$1.75

GOLDEN GATE BOOK

AN ELEMENTARY INTRODUCTION TO THE THEORY OF PROBABILITY

B. V. GNEDENKO

A. YA. KHINCHIN

Edited by J. B. Roberts

GNEDENKO KHINCHIN An Elementary Introduction to the THEORY OF PROBABILITY

B. V. Gnedenko and A. Ya. Khinchin, professors in the Institute of Mathematics at the University of Moscow, are two of Russia's best known mathematicians. Both men are acknowledged authorities on probability theory and are prolific writers.

This introduction to probability theory and its applications differs from most texts in that it uses basic probability concepts to solve commonplace problems that would normally be handled by statistical methods. Since little attention is given to the norma! distribution curve, to methods of hypothesis testing, or to decision theory, the suggested analytical methods relate directly to the basic concepts accepted by the Russian probability school. Numerous examples deal primarily with industrial and military problems.

Professors Gnedenko and Khinchin do not presuppose for the reader any advanced mathematics or prior knowledge of probability. Hence, their book is particularly suitable for high school supplementary reading, for beginning college courses in probability, and for summer mathematics institutes.

The translator of this book, W. R. Stahl, is a radiobiologist affiliated with the School of Science at Oregon State College. Its editor, J. B. Roberts, is an associate professor of mathematics at Reed College.

W. H. FREEMAN AND COMPANY
SAN FRANCISCO AND LONDON

FREEMAN

An Elementary Introduction to the Theory of Probability

A Series of Undergraduate Books in Mathematics

R. A. Rosenbaum, EDITOR

Introduction to Matrices and Linear Transformations
Daniel T. Finkbeiner, II

Introduction to Probability and Statistics
Henry L. Alder and Edward B. Roessler

GOLDEN GATE BOOKS

A Concrete Approach to Abstract Algebra
W. S. Sawyer

A Modern View of Geometry
Leonard M. Blumenthal

Sets, Logic, and Axiomatic Theories
Robert R. Stoll

The Solution of Equations in Integers
A. O. Gelfond

An Elementary Introduction to the Theory of Probability
B. V. Gnedenko and A. Ya. Khinchin

An Elementary Introduction to the Theory of Probability

B. V. GNEDENKO, *University of Moscow*

A. YA. KHINCHIN, *University of Moscow*

Translated by W. R. Stahl

Edited by J. B. Roberts

W. H. FREEMAN AND COMPANY
SAN FRANCISCO AND LONDON

Printed in the United States of America.
Library of Congress Catalogue Card Number: 61-10868.

CONTENTS

PREFACE TO THE FIRST EDITION

Knowledge of the theoretical basis of one or another mathematical science always makes possible a more intelligent and fruitful practical application of the conclusions of such a science. The theory of probability is of such scope that it is used by a large number of military commanders (and occasionally by line personnel) and by officials in industry, agriculture, economics, and so on, whose mathematical training is very limited.

It is the goal of our book to acquaint such persons with the basic concepts of the theory of probability and methods of probability computation in the most useful possible way. The book is fully readable by anyone who has completed ten-year intermediate schools; it is almost fully accessible to those having completed only seven-year schools. Practical examples are given throughout the book. However, our choice of examples was governed primarily not by their practical importance, but rather by their value as illustrations of the corresponding theoretical material.

Moscow, 7 January, 1945

PREFACE TO THE SECOND EDITION

This edition differs from the first only in minor details. We have added a "conclusions" section and made a small number of clarifying remarks in the text.

Kiev–Moscow, 7 November, 1949

PREFACE TO THE THIRD EDITION

The third edition differs hardly at all from previous ones. Only a few unimportant corrections were added.

7 April, 1952

PREFACE TO THE FOURTH EDITION

The fourth edition differs hardly at all from previous ones. Only a few unimportant corrections were added.

28 January, 1957

Chapter 1

The Probability of Events

§1. The Concept of Probability

When we say that a certain marksman hits the target 92% of the time we mean that of 100 firings carried out under certain conditions (the same target, the same distance, the same rifle, and so on), on the *average*, about 92 will hit the target (and 8 will miss it). Naturally, there need not be 92 hits in every 100 tries; sometimes there will be 91 or 90, other times 93 or 94. Sometimes the number may even be substantially more or less than 92, but, on the *average*, with many repeated firing trials under unchanged conditions, the percentage of hits will be constant as long as there is no important alteration with the passage of time (for instance, the marksman might improve his capability, achieving a score of 95 or higher). Experiments show that this marksman will obtain scores near 92 in most trials; scores of less than 88 or more than 96 may occur, but only infrequently. The figure 92%, representing the mastery of the marksman, is usually very *stable*. That is, the percentage of hits in most trials (under the same conditions) will be practically identical for a given marksman, and only rarely will his score deviate significantly from the average value.

Let us consider another example. In a certain industrial plant it is noted that under given conditions an average of 1.6% of manufactured items do not meet the quality standards and are rejected. This means that in a lot of, say, 1000 unscreened items there will be about 16 rejects. Sometimes, of course, the number of rejects may be higher or lower, but, on the average, the number will be close to 16. Obviously, we assume unaltered manufacturing conditions (organization of the technological process, raw materials, experience of workers, and so on).

Any number of such examples can be found. In all such cases we observe that for *uniform*, *large-scale* operations (repeated firings, mass production, and so on) the percentage of some occurrence (hitting the target, failure to meet manufacturing standards, and so on) remains practically the same under given conditions, and only rarely does it deviate significantly from the average. Therefore, we can state that this average value is characteristic of the repetitious operation (under given, strictly established conditions). The percentage of hits evaluates the mastery of the marksman; the percentage of rejects evaluates the quality of production. It is apparent that such averages are very important in the most varied domains: military problems, technology, economics, physics, chemistry. They allow us not only to evaluate many events which have already occurred, but also to predict the outcome of some large-scale trial in the future.

If under certain conditions a marksman hits the target 92 times out of 100 shots, we say that for this marksman under these conditions the *probability of a hit* is 92% (or 92/100, or 0.92). If under certain conditions there are 16 rejects out of every 1000 finished items produced by a particular plant, we say that the *probability of producing a reject* is 1.6%, or 0.016, for that manufacturing process.

What do we mean when we refer to the probability of an event in a certain repetitious series of occurrences? It is not difficult to answer this question. Such a mass trial always involves the frequent repetition of single similar operations (in marksmanship—of individual firings; in mass production—of the manufacture of particular items). We are concerned with a particular result of a single operation (accuracy of a single shot, nonuniformity of a specific manufactured item), but most of all with the number of such results in some mass trial or operation (how many shots hit the target, how many items will be rejected). The percentage (or portion) of such "successful"* events in a certain mass trial is precisely what we call the *probability* of the event under consideration. Always keep in mind that the probability of some event (outcome) has meaning only under strictly defined conditions which apply to the mass trial. Any basic alterations in these conditions will usually change the probability of interest to us.

If a mass operation is such that event A (for instance, hitting the target) is observed an average of a times among b single trials (firings), the probability for event A under the given conditions is a/b, or $(100a/b)\%$. Therefore, we can say that the *probability* of a successful outcome for a single trial equals the *proportion of such successful outcomes to the number of all single trials* included in the particular mass trial or operation. It is apparent that if the probability of some event is a/b, then in any series composed of b trials the event may occur more or less often than a times; it will occur a times only on the average. But in most such series of b operations the number of times that A occurs will be close to a, *especially if b is a large number.*

Example 1. In a certain city during the first quarter of the year there were born

in January: 145 boys and 135 girls;
in February: 142 boys and 136 girls;
in March: 152 boys and 140 girls.

*For the second illustration it would be more appropriate to say "unsuccessful." However, in the theory of probability it is customary to call those results "successful" which involve the occurrence of the event which interests us in the given problem.

How large is the probability of the birth of a boy?

$$\text{In January: } \frac{145}{280} \approx 0.518 \approx 51.8\%.$$

$$\text{In February: } \frac{142}{278} \approx 0.511 \approx 51.1\%.$$

$$\text{In March: } \frac{152}{292} \approx 0.520 \approx 52.0\%.$$

We observe that the average arithmetic ratio for individual months is close to 0.516; therefore, the desired probability under the given circumstances is about 0.516, or 51.6%. This value is well known in demography—the science which studies population dynamics. Under normal conditions the proportion of male births will not differ greatly from this value.

Example 2. In the beginning of the nineteenth century Robert Brown, the Scottish botanist, demonstrated a remarkable phenomenon now known as Brownian motion. In effect, microscopic particles suspended in a liquid (that is, in a state of passive equilibrium) exhibit a vibratory motion, with no apparent cause.

For a long time the reason for this seemingly self-induced motion was unknown. Eventually the kinetic theory of gases offered a simple and complete explanation. The motion of particles suspended in a liquid is due to collisions between the particles and molecules of the liquid. The kinetic theory of gases makes it possible to compute the probability that a given minute volume of liquid will not contain a single particle of suspended material or that one, two, three, or more particles will be present. A series of experiments was carried out to test this theory.

The following are the results of 518 observations made by the Swedish physicist Svedberg on microscopic particles of gold suspended in water. The volume of space under study contained no particles in 112 cases; one particle in 168 cases; two particles in 130 cases; three particles in 69 cases; four particles in 32 cases; five particles in five cases; six particles in one case; and seven particles in one case. The proportion of observations revealing a given number of particles is equal to:

0 particles: $\frac{112}{518} \approx 0.216$; 4 particles: $\frac{32}{518} \approx 0.062$;

1 particle: $\frac{168}{518} \approx 0.325$; 5 particles: $\frac{5}{518} \approx 0.010$;

2 particles: $\frac{130}{518} \approx 0.251$; 6 particles: $\frac{1}{518} \approx 0.002$;

3 particles: $\frac{69}{518} \approx 0.133$; 7 particles: $\frac{1}{518} \approx 0.002$.

The results of the observations corresponded very well with the theoretically predicted probabilities.

§2. Impossible and Certain Events

It is apparent that the probability of an event is always a positive number or zero. It cannot be larger than unity, because in the fraction which determines the probability the numerator may not be greater than the denominator (the number of successful trials cannot be greater than the number of all trials).

Let us agree that $P(A)$ will stand for the probability of event A. No matter what this event,

$$0 \leq P(A) \leq 1.$$

The greater $P(A)$, the more frequently will event A occur. For instance, the higher the probability of a marksman striking the target, the more often will he have successful trials and the greater his mastery. If the probability of an event is very low, the event occurs infrequently; if $P(A) = 0$, event A either never occurs or occurs so infrequently that it may be considered impossible. On the other hand, if $P(A)$ is close to unity, the numerator of the fraction expressing this probability is close to the denominator; that is, the vast majority of trials are successful. In this case, the event occurs a majority of the time; if $P(A) = 1$, event A occurs always or nearly always, and from the practical viewpoint it may be considered as "certain" (we can count on this event happening). If $P(A) = \frac{1}{2}$, the event A will occur in about 50% of all instances; this means that successful trials are observed about as often as unsuccessful trials. If $P(A) > \frac{1}{2}$, event A occurs more than half the time, and if $P(A) < \frac{1}{2}$, event A occurs less than half the time.

How small must a probability be for us to consider an event practically impossible? There is no general answer to this question, because everything depends on the importance of the event we are considering. Thus, 0.01 is not a large number. If we have a group of shells and 0.01 represents the probability a shell will not explode on hitting the ground, this means that about 1% of the firings will be without effect. This is tolerable. But if we have a parachute which fails to open with a probability of 0.01, the situation is naturally not acceptable, since it means that one out of every hundred jumps will lead to the loss of a man's life. These examples demonstrate that for each particular problem we must decide beforehand, on the basis of practical considerations, how low may be the probability of an event in order that we may disregard its possible occurrence without fear of the consequences.

§3. A Problem

One marksman produces a score of 80% hits, and another marksman (under the same conditions of firing) produces a score of 70%. What is the probability of the target being hit if both marksmen shoot at it simultaneously? A hit is counted even if one of two bullets strikes it.

First means of solution. Suppose that 100 dual firings take place. In about 80 firings the target will be struck by the first marksman; in about 20 firings he will miss the target. Now, since the second marksman strikes the target an average of

70 times out of 100, or 7 times out of 10, we may expect that in those 20 instances when the first marksman misses the target the second marksman will hit it about 14 times. Thus, considering the entire 100 firings, the target will be hit about $80 + 14 = 94$ times, and therefore the probability of hits for simultaneous firing by the two marksmen will be 94%, or 0.94.

Second means of solution. Let us again assume 100 dual firings. We have already noted that under these conditions the first marksman will miss about 20 times. Since the second marksman misses about 30 of 100 shots, or 3 out of 10, we may expect that in the 20 instances where the first marksman misses there will be about 6 in which the second marksman misses also. In each of these 6 instances the target will remain untouched, and in the remaining 94 at least one of the marksmen will be successful and the target will be struck. Again we conclude that with dual firing the target will be hit about 94 times out of 100. That is, the probability of a hit is 94%, or 0.94.

This problem is a very simple one. Nevertheless it leads to a very important conclusion: There may be cases in which it is very useful to find the probability of certain complex events by using the probabilities of simpler ones. Actually, such cases arise very frequently, not only in military applications, but in all types of research and practical work in which repetitious events occur. It would be most inconvenient to seek a special means of solution for each new problem encountered. Scientists always attempt to lay down general rules which will allow a mechanical or nearly mechanical solution of different problems that are similar to each other. In dealing with repeated events, the science which treats the formulation of such general rules is called the *theory of probability*. This book presents the first principles of that science. Since the theory of probability is a branch of mathematics, its method is one of accurate deduction, and among its tools are formulas, tables, and graphs.

Chapter 2

Rules for Adding Probabilities

§4. Derivation of the Rules for Adding Probabilities

The simplest and most important of the general rules used in calculating probabilities is that covering *the addition of probabilities*; this will now be examined.

When shooting at the target shown in Figure 1 from a given distance a marksman has a certain probability of hitting area 1, 2, 3, 4, 5, or 6. Suppose that for some particular marksman the probability of hitting area 1 is 0.24, and the probability of hitting area 2 is 0.17. As we have learned, this means that of 100 bullets fired by this marksman 24 bullets, on the average, will hit area 1, and 17 will hit area 2.

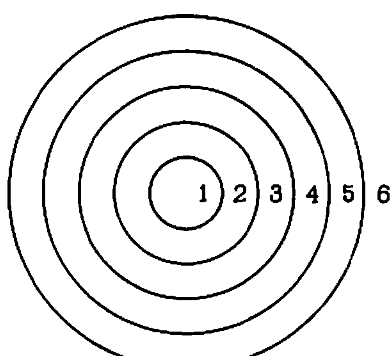

Figure 1

Suppose that in a certain shooting competition it is considered "excellent" if a bullet strikes area 1 and "good" if it strikes area 2. What is the probability that the marksman will score excellent or good?

This question is easy to answer. Out of every 100 firings by the marksman about 24 will hit area 1 and about 17 will hit area 2, which makes a total of $24 + 17 = 41$ bullets which will land in either area 1 or 2. Therefore, the needed probability is $0.41 = 0.24 + 0.17$. The probability that the marksman will score excellent or good is thus equal to the sum of the probabilities for excellent and good shots.

Let us consider another example. A passenger is waiting for trolley No. 26 or No. 16 at a trolley stop which handles trolleys of four designations: Nos. 16, 22, 26, and 31. Assuming that trolleys covering different routes appear equally often, on the average, let us deduce the probability that the first trolley arriving at the stop will be of the designation required by the passenger.

Clearly, the probability that that No. 16 will get there first is $\frac{1}{4}$; similarly, the probability that No. 26 will get there first is $\frac{1}{4}$. The needed probability is apparently $\frac{1}{2}$, since $\frac{1}{2} = \frac{1}{4} + \frac{1}{4}$, and therefore we may say that the probability that No. 16 or No. 26 will arrive first is equal to the sum of the probabilities that No. 16 or No. 26 will get there first.

We may now carry out a general deduction. For a certain mass trial it is known that in every series of b individual operations there will be found, on the average,

a_1 occurrences of a certain result A_1,
a_2 occurrences of a certain result A_2,
a_3 occurrences of a certain result A_3,

and so on. Stated differently,

the probability of event A_1 equals a_1/b;
the probability of event A_2 equals a_2/b;
the probability of event A_3 equals a_3/b;

and so on. What is the probability that in a certain individual trial one (which one is immaterial) of events A_1, A_2, A_3, $\cdots$ will occur?

The event under consideration may be called "A_1 or A_2 or A_3 or $\cdots$" (the sequence of dots here (as elsewhere) means "and so on"). This event occurs $a_1 + a_2 + a_3 + \cdots$ times in a series of b operations. Thus its probability is

$$\frac{a_1 + a_2 + a_3 + \cdots}{b} = \frac{a_1}{b} + \frac{a_2}{b} + \frac{a_3}{b} + \cdots,$$

which may be written in the following form:

$$P(A_1 \text{ or } A_2 \text{ or } A_3 \text{ or } \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots.$$

In this case, as in our examples and in our general derivation, we always assume that any two of the considered events (for instance, A_1 and A_2) are *mutually exclusive*; that is, they cannot occur in the same particular trial. For instance, in the example of the trolley stop, a vehicle cannot follow both the desired and undesired routes simultaneously—either it fulfills the passenger's needs or it does not.

This assumption of mutual exclusiveness of individual outcomes is most important. Without it the rule for addition of probabilities becomes incorrect, and its application leads to gross errors. Let us consider the problem we solved in §3. We were specifically seeking the probability that, with dual firings, either the first or the second marksman would hit the target. We found the probability of a hit for the first marksman to be 0.8 and that for the second 0.7. If we attempted to apply the addition rule to this situation, we would have to conclude that the probability sought was 1.5, a clearly unacceptable result, since we know that the probability for an event cannot exceed unity. This incorrect and meaningless answer results from using the addition rule when it does not apply: the two events under consideration are compatible, that is, not mutually exclusive, since it is perfectly possible that both marksmen will hit the target in the same double shot. A notable number of mistakes made by persons beginning to compute probabilities are due to such erroneous applications of the addition rule. It is therefore essential to avoid this error and to make certain each time the addition rule is applied that all of the events under consideration are really mutually exclusive.

We can now present a general formulation of the addition rule.

Addition rule. The probability of occurrence in a certain trial of one of the events $A_1, A_2, \cdots, A_n$ is equal to the sum of the probabilities of these events, provided each pair of events is mutually exclusive.

§5. The Complete System of Events

During the third governmental lottery bond issue for the recovery and development of the peoples' economy, lasting for a period of 20 years, one-third of the funds was classed as profit, and the other two-thirds went into circulation and was returned at the nominal rate. That is, for this issue every bond had a probability for profit of $\frac{1}{3}$ and a probability for return to circulation of $\frac{2}{3}$. Profit and return to circulation are complimentary events; that is, they are two occurrences, of which one and only one must occur in each instance. The sum of their probabilities is

$$\frac{1}{3} + \frac{2}{3} = 1,$$

and this is no accident. In general, if A_1 and A_2 are two complimentary events, and if in a series of b trials event A_1 occurs a_1 times and event A_2 occurs a_2 times, it is apparent that $a_1 + a_2 = b$. However,

$$P(A_1) = \frac{a_1}{b}, \qquad P(A_2) = \frac{a_2}{b},$$

and therefore

$$P(A_1) + P(A_2) = \frac{a_1}{b} + \frac{a_2}{b} = \frac{a_1 + a_2}{b} = 1.$$

We can obtain the same result by applying the addition rule, since complimentary events are mutually exclusive; therefore

$$P(A_1) + P(A_2) = P(A_1 \text{ or } A_2).$$

In this instance event "A_1 or A_2" is certain, since an evaluation of possible outcomes reveals that nothing else is possible; therefore the probability for this event is unity, and we can again write

$$P(A_1) + P(A_2) = 1.$$

The sum of probabilities for two complimentary events is equal to unity.

Indeed, for each complete system of events in which each pair of events is mutually exclusive the addition rule states

$$P(A_1) + P(A_2) + \cdots + P(A_n) = P(A_1 \text{ or } A_2 \text{ or } \cdots \text{ or } A_n).$$

However, the righthand side of this equation is equal to unity, since it is the probability of a certain event. Therefore, considering the whole system,

$$P(A_1) + P(A_2) + \cdots + P(A_n) = 1,$$

which is what we wished to prove.

Example 1. Of every 100 shots at the target shown in Figure 1 the following score is obtained, on the average:

44 strikes in area 1,
30 strikes in area 2,
15 strikes in area 3,
6 strikes in area 4,
4 strikes in area 5,
1 strike in area 6.

Here, $44 + 30 + 15 + 6 + 4 + 1 = 100$. These six outcomes of a shot apparently represent the complete system of events.

Correspondingly their probabilities are

$$0.44, \quad 0.30, \quad 0.15, \quad 0.06, \quad 0.04, \quad 0.01,$$

and we see that

$$0.44 + 0.30 + 0.15 + 0.06 + 0.04 + 0.01 = 1.$$

Bullets striking area 6 partially or completely miss the target and cannot be counted. However, this does not interfere with computation of the probability of striking this area; we need only subtract from unity the probabilities of hitting all the other areas.

Example 2. Statistics reveal that, in a certain textile mill, of every 100 loom stoppages requiring follow-up attention the causes are distributed, on the average, as follows:

22 are due to breakage of the warp thread,
31 are due to breakage of the woof thread,
27 are due to replacement of the shuttle,
3 are due to breakage of the runner,

and the rest of stoppages occur for miscellaneous reasons. We see that, aside from miscellaneous reasons for loom stoppage, there are four primary causes, whose probabilities are

$$0.22, \quad 0.31, \quad 0.27, \quad 0.03.$$

The sum of these probabilities is 0.83. Including the miscellaneous category, the reasons for loom stoppage constitute a complete system of events. Therefore the probability of loom stoppage due to miscellaneous causes is

$$1 - 0.83 = 0.17.$$

§6. Examples

The above theorems dealing with a complete system of events are often successfully used as a basis for the so-called *a priori* (preexperimental) computation of probabilities. Suppose that artillery fire is directed on a rectangular area, which can be broken down into six equal squares as shown in Figure 2. Often in such an application the area of interest is subject to an equal likelihood of hits. Therefore there are no grounds for supposing that any one of the six squares will be struck more often than the others. We therefore conclude that, on the average, each of the six squares will be hit equally often, that is, the probabilities p_1, p_2, p_3, p_4, p_5, p_6 for strikes on these squares are all equal. If we assume that each firing will necessarily involve one of the squares, it follows that each of the p values will be $\frac{1}{6}$. This is so because all the probabilities are the same and must total unity, as we have seen. Of course, this conclusion, which is based on a series of assumptions, requires experimental corroboration. However, since positive experimental results have been obtained in many similar situations, it is quite practical to rely on theoretical conclusions even before experimental confirmation. In such cases it is usually stated that a given trial involves n different *equiprobable* outcomes (in our example, firing at the area shown in Figure 2 must have as a result the hitting of one of the six squares), and the probability of each of the n outcomes is equal to $1/n$. This type of *a priori* computation is important because it often allows us to foresee the probability of an event under conditions whereby execution of a mass trial is either completely impossible or extremely difficult.

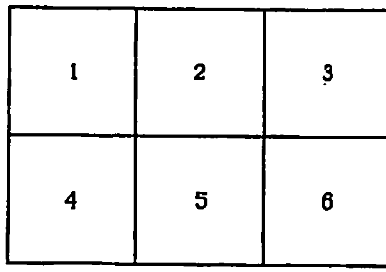

Figure 2

Example 1. Our governmental lottery bonds are usually identified with five-digit numbers. Suppose we wish to find the probability that the last digit of a randomly chosen winning series is 7 (for instance, No. 59607). According to our method for determining probability, we should examine a large number of circulation lists and count how many of the winning cards have numbers ending in 7. The ratio of that number to the total count of winning numbers will be the desired probability. However, since we have every reason to suppose that any of the ten digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 is as likely to appear in the last place of a winning number as any other, we can, without hesitation, state that the sought-for probability is 0.1. The correctness of this theoretical "hypothesis" can be easily proved. Carrying out all the necessary counting on some single lottery form, we will find that any of the ten digits will occur in the last position in about $\frac{1}{10}$ of all cases.

Example 2. A telephone line connecting two points A and B which are 2 kilometers apart is broken at an unknown location. What is the probability that the break occurred no farther than 450 meters from point A? Mentally dividing the phone line into meter-long sections and noting the essential identity of all such sections, we may assume that the probability of breakage is equal for every meter section. We then find, as in Example 1, the required probability:

$$\frac{450}{2000} = 0.225.$$

Chapter 3

Conditional Probabilities and the Rule for Multiplication

§7. The Concept of Conditional Probability

Electric light bulbs are manufactured in two plants, the first of which produces 70%, and the second 30%, of all required bulbs. Bulbs are considered standard quality if they will burn no less than 1200 hours, otherwise they are considered substandard. Of every 100 bulbs made in the first plant, an average of 83 are standard quality, and of every 100 bulbs from the second plant, 63 are standard quality. Using this data, we can easily compute that, on the average, of every 100 bulbs acquired by consumers, there will be 77 standard ones $[(0.83)\cdot(70) + (0.63)\cdot(30) = 77]$. Therefore the probability of buying a standard bulb is 0.77. However, if we assume that the bulbs in a certain store originated in the first plant, the probability of getting a standard lamp is

$$\frac{83}{100} = 0.83.$$

This example shows that the addition to the general circumstances governing the trial (the purchase of bulbs) of specific new conditions (information on where a bulb was made) may alter the probability for the outcome. The development of a concept of probability requires that the totality of conditions governing a given mass trial be precisely determined. By adding some new condition to this totality we generally alter it significantly. The mass trial is then carried out under new conditions. Actually, it is a different type of trial, and therefore the probability for one or another outcome of the trial is different than before.

Thus we have two different probabilities for the same event —the purchase of a standard bulb—which are computed under different conditions. Before we specify additional circumstances (that is, we do not state where the bulb was made) we have an *unconditional probability* for the purchase of a standard bulb, equal to 0.77. However, when we apply the added condition (that is, we specify that the bulb was made in the first plant) we obtain a *conditional probability* of 0.83. Designating as event A the purchase of a standard bulb and as event B its manufacture by the first plant, it is customary to let $P(A)$ represent the unconditional probability of event A and to let $P_B(A)$ represent the same probability with the condition imposed by event B, that is, that the bulb was manufactured by the first plant. We therefore obtain,

$$P(A) = 0.77, \qquad P_B(A) = 0.83.$$

Inasmuch as we can discuss the probability for the outcome of some particular trial only under clearly defined conditions, strictly speaking every probability is conditional; in the literal sense of the words, an unconditional probability cannot exist. However, in most actual situations the circumstances are such that all the possible outcomes under consideration have a common set of conditions which are essential for all outcomes and may be designated by K. Now if in the computation of a probability no conditions are specified except the set K, we call the probability *unconditional*. The term *conditional* is applied to probabilities computed on the assumption that supplementary, clearly defined conditions have been added to the set K which applies to all trials.

Thus in our example we assume that bulbs are manufactured under certain conditions which are the same for all bulbs released for sale. This assumption is so unavoidable and self-apparent there was no need to mention it in the formulation of the problem. If we do not apply additional specifications to a particular bulb, we consider the probability for any given trial as being unconditional. However, if in addition to these conditions we specify some other supplementary requirements, the probabilities meeting these requirements will be conditional.

Example 1. Referring to the problem given at the beginning of this section, the probability for the production of a substandard bulb by the second plant is evidently 0.3. Suppose it is established that a certain bulb is standard quality. What is the probability that this bulb was produced in the second plant?

Of every 1000 bulbs released for sale, on the average 770 will be of standard quality; of these, 581 will come from the first plant and 189 from the second, as is easily deduced from the following. Of every 1000 bulbs, 700, on the average, are manufactured by the first plant. Of every 100 bulbs from the first plant, an average of 83 are of standard quality. Therefore, of the 700 bulbs from the first plant $7 \cdot 83 = 581$ will be standard quality. The remaining 189 bulbs of standard quality are produced by the second plant. Therefore, the probability that the bulb was made by the second plant is

$$\frac{189}{770} \approx 0.245.$$

This is the conditional probability for manufacture of the bulb by the second plant computed on the assumption that the given bulb was standard. Using our former notation, we may write

$$P(\bar{B}) = 0.3, \qquad P_A(\bar{B}) \approx 0.245,$$

where the notation $\bar{B}$ means non-occurrence of event B.

Example 2. Observations carried out in a certain geographical area for many years indicate that of 100,000 children reaching the age of 10 years, an average of 82,277 will live to 40 years of age, and 37,977 will live to 70 years. What is the probability that if a man reaches 40 years of age he will also survive to 70 years?

Since of the 82,277 individuals reaching 40 years an average of 37,977 will reach 70 years, the probability that a person of 40 years will reach 70 years is given by

$$\frac{37{,}977}{82{,}277} \approx 0.46.$$

Let event A be survival to 70 years by a child reaching 10 years, and let event B be survival to 40 years by the same child. We can then say

$$P(A) = 0.37977 \approx 0.38, \qquad P_B(A) \approx 0.46.$$

§8. Derivation of the Rule for Multiplication of Probabilities

Let us return to the notation of the first example of the preceding section. Of every 1000 bulbs placed on sale, 300 on the average, are produced by the second plant, and of these an average of 189 meet standard requirements. We see that the probability of production of the bulb by the second plant (event $\bar{B}$) is

$$P(\bar{B}) = \frac{300}{1000} = 0.3,$$

and the probability of its being of standard quality, given that it was made in the second plant, is

$$P_{\bar{B}}(A) = \frac{189}{300} = 0.63.$$

Since of every 1000 bulbs 189 are produced by the second plant and are also of standard quality, the probability for the joint occurrence of events A and $\bar{B}$ is

$$P(A \mathbin{\&} \bar{B}) = \frac{189}{1000} = \frac{300}{1000} \cdot \frac{189}{300} = P(\bar{B}) \cdot P_{\bar{B}}(A).$$

This "multiplication rule" is easy to apply to the general case. Suppose that in every series of n trials result B occurs an average of m times and that in every series of m such trials where result B is observed result A is observed l times. Then in every series of n trials the simultaneous occurrence of events A and B will be observed an average of l times. Therefore,

$$P(B) = \frac{m}{n}, \qquad P_B(A) = \frac{l}{m},$$
$$P(A \mathbin{\&} B) = \frac{l}{n} = \frac{m}{n} \cdot \frac{l}{m} = P(B) \cdot P_B(A). \tag{3.1}$$

Multiplication Rule. The probability for the joint occurrence of two events is equal to the product of the probability of the first event and the conditional probability of the second event, computed on the assumption that the first has taken place.

It becomes apparent that we can consider as happening first either of two given events, so there is just as much justification for

$$P(A \mathbin{\&} B) = P(A) \cdot P_A(B) \tag{3.1'}$$

as for *(3.1)*, from which we obtain the important relationship

$$P(A) \cdot P_A(B) = P(B) \cdot P_B(A). \tag{3.2}$$

In the example,

$$P(A \,\&\, \bar{B}) = \frac{189}{1000}, \qquad P(A) = \frac{77}{100}, \qquad P_A(\bar{B}) = \frac{189}{770},$$

which confirms *(3.1′)*.

Example. At a certain plant, 96% of manufactured items are judged to be satisfactory (event A); of every 100 satisfactory items an average of 75 turned out to be "first-class" (event B). Find the probability that an item produced in this plant will be first-class.

We are seeking $P(A \,\&\, B)$, since for an item to be first-class it must be satisfactory (event A) and first-class (event B). Using the stated facts,

$$P(A) = 0.96, \qquad P_A(B) = 0.75.$$

Therefore, applying *(3.1′)*,

$$P(A \,\&\, B) = (0.96)(0.75) = 0.72$$

§9. Independent Events

The testing of two spools of thread prepared on different machines shows that a standard length of thread from the first spool has a probability of 0.84 for withstanding a standard test load, and that from the second withstands the load with a probability of 0.78. (If the standard load is equal to, say, 400 grams, the test shows that of 100 samples from the first spool an average of 84 will carry the load, and 16 will not.) What is the probability that two samples of thread, taken from two different spools, will both tolerate the standard test load?

Let A stand for the result that a sample taken from the first spool tolerates the standard load, and let B have the same meaning for a sample from the second spool. Since we are seeking $P(A \,\&\, B)$ we can apply the multiplication rule:

$$P(A \,\&\, B) = P(A) \cdot P_A(B).$$

Here, it is apparent that $P(A) = 0.84$; but what is the value of $P_A(B)$? In accordance with the general definition of conditional probabilities, the latter is the probability that a sample from the second spool will pass the load test, provided this test was passed by a sample from the first spool. But, clearly, event B does not depend on whether event A did or did not take place. If for no other reason, this is true because these tests can be carried out simultaneously, using samples of thread taken from entirely different spools prepared on separate machines. From the practical viewpoint, the percentage of tests in which thread from the second spool will withstand the test load does not depend on the strength of the samples from the first spool; that is,

$$P_A(B) = P(B) = 0.78,$$

from which it follows that

$$P(A \& B) = P(A) \cdot P(B) = (0.84)(0.78) = 0.6552.$$

The circumstances that make this illustration different from all the preceding ones are that here the probability of result B does not change when we add, to the general conditions, the requirement that event A has taken place. That is, the conditional probability $P_A(B)$ is equal to the unconditional probability $P(B)$. Here we can say concisely that *event B does not depend on event A*.

It is easy to see that if B does not depend on A, then A also does not depend on B. From *(3.2)* we see that if $P_A(B) = P(B)$, then $P_B(A) = P(A)$, and this shows that event A does not depend on event B. Therefore, the independence of two events is a *joint* or *mutual* property. We note that for mutually independent events the multiplication rule has the particularly simple form

$$P(A \& B) = P(A) \cdot P(B). \tag{3.3}$$

Every time we applied the addition rule it was necessary to first establish the mutual independence of the events in question, and similarly whenever *(3.3)* is used we must ascertain that events A and B are really independent. Neglecting this requirement results in a large number of errors. If events A and B are mutually dependent, then *(3.3)* is not correct, and it must be replaced by the more general *(3.1)* or *(3.1')*.

Rule *(3.3)* is easily applied to situations in which we seek a probability for the occurrence of not two, but three or more mutually independent events. Suppose, for example, that we have three mutually independent events A, B and C (this means that the probability of any one of these taking place does not depend on whether the remaining two have taken place). Since events A, B and C are mutually independent, *(3.3)* yields

$$P(A \,\&\, B \,\&\, C) = P(A \,\&\, B) \cdot P(C).$$

Replacing $P(A \,\&\, B)$ by its value from *(3.3)*, we find

$$P(A \,\&\, B \,\&\, C) = P(A) \cdot P(B) \cdot P(C). \tag{3.4}$$

It is apparent that a similar rule would hold when the result under consideration consists of any number of events, provided only that they are mutually independent (that is, the probability of any of them does not depend on the occurrence or non-occurrence of any of the others).

The probability for the joint occurrence of any number of mutually independent events is equal to the product of the probabilities of these events.

Example 1. A worker tends three machines. The probability that in any given hour the first machine will not require the worker's attention is 0.9; for the second machine it is 0.8; for the third it is 0.85. Find the probability that during a given hour none of the machines will require the worker's attention.

Accepting that the machines work independently of each other, we find from *(3.4)* that the required probability is

$$(0.9)(0.8)(0.85) = 0.612.$$

Example 2. Under the conditions of Example 1, find the probability that at least one of the three machines will not require the worker's attention during a given hour.

Here we are dealing with a probability of the form $P(A$ or B or $C)$, and therefore we think first of the addition rule. However, we immediately conclude that this rule is not applicable, since any two of the three events under consideration are mutually compatible (nothing prevents two of the machines from working normally during a given hour). Aside from this we also note that the sum of the three given probabilities markedly exceeds unity and therefore cannot represent any probability.

To solve this problem, we note the probability that a machine will require the worker's attention is equal to 0.1 for the first machine, 0.2 for the second, and 0.15 for the third. Since these three events are mutually independent, the probability that all three events will take place is given by *(3.4)*:

$$(0.1)(0.2)(0.15) = 0.0003.$$

But the events "all three machines will require attention" and "at least one of the three will work quietly" represent a pair of mutually incompatible events. Therefore the sum of their probabilities must be equal to unity, and the desired probability is $1 - 0.0003 = 0.9997$. When the probability of an event is this close to unity it can practically be considered a certainty. This means that almost always during a given hour at least one machine will be operating normally.

Example 3. Under certain conditions the probability of shooting down an enemy aircraft by one rifle shot is 0.004.

Find the probability of destroying an enemy plane with simultaneous fire from 250 rifles.

With a single shot there is a probability of $1 - 0.004 = 0.996$ that the airplane will not be downed. The probability that it will not be downed by all 250 shots is equal, in accordance with the rule for multiplication of independent events, to the product of 250 factors, each of which is 0.996, that is, equal to 0.996^{250}. On the other hand, the probability that at least one of the 250 shots will down the airplane is equal to

$$1 - 0.996^{250}.$$

Detailed computation, which will not be presented here, shows that the required number is about $\frac{5}{8}$. Thus, though the probability of downing an enemy plane by a single rifle shot is exceedingly small (0.004), the simultaneous firing of a large number of rifles leads to a probability for the desired result which is quite appreciable.

The concept we have used in the two preceding examples may easily be generalized, and it leads to an important rule. In both examples we dealt with the probability $P(A_1 \text{ or } A_2 \text{ or} \cdots \text{ or } A_n)$ of the occurrence of at least one of several mutually independent events $A_1, A_2, \cdots, A_n$. If we let $\bar{A}_k$ designate the non-occurrence of event A_k (that is, A_k will not take place), then events A_k and $\bar{A}_k$ are mutually incompatible, and

$$P(A_k) + P(\bar{A}_k) = 1.$$

On the other hand, events $\bar{A}_1, \bar{A}_2, \cdots, \bar{A}_n$ are clearly mutually independent, and therefore

$$\begin{aligned} P(\bar{A}_1 \,\&\, \bar{A}_2 \,\&\, \cdots \,\&\, \bar{A}_n) &= P(\bar{A}_1) \cdot P(\bar{A}_2) \cdots P(\bar{A}_n) \\ &= [1 - P(A_1)] \cdot [1 - P(A_2)] \cdot \cdots \cdot [1 - P(A_n)]. \end{aligned}$$

Finally, events $(A_1 \text{ or } \cdots \text{ or } A_n)$ and $(\bar{A}_1 \,\&\, \bar{A}_2 \,\&\, \cdots \,\&\, \bar{A}_n)$ are also mutually exclusive (only one of two things can occur: either at least one of events A_k will occur or all the events of $\bar{A}_k$ will occur). Therefore

$$P(A_1 \text{ or } A_2 \text{ or} \cdots \text{or } A_n) = 1 - P(\bar{A}_1 \,\&\, \bar{A}_2 \,\&\, \cdots \,\&\, \bar{A}_n)$$
$$= 1 - [1 - P(A_1)] \cdot [1 - P(A_2)] \cdots [1 - P(A_n)]. \qquad (3.5)$$

This important equation, which makes possible the computation of the probability of the occurrence of *at least one* of events $A_1, A_2, \cdots, A_n$, using the probability data for these events, is true when and only when these events are mutually independent. Quite commonly, all events A_k have the same probability p (as in Example 3), and then

$$P(A_1 \text{ or } A_2 \text{ or} \cdots \text{ or } A_n) = 1 - (1 - p)^n. \qquad (3.6)$$

Example 4. Under certain conditions of firing at an aerial target a shot is considered successful if the firing point of the charge is not removed more than 10 meters from the center of the target in any of three directions (height, breadth, or length). That is, in order for a shell to be effective it must explode within a cube whose center corresponds to the center of the target and whose edge is equal to 20 meters. Suppose the probability of a deviation greater than 10 meters in a given direction amounts to

$$p_1 = 0.08 \text{ for length,}$$
$$p_2 = 0.12 \text{ for breadth,}$$
$$p_3 = 0.1 \text{ for height.}$$

What is the probability P that a shot will be ineffective?

In order for a shot to be ineffective its firing point must deviate from the center of the target by more than 10 meters in at least one of the three directions. Inasmuch as these three events may normally be considered independent (since they are basically due to different causes), we may apply *(3.5)* for the solution of this problem, which gives

$$P = 1 - (1 - p_1) \cdot (1 - p_2) \cdot (1 - p_3) \approx 0.27.$$

We can conclude therefore that about 73 of every 100 shots will be effective.

Chapter 4

Consequences of Addition and Multiplication Rules

§ 10. The Derivation of Certain Inequalities

Let us return to the example involving light bulbs considered in Chapter 3. We shall use the following notation:

A = bulb of standard quality,
$\bar{A}$ = bulb of substandard quality,
B = bulb produced by the first plant,
$\bar{B}$ = bulb produced by the second plant.

It is apparent that events A and $\bar{A}$, as well as events B and $\bar{B}$, represent a pair of incompatible events.

If a bulb is standard quality (A), either it was produced by the first plant (A & B) or it was produced by the second plant (A & $\bar{B}$); since these two events are obviously incompatible, the addition rule gives

$$P(A) = P(A \,\&\, B) + P(A \,\&\, \bar{B}). \qquad (4.1)$$

Similarly, we find

$$P(B) = P(A \,\&\, B) + P(\bar{A} \,\&\, B). \qquad (4.2)$$

Finally, let us consider event (A or B), which may occur in three different ways:

$$1)\ A \,\&\, B, \qquad 2)\ A \,\&\, \bar{B}, \qquad 3)\ \bar{A} \,\&\, B.$$

Of these three possibilities, any two are incompatible; therefore the addition rule yields

$$P(A \text{ or } B) = P(A \,\&\, B) + P(A \,\&\, \bar{B}) + P(\bar{A} \,\&\, B). \quad (4.3)$$

Adding equations *(4.1)* and *(4.2)*, and noting *(4.3)*, we find

$$P(A) + P(B) = P(A \,\&\, B) + P(A \text{ or } B),$$

from which we obtain

$$P(A \text{ or } B) = P(A) + P(B) - P(A \,\&\, B). \quad (4.4)$$

We have derived a very important result. Though it is based on a concrete example, the example is sufficiently general that the conclusion may be taken as valid for any pair of events A and B. Up to now we have obtained expressions for the probability $P(A$ or $B)$ for the typical situation involving some relationship between A and B (we first assumed them to be mutually incompatible and later assumed them to be mutually independent). Equation *(4.4)* is applicable, without further restrictions, to any pair of events A and B. However, we should not disregard one basic difference between equation *(4.4)* and the previous equations. Previously, we expressed the probability $P(A$ or $B)$ in terms of $P(A)$ and $P(B)$ only, and knowing the probability of events A and B we could always find a unique value for the probability of event (A or B). In *(4.4)* the matter is different: in order to compute $P(A$ or $B)$ using *(4.4)* we must know $P(A \,\&\, B)$ as well as $P(A)$ and $P(B)$ (that is, the probability for the joint occurrence of events A and B). Finding this probability for the general case—wherein there is an arbitrary relationship between A and B—is usually less easy than finding

$P(A \text{ or } B)$. Therefore *(4.4)* is used infrequently for practical computations though its theoretical importance is great.

Let us demonstrate that *(4.4)* allows the derivation of the equations we have obtained for particular applications. If events A and B are mutually incompatible, event $(A \;\&\; B)$ is impossible, and $P(A \;\&\; B) = 0$, and *(4.4)* becomes

$$P(A \text{ or } B) = P(A) + P(B)$$

(the addition rule). If events A and B are mutually independent, we find, from *(3.3)*,

$$P(A \;\&\; B) = P(A) \cdot P(B),$$

and *(4.4)* yields

$$\begin{aligned} P(A \text{ or } B) &= P(A) + P(B) - P(A) \cdot P(B) \\ &= 1 - [1 - P(A)] \cdot [1 - P(B)], \end{aligned}$$

that is, equation *(3.5)* (for $n = 2$).

We can now derive an important result from *(4.4)*. Since in all instances $P(A \;\&\; B) \geq 0$, it follows from *(4.4)* that in all instances

$$P(A \text{ or } B) \leq P(A) + P(B). \tag{4.5}$$

This inequality may be applied to any number of events. For example, with three events, owing to *(4.5)*,

$$P(A \text{ or } B \text{ or } C) \leq P(A \text{ or } B) + P(C) \leq P(A) + P(B) + P(C).$$

In the same way we obtain the general result:

The probability for the occurrence of at least one of several events never exceeds the sum of the probabilities for these events.

The equality sign applies only when each pair of the given events is mutually incompatible.

§ 11. The Formula of Total Probability

Let us return to the example involving light bulbs and make use of the notation in §7. The probability for a standard quality bulb, given the condition that it was made in the second plant is, as noted before,

$$P_{\bar{B}}(A) = \frac{189}{300} = 0.63,$$

and the probability for a standard bulb, given the condition that it was made in the first plant, is

$$P_B(A) = \frac{581}{700} = 0.83.$$

Let us assume that these two values are known; let us also assume that the probability the bulb was made by the first plant is

$$P(B) = 0.7,$$

and that the probability it was made by the second plant is

$$P(\bar{B}) = 0.3.$$

What is the unconditional probability $P(A)$, that is, the probbility that a given bulb, chosen without consideration of its origin, will be satisfactory?

To solve this problem we will proceed as follows.

Let E stand for the double event: (1) the bulb was made by the first plant, and (2) it is standard. Let F stand for the same event relative to the second plant. Since every bulb is made by either the first or the second plant, event A is equivalent to event (E or F), and since events E and F are mutually incompatible, the addition rule gives

$$P(A) = P(E) + P(F). \tag{4.6}$$

On the other hand, in order for event E to occur it is necessary that: (1) the lamp be produced by the first plant (B), and

(2) it be standard (A). Therefore event E is equivalent to event (B & A), and the multiplication rule gives

$$P(E) = P(B) \cdot P_B(A).$$

In an exactly analogous way we find

$$P(F) = P(\bar{B}) \cdot P_{\bar{B}}(A).$$

These, substituted in *(4.6)*, yield

$$P(A) = P(B) \cdot P_B(A) + P(\bar{B}) \cdot P_{\bar{B}}(A).$$

Substituting numerical values, we find $P(A) = 0.77$. This solves the problem stated.

Example. In the planting of wheat a seed mixture is prepared from stock I, with small additions of stocks II, III, and IV. Let us take a single seed from this mixture, and let us designate the event that this seed is of stock I by A_1; that it is of stock II by A_2; stock III by A_3; stock IV by A_4. The probabilities that a randomly chosen seed will be of a given stock are known to be

$$P(A_1) = 0.96, \qquad P(A_3) = 0.02,$$
$$P(A_2) = 0.01, \qquad P(A_4) = 0.01.$$

(The sum of these four numbers is unity, which must be true of any total system of events.)

The probability that a fruit or tassel bearing 50 or more grains will grow from a seed is

0.50 for seeds of stock I,
0.15 for seeds of stock II,
0.20 for seeds of stock III,
0.05 for seeds of stock IV.

What is the unconditional probability that a tassel will bear no less than 50 grains?

Let K be the event that a tassel will bear no less than 50 grains. Then, from the given conditions,

$$P_{A_1}(K) = 0.50, \qquad P_{A_2}(K) = 0.20,$$
$$P_{A_3}(K) = 0.15, \qquad P_{A_4}(K) = 0.05.$$

The problem is to determine $P(K)$. Let E_1 stand for the event that a seed will be of stock A_1 and that a tassel growing from it will contain no less than 50 grains, which makes E_1 equivalent to event $(A_1 \ \& \ K)$. In a similar manner,

let E_2 stand for event $(A_2 \ \& \ K)$;
let E_3 stand for event $(A_3 \ \& \ K)$;
let E_4 stand for event $(A_4 \ \& \ K)$.

Obviously, for the occurrence of event K it is necessary that one of the events E_1, E_2, E_3, E_4 occurs. Since any two of these events are mutually incompatible, the addition rule yields

$$P(K) = P(E_1) + P(E_2) + P(E_3) + P(E_4). \tag{4.7}$$

On the other hand, the multiplication rule yields

$$\begin{aligned}
P(E_1) &= P(A_1 \ \& \ K) = P(A_1) \cdot P_{A_1}(K), \\
P(E_2) &= P(A_2 \ \& \ K) = P(A_2) \cdot P_{A_2}(K), \\
P(E_3) &= P(A_3 \ \& \ K) = P(A_3) \cdot P_{A_3}(K), \\
P(E_4) &= P(A_4 \ \& \ K) = P(A_4) \cdot P_{A_4}(K).
\end{aligned}$$

Substituting these expressions in *(4.7)*, we find

$$\begin{aligned}
P(K) = P(A_1) \cdot P_{A_1}(K) &+ P(A_2) \cdot P_{A_2}(K) \\
&+ P(A_3) \cdot P_{A_3}(K) + P(A_4) \cdot P_{A_4}(K),
\end{aligned}$$

which is the solution to the problem. Substituting numerical values, we find

$$P(K) = 0.486.$$

The two examples we have considered lead to an important general conclusion which we can now state and prove. Assume that a given trial allows the results $A_1, A_2, \cdots, A_n$, which represent a complete system of events (this means that any two of these events are mutually exclusive and that some one of them must occur). Then for any possible result K of this operation we have the relationship

$$P(K) = P(A_1) \cdot P_{A_1}(K) + P(A_2) \cdot P_{A_2}(K) + \cdots + P(A_n) \cdot P_{A_n}(K) \qquad (4.8)$$

This rule is usually called the *formula for full probability.* It is proven by using the same approach applied in the above two examples: the occurrence of event K requires the occurrence of one of the events A_i & K, and therefore the addition rule yields

$$P(K) = \sum_{i=1}^{m} P(A_i \,\&\, K). \qquad (4.9)$$

The multiplication rule yields

$$P(A_i \,\&\, K) = P(A_i) \cdot P_{A_i}(K).$$

Substituting these expressions in *(4.9)* we obtain *(4.8)*.

§ 12. Bayes' Rule

The equations of the preceding section allow us to derive an important rule which is widely applicable. We shall begin with a formal derivation, then discuss the practical meaning of the final equation.

Again let us assume that events $A_1, A_2, \cdots, A_n$ represent a complete system of events for a given trial or operation. If K stands for an arbitrary result of this trial, the multiplication rule yields

$$P(A_i \text{ or } K) = P(A_i) \cdot P_{A_i}(K) = P(K) \cdot P_K(A_i) \quad (1 \leq i \leq n),$$

from which

$$P_K(A_i) = \frac{P(A_i) \cdot P_{A_i}(K)}{P(K)} \qquad (1 \leq i \leq n).$$

Expressing the denominator of the resulting fraction in the form of the total probability equation, *(4.8)*, we have

$$P_K(A_i) = \frac{P(A_i) \cdot P_{A_i}(K)}{\sum_{r=1}^{n} P(A_r) \cdot P_{A_r}(K)} \qquad (1 \leq i \leq n). \qquad (4.10)$$

This is *Bayes' rule*, which has many practical uses in probability computation. It is most often used in situations illustrated by the following example.

Suppose firing is directed at a target located on a straight segment MN (Figure 3), which we shall mentally divide into five segments a, b', b'', c', c''. Let us assume that the exact position of the target is unknown. We know only the probabilities that it lies in one or another of the five segments:

$$P(a) = 0.48,$$
$$P(b') = P(b'') = 0.21,$$
$$P(c') = P(c'') = 0.05,$$

where each a, b', b'', c', c'' stands for the event that the target is located in that particular segment. (The total of these values is equal to unity.) Since the highest probability is that for segment a, we would aim our fire at this segment. However, because of unavoidable errors in firing, the target may also be damaged even when it is in a segment other than a. Let the probability for hitting the target (event K) be

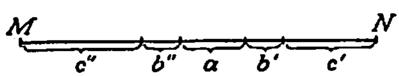

Figure 3

$P_a\ (K) = 0.56$ if the target lies in segment a;
$P_{b'}\ (K) = 0.18$ if the target lies in segment b;
$P_{b''}(K) = 0.16$ if the target lies in segment b'';
$P_{c'}\ (K) = 0.06$ if the target lies in segment c';
$P_{c''}(K) = 0.02$ if the target lies in segment c''.

Suppose now that firing is completed and the target is found to be hit (event K took place). As a result, the probabilities for various locations of the target that we used before [that is, the values of $P(a)$, $P(b')$, $\cdots$] must be reappraised. The qualitative nature of this reappraisal is apparent without any computations. We shot at segment a and hit the target; clearly, under these conditions the probability value $P(a)$ must be increased. However, we wish an accurate quantitative expression for the probabilities following our firing, that is, we wish to find the exact values of $P_K(a)$, $P_K(b')$, $\cdots$, representing the various possible locations of the target with the condition that the completed firing has damaged the target. Bayes' rule, *(4.10)*, provides the answer:

$$P_K(a) = \frac{P(a) \cdot P_a(K)}{\begin{array}{c}P(a) \cdot P_a(K) + P(b') \cdot P_{b'}(K) + P(b'') \cdot P_{b''}(K) \\ + P(c') \cdot P_{c'}(K) + P(c'') \cdot P_{c''}(K)\end{array}} \approx 0.8.$$

We see that $P_K(a)$ is really larger than $P(a)$.

In a similar manner we easily find the probability of $P_K(b')$, $\cdots$ corresponding to other portions of the target. For computational purposes it is useful to note that the expressions for these probabilities obtained by using Bayes' rule differ from each other only in their numerators; their denominators remain the same and are $P(K) \approx 0.34$.

The general situation may be outlined as follows. The conditions of a trial or operation under consideration contain certain unknown elements about which we may frame n different hypotheses (A_1, A_2, $\cdots$, A_n) which constitute a complete system of events. For certain reasons, we know the probabilities $P(A_i)$ of these hypotheses before testing. We also know that hypothesis A_i "communicates" to a certain event K (for instance, hitting the target) a probability $P_{A_i}(K)$, $(1 \leq i \leq n)$ [$P_{A_i}(K)$ is the probability of the event K, computed under the condition that hypothesis A_i is correct]. If event K takes place as the result of an experiment, the probability for hypothesis A_i, must be reappraised, and the problem is then to find the new probability $P_K(A_i)$ of these hypotheses. Bayes' rule gives the needed answer.

In artillery practice use is made of "pre-firing" (trial firing), the purpose of which is to improve knowledge of the firing conditions. In addition to target position, more knowledge may be desired about other circumstances that influence the effectiveness of the firing (specifically, practical techniques). Often pre-firing is carried out several times, and new probabilities for a certain hypothesis must be computed for each set of data. Bayes' rule always provides a ready solution.

Let us condense our notation for the general situation as follows:

$$P(A_i) = P_i,$$
$$P_{A_i}(K) = p_i \qquad (1 \leq i \leq n).$$

Bayes' rule takes on the simple form

$$P_K(A_i) = \frac{P_i p_i}{\sum_{r=1}^{n} P_r p_r}.$$

Let us assume that s pre-firings have been conducted, and that result K occurred m times and failed to occur $s - m$ times. Let K^* stand for this result obtained on the basis of s firings. We may assume that the results of individual firings represent mutually independent events. If hypothesis A_i is correct, the probability of the result K is equal to p_i, and therefore the probability of the opposite event (non-occurrence of K) is equal to $1 - p_i$.

Using the rule for multiplication of independent events, the probability that result K occurred for a certain m firings is $p_i^m(1 - p_i)^{s-m}$. Since the m firings which led to the event K may be any of the s total firings that took place, event K^* may occur in C_m^s different ways. Therefore, the rule for addition of probabilities yields

$$P_{A_i}(K^*) = C_m^s p_i^m (1 - p_i)^{s-m} \qquad (1 \leq i \leq n),$$

and Bayes' rule yields

$$P_{K^*}(A_i) = \frac{P_i p_i^m (1 - p_i)^{s-m}}{\sum\limits_{r=1}^{n} P_r p_r^m (1 - p_r)^{s-m}} \qquad (1 \le i \le n), \quad (4.11)$$

which is the required solution of the problem.

Obviously, such problems arise not only in artillery practice but also in other fields of human endeavor.

Example 1. Using again the problem posed at the beginning of this section let us find the probability of the target lying in segment a if two successive firings into this segment are hits.

Letting K^* stand for a double hit of the target, *(4.11)* yields

$$P_{K^*}(a) = \frac{P(a) \cdot [P_a(K)]^2}{P(a) \cdot [P_a(K)]^2 + P(b') \cdot [P_{b'}(K)]^2 + \cdots}$$

It is left to the reader to carry out the rather simple computation and assure himself that as a result of a double hit the probability of the target lying in segment a becomes even higher.

Example 2. The probability that a standard light bulb will be produced by a certain plant is equal to 0.96. Suppose that a new, simplified system of inspection is under consideration. (The need for simplified checking is often encountered in practice. For example, if every manufactured electric light bulb were subjected to, say, 1200 hours of burning before being released for sale, the user would receive only burned-out or nearly burned-out bulbs. It is necessary to use some other method of checking, such as a simple test of ability to light). The new inspection system gives a positive result with a probability of 0.98 for standard quality bulbs and a positive result with a probability of 0.05 for substandard bulbs. What is the probability that a bulb that passes the new checking system two times is really of standard quality?

Here the complete system of hypotheses is composed of two contrasting events: (1) the bulb meets the standard; (2) the bulb does not meet the standard. The probabilities of these hypotheses before the experiment are $P_1 = 0.96$ and $P_2 = 0.04$. For the first hypothesis the probability that the bulb meets the standard is $p_1 = 0.98$, and for the second it is $p_2 = 0.05$. Following the double testing, the probability of the first hypothesis is, from *(4.11)*,

$$\frac{P_1p_1^2}{P_1p_1^2 + P_2p_2^2} = \frac{(0.96)(0.98)^2}{(0.96)(0.98)^2 + (0.04)(0.05)^2} \approx 0.9999.$$

Note that if the bulb has passed the testing involved in this example, our assumption that it is standard will be in error only one time in ten thousand, which fully satisfies practical needs.

Example 3. A patient is thought to have one of three diseases A_1, A_2, or A_3, whose probabilities under the given conditions are

$$P_1 = \tfrac{1}{2}, \qquad P_2 = \tfrac{1}{6}, \qquad P_3 = \tfrac{1}{3}.$$

A test is carried out to help the diagnosis, and it yields a positive result with a probability of 0.1 for disease A_1, a probability of 0.2 for disease A_2, and a probability of 0.9 for disease A_3. The test is conducted five times, and the results are positive four times and negative once. What is the probability of each disease after this testing?

For disease A_1 the probability of the given outcome is, from the multiplication rule, $p_1 = C_4^5(0.1)^4 \cdot (0.9)$; for disease A_2 the probability is $p_2 = C_4^5(0.2)^4 \cdot (0.8)$; for disease A_3 the probability is $p_3 = C_4^5(0.9)^4 \cdot (0.1)$. Using Bayes' rule, we find that, subsequent to the testing, the probability for disease A_1 is

$$\frac{P_1p_1}{P_1p_1 + P_2p_2 + P_3p_3}$$

$$= \frac{\frac{1}{2}(0.1)^4 \cdot (0.9)}{\frac{1}{2}(0.1)^4 \cdot (0.9) + \frac{1}{6}(0.2)^4 \cdot (0.8) + \frac{1}{3}(0.9)^4 \cdot (0.1)} \approx 0.002;$$

the probability for disease A_2 is

$$\frac{P_2p_2}{P_1p_1 + P_2p_2 + P_3p_3}$$

$$= \frac{\frac{1}{6}(0.2)^4 \cdot (0.8)}{\frac{1}{2}(0.1)^4 \cdot (0.9) + \frac{1}{6}(0.2)^4 \cdot (0.8) + \frac{1}{3}(0.9)^4 \cdot (0.1)} \approx 0.01;$$

the probability for disease A_3 is

$$\frac{P_3p_3}{P_1p_1 + P_2p_2 + P_3p_3}$$

$$= \frac{\frac{1}{3}(0.9)^4 \cdot (0.1)}{\frac{1}{2}(0.1)^4 \cdot (0.9) + \frac{1}{6}(0.2)^4 \cdot (0.8) + \frac{1}{3}(0.9)^4 \cdot (0.1)} \approx .988.$$

Since the three events A_1, A_2, A_3 also constitute a total probability system after the testing, we may, as a computational check, add the three values to ascertain that their total is equal to unity.

Chapter 5

Bernoulli's Method

§ 13. Examples

Example 1. The fibers of a given type of cotton exceed a length of 45 millimeters 75% of the time on the average; 25% of the time they are shorter than or equal to this length. Find the probability that of three randomly chosen fibers two will be shorter and one will be longer than 45 millimeters.

Let us denote by A the event that a selected fiber is less than 45 millimeters and by B the event that it is greater than 45 millimeters. Evidently,

$$P(A) = \frac{3}{4}, \qquad P(B) = \frac{1}{4}.$$

Let us now agree that the notation AAB will stand for the complex event: the first two selections are less than 45 millimeters and the third is greater than 45 millimeters. The meaning of BBA, ABA, ... is clear. The problem is to compute the probability for event C, the situation that two out of three fibers will be less than 45 millimeters and one will be greater than 45 millimeters. It is apparent that one of the following distributions will have to occur for this to take place:

$$AAB, \qquad ABA, \qquad BAA. \tag{5.1}$$

Since any two of these three distributions are mutually incompatible, the addition rule yields

$$P(C) = P(AAB) + P(ABA) + P(BAA).$$

All three terms on the right are equal to each other, since the choice of fibers may be assumed to be mutually independent events. According to the multiplication rule for independent events, the probability for each of the distributions in *(5.1)* is equal to the product of three quantities, of which two are $P(A) = \frac{3}{4}$ and one is $P(B) = \frac{1}{4}$. Since the probability of each of the three distributions in *(5.1)* is $(\frac{3}{4})^2 \cdot \frac{1}{4} = \frac{9}{64}$, we have

$$P(C) = 3 \cdot \frac{9}{64} = \frac{27}{64},$$

which is the solution to the problem.

Example 2. Observations over many years have shown that of every 1000 newborns 515 are boys and 485 are girls. A certain family has six children. Find the probability that in this family there are no more than two girls.

For the event to occur the family must have no girls, one girl, or two girls. The probabilities for these individual situations may be written P_0, P_1, P_2. We see from the addition rule that the desired probability is

$$P = P_0 + P_1 + P_2. \tag{5.2}$$

The probability that any child is a boy is 0.515, and the probability that any child is a girl is 0.485. It is easiest to find P_0, which is the probability that all the children in the family are boys. Since the birth of a particular child is an event independent of all other births, we see from the multiplication rule that the probability for a family with six boys is equal to the product of six quantities, each equal to 0.515:

$$P_0 = (0.515)^6 \approx 0.018.$$

Let us turn to the calculation of P_1, the probability that of the six children in the family one is a girl. This event may take place in six different ways, depending on whether the girl was born as the first child, the second child, and so on. Let us consider for instance, that the girl was the fourth child. The probability for this particular event is equal to the product of six quantities, five of which are 0.515 and one of which (located in the fourth position) is 0.485: $(0.515)^5 \cdot (0.485)$; every other of five other possible birth sequences has exactly the same probability. Applying the addition rule, we find that the probability P_1 is equal to six times this product:

$$P_1 = 6 \cdot (0.515)^5 \cdot (0.485) \approx 0.105.$$

Consider now the calculation of P_2, the probability that two children are girls and four are boys. This event can occur in several different ways. One possibility is that the second and fifth children are girls and the rest are boys. Applying the multiplication rule, we find that the probability for each of these possibilities is $(0.515)^4 \cdot (0.485)^2$. Applying the addition rule, we find that P_2 is equal to this quantity multiplied by the number of all possible sequences of the type under consideration. The entire problem hinges on the determination of the latter number.

Each sequence is characterized by the fact that two of the six children are girls, and therefore each is equal to the number of different ways that two children can be chosen out of a group of six. The number of such choices is the number of ways two things can be chosen out of six, that is,

$$C_2^6 = \frac{6 \cdot 5}{2 \cdot 1} = 15.$$

Therefore,

$$P_2 = C_2^6(0.515)^4 \cdot (0.485)^2 = 15 \cdot (0.515)^4 \cdot (0.485)^2 \approx 0.247.$$

Combining the results, we obtain

$$P = P_0 + P_1 + P_2 \approx 0.018 + 0.105 + 0.247 = 0.370.$$

Thus in something under four times out of ten, six-children families will not comprise more than one-third girls and therefore not less than two-thirds boys.

§ 14. Bernoulli's Formulas

In the preceding section we dealt with *repeated trials* involving the possible occurrence of event A. Henceforth we shall use the word "trial" in a very broad sense. In discussing firing at a certain target we shall consider each shot as a trial. In analyzing tests on light bulbs for burning lifetime we shall consider a check on a particular bulb a trial. In investigating the frequency of births in regard to sex, weight, or height we shall consider the study of one specific infant a trial. Generally speaking, a trial will be considered the occurrence of a certain event of interest to us.

We have now come to one of the most important phases of probability theory, an aspect which has much practical application in different branches of knowledge and is also of great significance for probability theory as a mathematical science. We shall consider a sequence of mutually independent trials—trials of such a nature that the probability for one outcome or another in each trial does not depend on past or future outcomes. In each such trial there may occur (or fail to occur) a given event A, which has a probability p regardless of the number of the trial. This approach is known as *Bernoulli's principle*; it was first systematically pursued by the Swiss scientist Jacob Bernoulli, who lived in the late seventeenth century. We have already applied Bernoulli's principle in the preceding examples (see §13). We shall now solve a general problem for which all the earlier examples in this chapter are special applications.

The Problem. Under certain conditions the probability for the occurrence of event A in a given trial is equal to p. Find the probability that a series of n independent trials will involve k occurrences and $n-k$ non-occurrences of event A.

The events whose probabilities are sought can take the form of a number of different combinations. In order to obtain a given combination we must arbitrarily choose k particular trials and assume that event A occurred only in these k trials and that it did not occur in the remaining $n-k$ trials. Therefore each such combination requires the occurrence of n specific results, including k occurrences and $n-k$ non-occurrences of event A. The multiplication rule enables us to deduce that the probability of each combination is equal to

$$p^k(1-p)^{n-k}.$$

The number of different possible combinations is equal to the number of different arrangements, each composed of k trials, that may be made up in n trials; that is, it is equal to C_k^n. Using the addition rule and the well-known formula for combinations,

$$C_k^n = \frac{n(n-1)\cdots[n-(k-1)]}{k(k-1)\cdots 2\cdot 1},$$

we find that the desired probability for k appearances of event A during n independent trials is equal to

$$P_n(k) = \frac{n(n-1)\cdots[n-(k-1)]}{k(k-1)\cdots 2\cdot 1}p^k(1-p)^{n-k}, \quad (5.3)$$

which is the solution to the problem. Often it is more convenient to use the expression C_k^n in a somewhat different form. Multiplying the numerator and denominator by the product $(n-k)(n-k-1)\cdots 2\cdot 1$, we obtain

$$C_k^n = \frac{n(n-1)\cdots 2\cdot 1}{k(k-1)\cdots 2\cdot 1\cdot (n-k)[n-(k+1)]\cdots 2\cdot 1}.$$

If for purposes of brevity we let $m!$ stand for the product of all integers from 1 to m inclusive, we have

$$C_k^n = \frac{n!}{k!(n-k!)}\cdot$$

For $P_n(k)$ this yields

$$P_n(k) = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}. \qquad (5.4)$$

Formulas *(5.3)* and *(5.4)* are usually called *Bernoulli's formulas.* When n and k have large values the calculation of $P_n(k)$ using Bernoulli's formulas becomes somewhat tedious, since the factorials $n!$, $k!$, and $(n-k)!$ are large and cumbersome to compute. In such calculations wide use is made of specially prepared tables of factorials and also of certain approximations.

Example. The probability that the water consumption of a certain industrial plant will be normal (not over a certain number of liters per day) is equal to $\frac{3}{4}$. Find the probabilities that during a six-day period water consumption will be normal for 0, 1, 2, 3, 4, 5, and 6 days.

Letting $P_6(k)$ stand for the probability that the water consumption will not exceed the normal limit during k days out of six, we find, from formula *(5.3)*, using $p = \frac{3}{4}$,

$$P_6(6) = \left(\frac{3}{4}\right)^6 = \frac{3^6}{4^6};$$

$$P_6(5) = 6\cdot\left(\frac{3}{4}\right)^5\cdot\frac{1}{4} = \frac{6\cdot 3^5}{4^6};$$

$$P_6(4) = C_4^6\cdot\left(\frac{3}{4}\right)^4\cdot\left(\frac{1}{4}\right)^2 = C_2^6\frac{3^4}{4} = \frac{6\cdot 5}{2\cdot 1}\cdot\frac{3^4}{4^6} = \frac{15\cdot 3^4}{4^6};$$

$$P_6(3) = C_3^6\cdot\left(\frac{3}{4}\right)^3\cdot\left(\frac{1}{4}\right)^3 = \frac{6\cdot 5\cdot 4}{3\cdot 2\cdot 1}\cdot\frac{3^3}{4^6} = \frac{20\cdot 3^3}{4^6};$$

$$P_6(2) = \frac{6\cdot 5}{2\cdot 1}\cdot\left(\frac{3}{4}\right)^2\cdot\left(\frac{1}{4}\right)^4 = \frac{15\cdot 3^2}{4^6}.$$

$$P_6(1) = 6\cdot\left(\frac{3}{4}\right)\cdot\left(\frac{1}{4}\right)^5 = \frac{6\cdot 3}{4^6}$$

It is apparent that $P_6(0)$ (the probability that there will be excessive consumption on every one of the six days) is $(\frac{1}{4})^6$. All six probabilities are expressed as fractions with the same denominator, $4^6 = 4096$, which is helpful for shortening the calculations. Computation yields the following:

$$P_6(6) \approx 0.18; \qquad P_6(5) \approx 0.36; \qquad P_6(4) \approx 0.30;$$
$$P_6(3) \approx 0.13; \qquad P_6(2) \approx 0.03; \qquad P_6(1) \approx P_6(0) \approx 0.$$

We note that most probably the water consumption will exceed the normal limit for one or two days out of six, and that excessive consumption for five or six days $[P_6(1) + P_6(0)]$ practically never occurs.

§ 15. Calculation of the Most Probable Number of Occurrences of an Event

The foregoing example shows that the probability of normal water consumption during exactly k days first increases as k grows, reaches a maximum, then declines. This is shown in Figure 4, a curve of $P_6(k)$ versus k.

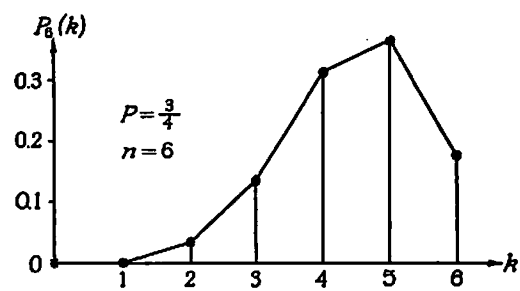

Figure 4

A more striking curve results from plotting $P_n(k)$ against k for large n. Figure 5 shows such a curve for $n = 15$ and $p = \frac{1}{2}$.

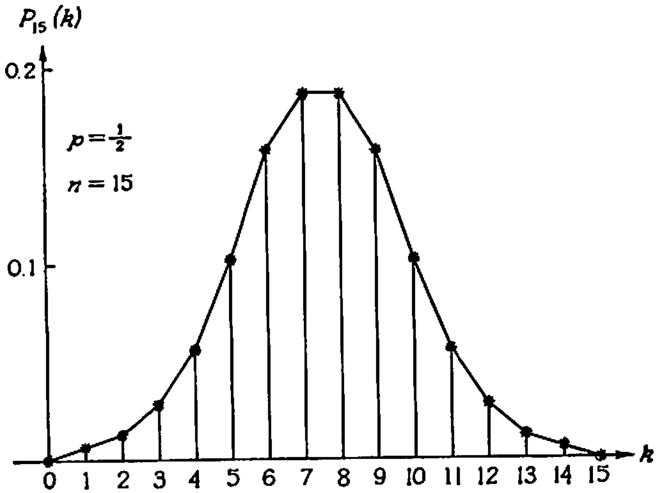

Figure 5

In practical applications we must sometimes predict the most likely frequency for an event; that is, we must predict for what k value the probability $P_n(k)$ is largest (in such applications the values of p and n have to be given). Bernoulli's formula makes it possible to find a simple answer to all such problems.

Let us first calculate the value of the fraction

$$\frac{P_n(k+1)}{P_n(k)}.$$

Using formula *(5.4)*, we find

$$P_n(k+1) = \frac{n!}{(k+1)!(n-k-1)!} p^{k+1}(1-p)^{n-k-1}. \quad (5.5)$$

Combining *(5.3)* and *(5.5)*, we obtain

$$\frac{P_n(k+1)}{P_n(k)} = \frac{n!\,k!\,(n-k)!\,p^{k+1}(1-p)^{n-k-1}}{(k+1)!(n-k-1)!\,n!\,p^k(1-p)^{n-k}}$$

$$= \frac{n-k}{k+1} \cdot \frac{p}{1-p}.$$

The probability $P_n(k+1)$ will be larger, equal to, or smaller than the probability $P_n(k)$, depending on whether the ratio $[P_n(k+1)]/[P_n(k)]$ is larger, equal to, or smaller than unity. As we have seen, this reduces to the question of which of the following three relationships will be correct:

$$\frac{n-k}{k+1}\cdot\frac{p}{1-p} > 1,$$

$$\frac{n-k}{k+1}\cdot\frac{p}{1-p} = 1,$$

$$\frac{n-k}{k+1}\cdot\frac{p}{1-p} < 1. \tag{5.6}$$

Now if we wish to determine what values of k satisfy the inequality $P_n(k+1) > P_n(k)$, we must show when the inequality

$$\frac{n-k}{k+1}\cdot\frac{p}{1-p} > 1$$

will apply, or when

$$(n-k)p > (k+1)(1-p),$$

from which we obtain

$$np - (1-p) > k.$$

Thus as long as k does not exceed the value $np - (1-p)$ $P_n(k+1) > P_n(k)$ always; that is, the probability $P_n(k)$ will always increase with increasing k. For instance, for the data given in Figure 5, $p = \frac{1}{2}$, $n = 15$, $np - (1-p) = 7$; this means that as long as $k < 7$ (for all values of k from 0 to 6 inclusive) $P_n(k+1)$ is greater than $P_n(k)$. This conclusion is corroborated by the curve. In a similar manner, using the other two relationships in *(5.6)*, we find

$$P_n(k+1) = P_n(k) \text{ if } k = np - (1-p),$$

and

$$P_n(k+1) < P_n(k) \text{ if } k > np - (1-p).$$

Thus, as soon as k increases beyond the threshold $np - (1 - p)$ the probability $P_n(k)$ starts to fall and continues to do so through $P_n(n)$.

We see, first, that the behavior of $P_n(k)$, as used in the foregoing examples, follows a general law (first a rise, then a decline) that is applicable to all cases. Moreover, we may now solve the problem stated originally—the determination of the most likely value of k. Let us designate this most likely value by k_0. Then

$$P_n(k_0 + 1) \leq P_n(k_0),$$

and therefore

$$k_0 \geq np - (1 - p).$$

On the other hand,

$$P_n(k_0 - 1) \leq P_n(k_0),$$

and we find

$$k_0 - 1 \leq np - (1 - p)$$

or

$$k_0 \leq np - (1 - p) + 1 = np + p.$$

Therefore the most likely value k_0 of k must satisfy the double inequality

$$np - (1 - p) \leq k_0 \leq np + p; \qquad (5.7)$$

the interval between $np - (1 - p)$ and $np + p$, within which k_0 must lie, has a magnitude of 1, as can be shown by a simple calculation. It follows that if either limit of this interval [for instance, $np - (1 - p)$] is not a whole number, then between these limits there will lie one and only one whole number, which completely determines k_0. This example is typical—for $p < 1$ the value of $np - (1 - p)$ will be a whole number only rarely.

In such an exceptional situation the inequality *(5.7)* yields two values for k_0, $np - (1 - p)$ and $np + p$, which differ from each other by unity. Both values will be the most likely ones—

they are equal, and their probabilities are higher than those for any other values of k. See Figure 5, where $p = \frac{1}{2}$, $n = 15$, and therefore $np - (1 - p) = 7$, $np + p = 8$. The most likely values of k (frequency of the event) are 7 and 8, and the probabilities at these values are identical and approximately equal to 0.196 (all of which is apparent from the diagram).

Example 1. Many years of observation in a certain city have shown that the probability of rain on the first day of July is equal to $\frac{4}{17}$. Find the most likely value for the number of July firsts that will have rain during the next 50 years.

Here, $n = 50$ and $p = \frac{4}{17}$, so

$$np - (1 - p) = 50 \cdot \frac{4}{17} - \frac{13}{17} = 11.$$

The answer is a whole number, which means we are again dealing with an exceptional situation. The two most likely values for rainy July firsts are equal and are 11 and 12.

Example 2. Artillery piece No. 1 is capable of firing 60 shells in a given time interval, with a probability of 0.7 for a hit on each firing. Artillery piece No. 2 fires 50 shells in the same interval, but its probability for a hit on each firing is 0.8. Which of the two guns is likely to score the most hits?

For gun No. 1:

$$n = 60, \qquad p = 0.7, \qquad np - (1 - p) = 41.7, \qquad k_0 = 42.$$

For gun No. 2:

$$n = 50, \qquad p = 0.8, \qquad np - (1 - p) = 39.8, \qquad k_0 = 40.$$

The most probable hit frequency for gun No. 1 is slightly higher than that for gun No. 2.

In practical situations the number n is very great (massive firing, mass industrial production, and so on). In such situations the product np will also be a very large number, provided only that the probability p is not exceedingly small.

In the expressions $np - (1 - p)$ and $np + p$, between which limits the most probable frequency of occurrence must lie, the second terms ($1 - p$ and p) are less than one. Therefore both values of both expressions, and consequently the most probable frequency bracketed by them, must be close to np. Thus, if 1000 shells are fired with a probability of a hit of 0.74 on each shot, $(1000) \cdot (0.74) = 740$ may be taken as the most probable hit frequency.

This conclusion may be presented in a more precise form. If k_0 stands for the most probable frequency for a particular event found in n trials, and $k_0 n$ is the most probable fraction of the n trials in which k will occur, the inequality *(5.7)* yields

$$p - \frac{1 - p}{n} \leq \frac{k_0}{n} \leq p + \frac{p}{n}.$$

Suppose that while using the same value p for the frequency of an event in individual trials we steadily increase the number of trials n (this will also lead to an increase in the probable frequency k_0). The fractions $(1 - p)/n$ and p/n, which appear in the left and right portions of the above inequality, will become smaller and smaller. This means that when n is large these fractions may be neglected—the left and right portions of the inequality, and therefore the fraction k_0/n between them, become equal to p. *With a large number of trials the most probable frequency for the occurrence of an event is practically equal to the probability for the occurrence of the event in a single trial.* Thus, if under certain firing conditions the probability for a hit is 0.84, with a large number of firings it is most probable that about 84% of the shells will strike the target. Of course this does not mean that the probability of exactly 84% hits will be high. On the contrary, even with a large number of firings this "most probable frequency" will be very small (in Figure 5 we saw that the highest probability was 0.196, and we were discussing only 15 trials; with a large number of trials the value would have been much smaller). The probability is the "highest" only in a relative sense. The likelihood of 84% hits is greater than the likelihood of 83% hits.

With very many firings the probability of a particular number of hits is of little interest. For instance, if 200 mortar shells are fired, it is hardly worthwhile to calculate the probability that precisely 137 of them will hit the target; whether there are 137 or 136 or 138 or even 140 hits is immaterial. On the other hand, there is great practical importance in knowing whether more than 100 shots out of 200 will strike the target, or if the number of hits will be between 100 and 125, or if the number of hits will be less than 50, and so forth. How should such probabilities be expressed? Suppose, for example, we wish to find the probability that the number of hits will be between 100 and 120 (including 120). More precisely, we seek the probability for the expression

$$100 < k \leq 120,$$

where k is the number of hits. In order for k to fall within the range covered by this inequality it must be equal to one of the numbers 101, 102, . . . , 120. Applying the addition rule, we find the required probability:

$$P(100 < k \leq 120) = P_{200}(101) + P_{200}(102) + \cdots + P_{200}(120).$$

To obtain this figure directly, we would have to compute 20 separate probabilities of the type $P_n(k)$, using *(5.3)*. Since with large numbers the necessary calculations are tedious, such sums are never calculated directly in practice. Instead, approximation formulas and tables based on advanced mathematical analyses are used. However, even with probabilities of the type $P(100 < k \leq 120)$ application of rather simple principles permits a solution by the method of successive approximations, a technique we shall discuss in the next chapter.

Chapter 6

Bernoulli's Theorem

§ 16. The Contents of Bernoulli's Theorem

Let us again look at Figure 5, in which the probabilities $P_{15}(k)$ for various values of k (the number of times a given event will occur) are shown as vertical lines. The probability applying to a certain *range* of values of k, that is, the probability that the frequency of occurrence will be equal to one of the numbers in a specific range, is given by the sum of the probabilities for all of the frequencies in this range, thus satisfying the addition rule (the probability is equal to the sum of the lengths of all the vertical lines within the range). Figure 5 clearly shows that this sum is a very different quantity for identical segments of the abscissa located in various places. Thus segments $2 \leq k < 5$ and $7 \leq k < 10$ are of identical length, and the probability for each is given by the sum of three vertical lines, but the probability for the latter segment is much greater than the probability for the former. We have seen that all graphs of probability $P_n(k)$ are basically similar to that which is shown in Figure 5. With increasing k the magnitude $P_n(k)$ first increases, then passes through a peak, then falls. Therefore, when there are two ranges of k (using identical portions of the abscissa) the higher probability will always be associated with the one lying closer to the most probable frequency k_0. A segment containing k_0 will always yield a higher probability than any other segment of the same length. Much more can be said about this matter. In particular, there are $n + 1$ possible values of k in n trials $(0 \leq k \leq n)$.

Let us consider a segment that contains k_0 in its center but which contains a very small proportion of all the given values of k, say, one-hundredth of them. If the total number of trials n is made very large, this short segment acquires a very high probability, and all the others taken together have a very low probability. Thus, though the segment is insignificantly small in comparison with n (it occupies only one-hundredth of the abscissa on the graph), the sum of the vertical lines lying within it is greater than the sum of all the remaining vertical lines. The reason is that the vertical lines in the central region of the graph are much longer than those at the ends. Therefore, for large values of n the graph of $P_n(k)$ will take on the form shown in Figure 6.

If we are dealing with a long series of n trials, there is a probability close to unity that the number k of events A will be very close to its most probable value and will differ from the most probable value only by a very small portion of the n trials.

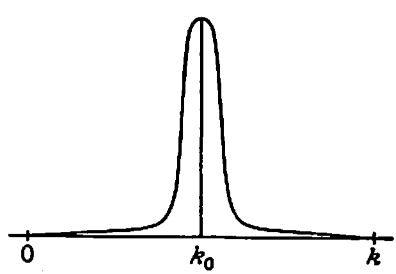

Figure 6

This hypothesis, known as Bernoulli's theorem and discovered at the beginning of the eighteenth century, represents one of the most important laws of probability theory. Until the middle of the nineteenth century all proofs of Bernoulli's theorem required complex mathematical techniques. The great Russian mathematician P. L. Chebyshev was the first to establish a simple, concise basis for this law. We shall now discuss Chebyshev's remarkable proof.

§17. The Proof of Bernoulli's Theorem

We have seen that with a large number n of trials the most probable frequency k_0 for the occurrence of an event A differs very little from the magnitude np, where p, as always, represents the probability for the event A on an individual trial. Therefore it will be sufficient to prove that with a large number of trials the value k for the frequency of event A will show a very high probability of differing only slightly from np—it will differ by not more than a quantity equal to any specified fraction of n (for instance, $0.01n$ or $0.001n$) or, in general, by not more than ϵn, where ϵ is a quantity of any desired smallness. That is, we must prove that the probability

$$P(|k - np| > \epsilon n) \tag{6.1}$$

will become indefinitely small with a sufficiently large n.

The addition rule indicates that the probability given by *(6.1)* is equal to the sum of the probabilities $P_n(k)$ for all those values of k which differ from np by more than ϵn. In Figure 7, a general diagram, this sum is shown as the sum of all vertical lines lying outside segment AB, that is, lying on both sides of np. Inasmuch as the total sum of all the vertical lines (the sum of probabilities in an entire probability system) is equal to unity, the greatest fraction of this total, which approaches unity, will be associated with segment AB, and only a very small fraction will be associated with the regions outside this segment. It follows that

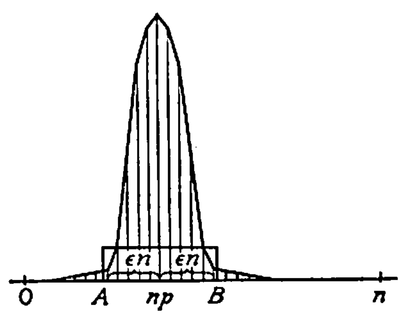

Figure 7

$$P(|k - np| > \epsilon n) = \sum_{|k-np|>\epsilon n} P_n(k) \tag{6.2}$$

We now turn to Chebyshev's proof. Since for each term of the above sum

$$\left|\frac{k - np}{\epsilon n}\right| > 1,$$

and, therefore,

$$\left(\frac{k - np}{\epsilon n}\right)^2 > 1,$$

the sum will increase if each of its terms $P_n(k)$ is replaced by the expression

$$\left(\frac{k - np}{\epsilon n}\right)^2 P_n(k).$$

It follows that

$$P(|k - np| > \epsilon n) < \sum_{|k-np|>\epsilon n} \left(\frac{k - np}{\epsilon n}\right)^2 P_n(k)$$

$$= \frac{1}{\epsilon^2 n^2} \sum_{|k-np|>\epsilon n} (k - np)^2 P_n(k).$$

Furthermore, it is evident that the latter sum will be increased still more if, in addition to those terms already present, we add new ones which will allow k to cover not only the range from $np - \epsilon n$ to $np + \epsilon n$, but the entire series of possible values it can take on (the whole series of numbers from 0 to n inclusive). This yields a greater value, and

$$P(|k - np| > \epsilon n) < \frac{1}{\epsilon^2 n^2} \sum_{k=0}^{n} (k - np)^2 P_n(k). \qquad (6.3)$$

This sum differs usefully from all preceding ones in that its value may be calculated exactly. Chebyshev's method consists precisely of replacing the hard-to-calculate sum *(6.2)* by *(6.3)*, which is amenable to exact evaluation. Even though the necessary calculation may seem very difficult, it can be carried out by anyone who has a knowledge of algebra. We have already made full use of Chebyshev's excellent contribution since it consists specifically in the substitution of *(6.3)* for *(6.2)*.

First we find that

$$\sum_{k=0}^{n} (k - np)^2 P_n(k) = \sum_{k=0}^{n} k^2 P_n(k) - 2np \sum_{k=0} kP_n(k)$$
$$+ n^2p^2 \sum_{k=0}^{n} P_n(k). \qquad (6.4)$$

Of the three sums on the right side, the last is equal to unity, since it represents the probability for a complete system of events. Therefore we must only calculate the sums

$$\sum_{k=0}^{n} kP_n(k) \text{ and } \sum_{k=0}^{n} k^2 P_n(k).$$

Since the terms corresponding to $k = 0$ are equal to zero, we can start the summation with $k = 1$.

1) For the calculation of both sums let us express $P_n(k)$ using *(5.4)*. This gives

$$\sum_{k=1}^{n} kP_n(k) = \sum_{k=1}^{n} \frac{kn!}{k!(n - k)!} p^k(1 - p)^{n-k}.$$

Since $n! = n(n - 1)!$ and $k! = k(k-1)!$, we obtain

$$\sum_{k=1}^{n} kP_n(k)$$
$$= np \sum_{k=1}^{n} \frac{(n - 1)!}{(k - 1)![(n - 1) - (k - 1)]!} p^{k-1}(1 - p)^{(n-1)-(k-1)};$$

or, letting $k - 1 = l$ in the sum on the right side and noting that l varies from 0 to $n - 1$ when k varies from 1 to n, we obtain

$$\sum_{k=1}^{n} kP_n(k) = np \sum_{l=0}^{n-1} \frac{(n - 1)!}{l!(n - 1 - l)!} p^l(1 - p)^{n-1-l} = np \sum_{l=0}^{n-1} P_{n-1}(l).$$

A little reflection shows that the latter sum, $\sum_{l=0}^{n-1} P_{n-1}(l)$, is equal to unity, since it is the sum of probabilities for a complete system of events (all possible frequencies of occurrence for the event l, in $n-1$ trials). Therefore for the sum $\sum_{k=0}^{n} kP_n(k)$ we obtain the very simple expression

$$\sum_{k=0}^{n} kP_n(k) = np. \tag{6.5}$$

2) For the computation of the second sum we first find the magnitude $\sum_{k=1}^{n} k(k-1)P_n(k)$. Since the term corresponding to $k = 1$ is equal to zero, the summation may be started with $k = 2$. Noting that $n! = n(n-1)(n-2)!$ and $k! = k(k-1)(k-2)!$, and setting $k - 2 = m$, as above, we obtain

$$\sum_{k=1}^{n} k(k-1)P_n(k) = \sum_{k=2}^{n} k(k-1)P_n(k)$$

$$= \sum_{k=2}^{n} \frac{k(k-1)n!}{k!(n-k)!} p^k(1-p)^{n-k}$$

$$= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)![(n-2)-(k-2)]!} p^{k-2}(1-p)^{(n-2)-(k-2)}$$

$$= n(n-1)p^2 \sum_{m=0}^{n-2} \frac{(n-2)!}{m!(n-2-m)!} p^m(1-p)^{n-2-m}$$

$$= n(n-1)p^2 \sum_{m=0}^{n-2} P_{n-2}(m) = n(n-1)p^2. \tag{6.6}$$

The last sum is equal to unity, since it is the probability for a certain complete system of events, that is, all possible frequencies for the event in $n-2$ trials. Equations *(6.5)* and *(6.6)* yield

$$\sum_{k=1}^{n} k^2P_n(k) = \sum_{k=1}^{n} k(k-1)P_n(k) + \sum_{k=1}^{n} kP_n(k)$$

$$= n(n-1)p^2 + np = n^2p^2 + np(1-p). \tag{6.7}$$

Both of the required sums have been computed. Substituting the results of *(6.5)* and *(6.7)* into *(6.4)*, we obtain

$$\sum_{k=0}^{n} (k - np)^2 P_n(k) = n^2p^2 + np(1 - p) - 2np \cdot np + n^2p^2$$
$$= np(1 - p).$$

Substituting this into inequality *(6.3)*, we obtain

$$P(|k - np| > \epsilon n) < \frac{np(1 - p)}{\epsilon^2 n^2} = \frac{p(1 - p)}{\epsilon^2 n}, \qquad (6.8)$$

which proves all that was required. We could have made ϵ as small as desired, but once a certain ϵ is chosen it is kept unchanged. However, when n is used in the above manner it can be made as large as desired, and therefore the fraction $[p(1 - p)]/\epsilon^2 n$ will become as small as desired, since the numerator remains constant.

If $p = 0.75$, then $1 - p = 0.25$, and

$$p(1 - p) = 0.1875 < 0.2.$$

If $\epsilon = 0.01$, inequality *(6.8)* yields

$$P\left(\left|k - \frac{3}{4}n\right| > \frac{1}{100}n\right) < \frac{0.2}{0.0001 \cdot n} = \frac{2000}{n}.$$

If $n = 200{,}000$, then

$$P(|k - 150{,}000| > 2000) < 0.01.$$

This would allow us to draw the following practical conclusion. At a given industrial plant a production technique results in 75% of the produced items possessing a particular quality (for instance, being "first line"). Among 200,000 items there is a probability in excess of 0.99 (that is, it is practically a certainty) that between 148,000 and 152,000 items will have the given quality.

It is necessary to make two comments:

1) Inequality *(6.8)* gives a rather crude value for the prob-

ability $P(|k - np| > \epsilon n)$; actually this probability is considerably lower, especially for large values of n. In practice, more accurate estimates are used, even though they are considerably more complex.

2) The estimate given by *(6.8)* becomes much more accurate for values of p very close to 0 or to 1.

Referring to the above example, if the probability that a manufactured item will possess a certain quality is 0.95, then $1 - p = 0.05$, and $p(1 - p) < 0.05$. Setting $\epsilon = 0.005$ and $n = 200{,}000$, we find

$$\frac{p(1-p)}{\epsilon^2 n} < \frac{(0.05)\cdot(1{,}000{,}000)}{(25)\cdot(200{,}000)} = 0.01,$$

as before. But here ϵn is equal not to 2000 but to 1000. We conclude, using $np = 190{,}000$, that from the viewpoint of practical accuracy the number of items possessing the given quality will lie between 189,000 and 191,000, with a total run of 200,000. Thus for $p = 0.95$ inequality *(6.8)* practically guarantees that for the expected number of items having a given quality the range will be one-half that found for $p = 0.75$; otherwise,

$$P(|k - 190{,}000| > 1000) < 0.01.$$

Problem. It is known that 25% of the workers in a certain labor force have a high-school education. Suppose that 200,000 workers are randomly chosen for a certain project. Find (1) the most probable percentage of workers with a high-school education among the 200,000, and (2) the probability that the true (actual) number of such workers will not differ from the most probable value by 1.6%.

Here, the probability of having a high-school education is equal to $\frac{1}{4}$ for each of the 200,000 randomly chosen workers (this circumstance really defines the meaning of "randomly chosen"). Therefore,

$n = 200{,}000,$

$$p = \frac{1}{4}; \qquad k_0 = np = 50{,}000; \qquad p(1-p) = \frac{3}{16}.$$

We seek the probability that $|k - np| < 0.016np$, or $|k - np| < 800$, where k is the number of workers with a high-school education.

Let us choose ϵ such that $\epsilon n = 800$:

$$\epsilon = \frac{800}{n} = 0.004.$$

Equation *(6.8)* yields

$$P(|k - 50{,}000| > 800) < \frac{3}{16 \cdot (0.000016) \cdot 200{,}000} \approx 0.06,$$

from which we find

$$P(|k - 50{,}000| < 800) > 0.94.$$

The desired most probable frequency is 50,000, and the desired probability is greater than 0.94. Actually, the probability in question is much closer to unity.

Chapter 7

Random Variables and the Distribution Law

§ 18. The Concept of a Random Variable

In our preceding work we have often encountered magnitudes whose numerical values cannot be stated once and for all because they vary under the influence of accidental circumstances. Thus, the number of males in one-hundred births will not be the same for each group of 100 births. Or, the length of a cotton fiber varies considerably not only from one geographical area to another, but also on the same bush and in the same clump.

Let us consider several other examples of magnitudes of this type.

1. When a gun is fired at a certain target under unchanging conditions the shells hit different points, a phenomenon called "scattering." The distance between the gun location and the impact point is a varying quantity, owing to factors that do not lend themselves to analysis. These are termed *random* factors—they have varying numerical magnitudes.

2) The velocity of a gas molecule is not constant, but varies because of collisions with other molecules. Inasmuch as any gas molecule may or may not suffer a collision with another

molecule, the variation in its velocity shows a *random* character.

3) The number of meteorites falling on the earth and reaching its surface during the course of a year is not constant, but is subject to marked variations owing to a series of circumstances that have a *random* character.

4) The weight of wheat grains grown in a certain area is not a specific number, but varies from grain to grain. Since it is impossible to consider the effects of *all* possible factors (the quality of soil, the number of daylight hours, the water supply, and so on) that affect grain size, the weight is a magnitude which is influenced by *chance*.

In spite of the diversity of these examples, they do have one feature in common. Each is concerned with a magnitude which in one way or another characterizes the result of some operation or trial (for instance, the counting of meteorites or the measurement of cotton fiber lengths). Each such magnitude has different values in successive trials, no matter how uniform the conditions, because of random factors that are beyond our control. In probability theory this type of magnitude is called a *random variable.* The previous examples are sufficient to demonstrate the importance of random variables in the application of probability theory to a great variety of theoretical and applied problems.

When we "know" a given random variable we do not necessarily know its numerical value. For example, if we know that a shell has hit 926 meters from its firing point, this distance has already acquired a definite numerical value and has ceased to be a random variable. What, then, must we know about a magnitude in order to characterize it fully as a *random magnitude*? First we must know all the numerical values the magnitude can have. Thus, if under certain conditions the shortest artillery shell range is 904 meters and the longest is 982 meters, the distance a shell will travel from its firing point may have any value within these two limits. In the third example it is evident that the numerical value of meteorites reaching the surface of the earth in a year may be any whole nonnegative number, that is, 0, 1, 2, 3, and so on.

However, knowledge of the range of a random variable does not provide the information which can serve best for practical purposes. Thus, if in the second example we consider a gas at two different temperatures the possible numerical values of molecular velocities may be identical, and therefore a mere listing of these values would not allow any comparison of these temperatures. Nonetheless, different temperatures cause very real differences in the state of a gas; differences to which we get no clue from a simple listing of possible molecular velocities. If we wish to evaluate the temperature of a certain mass of gas and we are given only a tabulation of possible molecular velocities, it is natural to ask how often one velocity or another is actually observed. In other words, we really try to find the *probabilities* for the various possible random magnitudes of interest to us.

§ 19. The Concept of a Distribution Law

We will begin with a very simple example. Let us suppose someone is shooting at a target such as that shown in Figure 8. A hit in area *I* scores three points, a hit in area *II* scores two points, and a hit in area *III* scores one point. (One might think that no score should be given for area *III*, that is, missing the target proper. However, a point may be considered as merely an indication of shooting, and a person who aims badly would only be scored for a shot.)

Let us consider as a random variable the number of points obtained on a given single shot. The possible values for this variable are 1, 2, and 3, and we let p_1, p_2, and p_3 stand for the probabilities of obtaining these particular scores. Thus p_3, stands for the probability of hitting area *I*. Although the possible values of the random variables are the same for all persons shooting at the target, the probabilities p_1, p_2, and p_3 differ markedly from each other for different marksmen and actually are an indication of marksmanship. Thus a very good marksman might show, for example, $p_3 = 0.8$, $p_2 = 0.2$, and $p_1 = 0$; an average marksman might show $p_3 = 0.3$, $p_2 = 0.5$, and $p_1 = 0.2$; and a poor shooter might show $p_3 = 0.1$, $p_2 = 0.3$, and $p_1 = 0.6$. If in a certain shooting competition 12 rounds at a time are fired, the possible numbers of hits in any of the areas *I*, *II*, and *III* will be whole numbers ranging from 0 to 12 inclusive. This fact in itself, however, does not allow us to judge the quality of the shooting. To gain a better understanding of the quality, we must also know the probabilities for the given scores as well as the range of possible scores. That is, we must know the numbers indicating how many times out of 12 shots a given area will be hit.

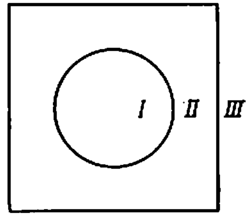

Figure 8

Clearly, this will hold in all similar situations—knowing the probabilities for various possible values of a random variable enables us to predict how often a certain wanted or unwanted outcome will occur and to evaluate the effectiveness or quality of the trial or operation associated with a particular random variable. Practical experience has shown that knowledge of the probabilities for all possible values of a random variable is indeed sufficient for answering any question dealing with this random variable as a quality criterion for some operation. Thus, in order to fully characterize one or another random variable it is necessary and sufficient to know

1) the range of all possible values of this variable;
2) the probability of all specific values in this range.

From this it is evident that a random variable can be conveniently presented in tabular form by listing horizontally, in some order, the possible values of the variable, then listing below each value the corresponding probability. The random variable for the above example involving points for hitting a target, specifically, the score of a good marksman, might be presented in the following manner.

Table I

1	2	3
0	0.2	0.8

In the general case, if $x_1, x_2, \cdots, x_n$ stand for the possible values of a random variable and the corresponding probabilities are given by $p_1, p_2, \cdots, p_n$, the resulting table looks as follows.

x_1	x_2	$\cdots$	x_n
p_1	p_2	$\cdots$	p_n

To give such a table (that is, to give all the possible values of the random variable with their probabilities) is equivalent to stating the distribution law for this random variable. A knowledge of the distribution law for a certain random variable makes it possible to solve all problems dealing with this variable.

Problem. The number of points scored on a single shot by a particular marksman follows the distribution given by table I; the point scores for another marksman show the following distribution.

Table II

1	2	3
0.2	0.5	0.3

Find the distribution law governing the sum of scores achieved by both shooters.

Since the sum we are considering is clearly a random variable, we may set up a random variable table. First we must consider all the possible results of dual firing by the marksmen. These data are listed below. The probability for each result is computed using the multiplication rule for independent events; x stands for the number of points scored by the first marksman, and y stands for the number of points scored by the second marksman.

No. of result	x	y	$x+y$	Probability of result
(1)	1	1	2	$(0)(0.2) = 0$
(2)	1	2	3	$(0)(0.5) = 0$
(3)	1	3	4	$(0)(0.3) = 0$
(4)	2	1	3	$(0.2)(0.2) = 0.04$
(5)	2	2	4	$(0.2)(0.5) = 0.10$
(6)	2	3	5	$(0.2)(0.3) = 0.06$
(7)	3	1	4	$(0.8)(0.2) = 0.16$
(8)	3	2	5	$(0.8)(0.5) = 0.40$
(9)	3	3	6	$(0.8)(0.3) = 0.24$

From the table we see that the sum of interest, $x+y$, may have the values 3, 4, 5, or 6; a value of 2 is not possible because its probability is zero (of course, for the sake of generality, we could consider the value 2 a possible magnitude for the sum $x+y$, with a probability of zero, as we did in Table I). For results (2) and (4) we obtain $x+y = 3$. Consequently, in order to obtain 3 for the sum $x+y$, either result (2) or result (4) must occur. Using the addition rule, we find that the probability for the latter is equal to the sum of the probabilities of these results, that is, it is equal to $0 + 0.04 = 0.04$. For the value $x+y = 4$ it is sufficient that one of results (3), (5), or (7) occur. Again using the addition rule, we find that the probability of this sum is $0 + 0.1 + 0.16 = 0.26$. In a similar manner we find that the probability for the sum $x + y = 5$ is

$$0.06 + 0.4 = 0.46,$$

and the probability for the sum $x + y = 6$, which can take place only with result (9), is equal to 0.24. Thus, for the random variable $x + y$ we obtain the following table of possible values.

Table III

3	4	5	6
0.04	0.26	0.46	0.24

Table III completely solves the problem.

The sum of all four probabilities given in Table III is equal to unity. Of course, every distribution law must have this property, since we are dealing with the sum of probabilities for all possible values of a random variable, that is, the sum of probabilities for a certain complete system of events. This property of distribution laws is a useful tool for checking the correctness of computations.

Chapter 8

Mean Values

§ 20. Determination of the Mean Value of a Random Variable

The two marksmen discussed in §19 may achieve scores of 3, 4, 5, or 6 points in dual firings. The probabilities for these four outcomes are given in Table III. We cannot answer the question "How many points do the two marksmen score in a given (dual) firing?" because different firings result in different scores. However, if we wish to evaluate the shooting accuracy of the two marksmen, we will be concerned not with the result of a single firing (whose result can be accidental or random), but with an *average* value for a whole series of shots. What then is the average score of the two marksmen? This question has been carefully considered and we can give it a clear answer.

We proceed as follows. If the two marksmen fire one-hundred dual rounds, we see from Table III that

about 4 dual rounds will yield a score of 3 points;
about 26 dual rounds will yield a score of 4 points;
about 46 dual rounds will yield a score of 5 points;
about 24 dual rounds will yield a score of 6 points.

Therefore every one-hundred dual firings will earn the marksmen a total number of points given by the sum

$$(3 \cdot 4) + (4 \cdot 26) + (5 \cdot 46) + (6 \cdot 24) = 490.$$

Dividing this by 100, we find that on the average one dual round earns 4.9 points. This is the answer to our question.

Note that instead of dividing the final sum (490) by 100 we could have divided each of the summands by 100 prior to their addition. The sum would then have yielded the average value per round immediately. It is easier to divide the second term of the product in each summand by 100; since these terms were obtained by multiplying the probabilities shown in Table III by 100 we may use the original probability values. Using the latter method, we obtain for the average number of points per round,

$$3 \cdot (0.04) + 4 \cdot (0.26) + 5 \cdot (0.46) + 6 \cdot (0.24) = 4.9.$$

The sum on the left is obtained directly from Table III in a very simple manner. Each of the possible values shown in the top line of the table is multiplied by the probability lying directly below it and all such products are added together.

Now let us apply the same principle to the general case. Suppose that a certain random variable is represented by the following table.

x_1	x_2	$\cdots$	x_k
p_1	p_2	$\cdots$	p_k

Note that the probability for the value x_1 of the magnitude x is given by p_1. This means that in a series of n trials the value x_1 will be encountered approximately n_1 times, where $n_1/n = p_1$. It follows that $n_1 = np_1$. In a similar manner the value x_2 will be found about $n_2 = np_2$ times, and so on. Thus a series of n trials will contain on the average

$$\begin{aligned} n_1 &= np_1 \text{ operations, where } x = x_1; \\ n_2 &= np_2 \text{ operations, where } x = x_2; \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots; \\ n_k &= np_k \text{ operations, where } x = x_k. \end{aligned}$$

The sum of values for the magnitude x in all the n trials will therefore be approximately equal to

$$x_1n_1 + x_2n_2 + \cdots + x_kn_k = n(x_1p_1 + x_2p_2 + \cdots + x_kp_k).$$

Consequently, the mean value $\bar{x}$ of the random variable, corresponding to a single trial and obtained from the above sum, after division by the number of trials n in the particular series, will be equal to

$$\bar{x} = x_1p_1 + x_2p_2 + x_3p_3 + \cdots + x_kp_k.$$

This leads us to the following important result.

To obtain the mean value for a random variable, we must multiply each of its possible values by the probability for this value and then sum all the resulting products.

Of what use is knowing the mean value of a random variable? In order to answer this question more clearly, we shall first consider several examples.

Example 1. The point scores of the two marksmen are random variables whose distribution laws are given by Table I (the first marksman) and Table II (the second marksman). These tables reveal that the first marksman is better than the second. Clearly, the probability for the highest score, 3 points, is much greater for the first marksman, whereas the second marksman shows a higher probability for a low score. However, this approach is unsatisfactory because it is of a purely qualitative character. It does not give us a measure or number which will directly appraise the shooting accuracy of either marksman, in the way, for example, that temperature directly measures the degree to which a physical object is heated. Without such a quantitative measure we may encounter a situation wherein direct inspection of the results does not yield an answer or gives an impression which is open to argument. For instance, if instead of Tables I and II we had the following tables, it would be difficult to decide, from a brief inspection, which of the two marksmen was better.

Table I′

1	2	3
0.4	0.1	0.5

For the first marksman

Table II′

1	2	3
0.1	0.6	0.3

For the second marksman

To be sure, the best score of 3 points is obtained more often by the first marksman than by the second, but the first marksman is also more likely to get the worst score of 1 point. In addition, a score of 2 points is more frequently obtained by the second marksman.

Using the rule given above, let us find the mean values for each of the two marksmen.

1) For the first marksman,

$$1 \cdot (0.4) + 2 \cdot (0.1) + 3 \cdot (0.5) = 2.1.$$

2) For the second marksman,

$$1 \cdot (0.1) + 2 \cdot (0.6) + 3 \cdot (0.3) = 2.2.$$

It is evident that the second marksman has a slightly higher point score on the average than the first. Practically, this means that with often-repeated firing the second marksman will do better than the first. We may now say with assurance that the second marksman shoots better; his average or mean score provides a convenient measure by which we can easily and without equivocation compare the skill of the two shooters.

Example 2. To achieve the best alignment in the assembly of an accurate mechanism, 1, 2, 3, 4, or 5 copies of a certain part may be required, depending on chance. Thus, the number of trials x necessary for proper assembly of the mechanism represents a random variable with the possible values 1, 2, 3, 4, and 5. Suppose the probabilities for these values are as given in the following table.

1	2	3	4	5
0.07	0.16	0.55	0.21	0.01

Let us assume that we must provide a given worker with sufficient parts to assemble 20 mechanisms (we assume that a part rejected during the assembly of one mechanism will not be used in others). We cannot use the above table directly to obtain some order-of-magnitude value for the required number of parts; it only tells us that different quantities are needed on different occasions. However, if we find the mean value $\bar{x}$ for the number of part fittings x needed for a correct assembly and multiply this mean value by 20, we will obtain a probable estimate of the needed figure. We find

$$\bar{x} = 1 \cdot (0.07) + 2 \cdot (0.16) + 3 \cdot (0.55) + 4 \cdot (0.21) + 5 \cdot (0.01) = 2.93$$

$$20\bar{x} = (2.93) \cdot 20 = 58.6 \approx 59.$$

If the worker is to have a small spare supply on hand, in the event his actual need for parts exceeds the probable need, it is practical to give him 60–65 parts.

In the two foregoing examples we dealt with situations in which a certain rough evaluation was required for a particular random variable. Such a rough estimate cannot be obtained from a quick look at a table—this tell us only that the random variable in question may take on certain values with certain probabilities. But by computing the *mean value* of the random variable from a table we obtain the required evaluation. The mean value is precisely that magnitude which our random variable will show on the average, for a lengthy series of trials. From the practical viewpoint, the mean value characterizes the random variable especially well when operations or trials are repeated many times.

Problem 1. A series of trials involving the occurrence of event A which always has the same probability p is carried out. The results of individual trials are independent of each other. Find the mean value for the occurrence of event A in a series of n trials.

The frequency of occurrence of event A in a series of n trials is a random variable with the possible values $0, 1, 2, \cdots, n$. The probability of the value k is given by

$$P_n(k) = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}.$$

Therefore the required mean value is

$$\sum_{k=0}^{n} kP_n(k).$$

We computed this sum in our proof of Bernoulli's theorem, and we found that it was equal to np. We also saw that the *most* probable frequency for A in n trials, with a large n, is close to np. Now we observe that the mean value for the frequency of A is precisely np for any n. Thus, in this case, the most probable value of the random variable corresponds to its mean value. However, we should not conclude that this correspondence holds for all random variables—in general, the most probable value for a random variable may differ markedly from its mean value. Consider, for example, the random variable with the following distribution law.

0	5	10
0.7	0.1	0.2

Here the most probable value is zero, and the mean value is 2.5.

Problem 2. A series of firings is carried out with a hit probability of 0.8. The firing is continued until the first hit, but the firing is not allowed to exceed four shots. Determine the mean value for the number of firings.

The number of firings necessary may be 1, 2, 3, or 4, and we have to find the probabilities for these four situations. When only one shot is made the target must be hit on that firing. The probability for this situation is

$$p_1 = 0.8.$$

In order for exactly two shots to take place the first must be a miss and the second a hit. Using the multiplication rule for independent events, we find that the probability for this situation is

$$p_2 = (1 - 0.8) \cdot (0.8) = 0.16.$$

If three shots are necessary, the first two must be misses and the third a hit. Therefore,

$$p_3 = (1 - 0.8)^2 \cdot (0.8) = 0.032.$$

Finally, if four shots are required, the first three must be misses (regardless of the result of the fourth shot). Therefore,

$$p_4 = (1 - 0.8)^3 = 0.008.$$

Thus the number of firings, as a random variable, is represented by the following distribution law.

1	2	3	4
0.8	0.16	0.032	0.008

The mean value of this variable is therefore equal to

$$1 \cdot (0.8) + 2 \cdot (0.16) + 3 \cdot (0.032) + 4 \cdot (0.008) = 1.248.$$

If a company of soldiers is to carry out 100 such firings, about $(1.248) \cdot 100$, or 125, shells will be required.

Problem 3. Each side of a certain square of ground area is shown, by aerial photography, to be equal to 350 meters. The quality of the aerial photography is indicated by the fact that an error of

0 meters has a probability of 0.42;
$\pm$ 10 meters has a probability of 0.16;*
$\pm$ 20 meters has a probability of 0.08;
$\pm$ 30 meters has a probability of 0.05.

Find the mean value of the *area* of the square.

Because of chance circumstances occurring in aerial photography, the measurement of a side of this ground area constitutes a random variable whose distribution law is given in the following table.

Table I

320	330	340	350	360	370	380
0.05	0.08	0.16	0.42	0.16	0.08	0.05

Using this table, we could immediately find the mean value of the magnitude in question. However, here we do not even

* This is to be understood as follows. An error of $+10$ meters and an error of -10 meters each have a probability of 0.16; this holds also for the other values.

have to apply our computational rule. Since inspection shows that equal errors in the plus or minus direction have equal probabilities, the mean value of the side of the square is equal to the observed value of 350 meters. More precisely, the expression for the mean value will contain the following terms:

$$\begin{aligned}(340 + 360)\cdot(0.16) &= [(350 + 10) + (350 - 10)]\cdot(0.16)\\ &= 2\cdot 350\cdot(0.16);\\ (330 + 370)\cdot(0.08) &= 2\cdot 350\cdot(0.08);\\ (310 + 380)\cdot(0.05) &= 2\cdot 350\cdot(0.05).\end{aligned}$$

The expression for the mean value is therefore

$$350\cdot[0.42 + 2\cdot(0.16) + 2\cdot(0.08) + 2\cdot(0.05)] = 350.$$

We might suppose, by applying the same consideration of symmetry, that the area of the ground square must be $350^2 = 122{,}500$ square meters. This would be true if the mean value of the square of a random variable were equal to its mean squared, but this is not so. The area of the ground square may have the following values:

$$320^2, \quad 330^2, \quad 340^2, \quad 350^2, \quad 360^2, \quad 370^2, \quad 380^2.$$

Which of these is the correct value? The answer depends on which of the seven magnitudes listed in Table I occurs. The probabilities for the seven possible values are, of course, the same as those shown in Table I. That is, the distribution law for the squares is as follows.

320^2	330^2	340^2	350^2	360^2	370^2	380^2
0.05	0.08	0.16	0.42	0.16	0.08	0.05

Consequently, the mean value is

$$320^2\cdot(0.05) + 330^2\cdot(0.08) + 340^2\cdot(0.16) + 350^2\cdot(0.42) + 360^2\cdot(0.16) + 370^2\cdot(0.08) + 380^2\cdot(0.05).$$

For the sake of shortening the computations, it is worthwhile to use the symmetry which is present. The method should be noted, since similar opportunities for simplification arise quite often. We may rewrite the given expression in the form

$$\begin{aligned}
&350^2 \cdot (0.42) + (340^2 + 360^2) \cdot (0.16) + (330^2 + 370^2) \\
&\qquad \cdot (0.08) + (320^2 + 380^2) \cdot (0.05) \\
&= 350^2 \cdot (0.42) + [(350 - 10)^2 + (350 + 10)^2] \cdot (0.16) \\
&\qquad + [(350 - 20)^2 + (350 + 20)^2] \cdot (0.08) \\
&\qquad + [(350 - 30)^2 + (350 + 30)^2] \cdot (0.05) \\
&= 350^2 [0.42 + 2 \cdot (0.16) + 2 \cdot (0.08) + 2 \cdot (0.05)] + 2 \\
&\qquad \cdot 10^2 \cdot (0.16) + 2 \cdot 20^2 \cdot (0.08) + 2 \cdot 30^2 \cdot (0.05) \\
&= 350^2 + 2 \cdot (16 + 32 + 45) = 122{,}686.
\end{aligned}$$

Using this method, we may carry out the entire calculation mentally.

Note that the mean value of the area of the square is a little larger (although the difference is not great) than the square of the mean value, that is, larger than $122{,}500 = 350^2$. We can easily prove that a basic principle is involved here: the mean value of the square of any random variable is always larger than the square of its mean. Assume, for instance, a random variable x which follows some entirely arbitrary distribution:

x_1	x_2	$\cdots$	x_k
p_1	p_2	$\cdots$	p_k

The quantity x^2 will follow the distribution shown below.

x_1^2	x_2^2	$\cdots$	x_k^2
p_1	p_2	$\cdots$	p_k

The mean values for these two quantities are

$$\bar{x} = x_1p_1 + x_2p_2 + \cdots + x_kp_k,$$
$$\overline{x^2} = x_1^2p_1 + x_2^2p_2 + \cdots + x_k^2p_k.$$

It is evident that

$$\overline{x^2} - (\bar{x})^2 = \overline{x^2} - 2(\bar{x})^2 + (\bar{x})^2,$$

and since $p_1 + p_2 + \cdots + p_k = 1$, the three terms on the right may be written as follows:

$$\overline{x^2} = \sum_{i=1}^{k} x_i^2 p_i,$$

$$2(\bar{x})^2 = 2(\bar{x})(\bar{x}) = 2\bar{x}\sum_{i=1}^{k} x_ip_i = \sum_{i=1}^{k} 2\bar{x}x_ip_i,$$
$$(\bar{x})^2 = (\bar{x})^2\sum_{i=1}^{k} p_i = \sum_{i=1}^{k} (\bar{x})^2p_i.$$

Therefore

$$\overline{x^2} - (\bar{x})^2 = \sum_{i=1}^{k} [x_i^2 - 2\bar{x}x_i + (\bar{x})^2]p_i = \sum_{i=1}^{k} (x_i - \bar{x})^2p_i.$$

Since all the terms of the sum on the right are nonnegative, we obtain

$$\overline{x^2} - (\bar{x})^2 > 0.$$

This completes the proof.

Chapter 9

Mean Values of Sums and Products

§ 21. The Theorem on Mean Values of Sums

Often we must find the mean value of the sum of two (or more) random variables whose individual means are known. Suppose, for example, that two industrial plants manufacture an identical product and it is known that on the average one plant produces 120 items a day and the other produces 180 items a day. Using these data, is it possible to find the mean number of items which may be expected from both plants taken together? Or, are these data insufficient, and must we know, in addition to the mean values, something else concerning these two random variables (for instance, must we know in full their distribution laws)?

It is very important to note: (1) in order to compute the mean value of a sum it is sufficient in all instances to know the mean values of the summands, and (2) the mean value of the sum is always given by the simplest possible expression, namely, *the mean value of a sum is always equal to the sum of the means of the summands*. Thus, if x and y are two perfectly arbitrary random variables, then

$$\overline{x+y} = \bar{x} + \bar{y}.$$

Here, x would be the number of items produced by the first plant alone, and y would be the number of items produced by the second plant alone; that is, $\bar{x} = 120, \bar{y} = 180$. Therefore,

$$\overline{x+y} = \bar{x} + \bar{y} = 300.$$

To prove this principle for the general case, let us suppose that the variables x and y follow the distributions shown below.

Table I

x_1	x_2	$\cdots$	x_k
p_1	p_2	$\cdots$	p_k

Table II

y_1	y_2	$\cdots$	y_l
q_1	q_2	$\cdots$	q_l

The possible values for $x + y$ are given by all possible sums of the form $x_i + y_j$, where $1 \leq i \leq k$ and $1 \leq j \leq l$. The probability for a value $x_i + y_j$, which we shall designate p_{ij}, is known—it is the probability for the dual event $x = x_i$ and $y = y_j$, that is, the probability that x will take on the value x_i while y takes on the value y_j. Now if we could consider these two events as being mutually independent, we could, using the multiplication rule, conclude that

$$p_{ij} = p_i q_j. \tag{9.1}$$

However, we shall certainly not assume immediately that these events are independent. Therefore, equation *(9.1)* will not apply generally, and the values in Tables I and II will not permit us to conclude anything regarding p_{ij}.

Using the general rule, we find that the mean value of the magnitude $x + y$ is equal to the sum of the products of all possible values of these magnitudes and their probabilities; that is,

$$\overline{x+y} = \sum_{i=1}^{k} \sum_{j=1}^{l} (x_i + y_j) p_{ij}$$
$$= \sum_{i=1}^{k} x_i \left(\sum_{j=1}^{l} p_{ij} \right) + \sum_{j=1}^{l} y_j \left(\sum_{i=1}^{k} p_{ij} \right). \tag{9.2}$$

Let us now carefully examine the sum $\sum_{j=1}^{l} p_{ij}$. This is the sum of the probabilities for all possible events of the form $x = x_i$ and $y = y_j$, in which i remains the same for all terms of the sum while j takes on all possible different values from 1 to l inclusive. Since the events $y = y_j$ are clearly not the same with different j's, we see, from the addition rule, that the sum $\sum_{j=1}^{l} p_{ij}$ is the probability for the occurrence of *any one of the l events* $(x = x_i, y = y_j$, for $j = 1, 2, \cdots, l)$.

However, to say "some one of the events $x = x_i, y = y_j$ $(1 \leq j \leq l)$ took place" is the same as saying "there occurred event $x = x_i$." Indeed: (1) if one of the events $x = x_i, y = y_j$ took place, the value of j being immaterial, then clearly event $x = x_i$ also took place; (2) on the other hand, if $x = x_i$ occurred, then y necessarily must have had one of the values $y_1, y_2, \cdots, y_l$, and this means that one of the events $x = x_i, y = y_j$ $(1 \leq j \leq l)$ took place. Therefore, $\sum_{j=1}^{l} p_{ij}$—the probability for the occurrence of any one of the events $x = x_i, y = y_j$ $(1 \leq j \leq l)$—is equal to the probability for the event x_i. That is,

$$\sum_{j=1}^{l} p_{ij} = p_i.$$

In an entirely analogous manner we may convince ourselves that

$$\sum_{i=1}^{k} p_{ij} = q_j.$$

Substituting these expressions in *(9.2)*, we find

$$\overline{x+y} = \sum_{i=1}^{k} x_i p_i + \sum_{j=1}^{l} y_j q_j = \bar{x} + \bar{y},$$

which is what we wished to prove.

Obviously, the theorem we have proved for two summands can be readily extended to three or more summands. Using the proof just given, we may write

$$\overline{x + y + z} = \overline{x + y} + \bar{z} = \bar{x} + \bar{y} + \bar{z},$$

and so on.

Example. A series of n firings with the probabilities $p_1, p_2, \cdots, p_n$ is carried out. Find the mean value for the number of hits.

In an individual firing the number of hits is a random variable, which may only take on the values 1 (hit) or 0 (miss). The probabilities for these occurrences on the first firing are therefore equal to p_1 and $1 - p_1$, as a result of which the mean value for a hit on the first firing must be

$$1 \cdot p_1 + 0 \cdot (1 - p_1) = p_1.$$

For the second firing the probability is equal to p_2, and so on. The overall number of hits is the sum of hits in the individual firings. Applying the rule for addition of means, we find that the mean value of the total score is therefore the sum of the mean hit frequencies for the individual shots, or

$$p_1 + p_2 + \cdots + p_n,$$

which is the solution to the problem.

In particular, if the hit probability is the same for all firings $(p_1 = p_2 = \cdots = p_n = p)$, the mean value of the total number of hits is equal to np. We have previously obtained this result in *(6.5)*. Compare the tedious calculations which would have been required using *(6.5)* with the simple conclusion, not requiring any calculation, which led us to the same answer in the above example. Moreover, we have gained not only in simplicity, but also in generality. Previously we considered the results of individual shots as being events which had to be *mutually independent*, and the first method was applicable only when this hypothesis was made. Now we can dispense with this hypothesis, since the rule for the addition of means, which we find in our new conclusions, holds without restrictions for any random variables. Thus, regardless of the possible relationship between individual firings, provided only that the probability for a hit p remains the same on all shots, the mean value for the total number of hits is equal to np.

§ 22. The Theorem on Mean Values of Products

The same problem we solved for the sum of random variables may often also arise in connection with their products. Again suppose that the random variables x and y follow the distributions shown in Tables I and II. The product xy is also a random variable whose possible values are represented by products having the form $x_i y_j$ $(1 \leq i \leq k, 1 \leq j \leq l)$. The probability for the value $x_i y_j$ is equal to p_{ij}. The problem consists of finding a rule which would always make it possible to express the mean value $\overline{xy}$ of the magnitude xy through the mean values of the multiplicands. However, in general, the solution of this problem is impossible. The magnitude $\overline{xy}$ is not, in general, uniquely determined by $\bar{x}$ and $\bar{y}$ (for the same values $\bar{x}$ and $\bar{y}$ different values of $\overline{xy}$ may occur). It follows that no general equation expressing $\overline{xy}$ in terms of $\bar{x}$ and $\bar{y}$ can be given.

There is one important instance when such an expression is possible and the relationship found is an exceedingly simple one. Let us agree to let the random variables x and y be called *mutually independent* if the events $x = x_i$ and $y = y_j$ remain mutually independent for all values of i and j, that is, if the circumstance that one of these two random variables has taken on some particular value in no way affects the distribution law for the other. If the magnitudes x and y are mutually independent in this sense, then

$$p_{ij} = p_i q_j (i = 1, 2, \cdots, k; j = 1, 2, \cdots, l),$$

owing to the multiplication rule for independent events. It follows that

$$\overline{xy} = \sum_{i=1}^{k}\sum_{j=1}^{l} x_i y_j p_{ij} = \sum_{i=1}^{k}\sum_{j=1}^{l} x_i y_j p_i q_j = \sum_{i=1}^{k} x_i p_i \sum_{j=1}^{l} y_j q_j = \bar{x} \cdot \bar{y}.$$

For mutually independent random variables the mean value of a product is equal to the product of the mean values.

This rule, which was developed for the multiplication of two random variables, may be extended to the product of any number of terms. It is necessary only that all the terms be mutually independent, that is, that the assignment of particular values to some of these terms does not influence the distribution laws governing the rest of them.

Example 1. Suppose we are required to measure the area of a rectangular section of ground by the use of aerial photos and that photos taken on a certain occasion give 50 meters and 72 meters as the sides. Let us assume that the distribution law governing the errors is not known, but that it is established that errors of a particular size are equally likely in the plus and minus directions. It is then apparent from symmetry considerations (and it can easily be proven—see Problem 3, p. 76, in Chapter 8) that the mean values of the side of the rectangle agrees with the actual measurements. If these two aerial measurements may be considered mutually independent random variables, the mean value for the area will be equal to the product of the means of the rectangle's sides; the area would be $50 \cdot 72 = 3600$. However, the measurements of the sides could, conceivably, be not mutually independent, as would be true if both measurements were made by the same poorly aligned instrument. If the length measurement gave a magnitude notably greater than the true dimension, we might suppose that the instrument yielded results which were always too large. This would increase the probability for high results for width measurements, and therefore the two values in question can not be considered mutually independent. In such a case, the mean value for the area cannot be assumed to be equal to the product of the means of the sides of the rectangle, and more information is necessary for a true estimate.

Example 2. Suppose that an electrical conductor, the resistance of which depends on chance circumstances, is carrying an electric current whose magnitude also depends on chance. It is known that the mean value of the resistance is 25 ohms and that the mean value of the current is 6 amperes. Compute the mean value of the voltage E across the conductor.

Ohm's law states

$$E = IR,$$

where I is the current and R is the resistance of the conductor. Since $\bar{R} = 25$ and $\bar{I} = 6$, and if we assume that the magnitudes R and I are mutually independent, we find

$$\bar{E} = \overline{RI} = 25 \cdot 6 = 150 \text{ volts.}$$

Chapter 10

Scattering and Mean (Standard) Deviations

§ 23. The Inadequacy of the Mean Value as a Characterization of a Random Variable

We have observed several times that the mean value of a random variable provides a first approximation of the variable. Frequently the mean is sufficient for the practical requirements at hand. For instance, when comparing the skill of two competing marksmen we are satisfied with merely knowing their respective point scores. Or, in comparing the efficiency of two different methods of placing mines in a given battlefield we need use only the mean value of the number of explosions causing harm to the enemy resulting from the two systems. In these cases we have gained an advantage by expressing the random variable as a single number, that is, as a mean value, instead of presenting its whole complex distribution law. Hence, it would seem, in such cases, that we are dealing not with a random magnitude, but with a magnitude that is clearly known.

However, much more often we encounter situations in which practically important characteristics of the random variable are not at all indicated by its mean value, and a more detailed knowledge of its distribution law is a necessity. A typical example of this nature is the flight distance of an artillery shell. Let x stand for the distance between the shell's firing point and its hitting point. With perfect aiming the mean value of this distance $\bar{x}$ is equal to the distance between the gun and the target. Let us assume that the correct distance $\bar{x}$ is known and that perfect aiming has been achieved. Where will the shells hit? Will many of them actually strike the target? Will some of them at least hit close to the target? Knowing the value $\bar{x}$ only does not allow us to answer any of these questions; we know only that sometimes a shell will fall short or long and that the likelihood of over- or under-shooting is about the same. This is true because the mean value for the distance corresponds to the actual distance from the firing point to the target. [The authors are assuming distance is symmetrically distributed about the mean value. (Ed.)] We do not know the answer to the most important question: Will the majority of the shells hit close to the center of the target, with a resulting high probability of damage, or will the majority hit either side of the target, causing a low incidence of hits? Both situations can occur with any given value for $\bar{x}$.

The behavior of two guns having the same $\bar{x}$ may be entirely different. One may exhibit a much greater "scatter" of shells than the other. This means that shells from one gun will, on the average, strike the ground much farther from the target, in both directions, than those from the other gun; the latter should therefore be considered a superior weapon. Yet the $\bar{x}$ for both guns is the same.

Let us consider another example. Suppose that two types of wheat are in use on a certain farm. Depending on chance circumstances (number of plantings, distribution of fertilizer, average hours of daylight, and so), the harvest per square meter is subject to large fluctuations and represents a random variable. Assuming that under identical conditions the mean harvest size for each type of wheat is the same—240 grams per square meter—can we judge the character of the seeds, knowing only this mean value? We cannot, since from the farmer's viewpoint the most valuable type of wheat is that which is least subject to stray influences of weather and other factors—the type which shows the least scatter in its harvest size. So we see that in the testing of one type of wheat or another for its growing ability the range of its possible yields is of no less importance than its mean harvest yield.

§ 24. Various ways of Measuring the Scatter of a Random Variable

The above examples, as well as analogous examples, show convincingly that in many situations wherein we must characterize a random variable thoroughly, presenting only the mean value is entirely inadequate. If only this is done, much needed data remain unknown. To obtain these data, either we must have on hand the entire distribution table for the variable, which from the practical view is always complex and cumbersome, or, in addition to the mean value of the variable, we must include in its description one or two other numbers of a general type in order that this smaller group of numbers, taken together, will give a description of the more essential aspects of the random variable which fulfills practical needs. Let us now consider how such a characterization can be accomplished.

The above examples have shown that in many situations the factor of greatest practical importance is the overall deviation from the mean by a given random variable (that is, the scattering or spread of its values). Are these values, for the most part, closely grouped around the mean (and therefore grouped among themselves), or are they usually quite far from the mean (and therefore sometimes differ widely from each other)? The following rough tables will make it possible to clearly distinguish these two situations. Consider two random variables that follow these distribution laws.

Table I

-0.01	$+0.01$
0.5	0.5

Table II

-100	$+100$
0.5	0.5

Both random variables have zero as their mean value. However, although the first always has values that are very close to zero (and are similar to each other), the second usually has values that differ very much from zero (and therefore differ from each other). For the first variable, knowledge of the mean also provides a rough estimate of its possible actual magnitudes; for the second variable, the mean is usually far removed from its possible real values and does not provide any useful estimate. It is customary to say that the *scatter* magnitude for the second variable is much greater than that for the first variable.

Now the problem is finding some number which will give a convenient *measure* of the scatter of a random variable. This number should at least roughly show how many gross deviations from the mean are to be expected. The deviation $x - \bar{x}$ of the random variable x from its mean value $\bar{x}$ is also a random variable; so is the absolute value $|x - \bar{x}|$ of this deviation, which gives the magnitude of the deviation without regard to sign. It would be desirable to have a value that would roughly characterize this random or chance deviation $|x - \bar{x}|$, which would indicate about how great the deviation should normally be. There are many ways of obtaining such a value; the following three are the most commonly used.

1. *The Mean Deviation.* To obtain a rough estimate of the random variable $|x - \bar{x}|$, it is most natural to take its mean value $\overline{|x - \bar{x}|}$. This mean value of the absolute magnitude of the deviation is called the *mean deviation* or *average deviation* of x. If the random variable x has the distribution table

x_1	x_2	$\cdots$	x_k
p_1	p_2	$\cdots$	p_k

the distribution table for the random variable $|x - \bar{x}|$ has the form

$\|x_1 - \bar{x}\|$	$\|x_2 - \bar{x}\|$	$\cdots$	$\|x_k - \bar{x}\|$
p_1	p_2	$\cdots$	p_k

where $\bar{x} = \sum_{i=1}^{k} x_i p_i$. The mean deviation M_x of the magnitude x is obtained from the formula

$$M_x = \overline{|x - \bar{x}|} = \sum_{i=1}^{k} |x_i - \bar{x}| p_i,$$

in which it is evident that $\bar{x} = \sum_{i=1}^{k} x_i p_i$ once again. For the variables shown in Tables I and II, $x = 0$, and we obtain $M_{xI} = 0.01$ and $M_{xII} = 100$. It should be noted, however, that both of these examples are trivial, since the absolute values of the deviations can take on only one value, which does not satisfy the definition of a random variable.

Let us also compute the mean deviations for the random variables determined by Tables I′ and II′ on p. 72. We have seen that the means for these variables are 2.1 and 2.2—they are very close to each other. The mean deviation for the first variable is

$$|1 - 2.1| \cdot (0.4) + |2 - 2.1| \cdot (0.1) + |3 - 2.1| \cdot (0.5) = 0.9,$$

and for the second it is

$$|1 - 2.2| \cdot (0.1) + |2 - 2.2| \cdot (0.6) + |3 - 2.2| \cdot (0.3) = 0.48.$$

Note that the mean deviation for the second variable is about half that for the first. Practically speaking, this means that though both marksmen clearly get the same mean scores, and in that sense can be considered equally skillful, the second marksman is much more *consistent* or *uniform* in his shooting, since his score shows much less scatter. The first marksman, though getting the same mean score, aims unevenly, often doing better and often doing worse than his average indicates.

2. *The Standard Deviation.** Measuring the value of deviation by the use of the mean deviation is natural, but usually it is tiresome, since computations and evaluations using absolute quantities are often complex and sometimes not even feasible. Therefore, in practice it is often preferable to use another measure of the deviation.

Just as the deviation $x - \bar{x}$ of a random variable x from its mean $\bar{x}$ is a random variable, so is its *square*, $(x - \bar{x})^2$. The distribution table for the latter is as follows.

$(x_1 - \bar{x})^2$	$(x_2 - \bar{x})^2$	$\cdots$	$(x_k - \bar{x})^2$
p_1	p_2	$\cdots$	p_k

The mean value of these squares is therefore

$$\sum_{i=1}^{k} (x_i - \bar{x})^2 p_i.$$

This magnitude gives us an idea of the approximate size of the *square* of the deviation $(x - \bar{x})$. Taking the square root of this quantity, we obtain

$$Q_x = \sqrt{\sum_{i=1}^{k} (x_i - \bar{x})^2 p_i},$$

which is a value that characterizes the size of the deviation and which is called the *standard deviation* of the random vari-

* In the Russian text this is called the "mean quadratic deviation." (Ed.)

able x. The square of Q_x, as given above, is known as the *variance** of the variable. We can see that this new estimate of the deviation has a somewhat more artificial character than does the mean deviation used above. In obtaining it we pursue a more indirect path, obtaining first a rough value for the *squared* deviation, then returning to the actual deviation by taking a square root. But, as we shall see in the next section, using the quantity Q_x greatly simplifies our computations. This is why statisticians commonly use the standard deviation in practice.

Example. Considering again the random variables determined by Tables I′ and II′ on p. 72, we obtain

$$Q^2_{xI'} = (1 - 2.1)^2 \cdot (0.4) + (2 - 2.1)^2 \cdot (0.1) + (3 - 2.1)^2 \cdot (0.5) = 0.89$$

and

$$Q^2_{xII'} = (1 - 2.2)^2 \cdot (0.1) + (2 - 2.2)^2 \cdot (0.6) + (3 - 2.2)^2 \cdot (0.3) = 0.36,$$

from which it follows that

$$Q_{xI'} = \sqrt{0.89} \approx 0.94 \text{ and } Q_{xII'} = 0.6.$$

The mean deviations for the same random variables were

$$M_{xI'} = 0.9 \text{ and } M_{xII'} = 0.48.$$

We observe that the standard deviation, as well as the mean deviation, is much higher for the first of these variables than for the second. Whether we measure the scatter using the mean or the standard deviation, we reach the same conclusion—the first of the two variables shows more spread than does the second.

In both instances the standard deviation is larger than the

* In the Russian text this is called the "dispersion." (Ed.)

mean deviation; this is true for any random variable. Actually, the variance Q_x^2, as the mean of the square of the magnitude $|x - \bar{x}|$, cannot be less than the square of the mean value M_x for $|x - \bar{x}|$, because of the rule given on p. 79. Thus, from $Q_x^2 \geq M_x^2$, it follows that $Q_x \geq M_x$.

3. The Average (Probable) Deviation. Frequently, especially in military applications, use is made of another way of estimating the scatter. We shall demonstrate this method using an example from artillery firing.

Suppose that artillery is fired from point O in the direction OX (Figure 9). The distance x from the point of firing to the point where the shell hits the ground is a random variable whose mean value is given by the location of the "hit center" C, with $OC = \bar{x}$. Individual shells are scattered closer or farther from this hit center.

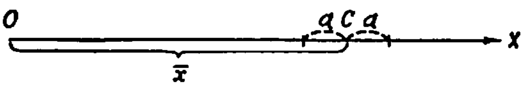

Figure 9

The deviation $x - \bar{x}$ of the random variable under consideration (the flight distance of the shell) from its mean also represents the deviation of the point where the shell hits from the hit center C. Thus any estimate of the magnitude $|x - \bar{x}|$ must give some indication of the degree of scatter or spread of the shells around the center C and therefore is an important parameter showing the accuracy of the firing.

If we lay off a very small segment α to the right and left of the central point C, we find that only a few of the shells will land within the resulting line segment 2α, whose center is C. In other words, the probability that $|x - \bar{x}| < \alpha$, for small α, will be very small. However, suppose we now widen this segment by increasing the magnitude of α (which can be chosen as we wish). The larger we make the line segment the greater will be the percentage of shells that will fall within it, and therefore the higher will be the probability for any one shell landing within the segment. If α is very large, practically all the shells will fall within the segment. Therefore, with a gradual growth of the quantity α, the probability for the inequality

$$|x - \bar{x}| < \alpha$$

increases from zero to unity. At first, with small α, it is more likely that

$$|x - \bar{x}| > \alpha.$$

(that is, that the shell will fall outside the segment); later, with larger α, it becomes more probable that $|x - \bar{x}| < \alpha$ (that is, that the shell will land within the segment). It follows that somewhere in the transition between small and large values for α there must occur a value α_0 for which the shell has an equal probability of landing within or without the line segment $2\alpha_0$, which is drawn around the center C. That is to say, under these conditions the inequalities

$$|x - \bar{x}| < \alpha_0,$$

$$|x - \bar{x}| > \alpha_0$$

are equally probable, and therefore they each have the probability $\frac{1}{2}$ (providing we agree the probability for landing on the exact dividing line $|x - \bar{x}| = \alpha$ is insignificant). With $\alpha < \alpha_0$ the second inequality is more likely, and with $\alpha > \alpha_0$ the first is more likely. So there exists a definite value in reference to which the absolute value of the deviation is equally likely to be either greater or smaller.

How large is α_0? The answer depends on the quality of the artillery firing. Clearly, the value α_0 may serve as a measure of shell scatter in a way similar to utilization of the mean or standard deviation. For example, if α_0 is very small, this means that half of all shells fired will land within a very small portion of the line close to C, which indicates a rather small scatter. On the other hand, if α_0 is large, then even with a large interval containing C we must still expect that half of all shells fired will fall outside these limits. Here, the shell firing clearly shows a marked spread around the center.

The value α_0 is usually called the *average* or *probable deviation* of the magnitude x. Therefore the average or probable deviation of the random variable x is a quantity for which it is equally likely that the actual deviation $x - \bar{x}$ will be larger or smaller. Although the probable deviation of x, which we shall now write E_x, is less convenient for mathematical computations than the mean deviation M_x, and certainly much less convenient than the standard deviation Q_x, it is customary, in artillery practice, to use specifically E_x. We shall presently learn why this does not present any real difficulties in practice.

§ 25. A Theorem Concerning the Standard Deviation

Let us now convince ourselves that the standard deviation really does possess special properties which force us to choose it in preference to all other characteristic measures of the deviation, such as the mean deviation, the probable deviation, or others. As will be seen below, from the practical viewpoint the following problem is of particular importance.

Suppose we have several random variables $x_1, x_2, \cdots, x_n$ with standard deviations $q_1, q_2, \cdots, q_n$. Let $x_1 + x_2 + \cdots + x_n = X$. How can we find the standard deviation Q of the quantity X if we are given $q_1, q_2, \cdots, q_n$, and if we assume that the random variables x_i $(1 \leq i \leq n)$ are mutually independent?

Using the rule for the addition of mean values, we have

$$\bar{X} = \bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_n,$$

and therefore,

$$X - \bar{X} = (x_1 - \bar{x}_1) + (x_2 - \bar{x}_2) + \cdots + (x_n - \bar{x}_n),$$

from which we obtain

$$(X - \bar{X})^2 = \left[\sum_{i=1}^{n} (x_i - \bar{x}_i)\right]^2$$

$$= \sum_{i=1}^{n} (x_i - \bar{x}_i)^2 + \sum_{\substack{i=1 \\ i \neq k}}^{n} \sum_{k=1}^{n} (x_i - \bar{x}_i)(x_k - \bar{x}_k). \quad (10.1)$$

Let us note that

$$\overline{(X - \bar{X})^2} = Q^2, \qquad \overline{(x_i - \bar{x}_i)^2} = q_i^2 \qquad (1 \leq i \leq n).$$

Using the rule for the addition of mean values, we find

$$Q^2 = \sum_{i=1}^{n} q_i^2 + \sum_{\substack{i=1 \\ i \neq k}}^{n} \sum_{k=1}^{n} \overline{(x_i - \bar{x}_i)(x_k - \bar{x}_k)}. \qquad (10.2)$$

However, in accordance with the assumptions, since the magnitudes x_i and x_k are mutually independent for $i \neq k$, the rule for multiplication of mean values of mutually independent variables, with $i \neq k$, yields

$$\overline{(x_i - \bar{x}_i) \cdot (x_k - \bar{x}_k)} = \overline{(x_i - \bar{x}_i)} \cdot \overline{(x_k - \bar{x}_k)}.$$

In this equality both terms on the right are equal to zero, since, for example,

$$\overline{(x_i - \bar{x}_i)} = \bar{x}_i - \bar{x}_i = 0.$$

It follows that in the last sum of *(10.2)* each of the terms taken individually is zero, which leads to

$$Q^2 = \sum_{i=1}^{n} q_i^2,$$

that is, *the variance of the sum of mutually independent random variables is equal to the sum of their variances.*

Thus, for mutually independent random variables we have, in addition to the rule for the addition of mean values, the very important *rule for addition of variances.* Applying this to standard deviations, we obtain

$$Q = \sqrt{\sum_{i=1}^{n} q_i^2}. \tag{10.3}$$

That the standard deviation of a sum may be expressed in terms of the standard deviations of its summands, providing they are mutually independent, is one of the most important properties of standard deviations, and it distinguishes them from mean, probable, or other deviations.

Example 1. For n firings with one and the same probability p of a hit, the mean value for hits is equal to np (as shown on p. 83). To obtain a rough estimate of how large the actual deviation may be from its mean value, we shall first find the standard deviation of the hit frequency. This is most easily done using *(10.3).*

The number of hits in n firings may be considered (see p. 83) as the sum of hits resulting from individual firings. Since the latter are mutually independent random variables, using the addition rule for variances we may apply *(10.3)* for the computation of the standard deviation Q of the total number of hits; here $q_1, q_2, \cdots, q_n$ refer to the standard deviations of the hit frequency in individual firings. Now the number of hits x_i on the ith firing is given by the following table.

1	0
p	$1-p$

Therefore, $\bar{x}_i = p$ and $q_i^2 = \overline{(x - \bar{x}_i)^2} = (1-p)^2 p + p^2 (1-p)$ $= p(1-p)$; consequently,

$$Q = \sqrt{\sum_{i=1}^{n} q_i^2} = \sqrt{np(1-p)}.$$

This completes the solution. Comparing the mean value np of hit frequency with its standard deviation $\sqrt{np(1-p)}$, we note that with large values of n (that is, with large numbers of firings) the latter becomes much smaller than the former, of which it represents only a small fraction. Thus with $n = 900$, $p = \frac{1}{2}$ the mean value of the hit frequency is equal to 450, and the standard deviation is equal to $\sqrt{900 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 15$. So, roughly speaking, the actual number of hits should differ by only 3–4% from the mean value.

Example 2. Suppose we are concerned with the assembly of a certain mechanism composed of n parts, which are arranged in sequence on a certain shaft and retained at the ends of the shaft by a certain holder (see Figure 10). The length of each part may deviate somewhat from its respective standard, and therefore it represents a random variable. We can assume that these random variables are mutually independent. If the mean lengths of the parts and the deviations of the lengths are given by $a_1, a_2, \cdots, a_n$ and $q_1, q_2, \cdots, q_n$, respectively, the mean value and standard deviation of a long sequence composed of n parts are equal to

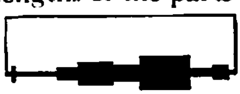

Figure 10

$$a = \sum_{k=1}^{n} a_k \text{ and } q = \sqrt{\sum_{k=1}^{n} q_k^2}.$$

In particular, if $n = 9$, $a_1 = a_2 = \cdots = a_9 = 10$ cm and $q_1 = q_2 = \cdots = q_9 = 0.2$ cm, then $a = 90$ cm and $q = \sqrt{9 \cdot (0.2)^2} = 0.6$ cm.

Thus we see that if on the average the length of each part deviates 2% from its mean, the length of the entire assembly will differ from its mean by only about $\frac{2}{3}$%. This result—the decrease in percentage error on addition of random variables—plays a very important role in the assembly of exact mechanisms. If there was no mutual compensation of deviations from normal size in components, the situations would be frequently encountered wherein the holder could not encompass the assembly of parts or the clearances would be too large. In either instance the entire mechanism would have to be rejected. Now reducing such rejects by decreasing the individual "pass" tolerances, that is, by lowering the allowable limits of deviation for each part from its standard, would not be profitable, since even a comparatively small increase in the precision of a part greatly increases its cost. (In recent times technological thinking has made use of an "allowable tolerance" theory, which is based on the concepts and conclusions of probability theory. This theory is now actively being developed by Soviet scientists.)

Example 3. Suppose that n measurements are made of a certain dimension under identical conditions. As the result of a whole series of circumstances (position of the apparatus, the observer, variations in the state of air, presence of dust, and so on), successive measurements will, in general, give different results; therefore we are dealing with a *random error* of measurement. Let us designate the results of measurements as $x_1, x_2, \cdots, x_n$, the subscript attached to the x being the number of the measurement. The mean value for all these measurements is $\bar{x}$. It is also natural to assume that the standard deviation q is the same for all the measurements, since they are accomplished under unchanging conditions. Finally, as usual, we shall consider $x_1, x_2, \cdots, x_n$ as being mutually independent.

Now let us examine the arithmetic mean of n measurements:

$$\xi = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

This is a random variable. Let us find its mean value and standard deviation. According to the addition rule,

$$\begin{aligned}\bar{\xi} &= \frac{1}{n}\overline{(x_1 + x_2 + \cdots + x_n)} \\ &= \frac{1}{n}(\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_n) = \frac{1}{n}(n\bar{x}) = \bar{x};\end{aligned}$$

that is, the mean value, as was evident at the start, is the same as the mean for each individual measurement. Further, the standard deviation of the sum $x_1 + x_2 + \cdots + x_n$, according to the rule for addition of variances *(10.3)*, is

$$Q = \sqrt{nq^2} = q\sqrt{n},$$

and the standard deviation of the quantity ξ, which represents $1/n$ of this sum, is

$$\frac{Q}{n} = \frac{q}{\sqrt{n}}.$$

We thus reach a very important conclusion. The arithmetic mean of n mutually independent and identically distributed randon variables has:

1) a mean value, which is the same as each of its component means;

2) a standard deviation, which is $\sqrt{n}$ times smaller than that of its components.

Thus if the mean value of a certain distance x is 200 meters and its standard deviation is $q = 5$ meters, the arithmetic mean of a hundred such measurements will naturally have the same mean value of 200 meters, but its standard deviation will be $\sqrt{100} = 10$ times smaller (it will be $q/10 = 0.5$ meters). Therefore we have grounds for supposing that the arithmetic mean of 100 actual measurements will be much closer to the mean value, 200 meters, than will the result of one or another individual measurement.

The arithmetic mean of a large number of mutually independent quantities shows a much smaller variance than do each of these quantities taken separately.

Chapter 11

The Law of Large Numbers

§ 26. Chebyshev's Inequality

We have stated a number of times that knowledge of one of the measures of average deviation of a random variable (for instance, its standard deviation) permits a rough idea of how much the actual values of this variable will deviate from its mean value. However, in itself this measurement does not allow any quantitative evaluation and does not make it possible to compute even approximately how frequently large deviations will be encountered.

The above considerations allow us to make the following simple argument, which was first carried out by Chebyshev. We will start with the expression for the variance of a random variable x (p. 92):

$$Q_x^2 = \sum_{i=1}^{k} (x_i - \bar{x})^2 p_i.$$

Suppose α is any positive number. Now if in the above sum we discard all terms in which $|x_i - \bar{x}| \leq \alpha$, and retain only those in which $|x_i - \bar{x}| > \alpha$, this can only make the sum smaller:

$$Q_x^2 \geq \sum_{|x_i - \bar{x}| > \alpha} (x_i - \bar{x})^2 p_i.$$

However, this sum will decrease still more if in each of its terms we replace the term $(x_i - \bar{x})^2$ by the lesser quantity α^2:

$$Q_x^2 \geq \alpha^2 \sum_{|x_i - \bar{x}| > \alpha} p_i.$$

The sum now appearing on the right is the sum of probabilities for all those values x_i of the random variable x which deviate from $\bar{x}$ in either direction by a quantity larger than α; according to the addition rule this is the probability of the event that the variable x will take on any of those values. That is to say, it is the probability $P(|x - \bar{x}| > \alpha)$ that the actual deviation will be larger than α. It follows that

$$P(|x - \bar{x}| > \alpha) \leq \frac{Q_x^2}{\alpha^2}, \qquad (11.1)$$

which allows us to evaluate the probability for a deviation larger than any given quantity α, providing only that we know the standard deviation Q_x. To be sure, the evaluation given by "Chebyshev's inequality," *(11.1)*, is often very rough, but nonetheless it can often be used directly in practice. It need hardly be noted that its theoretical significance is great.

At the end of the preceding section we considered the following situation: The mean value of a series of measurements is 200 meters, and the associated standard deviation is 5 meters. Now under these conditions the probability of actually getting a deviation of 3 meters is very substantial. In fact, we might suppose that it is larger than $\frac{1}{2}$, but the precise value can, of course, be found only if we have full knowledge of the distribution law governing the measurements. On the other hand, we note that the arithmetic mean ξ of 100 measurements has a standard deviation of only 0.5 meter. It follows from inequality *(11.1)* that

$$P(|\xi - 200| > 3) \leq \frac{(0.5)^2}{3^2} = \frac{1}{36} \approx 0.03.$$

Thus with an arithmetic mean of 100 measurements the probability of getting a deviation of 3 meters becomes very small (it is even smaller than the figures we have shown, and a deviation of this size can practically be disregarded).

In Example 1 of chapter 10 the data indicated a mean of 450 hits in 900 firings, with a standard deviation of 15. If we wish to find the probability that the actual number of hits will be within the limits of, say, 400–500 (that is, $|m - 450| \leq 50$), Chebyshev's inequality states

$$P(|m-450| \leq 50) = 1 - P(|m-450| > 50) \geq 1 - \frac{15^2}{50^2} = 0.91.$$

Actually the true probability is substantially greater than this.

§ 27. The Law of Large Numbers

Suppose we are dealing with n mutually independent random variables $x_1, x_2, \cdots, x_n$, all of which have the same mean value a and standard deviation q. The arithmetic mean of these variables is

$$\xi = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

As shown previously the mean value is equal to a, and the standard deviation is equal to $q/\sqrt{n}$. Under these conditions for any positive value of α, Chebyshev's inequality states

$$P(|\xi - a| > \alpha) \leq \frac{q^2}{\alpha^2 n}. \tag{11.2}$$

Suppose we are considering an arithmetic mean n for measurements of a certain dimension, for which the values given above apply ($q = 5$ meters, $a = 200$ meters). This yields

$$P(|\xi - 200| > \alpha) \leq \frac{25}{\alpha^2 n}.$$

We may choose α very small, for instance, $\alpha = 0.5$ meter; then

$$P(|\xi - 200| > 0.5) \leq \frac{100}{n}.$$

If the number of measurements n is made very large, the right side of this inequality becomes very small. For example, if $n = 10{,}000$, the right side will be 0.01, and for the arithmetic mean of 10,000 measurements

$$P(|\xi - 200| > 0.5) \leq 0.01.$$

If we agree to disregard very unlikely occurrences, for 10,000 measurements we can say with assurance that the arithmetic mean will not deviate from 200 meters by more than 50 centimeters. If we wished to achieve even higher accuracy, for instance, to within 10 centimeters, we could let $\alpha = 0.1$ meter, which would result in

$$P(|\xi - 200| > 0.1) \leq \frac{25}{0.01n} = \frac{2500}{n}.$$

In order for the right side of this inequality to be less than 0.01 we would have to carry out not 10,000 but 250,000 measurements. No matter how minute an α is chosen it is still generally possible to make the right side of inequality *(11.2)* as small as needed—it is necessary only to make n sufficiently large. Thus, with a large enough n the inequality

$$|\xi - a| \leq \alpha$$

may be considered as true with probability as close to unity as desired.

If the random variables $x_1, x_2, \cdots, x_n$ are mutually independent and if they all have the same mean and standard deviation, then the quantity

$$\xi = \frac{x_1 + x_2 + \cdots x_n}{n},$$

for sufficiently large n, will differ by as little as specified from the com-

mon mean with a probability as close to unity as desired.

This represents a commonplace application of one of the most basic theorems of probability theory, the so-called *law of large numbers*, which was formulated in the middle of the last century by the great Russian mathematician Chebyshev. The deeper significance of this remarkable law is that while an individual random variable may take on values that are quite far from its mean value (that is, show sizeable scattering), the arithmetic mean of a large number of random variables behaves very differently: it shows very little scatter and only takes on those values very close to its mean with an overwhelmingly high probability. The reason for this is, of course, that when we take an arithmetic mean the random deviations in one direction or the other cancel each other, and the net deviation is usually small.

An important and frequently applied result of this theorem of Chebyshev's is the technique of evaluating the quality of a large amount of some material by taking a comparatively small sample. For instance, the quality of cotton in bales is judged by taking several small clutches of fibers randomly from several points in the bale. The quality of a large shipment of grain may be ascertained by taking several modest samples dipped at random locations in the storage bin (the sampling cup may hold only 100–200 grams, whereas the bin may contain tens or even hundreds of tons of grain). The judgement of the product made on this basis is an accurate one, since, for example, though the grain sampling cup may represent a small fraction of the total load of grain, it contains quite a substantial number of grains. Because of the law of large numbers this method gives a sufficiently accurate determination of the average weight of a single grain and therefore of the quality of the entire bin. Similarly, a large bale of cotton may be evaluated on the basis of a small clutch sample, containing perhaps several hundred fibers which weigh no more than a tenth of a gram.

§ 28. Proof of the Law of Large Numbers

Up to this point we have considered only those situations in which all the variables $x_1, x_2, \cdots$ had one and the same mean and standard deviation. However, the law of large numbers is also applicable in much broader circumstances. We shall now consider the situation in which the mean values of the variables $x_1, x_2, \cdots$ may be any numbers whatever (which shall be designated $a_1, a_2, \cdots$). The mean value of the quantity

$$\xi = \frac{1}{n}(x_1 + x_2 \cdots + + x_n)$$

will evidently be

$$A = \frac{1}{n}(a_1 + a_2 + \cdots + a_n).$$

Moreover, because of Chebyshev's inequality, *(11.1)*, for any positive value of α

$$P(|\xi - A| > \alpha) \leq \frac{Q_\xi^2}{\alpha^2}. \tag{11.3}$$

The whole problem centers about finding the value of Q_ξ^2, whose evaluation is almost as easy as it was in the simplest case considered above. The quantity Q_ξ^2 is the variance of the magnitude ξ, which consists of the sum of n mutually independent magnitudes divided by n (it should be noted that the hypothesis of mutual independence is retained). Applying the rule for addition of variances, we obtain

$$Q_\xi^2 = \frac{1}{n^2}(q_1^2 + q_2^2 + \cdots + q_n^2),$$

in which $q_1, q_2, \cdots$ stand for the corresponding standard deviations of the variables $x_1, x_2, \cdots$ We may now assume that, in general, these standard deviations also differ among themselves. However, we shall also assume that no matter how many magnitudes we measure, their standard deviation will never exceed a certain positive quantity b. In practice this condition is always satisfied, since we add more or less similar quantities whose degree of spread is not very great. Thus we assume that $q_i < b$ $(i = 1, 2, \cdots)$, which makes the last equation assume the form

$$Q_\xi^2 < \frac{1}{n^2} nb^2 = \frac{b^2}{n}.$$

We conclude, from this relation and inequality *(11.3)*,

$$P(|\xi - A| > \alpha) < \frac{b^2}{n\alpha^2}.$$

No matter how small the value of α, with a sufficiently large number n of random variables the right side of this inequality may be made small without limit. This proves the law of large numbers in the general case under consideration.

If the variables $x_1, x_2, \cdots$ are mutually independent and their standard deviations are all smaller than a fixed positive number, then for sufficiently large n the arithmetic mean

$$\xi = \frac{1}{n}(x_1 + x_2 + \cdots + x_n),$$

with probability as close to unity as desired, differs in absolute value by as little as desired from its mean value. This is a statement of the law of large numbers in its general form, which is attributed to Chebyshev.

We shall now turn our attention to a special circumstance. Suppose that a person is making measurements of a certain quantity a. Repetition of the measurement under identical circumstances yields the values $x_1, x_2, \cdots, x_n$, which are not entirely identical. Now in order to approximate the value a he takes the arithmetic mean

$$a \approx \frac{1}{n}(x_1 + x_2 + \cdots + x_n).$$

Can we be assured of obtaining a value for a with any desired accuracy by using a sufficiently large number of trials?

We can, providing the measurements do not involve some systematic error, that is, providing

$$\bar{x}_k = a \text{ (with } k = 1, 2, \cdots, n),$$

and providing the measured quantities themselves are not indeterminate. For example, a systematic error may occur if the recording instrument fails to read the true value of the measured quantity on its scale. Now if the construction of the instrument is such that it cannot give an answer with an accuracy greater than a certain quantity δ (because the widths of the scale divisions are equal to δ, for example) we cannot obtain an accuracy of better than $\pm\delta$. Clearly, the arithmetic mean will also show a certain indeterminacy δ, which is the same as that for each x_k.

From these remarks it follows that if an apparatus gives an answer for a measurement with a certain indeterminancy δ, all attempts to obtain a more accurate estimate of a using the law of large numbers are doomed to failure, and any computations carried out in such an attempt are a waste of time.

Chapter 12

The Normal Laws

§ 29. Statement of the Problem

We have seen that a notable number of natural phenomena, and also man-made events such as military operations, involve random variables in a most basic way. Often before such a phenomena, process, or operation is completed our only knowledge of the random variables involved will be their distribution laws; that is, we may only have tables showing values of the variables and the associated probabilities for these values. If a variable can take on an indefinitely large number of discernibly different values (as it can for the flight distance of a shell, the error of a measurement, and so on), it is preferable to give probabilities not for specific values of the variable but for ranges of values (for example, the probability that a shell will travel between 720 and 740 meters, between 740 and 760 meters, and so on). This approach does not alter the essence of the matter. In order to make use of a random variable for the measurement of particular quantities of interest we must obtain the most accurate representation of its distribution law possible.

We might try to obtain the distribution laws for each random variable individually, declining to draw any conclusions or make any guesses of a general nature. In this case we would approach each specific random variable without any preconceptions and try to deduce all the features of its distribution law by a purely experimental method. However, this would place before us a task of nearly insurmountable difficulty. In every new problem we would have to carry out a large number of trials just to obtain the most essential features of the new and unfamiliar distribution law. For this reason scientists have for some time attempted to find general types of distribution laws which could reasonably be assumed to cover, if not all, at least broad classes of the random variables encountered in practice. Certain of these types of laws were established theoretically long ago and then confirmed by experiment. It is easy to see the value in being able to predict the type of distribution law that a newly encountered random variable will follow on the basis of theoretical considerations and past experience. When such a guess proves to be correct we usually need have only the results of a very modest number of experiments or observations in order to establish all the necessary features of the corresponding distribution law.

Theoretical investigations have shown that in a large number of practical instances we have adequate justification for assuming that certain characteristic frequency distributions will apply. Such distributions are called *normal laws*. We shall consider these laws briefly in this chapter, without proofs or precise formulations, in order to avoid inappropriate complexity.

Among the random variables met in practice, a good many characteristically involve "random mistakes" or "random errors," or, at any rate, are easily analyzed in such terms. Suppose, for example, that we are studying the flight distance x of a shell fired from a certain artillery piece. We naturally assume that there is a certain normal, average distance x_0, on whose basis we set the aiming devices. The difference $x - x_0$

then represents a "mistake" or "error" in the distance, and a study of the random variable x becomes a study of the "random error" $x - x_0$.

However, such an error, whose magnitude varies from one firing to another, usually depends on a great many causes which act independent of each other. There may be accidental vibrations of the gun barrel, unavoidable (though perhaps small) fluctuations in the weight and shape of the shell, chance variations in atmospheric pressure (which alter the resistance of the air), accidental mistakes in settings (as when a setting is made anew before each shot or before a small series of firings). All these factors and many others may produce errors in the flight distance. Such particular errors represent mutually independent random variables, and, moreover, *the effect of any one of them is only a minute portion of the entire effect.* Now the final error $x - x_0$, that we are considering, is simply the sum of all such random variations resulting from different causes. Thus, in the above example, the error of interest is the sum of a large number of mutually independent random variables. Clearly, similar circumstances are present for most random errors encountered in practice.

Theoretical considerations that cannot be presented here show that the distribution law for any random variable which can be considered as the sum of a large number of mutually independent random variables will correspond closely to a certain specific distribution law, regardless of the nature of the component effects, *provided only that each individual component is small in respect to the whole sum.* The law in question is specifically the "*normal law*" (some additional comments on which are given on p. 118).

Thus we can state that a substantial portion of all random variables encountered in practice (all those involving errors made up of a large number of mutually independent errors) approximate a normal distribution. We must now acquaint ourselves with the basic characteristics of normal distributions.

§ 30. The Concept of a Distribution Curve

In §15 we had occasion to present a distribution law in graphical form. This is a very useful technique, since it allows us to see at a glance, without studying tables, the important features of the distribution law under consideration. Such diagrams are made as follows: Various possible values of the random variable under consideration are laid off on the horizontal axis, making use of a zero center point with positive values on the right and negative values on the left (see Figure 11). The probability for each possible value is indicated on the vertical axis above the corresponding point on the horizontal axis. Scales are selected for both axes in such a way that the entire diagram will be of convenient size and shape. One glance at Figure 11 convinces us that the random variable described has a most probable value x_5 (a negative value) and that the probability drops off continuously (and quite sharply)

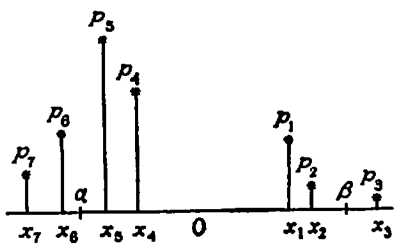

Figure 11

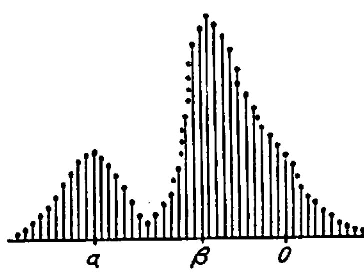

Figure 12

as we move away from x_5 on the horizontal axis. The probability that the variable will have a value lying within a certain interval (α, β) is equal to the sum of the probabilities for all the possible values within this interval, as we know from the rule for addition of probabilities. This would be shown graphically by summing the lengths of all the vertical lines corresponding to the values in the interval. In Figure 11, $P(\alpha < x < \beta) = p_1 + p_2 + p_4 + p_5$.

If, as often happens in practice, the number of possible values is very great, the use of an inconveniently wide horizontal axis may be avoided by choosing a very fine scale. This causes the possible values to be densely spaced (Figure 12), and the upper margin of the vertical lines appears to fuse into a single continuous curve, which is called the *distribution curve* for the particular random variable. Here also, the probability for the inequality $\alpha < x < \beta$ is shown graphically as the sum of lengths of the vertical lines lying in the interval (α, β). Let us now assume that the distance between any two possible neighboring values is always equal to unity. This would obtain if possible values corresponded to a sequence of whole numbers, which can practically always be actually accomplished by choosing a sufficiently small scale unit. Then the length of each vertical line is numerically equal to the *area of the rectangle* made up of the vertical line and the unit distance to an adjacent line (Figure 13).

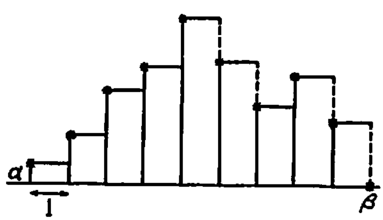

Figure 13

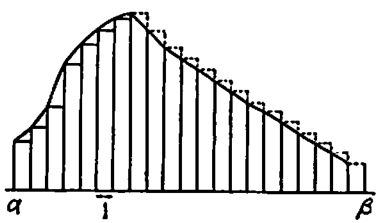

Figure 14

It follows that the probability for the inequality $\alpha < x < \beta$ may be represented graphically by the sum of the areas of all the rectangles situated above the interval. However, if the possible values are very close together, as in Figure 12, the sum of these rectangles will be nearly equal to the area of the curve bounded by the horizontal axis below, the distribution curve above, and the vertical lines above α and β on the sides (Figure 14). Here, as before, the distance between adjacent lines is one unit. Therefore, if we use a diagram such as that shown in Figure 14, the probability for obtaining a value of a certain random variable in any interval of values is simply and conveniently given by the area lying within this segment under the distribution curve. When the distribution law for a random variable is shown by a curve of this type we ordinarily dispense with the actual vertical lines, which are unnecessary and tend to obscure the diagram. In such a diagram the very question of the probability for individual values of the variable loses meaning. If a great many values are possible (and this assumption underlies all distribution curves), the probability for individual values will usually be minute (almost equal to zero) and therefore of no interest. For instance, when considering artillery fire it is not at all important to know the probability of the situation in which a shell strikes the ground precisely 473 centimeters from the center of the target. On the other hand, it is of value to know the likelihood of deviations within the range of 3–5 meters from the target. And it is the same for all similar situations: If the random variable can take on a great many different values, we are not interested in knowing the probabilities of these individual values, but rather the probabilities for a value falling in an entire interval. The probabilities for such intervals are clearly and directly found, as we have just discussed, from the area under the curve in the corresponding diagram.

§ 31. Properties of Normal Distribution Curves

A variable following a normal curve may always take on an infinite number of possible values. For this reason it is convenient to represent normal laws graphically as continuous curves.

Figure 15 shows several frequency distributions that follow a normal law. In spite of obvious differences in appearance, all the curves show certain common traits:

1) All of them have a single peak and fall-off continuously

on either side of the peak. This means that as the magnitude of the random variable diverges from its most probable value there is a corresponding decrease in probability.

2) All the curves are symmetrical with respect to the vertical axis drawn through the peak. This indicates that values of the random variable at equal distances on either side of the most probable value have equal probabilities.

3) All the curves are bell-shaped: close to the peak they are convex upward, and at some distance from it they change direction and are concave upward. The distance from the peak to this inflection point varies for specific curves. (For readers acquainted with the elements of higher mathematics we mention that the equation of the normal law has the form

$$y = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-a)^2}{2\sigma^2}},$$

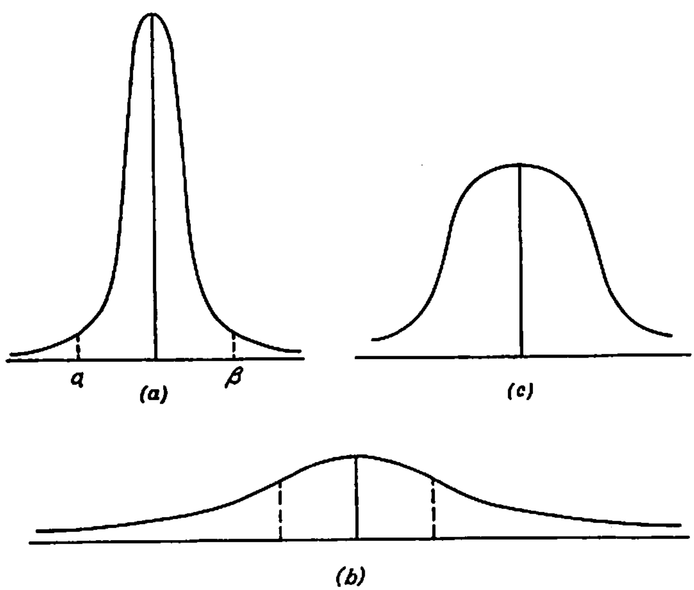

Figure 15

where $e = 2.71828 \cdots$, the base of natural logarithms, $\pi = 3.14159 \cdots$, the ratio of the circumference of a circle to its diameter, and a and σ^2 represent the mean value and variance of the random variable. Knowledge of the analytic form of the normal law may make it much easier for the reader to understand what follows in the book. However, the remaining material can also be understood by someone who has not studied higher mathematics.)

In what way, then, do different normal curves differ from each other? To answer this question, we must first recall that the total area below any frequency distribution curve must be equal to unity, since this area corresponds to the probability that the random variable will have some one of its possible values, which, of course, is a certainty. Therefore, various distribution curves differ from each other in that this total area is divided differently along the baseline. In regard to normal curves, such as those shown in Figure 15, the important matter is what portion of the total area is concentrated in segments that are close to the most probable value and what portion corresponds to segments which are less probable. For the curve shown in Figure 15(a), practically the whole area is concentrated very close to the most probable value. This means that in the vast majority of cases the random variable, and therefore the actual magnitude, will take on values close to the most probable value. Now, in view of the symmetry noted above, for a normal law the most probable value will always correspond to the mean value, and we can state that a random variable following the curve shown in Figure 15(a) has little spread; that is, it has a small variance and standard deviation.

On the other hand, in Figure 15(b) the area concentrated near the most probable value represents only a small portion of the total area. (This becomes apparent if we draw a segment (α, β) of identical length on the curves shown in Figure 15(a) and 15(b) and measure the area lying above each segment.) In this case it is quite likely, therefore, that the random variable will have a value which differs appreciably from the most probable one. This variable has a marked spread, and its variance and standard deviation are therefore great. Obviously, the curve shown in Figure 15(c) occupies an intermediate position between those shown in Figure 15(a) and 15(b).

In order to familiarize ourselves with the whole class of normal laws most quickly, so as to be able to apply them, we shall make use of two basic properties of these laws. We shall describe these properties in detail, but we shall not attempt to prove them since such a proof involves a precise definition of the normal law, and this calls for a knowledge of higher mathematics on the part of the reader.

Property 1. If the variable x follows the normal distribution law, then:

1) For any variable $c > 0$ and d the magnitude $cx + d$ also follows a normal distribution law;

2) Conversely, given a normal law we can find a (unique) pair of magnitudes $c > 0$ and d such that the quantity $cx + d$ will follow that normal law. Thus, if the random variable x follows a normal distribution law, the distribution law describing the variable $cx + d$ will also be a normal law for all possible values of $c > 0$ and d.

Property 2. If the random variables x and y are mutually independent and follow a normal distribution law, their sum $z = x + y$ also follows a normal distribution law.

Having accepted these two basic properties without proof, we may derive rigorously a series of properties of normal laws that are of particular importance in practice.

I. *For any two numbers a and $q > 0$ there exists a unique normal distribution with a mean value of a and a standard deviation q.*

Suppose, for instance, that x is a random variable distributed by a normal law, with the mean value $\bar{x}$ and standard deviation Q_x. Using property 1, our assertion will be proved if we can show that there exists a unique pair of numbers $c > 0$ and d

which satisfy the requirement that the variable $cx + d$ has the mean value a and the standard deviation q. If the table for variable x is as follows,

Table I

x_1	x_2	$\dots$	x_n
p_1	p_2	$\dots$	p_n

then the variable $cx + d$ (where $c > 0$ and d is any constant) will correspond to the table.

$cx_1 + d$	$cx_2 + d$	$\dots$	$cx_n + d$
p_1	p_2	$\dots$	p_n

Evidently,

$$\sum_k x_k p_k = \bar{x}, \sum_k (x_k - \bar{x})^2 p_k = Q_x^2.*$$

Our requirements may be reduced to the two conditions:

$$\sum_k (cx_k + d)p_k = a \text{ and } \sum_k (cx_k + d - a)^2 p_k = q^2.$$

The first yields

$$c\sum_k x_k p_k + d\sum_k p_k = a$$

or

$$c\bar{x} + d = a, \tag{12.1}$$

and the second yields

$$\sum_k (cx_k + d - c\bar{x} - d)^2 p_k = c^2 \sum_k (x_k - \bar{x})^2 p_k = c^2 Q_x^2 = q^2,$$

* The symbol $\sum_k$ is an abbreviation for $\sum_{k=1}^{n}$.

from which (since $c > 0$)

$$c = \frac{q}{Q_x}. \tag{12.2}$$

Therefore, we find from *(12.1)*,

$$d = a - c\bar{x} = a - \frac{q\bar{x}}{Q_x}. \tag{12.3}$$

Thus, given the numbers a and q it is always possible to find the values c and d from *(12.2)* and *(12.3)*, and moreover in a unique way. The variable $cx + d$ follows a normal distribution with the mean a and standard deviation q. Our assertion is therefore fully proved.

If we do not confine ourselves to normal laws, but consider instead all possible distribution laws, specification of the mean and standard deviation for a random variable provides us with relatively little information about the distribution, since there exist a great many distribution curves, all different, that possess the same mean and variance. In general, stating the mean and variance provides only an approximate characterization of the distribution law for a particular random variable.

However, the situation is otherwise if we confine ourselves to dealing with normal distributions only. As just demonstrated, any choice of a mean value and variance for a random variable is compatible with its fitting a normal curve. On the other hand, and this is a most important point, if we have some basis for supposing in advance that a certain variable follows one of the normal laws, then by stating its mean and variance we uniquely and fully determine its applicable normal distribution, which means the random variable is fully characterized. In particular, knowing the mean value and variance of such a variable, we may compute the probability that the variable will actually take on any arbitrary magnitude within its range.

II. *The relationship of the mean and probable deviation to the standard deviation is the same for all normal distributions.*

Suppose we are dealing with any two normal distributions, and suppose that x is a random variable following the first distribution. Owing to basic property *(12.1)* there exist constants $c > 0$ and d such that the magnitude $cx + d$ will follow the second distribution. Let Q_x and E_x stand for the standard deviation and mean (or probable) deviation of the first variable, and let q and e stand for the same parameters for the second. From the equations for the probable deviation we obtain

$$P\{|(cx + d) - (c\bar{x} + d)| < e\} = \frac{1}{2},$$

or

$$P\{c|x - \bar{x}| < e\} = \frac{1}{2},$$

and, finally,

$$P\left(|x - \bar{x}| < \frac{e}{c}\right) = \frac{1}{2}.$$

From this, using the equation for the probable deviation once again, we find that e/c is the probable deviation of the variable x; that is,

$$\frac{e}{c} = E_x,$$

or

$$\frac{e}{E_x} = c.$$

Equation *(12.2)* shows that

$$\frac{e}{E_x} = \frac{q}{Q_x},$$

or

$$\frac{e}{q} = \frac{E_x}{Q_x};$$

that is, the relationship of the probable deviation to the standard deviation is identical for the two distributions.

Since the above two laws were assumed to be any two arbitrarily chosen normal laws, our original assertion is proved.

The ratio e/q is thus an absolute constant; we shall designate it by λ. Calculations show that

$$\lambda = \sqrt{\frac{2}{\pi}} \approx 0.674.$$

It follows that for any normal distribution

$$e = \sqrt{\frac{2}{\pi}} \cdot q.$$

In view of this exceedingly simple relationship between the quantities e and q, for variables that follow a normal distribution, it is of no practical importance which of two scattering parameters we use in a given situation. We observed above that in general (that is, not only when considering quantities which follow normal distributions), the standard deviation possesses a whole series of simple properties that are not applicable to other similar measures and which dictate the choice of this parameter specifically as a measure of variance from both the theoretical and practical viewpoint. Nonetheless, we also noted that artillery officers almost always use the mean deviation. We see now why this tradition does not cause any harm: random variables normally encountered in artillery science almost always follow normal distributions, and, because of the proportionality mentioned above, the choice of one parameter or another of variance is almost insignificant.

III. *Let x and y be two mutually independent random variables which follow normal laws, and let $z = x + y$. It follows that*

$$E_z = \sqrt{E_x^2 + E_y^2},$$

where E_x, E_y and E_z refer respectively to the probabilities of deviation for the variables x, y, and z.

As discussed in §25, an analogous formula for the standard deviation is applicable regardless of the distribution law that the variables x and y follow. Should they happen to follow a normal law, then the magnitude z also follows a normal law, as stated in fundamental property 2 (p. 118). Therefore, because of the preceding property II,

$$E_x = \lambda Q_x, \qquad E_y = \lambda Q_y, \qquad E_z = \lambda Q_z,$$

and, therefore,

$$E_z = \lambda \sqrt{Q_x^2 + Q_y^2} = \sqrt{(\lambda Q_x)^2 + (\lambda Q_y)^2} = \sqrt{E_x^2 + E_y^2}.$$

With normal distributions one of the most important properties of standard deviations is carried over directly to the probable (and mean) deviations.

§32. Illustrative Problems

Let us agree to define a *basic normal distribution law* as one for which the mean value is zero and the variance is unity. If x is a random variable following the basic normal distribution, we can, for the sake of brevity, write

$$P\{|x| < a\} = \Phi(a)$$

for any positive value of a. Thus, $\Phi(a)$ is the probability that the variable x, which follows the basic normal distribution, will not have an absolute value in excess of a. A very accurate table has been compiled for the quantity $\Phi(a)$, which gives values for various a's. It is an indispensible aid for anyone who deals with probability computations, and it is found in every book dealing with probability theory (see p. 138). With a table of the function $\Phi(a)$ in hand we may easily and accurately carry out necessary calculations for any magnitude that follows a normal distribution law. We shall now use examples to show how this is done.

Problem I. A random variable follows a normal distribution law with a mean of $\bar{x}$ and a standard deviation Q_x. Find the probability that the absolute value of the deviation $x - \bar{x}$ will not exceed a certain quantity a.

Let z be a random variable distributed in accordance with the basic normal law. According to fundamental property 1 (see p. 118), we can find quantities $c > 0$ and d such that the sum $cz + d$ will have the mean $\bar{x}$ and standard deviation Q_x, that is, such that it follows the same normal law as the given variable x. Therefore,

$$\begin{aligned} P(|x - \bar{x}| < a) &= P|(cz + d) - (c\bar{z} + d)| < a) \\ &= P(c|z - \bar{z}| < a). \end{aligned}$$

However, because of *(12.2)* we note that $c = Q_x/Q_z = Q_x$, because $Q_z = 1$ (for the basic normal distribution the variance is equal to unity). It follows that

$$\begin{aligned} P(|x - \bar{x}| < a) &= P(Q_x|z - \bar{z}| < a) \\ &= P\left(|z| < \frac{a}{Q_x}\right) \\ &= \Phi\left(\frac{a}{Q_x}\right). \end{aligned} \qquad (12.4)$$

This solves the problem, since the value of $\Phi(a/Q_x)$ may be found directly from the table. Using the table and *(12.4)*, we may easily calculate the probability for any maximal deviation of a variable that follows any normal distribution.

Example 1. A given mechanical part is made on a lathe. Its length represents a random variable, that follows a normal law and has a mean value of 20 centimeters with a variance of 0.2 centimeter. Find the probability that the length of the part will be between 19.7 and 20.3 centimeters, that is, that the deviation in either direction will not exceed 0.3 centimeters.

Because of *(12.4)* and the table,

$$P\{|x - 20| < 0.3\} = \Phi\left(\frac{0.3}{0.2}\right) = \Phi(1.5) = 0.866.$$

Hence, about 87% of all parts manufactured under the cited conditions will have a length of from 19.7 to 20.3 centimeters, and the remaining 13% will show a larger deviation from the mean.

Example 2. Using the conditions in Example 1, state what accuracy of part size may be guaranteed with a probability of 0.95.

Evidently the problem consists in finding such a positive value a for which

$$P\{|x - 20| < a\} > 0.95.$$

The computation in Example 1 demonstrated that $a = 0.3$ is inadequate for such a guarantee, since the left portion of the above inequality will be less than 0.87. From *(12.4)* we find

$$P\{|x - 20| < a\} = \Phi\left(\frac{a}{0.2}\right) = \Phi(5a).$$

We must first find in the table a value for $5a$ for which

$$\Phi(5a) > 0.95.$$

Such will be the case with

$$5a > 1.97,$$

and therefore a must be greater than 0.394. Thus we may guarantee that the deviation will not exceed 0.4 centimeters in 95 cases out of 100.

Example 3. In certain practical situations it is frequently accepted that for accurate results a random variable x, which follows a normal distribution, should not show a deviation greater than three times the standard deviation Q_x. What is the justification for this viewpoint?

Equation *(12.4)* and the table show that

$$P\{|x - \bar{x}| < 3Q_x\} = \Phi(3) > 0.997,$$

and therefore

$$P\{|x - \bar{x}| > 3Q_x\} < 0.003.$$

Practically speaking, this means that a deviation whose absolute value exceeds three times Q_x will be encountered, on the average, less than three times in a 1000 trials. Whether we can disregard such a probability is a decision that must be made on the basis of the problem at hand and cannot be settled dogmatically.

Let us note that the relationship $P\{|x - \bar{x}| < 3Q_x\} = \Phi(3)$ is evidently nothing more than a particular case of the equation

$$P\{|x - \bar{x}| < aQ_x\} = \Phi(a), \tag{12.5}$$

which follows from *(12.4)* and applies for any random variable x that is normally distributed.

Example 4. For an artillery shell weighing 8.4 kilograms it is found that the absolute value of the deviation exceeds 50 grams an average of three times in a hundred. Assuming that the weight of shells follows a normal distribution law, find the magnitude of the probable deviation.

We are given that

$$P(|x - 8.4| > 0.05) = 0.03,$$

where x is the weight of a randomly chosen shell. It follows that

$$0.97 = P(|x - 8.4| < 0.05) = \Phi\left(\frac{0.05}{Q_x}\right).$$

The table shows that $\Phi(a) = 0.97$ for $a \approx 2.12$. Therefore,

$$\frac{0.05}{Q_x} \approx 2.12,$$

from which we find

$$Q_x \approx \frac{0.05}{2.12}.$$

The probable deviation, as we know, is equal to

$$E_x = 0.674 Q_x \approx 0.0155 \text{ kilograms} = 15.5 \text{ grams}.$$

Example 5. In the firing of an artillery piece the deviation of the shell from the target is due to three independent causes: (1) errors in determining the true position of the target; (2) errors in aiming; and (3) errors that vary from shot to shot, such as weight of the shell, atmospheric conditions, and so on. Assuming that all three types of errors follow a normal distribution law with a mean value of zero and probable deviations equal respectively to 24 meters, 8 meters, and 12 meters, find the probability that the total deviation from the target will be 40 meters.

Since the probable deviation of the total error x is given by property III (p. 122) as

$$\sqrt{24^2 + 8^2 + 12^2} = 28 \text{ meters},$$

the standard deviation of the total error is

$$\frac{28}{0.674} = 41.5,$$

and therefore

$$P(|x| < 40) = \Phi\left(\frac{40}{41.5}\right) \approx \Phi(0.964) = 0.665.$$

Deviations not exceeding 40 meters will therefore be encountered in about $\frac{2}{3}$ of all firings.

Problem II. The random variable x is distributed in accordance with a normal law whose mean value is $\bar{x}$ and whose standard deviation is Q_x. Find the probability that the absolute value of the deviation $x - \bar{x}$ will lie between the magnitudes a and b $(0 < a < b)$.

Applying the addition rule, we find

$$P(|x - \bar{x}| < b) = P(|x - \bar{x}| < a) + P(a < |x - \bar{x}| < b),$$

and therefore

$$P(a < |x - \bar{x}| < b) = P(|x - \bar{x}| < b) - P(|x - \bar{x}| < a)$$
$$= \Phi\left(\frac{b}{Q_x}\right) - \Phi\left(\frac{a}{Q_x}\right), \qquad (12.6)$$

which is the solution to the problem.

For the majority of practical problems the table of values for the function $\Phi(a)$ is a very strong tool. Often we need only compute the probability for the quantity $|x - \bar{x}|$ falling into a fairly wide range; therefore, in practice, in addition to having the "full" table on hand it is also worthwhile to use a condensed version, which may be easily made up from the full table by applying *(12.6)*.

We illustrate how we may construct such a short table, which will be much cruder than the full table at the back of the book, but which will be quite adequate for many applications. Let us divide the range of values of $|x - \bar{x}|$ into five segments: (1) from zero to 0.32 Q_x; (2) from 0.32 Q_x to 0.69 Q_x; (3) from 0.69 Q_x to 1.15 Q_x; (4) from 1.15 Q_x to 2.58 Q_x; (5) greater than 2.58 Q_x.

Applying *(12.4)*, we see that

$$P(|x - \bar{x}| < 0.32\ Q_x) = \Phi(0.32) \approx 0.25;$$
$$P(0.32\ Q_x < |x - \bar{x}| < 0.69\ Q_x) = \Phi(0.69) - \Phi(0.32) \approx 0.25;$$
$$P(0.69\ Q_x < |x - \bar{x}| < 1.15\ Q_x) = \Phi(1.15) - \Phi(0.69) \approx 0.25;$$
$$P(1.15\ Q_x < |x - \bar{x}| < 2.58\ Q_x) = \Phi(2.58) - \Phi(1.15) \approx 0.24;$$
$$P(|x - \bar{x}| > 2.58\ Q_x) = 1 - \Phi(2.58) \approx 0.01.$$

The results of these computations can be shown on a chart such as that in Figure 16.

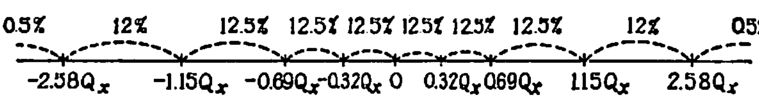

Figure 16

In this figure an infinitely long line is divided into ten segments—five positive and five negative. Above each segment is given the percentage of situations in which actual deviations will fall within that range. For example, using the above computations, we find that about 25% of all deviations must lie in the range $-1.15\ Q_x$ to $-0.69\ Q_x$ and $+0.69\ Q_x$ to $+1.15\ Q_x$, taken together. Because of the symmetry in the normal law, deviations will occur approximately equally often in both directions, so positive and negative deviations will each amount to about 12.5% of the total. Having at hand this or an even simpler scheme, we may immediately get an idea of the basic aspects of the deviations for a given random variable that follows a normal distribution and has an arbitrary mean and standard deviation.

Finally, let us consider the calculation of a probability that a random variable obeying a normal law will fall into a certain arbitrarily chosen range of values.

Problem III. Knowing that the random variable x follows a normal distribution (with a mean of $\bar{x}$ and a standard deviation Q_x, use the table to compute the probability applicable to the inequality $a < x < b$, where a and b $(a < b)$ are two arbitrary numbers.

We shall have to consider three cases, which are determined by the magnitudes of a and b relative to $\bar{x}$.

First Case: $\bar{x} \leq a \leq b$.

Using the addition rule, we find

$$P(\bar{x} < x < b) = P(\bar{x} < x < a) + P(a < x < b).$$

From this we obtain

$$\begin{aligned} P(a < x < b) &= P(\bar{x} < x < b) - P(\bar{x} < x < a) \\ &= P(0 < x - \bar{x} < b - \bar{x}) \\ &\qquad - P(0 < x - \bar{x} < a - \bar{x}). \end{aligned}$$

But with any $\alpha > 0$ the symmetry of the normal law indicates

$$
\begin{aligned}
P(0 < x - \bar{x} < \alpha) &= P(-\alpha < x - \bar{x} < 0) \\
&= \frac{1}{2} P(-\alpha < x - \bar{x} < \alpha) \\
&= \frac{1}{2} P(|x - \bar{x}| < \alpha) \\
&= \frac{1}{2} \Phi\left(\frac{\alpha}{Q_x}\right),
\end{aligned}
\tag{12.7}
$$

and therefore

$$
P(a < x < b) = \frac{1}{2}\left[\Phi\left(\frac{b - \bar{x}}{Q_x}\right) - \Phi\left(\frac{a - \bar{x}}{Q_x}\right)\right].
$$

Second Case: $a \leq \bar{x} \leq b$.

We find

$$
\begin{aligned}
P(a < x < b) &= P(a < x < \bar{x}) + P(\bar{x} < x < b) \\
&= P(a - \bar{x} < x - \bar{x} < 0) \\
&\qquad + P(0 < x - \bar{x} < b - \bar{x}) \\
&= \frac{1}{2}\Bigg[\Phi\left(\frac{\bar{x} - a}{Q_x}\right) + \Phi\left(\frac{b - \bar{x}}{Q_x}\right),
\end{aligned}
$$

which follows from (12.7).

Third Case: $a \leq b \leq \bar{x}$.

We find

$$
P(a < x < \bar{x}) = P(a < x < b) + P(b < x < \bar{x}),
$$

from which we obtain

$$
\begin{aligned}
P(a < \bar{x} < b) &= P(a < x < \bar{x}) - P(b < x < \bar{x}) \\
&= P(a - \bar{x} < x - \bar{x} < 0) \\
&\qquad - P(b - \bar{x} < x - \bar{x} < 0) \\
&= \frac{1}{2}\left[\Phi\left(\frac{\bar{x} - a}{Q_x}\right) - \Phi\left(\frac{\bar{x} - b}{Q_x}\right)\right].
\end{aligned}
$$

We have solved the problem for all three cases. We see that for a random variable following some normal distribution law our table makes it possible to find the probability of the variable falling into any range or segment of values and thus is a complete characterization of its distribution law.

In order to learn how such computations are carried out in practice let us consider the following example.

Example. Firing is carried out from point O along a line OX. The mean flight distance of a shell is 1200 meters. Assuming that this distance will obey a normal distribution with a standard deviation of 40 meters, find the percentage of shells that will overshoot the target by 60–80 meters.

In order for a shell to overshoot in this manner we must have $1260 < H < 1280$. Applying the concluding equation of the first case in Problem III, we find

$$P(1260 < H < 1280) = \frac{1}{2}\Phi\left[\left(\frac{1280 - 1200}{40}\right) - \Phi\left(\frac{1260 - 1200}{40}\right)\right]$$

$$= \frac{1}{2}\left[\Phi(2) - \Phi(1.5);\right]$$

from the table we obtain

$$\Phi(2) \approx 0.955, \qquad \Phi(1.5) \approx 0.866,$$

from which we obtain $P(1260 < H < 1280) \approx 0.044$. Thus, this degree of overshoot will occur in about 4% of firings.

Conclusions

During the last ten years probability theory has become one of the most rapidly developing branches of mathematical science. New theoretical findings open up added possibilities for the scientific use of probability techniques. More refined, more detailed studies of natural phenomena stimulate probability theory experts to seek new methods and new laws that will fit the data. Probability theory is a mathematical science that is not aloof from practical life and the demands of other sciences; it proceeds apace with the general progress of natural science and technology. But the reader must not erroneously assume that probability theory is nothing but a handmaiden or practical tool for the solution of practical problems. It certainly is not, and during the last 30 years this theory has grown into a substantial mathematical science with its own problems and methods of investigation. At the same time the most important and pressing problems of probability theory are applicable to the solution of pressing problems in natural science.

The origins of probability theory date to the middle of the seventeenth century and are associated with the names of Fermat (1601–1665), Pascal (1623–1662), and Huygens (1625–1695). In the works of these scientists we discover rudimentary forms of concepts such as the probability of a chance event and the mathematical expectation of a random magnitude. The starting point for their studies were problems arising in connection with games of chance. However, they clearly understood the importance of their new concepts for the study of nature and, for example, in his work "Calculations in Games of Chance," Huygens wrote "the reader will observe that we are dealing not only with games but rather with the foundations of a new and very interesting theory." Among later scientists who exerted a major influence on the development of probability theory were Bernoulli (1654–1705), whose name we have already learned in this text, DeMoivre (1667–1754), Bayes (died, 1763), Laplace (1749–1827), Gauss (1777–1855), and Poisson (1781–1840).

The subsequent powerful growth of probability theory is closely linked with the traditions and achievements of Russian science. During the last century probability theory reached a dead end in the Western world. The Russian master mathematician P. L. Chebyshev found a new technique for developing the theory—a multisided study of the sequential properties of independent random variables. Chebyshev, and his students A. A. Markov and A. M. Lyapounov, obtained fundamental results along these lines (the law of large numbers, Lyapounov's theorem).

The reader is already familiar with the law of large numbers, and our immediate goal will now be to give some idea of another important addition to the theory of probability which has been given the name of Lyapounov's theorem (otherwise known as the Central Limit Theorem of probability theory).

The reason for the great importance of the latter is that a notable number of phenomena whose outcome depends on chance follow a course that is basically as follows: the phenomenon under study is subject to the action of a very large number of independently acting random causes, each of which exerts only a very slight effect on the course of the action as a whole. The effect of each of these causes may be attributed to the random variables $\xi_1, \xi_2, \cdots, \xi_n$, and their total effect is equal to their sum $s_n = \xi_1 + \xi_2 + \cdots + \xi_n$. Since determination of the size of each of these effects (in other words, finding the distribution functions for the variables ξ_i), or even a simple listing of the latter, is practically impossible, it is very important to develop a method that elucidates their combined total effects, regardless of the nature of each individual component.

The usual methods of approach are powerless before this problem, and in their place we must find another technique, in which the large number of factors acting on the final effect will not be an obstacle but rather an aid for solving the problem at hand. Such a technique must compensate for our inadequate knowledge of each separately acting cause by making use of their large numbers. The Central Limit Theorem, which was established in the main through the work of academician P. L. Chebyshev (1821–1894), A. A. Markov (1856–1922), and A. M. Lyapounov (1857–1918) asserts that if the effective causes $\xi_1, \xi_2, \cdots, \xi_n$ are mutually independent and are present in very large number n, which means that the effect of any one individual cause will be slight in respect to the total effect, then the distribution law followed by the sum s_n will differ only insignificantly from the normal distribution law.

Let us consider examples of some phenomena that follow this outline.

During artillery fire unavoidable deviations of the hitting points of shells cause the well-known phenomenon of scatter about the target. Since the scatter is the result of a large number of independent effects (for example, inaccuracy in the barrel alignment, unbalanced warheads, variations in properties of the metals and chemicals making up the shells, small errors in aiming which are hardly noticeable to the eye, insignificant variations in atmospheric parameters during successive shots, and many others), each of which has only a minute influence on the shell trajectory, it follows from Lyapounov's theorem that the scatter will obey a normal law. This circumstance is noted in analyses of artillery firing and is used as the basis for related calculations.

Whenever we carry out some observations in order to evaluate some physical constant the result of our measurement is unavoidably influenced by a great number of factors, each of which cannot be evaluated individually but does contribute to the error. This may involve inaccuracies in the measuring instrument, whose reading may change because of various atmospheric, thermal, mechanical, and other factors. We must also consider errors by the observer connected with peculiarities of his vision or hearing, or related to psychic tensions or fatigue. The actual error of measurements is thus the total result of a great number of minor, independent—we might say, elementary—errors that vary from situation to situation. As a result of Lyapounov's theorem, we may again assert that the observation errors will follow a normal distribution.

Any number of similar examples can be found: the position and velocity of gas molecules, as determined by a large number of collisions within the gas; the quantity of an accurately weighed substance; deviations of precise mechanical parts from required dimensions during mass production; the distribution of weights and heights of animals, plants, certain organs; and so on.

The emergence of physical statistics and also the requirements of various branches of technology have given probability theory a large number of completely new problems that do not fit into the framework of classical theory. Although physical scientists or technologists were interested in studying some *process*, that is, an event taking place in time, the probability theory possessed neither the general means nor perfected specific methods for solving the problems arising in the study of continuous processes. It became necessary to develop a general theory for *chance processes*, that is, a theory dealing with random variables or magnitudes that were *dependent on one or several continuously varying parameters.*

We will list several problems which bring to light random variables which show a change with the passage of time. Imagine that we have decided to trace the course of a certain molecule in a gas or liquid. At any given time the molecule may collide with other molecules, thereby changing its velocity and direction of motion. Such an alteration in state of the molecule is subject to the action of chance at each moment. The study of a whole series of physical phenomena requires precisely the ability to compute the probability that a certain number of molecules will traverse a given distance in a certain period of time. Thus, for instance, if two gases or two liquids are placed in contact, there commences a mutual interchange of molecules between the fluids, that is, diffusion takes place. How quickly does the diffusion process occur? What laws does it follow? When will the resulting mixture of gas become essentially uniform? All these questions are answered by statistical diffusion theory, at whose basis there lies computations of probabilities and the study of random or chance processes. Evidently, similar problems are encountered in chemistry when we study the processes of chemical reactions. What portion of the molecules have already entered into a reaction as time passes; when is the reaction essentially at an end?

A most important group of processes involves radioactive disintegration. This phenomena consists in the fact that atoms of a radioactive substance break down and turn into atoms of another element. Each disintegration occurs instantaneously, similar to an explosion with a certain release of energy. Numerous observations have established that atomic disintegration occurs as a chance process in time and is independent of other disintegrations (provided the mass of the radioactive substance is not too great). In studying the process of radioactive breakdown it is necessary to determine the probability that a given number of atoms will disintegrate in a certain time interval. This problem is a typical problem of random-process theory. Formally speaking, if we consider only the mathematical analysis of this phenomenon, the same equations apply also to other types of processes, for example, the number of calls received at a telephone switching center from subscribers, Brownian movement, the breakage of threads on a knitting machine.

The beginnings of the general theory of chance processes were laid down by the Soviet mathematicians A. N. Kolmogorov and A. Ya. Khinchin in the early thirties. Somewhat earlier, in the first decade of the present century, A. A. Markov began the investigation of sequences of dependent random variables, which have come to be known as Markov chains. During the twenties, physicists turned the theory which he developed as a purely mathematical discipline into a useful tool for studying nature. Since that time many scientists (S. N. Bernstein, V. I. Romanovsky, A. N. Kolmogorov, J. Hadamard, M. Fréchet, and others) have made important contributions to the theory of Markov chains.

In the twenties A. N. Kolmogorov, E. E. Slutsk,, A. Ya. Khinchin, and Paul Lévy demonstrated a close connection between probability theory and the mathematical disciplines concerned with the general concept of functions (theory of sets and theory of functions of a real variable). Somewhat earlier these same ideas had been conceived by E. Borel. The discovery of this connection proved to be most fruitful, and specifically from this viewpoint it became possible to find a final solution to the classical problems framed by Chebyshev.

Finally, we should note the works of S. N. Bernstein, A. N. Kolmogorov, and von Mises on the construction of a logical basis for probability theory which could encompass the various problems presented by the natural sciences and other fields of endeavor.

Soviet science occupies a distinguished position in the current energetic development of probability theory.

TABLE OF VALUES OF THE FUNCTION $\Phi(a)$

a	$\Phi(a)$	a	$\Phi(a)$	a	$\Phi(a)$	a	$\Phi(a)$	a	$\Phi(a)$
0.00	0.000	0.60	0.451	1.20	0.770	1.80	0.928	2.40	0.984
0.01	0.008	0.61	0.458	1.21	0.774	1.81	0.930	2.41	0.984
0.02	0.016	0.62	0.465	1.22	0.778	1.82	0.931	2.42	0.984
0.03	0.024	0.63	0.471	1.23	0.781	1.83	0.933	2.43	0.985
0.04	0.032	0.64	0.478	1.24	0.785	1.84	0.934	2.44	0.985
0.05	0.040	0.65	0.484	1.25	0.789	1.85	0.936	2.45	0.986
0.06	0.048	0.66	0.491	1.26	0.792	1.86	0.937	2.46	0.986
0.07	0.056	0.67	0.497	1.27	0.796	1.87	0.939	2.47	0.986
0.08	0.064	0.68	0.504	1.28	0.800	1.88	0.940	2.48	0.987
0.09	0.072	0.69	0.510	1.29	0.803	1.89	0.941	2.49	0.987
0.10	0.080	0.70	0.516	1.30	0.806	1.90	0.943	2.50	0.988
0.11	0.088	0.71	0.522	1.31	0.810	1.91	0.944	2.51	0.988
0.12	0.096	0.72	0.528	1.32	0.813	1.92	0.945	2.52	0.988
0.13	0.103	0.73	0.535	1.33	0.816	1.93	0.946	2.53	0.989
0.14	0.111	0.74	0.541	1.34	0.820	1.94	0.948	2.54	0.989
0.15	0.119	0.75	0.547	1.35	0.823	1.95	0.949	2.55	0.989
0.16	0.127	0.76	0.553	1.36	0.826	1.96	0.950	2.56	0.990
0.17	0.135	0.77	0.559	1.37	0.829	1.97	0.951	2.57	0.990
0.18	0.143	0.78	0.565	1.38	0.832	1.98	0.952	2.58	0.990
0.19	0.151	0.79	0.570	1.39	0.835	1.99	0.953	2.59	0.990
0.20	0.159	0.80	0.576	1.40	0.838	2.00	0.955	2.60	0.991
0.21	0.166	0.81	0.582	1.41	0.841	2.01	0.956	2.61	0.991
0.22	0.174	0.82	0.588	1.42	0.844	2.02	0.957	2.62	0.991
0.23	0.182	0.83	0.593	1.43	0.847	2.03	0.958	2.63	0.991
0.24	0.190	0.84	0.599	1.44	0.850	2.04	0.959	2.64	0.992
0.25	0.197	0.85	0.605	1.45	0.853	2.05	0.960	2.65	0.992
0.26	0.205	0.86	0.610	1.46	0.856	2.06	0.961	2.66	0.992
0.27	0.213	0.87	0.616	1.47	0.858	2.07	0.962	2.67	0.992
0.28	0.221	0.88	0.621	1.48	0.861	2.08	0.962	2.68	0.993
0.29	0.228	0.89	0.627	1.49	0.864	2.09	0.963	2.69	0.993
0.30	0.236	0.90	0.632	1.50	0.866	2.10	0.964	2.70	0.993
0.31	0.243	0.91	0.637	1.51	0.867	2.11	0.965	2.72	0.993
0.32	0.251	0.92	0.642	1.52	0.871	2.12	0.966	2.74	0.994
0.33	0.259	0.93	0.648	1.53	0.874	2.13	0.967	2.76	0.994
0.34	0.266	0.94	0.653	1.54	0.876	2.14	0.968	2.78	0.995
0.35	0.274	0.95	0.658	1.55	0.879	2.15	0.968		
0.36	0.281	0.96	0.663	1.56	0.881	2.16	0.969	2.80	0.995
0.37	0.289	0.97	0.668	1.57	0.884	2.17	0.970	2.82	0.995
0.38	0.296	0.98	0.673	1.58	0.886	2.18	0.971	2.84	0.995
0.39	0.303	0.99	0.678	1.59	0.888	2.19	0.971	2.86	0.996
								2.88	0.996
0.40	0.311	1.00	0.683	1.60	0.890	2.20	0.972		
0.41	0.318	1.01	0.688	1.61	0.893	2.21	0.973	2.90	0.996
0.42	0.326	1.02	0.692	1.62	0.895	2.22	0.974	2.92	0.996
0.43	0.333	1.03	0.697	1.63	0.897	2.23	0.974	2.94	0.997
0.44	0.340	1.04	0.702	1.64	0.899	2.24	0.975	2.96	0.997
0.45	0.347	1.05	0.706	1.65	0.901	2.25	0.976	2.98	0.997
0.46	0.354	1.06	0.711	1.66	0.903	2.26	0.976		
0.47	0.362	1.07	0.715	1.67	0.905	2.27	0.977	3.00	0.997
0.48	0.369	1.08	0.720	1.68	0.907	2.28	0.977	3.10	0.998
0.49	0.376	1.09	0.724	1.69	0.909	2.29	0.978	3.20	0.999
								3.30	0.999
0.50	0.383	1.10	0.729	1.70	0.911	2.30	0.979	3.40	0.999
0.51	0.390	1.11	0.733	1.71	0.913	2.31	0.979		
0.52	0.397	1.12	0.737	1.72	0.915	2.32	0.980	3.50	0.9995
0.53	0.404	1.13	0.742	1.73	0.916	2.33	0.980	3.60	0.9997
0.54	0.411	1.14	0.746	1.74	0.918	2.34	0.981	3.70	0.9998
0.55	0.418	1.15	0.750	1.75	0.920	2.35	0.981	3.80	0.99986
0.56	0.425	1.16	0.754	1.76	0.922	2.36	0.982	3.90	0.99990
0.57	0.431	1.17	0.758	1.77	0.923	2.37	0.982	4.00	0.99994
0.58	0.438	1.18	0.762	1.78	0.925	2.38	0.983		
0.59	0.445	1.19	0.766	1.79	0.927	2.39	0.983	5.00	0.9999994

INDEX