S$ a real number p_i, called its probability, such that

$$\text{(i)}\ \ p_i \geqq 0 \quad \text{and} \quad \text{(ii)}\ \ p_1 + p_2 + \cdots = \sum_{i=1}^{\infty} p_i = 1$$

The probability $P(A)$ of any event A is then the sum of the probabilities of its points.

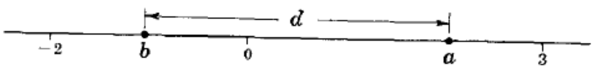
Example 10: Consider the sample space $S = \{1, 2, 3, \ldots, \infty\}$ of the experiment of tossing a coin till a head appears; here n denotes the number of times the coin is tossed. A probability space is obtained by setting

$$p(1) = \tfrac{1}{2}, \quad p(2) = \tfrac{1}{4}, \quad \ldots, \quad p(n) = 1/2^n, \quad \ldots, \quad p(\infty) = 0$$

The only uncountable sample spaces S which we will consider here are those with some finite geometrical measurement $m(S)$ such as length, area or volume, and in which a point is selected at random. The probability of an event A, i.e. that the selected point belongs to A, is then the ratio of $m(A)$ to $m(S)$; that is,

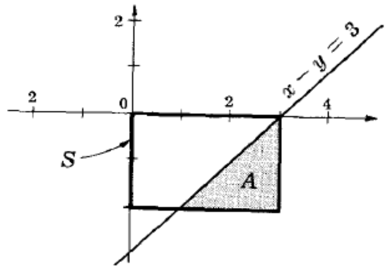
$$P(A) = \frac{\text{length of } A}{\text{length of } S} \quad \text{or} \quad P(A) = \frac{\text{area of } A}{\text{area of } S} \quad \text{or} \quad P(A) = \frac{\text{volume of } A}{\text{volume of } S}$$

Example 11: On the real line $\mathbf{R}$, points a and b are selected at random such that $-2 \leqq b \leqq 0$ and $0 \leqq a \leqq 3$, as shown below. Find the probability p that the distance d between a and b is greater than 3.



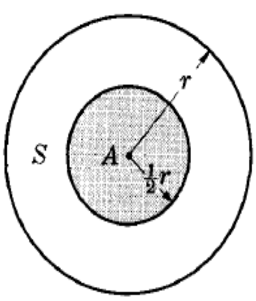
The sample space S consists of the ordered pairs (a, b) and so forms the rectangular region shown in the adjacent diagram. On the other hand, the set A of points (a, b) for which $d = a - b > 3$ consists of those points of S which lie below the line $x - y = 3$, and hence forms the shaded area in the diagram. Thus

$$p = P(A) = \frac{\text{area of } A}{\text{area of } S} = \frac{2}{6} = \frac{1}{3}$$



example 12

A point is selected at random inside a circle. Find the probability p that the point is closer to the center of the circle than to its circumference.



Let S denote the set of points inside the circle with radius r, and let A denote the set of points inside the concentric circle of radius $\frac{1}{2}r$. (Thus A consists precisely of those points of S which are closer to its center than to its circumference.) Accordingly,

$$p = P(A) = \frac{\text{area of } A}{\text{area of } S} = \frac{\pi(\frac{1}{2}r)^2}{\pi r^2} = \frac{1}{4}$$