CHAPTER 1

THE MEANING OF PROBABILITY

1-1 INTRODUCTION

The theory of probability deals with averages of mass phenomena occurring sequentially or simultaneously: electron emission, telephone calls, radar detection, quality control, system failure, games of chance, statistical mechanics, turbulence, noise, birth and death rates, and queueing theory, among many others.

It has been observed that in these and other fields certain averages approach a constant value as the number of observations increases and this value remains the same if the averages are evaluated over any subsequence specified before the experiment is performed. In the coin experiment, for example, the percentage of heads approaches 0.5 or some other constant, and the same average is obtained if we consider every fourth, say, toss (no betting system can beat the roulette).

The purpose of the theory is to describe and predict such averages in terms of probabilities of events. The probability of an event A is a number $P(A)$ assigned to this event. This number could be interpreted as:

> If the experiment is performed n times and the event A occurs n_A times, then, with a high degree of certaints, the relative frequency n_A/n of the occurrence of A is close to $P(A)$:
>
> $$P(A) \simeq n_A/n \qquad \text{...(1.1)}$$
>
> provided that n is sufficiently large

This interpretation is imprecise: The terms "with a high degree of certainty," "close," and "sufficiently large" have no clear meaning. However, this lack of precision cannot be avoided. If we attempt to define in probabilistic terms the "high degree of certainty" we shall only postpone the inevitable conclusion that probability, like any physical theory, is related to physical phenomena only in inexact terms. Nevertheless, the theory is an exact discipline developed logically from clearly defined axioms, and when it is applied to real problems, it *works*.

OBSERVATION, DEDUCTION, PREDICTION. In the applications of probability to real problems, these steps must be clearly distinguished:

Step 1 (physical) We determine by an inexact process the probabilities $P(A_i)$ of certain events A_i.

This process could be based on the relationship (1-1) between probability and observation: The probabilistic data $P(A_i)$ equal the observed ratios n_{A_i}/n. It could also be based on "reasoning" making use of certain symmetries: If, out of a total of N outcomes, there are N_A outcomes favorable to the event A, then $P(A) = N_A/N$.

For example, if a loaded die is rolled 1000 times and five shows 200 times, then the probability of *five* equals 0.2. If the die is fair, then, because of its symmetry, the probability of *five* equals $1/6$.

Step 2 (conceptual) We assume that probabilities satisfy certain axioms, and by deductive reasoning we determine from the probabilities $P(A_i)$ of certain events A_i the probabilities $P(B_j)$ of other events B_j.

For example, in the game with a fair die we deduce that the probability of the event *even* equals $3/6$. Our reasoning is of the form:

$$\text{If P (1)} = \ldots = \text{P (6)} = \frac{1}{6} \text{ then P (even)} = \frac{3}{6}$$

Step 3 (physical) We make a physical prediction based on the numbers $P(B_j)$ so obtained.

This step could rely on (1-1) applied in reverse: If we perform the experiment n times and an event B occurs n_B times, then $n_B \simeq nP(B)$.

If, for example, we roll a fair die 1000 times, our prediction is that even will show about 500 times.

We could not emphasize too strongly the need for separating these three steps in the solution of a problem. We must make a clear distinction between the data that are determined empirically and the results that are deduced logically.

Steps 1 and 3 are based on *inductive reasoning*. Suppose, for example, that we wish to determine the probability of heads of a given coin. Should we toss the coin 100 or 1000 times? If we toss it 1000 times and the average number of heads equals 0.48 , what kind of prediction can we make on the basis of this observation? Can we deduce that at the next 1000 tosses the number of heads will be about 480? Such questions can be answered only inductively.

In this book, we consider mainly step 2, that is, from certain probabilities we derive *deductively* other probabilities. One might argue that such derivations are mere tautologies because the results are contained in the assumptions. This is true in the same sense that the intricate equations of motion of a satellite are included in Newton's laws.

To conclude, we repeat that the probability $P(A)$ of an event A will be interpreted as a number assigned to this event as mass is assigned to a body or resistance to a resistor. In the development of the theory, we will not be concerned about the "physical meaning" of this number. This is what is done in circuit analysis, in electromagnetic theory, in classical mechanics, or in any other scientific discipline. These theories are, of course, of no value to physics unless they help us solve real problems. We must assign specific, if only approximate, resistances to real resistors and probabilities to real events (step 1); we must also give physical meaning to all conclusions that are derived from the theory (step 3). But this link between concepts and observation must be separated from the purely logical structure of each theory (step 2).

As an illustration, we discuss in Example 1-1 the interpretation of the meaning of resistance in circuit theory.

EXAMPLE 1-1 ▶ A resistor is commonly viewed as a two-terminal device whose voltage is proportional to the current

$$R = \frac{v(t)}{i(t)} \qquad ...(1.2)$$

This, however, is only a convenient abstraction. A real resistor is a complex device with distributed inductance and capacitance having no clearly specified terminals. A relationship of the form (1-2) can, therefore, be claimed only within certain errors, in certain frequency ranges, and with a variety of other qualifications. Nevertheless, in the development of circuit theory we ignore all these uncertainties. We assume that the resistance R is a precise number satisfying (1-2) and we develop a theory based on (1-2) and on Kirchhoff's laws. It would not be wise, we all agree, if at each stage of the development of the theory we were concerned with the true meaning of R. ◀

1-2 THE DEFINITIONS

In this section, we discuss various definitions of probability and their roles in our investigation.

Axiomatic Definition

We shall use the following concepts from set theory (for details see Chap. 2): The certain event S is the event that occurs in every trial. The union $A \cup B \equiv A + B$ of two events A and B is the event that occurs when A or B or both occur. The intersection $A \cap B \equiv AB$ of the events A and B is the event that occurs when both events A and B occur. The events A and B are *mutually exclusive* if the occurrence of one of them excludes the occurrence of the other.

We shall illustrate with the die experiment: The certain event is the event that occurs whenever any one of the six faces shows. The union of the events even and less than 3 is the event *1 or 2 or 4 or 6* and their intersection is the event 2 . The events even and *odd* are mutually exclusive.

The axiomatic approach to probability is based on the following three postulates and on nothing else: The probability $P(A)$ of an event A is a non-negative number assigned to this event:

$$P(A) \geq 0 \qquad ...(1.3)$$

The probability of the certain event equals 1:

$$P(S) = 1 \qquad ...(1.4)$$

If the events A and B are mutually exclusive, then

$$P(A \cup B) = P(A) + P(B) \quad ...(1.5)$$

This approach to probability is relatively recent (A.N. Kolmogorov,[1] 1933). However, in our view, it is the best way to introduce a probability even in elementary courses. It emphasizes the deductive character of the theory, it avoids conceptual ambiguities, it provides a solid preparation for sophisticated applications, and it offers at least a beginning for a deeper study of this important subject.

The axiomatic development of probability might appear overly mathematical. However, as we hope to show, this is not so. The elements of the theory can be adequately explained with basic calculus.

Relative Frequency Definition

The relative frequency approach is based on the following definition: The probability $P(A)$ of an event A is the limit

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n} \quad ...(1.6)$$

where n_A is the number of occurrences of A and n is the number of trials. This definition appears reasonable. Since probabilities are used to describe relative frequencies, it is natural to define them as limits of such frequencies. The problem associated with a priori definitions are eliminated, one might think, and the theory is founded on observation.

However, although the relative frequency concept is fundamental in the applications of probability (steps 1 and 3), its use as the basis of a deductive theory (step 2) must be challenged. Indeed, in a physical experiment, the numbers n_A and n might be large but they are only finite; their ratio cannot, therefore, be equated, even approximately, to a limit If (1-6) is used to define $P(A)$, the limit must be accepted as a hypothesis, not as a number that can be determined experimentally.

Early in the century, Von Mises[2] used (1-6) as the foundation for a new theory. At that time, the prevailing point of view was still the classical, and his work offered a welcome alternative to the a priori concept of probability, challenging its metaphysical implications and demonstrating that it leads to useful conclusions mainly because it makes implicit use of relative frequencies based on our collective experience. The use of (1-6) as the basis for deductive theory has not, however, enjoyed wide acceptance even though (1-6) relates $P(A)$ to observed frequencies. It has generally been recognized that the axiomatic approach (Kolmogorov) is superior.

We shall venture a comparison between the two approaches using as illustration the definition of the resistance R of an ideal resistor. We can define R as a limit

$$R = \lim_{n \to \infty} \frac{e(t)}{i_n(t)}$$

[1]A.N. Kolmogrov: Grundbegriffe der Wahrscheinlichkeits Rechnung, *Ergeb. Math und ihrer Grensg. Vol. 2 1933.*

[2]Richard Von Mises: *Probability, Statistics and Truth*, English edition, H. Geiringer, ed., G. Allen and Unwin Ltd., London, 1957.

where $e(t)$ is a voltage source and $i_n(t)$ are the currents of a sequence of real resistors that tend in some sense to an ideal two-terminal element. This definition might show the relationship between real resistors and ideal elements but the resulting theory is complicated. An axiomatic definition of R based on Kirchhoff's laws is, of course, preferable.

Classical Definition

For several centuries, the theory of probability was based on the classical definition. This concept is used today to determine probabilistic data and as a working hypothesis. In the following, we explain its significance.

According to the classical definition, the probability $P(A)$ of an event A is determined a priori without actual experimentation: It is given by the ratio

$$P(A) = \frac{N_A}{N} \qquad ...(1.7)$$

where N is the number of *possible* outcomes and N_A is the number of outcomes that are *favorable* to the event A.

In the die experiment, the possible outcomes are six and the outcomes favorable to the event *even* are three; hence $P(\mathit{even}) = 3/6$.

It is important to note, however, that the significance of the numbers N and N_A is not always clear. We shall demonstrate the underlying ambiguities with Example 1-2.

EXAMPLE 1-2

▶ We roll two dice and we want to find the probability p that the sum of the numbers that show equals 7.

To solve this problem using (1-7), we must determine the numbers N and N_A. *(a)* We could consider as possible outcomes the 11 sums $2,3,\ldots,12$. Of these, only one, namely the sum 7, is favorable; hence $p = 1/11$. This result is of course wrong. *(b)* We could count as possible outcomes all pairs of numbers not distinguishing between the first and the second die. We have now 21 outcomes of which the pairs $(3,4),(5,2)$, and $(6,1)$ are favorable. In this case, $N_A = 3$ and $N = 21$; hence $p = 3/21$. This result is also wrong. *(c)* We now reason that the above solutions are wrong because the outcomes in *(a)* and *(b)* are not *equally likely*. To solve the problem "correctly," we must count all pairs of numbers distinguishing between the first and the second die. The total number of outcomes is now 36 and the favorable outcomes are the six pairs $(3,4),(4,3),(5,2)$, $(2,5),(6,1)$, and $(1,6)$; hence $p = 6/36$. ◀

Example 1-2 shows the need for refining definition (1-7). The improved version reads as follows:

> The probability of an event equals the ratio of its favorable outcomes to the total number of outcomes provided that all outcomes are *equally likely*.

As we shall presently see, this refinement does not eliminate the problems associated with the classical definition.

Notes: 1. The classical definition was introduced as a consequence of the *principle of insufficient reason*[3] : "In the absence of any prior knowledge, we *must* assume that the events A_i have equal probabilities." This conclusion is based on the subjective interpretation of probability as a measure of *our state of knowledge* about the events A_i. Indeed, if it were not true that the events A_i have the same probability, then changing their indices we would obtain different probabilities without a change in the state of our knowledge.

2. As we explain in Chap. 14, the principle of insufficient reason is equivalent to the *principle of maximum entropy.*

CRITIQUE. The classical definition can be questioned on several grounds.

A. The term *equally likely* used in the improved version of (1-7) means, actually, *equally probable*. Thus, in the definition, use is made of the concept to be defined. As we have seen in Example 1-2 this often leads to difficulties in determining N and N_A.

B. The definition can be applied only to a limited class of problems. In the die experiment, for example, it is applicable only if the six faces have the same probability. If the die is loaded and the probability of *four equals* 0.2, say, the number 0.2 cannot be derived from (1-7).

C. It appears from (1-7) that the classical definition is a consequence of logical imperatives divorced from experience. This, however, is not so. We accept certain alternatives as equally likely because of our collective experience. The probabilities of the outcomes of a fair die equal $1/6$ not only because the die is symmetrical but also because it was observed in the long history of rolling dice that the ratio n_A/n in (1-1) is close to $1/6$. The next illustration is, perhaps, more convincing:

We wish to determine the probability p that a newborn baby is a boy. It is generally assumed that $p = 1/2$; however, this is not the result of pure reasoning. In the first place, it is only approximately true that $p = 1/2$. Furthermore, without access to long records we would not know that the boy-girl alternatives are equally likely regardless of the sex history of the baby's family, the season or place of its birth, or other conceivable factors. It is only after long accumulation of records that such factors become irrelevant and the two alternatives are accepted as equally likely.

D. If the number of possible outcomes is infinite, then to apply the classical definition we must use length, area, or some other measure of infinity for determining the ratio N_A/N in (1-7). We illustrate the resulting difficulties with the following example known as the *Bertrand paradox*.

EXAMPLE 1-3

BERTRAND PARADOX

▶ We are given a circle C of radius r and we wish to determine the probability p that the length l of a "randomly selected" cord AB is greater than the length $r\sqrt{3}$ of the inscribed equilateral triangle.

[3] H. Bernoulli. *Ars Conjectandi*, 1713.

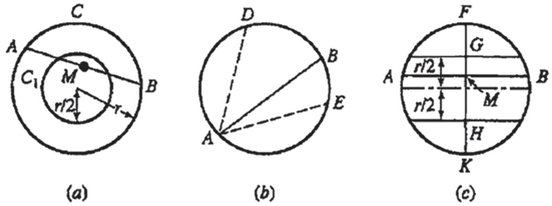

Figure 1.1

We shall show that this problem can be given at least three reasonable solutions.

I. If the center M of the cord AB lies inside the circle C_1 of radius $r/2$ shown in Fig. 1-1***a***, then $l > r\sqrt{3}$. It is reasonable, therefore, to consider as favorable outcomes all points inside the circle C_1 and as possible outcomes all points inside the circle C. Using as measure of their numbers the corresponding areas $\pi r^2/4$ and πr^2, we conclude that

$$p = \frac{\pi r^2/4}{\pi r^2} = \frac{1}{4}$$

II. We now assume that the end A of the cord AB is fixed. This reduces the number of possibilities but it has no effect on the value of p because the number of favorable locations of B is reduced proportionately. If B is on the 120° arc DBE of Fig. 1-1b, then $l > r\sqrt{3}$. The favorable outcomes are now the points on this arc and the total outcomes all points on the circumference of the circle C. Using as their measurements the corresponding lengths $2\pi r/3$ and $2\pi r$, we obtain

$$p = \frac{2\pi r/3}{2\pi r} = \frac{1}{3}$$

III. We assume finally that the direction of AB is perpendicular to the line FK of Fig. 1-1c. As in II this restriction has no effect on the value of p. If the center M of AB is between G and H, then $l > r\sqrt{3}$. Favorable outcomes are now the points on GH and possible outcomes all points on FK. Using as their measures the respective lengths r and $2r$, we obtain

$$p = \frac{r}{2r} = \frac{1}{2}$$

◀

We have thus found not one but three different solutions for the same problem! One might remark that these solutions correspond to three different experiments. This is true but not obvious and, in any case, it demonstrates the ambiguities associated with the classical definition, and the need for a clear specification of the outcomes of an experiment and the meaning of the terms "possible" and "favorable."

VALIDITY. We shall now discuss the value of the classical definition in the determination of probabilistic data and as a working hypothesis.

A. In many applications, the assumption that there are N equally likely alternatives is well established through long experience. Equation (1-7) is then accepted as self-evident. For example, "If a ball is selected at random from a box containing *m* black and *n* white balls, the probability that it is white equals $n/(m + n)$," or, "If a call occurs at random in the time interval (0, T), the probability that it occurs in the interval (t_1, t_2) equals (t_2, t_1) /T." Such conclusions are of course, valid and useful; however, their validity rests on the meaning of the word *random*. The conclusion of the last example that "the unknown probability equals $(t_2 - t_1)$ /T" is not a consequence of the "randomness" of the call. The two statements are merely equivalent and they follow not from a priori reasoning but from past records of telephone calls.

B. In a number of applications it is impossible to determine the probabilities of various events by repeating the underlying experiment a sufficient number of times. In such cases, we have no choice but to *assume* that certain alternatives are equally likely and to determine the desired probabilities from (1-7). This means that we use the classical definition as a *working* hypothesis. The hypothesis is accepted if its observable consequences agree with experience, otherwise it is rejected. We illustrate with an important example from statistical mechanics.

EXAMPLE 1-4

▶ Given n particles and $m > n$ boxes, we place at random each particle in one of the boxes. We wish to find the probability p that in n preselected boxes, one and only one particle will be found.

Since we are interested only in the underlying assumptions, we shall only state the results (the proof is assigned as Prob. 4-34). We also verify the solution for $n = 2$ and $m = 6$. For this special case, the problem can be stated in terms of a pair of dice: The $m = 6$ faces correspond to the m boxes and the $n = 2$ dice to the n particles. We assume that the preselected faces (boxes) are 3 and 4.

The solution to this problem depends on the choice of possible and favorable outcomes We shall consider these three celebrated cases:

MAXWELL-BOLTZMANN STATISTICS

If we accept as outcomes all possible ways of placing n particles in m boxes distinguishing the identity of each particle, then

$$p = \frac{n!}{m^n}$$

For $n = 2$ and $m = 6$ this yields $p = 2/36$. This is the probability for getting 3, 4 in the game of two dice.

BOSE-EINSTEIN STATISTICS

If we assume that the particles are not distinguishable, that is, if all their permutations count as one, then

$$p = \frac{(m-1)!n!}{(n+m-1)!}$$

For $n = 2$ and $m = 6$ this yields $p = 1/21$. Indeed, if we do not distinguish between the two dice, then $N = 21$ and $N_A = 1$ because the outcomes 3,4 and 4, 3 are counted as one.

FERMI-DIRAC STATISTICS

If we do not distinguish between the particles and also we assume that in each box we are allowed to place at most one particle, then

$$p = \frac{n!(m-n)!}{m!}$$

For $n = 2$ and $m = 6$ we obtain $p = 1/15$. This is the probability for 3,4 if we do not distinguish between the dice and also we ignore the outcomes in which the two numbers that show are equal.

One might argue, as indeed it was in the early years of statistical mechanics, that only the first of these solutions is logical. The fact is that in the absence of direct or indirect experimental evidence this argument cannot be supported. The three models proposed are actually only *hypotheses* and the physicist accepts the one whose consequences agree with experience. ◀

C. Suppose that we know the probability $P(A)$ of an event A in experiment 1 and the probability $P(B)$ of an event B in experiment 2 . In general, from this information we cannot determine the probability $P(AB)$ that both events A and B will occur. However, if we know that the two experiments are *independent*, then

$$P(AB) = P(A)P(B) \qquad ...(1.8)$$

In many cases, this independence can be established a priori by reasoning that the outcomes of experiment 1 have no effect on the outcomes of experiment 2. For example, if in the coin experiment the probability of *heads* equals $1/2$ and in the die experiment the probability of even equals $1/2$, then, we conclude "logically," that if both experiments are performed, the probability that we get *heads* on the coin and *even* on the die equals $1/2 \times 1/2$. Thus, as in (1-7), we accept the validity of (1-8) as a logical necessity without recourse to (1-1) or to any other direct evidence.

D. The classical definition can be used as the basis of a deductive theory if we accept (1-7) as an *assumption*. In this theory, no other assumptions are used and postulates (1-3) to (1-5) become theorems. Indeed, the first two postulates are obvious and the third follows from (1-7) because, if the events A and B are mutually exclusive, then $N_{A+B} = N_A + N_B$; hence

$$P(A \cup B) = \frac{N_{A+B}}{N} = \frac{N_A}{N} + \frac{N_B}{N} = P(A) + P(B)$$

As we show in (2-25), however, this is only a very special case of the axiomatic approach to probability.

1-3 PROBABILITY AND INDUCTION

In the applications of the theory of probability we are faced with the following question: Suppose that we know somehow from past observations the probability $P(A)$ of an event A in a given experiment. What conclusion can we draw about the occurrence of this event in a *single* future performance of this experiment? (See also Sec. 8-1.)

We shall answer this question in two ways depending on the size of $P(A)$: We shall give one kind of an answer if $P(A)$ is a number distinctly different from 0 or 1 , for example 0.6 , and a different kind of an answer if $P(A)$ is close to 0 or 1 , for example 0.999. Although the boundary between these two cases is not sharply defined, the corresponding answers are fundamentally different.

Case 1 Suppose that $P(A)=0.6$. In this case, the number 0.6 gives us only a "certain degree of confidence that the event A will occur." The known probability is thus used as a "measure of our belief" about the occurrence of A in a single trial. This interpretation of $P(A)$ is subjective in the sense that it cannot be verified experimentally. In a single trial, the event A will either occur or will not occur. If it does not, this will not be a reason for questioning the validity of the assumption that $P(A)=0.6$.

Case 2 Suppose, however that $P(A)=0.999$. We can now state with practical certainty that at the next trial the event A will occur. This conclusion is objective in the sense that it can be verified experimentally. At the next trial the event A must occur. If it does not, we must seriously doubt, if not outright reject, the assumption that $P(A)=0.999$.

The boundary between these two cases, arbitrary though it is (0.9 or 0.999999 ?), establishes in a sense the line separating "soft" from "hard" scientific conclusions. The theory of probability gives us the analytic tools (step 2) for transforming the "subjective" statements of case 1 to the "objective" statements of case 2. In the following, we explain briefly the underlying reasoning.

As we show in Chap. 3, the information that $P(A)=0.6$ leads to the conclusion that if the experiment is performed 1000 times, then "almost certainly" the number of times the event A will occur is between 550 and 650 . This is shown by considering the repetition of the original experiment 1000 times as a *single* outcome of a new experiment. In this experiment the probability of the event

$$A_1=\{\text{ the number of times } A \text{ occurs is between} 550 \text{ and} 650\}$$

equals 0.999 (see Prob. 4-25). We must, therefore, conclude that (case 2) the event A_1 will occur with practical certainty.

We have thus succeeded, using the theory of probability, to transform the "subjective" conclusion about A based on the *given* information that $P(A)=0.6$, to the "objective" conclusion about A_1 based on the *derived* conclusion that $P(A_1)=0.999$. We should emphasize, however, that both conclusions rely on inductive reasoning. Their difference, although significant, is only quantitative. As in case 1, the "objective" conclusion of case 2 is not a certainty but only an inference. This, however, should not surprise us; after all, no prediction about future events based on past experience can be accepted as logical certainty.

Our inability to make categorical statements about future events is not limited to probability but applies to all sciences. Consider, for example, the development of classical mechanics. It was *observed* that bodies fall according to certain patterns, and on this evidence Newton formulated the laws of mechanics and used them to *predict* future events. His predictions, however, are not logical certainties but only plausible inferences. To "prove" that the future will evolve in the predicted manner we must invoke metaphysical causes.

1-4 CAUSALITY VERSUS RANDOMNESS

We conclude with a brief comment on the apparent controversy between causality and randomness. There is no conflict between causality and randomness or between determinism and probability if we agree, as we must, that scientific theories are not *discoveries* of the laws of nature but rather *inventions* of the human mind. Their consequences are presented in deterministic form if we examine the results of a single trial; they are presented as probabilistic statements if we are interested in averages of many trials. In both cases, all statements are qualified. In the first case, the uncertainties are of the form "with certain errors and in certain ranges of the relevant parameters"; in the second, "with a high degree of certainty if the number of trials is large enough." In the next example, we illustrate these two approaches.

EXAMPLE 1-5

▶ A rocket leaves the ground with an initial velocity v forming an angle θ with the horizontal axis (Fig. 1-2). We shall determine the distance $d = OB$ from the origin to the reentry point B.

From Newton's law it follows that

$$d = \frac{v^2}{g} sin 2\theta \qquad ...(1.9)$$

This seems to be an unqualified consequence of a causal law; however, this is not so. The result is approximate and it can be given a probabilistic interpretation.

Indeed, (1-9) is not the solution of a real problem but of an idealized model in which we have neglected air friction, air pressure, variation of g, and other uncertainties in the values of v and θ. We must, therefore, accept (1-9) only with qualifications. It holds within an error ε provided that the neglected factors are smaller than δ.

Suppose now that the reentry area consists of numbered holes and we want to find the reentry hole. Because of the uncertainties in v and θ, we are in no position to give a deterministic answer to our problem. We can, however, ask a different question: If many rockets, nominally with the same velocity, are launched, what percentage will enter the nth hole? This question no longer has a causal answer; it can only be given a random interpretation.

Thus the same physical problem can be subjected either to a deterministic or to a probabilistic analysis. One might argue that the problem is inherently deterministic because the rocket has a precise velocity even if we do not know it. If we did, we would know exactly the reentry hole. Probabilistic interpretations are, therefore, necessary because of our ignorance.

Such arguments can be answered with the statement that the physicists are not concerned with what is true but only with what they can observe. ◀

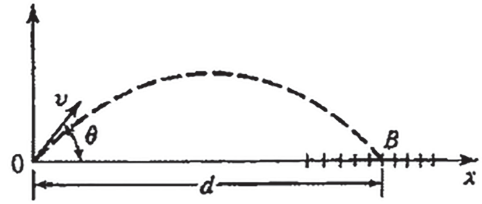

Figure 1.2

Historical Perspective

Probability theory has its humble origin in problems related to gambling and games of chance. The origin of the theory of probability goes back to the middle of the 17th century and is connected with the works of Pierre de Fermat (1601-1665), Blaise Pascal (1623-1662), and Christian Huygens (1629-1695). In their works, the concepts of the probability of a stochastic event and the expected or mean value of a random variable can be found. Although their investigations were concerned with problems connected with games of chance, the importance of these new concepts was clear to them, as Huygens points out in the first printed probability text[4] (1657) *On Calculations in Games of Chance*: "The reader will note that we are dealing not only with games, but also that the foundations of a very interesting and profound theory are being laid here." Later, Jacob Bernoulli (1654-1705), Abraham De Moivre (1667-1754), Rev. Thomas Bayes (1702-1761), Marquis Pierre Simon Laplace (1749-1827), Johann Friedrich Carl Gauss (1777-1855), and Siméon Denis Poisson (1781-1840) contributed significantly to the development of probability theory. The notable contributors from the Russian school include P.L. Chebyshev (1821-1894), and his students A. Markov (1856-1922) and A.M. Lyapunov (1857-1918) with important works dealing with the law of large numbers.

The deductive theory based on the axiomatic definition of probability that is popular today is mainly attributed to Andrei Nikolaevich Kolmogorov, who in the 1930s along with Paul Levy found a close connection between the theory of probability and the mathematical theory of sets and functions of a real variable. Although Emile Borel had arrived at these ideas earlier, putting probability theory on this modern frame work is mainly due to the early 20 th century mathematicians.

Concluding Remarks

In this book, we present a deductive theory (step 2) based on the axiomatic definition of probability. Occasionally, we use the classical definition but only to determine probabilistic data (step 1).

To show the link between theory and applications (step 3), we give also a relative frequency interpretation of the important results. This part of the book, written in small print under the title *Frequency interpretation*, does not obey the rules of deductive reasoning on which the theory is based.

---

[4]Although the ecentric scholar (and gambler) Girolamo Cardano (1501-1576) had written. *The Book of Games and Chance* around 1520, it was not published until 1663. Cardano had left behind 131 printed works and 111 additional manuscripts.