A. Ya. Khinchin
1894–1959

An ELEMENTARY INTRODUCTION *to the* THEORY *of* PROBABILITY

By

B. V. GNEDENKO

and

A. Ya. KHINCHIN

AUTHORIZED EDITION

Translated from the Fifth Russian Edition by

LEO F. BORON
The Pennsylvania State University

with the editorial collaboration of

SIDNEY F. MACK
The Pennsylvania State University

DOVER PUBLICATIONS, INC.
NEW YORK

FOREWORD TO THE FIFTH SOVIET EDITION

The present edition was prepared by me after the death of A. Ya. Khinchin, an eminent scientist and teacher. Many of the ideas and results in the modern development of the theory of probability are intimately connected with the name of Khinchin. The systematic utilization of the methods of set theory and the theory of functions of a real variable in the theory of probability, the construction of the foundations of the theory of stochastic processes, the extensive development of the theory of the summation of independent random variables, and also the construction of a new approach to the problems of statistical physics and the elegant system of its discussion—all this is due to Aleksandr Yakovlevich Khinchin. He shares with S. N. Bernshtein and A. N. Kolmogorov the honor of creating the Soviet school of probability theory, which plays an outstanding role in modern science. I consider myself fortunate to have been his student.

We wrote this booklet in the period of the victorious conclusion of the Great Patriotic War; this was naturally reflected in the elementary formulation of military problems which we used as examples. Now—fifteen years after the victory—in days when the entire country is covered with forests of new construction, it is natural to extend the subject matter in the examples to illustrate the general theoretical situation. It is for this reason therefore that, not changing the discussion and elementary character of the book, I have allowed myself the privilege of replacing a large number of examples by new ones. The same changes, with some negligible exceptions, were introduced by me also in the French edition of our booklet (Paris, 1960).

Moscow, October 6, 1960 B. V. Gnedenko

FOREWORD TO THE FIRST SOVIET EDITION

Acquaintance with the theoretical foundations of a mathematical science always enables one to apply more knowledgeably and actively the results of this science in practice. Likewise, in the area of probability theory, the situation is such that a large number of leaders (and occasionally also rank and file workers) in the military, in industry, agricultural economy, economy, etc., whose mathematical training is very limited, must deal with the practical applications of this science.

The present little book has as its aim to acquaint, in the most accessible form, the workers of this group with the fundamental concepts of probability theory and the methods of probability calculations. This booklet is completely accessible to all those who have completed the 10-year secondary school [ages 7–17 in the USSR]; it is almost entirely accessible to those who have completed the 7-year school also [ages 7–14 in the USSR]. In almost all its parts, the book is constructed on the basis of concrete, practical examples; in the choice of these examples, however, we were guided primarily not by their practical reality but by the illustrative value for the mastery of the corresponding theoretical situations.

Moscow, January 7, 1945

CONTENTS

PART I. PROBABILITIES

PART II. RANDOM VARIABLES

PART I

PROBABILITIES

CHAPTER I

THE PROBABILITY OF AN EVENT

§ 1. The concept of probability

When we say that under given conditions of firing a marksman has 92% success we mean that of 100 shots fired by him under certain well-defined conditions (e.g., the same target at a prescribed distance, the same firearm, and so on), there are approximately 92 successes (and hence about 8 failures) *on the average*. Of course, there will not be exactly 92 successful shots out of every 100; sometimes there will be 91 or 90 of them, sometimes there will be 93 or 94; at times the number of successes can even be noticeably less or noticeably greater than 92; but *on the average* after many repetitions of shots under the same conditions, this percentage of target hits will remain unchanged as long as with the passage of time no essential changes take place in the firing conditions. (Otherwise, for example, our marksman could increase his mastery, and thereby increase the average percentage of target hits to 95 or higher.) And experience shows that for such a marksman, the number of successful shots per hundred will be close to 92; those hundreds, in which, for example, this number is less than 88 or greater than 96, although these will be encountered, will occur comparatively rarely. The figure 92% which serves as an index of mastery of our marksman is usually very *stable*; i.e., the percentage of target hits in the majority of shots (under the same conditions) will be almost the same for a given marksman—deviating rather significantly from its average value only in rare, exceptional cases.

Let us consider still another example. It is observed in a certain factory that under given conditions on the average 1.6% of the manufactured articles do not satisfy the standard and are rejected. This means that in a collection, say, of 1000 articles which have not yet been subjected to inspection, there will be approximately 16 which are unusable. Sometimes, of course, the number of rejected articles will be somewhat greater, sometimes somewhat less, but on the average this number will be close to 16, and in the majority of collections of 1000 articles it will also be close to 16. It is understood that here also we assume that the conditions of production are invariant; i.e., the organization of the technological process, equipment, raw materials, qualification of workers, and so on, remain the same.

Clearly, one could introduce any number of such examples. In all these cases, we see that in *homogeneous, numerous* operations performed under prescribed conditions (repeated firings, the mass production of articles, and so on), the percentage of a certain type of event which is important to us (hitting the target, the fact that articles do not meet a fixed standard, and so on) will almost always remain approximately unchanged, only in rare cases deviating somewhat significantly from some average figure. One can therefore say that this average figure is a characteristic index of the given operation (under prescribed, strictly established conditions). The percentage of target hits describes for us the mastery of the marksman, the percentage of rejects gives us an estimate of how much of the production is of good quality. It is therefore self-evident that the knowledge of such indices is very important in the most diverse areas: in military operations, technology, economy, physics, chemistry, and other fields; for it enables us not only to estimate the outcome of mass phenomena which have already occurred but also to foresee the outcome of a mass operation in the future.

If, under given firing conditions, a marksman hits the target on the average 92 times out of 100 shots, we say that for this marksman and under these conditions the *probability of hitting the target* is 92% (or 92/100 or 0.92). If, under given conditions, on the average of every 1000 finished articles in a certain factory there are 16 rejects, then we say that the *probability of manufacturing a reject* is 0.016 or 1.6% for the given production.

But in general what do we call the probability of an event in a given mass operation? It is now not difficult to answer this question. A mass operation always consists in the repetition of a large number of identical individual operations (e.g., firing—of individual shots, mass production—the manufacture of individual articles, and so on). We are interested in a well-defined result of individual operations (hitting the target in a single shot, the fact that an individual article is nonstandard, and so forth), and above all in the number of such results in some mass operation (how many shots will hit the target, how many articles will be rejected, and so on). The percentage or, in general, the fractional part of such "successful" results in a given mass operation will be called the *probability* of this result—this is of importance to us. In the second example it would be more appropriate to say "unsuccessful" results. However, in the theory of probability it is conventional to call those results which lead to the realization of the event which interests us in a problem "successful." In this connection, one must always have in view that the question of the probability of an event (result) has meaning only under precisely defined conditions in which our mass operation proceeds. Every essential variation of these conditions causes, as a rule, a change in the probability of the event under consideration.

If the mass operation is such that event A (for example, hitting the target) is observed on the average a times in b individual operations (shots), then the probability of the event A under the given conditions is $\frac{a}{b}$ $\left(\text{or } \frac{100a}{b}\%\right)$. We can therefore say that the *probability* of a "successful" result of an individual operation is the *ratio of the number of such "successful" results observed to the number of these individual operations* constituting the prescribed mass operation. It is self-evident that if the probability of some event equals a/b, then in every collection of b individual operations this event can possibly occur more than a times and less than a times—it is only *on the average* that it occurs approximately a times. And in the majority of many such collections of b operations the number of occurrences of event A will be close to a—*particularly, if* b *is a large number.*

Example 1. During the first quarter of the year, in a certain city there were born:

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

What is the probability that a boy is born? The fractional part of boy births is:

$$\frac{145}{280} \approx 0.518 = 51.8\% \text{ in January}$$

$$\frac{142}{278} \approx 0.511 = 51.1\% \text{ in February}$$

$$\frac{152}{292} \approx 0.520 = 52.0\% \text{ in March.}$$

We see that the arithmetic average of the fractional parts for the individual months is close to the number $0.516 = 51.6\%$, so the probability sought, under the given conditions, is approximately 0.516 or 51.6%. This number is well known in demography (which is the science whose domain is the study of population dynamics); it appears that the fractional part of boy births under usual conditions will not deviate significantly from this number during various periods of time.

EXAMPLE 2. At the beginning of the last century there was discovered a remarkable phenomenon, which received the name Brownian movement (after the English botanist Brown who discovered it). This phenomenon is that very fine particles of matter suspended in a liquid are in chaotic motion which is executed without any visible causes. For a long time the reason for this apparently spontaneous motion could not be clarified, until the kinetic theory of gases gave a simple and complete explanation: the movement of particles suspended in a liquid results from the collision of molecules of the liquid against these particles. The kinetic theory of gases enables one to calculate the probability that in a given volume of liquid there will not be a single particle of suspended matter, the probability that there will be one, two, three, and so on, such particles. A number of experiments were carried out with the purpose of verifying the predications of the theory.

We present the results of 518 observations, made by the Swedish chemist Svedberg, of very fine particles of gold suspended in water. It was found that in the portion of space under observation, not a single particle was observed 112 times, 1 particle was observed 168 times, 2 particles 130 times, 3 particles 69 times, 4 particles 32 times, 5 particles 5 times, 6 particles once, and finally, 7 particles once. The fractional part of the observed number of particles equals

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

$$1 \text{ particle: } \frac{168}{518} \approx 0.325 \qquad 5 \quad " \quad \frac{5}{518} \approx 0.010$$

$$2 \text{ particles: } \frac{130}{518} \approx 0.251 \qquad 6 \quad " \quad \frac{1}{518} \approx 0.002$$

$$3 \quad " \quad \frac{69}{518} \approx 0.133 \qquad 7 \quad " \quad \frac{1}{518} \approx 0.002.$$

The results of the observations, as it turned out, coincided very well with the theoretically predicted probabilities.

EXAMPLE 3. In a number of problems which are important in practice, it is essential to know how frequently certain letters of the

Russian alphabet can occur in a text. Thus, for example, it is irrational to stock up the same number of all letters in forming a typographical font, since certain letters in the text are encountered significantly more frequently than others. Therefore, one strives to have a larger number of the letters which are encountered more frequently. Investigations performed on literary texts led to an estimate of the frequency of occurrence of the letters in the Russian alphabet, including the spaces between letters, which is summarized in the following table[1] (set up in the order of decreasing relative frequency of occurrence).

Thus, the indicated investigations show that on the average of 1000 spaces and letters selected at random in a text, the letter "ф" will occur in two places, the letter "к" in twenty-eight places, the letter "о" in ninety places, and there will be spaces between letters in one hundred and seventy-five places. These data are sufficiently valuable information for forming stock fonts.

In recent years similar investigations, no longer restricted to the statistics of letters in Russian texts, are beginning to be used extensively for the explanation of the peculiarities of the Russian language, and also of the literary style of various authors.

Letter	space	о	е, ё	а	и	т	н
Relative frequency	0.175	0.090	0.072	0.062	0.062	0.053	0.053
Letter	с	р	в	л	к	м	д
Relative frequency	0.045	0.040	0.038	0.035	0.028	0.026	0.025
Letter	п	у	я	ы	з	ь, ъ	б
Relative frequency	0.023	0.021	0.018	0.016	0.016	0.014	0.014
Letter	г	ч	й	х	ж	ю	ш
Relative frequency	0.013	0.012	0.010	0.009	0.007	0.006	0.006
Letter	ц	щ	э	ф			
Relative frequency	0.004	0.003	0.002	0.002			

Similar data relative to telegraph communications can be used for the creation of the most economical telegraph codes which would allow one to transmit messages by means of a smaller number of signs and, therefore, more rapidly. It has become clear that the telegraph codes utilized now are not sufficiently economical.

[1] This little table was adapted by the first-named author from the extraordinarily popular booklet *Probability and Information* by A. M. Yaglom and I. M. Yaglom, 2nd ed., Fizmatgiz, 1960.

§ 2. Impossible and certain events

The probability of an event, obviously, is always a positive number or zero. It cannot be greater than unity because in the fraction by which it is defined the numerator cannot be greater than the denominator, for the number of "successful" operations cannot be greater than the number of all operations undertaken.

We agree to denote the probability of the event A by $P(A)$. Whatever this event is, we have

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

The larger $P(A)$ is, the more often the event A occurs. For example, the greater the probability that a marksman hits the target, the more often does he have successful shots. If the probability of an event is very small, then it occurs rarely; if $P(A)=0$, then the event either never occurs or it occurs very rarely, so that in practice one can consider it to be *impossible*. In contrast, if $P(A)$ is close to unity, then in the fraction by which this probability is expressed, the numerator is close to the denominator, i.e., the overwhelming majority of operations are "successful"; if $P(A)=1$, then the event A occurs always or almost always, so that in practice one can assume it to be, as one says, "certain," i.e., one can assume that its occurrence is *certain*. If $P(A)=1/2$, then the event A occurs in approximately half of all cases; this means that "successful" operations are observed approximately as often as "unsuccessful" ones. If $P(A)>1/2$, then the event A occurs more frequently than it does not occur; for $P(A)<1/2$, we have the reverse phenomenon.

How small must the probability of an event be before we can assume it to be, in practice, impossible? It is impossible to give a general answer to this question because everything depends on how important the event is with which we are dealing. Thus, 0.01 is a small number. If we have a supply of shells and 0.01 is the probability that a shell will not explode upon falling, then this means that approximately 1% of the shots will be ineffective. One can reconcile oneself to this! But if we have a parachute and the probability that in a jump it will not open is 0.01, then it is of course impossible to reconcile oneself with this under any circumstances, because this means that in one out of a hundred jumps the valuable life of a parachutist will be lost. These examples show that in every individual problem we must establish in advance, on the basis of practical considerations, how small the probability of an event ought to be in order that we can consider

it to be impossible and of insignificant consequence to the undertaking at hand.

§ 3. Problem

PROBLEM. One marksman has 80% as his average of target hits and another (under the same firing conditions) has 70%. Find the probability of destroying the target if both marksman shoot at it simultaneously. The target is assumed to be destroyed if at least one of the two bullets hits it.

First method of solution. We assume that 100 double shots are fired. The target will be destroyed by the first marksman in approximately 80 of them. There remain about 20 shots in which this marksman misses. Since the second marksman destroys the target on the average 70 times in 100 shots and hence 7 times in 10 shots, we can expect that in these 20 shots in which the first marksman misses, the second succeeds in destroying the target approximately 14 times. Thus, in all 100 shots, the target turns out to be destroyed approximately $80+14=94$ times. The probability of destroying the target under the simultaneous fire of both marksmen is therefore equal to 94% or 0.94.

Second method of solution. We again assume that 100 double shots are fired. We have already seen that in this connection the first marksman has approximately 20 misses. Since the second marksman has approximately 30 misses per hundred shots and hence 3 misses per ten shots, one can expect that among those 20 shots in which the first marksman misses, there will be approximately 6 in which the second will also miss. In each of these 6 shots, the target will remain undestroyed and in each of the remaining 94 shots at least one of the marksmen will shoot successfully and hence the target will be destroyed. We again arrive at the result that for a double firing the target will be destroyed in approximately 94 cases in 100; i.e., that the probability of destruction is 94% or 0.94.

The problem we considered is very simple. But, nonetheless, it already leads us to a very important result: there are cases when it is useful to know how to find, knowing the probabilities of certain events, the probabilities of other, more complicated events. In fact, there are very many cases like this not only in military operations but also in every science and in every practical activity where we encounter mass phenomena.

Of course, it would be very inconvenient to search for the particular method of solution for every new problem of this sort encountered.

Science always endeavors to form general rules, the knowledge of which would readily permit one to solve mechanically or almost mechanically individual problems which are similar to one another. In the area of mass phenomena, the science which takes upon itself the formulation of such general rules is called the *theory of probability.* The first principles of this science will be given in this book.

The theory of probability is one chapter of mathematical science, like arithmetic or geometry. Therefore, its path is the path of precise reasoning, and formulas, tables, diagrams, and so on, serve as its tools.

CHAPTER 2

RULE FOR THE ADDITION OF PROBABILITIES

§ 4. Derivation of the rule for the addition of probabilities

The simplest and most important rule used in the calculation of probabilities is the *addition rule,* which we shall now consider.

In firing at a target, depicted in Fig. 1, for every marksman standing at a prescribed distance, there is a certain probability of hitting each of the regions 1, 2, 3, 4, 5, 6. Suppose that for some marksman the probability of hitting region 1 is 0.24 and that the probability of hitting region 2 is 0.17. As we already know, this means that of one hundred bullets shot by this marksman, 24 bullets (on the average) hit region 1 and 17 bullets hit region 2.

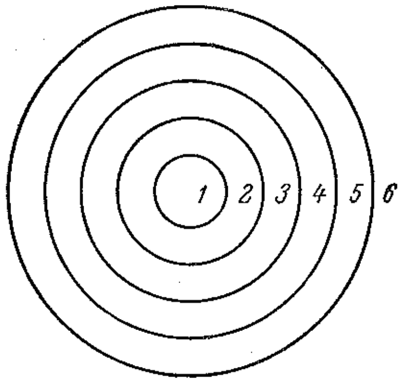

FIG. 1

Suppose that, in some competition, a shot is adjudged "exceptional" if the bullet falls into region 1 and "good" if it falls into region 2. What is the probability that the marksman's shot is either good or exceptional?

It is easy to answer this question. Of every hundred bullets shot by the marksman, approximately 24 fall into region 1 and approximately 17 into region 2. This means that of every hundred bullets there will be approximately $24+17=41$ which will fall into either region 1 or into region 2. The probability sought therefore equals $0.41=0.24+0.17$. Consequently, *the probability that the shot will be either exceptional or good equals the sum of the probabilities of the exceptional and good shots.*

Let us consider still another example. A passenger is waiting for trolley No. 26 or No. 16 at a trolley stop at which trolleys with one of the four route Nos. 16, 22, 26, and 31 stop. Assuming that the trolleys of all routes appear on the average equally frequently, find the probability that the first trolley appearing at the stop will have the route needed by the passenger.

Clearly, the probability that trolley No. 16 will be the first to appear at the stop equals 1/4; the probability that trolley No. 26 will be the first is the same. So, the probability sought is obviously equal to 1/2. But

$$1/2 = 1/4+1/4;$$

therefore we can say that the probability that trolley No. 16 or trolley No. 26 will appear first equals the sum of the probabilities of the appearance of trolley No. 16 and trolley No. 26.

We can now carry out the general discussion. In the performance of a certain mass operation, it was established that in every series of b individual operations on the average

a certain result A_1 is observed a_1 times
" " A_2 " a_2 "
" " A_3 " a_3 "

and so forth. In other words,

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

and so on. How great is the probability that, in some individual operation, one of the results $A_1, A_2, A_3, \ldots$ occurs, it being immaterial which one?

The event of interest can be called "A_1 or A_2 or A_3 or ...." (Here and in other similar cases the ellipsis dots [...] denote "and so forth.") In a series of b operations, this event occurs $a_1+a_2+a_3+\ldots$ times; this means that the probability sought equals

$$\frac{a_1+a_2+a_3+\ldots}{b} = \frac{a_1}{b}+\frac{a_2}{b}+\frac{a_3}{b}+\ldots$$

which can be written as the following formula:

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

In this connection, in our examples as well as in our general discussion, we always assume that any two of the results considered (for instance, A_1 and A_2) are *mutually incompatible*, i.e., they cannot be observed together in the same individual operation. For instance, the trolley arriving cannot simultaneously be from a needed and not-needed route—it either satisfies the requirement of the passenger or it does not. This assumption concerning the mutual incompatibility of the individual results is very important, for without it the addition rule becomes invalid and its application leads to serious errors. We consider, for example, the problem we solved at the end of the preceding section (see page 9). There we even found the probability that for a double shot either one or the other shot will hit the target, in which connection for the first marksman the probability of hitting the target equals 0.8 and for the second 0.7. If we wished to apply the addition rule to the solution of this problem, then we would at once have found that the probability sought equals $0.8+0.7=1.5$ which is manifestly absurd since we already know that the probability of an event cannot be greater than unity. We arrived at this invalid and meaningless answer because we applied the addition law to a case where one must not apply it: the two results we are dealing with in this problem are *mutually compatible*, inasmuch as it is entirely possible that both marksmen destroy the target with the same double shot. A significant portion of errors which novices make in the computation of probabilities is due in fact to such an invalid application of the addition rule. It is therefore necessary to guard carefully against this error and verify in every application of the addition rule whether in fact, among those events to which we wish to apply it, every pair is mutually incompatible.

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

ADDITION RULE. *The probability of occurrence in a certain operation of any one of the results* $A_1, A_2, \ldots, A_n$ *(it being immaterial which one) is equal to the sum of the probabilities of these results, provided that every pair of them is mutually incompatible.*

§ 5. Complete system of events

In the Third (Soviet) Government Loan (TSGL) for the reconstruction and development of the national economy, in the course of the twenty-year period of its operation, a third of the bonds win and the remaining two-thirds are drawn in a lottery and are paid off at the nominal rate. In other words, for this loan each bond has a probability equal to 1/3 of winning and a probability equal to 2/3 of being drawn in a lottery. Winning and being drawn in a lottery are *complementary events*; i.e., they are two events such that one and only one of them must necessarily occur for every bond. The sum of their probabilities is

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

and this is not accidental. In general, if A_1 and A_2 are two complementary events and if in a series of b operations the event A_1 occurs a_1 times and the event A_2 occurs a_2 times, then, obviously, $a_1+a_2=b$. But

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

so that

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

This same result can also be obtained from the addition rule: since complementary events are mutually incompatible, we have

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

But the event "A_1 or A_2" is a certain event since it follows from the definition of complementary events that it certainly must occur; therefore, its probability equals unity and we again obtain

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

The sum of the probabilities of two complementary events equals unity.

This rule admits of a very important generalization which can be proved by the same method. Suppose we have n events $A_1, A_2, \ldots, A_n$ (where n is an arbitrary positive integer) such that in each individual operation one and only one of these events must necessarily occur; we agree to call such a group of events a *complete system*. In particular, every pair of complementary events, obviously, constitutes a complete system.

The sum of the probabilities of events constituting a complete system is equal to unity.

In fact, according to the definition of a complete system, any two events in this system are mutually incompatible, so that the addition rule yields

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

But the right member of this equality is the probability of a certain event and it therefore equals unity; thus, for a complete system, we have

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

which was to be proved.

EXAMPLE 1. Of every 100 target shots (target depicted in Fig. 1 on page 11), a marksman has on the average

44	hits in	region	1
30	,,	,,	2
15	,,	,,	3
6	,,	,,	4
4	,,	,,	5
1	hit	,,	6

($44+30+15+6+4+1=100$). These six firing results obviously constitute a complete system of events. Their probabilities are equal to

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

respectively; we have

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

Shots falling completely or partially into the region 6 do not hit the target at all and cannot be considered; this does not, however, hinder finding the probability of falling into this region, for which it is sufficient to subtract from unity the sum of the probabilities of falling into all the other regions.

EXAMPLE 2. Statistics show that at a certain weaving factory, of every hundred stoppages of a weaving machine requiring the subsequent work of the weaver, on the average,

22	occur due to	a break in the warp thread	
31	,, ,,	,, ,, woof	,,
27	,, ,,	change in the shuttle	
3	occur due to	breakage of the shuttlecock	

and the remaining stoppages of the machine are due to other reasons.

We see that besides other reasons for the stoppage of the machine, there are four definite reasons whose probabilities are equal to

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

respectively. The sum of these probabilities equals 0.83. Together with the other reasons, the reasons pointed out for stoppage of the machine constitute a complete system of events; therefore, the probability of stoppage of the machine from other causes equals

$$1-0.83 = 0.17.$$

§ 6. Examples

We frequently successfully base the so-called a priori, i.e., pre-trial, calculation of probabilities on the theorem concerning a complete system of events which we have established. Suppose, for example, that we are studying the falling of cosmic particles into a

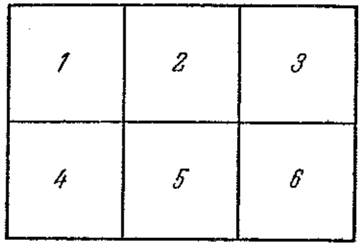

FIG. 2

small area of rectangular form (see Fig. 2)—this area being subdivided into the 6 equal squares numbered in the figure. The subareas of interest find themselves under the same conditions and therefore there is no basis for assuming that particles will fall into any one of these six squares more often than another. We therefore assume that on the average particles will fall into each of the six squares equally frequently, i.e., that the probabilities $p_1, p_2, p_3, p_4, p_5, p_6$ of falling into these squares are equal. If we assume that we are interested only in particles which fall into this area, then it will follow from this that each of the numbers p equals $1/6$, inasmuch as these numbers are equal and their sum equals unity by virtue of the theorem we proved above. Of course this result, which is based on a number of assumptions, requires experimental verification for its affirmation. We have, however, become so accustomed in such cases to obtaining excellent agreement between our theoretical assumptions and their experimental verifications that we can depend

on the theoretically deduced probabilities for all practical purposes. We usually say in such cases that the given operation can have n distinct, mutually *equi-probable* results (thus, in our example of cosmic particles falling into an area, depicted in Fig. 2, the result is that the particle falls into one of the six squares). The probability of each of these n results is equal in this case to $1/n$. The importance of this type of a priori reasoning is that in many cases it enables us to foresee the probability of an event under conditions where its determination by repetitive operations is either absolutely impossible or extremely difficult.

EXAMPLE 1. In the case of government loan bonds, the numbers of a series are usually expressed by five-digit numbers. Suppose we wish to find the probability that the last digit, taken at random from a winning series, equals 7 (as, for example, in the series No. 59607). In accordance with our definition of probability, we ought to consider, for this purpose, a long series of lottery tables and calculate how many winning series have numbers ending in the digit 7; the ratio of this number to the total number of winning series will then be the probability sought. However, we have every reason to assume that any one of the ten digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 has as much of a chance to appear in the last place in a number of the winning series as any other. Therefore, without any hesitation, we make the assumption that the probability sought equals 0.1. The reader can easily verify the legitimacy of this theoretical "foresight": carry out all necessary calculations within the framework of any one lottery table and verify that in reality each of the 10 digits will appear in the last place in approximately 1/10 of all cases.

EXAMPLE 2. A telephone line connecting two points A and B at a distance of 2 km. broke at an unknown spot. What is the probability that it broke no farther than 450 m. from the point A? Mentally subdividing the entire line into individual meters, we can assume, by virtue of the actual homogeneity of all these parts, that the probability of breakage is the same for every meter. From this, similar to the preceding, we easily find that the required probability equals

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

CHAPTER 3

CONDITIONAL PROBABILITIES AND THE MULTIPLICATION RULE

§ 7. The concept of conditional probability

Electric light bulbs are manufactured at two plants—the first plant furnishes 70% and the second 30% of all required production of bulbs. At the first plant, among every 100 bulbs 83 are on the average standard,[1] whereas only 63 per hundred are standard at the second plant.

It can easily be computed from these data that on the average each set of 100 electric light bulbs purchased by a consumer will contain 77 standard bulbs and, consequently, the probability of buying a standard bulb equals 0.77.[2] But we shall now assume that we have made it clear that the bulbs on stock in a store were manufactured at the first plant. Then the probability that the bulb is standard will change—it will equal $83/100 = 0.83$.

The example just considered shows that the addition to the general conditions under which an operation takes place (in our case this is the purchase of the bulbs) of some essentially new condition (in our example this is knowledge of the fact that the bulb was produced by one or the other of the plants) can change the probability of some result of an individual operation. But this is understandable; for the very definition of the concept of probability requires that the totality of conditions under which a given mass operation occurs be precisely defined. By adding any new condition to this collection of conditions we, generally speaking, change this collection in an essential way. Our mass operation takes place after this addition under new conditions; in reality, this is already another operation, and therefore the probability of some result in it will no longer be the same as that under the initial conditions.

We thus have two distinct probabilities of the same event—i.e., the purchase of a standard bulb—but these probabilities are calculated under different conditions. As long as we do not set down an additional condition (e.g., not considering where the bulb was manufactured), we take the *unconditional probability* of purchasing a standard bulb as equal to 0.77; but upon placing an additional condition (that the bulb was manufactured in the first plant) we obtain the *conditional probability* 0.83, which differs somewhat from the preceding. If we denote by A the event of purchasing a standard bulb and by B the event that it was manufactured in the first plant, then we usually denote by $P(A)$ the unconditional probability of event A and by $P_B(A)$ the probability of the same event under the condition that event B has occurred, i.e., that the bulb was manufactured by the first plant. We thus have $P(A) = 0.77$, $P_B(A) = 0.83$.

[1] In this regard, we call a bulb "standard" (i.e., it meets certain *standard* requirements) if it is capable of functioning no less than 1200 hours; otherwise, the bulb will be called substandard.

[2] In fact, we have $0.83 \cdot 70 + 0.63 \cdot 30 = 77$.

Since one can discuss the probability of a result of a given operation only under certain precisely defined conditions, every probability is, strictly speaking, a conditional probability; unconditional probabilities cannot exist in the literal sense of this word. In the majority of concrete problems, however, the situation is such that at the basis of all operations considered in a given problem there lies some well-defined set of conditions K which are assumed satisfied for all operations. If in the calculation of some probability no other conditions except the set K are assumed, then we shall call such a probability *unconditional*; the probability calculated under the assumption that further precisely prescribed conditions, besides the set of conditions K common to all operations, are satisfied will be called *conditional*.

Thus, in our example, we assume, of course, that the manufacture of a bulb occurs under certain well-defined conditions which remain the same for all bulbs which are placed on sale. This assumption is so unavoidable and self-evident that in the formulation of problems we did not even find it necessary to mention it. If we do not place any additional conditions on the given bulb, then the probability of some result in the testing of the bulb will be called unconditional. But if, over and above these conditions, we make still other, additional requirements, then the probabilities computed under these requirements will now be conditional.

Example 1. In the problem we described at the beginning of the present section, the probability that the bulb was manufactured by the second plant obviously equals 0.3. It is established that the bulb is of standard quality. After this observation, what is the probability that this bulb was manufactured at the second plant?

Among every 1000 bulbs put on the market, on the average 770 bulbs are of standard quality—and of this number 581 bulbs came from the first plant and 189 bulbs came from the second.[1] After making this observation, the probability of issuing a bulb by the second plant therefore becomes $189/770 \approx 0.245$. This is the conditional probability of issuing a bulb by the second plant, calculated under the assumption that the given bulb is standard. Using our previous notation, we can write $P(\bar{B}) = 0.3$ and $P_A(\bar{B}) \approx 0.245$, where the event $\bar{B}$ denotes the nonoccurrence of the event B.

EXAMPLE 2. Observations over a period of many years carried out in a certain region showed that among 100,000 children who have attained the age of 9, on the average 82,277 live to 40 and 37,977 live to 70. Find the probability that a person who attains the age 40 will also live to 70.

Since on the average 37,977 of the 82,277 forty-year-olds live to 70, the probability that a person aged 40 will live to 70 equals $37{,}977/82{,}277 \approx 0.46$.

If we denote by A the first event (that a nine-year-old child lives to 70) and by B the second event (that this child attains the age 40), then obviously, we have $P(A) = 0.37{,}977 \approx 0.38$ and $P_B(A) \approx 0.46$.

§ 8. Derivation of the rule for the multiplication of probabilities

We now return to the first example in the preceding section. Among every 1000 bulbs placed on the market, on the average 300 were manufactured at the second plant, and among these 300 bulbs on the average 189 are of standard quality. We deduce from this that the probability that the bulb was manufactured at the second plant (i.e., event $\bar{B}$) equals $P(\bar{B}) = 300/1000 = 0.3$ and the probability that it is of standard quality, under the condition that it was manufactured at the second plant, equals $P_{\bar{B}}(A) = 189/300 = 0.63$.

Since, out of every 1000 bulbs, 189 were manufactured at the second plant and are at the same time of standard quality, the probability of the simultaneous occurrence of the events A and $\bar{B}$ equals

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

[1] This can easily be calculated as follows. Among every 1000 bulbs, on the average 700 were manufactured at the first plant, and among every 100 bulbs from the first plant on the average 83 are of standard quality. Consequently, among 700 bulbs from the first plant, on the average $7\cdot 83 = 581$ will be of standard quality. The remaining 189 bulbs of standard quality were produced at the second plant.

This "multiplication rule" can also be easily extended to the general case. Suppose in every sequence of n operations, the result B occurs on the average m times, and that in every sequence of m such operations in which the result B is observed, the result A occurs l times. Then, in every sequence of n operations, the simultaneous occurrence of the events B and A will be observed on the average l times. Thus,

$$P(B) = \frac{m}{n}, \qquad P_B(A) = \frac{l}{m},$$

$$P(A \text{ and } B) = \frac{l}{n} = \frac{m}{n}\cdot\frac{l}{m} = P(B)\cdot P_B(A). \tag{1}$$

MULTIPLICATION RULE. *The probability of the simultaneous occurrence of two events equals the product of the probability of the first event with the conditional probability of the second, computed under the assumption that the first event has occurred.*

It is understood that we can call either of the two given events the first so that on an equal basis with formula (1) we can also write

$$P(A \text{ and } B) = P(A)\cdot P_A(B), \tag{1'}$$

from which we obtain the important relation:

$$P(A)\cdot P_A(B) = P(B)\cdot P_B(A). \tag{2}$$

In our example, we had

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

and this shows that formula (1′) is satisfied.

EXAMPLE. At a certain enterprise, 96% of the articles are judged to be usable (event A); out of every hundred usable articles, on the average 75 turn out to be of the first sort (event B). Find the probability that an article manufactured at this enterprise is of the first sort.

We seek $P(A \text{ and } B)$ since, in order that an article be of the first sort, it is necessary that it be usable (event A) and of the first sort (event B).

By virtue of the conditions of the problem, $P(A) = 0.96$ and $P_A(B) = 0.75$. Therefore, on the basis of formula (1′), $P(A \text{ and } B) = 0.96 \cdot 0.75 = 0.72$.

§ 9. Independent events

Two skeins of yarn, manufactured on different machines, were tested for strength. It turned out that a sample of prescribed length taken from the first skein held a definite standard load with probability 0.84 and that from the second skein with probability 0.78.[1] Find the probability that two samples of yarn, taken from two different skeins, are both capable of supporting the standard load.

We denote by A the event that the sample taken from the first skein supports the standard load and by B the analogous event for the sample from the second skein. Since we are seeking $P(A \text{ and } B)$, we apply the multiplication rule:

$$P(A \text{ and } B) = P(A) \cdot P_A(B).$$

Here we obviously have $P(A) = 0.84$; but what is $P_A(B)$? According to the general definition of conditional probabilities, this is the probability that the sample of yarn from the second skein will support the standard load if the sample from the first skein supported such a load. But the probability of event B does not depend on whether or not event A has occurred, for these tests can be carried out simultaneously and the yarn samples are chosen from completely unrelated skeins, manufactured on different machines. In practice, this means that the percentage of trials in which the yarn from the second skein supports the standard load does not depend on the strength of the sample from the first skein; i.e.,

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

It follows from this that

$$P(A \text{ and } B) = P(A) \cdot P(B) = 0.84 \cdot 0.78 = 0.6552.$$

The peculiarity which distinguishes this example from the preceding ones consists, as we see, in that here the probability of the result B is not changed by the fact that to the general conditions we add the requirement that the event A occur. In other words, the conditional probability $P_A(B)$ equals the unconditional probability $P(B)$. In this case we will say, briefly, that *the event* B *does not depend on the event* A.

It can easily be verified that if B does not depend on A, then A also does not depend on B. In fact, if $P_A(B) = P(B)$, then by virtue of formula (2) $P_B(A) = P(A)$ and this means that the event A does not depend on the event B. Thus, the independence of two events is a *mutual* (or *dual*) property. We see that for mutually independent events, the multiplication rule has a particularly simple form:

$$P(A \text{ and } B) = P(A) \cdot P(B). \tag{3}$$

[1] If the standard load equals, say, 400 grams, then this means the following: among 100 samples taken from the first skein, 84 samples on the average support such a load and 16 do not support it and break.

As in every application of the addition rule it is necessary to establish in advance the mutual incompatibility of the given events, so in every application of rule (3) it is necessary to verify that the events A and B are mutually independent. Disregard for these instructions leads to errors. If the events A and B are mutually dependent, then formula (3) is not valid and must be replaced by the more general formula (1) or 1′).

Rule (3) is easily generalized to the case of seeking the probability of the occurrence of not two, but of three or more mutually independent events. Suppose, for example, that we have three mutually independent events A, B, C (this means that the probability of any one of them does not depend on the occurrence or the nonoccurrence of the other two events). Since the events A, B and C are mutually independent, we have, by rule (3):

$$P(A \text{ and } B \text{ and } C) = P(A \text{ and } B) \cdot P(C).$$

Now if we substitute here for $P(A \text{ and } B)$ the expression for this probability from formula (3), we find:

$$P(A \text{ and } B \text{ and } C) = P(A) \cdot P(B) \cdot P(C). \tag{4}$$

Clearly, such a rule holds in the case when the set under consideration contains an arbitrary number of events as long as these events are mutually independent (i.e., the probability of each of them does not depend on the occurrence or nonoccurrence of the remaining events).

The probability of the simultaneous occurrence of any number of mutually independent events equals the product of the probabilities of these events.

Example 1. A worker operates three machines. The probability that for the duration of one hour a machine does not require the attention of the worker equals 0.9 for the first machine, 0.8 for the second, and 0.85 for the third. Find the probability that for the duration of an hour none of the machines requires the worker's attention.

Assuming that the machines work independently of each other, we find, by formula (4), that the probability sought is

$$0.9 \cdot 0.8 \cdot 0.85 = 0.612.$$

Example 2. Under the conditions of Example 1, find the probability that at least one of the three machines does not require the attention of the worker for the duration of one hour.

In this problem, we are dealing with a probability of the form $P(A \text{ or } B \text{ or } C)$ and, therefore, we of course think first of all of the addition rule. However, we soon realize that this rule is not applicable in the present case inasmuch as any two of the three events considered can occur simultaneously; for nothing hinders any two machines from working without being given attention for the duration of the same hour. Moreover, independently of this line of reasoning, we at once see that the sum of the three given probabilities is significantly larger than unity and hence we cannot compute the probability in this way.

To solve the problem as stated, we note that the probability that a machine requires the attention of the worker equals 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 these events are realized equals

$$0.1 \cdot 0.2 \cdot 0.15 = 0.0003,$$

according to rule (4). But the events "all three machines require attention" and "at least one of the three machines operates without receiving attention" clearly represent a pair of complementary events. Therefore, the sum of their probabilities equals unity and, consequently, the probability sought equals $1-0.0003=0.9997$. When the probability of an event is as close to unity as this, then this event can in practice be assumed to be certain. This means that almost always, in the course of an hour, at least one of the three machines will operate without receiving attention.

Example 3. Under certain definite conditions, the probability of destroying an enemy's plane with a rifle shot equals 0.004. Find the probability of destroying an enemy plane when 250 rifles are fired simultaneously.

For each shot, the probability is $1-0.004=0.996$ that the plane will not be downed. The probability that it will not be downed by all 250 shots equals, according to the multiplication rule for independent events, the product of 250 factors each of which equals 0.996, i.e., it is equal to $(0.996)^{250}$. And the probability that at least one of the 250 shots proves to be sufficient for downing the plane is therefore equal to

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

A detailed calculation, which will not be carried out here, shows that this number is approximately equal to 5/8. Thus, although the probability of downing an enemy plane by one rifle shot is negligibly small—0.004—with the simultaneous firing from a large number of rifles, the probability of the desired result becomes very significant.

The line of reasoning which we utilized in the last two examples can easily be generalized and leads to an important general rule. In both cases, we were dealing with the probability $P(A_1 \text{ or } A_2 \text{ or } A_3 \ldots \text{ or } A_n)$ of the occurrence of at least one of several mutually independent events $A_1, A_2, \ldots, A_n$. If we denote by $\bar{A}_k$ the event that A_k will not occur, then the events A_k and $\bar{A}_k$ are complementary, so that

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

On the other hand, the events $\bar{A}_1, \bar{A}_2, \ldots, \bar{A}_n$ are obviously mutually independent so that

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

Finally, the events (A_1 or A_2 or ... or A_n) and ($\bar{A}_1$ and $\bar{A}_2$ and ... and $\bar{A}_n$) obviously are complementary; that is, one of the following: either at least one of the events A_k occurs or all the events $\bar{A}_k$ occur. Therefore,

$$\begin{aligned} P(A_1 \text{ or } A_2 \text{ or } \ldots \text{ or } A_n) &= 1-P(\bar{A}_1 \text{ and } \bar{A}_2 \text{ and } \ldots \text{ and } \bar{A}_n) \\ &= 1-[1-P(A_1)]\cdot[1-P(A_2)] \ldots [1-P(A_n)]. \qquad (5) \end{aligned}$$

This important formula, which enables one to calculate the probability of the occurrence *of at least one* of the events $A_1, A_2, \ldots, A_n$ on the basis of the given probabilities of these events, is valid if, and only if, these events are mutually independent. In the particular case when all the events A_k have the same probability p (as was the case in Example 3, above) we have:

$$P(A_1 \text{ or } A_2 \text{ or } \ldots \text{ or } A_n) = 1-(1-p)^n. \qquad (6)$$

Example 4. An instrument part is being lathed in the form of a rectangular parallelepiped. The part is considered usable if the length of each of its edges deviates by no more than 0.01 mm. from prescribed dimensions. If the probability of deviations exceeding 0.01 mm. is

$p_1 = 0.08$ along the length of the parallelepiped
$p_2 = 0.12$ " width " "
$p_3 = 0.10$ " height " "

find the probability P that the part is not usable.

For the part to be unusable, it is necessary that at least in one of the three directions the deviation from the prescribed dimension exceed

0.01 mm. Since these three events can usually be assumed mutually independent (because they are basically due to different causes), to solve the problem we can apply formula (5); this yields

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

Consequently, we can assume that of every 100 parts approximately 73 on the average turn out to be usable.

CHAPTER 4

CONSEQUENCES OF THE ADDITION AND MULTIPLICATION RULES

§ 10. Derivation of certain inequalities

We turn again to the electric light bulb example of the preceding chapter (see page 18). We introduce the following notation for events:

A—the bulb is of standard quality
$\bar{A}$— the bulb is of substandard quality
B—the bulb was manufactured at the first plant
$\bar{B}$—the bulb was manufactured at the second plant.

Obviously, events A and $\bar{A}$ constitute a pair of complementary events; the events B and $\bar{B}$ form a pair of the same sort.

If the bulb is of standard quality (A), then either it was manufactured by the first plant (A and B) or by the second (A and $\bar{B}$). Since the last two events, evidently, are incompatible with one another, we have, according to the addition rule

$$P(A) = P(A \text{ and } B)+P(A \text{ and } \bar{B}). \quad (1)$$

In the same way, we find that

$$P(B) = P(A \text{ and } B)+P(\bar{A} \text{ and } B). \quad (2)$$

Finally, we consider the event (A or B); we obviously have the following three possibilities for its occurrence:

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

Of these three possibilities, any two are incompatible with one another; therefore, by the addition rule, we have

$$P(A \text{ or } B) = P(A \text{ and } B)+P(A \text{ and } \bar{B})+P(\bar{A} \text{ and } B). \quad (3)$$

Adding equalities (1) and (2) memberwise and taking equality (3) into consideration, we easily find that

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

from which it follows that

$$P(A \text{ or } B) = P(A)+P(B)-P(A \text{ and } B). \quad (4)$$

We have arrived at a very important result. Although we carried out our reasoning for a particular example, it was so general that the result can be considered established for any pair of events A and B. Up to this point, we obtained expressions for probabilities $P(A \text{ or } B)$ only under very particular assumptions concerning the connection between the events A and B (we first assumed them to be incompatible and, later, to be mutually independent). Formula (4) which we just obtained holds without any additional assumptions for an arbitrary pair of events A and B. It is true that we must not forget one essential difference between formula (4) and our previous formulas. In previous formulas, the probability $P(A \text{ or } B)$ was always expressed in terms of the probabilities $P(A)$ and $P(B)$, so that, knowing only the probabilities of the events A and B, we were always able to determine the probability of the event (A or B) uniquely. The situation is different in formula (4): to compute the quantity $P(A \text{ or } B)$ by this formula it is necessary to know, besides $P(A)$ and $P(B)$, the probability $P(A \text{ and } B)$, i.e., the probability of the simultaneous occurrence of the events A and B. To find this same probability in the general case, with arbitrary connection between the events A and B, is usually no easier than to find $P(A \text{ or } B)$; therefore, for practical calculations we seldom use formula (4) directly—but it is, nonetheless, of very great theoretical significance.

We shall first convince ourselves that our previous formulas can easily be obtained from formula (4) as special cases. If the events A and B are mutually incompatible, then the event (A and B) is impossible—hence, $P(A \text{ and } B)=0$—and formula (4) leads to the relation

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

i.e., to the addition law. If the events A and B are mutually independent, then, according to formula (3) on page 22, we have

$$P(A \text{ and } B) = P(A)\cdot P(B),$$

and formula (4) yields

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

Thus, we obtain formula (5) on page 25 (for the case $n=2$).

Furthermore, we deduce an important corollary from formula (4). Since $P(A \text{ and } B)\geq 0$ in all cases, it follows from formula (4) in all cases that

$$P(A \text{ or } B) \leq P(A)+P(B). \tag{5}$$

This inequality can easily be generalized to any number of events. Thus, for instance, in the case of three events, we have, by virtue of (5),

$$\begin{aligned} P(A \text{ or } B \text{ or } C) &\leq P(A \text{ or } B)+P(C) \\ &\leq P(A)+P(B)+P(C), \end{aligned}$$

and, clearly, one can proceed in the same way from three events to four, and so on. We obtain the following general result:

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

In this connection, the equality sign holds only in the case when every pair of the given events is mutually incompatible.

§ 11. Formula for total probability

We return once more to the bulb example on page 18 and use, for the various results of the experiments, the notation introduced on page 27. The probability that a bulb is of standard quality under the condition that it was manufactured at the second plant equals, as we have already seen more than once,

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

and the probability of the same event under the condition that the bulb was manufactured at the first plant is

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

Let us assume that these two numbers are known and that we also know that the probability that the bulb was manufactured at the first plant is

$$P(B) = 0.7$$

and at the second plant is

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

It is required that one find the unconditional probability $P(A)$, i.e., the probability that a random bulb is of standard quality, without any assumptions concerning the place where it was manufactured.

In order to solve this problem, we shall reason as follows. We denote by E the joint event consisting of 1) that the bulb was issued by the first plant and 2) that it is standard, and by F the analogous event for the second plant. Since every standard bulb is manufactured by the first or second plant, the event A is equivalent to the event "E or F" and since the events E and F are mutually incompatible, we have, by the addition law

$$P(A) = P(E)+P(F). \quad (6)$$

On the other hand, in order that the event E hold, it is necessary 1) that the bulb be manufactured by the first plant (B) and 2) that it be standard (A); therefore, the event E is equivalent to the event "B and A," from which it follows, by the multiplication rule, that

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

In exactly the same way we find that

$$P(F) = P(\overline{B})\cdot P_{\overline{B}}(A),$$

and, substituting these expressions into equality (6), we have

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

This formula solves the problem we posed. Substituting the given numbers, we find that $P(A) = 0.77$.

EXAMPLE. For a seeding, there are prepared wheat seeds of the variety I containing as admixture small quantities of other varieties—II, III, IV. We take one of these grains. The event that this grain is of variety I will be denoted by A_1, that it is of variety II by A_2, of variety III by A_3, and, finally, of variety IV by A_4. It is known that the probability that a grain taken at random turns out to be of a certain variety equals:

$$P(A_1) = 0.96;\quad P(A_2) = 0.01;\quad P(A_3) = 0.02\quad P(A_4) = 0.01.$$

(The sum of these four numbers equals unity, as it should in every case of a complete system of events.)

The probability that a spike containing no less than 50 grains will grow from the grain equals:

1) 0.50 for a grain of variety I
2) 0.15 " " II
3) 0.20 " " III
4) 0.05 " " IV.

It is required that one find the unconditional probability that the spike has no less than 50 grains.

Let K be the event that the spike contains no less than 50 grains; then, by the condition of the problem, we have

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

Our problem is to determine $P(K)$. We denote by E_1 the event that the grain turns out to be of variety I and that the spike growing from it will contain no less than 50 grains, so that E_1 is equivalent to the event (A_1 and K); in the same way, we denote

the event (A_2 and K) by E_2
the event (A_3 and K) by E_3
the event (A_4 and K) by E_4.

Obviously, for the event K to occur it is necessary that one of the events E_1, E_2, E_3, or E_4 occur and since any pair of these events is mutually incompatible, we obtain, by the addition rule

$$P(K) = P(E_1)+P(E_2)+P(E_3)+P(E_4). \quad (7)$$

On the other hand, according to the multiplication rule, we have

$$P(E_1) = P(A_1 \text{ and } K) = P(A_1)\cdot P_{A_1}(K)$$
$$P(E_2) = P(A_2 \text{ and } K) = P(A_2)\cdot P_{A_2}(K)$$
$$P(E_3) = P(A_3 \text{ and } K) = P(A_3)\cdot P_{A_3}(K)$$
$$P(E_4) = P(A_4 \text{ and } K) = P(A_4)\cdot P_{A_4}(K).$$

Substituting these expressions into formula (7), we find that

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

which obviously solves our problem. Substituting the given numbers into the last equation, we find that

$$P(K) = 0.486.$$

The two examples which we considered here in detail bring us to an important general rule which we can now formulate and prove without difficulty. Suppose a given operation admits of the results $A_1, A_2, \ldots, A_n$ and that these form a complete system of events. (Let us recall that this means that any two of these events are mutually incompatible and that some one of them must necessarily occur.) Then for an arbitrary possible result K of this operation, the relation

$$P(K) = P(A_1)\cdot P_{A_1}(K)+P(A_2)\cdot P_{A_2}(K)+\ldots+P(A_n)\cdot P_{A_n}(K) \quad (8)$$

holds. Rule (8) is usually called the "*formula for total probability.*" Its proof is carried out exactly as in the two examples we considered above: first, the occurrence of the event K requires the occurrence of one of the events "A_i and K" so that, by the addition rule, we have

$$P(K) = \sum_{i=1}^{n} P(A_i \text{ and } K); \tag{9}$$

second, by the multiplication rule,

$$P(A_i \text{ and } K) = P(A_i) \cdot P_{A_i}(K);$$

substituting these expressions into equation (9) we arrive at formula (8).

§ 12. Bayes's formula

The formulas of the preceding section enable us to derive an important result having numerous applications. We start with a formal derivation, postponing an explanation of the real meaning of the final formula until we consider examples.

Again, let the events $A_1, A_2, \ldots, A_n$ form a complete system of results of some operation. Then, if K denotes an arbitrary result of this operation, we have, by the multiplication rule

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

from which it follows that

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

or, expressing the denominator of the fraction obtained according to the formula for total probability (8) in the preceding section, we find that

$$P_K(A_i) = \frac{P(A_i) \cdot P_{A_i}(K)}{\sum_{r=1}^{n} P(A_r) \cdot P_{A_r}(K)} \quad (1 \leq i \leq n). \tag{10}$$

This is *Bayes's formula*, which has many applications in practice in the calculation of probabilities. We apply it most frequently in situations illustrated by the following example.

Suppose a target situated on a linear segment MN (see Fig. 3) is being fired upon; we imagine the segment MN to be subdivided into five small subsegments a, b', b'', c', c''. We assume that the precise position of the target is not known; we only know the probability that the target lies on one or another of these subsegments. We suppose

where now a, b', b'', c', c'' denote the following events: the target lies in the segment a, b', b'', c', c'', respectively. (Note that the sum of these numbers equals unity.) The largest probability corresponds to the segment a toward which we therefore, naturally, aim our shot.

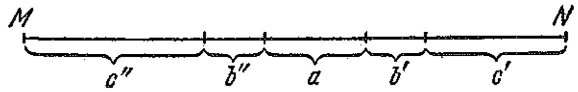

FIG. 3

However, due to unavoidable errors in firing, the target can also be destroyed when it is not in a but in any of the other segments. Suppose the probability of destroying the target (event K) is

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

We assume that a shot has been fired and that the target was destroyed (i.e., event K occurred). As a result of this, the probabilities of the various positions of the target which we had earlier [i.e., the numbers $P(a)$, $P(b'), \ldots$] must be recalculated. The qualitative aspect of this revised calculation is clear without any computations, for we shot at the segment a and hit the target—it is clear that the probability $P(a)$ in this connection must increase. Now we wish to compute exactly and quantitatively the new value due to our shot; i.e., we wish to find an exact expression for the probabilities $P_K(a)$, $P_K(b'), \ldots$ of the various possible positions of the target under the condition that the target was destroyed by the shot fired. Bayes's formula (10) at once gives us the answer to this problem. Thus,

$$P_K(a) = \{P(a) \cdot P_a(K)\}/\{P(a) \cdot P_a(K) + P(b') \cdot P_{b'}(K) + P(b'') \cdot P_{b''}(K) + P(c') \cdot P_{c'}(K) + P(c'') \cdot P_{c''}(K)\} \approx 0.8;$$

we see that $P_K(a)$ is in fact larger than $P(a)$.

We easily find the probabilities $P_K(b'), \ldots$ for the other positions of the target in a similar manner. For the calculations, it is useful to note that the expressions given for these probabilities by Bayes's

The general scheme of this type of situation can be described as follows. The conditions of the operation contain some unknown element with respect to which n distinct "hypotheses" can be made: $A_1, A_2, \ldots, A_n$ which form a complete system of events. For one reason or another we know the probabilities $P(A_i)$ of these hypotheses to be tested; it is also known that the hypothesis A_i "conveys" a probability $P_{A_i}(K)$ $(1 \le i \le n)$ to some event K (for instance, hitting a target). Here, $P_{A_i}(K)$ is the probability of the event K calculated under the condition that the hypothesis A_i is true. If, as the result of a trial, event K has occurred, then this requires a re-evaluation of the probability of the hypothesis A_i and the problem consists in finding the new probabilities $P_K(A_i)$ of these hypotheses; Bayes's formula gives the answer.

In artillery practice, so-called test-firings are carried out which have for their purpose making more precise our knowledge of the firing conditions. In this regard, not only the position of the target can serve as the unknown element whose effect is required to be made precise, but also any other element in the firing conditions which influences the effectiveness of the results (in particular, some peculiarity of the fire-arm used). It often happens that not one such shot is fired but, rather, several, and the problem posed is to calculate the new probabilities of the hypotheses on the basis of the firing results obtained. In all such cases, Bayes's formula also easily solves the problems.

For the sake of brevity in writing, we shall set, in the general scheme considered by us,

$$P(A_i) = P_i \text{ and } P_{A_i}(K) = p_i \quad (1 \le i \le n),$$

so that Bayes's formula has the simple form

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

We assume that s test shots have been fired, in which connection the result K occurred m times and did not occur $s-m$ times. We denote by K^* the result obtained from a series of s shots. We can assume that the results of individual shots constitute mutually independent events. If the hypothesis A_i is valid, the probability of the result K equals p_i and, hence, the probability of the complementary event that K does not occur equals $1-p_i$.

The probability that the result K occurred for the definite m shots equals $p_i^m(1-p_i)^{s-m}$ according to the multiplication rule for independent events. Since the m shots in which the result K occurred can be any of the s fired, the event K^* can be realized in C_s^m incompatible ways. Thus, according to the rule for the addition of probabilities, we have

$$P_{A_i}(K^*) = C_s^m p_i^m (1-p_i)^{s-m} \quad (1 \le i \le n),$$

and Bayes's formula yields

$$P_{K^*}(A_i) = \frac{P_i p_i^m (1-p_i)^{s-m}}{\sum_{r=1}^{n} P_r p_r^m (1-p_r)^{s-m}} \quad (1 \le i \le n), \tag{11}$$

which solves the problem posed. Of course, such problems arise not only in artillery practice, but also in other areas of human activity.

Example 1. Referring to the problem we considered in the beginning of the present section, we now seek the probability that the target lies in the segment a if two successive shots at this segment yielded hits.

Denoting by K^* the event of hitting the target twice, we have, according to formula (11)

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

We leave it to the reader to carry out the uncomplicated calculation and verify that as a result of hitting the target twice the probability that the target is situated in the segment a has been increased still more.

Example 2. The probability that in a certain production process the articles satisfy a prescribed standard equals 0.96. A simplified system of testing[1] is suggested which for the articles satisfying the standard yield a positive result with probability 0.98 and for articles which do not satisfy the standard a positive result with a probability 0.05. What is the probability that the articles which endure the simplified test twice satisfy the standard?

Here, a complete system of hypotheses consists of two complementary events: 1) that the article satisfies the standard, or 2) that the article does not satisfy the standard. The probabilities of these hypotheses are, before the test, equal to $P_1=0.96$ and $P_2=0.04$,

[1] The necessity for a simplified control is encountered very frequently in practice. For instance, if upon dispensing electric light bulbs all of them were subjected to testing for their ability to burn for a period, say, of not less than 1200 hours, then the consumer would obtain only burnt-out or almost burnt-out bulbs. Thus one must replace the test for period of burning by other tests—for example, testing the bulb for lighting up.

respectively. Under the first hypothesis, the probability that the article endures the test equals $p_1=0.98$ and, under the second hypothesis, the probability equals $p_2=0.05$. After a two-fold test, the probability of the first hypothesis is equal, on the basis of formula (11), to

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

We see that if the article endured the test indicated in the conditions of the problem, then we can make an error only once in ten thousand cases assuming that it is standard. This, of course, completely satisfies the requirements in practice.

EXAMPLE 3. In an examination of a patient, it is suspected that he has one of three illnesses: A_1, A_2, A_3. Their probabilities, under prescribed conditions, are

$$P_1=1/2,\quad P_2=1/6,\quad P_3=1/3,$$

respectively. In order to make the diagnosis more precise, some analysis is specified which yields a positive result with probability 0.1 in the case of illness A_1, with probability 0.2 in the case of illness A_2, and with probability 0.9 in the case of illness A_3. The analysis was carried out five times and yielded a positive result four times and a negative result once. It is required that one find the probability of each of the illnesses after the analysis.

In the case of illness A_1, the probability of the indicated results of the analyses is equal, by the multiplication rule, to $p_1=C_5^4(0.1)^4\cdot 0.9$. For the second hypothesis, this probability equals $p_2=C_5^4(0.2)^4\cdot 0.8$ and for the third it is equal to $p_3=C_5^4(0.9)^4\cdot 0.1$.

According to Bayes's formula, we find that after the analyses the probability of illness A_1 turns out to be equal to

$$\frac{P_1p_1}{P_1p_1+P_2p_2+P_3p_3}$$
$$=\frac{(1/2)\cdot(0.1)^4\cdot 0.9}{(1/2)\cdot(0.1)^4\cdot 0.9+(1/6)\cdot(0.2)^4\cdot 0.8+(1/3)\cdot(0.9)^4\cdot 0.1}\approx 0.002;$$

the probability of illness A_2 is

$$\frac{P_2p_2}{P_1p_1+P_2p_2+P_3p_3}$$
$$=\frac{(1/6)\cdot(0.2)^4\cdot 0.8}{(1/2)\cdot(0.1)^4\cdot 0.9+(1/6)\cdot(0.2)^4\cdot 0.8+(1/3)\cdot(0.9)^4\cdot 0.1}\approx 0.01;$$

and for illness A_3 it is

$$\frac{P_3p_3}{P_1p_1+P_2p_2+P_3p_3}$$
$$=\frac{(1/3)\cdot(0.9)^4\cdot 0.1}{(1/2)\cdot(0.1)^4\cdot 0.9+(1/6)\cdot(0.2)^4\cdot 0.8+(1/3)\cdot(0.9)^4\cdot 0.1}\approx 0.988.$$

Since these three events A_1, A_2, A_3 form, even after the test, a complete system of events, we can as a check on the calculation carried out add the three numbers obtained and verify that their sum is equal to unity, as before.

CHAPTER 5

BERNOULLI'S SCHEME

§ 13. Examples

Example 1. Among fibers of cotton of a definite sort 75% on the average have lengths less than 45 mm. and 25% have lengths greater than (or equal to) 45 mm. Find the probability that of three fibers taken at random two will be shorter than and one will be longer than 45 mm.

We denote the event of choosing a fiber of length less than 45 mm. by A and the event of choosing a fiber of length greater than 45 mm. by B; it is then clear that

$$P(A) = 3/4; \quad P(B) = 1/4.$$

We shall further agree to denote the following compound event by AAB: the first two fibers chosen are shorter than 45 mm. and the third fiber is longer than 45 mm. It is clear what the meaning of the schemes BBA, ABA, and so on, will be. Our problem is to compute the probability of the event C: that of three fibers two are shorter than 45 mm. and one fiber is longer than 45 mm. Evidently, for this to happen one of the following schemes must be realized:

$$AAB, ABA, BAA. \tag{1}$$

Since any two of these three results are mutually incompatible we have, by the addition rule

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

All three terms in the right member are equal inasmuch as the results of the choice of the fibers can be assumed to be mutually independent events. The probability of each of the schemes (1), according to the multiplication rule for probabilities of independent events, is representable as the product of three factors of which two equal $P(A) = 3/4$ and one equals $P(B) = 1/4$. Thus, the probability of each of the three schemes (1) equals

$$(3/4)^2 \cdot (1/4) = 9/64,$$

and, consequently,

$$P(C) = 3 \cdot (9/64) = 27/64,$$

which is the solution of our problem.

Example 2. As the result of observations extending over many decades it was found that of every 1000 newly born children on the average there are born 515 boys and 485 girls. In a certain family there are six children. Find the probability that there are no more than two girls among them.

For the occurrence of the event whose probability we are seeking, it is necessary that there be either 0 or 1 or 2 girls. The probabilities of these particular events will be denoted by P_0, P_1, P_2, respectively. It is clear that, according to the rule for the addition of probabilities, the probability sought is

$$P = P_0+P_1+P_2. \tag{2}$$

For each child, the probability that it is a boy equals 0.515 and, hence, the probability that it is a girl equals 0.485.

P_0 is the easiest to find; this is the probability that all the children in the family are boys. Since the birth of a child of either sex can be considered as independent of the sex of the remaining children, the probability, according to the rule for the multiplication of probabilities, that all six children are boys is equal to the product of six factors each equal to 0.515, i.e.,

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

We now go over to the calculation of P_1, i.e., the probability that of the six children in the family one child is a girl and the remaining five are boys. This event can occur in six different ways depending on which child in the order of birth is a girl (i.e., first, second, etc.). We consider any of the possible ways of this event, for example the one that a girl is born as the fourth child. The probability of this possibility, according to the multiplication rule, equals the product of six factors of which five equal 0.515 and the sixth (situated in the fourth place) equals 0.485; i.e., this probability equals $(0.515)^5 \cdot 0.485$. This is also the probability of each of the other five possibilities of the event which interests us at the moment; therefore, the probability P_1 of this event is equal, according to the addition rule, to the sum of six numbers each equal to $(0.515)^5 \cdot 0.485$, i.e.,

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

We now turn to the calculation of P_2 (i.e., the probability that two of the children are girls and four are boys). Analogous to what precedes, we at once note that this event admits of a whole series of possibilities. One of the possibilities will be, for instance, the following: the second and fifth child in order of birth are girls and the remainder are boys. The probability of each of the possibilities, according to the multiplication rule, equals $(0.515)^4 \cdot (0.485)^2$ and, consequently, P_2 equals, by the addition rule, the number $(0.515)^4 \cdot (0.485)^2$, multiplied by the number of all possibilities of the type considered; the entire problem thus reduces to the determination of this last number.

Each of the possibilities is characterized by the fact that of six children two are girls and the remainder are boys; the number of different possibilities consequently equals the number of distinct choices of two children from the six at hand. The number of such choices equals the number of combinations of six distinct objects taken two at a time; i.e., $C_6^2 = (6 \cdot 5)/(2 \cdot 1) = 15$. Thus,

$$P_2 = C_6^2 \cdot (0.515)^4 \cdot (0.485)^2 = 15 \cdot (0.515)^4 \cdot (0.485)^2 \approx 0.247.$$

Combining the results obtained above, we have

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

Thus, in about 37% of the families having six children we will find fewer than three girls and, hence, more than three boys among the children.

§ 14. The Bernoulli formulas

In the preceding section, we became acquainted by means of a number of examples with the scheme of *repeated trials*, in each of which an event A can be realized. We attribute a very broad and variegated sense to the word "trial." Thus, if we fire at a certain target, by a trial we shall understand each individual shot. If we test electric light bulbs for length of burning, then a trial will be understood to be the testing of each bulb. If we are studying the composition of newly born children by sex, weight, or height, then a trial will be understood to be the investigation of an individual child. In general, by a trial we shall in what follows understand the realization of certain conditions in the presence of which some event of interest to us can occur.

We have now arrived at the consideration of one of the important schemes in the theory of probability having, besides application in various branches of knowledge, great significance also in probability theory itself as a mathematical science. This scheme consists in

considering a sequence of mutually independent trials, i.e., of such trials for which the probability of some result in each of them does not depend on what results occurred or will occur in the remainder. In each of these trials, there can occur (or not occur) some event A with probability p which does not depend on the number of trials. The scheme just described has received the name *Bernoulli scheme* since the origin of its systematic study can be traced back to the renowned Swiss mathematician Jacob Bernoulli, who lived at the end of the seventeenth century.

We have already dealt with the Bernoulli scheme in our examples; in order to convince ourselves of this, it is sufficient to recall the examples of the preceding section. We shall now solve the following general problem; all the examples we considered up to this point in this chapter were particular cases of this.

Problem. Under certain conditions, the probability that the event A occurs in every trial equals p; find the probability that a sequence of n independent trials yields k occurrences and $n-k$ nonoccurrences of the event A.

The event whose probability is sought splits into a number of possibilities; in order to obtain one definite possibility, we must arbitrarily choose from the given sequence any k trials and assume that the event A occurred for precisely these k trials and that A did not occur for the remaining $n-k$. Thus, every such possibility requires the occurrence of n definite results—in this number k occurrences and $n-k$ nonoccurrences of the event A. By the multiplication rule, we find that the probability of each definite possibility equals

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

The number of different possibilities equals the number of different sets of k trials each of which can be constructed from n distinct trials, i.e., it is equal to C_n^k. Applying the addition rule and the known formula for the number of combinations of n objects taken k at a time,

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

we find that the probability of k occurrences of the event A with n independent trials equals

$$P_n(k) = \frac{n(n-1)\ldots[n-(k-1)]}{k(k-1)\ldots 2\cdot 1} p^k(1-p)^{n-k}, \qquad (3)$$

which solves our problem. It is frequently more convenient to represent the expression C_n^k in a somewhat different form; multiplying the numerator and denominator by the product $(n-k)[n-(k+1)] \ldots 2\cdot 1$, we obtain

$$C_n^k = \frac{n(n-1)\ldots 2\cdot 1}{k(k-1)\ldots 2\cdot 1(n-k)\cdot[n-(k+1)]\ldots 2\cdot 1},$$

or, denoting for brevity the product of all integers from 1 to m inclusively by $m!$,

$$C_n^k = \frac{n!}{k!(n-k)!}.$$

For $P_n(k)$, this yields

$$P_n(k) = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}. \tag{4}$$

Formulas (3) and (4) are usually called *Bernoulli's formulas.* For large values of n and k, the computation of $P_n(k)$ according to these formulas is rather difficult since the factorials $n!$, $k!$, $(n-k)!$ are very large numbers which are rather cumbersome to evaluate. Therefore, in calculations of this type specially compiled tables of factorials as well as various approximation formulas are extensively used.

EXAMPLE. The probability that the consumption of water at a certain factory is normal (i.e., it is not more than a prescribed number of liters every twenty-four hours) equals 3/4. Find the probability that in the next 6 days the consumption of water will be normal in the course of 0, 1, 2, 3, 4, 5, 6 days.

Denoting by $P_6(k)$ the probability that in the course of k days out of 6 the consumption of water will be normal, we find, by formula (3) (where we must set $p=3/4$), that

$$P_6(6) = (3/4)^6 = 3^6/4^6,$$

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

$$P_6(4) = C_6^4\cdot(3/4)^4\cdot(1/4)^2 = C_6^2\frac{3^4}{4^6} = \frac{6\cdot 5}{2\cdot 1}\cdot\frac{3^4}{4^6} = \frac{15\cdot 3^4}{4^6},$$

$$P_6(3) = C_6^3\cdot(3/4)^3\cdot(1/4)^3 = \frac{6\cdot 5\cdot 4}{3\cdot 2\cdot 1}\cdot\frac{3^3}{4^6} = \frac{20\cdot 3^3}{4^6},$$

$$P_6(2) = \frac{6\cdot 5}{2\cdot 1}(3/4)^2\cdot(1/4)^4 = \frac{15\cdot 3^2}{4^6},$$

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

finally, we evidently have $P_6(0)$ (i.e., the probability that there is excessive consumption in each of the 6 days) equal to $1/4^6$. All six

probabilities are expressed as fractions with the same denominator, $4^6=4096$; we use this, of course, to shorten our calculations. These yield

$$P_6(6) \approx 0.18;\quad P_6(5) \approx 0.36;\quad P_6(4) = 0.30;$$
$$P_6(3) \approx 0.13;\quad P_6(2) \approx 0.03;\quad P_6(1) = P_6(0) \approx 0.00.$$

We see that it is most probable that there will be an excessive consumption of water in the course of one or two days of the six and that the probability of excessive consumption in the course of five or six days, i.e., $P_6(1)+P_6(0)$, practically equals zero.

§ 15. The most probable number of occurrences of an event

The example which we just considered shows that the probability of a normal consumption of water in the course of exactly k days with increasing k at first increases and then, having attained its largest value, begins to decrease; this is most clearly seen if the variation of the probability $P_6(k)$ with increasing k is expressed geometrically in the form of a diagram, shown in Fig. 4. A still clearer picture is

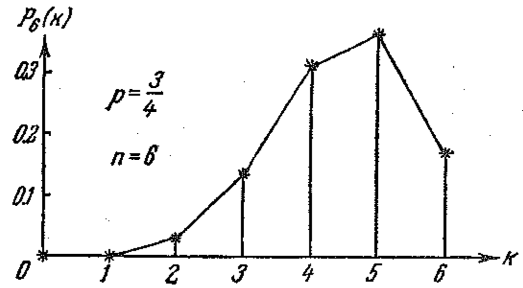

FIG. 4

given by diagrams of the variation of the quantity $P_n(k)$ as k increases when the number n becomes larger; thus, for $n=15$ and $p=1/2$, the diagram has the form shown in Fig. 5.

In practice, it is sometimes required to know what number of occurrences of the event is *most probable*, i.e., for what number k the probability $P_n(k)$ is the largest. (In this connection, it is, of course, assumed that p and n are prescribed.) The Bernoulli formulas allow us in all cases to find a simple solution of this problem; we shall occupy ourselves with this now.

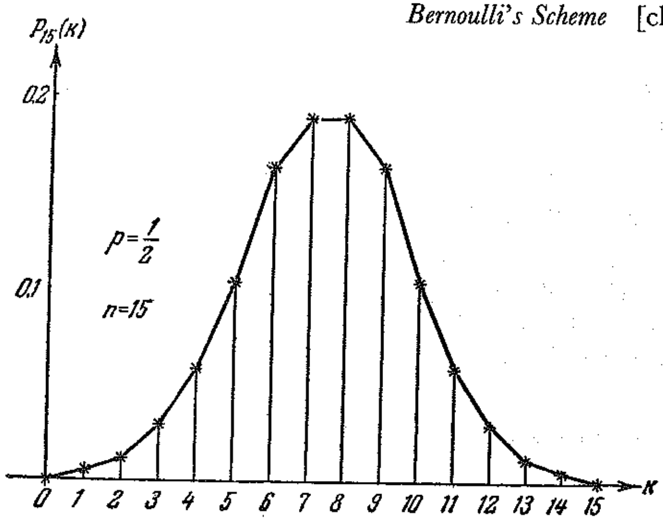

FIG. 5

We first calculate the magnitude of the ratio $P_n(k+1)/P_n(k)$. By virtue of formula (4),

$$P_n(k+1) = \frac{n!}{(k+1)!(n-k-1)!} p^{k+1}(1-p)^{n-k-1}, \qquad (5)$$

and, from formulas (3) and (5), we have

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

The probability $P_n(k+1)$ will be larger than, equal to, or less than the probability $P_n(k)$ depending on whether or not the ratio $P_n(k+1)/P_n(k)$ is larger than, equal to, or less than unity, and the latter, as we see, reduces to the question of which of the three relations

$$\frac{n-k}{k+1}\cdot\frac{p}{1-p} > 1, \quad \frac{n-k}{k+1}\cdot\frac{p}{1-p} = 1, \quad \frac{n-k}{k+1}\cdot\frac{p}{1-p} < 1 \qquad (6)$$

is valid. If we wish, for example, to determine the values of k for which the inequality $P_n(k+1) > P_n(k)$ is satisfied, then we must recognize for what values of k the inequality

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

or

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

holds. From this we obtain

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

thus, as long as k increases but does not attain the value $np-(1-p)$, we will always have $P_n(k+1) > P_n(k)$. Thus, with increasing k, the probability $P_n(k)$ will always increase. For example, in the scheme to which the diagram in Fig. 5 corresponds, we have $p=1/2$, $n=15$, $np-(1-p)=7$; this means that as long as $k<7$ (i.e., for all k from 0 to 6 inclusively), we have $P_n(k+1) > P_n(k)$. The diagram substantiates this.

In precisely the same way, starting with the other two relations in (6), we find that

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

and

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

thus, as soon as the number k exceeds the bound $np-(1-p)$, the probability $P_n(k)$ begins to decrease and will decrease to $P_n(n)$.

This derivation first of all convinces us that the behavior of the quantity $P_n(k)$ considered by us in the examples is a general law which holds in all cases: as the number k increases, $P_n(k)$ first increases and then decreases. But, more than this, this result also allows us to solve quickly the problem we have set for ourselves—i.e., to determine the most probable value of the number k. We denote this most probable value of the number k by k_0. Then

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

from which it follows, according to what precedes, that

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

On the other hand,

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

from which, according to what precedes, the inequality

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

or

$$k_0 \le np-(1-p)+1 = np+p$$

must hold. Thus, the most probable value k_0 of the number k must satisfy the double inequality

$$np-(1-p) \le k_0 \le np+p. \qquad (7)$$

The interval from $np-(1-p)$ to $np+p$, in which the number k_0 must therefore lie, has length 1 as can be shown by a simple calculation; therefore, if either of the endpoints of this interval, for instance the number $np-(1-p)$, is not an integer, then between these endpoints there will necessarily lie one, and only one, integer and k_0 will be uniquely determined. We ought to consider this case to be normal; for, p is less than 1, and therefore only in exceptional cases will the quantity $np-(1-p)$ be an integer. In this exceptional case, inequalities (7) yield two values for the number k_0: $np-(1-p)$ and $np+p$, which differ from one another by unity. Those two values will also be the most probable; their probabilities will be equal and exceed the probability of any other value of the number k. This exceptional case holds, for instance, in the scheme expressed by the diagram in Fig. 5; here, $n=15$, $p=1/2$ and hence $np-(1-p)=7$, $np+p=8$; the numbers 7 and 8 serve as the most probable values of the number k of occurrences of the event; their probabilities are equal to one another, each of them being approximately equal to 0.196. (All this can be seen on the diagram.)

EXAMPLE 1. As the result of observations over a period of many years, it was discovered, for a certain region, that the probability that rain falls on July 1 equals 4/17. Find the most probable number of rainy July 1's for the next 50 years. Here, $n=50$, $p=4/17$, and

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

As this number turned out to be an integer, it means we are dealing with the exceptional case; the most probable value of the number of rainy days will be the numbers 11 and 12 which are equally probable.

EXAMPLE 2. In a physics experiment, particles of a prescribed type are being observed. Under fixed conditions, during an interval of time of definite length, on the average 60 particles appear and each of them has—with a probability 0.7—a velocity greater than v_0. Under other conditions, during the same interval of time there appear on the average only 50 particles, but for each of them the probability of having a velocity exceeding v_0 equals 0.8. Under what conditions of the experiment will the most probable number of particles having a velocity exceeding v_0 be the greatest?

Under the first conditions of the experiment,

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

For the second conditions of the experiment,

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

We see that the most probable number of "fast" particles under the first conditions of the experiment is somewhat larger than under the second.

In practice, we often encounter the situation when the number n is very large; e.g., in the case of mass firing, the mass production of

articles, and so on. In this case the product np will also be a very large number provided the probability p is not unusually small. And since in the expressions $np-(1-p)$ and $np+p$, between which lie the most probable number of occurrences of the event, the quantities p and $1-p$ are less than unity, we see that both these expressions and hence the most probable number of occurrences of the event are all close to np. Thus, if the probability of completing a telephone connection in less than 15 seconds equals 0.74, then we can take $1000\cdot 0.74$ as the most probable number of connections, among every 1000 calls coming into the central exchange, made in less than 15 seconds.

This result can be given a still more precise form. If k_0 denotes the most probable number of occurrences of the event in n trials, then k_0/n is the most probable "fraction" of occurrences of the event for the same n trials; inequalities (7) yield

$$p-\frac{1-p}{n} \leq \frac{k_0}{n} \leq p+\frac{p}{n}. \tag{8}$$

Let us assume that, leaving the probability p of the occurrence of the event for an individual trial invariant, we shall increase indefinitely the number of trials n. (In this connection we, of course, also increase the most probable number of occurrences k_0.) The fractions $(1-p)/n$ and p/n, appearing in the left and right members of the inequalities (8) above will become smaller and smaller; this means that, for large n, these fractions can be disregarded. We can now consider both the left and right members of the inequalities (8) and hence also the fraction k_0/n contained between them to be equal to p. Thus, *the most probable ratio of occurrences of the event—provided there are a large number of trials—is practically equal to the probability of the occurrence of the event in an individual trial.*

For example, if for certain measurements the problem of making in an individual measurement an error comprised between α and β equals 0.84, then for a large number of measurements one can expect with the greatest probability errors comprised between α and β in approximately 84% of the cases. This does not mean, of course, that the probability of obtaining exactly 84% of such errors will be large; on the contrary, this "largest probability" itself will be very small in a large number of measurements (thus, we saw in the scheme in Fig. 5 that the largest probability turned out to be equal to 0.196 where we were dealing with 15 trials altogether; for a large number of trials it is significantly less). This probability is the largest only in the comparative sense: the probability of obtaining 84% of the measurements with errors comprised between α and β is larger than the probability of obtaining 83% or 86% of such measurements.

On the other hand, it is easily understandable that in extended series of measurements the probability of a certain individual number of errors of a given quantity cannot be of significant interest. For example, if we carry out 200 measurements, then it is doubtful whether it is expedient to calculate the probability that exactly 137 of them will be measurements with the prescribed precision because in practice it is immaterial whether the number is 137 or 136 or 138 or even, for instance, 140. In contrast, questions of the probability that the number of measurements for which the error is between prescribed bounds will be more than 100 of the 200 measurements made or that this number will be somewhere between 100 and 125 or that it will be less than 50, and so on, are certainly of practical interest. How should we express this type of probability? Suppose we wish, for example, to find the probability that the number of measurements will be between 100 and 120 (including 120); more specifically, we will seek the probability of satisfying the inequality

$$100 < k \leq 120,$$

where k is the number of measurements. For these inequalities to be realized, it is necessary that k be equal to one of the twenty numbers 101, 102, . . ., 120. According to the addition rule, this probability equals

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

To calculate this sum directly, we would have first to compute 20 individual probabilities of the type $P_n(k)$ according to formula (3); for such large numbers, such calculations present insurmountable difficulties. Therefore, sums of the form obtained are never computed by means of direct calculations in practice. For this purpose there exist suitable approximation formulas and tables. The composition of these formulas and tables is based on complicated methods of mathematical analysis, which we shall not touch upon here. However, concerning probabilities of the type $P(100 < k \leq 120)$ one can obtain information by simple lines of reasoning in many cases which lead to the complete solution of the problem posed. We shall discuss this problem in the following chapter.

CHAPTER 6

BERNOULLI'S THEOREM

§ 16. Content of Bernoulli's theorem

Let us take another good look at the diagram in Fig. 5 (on page 44), where the probabilities of various values of the number k of occurrences of the event under consideration are the numbers $P_{15}(k)$, which are depicted by the vertical lines. The probability assigned to some *segment* of values of k (the probability that the *number* of occurrences of the event of interest to us turns out to be equal to some one of the numbers of this segment) is equal, according to the addition rule, to the sum of the probabilities of all the numbers of this segment; i.e., it is equal to the sum of the lengths of all vertical lines situated over this segment. Pictorially, the figure shows that this sum is quite different for various segments of the same length. Thus, the segments $2 \leq k < 5$ and $7 \leq k < 10$ have the same length; the probability of each of them is expressed by the sum of the lengths of three vertical lines, and we see that for the second segment this sum is significantly larger than for the first. We already know that the diagrams of the probabilities $P_n(k)$ have, for all n, basically, the same form as the diagram in Fig. 5; i.e., the quantity $P_n(k)$ at first increases with increasing k and then, after passing through its largest value, it decreases. It is therefore clear that of the two segments of values of the number k having the same length, the one situated nearer the most probable value, k_0, will in all cases have the largest probability. In particular, on the segment having the number k_0 as its center we will always have a greater probability than on any other segment of the same length.

But it turns out that much more can be said in this regard. There are in all $n+1$ possible values of the number k of occurrences of the event in n trials ($0 \leq k \leq n$). We take the segment having center at k_0 and containing only a small fractional part, for example one hundredth, of the possible values of the number k. It then turns out that if the total number n of trials is very large, the predominant probability will correspond to this segment and all other values of the number k taken together have a negligibly small probability. Thus, although the segment we chose is negligibly small in comparison with n (on the figure it occupies in all a one-hundredth part of the entire length of the diagram), nevertheless, the sum of the vertical lines situated over it will be significantly larger than the sum of all remaining vertical lines. The reason for this lies in the fact that the lines in the central part of the diagram are many times larger than the lines situated near the ends. Thus, for large n the diagram of the quantity $P_n(k)$ has a form which is approximately that shown in Fig. 6.

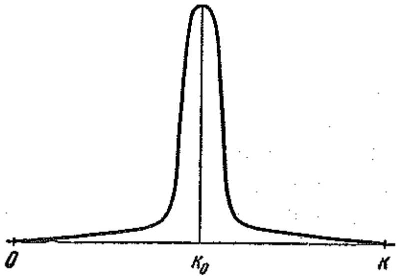

Fig. 6

In practice, this obviously means the following: *if we perform a series of a large number* n *of trials, then we can expect with a probability close to unity that the number* k *of occurrences of the event* A *will be very close to its most probable value, differing from the latter only by an insignificant fractional part of the total number* n *of trials made.*

This proposition, known under the name of *Bernoulli's theorem* and discovered at the beginning of the eighteenth century, is one of the important laws of probability theory. Up to the middle of the last century, all proofs of this theorem required complicated mathematical means and the great Russian mathematician P. L. Chebyshev was the first to find a very simple and short derivation of this law; we now present Chebyshev's remarkable proof.

§ 17. Proof of Bernoulli's theorem

We already know that for a large number n of trials, the most probable number k_0 of occurrences of the event A differs very little from the quantity np, where p, as always, denotes the probability of the event A for an individual trial. It is therefore sufficient for us to prove that, for a large number of trials, with very high probability the number k of occurrences of the event A will differ from np by very little —by not more than an arbitrarily small fractional part of the number n (not more, for example, than by 0.01 n or 0.001 n, or, in general, not

more than by εn where ε is an arbitrarily small number). In other words, we must show that the probability

$$P(|k-np| > \varepsilon n) \tag{1}$$

will be as small as we please for sufficiently large n.

In order to verify this, we note that according to the law of addition, probability (1) equals the sum of the probabilities $P_n(k)$ for all those

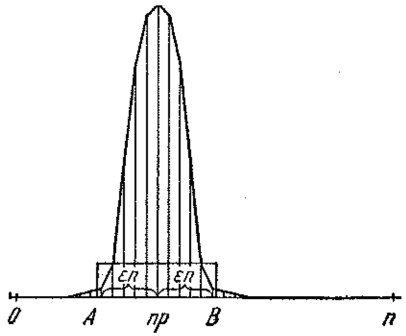

Fig. 7

values of the number k which lie at a distance not more than εn from np; in our typical diagram (Fig. 7), this sum is expressed by the sum of the lengths of all vertical lines lying exterior to the segment $\overline{AB}$—to the right as well as to the left of it. Since the sum total of all the vertical lines (being the sum of the probabilities of a complete system of events) equals unity, this means that the overwhelming portion (almost equal to unity) of this sum corresponds to the segment $\overline{AB}$ and only a negligibly small part of it corresponds to the regions lying exterior to this segment.

Thus,

$$P(|k-np| > \varepsilon n) = \sum_{|k-np|>\varepsilon n} P_n(k). \tag{2}$$

We now turn to Chebyshev's line of reasoning. Since in every term of the sum written down we have

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

and hence

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

we can only increase this sum if each of its terms $P_n(k)$ is replaced by the expression

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

Therefore,

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

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

Furthermore, it is obvious that the last sum is increased still more if further new terms are added to the terms it already has, forcing the number k to range over not only the parts to the left of $np-\varepsilon n$ and to the right of $np+\varepsilon n$, but over the entire series of values which are possible for it, i.e., the entire series of numbers from 0 to n inclusive. We thus obtain, a fortiori,

$$P(|k-np| > \varepsilon n) < \frac{1}{\varepsilon^2 n^2} \sum_{k=0}^{n} (k-np)^2 P_n(k). \qquad (3)$$

The latter sum differs advantageously from all the preceding sums in that it can be computed precisely; the Chebyshev method thus consists of replacing sum (2), which is difficult to estimate, by the sum (3), which admits of an exact computation.

We now proceed to make this calculation; no matter how long it may appear to take us, these are simply difficulties of a technical nature which anyone who knows algebra can handle. The remarkable idea of Chebyshev has already been completely utilized by us, as it consisted, namely, in the transition from equality (2) to inequality (3).

First of all, we easily find that

$$\sum_{k=0}^{n} (k-np)^2 P_n(k) = \sum_{k=0}^{n} k^2 P_n(k) - 2np \sum_{k=0}^{n} kP_n(k) + n^2p^2 \sum_{k=0}^{n} P_n(k). \qquad (4)$$

Of the three sums in the right member, the last is equal to unity since it is the sum of the probabilities of a complete system of events. This means that it only remains for us to calculate the sums

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

In this connection, in both sums the terms corresponding to $k=0$ are equal to zero so that one can start the summation with $k=1$.

1) To calculate both sums, we express $P_n(k)$ according to formula (4), Chapter 5 (see page 42). We find that

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

since, obviously, $n! = n(n-1)!$ and $k! = k(k-1)!$, we find that

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

or, setting $k-1=l$ in the sum in the right member and noting that l varies from 0 to $n-1$ as k varies from 1 to n,

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

$$= np \sum_{l=0}^{n-1} P_{n-1}(l).$$

The last sum, i.e., $\sum_{l=0}^{n-1} P_{n-1}(l)$, of course, equals unity because it is the sum of the probabilities of a complete system of events—all possible numbers of occurrences of the event l for $n-1$ trials. Thus, for the sum $\sum_{k=0}^{n} kP(k)$, we obtain the very simple expression

$$\sum_{k=0}^{n} kP_n(k) = np. \qquad (5)$$

2) To calculate the second sum, we first find the quantity $\sum_{k=1}^{n} k(k-1)P_n(k)$; since the term corresponding to $k=1$ is obviously equal to zero, the summation can begin with the value $k=2$. Noting that $n! = n(n-1)(n-2)!$ and that $k! = k(k-1)(k-2)!$, we easily conclude, setting $k-2=m$, similarly to what we did before, that

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

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

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

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

$$= n(n-1)p^2 \sum_{m=0}^{n-2} P_{n-2}(m) = n(n-1)p^2, \qquad (6)$$

because the last sum is again equal to unity being the sum of the probabilities of a complete system of events—all possible numbers of occurrences of the events for $n-2$ trials.

Finally, formulas (5) and (6) yield

$$\sum_{k=1}^{n} k^2 P_n(k) = \sum_{k=1}^{n} k(k-1)P_n(k) + \sum_{k=1}^{n} kP_n(k)$$
$$= n(n-1)p^2 + np = n^2p^2 + np(1-p). \qquad (7)$$

Now, both of the sums that we needed have been computed. Substituting results (5) and (7) into relation (4), we find finally that

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

Substituting this simple expression we just derived into inequality (3), we obtain

$$P(|k-np| > \varepsilon n) < \frac{np(1-p)}{\varepsilon^2 n^2} = \frac{p(1-p)}{\varepsilon^2 n}. \qquad (8)$$

This inequality completes the proof of everything required. In fact, it is true that we could have taken the number ε arbitrarily small; however, having chosen it, we do not change it any more. But the number n of trials in the sense of our assertion can be arbitrarily large. Therefore, the fraction $p(1-p)/(\varepsilon^2 n)$ can be assumed to be as small as we please, since with increasing n its denominator can be made arbitrarily large whereas the numerator at the same time remains unchanged.

For example, let $p = 0.75$, so that

$$1-p = 0.25 \quad \text{and} \quad p(1-p) = 0.1875 < 0.2;$$

choose $\varepsilon = 0.01$; then inequality (8) yields

$$P\left(\left|k - \frac{3}{4}n\right| > 0.01n\right) < \frac{0.2}{0.0001 \cdot n} = \frac{2000}{n}.$$

If, for instance, we take $n = 200{,}000$, then

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

In practice, this means, for example, the following: if in some production process, under fixed operating conditions, 75% on the average of the articles possess a certain property (for example, they belong to the first sort), then of 200,000 articles, from 148,000 to 152,000 articles will possess this property with a probability exceeding 0.99 (i.e., almost certainly).

In regard to this matter we must make two observations:

1. Inequality (8) yields a very rough estimate of the probability $P(|k-np| > \varepsilon n)$; in fact, this probability is significantly smaller—especially for large values of n. In practice, we therefore make use of more precise estimates whose derivation is, however, considerably more complicated.

2. The estimate, given by inequality (8), becomes significantly more precise when the probability p is very small—or just the opposite—very close to unity. Thus, if in the example we have just introduced, the probability that the article possesses a certain property equals $p = 0.95$, then $1-p = 0.05$, and $p(1-p) < 0.05$. Therefore, choosing $\varepsilon = 0.005$, $n = 200{,}000$, we find that

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

just as before. But now εn is not equal to 2000 but only to 1000; from this (since $np = 190{,}000$) we conclude that with practical certainty the number of articles possessing the property under consideration will, for a total number of 200,000 articles, lie between 189,000 and 191,000. Thus, inequality (8) practically guarantees us that the number of articles possessing the property concerned will be in an interval for $p = 0.95$ of half the length of that for $p = 0.75$, because we have here

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

PROBLEM. It is known that one-fourth of the workers in a particular branch of industry have an elementary school education. For a certain investigation, 200,000 workers are chosen at random. Find 1) the most probable value of the number of workers with an elementary school education among the 200,000 workers chosen and 2) the probability that the true (actual) number of such workers deviates from the most probable number by no more than 1.6%.

In the solution of this problem, we start with the fact that the probability of having an elementary education equals one-fourth for each of the 200,000 workers chosen at random. (This is precisely the key to the meaning of the phrase "at random.") Thus, in our problem, we have

$$n = 200{,}000, \quad p = 1/4, \quad k_0 = np = 50{,}000, \quad p(1-p) = 3/16.$$

We are seeking the probability that $|k-np| < 0.016np$ or that $|k-np| < 800$, where k is the number of workers with an elementary school education. We choose ε so as to have $\varepsilon n = 800$; from this we find that $\varepsilon = 800/n = 0.004$. Formula (8) yields

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

from which it follows that

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

Answer. The most probable value, which is what we are looking for, equals 50,000; the probability sought is greater than 0.94. (Actually, the probability sought is significantly closer to unity.)

PART II

RANDOM VARIABLES