INTRODUCTION TO
MATHEMATICAL PROBABILITY

INTRODUCTION

TO

MATHEMATICAL PROBABILITY

BY

J. V. USPENSKY

Professor of Mathematics, Stanford University

McGRAW-HILL BOOK COMPANY, INC.

NEW YORK AND LONDON

1937

PRINTED IN THE UNITED STATES OF AMERICA

IX

PREFACE

This book is an outgrowth of lectures on the theory of probability which the author has given at Stanford University for a number of years. At first a short mimeographed text covering only the elementary parts of the subject was used for the guidance of students. As time went on and the scope of the course was gradually enlarged, the necessity arose of putting into the hands of students a more elaborate exposition of the most important parts of the theory of probability. Accordingly a rather large manuscript was prepared for this purpose. The author did not plan at first to publish it, but students and other persons who had opportunity to peruse the manuscript were so persuasive that publication was finally arranged.

The book is arranged in such a way that the first part of it, consisting of Chapters I to XII inclusive, is accessible to a person without advanced mathematical knowledge. Chapters VII and VIII are, perhaps, exceptions. The analysis in Chapter VII is rather involved and a better way to arrive at the same results would be very desirable. At any rate, a reader who does not have time or inclination to go through all the intricacies of this analysis may skip it and retain only the final results, found in Section 11. Chapter VIII, though dealing with interesting and historically important problems, is not important in itself and may without loss be omitted by readers. Chapters XIII to XVI incorporate the results of modern investigations. Naturally they are more complex and require more mature mathematical preparation.

Three appendices are added to the book. Of these the second is by far the most important. It gives an outline of the famous Tshebysheff-Markoff method of moments applied to the proof of the fundamental theorem previously established by another method in Chapter XIV.

No one will dispute Newton's assertion: "In scientiis addiscendis exempla magis prosunt quam praecepta." But especially is it so in the theory of probability. Accordingly, not only are a large number of illustrative problems discussed in the text, but at the end of each chapter a selection of problems is added for the benefit of students. Some of them are mere examples. Others are more difficult problems, or even important theorems which did not find a place in the main text. In all such cases sufficiently explicit indications of solution (or proofs) are given.

The book does not go into applications of probability to other sciences. To present these applications adequately another volume of perhaps larger size would be required.

No one is more aware than the author of the many imperfections in the plan of this book and its execution. To present an entirely satisfactory book on probability is, indeed, a difficult task. But even with all these imperfections we hope that the book will prove useful, especially since it contains much material not to be found in other books on the same subject in the English language.

J. V. USPENSKY.

STANFORD UNIVERSITY,
September, 1937.

CONTENTS

	Page
Preface	v
Introduction	1
Scope of Probability—Necessity and Chance—Genesis of Idea of Probability—Classical Definition of Mathematical Probability—Its Limitations—Short Sketch of Historical Development of Theory of Probability.	
CHAPTER I	
Computation of Probability by Direct Enumeration of Cases	14
Simple Problems Solved by Direct Application of Definition—Permutations, Arrangements and Combinations—Some Illustrative Problems—Problems for Solution.	
CHAPTER II	
Theorems of Total and Compound Probability	27
Mutually Exclusive or Incompatible Events—Theorem of Total Probability in Restricted and Generalized Form—Conditional Probability—Theorem of Compound Probability—Independent Events—Compound Probability of Independent Events—Problems Illustrating the Combined Use of Theorem of Total and Compound Probabilities—Problems for Solution.	
CHAPTER III	
Repeated Trials	44
Independent and Dependent Trials—Bernoullian Trials—Most Probable Number of Successes—Approximate Formulas in Case of a Large Number of Trials—Markoff's Method of Continued Fractions—Problems for Solution.	
CHAPTER IV	
Probabilities of Hypotheses and Bayes' Theorem	60
Bayes' Formula for Probability a Posteriori—Probability of Future Events—Estimation of Probabilities from Observed Frequencies—Criticism of Underlying Assumptions—Problems for Solution.	
CHAPTER V	
Use of Difference Equations in Solving Problems of Probability	74
Simple Chains of Trials—Problem of Runs—Examples Requiring Use of Partial Difference Equations—Problems for Solution.	
CHAPTER VI	
Bernoulli's Theorem	96
Statement and Proof of Bernoulli's Theorem—Theorem of Cantelli—Significance of Bernoulli's Theorem—Application to Games of Chance—	

PAGE

Experimental Verification of Bernoulli's Theorem—Appendix: Buffon's Needle Problem—Problems for Solution.

CHAPTER VII

APPROXIMATE EVALUATION OF PROBABILITIES IN BERNOULLIAN CASE 119

Statement of Problem—Analytical Representation of Required Probability—Analysis Leading to Final Approximate Formula—Numerical Examples—Problems for Solution.

CHAPTER VIII

FURTHER CONSIDERATIONS ON GAMES OF CHANCE. 139

Problems Concerning Ruin of Gamblers When Number of Games is Unlimited—Stakes Equal—Case of Unequal Stakes—Probability of Ruin in Case of a Limited Number of Games—Problems for Solution.

CHAPTER IX

MATHEMATICAL EXPECTATION. 161

Stochastic Variables—Definition of Mathematical Expectation—Mathematical Expectation of Sums—Examples—Mathematical Expectation of Products—Dispersion and Standard Deviation—Dispersion of Sums—Examples—Problems for Solution.

CHAPTER X

THE LAW OF LARGE NUMBERS 182

Tshebysheff's Lemma—Law of Large Numbers—Sufficient Conditions for Its Validity—Some Experimental Results—Case of Infinite Dispersion: Theorem of Markoff—Law of Large Numbers for Identical Variables: Theorem of Khintchine—Problems for Solution.

CHAPTER XI

APPLICATION OF THE LAW OF LARGE NUMBERS 208

Poisson's Theorem—Divergence Coefficient—Lexis' and Poisson's Cases—Tschuprow's Theorem—Examples of Dependent Trials—Experimental Results for Simple Chains—Problems for Solution.

CHAPTER XII

PROBABILITIES IN CONTINUUM 235

Notion of Probability for Continuous Variables—Case of One Variable—Density of Probability—Distribution Function, Moments and Characteristic Function—Examples—Case of Two or More Variables—Geometrical Problems—Problems for Solution.

CHAPTER XIII

THE GENERAL CONCEPT OF DISTRIBUTION 260

Increasing Functions and Stieltjes' Integrals—Representation of Distribution Function by Stieltjes' Integral—Moments and Characteristic Function—Inequalities for Moments—Composition of Distribution Functions—Examples—Determination of Distribution When Its Characteristic Function Is Known—Composition of Characteristic Functions—Examples—Problems for Solution.

CHAPTER XIV

PAGE

FUNDAMENTAL LIMIT THEOREMS 283

Introductory Remarks—Laplace-Liapounoff's Theorem—Fundamental Lemma—Proof of Laplace-Liapounoff's Theorem—Particular Cases—Estimation of Error Term—Hypothesis of Elementary Errors—Examples Concerning Dependent Variables—Problems for Solution.

CHAPTER XV

NORMAL DISTRIBUTION IN TWO DIMENSIONS. LIMIT THEOREM FOR SUMS OF INDEPENDENT VECTORS. ORIGIN OF NORMAL CORRELATION 308

Properties of Two-dimensional Normal Distribution—Limit Theorem for Sums of Independent Vectors—Application to Trials with Three or More Alternatives—Pearson's "χ^2-Test"—Hypothetical Explanation of Normal Correlation—Problems for Solution.

CHAPTER XVI

DISTRIBUTION OF CERTAIN FUNCTIONS OF NORMALLY DISTRIBUTED VARIABLES . 331

Distribution of Sums of Squares—Distribution of Sums of Squares of Deviations from the Mean—"Student's" Distribution—Distribution of the Correlation Coefficient of a Sample—Problems for Solution.

APPENDIX I

EULER'S SUMMATION FORMULA—STIRLING'S FORMULA—SOME DEFINITE INTEGRALS 347

APPENDIX II

METHOD OF MOMENTS AND ITS APPLICATIONS 356

Introductory Remarks—Continued Fractions Associated with Power Series—Continued Fractions Associated with $\int_a^b \frac{d\varphi(x)}{z-x}$. Properties of Denominators of Their Convergents—Equivalent Point Distributions—Tshebysheff's Inequalities—Application to Normal Distribution—Tshebysheff-Markoff's Fundamental Theorem—Application to Sums of Independent Variables.

APPENDIX III

ON A GAUSSIAN PROBLEM 396

Statement of the Problem—Analytical Expression for $P_n(x)$ and Its Derivative—More General Recurrence Relation—Main Inequalities—Solution of the Gaussian Problem.

TABLE OF PROBABILITY INTEGRAL 407

INDEX 409

INTRODUCTION TO MATHEMATICAL PROBABILITY

INTRODUCTION

Quanto enim minus rationis terminis comprehendi posse videbatur, quae fortuita sunt atque incerta, tanto admirabilior ars censebitur, cui ista quoque subjacent.—
CHR. HUYGENS,
De ratiociniis in ludo aleae.

1. It is always difficult to describe with adequate conciseness and clarity the object of any particular science; its methods, problems, and results are revealed only gradually. But if one must define the scope of the theory of probability the answer may be this: The theory of probability is a branch of applied mathematics dealing with the effects of *chance.* Here we encounter the word "chance," which is often used in everyday language but with rather indefinite meaning. To make clearer the idea conveyed by this word, we shall try first to clarify the opposite idea expressed in the word "necessity." Necessity may be logical or physical. The statement "The sum of the angles in a triangle is equal to two right angles" is a logical necessity, provided we assume the axioms of Euclidean geometry; for in denying the conclusion of the admitted premises, we violate the logical law of contradiction.

The following statements serve to illustrate the idea of physical necessity:

A piece of iron falls down if not supported.

Water boils if heated to a sufficiently high temperature.

A die thrown on a board never stands on its edge.

The logical structure of all these statements is the same: When certain conditions which may be termed "causes" are fulfilled, a definite effect occurs *of necessity.* But the nature of this kind of necessity is different from that of logical necessity. The latter, with our organization of mind, appears absolute, while physical necessity is only a result of extensive induction. We have never known an instance in which water, heated to a high temperature, did not boil; or a piece of iron did not fall down; or a die stood on its edge. For that reason we are led to believe that in the preceding examples (and in innumerable similar instances) the effect follows from its "cause" of necessity.

Instead of the term "physical necessity" we may introduce the abstract idea of "natural law." Thus, it is a "natural law" that the piece of iron left without support will fall down. Natural laws derived from extensive experiments or observations may be called "empirical laws" to distinguish them from theoretical laws. In all exact sciences which have reached a high degree of development, such as astronomy, physics, and chemistry, scientists endeavor to build up an abstract and simplified image of the infinitely complex physical world—an image which can be described in mathematical terms. With the help of hypotheses and some artificial concepts, it becomes possible to derive mathematically certain laws which, when applied to the world of reality, represent many natural phenomena with an amazing degree of accuracy. It is true that in the development of the sciences it sometimes becomes necessary to recast the previously accepted image of the physical world, but it is remarkable that the fundamental theoretical laws even then undergo but slight modification in substance or interpretation.

The chief endeavor of the exact sciences is the discovery of natural laws, and their formulation is of the greatest importance to the promotion of human knowledge in general and to the extension of our powers over natural phenomena.

Are the events caused by natural laws absolutely certain? No, but for all practical purposes they may be considered as certain. It is possible that one or another of the natural laws may fail, but such failure would constitute a real "miracle." However, granted that the possibility of miracles is consistent with the nature of scientific knowledge, actually this possibility may be disregarded.

2. If the preceding explanations throw a faint light upon the concept of necessity, it now remains to illuminate by comparison some characteristic features inherent in the concept of "chance." To say that chance is a denial of necessity is too vague a statement, but examples may help us to understand it better.

If a die is thrown upon a board we are *certain* that one of the six faces will turn up. But whether *a particular face* will show depends on what we call chance and cannot be predicted. Now, in the act of tossing a die there are some conditions known to us: first, that it is nearly cubic in shape; further, if it is a good die, its material is as nearly as possible homogeneous. Besides these known conditions, there are other factors influencing the motion of the die which are completely inaccessible to our knowledge. First among them are the initial position and the impulse imparted by the player's hand. These depend on an "act of will"—an agent which may act without any recognizable motivation—and therefore they are outside the domain of rational knowledge. Second, supposing the initial conditions known, the complexity of the resulting motion defies any possibility of foreseeing the final result.

Another example: If equal numbers of white and black balls, which do not differ in any respect except in color, are concealed in an urn, and we draw one of them blindly, it is certain that its color will be either white or black, but whether it will be black or white we cannot predict: that depends on chance. In this example we again have a set of known conditions: namely, that balls in equal numbers are white and black, and that they are not distinguishable except in color. But the final result depends on other conditions completely outside our knowledge. First, we know nothing about the respective positions of the white and black balls; second, the choice of one or the other depends on an act of will.

It is an observed fact that the numbers of marriages, divorces, births, deaths, suicides, etc., per 1,000 of population, in a country with nearly settled living conditions and during not too long a period of time, do not remain constant, but oscillate within comparatively narrow limits. For a given year it is impossible to predict what will be their numbers: that depends on chance. For, besides some known conditions, such as the level of prosperity, sanitation, and many other things, there are unnumbered factors completely outside our knowledge.

Many other examples of a similar kind can be cited to illustrate the notion of chance. They all possess a common logical structure which can be described as follows: an event A may materialize under certain known or "fixed" conditions, but not necessarily; for under the same fixed conditions other events $B, C, D, \ldots$ are also possible. The materialization of A depends also upon other factors completely outside our control and knowledge. Consequently, whether A will materialize or not under such circumstances cannot be foreseen; the materialization of A is due to chance, or, to express it concisely, A is a contingent event.

3. The idea of necessity is closely related to that of certainty. Thus it is "certain" that everybody will die in the due course of time. In the same way the idea of chance is related to that of *probability* or *likelihood*. In everyday language, the words "probability" and "probable" are used with different shades of meaning. By saying, "Probably it will rain tomorrow," we mean that there are more signs indicating rainy weather than fair for tomorrow. On the other hand, in the statement, "There is little probability in the story he told us," the word "probability" is used in the sense of credibility. But henceforth we shall use the word as equivalent to the *degree* of credence which we may place in the possibility that some contingent event may materialize. The "degree of credence" implies an almost instinctive desire to compare probabilities of different events or facts. That such comparison is possible one can gather from the following examples:

I live on the second floor and can reach the ground either by using the stairway or by jumping from the window. Either way I might be injured, though not necessarily. How do the probabilities of being injured compare in the two cases? Everyone, no doubt, will say that the probability of being injured by jumping from the window is "greater" than the probability of being injured while walking down the stairway. Such universal agreement might be due either to personal experience or merely to hearsay about similar experiences of other persons.

An urn contains an equal number of white and black balls that are similar in all respects except color. One ball is drawn. It may be either black or white. How do the probabilities of these two cases compare? One almost instinctively answers: "They are equal."

Now, if there are 10 white balls and 1 black ball in the urn, what about the probabilities of drawing a white or a black ball? Again one would say without hesitation that the probability of drawing a white ball is greater than that of drawing a black ball.

Thus, probability appears to be something which admits of comparisons in magnitude, but so far only in the same way as in the intensity of pain produced by piercing the skin with needles.

But it is a noteworthy observation that men instinctively try to characterize probabilities numerically in a naive and unscientific manner. We read regularly in the sporting sections of newspapers, predictions that in a coming race a certain horse has two chances against one to win over another horse, or that the chances of two football teams are as 10 to 7, etc. No doubt experts do know much about the respective horses and their riders, or the comparative strengths of two competing football teams, but their numerical estimates of chances have no other merit than to show the human tendency to assign numerical values to probabilities which most likely cannot be expressed in numbers.

It is possible that a man endowed with good common sense and ripe judgment can weigh all available evidence in order to compare the probabilities of the various possible outcomes and to direct his actions accordingly so as to secure profit for himself or for society. But precise conclusions can never be attained unless we find a satisfactory way to represent or to measure probabilities by numbers, at least in some cases.

4. As in other fields of knowledge, in attempting to measure probabilities by numbers, we encounter difficulties that cannot be avoided except by making certain ideal assumptions and agreements. In geometry (we speak of applied and not of abstract geometry), before explaining how lengths of rectilinear segments can be measured, we must first agree on criteria of equality of two segments. Similarly, in dealing with probability, the first step is to answer the question: When may two contingent events be considered as *equally probable* or, to use a more common expression, *equally likely?* From the statements of Jacob Bernoulli, one of the founders of the mathematical theory of probability, one can infer the following criterion of equal probability:

Two contingent events are considered as equally probable if, after taking into consideration all relevant evidence, one of them cannot be expected in preference to the other.

Certainly there is some obscurity in this criterion, but it is hardly possible to substitute any better one. To be perfectly honest, we must admit that there is an unavoidable obscurity in the principles of all the sciences in which mathematical analysis is applied to reality.

The application of Bernoulli's criterion to particular cases is beset with difficulties and requires good common sense and keen judgment. There is much truth in Laplace's statement: "La théorie des probabilités n'est au fond que le bon sens réduit au calcul."

To elucidate the nature of these difficulties, let us consider an urn filled with white and black balls, but in unknown proportion. The only evidence we have, namely, that there are both white and black balls in the urn, in this case appears insufficient for any conclusion about the respective probabilities of drawing a white or a black ball. We instinctively think of the numbers of the two kinds of balls, and, being in ignorance on this point, we are inclined to suspend judgment. But if we know that white and black balls are equal in number and distributed without any sort of regularity, this knowledge appears sufficient to assume the equality of the probabilities of drawing a white or a black ball. It is possible that, perhaps unconsciously, we are influenced by the commonly known fact that if we repeatedly draw a ball out of the urn many times, returning the ball each time before drawing again, the white and the black balls appear in nearly equal numbers.

If an urn contains a certain number of identical balls distinguished from one another by some characteristic signs, for example, by the numbers 1, 2, 3, . . . , the knowledge that the balls are identical and are distributed without regularity suffices in this case to cause us to conclude that the probabilities for drawing any of the balls should be considered as equal. Again, in so readily assuming this conclusion we may be influenced by the fact empirically observed (by ourselves or by others) that in a long series of drawings, with balls being restored to the urn after each withdrawal, the balls appear with nearly the same frequency.

An ordinary die is tossed. Should we consider the possible numbers of points 1, 2, 3, 4, 5, 6 as equally probable? To pronounce any judgment, we must know something about the die. If it is known that the die has a regular cubic shape and that its material is homogeneous, we readily agree on the equal probabilities of all the numbers of points 1, 2, 3, 4, 5, 6. And this a priori conclusion, based on Bernoulli's criterion, agrees with the observed fact that each number of points does appear nearly an equal number of times in a long series of throws, if the die is a good one. However, if we only know that the die has a regular shape, but not whether or not it is loaded, it is only sensible to suspend judgment.

These examples show that before trying to apply Bernoulli's criterion, we must have at our disposal some evidence the amount of which cannot be determined by any general rules. It may be also that the reason a priori must be supplemented by some empirical evidence. In some cases, lacking sufficient grounds to assert equal probabilities for two events, we may assume them as a hypothesis, to be kept until for some reason we are forced to abandon it.

5. Besides the ticklish question: When are we entitled to consider events as equally probable? there is another fundamental assumption required to make possible the measurement of probabilities by numbers.

Events $a_1, a_2, \ldots a_n$ form an exhaustive set of possibilities under certain fixed conditions S, if at least one of them must necessarily materialize. They are mutually exclusive if any two of them cannot materialize simultaneously. The fundamental assumption referred to consists in the possibility of subdividing results consistent with the conditions S into a number of exhaustive, mutually exclusive, and equally likely events, or cases (as they are commonly called):

$$a_1, a_2, \ldots a_n.$$

This being granted, the probability of any one of these cases is assumed to be $1/n$.

An event A may materialize in several mutually exclusive particular forms: $\alpha, \beta, \ldots \lambda$; that is, if A occurs, then one and only one of the events $\alpha, \beta, \ldots \lambda$ occurs also, and conversely the occurrence of one of these events necessitates the occurrence of A. Thus, if A consists in drawing an ace from a deck of cards, A may materialize in four mutually exclusive forms: as an ace of hearts, diamonds, clubs, or spades.

Let an event A be represented by its particular forms $a_1, a_2, \ldots a_m$, which together with other events $a_{m+1}, a_{m+2}, \ldots a_n$ constitute an exhaustive set of mutually exclusive and equally likely cases consistent with the conditions S. Events $a_1, a_2, \ldots a_m$ are called "cases favorable to A."

Definition of Mathematical Probability. *If, consistent with conditions S, there are n exhaustive, mutually exclusive, and equally likely cases, and m of them are favorable to an event A, then the mathematical probability of A is defined as the ratio m/n.*

In drawing a card from a full deck there are 52 and no more mutually exclusive and equally likely cases; 4 of them are favorable for drawing an ace; hence the probability of drawing an ace is $4/52 = 1/13$.

From an urn containing 10 white, 20 black, and 5 red balls, one ball is drawn. Here, distinguishing individual balls, we have 35 equally likely cases. Among them there are 10, 20, and 5 cases, favorable respectively to a white, a black, or a red ball. Hence the probabilities of drawing a white, a black, or a red ball are, respectively, $\frac{2}{7}$, $\frac{4}{7}$, and $\frac{1}{7}$.

In the first example, instead of 52 cases, we may consider only 13 cases according to the denominations of the cards. These cases being regarded as equally likely, there is only one of them favorable to an ace. The probability of drawing an ace is $\frac{1}{13}$. This observation makes it clear that the subdivision of all possible results into equally likely cases can be done in various ways. To avoid contradictory estimations of the same probability we must always observe the following rules:

Two events are equally likely if each of them can be represented by equal numbers of equally likely forms.

Two events are *not* equally likely if they are represented by unequal numbers of equally likely forms.

Thus, if two equally likely events are each represented by different numbers of their respective forms, then the latter cannot be considered as equally likely.

Each card is characterized by its denomination and the suit to which it belongs. Noting denominations, we distinguish 13 cases, but each of these is represented by 4 new cases according to the suit to which the card belongs. Altogether we have, then, 52 cases recognized as equally likely; hence, the above-mentioned 13 cases should be considered as equally likely.

In connection with the definition of mathematical probability, mention should be made of an important principle not always explicitly stated. If

$$a_1, a_2, \ldots a_m; \qquad b_1, b_2, \ldots b_p$$

are all mutually exclusive and equally likely cases consistent with certain conditions, and the indication of the occurrence of an event B makes cases $b_1, b_2, \ldots b_p$ impossible, cases $a_1, a_2, \ldots a_m$ still should be considered as equally likely. To illustrate this principle, consider an urn with six tickets bearing numbers 1, 2, . . . 6. Two tickets are drawn in succession. If nothing is known about the number of the first ticket, we still have six possibilities for the number of the second ticket, which we agree to consider as equally likely. But as soon as the number of the first ticket becomes known, then there are only five cases left concerning the number of the second ticket. According to the above principle we must consider these five cases as equally likely.

6. Probability as defined above is represented by a number contained between 0 and 1. In the extreme case in which the probability is 0, it indicates the impossibility of an event. On the contrary, in the other extreme case in which the probability is 1, the event is certain. When the probability is expressed by a number very near to 1, it means that the overwhelming majority of cases are favorable to the event. On the contrary, a probability near to 0 shows that the proportion of favorable cases is small.

From our experience we know that events with a small probability seldom happen. For instance, if the probability of an event is 1/1,000,000, the situation may be likened to the drawing of a white ball from an urn containing 999,999 black balls and a single white one. This white ball is practically lost among the majority of black balls, and for all practical purposes we may consider its extraction impossible. Similarly, the probability 999,999/1,000,000 may be considered, from a practical standpoint, as an indication of certainty. What limit for smallness of probability is to be set as an indication of practical impossibility? Evidently there is no general answer to this question. Everything depends on the risk we can face if, contrary to expectation, an event with a small probability should occur. Hence, the main problem of the theory of probability consists in finding cases in which the probability is very small or very near to 1. Instead of saying, "The probability is very near to 1," we shall say, "great probability," although, of course, the probability can never exceed 1.

7. The definition of mathematical probability in Sec. 5 is essentially the classical definition proposed by Jacob Bernoulli and adopted by Laplace and almost all the important contributors to the theory of probability. But, since the middle of the nineteenth century (Cournot, John Stuart Mill, Venn), and especially in our days, the classical definition has been severely criticized. Several attempts have been made to rear up the edifice of the mathematical theory of probability on quite a different definition of mathematical probability. It does not enter into our plan to criticize these new definitions, but, in the opinion of the author, many of them are self-contradictory. Modern attempts to build up the theory of probability as an axiomatic science may be interesting in themselves as mental exercises; but from the standpoint of applications the purely axiomatic science of probability would have no more value than, for example, would the axiomatic theory of elasticity.

The most serious objection to the classical definition is that it can be used only in very simple and comparatively unimportant cases like games of chance. This objection, stressed by von Mises, is in reality not a new one. It is one of the objections Leibnitz made against Jacob Bernoulli's views concerning the possibility of applications of the theory of probability to various important fields of human endeavor and not merely to games of chance.

It is certainly true that the classical definition cannot be directly applied in many important cases. But is it the fault of the definition or is it rather due to our ignorance of the innermost mechanisms which, apart from chance, contribute to the materialization or nonmaterialization of contingent events? It seems that this is what Jacob Bernoulli meant in his reply to Leibnitz:

Objiciunt primo, aliam esse rationem calculorum, aliam morborum aut mutationum aeris; illorum numerum determinatum esse, horum indeterminatum et vagum. Ad quod respondeo, utrumque respectu cognitionis nostrae aequi poni incertum et indeterminatum; sed quicquam in se et sua natura tale esse, non magis a nobis posse concipi, quam concipi potest, idem simul ab Auctore naturae creatum esse et non creatum: quaecumque enim Deus fecit, eo ipso dum fecit, etiam determinavit.[1]

8. A brilliant example of how the profound study of a subject finally makes it possible to apply the classical definition of mathematical probability is afforded in the fundamental laws of genetics (a science of comparatively recent origin, whose importance no one can deny), discovered by the Augustinian monk, Gregor Mendel (1822–1884). During eight years Mendel[2] conducted experimental work in crossing different varieties of the common pea plant with the purpose of investigating how pairs of contrasting characters were inherited. For the pea plant there are several pairs of such contrasting characters: round or wrinkled seeds, tallness or dwarfness, yellow or green pod color, etc. Let us concentrate our attention on a definite pair of contrasting characters, yellow or green pod color. Peas with green pod color always breed true. Also some peas with yellow color always breed true, while still others produce both varieties. True breeding pea plants constitute two pure races: A with yellow pod color and B with green pod color, while plants with yellow pods not breeding true constitute a hybrid race, C. Crossing plants of the race A with those of the race B and planting the seeds, Mendel obtained a first generation F_1 of hybrids. Letting plants of the first generation self-fertilize and again planting their seeds to produce the second generation F_2, Mendel found that in this generation there were 428 yellow pod plants and 152 green pod plants in the ratio 2.82:1. In regard to other contrasting characters the ratio of approximately 3:1 was observed in all cases. Later experimental work only confirmed Mendel's results. Thus, combined experiments of Correns, Tschermak, and others gave among 195,477 individuals of F_2, 146,802 yellow pod plants and 48,675 green pod plants, in the ratio 3.016:1.

[1] To understand the beginning of this statement see the translation from "Ars conjectandi" in Chap. VI, p. 105.

[2] Mendel's results were published in 1865, but passed completely unnoticed until in about 1900 the same facts were rediscovered by DeVries, Correns, and Tschermak. Modern genetics dates from about this time.

Mendel not only discovered such remarkable regularities, but also suggested a rational explanation of the observed ratio 3:1, which with some modifications is accepted even today. Bodies of plants and animals are built up of enormous numbers of cells, among which the reproductive cells, or gametes, differ from the remaining "somatic" cells in some important qualities. Cells are not homogeneous, but possess a definite structure. In somatic cells there are found bodies, called chromosomes, whose number is even and the same for the same species. Exactly half of this number of chromosomes is found in reproductive cells. Chromosomes are supposed to be seats of hypothetical "genes," which are considered as bearers of various heritable characters. A chromosome of one pure race A bearing a character a differs from the homologous chromosome of another pure race B bearing a contrasting character b in that they contain genes of different kinds. Since characters a and b are borne by definite chromosomes, the situation in regard to the two characters a and b is exactly the same as if gametes of both races contained just one chromosome. Let us represent them symbolically by $\odot$ and $\otimes$. In the act of fertilization a pair of paternal and maternal gametes conjugate and form a zygote, which by division and growth produces all cells of the filial generation. Certain of these cells become the germ cells and are set apart for the formation, by a complicated process, of gametes, one half of which contain the chromosome of the paternal type and the other half that of the maternal type.

According to this theory, in crossing two individuals belonging to races A and B, zygotes of the first generation F_1 will be of the type $\odot$—$\otimes$, and will produce gametes, in equal numbers, of the types $\odot$, $\otimes$. Now if two individuals of F_1 (hybrids) are crossed (or one individual self-fertilized as in the cases of some plants), one paternal gamete conjugates with one maternal, and for the resulting zygote there are four possibilities:

$$\odot\text{—}\odot \qquad \odot\text{—}\otimes \qquad \otimes\text{—}\odot \qquad \otimes\text{—}\otimes$$

These possibilities may be considered as equally probable, whence the probabilities for an individual of the generation F_2 to belong respectively to the races A, B, C are $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{2}$. Similarly, one easily finds that in crossing an individual of the race A with one of the hybrid race C, the probabilities of the offspring belonging to A or C are both equal to $\frac{1}{2}$.

It is easy now to offer a rational explanation of the Mendelian ratio 3:1. In the case of pea plants, individuals of the race A and hybrids are not distinguishable in regard to the color of their pods. Hence the probability of the offspring of a hybrid plant having yellow pods is $\frac{3}{4}$, while for the offspring to have green pods the probability is $\frac{1}{4}$. When the generation F_2 consists of a great many individuals, the theory of probability shows that the ratio of the number of yellow pod plants to the number of green pod plants is not likely to differ much from the ratio 3:1. In crossing plants of the race A with hybrids, the offspring, if numerous, will contain plants of race A or C, respectively, in a proportion which is not likely to differ much from 1:1. And this conclusion was experimentally verified by Mendel himself.

9. If in the case of the Mendelian laws the profound study of the mechanism of heredity together with hypothetical assumptions of the kind used in physics, chemistry, etc., paved the way for a rational explanation of observed phenomena on the basis of the theory of probability, in many other important instances we are still unable to reach the same degree of scientific understanding. Stability of statistical ratios observed in many cases suggests the idea that they should be explained on the basis of probability. For instance, it has been observed that the ratio of human male and female births is nearly 51:50 for large samples, and this is largely independent of climatic conditions, racial differences, living conditions in different countries, etc. Although the factors determining sex are known, yet some complications not sufficiently cleared up prevent estimation of probabilities of male and female births.

In all instances of the pronounced stability of statistical ratios we may believe that some day a way will be found to estimate probabilities in such cases. Therefore many applications of the theory of probability to important problems of other sciences are based on belief in the existence of the probabilities with which we are concerned. In other cases in which the theory of probability is used, we may have grave doubts as to whether this science is applied legitimately. The fact that many applications of probability are based on belief or faith should not discourage us; for it is better to do something, though it may be not quite reliable, than nothing. Only we must not be overconfident about the conclusions reached under such circumstances.

After all, is not faith at the bottom of all scientific knowledge? Physicists speak of electrons, which never have been seen and are known only through their visible manifestations. Electrons are postulated just to coordinate into a coherent whole a large variety of observed phenomena. Is not this faith? It must be, for according to Paul (Hebrews, 11:1), "Faith is the substance of things hoped for, the evidence of things not seen."

10. In concluding this introduction it remains to give a short account of the history of the theory of probability. Although ancient philosophers discussed at length the necessity and contingency of things, it seems that mathematical treatment of probability was not known to the ancients. Apart from casual remarks of Galileo concerning the correct evaluation of chances in a game of dice, we find the true origin of the science of probability in the correspondence between two great men of the seventeenth century, Pascal (1623–1662) and Fermat (1601–1665). A French nobleman, Chevalier de Méré, a man of ability and great experience in gambling, asked Pascal to explain some seeming contradictions between his theoretical reasoning and the observations gathered from gambling. Pascal solved this difficulty and attacked another problem proposed to him by de Méré. On hearing from Pascal about these problems, Fermat became interested in them, and in their private correspondence these two great men laid the first foundations of the science of probability. Bertrand's statement, "Les grands noms de Pascal et de Fermat décorent le berceau de cette science" cannot be disputed.

Huygens (1629–1695), a great Dutch scientist, became acquainted with the contents of this correspondence and, spurred on by the new ideas, published in 1654 a first book on probability, "De ratiociniis in ludo aleae," in which many interesting and rather difficult problems on probabilities in games of chance were solved. To him we owe the concept of "mathematical expectation" so important in the modern theory of probability.

Jacob Bernoulli (1654–1705) meditated on the subject of probability for about twenty years and prepared his great book, "Ars conjectandi," which, however, was not published until eight years after his death in 1713, by his nephew, Nicholas Bernoulli. Bernoulli envisaged the subject from the most general point of view, and clearly foresaw a whole field of applications of the theory of probability outside of the narrow circle of problems relating to games of chance. To him is due the discovery of one of the most important theorems known as "Bernoulli's theorem."

The next great successor to Bernoulli is Abraham de Moivre (1667-1754), whose most important work on probability, "The Doctrine of Chances," was first published in 1718 and twice reprinted in 1738 and in 1756. De Moivre does not contribute much to the principles, but this work is justly renowned for new and powerful methods for the solution of more difficult problems. Many important results, ordinarily attributed to Laplace and Poisson, can be found in de Moivre's book.

Laplace (1749–1827), whose contributions to celestial mechanics assured him everlasting fame in the history of astronomy, was very much interested in the theory of probability from the very beginning of his scientific career. After writing several important memoirs on the subject, he finally published, in 1812, his great work "Théorie analytique des probabilités," accompanied by a no less known popular exposition, "Essai philosophique sur les probabilités," destined for the general educated public. Laplace's work, on account of the multitude of new ideas, new analytic methods, and new results, in all fairness should be regarded as one of the most outstanding contributions to mathematical literature. It exercised a great influence on later writers on probability in Europe, whose work chiefly consisted in elucidation and development of topics contained in Laplace's book.

Thus in European countries further development of the theory of probability was somewhat retarded. But the subject took on important developments in the works of Russian mathematicians: Tshebysheff (1821–1894) and his former students, A. Markoff (1856–1922) and A. Liapounoff (1858–1918). Castelnuovo in his fine book "Calcolo delle probabilità" rightly regards the contributions to the theory of probability due to Russian mathematicians as the most important since the time of Laplace.

At the present time interest in the theory of probability is revived everywhere, but again the most outstanding recent contributions have been made in Russia, chiefly by three prominent mathematicians: S. Bernstein, A. Khintchine, and A. Kolmogoroff.

In closing this introduction it seems proper to quote the closing words of the "Essai philosophique sur les probabilités":

> On voit par cet Essai, que la théorie des probabilités n' est au fond, que le bon sens réduit au calcul: elle fait apprécier avec exactitude, ce que les ésprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte. Elle ne laisse rien d'arbitraire dans le choix des opinions et des partis à prendre, toutes les fois que l'on peut, à son moyen, déterminer le choix le plus avantageux. Par la, elle devient le supplément le plus heureux, à l'ignorance et à la faiblesse de l'ésprit humain. Si l'on considère les méthodes analytiques auxquelles cette théorie a donné naissance, la vérité des principes qui lui servent de base, la logique fine et délicate qu' exige leur emploi dans la solution des problèmes, les établissements d'utilité publique qui s'appuient sur elle, et l'extension qu'elle a reçue et qu'elle peut reçevoir encore, par son application aux questions les plus importantes de la Philosophie naturelle et des sciences morales; si l'on observe ensuite, que dans les choses mêmes qui ne peuvent être soumise au calcul, elle donne les aperçus les plus sûrs qui puissent nous guider dans nos jugements, et qu'elle apprend à se garantir des illusions qui souvent nous égarent; on verra qu'il n'est point de science plus digne de nos méditations.

References

JACOB BERNOULLI: "Ars conjectandi," 1713.
LAPLACE: "Essai philosophique sur les probabilités," Oeuvres VII, pp. v–cliii.
BERTRAND: "Calcul des probabilités," pp. vi–l, Paris, 1889.
POINCARÉ: "Calcul des probabilités," 2d ed., pp. 1–23, Paris, 1912.
KRIES: "Die Principien der Wahrscheinlichkeitsrechnung. Eine logische Untersuchung," Freiburg, 1896.
J. M. KEYNES: "A Treatise on Probability," London, 1921.
VON MISES: "Wahrscheinlichkeitsrechnung," Leipzig, 1931.
A. KOLMOGOROFF: "Grundbegriffe der Wahrscheinlichkeitsrechnung," Berlin, 1933.

CHAPTER I

COMPUTATION OF PROBABILITIES BY DIRECT ENUMERATION OF CASES

1. The probability of an event can be found by direct application of the definition when it is possible to make a complete enumeration of all equally likely cases, as well as of those favorable to that event. Here we shall consider a few problems, beginning with the simplest, to illustrate this direct method of evaluating probabilities.

Problem 1. Two dice are thrown. What is the probability of obtaining a total of 7 or 8 points?

Solution. Suppose we distinguish the dice by the numbers 1 and 2. There are 6 possible cases as to the number of points on the first die; and each of these cases can be accompanied by any of the 6 possible numbers of points on the second die. Hence, we can distinguish altogether $6 \times 6 = 36$ different cases. Provided the dice are ideally regular in shape and perfectly homogeneous, we have good reason to consider these 36 cases as equally likely, and we shall so consider them.

Next, let us find out how many cases are favorable to the total of 7 points. This may happen only in the following ways:

First Die	Second Die
1	6
2	5
3	4
4	3
5	2
6	1

Likewise, for 8 points:

First Die	Second Die
2	6
3	5
4	4
5	3
6	2

That is, out of the total number of 36 cases there are 6 cases favorable to 7 points and 5 cases favorable to 8 points; hence, the probability of obtaining 7 points is $6/36$ and the probability of obtaining 8 points is $5/36$.

2. Problem 2. A coin is tossed three times in succession. What is the probability of obtaining 2 heads? What is the probability of obtaining tails at least once?

Solution. In the first throw there are two possible cases: heads or tails. And if the coin is unbiased (which we assume is true) these two cases may be considered as equally likely. In two throws there are $2 \times 2 = 4$ cases; namely, both of the two possible cases in the first toss can combine with both of the possible cases in the second. Similarly, in three throws the number of cases will be $2 \times 2 \times 2 = 8$. To find the number of cases favorable to obtaining 2 heads, we must consider that this can happen only in three ways:

Heads	Heads	Tails
Heads	Tails	Heads
Tails	Heads	Heads

The number of favorable cases being 3, the probability of obtaining two heads is $3/8$.

To answer the second part of the question, we observe that there is only one case when tails does not turn up. Therefore, the number of cases favorable to obtaining tails at least once is $8 - 1 = 7$, so that the required probability is $7/8$.

3. Problem 3. Two cards are drawn from a deck of well-shuffled cards. What is the probability that both the extracted cards are aces?

Solution. Since there are 52 cards in the deck, there are 52 ways of extracting the first card. After the first card has been withdrawn, the second extracted card may be one of the remaining 51 cards. Therefore, the total number of ways to draw two cards is 52×51. All these cases may be considered as equally likely.

To find the number of cases favorable to drawing aces, we observe that there are 4 aces; therefore, there are 4 ways to get the first ace. After it has been extracted, there are 3 ways to get a second ace. Hence, the total number of ways to draw 2 aces, is 4×3, and the required probability is:

$$\frac{4 \times 3}{52 \times 51} = \frac{1}{13 \times 17} = \frac{1}{221}.$$

Problem 4. Two cards are drawn from a full pack, the first card being returned to the pack before the second is taken. What is the probability that both the extracted cards belong to a specified suit?

Solution. There are 52 ways of getting the first card. For the second drawing, there are also 52 ways, because by returning the first extracted card to the pack, the original number was restored. Under such circumstances, the total number of ways to extract two cards is 52×52. Now, because there are 13 cards in a suit, the number of cases favorable to obtaining two cards of a specified suit is 13×13.

Therefore, the required probability is given by:

$$\frac{13 \times 13}{52 \times 52} = \frac{1 \times 1}{4 \times 4} = \frac{1}{16}.$$

4. Problem 5. An urn contains 3 white and 5 black balls. One ball is drawn. What is the probability that it is black?

Solution. The total number of balls is 8. To distinguish them, we may imagine that they are numbered. As to the number on the ball drawn, there are 8 possible cases that may reasonably be considered as equally likely. Obviously, there are 5 cases favorable to the black color of the ball drawn. Therefore, the required probability is $5/8$.

By a slight modification of the last problem, we come to the following interesting situation:

Problem 6. The contents of the urn are the same as in the foregoing problem. But this time we suppose that one ball is drawn, and, *its color unnoted,* laid aside. Then another ball is drawn, and we are required to find the probability that it is black or white.

Solution. Suppose again that the balls are numbered, so that the white balls bear numbers 1, 2, and 3; and the black balls bear numbers 4, 5, 6, 7, 8. Obviously, there are 8 ways to get the first ball, and whatever it is, there remain only 7 ways to get the second ball. The total number of equally likely cases is $8 \times 7 = 56$.

It is a little more difficult to find the number of cases favorable to extracting a white or black ball in the second drawing. Suppose we are interested in the white color of the second ball. If the first ball drawn is a white one, it may bear one of the numbers 1 to 3. Whatever this number is, the second ball, if it is white, can bear only the two remaining numbers. Therefore, under the assumption that the first ball is a white one, the number of favorable cases is $3 \times 2 = 6$. Again, supposing that the first ball drawn is black, we have 5 possibilities as to its number, and, corresponding to any one of these possibilities, there are 3 possibilities as to the number of the white ball to be taken in the second drawing, so that the number of favorable cases now is $5 \times 3 = 15$. The number of all favorable cases is $6 + 15 = 21$. The required probability for the white ball is $21/56 = 3/8$. In the same way, we should find that the probability for the black ball is $5/8$. It is remarkable that these two probabilities are the same as if only a single ball had been drawn.

The situation is quite different if we know the color of the first ball. Suppose, for instance, that it is white. The total number of equally likely cases will then be $3 \times 7 = 21$; and the number of cases favorable to getting another white ball is $3 \times 2 = 6$, so that the probability in this case is $2/7$.

This last example shows clearly how much probability depends upon a given or known set of conditions.

5. Problem 7. Three boxes, identical in appearance, each have two drawers. The first box contains a gold coin in each drawer; the second contains a silver coin in each drawer; but the third contains a gold coin in one drawer and a silver coin in the other. (*a*) A box is chosen at random. What is the probability that it contains coins of different metals? (*b*) A box is chosen, one of its drawers opened, and a gold coin found. What is the probability that the other drawer contains a silver coin?

Solution. (*a*) Since nothing outwardly distinguishes one box from the other, we may recognize three equally likely cases, and among them is only one case of a box with coins of different metals. Therefore, we estimate the required probability as $\frac{1}{3}$.

(*b*) As to the second question, one is tempted to reason as follows: The fact that a gold coin was found in one drawer leaves only two possibilities as to the content of the other drawer; namely, that the coin in it is either gold or silver. Hence, the probability of a silver coin in the second drawer seems to be $\frac{1}{2}$. But this reasoning is fallacious. It is true that, when the gold coin is found in one drawer, there are only two possibilities left as to the content of the other drawer; but these possibilities *cannot be considered as equally likely.* To see this point clearly, let us distinguish the drawers of the first box by the numbers 1 and 2; those of the second box by the numbers 3 and 4; finally, in the third box, 5 will distinguish the drawer containing the silver coin, while 6 will represent the drawer with the gold coin.

Instead of three equally likely cases:

box 1, box 2, box 3

we now have six cases:

drawers 1, 2; drawers 3, 4; drawers 5, 6,

which, with reference to the fundamental assumptions, must be considered as equally likely. If nothing were known about the contents of the drawer which has been opened, the number of this drawer might be either 1, 2, 3, 4, 5, or 6. But as soon as the gold coin is discovered in it, cases 3, 4, and 5 become impossible, and there remain *three* equally likely assumptions as to the number of the opened drawer: it may be either 1 or 2 or 6. That leaves three cases, and in only one of them, namely, in case 6, will the other drawer contain a silver coin. Thus the answer to the second question (*b*) is $\frac{1}{3}$.

6. In the preceding problems the enumeration of cases did not present any difficulty. We are now going to discuss a few problems in which this enumeration is not so obvious but can be greatly simplified by the use of well-known formulas for the number of permutations, arrangements, and combinations.

Let m distinct objects be represented by the letters $a, b, c, \dots l$. Using all these objects, we can place them in different orders and form "permutations." For instance, if there are only three letters, a, b, and c, all the possible permutations are: *abc*, *acb*, *bac*, *bca*, *cab*, *cba*,—6 different permutations out of 3 letters. In general, the number of permutations P_m of m objects is expressed by

$$P_m = 1 \cdot 2 \cdot 3 \cdots m = m!$$

If n objects are taken out of the total number of m objects to form groups, attention being paid to the order of objects in each group, then these groups are called "arrangements." For instance, by taking two letters out of the four letters a, b, c, d, we can form the following 12 arrangements:

ab	*ba*	*ca*	*da*
ac	*bc*	*cb*	*db*
ad	*bd*	*cd*	*dc*

Denoting by the symbol A_m^n the number of arrangements of m objects taken n at a time, the following formula holds:

$$A_m^n = m(m-1)(m-2) \cdots (m-n+1).$$

Again, if we form groups of n objects taken out of the total number of m objects, this time paying no attention to the order of objects in the group, we form "combinations." For instance, following are the different combinations out of 5 objects taken 3 at a time:

abc	*abd*	*abe*	*acd*	*ace*
ade	*bcd*	*bce*	*bde*	*cde*

In general, the number of combinations out of m objects taken n at a time, which is usually denoted by the symbol C_m^n, is given by

$$C_m^n = \frac{m(m-1)(m-2)\cdots(m-n+1)}{1\cdot 2\cdot 3\cdots n}.$$

It is useful to recall that the same expression may also be exhibited as follows:

$$C_m^n = \frac{m!}{n!(m-n)!},$$

whence, by substituting $m - n$ instead of n, the useful formula

$$C_m^n = C_m^{m-n}$$

can be derived.

7. After these preliminary remarks, we can turn to the problems in which the foregoing formulas will often be used.

Problem 8. An urn contains a white balls and b black balls. If $\alpha + \beta$ balls are drawn from this urn, find the probability that among them there will be exactly α white and β black balls.

Solution. If we do not distinguish the order in which the balls come out of the urn, the total number of ways to get $\alpha + \beta$ balls out of the total number $a + b$ balls is obviously expressed by $C_{a+b}^{\alpha+\beta}$ and this is the number of all possible and equally likely cases in this problem. The number of ways to draw α white balls out of the total number a of white balls in the urn is C_a^α; and similarly C_b^β represents the number of ways of drawing β black balls out of the total number b of black balls. Now every group of α white balls combines with every possible group of β black balls to form the total of α white balls and β black balls, so that the number of ways to form all the groups containing α white balls and β black balls is $C_a^\alpha \cdot C_b^\beta$. This is also the number of favorable cases; hence, the required probability is

$$p = \frac{C_a^\alpha \cdot C_b^\beta}{C_{a+b}^{\alpha+\beta}},$$

or, in a more explicit form,

$$(1) \quad p = \frac{1 \cdot 2 \cdots (\alpha + \beta)}{1 \cdot 2 \cdots \alpha \cdot 1 \cdot 2 \cdots \beta} \cdot \frac{a(a-1) \cdots (a - \alpha + 1) \cdot b(b-1) \cdots (b - \beta + 1)}{(a+b)(a+b-1) \cdots (a + b - \alpha - \beta + 1)}.$$

Problem 9. An urn contains n tickets bearing numbers from 1 to n, and m tickets are drawn at a time. What is the probability that i of the tickets removed have numbers previously specified?

Solution. This problem does not essentially differ from the preceding one. In fact, i tickets with preassigned numbers can be likened to i white balls, while the remaining tickets correspond to the black balls. The required probability, therefore, can be obtained from the expression (1) by taking $a = i$, $b = n - i$, $\alpha = i$, $\beta = m - i$ and, all simplifications performed, will be given by

$$(2) \qquad p = \frac{m(m-1) \cdots (m - i + 1)}{n(n-1) \cdots (n - i + 1)}.$$

The conditions of this problem were realized in the French lottery, which was operated by the French royal government for a long time but discontinued soon after the Revolution of 1789. Similar lotteries continued to exist in other European countries throughout the nineteenth century. In the French lottery, tickets bearing numbers from 1 to 90 were sold to the people, and at regular intervals drawings for winning numbers were held in different French cities. At each drawing, 5 numbers were drawn. If a holder of tickets won on a single number, he received 15 times its cost to him. If he won on two, three, four, or five tickets, he could claim respectively 270, 5,500, 75,000, and, finally, 1,000,000 times their cost to him.

The numerical values of the probabilities corresponding to these different cases are worked out as follows: we must take $n = 90$, $m = 5$, and $i = 1, 2, 3, 4$, or 5 in the expression (2). The results are

Single ticket $\qquad \dfrac{5}{90} = \dfrac{1}{18}.$

Two tickets $\qquad \dfrac{5 \cdot 4}{90 \cdot 89} = \dfrac{2}{801}.$

Three tickets $\qquad \dfrac{5 \cdot 4 \cdot 3}{90 \cdot 89 \cdot 88} = \dfrac{1}{11748}.$

Four tickets $\qquad \dfrac{5 \cdot 4 \cdot 3 \cdot 2}{90 \cdot 89 \cdot 88 \cdot 87} = \dfrac{1}{511038}.$

Five tickets $\qquad \dfrac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86} = \dfrac{1}{43949268}.$

8. Problem 10. From an urn containing a white balls and b black ones, a certain number of balls, k, is drawn, and they are laid aside, their color unnoted. Then one more ball is drawn; and it is required to find the probability that it is a white or a black ball.

Solution. Suppose the k balls removed at first and the last ball drawn are laid on $k + 1$ different places, so that the last ball occupies the position at the extreme right. The number of ways to form groups of $k + 1$ balls out of the total number of $a + b$ balls, attention being paid to the order, is

$$(a + b)(a + b - 1) \cdots (a + b - k).$$

Such is the total number of cases in this problem, and they may all be considered as equally likely. To find the number of cases favorable to a white ball, we observe that the last place should be occupied by one of the a white balls. Whatever this white ball is, the preceding k balls form one of the possible arrangements out of $a + b - 1$ remaining balls taken k at a time. Hence, it is obvious that the number of cases favorable to a white ball is

$$a(a + b - 1) \cdots (a + b - k),$$

and therefore the required probability is given by

$$\frac{a}{a + b}$$

for a white ball. In a similar way we find the probability $b/(a+b)$ of drawing a black ball. These results show that the probability of getting white or black balls in this problem is the same as if no balls at all were removed at first. Here we have proof that the peculiar circumstances observed in Prob. 6 are general.

9. Problem 11. Two dice are thrown n times in succession. What is the probability of obtaining double six at least once?

Solution. As there are 36 cases in every throw and each case of the first throw can combine with each case of the second throw, and so on, the total number of cases in n throws will be 36^n. Instead of trying to find the number of favorable cases directly, it is easier to find the number of unfavorable cases; that is, the number of cases in which double-sixes would be excluded. In one throw there are 35 such cases, and in n throws there will be 35^n. Now, excluding these cases, we obtain $36^n - 35^n$ favorable cases; hence, the required probability is

$$p = 1 - \left(\tfrac{35}{36}\right)^n.$$

If one die were thrown n times in succession, the probability to obtain 6 points at least once would be

$$p = 1 - \left(\tfrac{5}{6}\right)^n.$$

Now, suppose we want to find the number of throws sufficient to assure a probability $> \frac{1}{2}$ of obtaining double six at least once. To this end we must solve the inequality

$$\left(\tfrac{35}{36}\right)^n < \tfrac{1}{2}$$

for n; whence we find

$$n > \frac{\log 2}{\log 36 - \log 35} = 24.6 \cdots.$$

It means that in 25 throws there is more likelihood to obtain double six at least once than not to obtain it at all. On the other hand, in 24 throws, we have less chance to succeed than to fail.

Now, if we dealt with a single die, we should find that in 4 throws there are more chances to obtain 6 points at least once than there are chances to fail.

This problem is interesting in a historical respect, for it was the first problem on probability solved by Pascal, who, together with his great contemporary Fermat, had laid the first foundations of the theory of probability. This problem was suggested to Pascal by a certain French nobleman, Chevalier de Méré, a man of great experience in gambling. He had observed the advantage of betting for double six in 25 throws and for one six (with a single die) in 4 throws. He found it difficult to understand because, he said, there were 36 cases for two dice and 6 cases for one die in each throw, and yet it is not true that 25:4 = 36:6. Of course, there is no reason for such an arbitrary conclusion, and the correct solution as given by Pascal not only removed any apparent paradoxes in this case, but it led to the same number, 25, observed by gamblers in their daily experience.

10. Problem 12. A certain number n of identical balls is distributed among N compartments. What is the probability that a certain specified compartment will contain h balls?

Solution. To find the number of all possible cases in this problem, suppose that we distinguish the balls by numbering them from 1 to n. The ball with the number 1 may fall into any of the N compartments, which gives N cases. The ball with the number 2 may also fall into any one of the N compartments; so that the number of cases for 2 balls will be $N \cdot N = N^2$. Likewise, for 3 balls the number of cases will be

$$N^2 \cdot N = N^3,$$

and for any number n of balls the number of cases will be N^n. To find the number of favorable cases, first suppose that a group of h specified balls falls into a designated compartment. The remaining $n - h$ balls may be distributed in any way among $N - 1$ remaining compartments. But the number of ways to distribute $n - h$ balls among $N - 1$ compartments is $(N - 1)^{n-h}$ and this becomes the number of all favorable cases in which a specified group of h balls occupies the designated compartment. Now, it is possible to form C_n^h such groups; therefore, the total number of favorable cases is given by

$$C_n^h \cdot (N - 1)^{n-h},$$

and the required probability will be

$$p_h = \frac{C_n^h \cdot (N - 1)^{n-h}}{N^n}.$$

In case n, N and h are large numbers, the direct application of this formula becomes difficult, and it is advisable to seek an approximate expression for p_h. To this end we write the preceding expression thus:

$$p_h = \frac{\left(\frac{n}{N}\right)^h}{1 \cdot 2 \cdot 3 \cdots h}\left(1 - \frac{1}{N}\right)^{n-h} P,$$

where

$$P = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{h-1}{n}\right)$$

Now, supposing $1 \leqq k \leqq h - 1$, we have

$$(a) \qquad \left(1 - \frac{k}{n}\right)\left(1 - \frac{h-k}{n}\right) = 1 - \frac{h}{n} + \frac{k(h-k)}{n^2} > 1 - \frac{h}{n}.$$

On the other hand,

$$k(h-k) \leqq \frac{h^2}{4},$$

and so

$$(b) \qquad \left(1 - \frac{k}{n}\right)\left(1 - \frac{h-k}{n}\right) \leqq \left(1 - \frac{h}{2n}\right)^2.$$

The inequalities (*a*) and (*b*) give simple lower and upper limits for P. For we can write P^2 thus:

$$P^2 = \prod_{k=1}^{h-1}\left(1 - \frac{k}{n}\right)\left(1 - \frac{h-k}{n}\right)$$

and then apply (*a*) or (*b*), which leads to these inequalities

$$P < \left(1 - \frac{h}{2n}\right)^{h-1}, \qquad P > \left(1 - \frac{h}{n}\right)^{\frac{h-1}{2}}.$$

Correspondingly, we have

$$p_h < \frac{\left(\frac{n}{N}\right)^h}{1 \cdot 2 \cdot 3 \cdots h}\left(1 - \frac{1}{N}\right)^{n-h}\left(1 - \frac{h}{2n}\right)^{h-1}$$

$$p_h > \frac{\left(\frac{n}{N}\right)^h}{1 \cdot 2 \cdot 3 \cdots h}\left(1 - \frac{1}{N}\right)^{n-h}\left(1 - \frac{h}{n}\right)^{\frac{h-1}{2}}.$$

Problem 13. What is the probability of obtaining a given sum s of points with n dice?

Solution. The number of all cases for n dice is evidently 6^n. The number of favorable cases is the same as the total number of solutions of the equation

$$(1) \qquad \alpha_1 + \alpha_2 + \cdots + \alpha_n = s$$

where $\alpha_1, \alpha_2, \cdots \alpha_n$ are integers from 1 to 6. This number can be determined by means of the following device: Multiplying the polynomial

$$(2) \qquad x + x^2 + x^3 + x^4 + x^5 + x^6$$

by itself, the product will consist of terms

$$x^{\alpha_1 + \alpha_2}$$

where α_1 and α_2 independently assume all integral values from 1 to 6. Collecting terms with the same exponent s, the coefficient of x^s will give the number of solutions of the equation

$$\alpha_1 + \alpha_2 = s,$$

α_1, α_2 being subject to the above mentioned limitations.

Similarly, multiplying the same polynomial (2) three times in itself and collecting terms with the same exponent s, the coefficient of x^s will give the number of solutions of equation (1) for $n = 3$. In general, the number of solutions of equation (1) for any n is the coefficient of x^s in the expanded polynomial

$$(x + x^2 + x^3 + x^4 + x^5 + x^6)^n.$$

Now we have identically

$$x + x^2 + x^3 + x^4 + x^5 + x^6 = \frac{x(1 - x^6)}{1 - x},$$

and by the binomial theorem

$$x^n(1 - x^6)^n = \sum_{l=0}^{n} (-1)^l C_n^l x^{n+6l}$$

$$(1 - x)^{-n} = \sum_{k=0}^{\infty} C_{n+k-1}^{n-1} x^k.$$

Multiplying these series we find the following expression as the coefficient of x^s:

$$\sum_{l=0}^{\frac{s-n}{6}} (-1)^l C_n^l C_{s-6l-1}^{n-1}$$

where summation extends over integers not exceeding $\frac{s-n}{6}\cdot$ The same sum represents the number of favorable cases. Dividing it by 6^n, we get the following expression for the probability of s points on n dice:

$$p = \frac{1}{6^n} \sum_{l=0}^{\frac{s-n}{6}} (-1)^l C_n^l C_{s-6l-1}^{n-1}.$$

The preceding problems suffice to illustrate how probability can be determined by direct enumeration of cases. For the benefit of students, a few simple problems without elaborate solutions are added here.

Problems for Solution

1. What is the probability of obtaining 9, 10, 11 points with 3 dice?
Ans. $\frac{25}{216}$, $\frac{27}{216}$, $\frac{27}{216}$.

2. What is the probability of obtaining 2 heads and 2 tails when 4 coins are thrown? *Ans.* $\frac{3}{8}$.

3. Two urns contain respectively 3 white, 7 red, 15 black balls, and 10 white, 6 red, 9 black balls. One ball is taken from each urn. What is the probability that they both will be of the same color? *Ans.* $\frac{207}{625}$.

4. What is the probability that of 6 cards taken from a full pack, 3 will be black and 3 red. *Ans.* $\frac{13000}{39151} = 0.332$ approximately.

5. Ten cards are taken from a full pack. What is the probability of finding among them (*a*) at least one ace; (*b*) at least two aces? *Ans.* $\frac{349}{595}$; $\frac{1257}{7735}$.

6. The face cards are removed from a full pack. Out of the 40 remaining cards, 4 are drawn. What is the probability that they belong to different suits?
Ans. $\frac{1000}{9139}$.

7. Under the same conditions, what is the probability that the 4 cards belong to different suits and different denominations? *Ans.* $\frac{504}{9139}$.

8. Five cards are taken from a full pack. Find the probabilities (*a*) that they are of different denominations; (*b*) that 2 are of the same denomination and 3 scattered; (*c*) that one pair is of one denomination and another pair of a different denomination, and one odd; (*d*) that 3 are of the same denomination and 2 scattered; (*e*) that 2 are of one denomination and 3 of another; (*f*) that 4 are of one denomination and 1 of another.
Ans. (*a*) $\frac{2112}{4165}$; (*b*) $\frac{1760}{4165}$; (*c*) $\frac{198}{4165}$; (*d*) $\frac{88}{4165}$; (*e*) $\frac{6}{4165}$; (*f*) $\frac{1}{4165}$.

9. What is the probability that 5 tickets taken in succession in the French lottery will present an increasing or decreasing sequence of numbers? *Ans.* $\frac{1}{60}$.

10. What is the probability that among 5 tickets drawn in the French lottery there is at least one with a one-digit number? *Ans.* $\frac{46282}{110983} = 0.417$.

11. Twelve balls are distributed at random among three boxes. What is the probability that the first box will contain 3 balls? *Ans.* $\frac{55.2^{11}}{3^{12}} = 0.2120$.

12. In Prob. 12 (page 22) what is the most probable number of balls in a specified box? *Ans.* The probability

$$p_h = \frac{C_n^h(N-1)^{n-h}}{N^n}$$

is the greatest if the integer h is determined by the conditions

$$\frac{n+1}{N} - 1 \leqq h \leqq \frac{n+1}{N}.$$

13. Apply these considerations to the case of $n = 200$, $N = 20$. *Ans.* Since $h = 10$ the inequalities on page 23 give

$$p_{10} < \frac{10^{10}}{10!}\left(1 - \frac{1}{20}\right)^{190}\left(1 - \frac{1}{40}\right)^{9}$$

$$p_{10} > \frac{10^{10}}{10!}\left(1 - \frac{1}{20}\right)^{190}\left(1 - \frac{1}{20}\right)^{\frac{9}{2}}.$$

To find an approximate value of

$$(1 - \tfrac{1}{20})^{190}$$

note that

$$(1 - \tfrac{1}{20})^{190} = e^{-190\left[\frac{1}{20}+\frac{1}{2\cdot 20^2}+\frac{1}{3\cdot 20^3}+\cdots\right]}.$$

To 3 decimals

$$p_{10} = 0.128.$$

14. Four different objects, 1, 2, 3, 4, are distributed at random on four places marked 1, 2, 3, 4. What is the probability that none of the objects occupies the place corresponding to its number? *Ans.* $\frac{3}{8}$.

15. Two urns contain, respectively, 1 black and 2 white balls, and 2 black and 1 white ball. One ball is transferred from the first urn into the second, after which a ball is drawn from the second urn. What is the probability that it is white?

Ans. $\frac{5}{12}$.

16. What is the probability of getting 20 points with 6 dice?

Ans. $\frac{4221}{46656}$ = 0.09047.

17. An urn contains a white and b black balls. Balls are drawn one by one until only those of the same color are left. What is the probability that they are white?

Ans. $\dfrac{a}{a+b}$.

18. In an urn there are n groups of p objects each. Objects in different groups are distinguished by some characteristic property. What is the probability that among $\alpha_1 + \alpha_2 + \cdots + \alpha_n$ objects ($0 \leqq \alpha_i \leqq p$; $i = 1, 2, \ldots n$) taken, there are α_1 of one group, α_2 of another, etc.? *Ans.* Let λ among the numbers $\alpha_1, \alpha_2, \ldots \alpha_n$ be equal to a, μ be equal to b, . . . σ be equal to l. The required probability is

$$\frac{n!}{\lambda!\mu!\cdots\sigma!}\,\frac{C_p^{\alpha_1}C_p^{\alpha_2}\cdots C_p^{\alpha_n}}{C_{np}^{\alpha_1+\alpha_2\cdots+\alpha_n}}.$$

Problem 8 is a particular case of this.

19. There are N tickets numbered 1, 2, . . . N of which n are taken at random and arranged in increasing order of their numbers: $x_1 < x_2 < \cdots < x_n$. What is the probability that $x_m = M$? *Ans.* $\dfrac{C_{M-1}^{m-1}C_{N-M}^{n-m}}{C_N^n}$.

CHAPTER II

THEOREMS OF TOTAL AND COMPOUND PROBABILITY

1. As the problems become more complex the difficulties in enumerating cases grow and often the computation of probabilities by direct application of definition becomes very involved. In many cases the complications can be avoided by use of two theorems which are fundamental in the theory of probability.

Before we can give a clear and exact statement of the first fundamental theorem, we must define what is meant by "mutually exclusive" or "incompatible" events. Events are called mutually exclusive or incompatible if the occurrence of one of them precludes the occurrence of all the others. For instance, the four events concerning the number of points on two dice

First Die	Second Die
1	4
2	3
3	2
4	1

are mutually exclusive because it is evident that as soon as one of them occurs, none of the others can materialize.

On the contrary, events are compatible if it is possible for them to materialize simultaneously. For instance, the events of 5 points on one die and 5 points on the other, are compatible, since in tossing two dice it is possible to get 5 points on each.

To denote the probability of an event A, we shall use the symbol (A). To denote the probability of A or B (or both) we shall use the symbol $(A + B)$. Dealing with several events $A, B, \ldots L$, the symbol

$$(A + B + \cdots + L)$$

will denote the probability of the occurrence of at least one of them. If $A, B, \ldots L$ are mutually exclusive events, this symbol represents the probability of the occurrence of one of them without specification as to which one.

2. Now we shall state the first fundamental theorem, called the "theorem of total probability" or "theorem of addition of probabilities," in the following way:

Theorem of Total Probability. *The probability for one of the mutually exclusive events $A_1, A_2, \ldots A_n$ to materialize, is the sum of the probabilities* *of these events.* In symbolical notations, it is expressed thus:

$$(A_1 + A_2 + \cdots + A_n) = (A_1) + (A_2) + \cdots + (A_n).$$

Proof. Let N be the number of all possible and equally likely cases out of which m_1 cases are favorable to the event A_1, m_2 cases are favorable to the event A_2, . . . , and finally, m_n cases are favorable to the event A_n. These cases are all different, since events $A_1, A_2, \ldots A_n$ are incompatible. The number of cases favorable to either A_1 or A_2, . . . or A_n is therefore

$$m_1 + m_2 + \cdots + m_n.$$

Hence, by definition

$$(A_1 + A_2 + \cdots + A_n) = \frac{m_1 + m_2 + \cdots + m_n}{N} = \frac{m_1}{N} + \frac{m_2}{N} + \cdots + \frac{m_n}{N}.$$

Again, by definition of probability,

$$\frac{m_1}{N} = (A_1); \qquad \frac{m_2}{N} = (A_2); \cdots \frac{m_n}{N} = (A_n),$$

and so finally

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

as stated.

3. It is important to know that the same theorem, stated in a slightly different form, is especially useful in applications. An event A can occur in several mutually exclusive forms, $A_1, A_2, \ldots A_n$, which may be considered as that many mutually exclusive events. Whenever A occurs, one of these events must occur, and conversely. Consequently, the probability of A is the same as the probability of one (unspecified) of its mutually exclusive forms. If, for instance, occurrence of 5 points on two dice is A, then this event occurs in 4 mutually exclusive forms, as tabulated above.

From the new point of view, the theorem of total probability can be stated thus:

Second Form of Theorem of Total Probability. *The probability of an event A is the sum of the probabilities of its mutually exclusive forms $A_1, A_2, \ldots A_n$; or, using symbols,*

$$(A) = (A_1) + (A_2) + \cdots + (A_n).$$

Probabilities $(A_1), (A_2), \ldots (A_n)$ are partial probabilities of incompatible forms of A. Since the probability A is their sum, it may be called a total probability of A. Hence the name of the theorem.

In the preceding example we saw that 5 points on two dice could be obtained in 4 mutually exclusive ways. Now the probability of any one of these ways is $1/36$; hence, by the preceding theorem, the probability of obtaining 5 points with two dice is

$$\tfrac{1}{36} + \tfrac{1}{36} + \tfrac{1}{36} + \tfrac{1}{36} = \tfrac{4}{36} = \tfrac{1}{9},$$

as it should be.

If events $A_1, A_2, \ldots A_n$ are not only mutually exclusive, but "exhaustive," which means that one of them must necessarily take place, the probability that one of them will happen is a certainty = 1, so that we must have

$$(A_1) + (A_2) + \cdots + (A_n) = 1.$$

An event which is *not* certain, may or may not happen; this constitutes two mutually exclusive cases. It is customary to call nonoccurrence of a certain event A as the "event opposite" to A, and we shall denote it by the symbol $\bar{A}$. Now A and $\bar{A}$ constitute two exhaustive and mutually exclusive cases. Hence, by the preceding remark

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

That is, if p is the probability of A

$$q = 1 - p$$

represents the probability that A will not occur.

4. If an event A is considered in connection witn another event B, the *compound* event AB consists in simultaneous occurrence of A *and* B. For three events A, B, C, the compound event ABC consists in simultaneous occurrence of A *and* B *and* C, and so on for any number of component events. We shall denote the probability of a compound event $AB \ldots L$ by the symbol

$$(AB \ldots L).$$

An event A can materialize in two mutually exclusive forms, namely, as A *and* B or A *and* $\bar{B}$. Hence, by the theorem of total probability

$$(A) = (AB) + (A\bar{B}).$$

Similarly

$$(B) = (BA) + (B\bar{A}),$$

or, since the symbol (BA) does not depend upon the order of letters,

$$(B) = (AB) + (\bar{A}B).$$

The sum $(A) + (B)$ can be expressed as

$$(A) + (B) = (AB) + [(A\bar{B}) + (\bar{A}B) + (AB)].$$

Again, by the theorem of total probabilities, the sum

$$(A\bar{B}) + (\bar{A}B) + (AB)$$

represents the probability $(A + B)$ of the occurrence of at least one of the events A or B. The preceding equation leads to the useful formula

$$(1) \qquad (A + B) = (A) + (B) - (AB)$$

which obviously is a generalization of the theorem of total probability; for $(AB) = 0$ if A and B are incompatible. Equation (1) can be used to derive an important inequality. Since $(A + B) \leqq 1$, it follows from (1) that

$$(AB) \geqq (A) + (B) - 1.$$

If B itself is a compound event A_1A_2, this inequality leads to

$$(AA_1A_2) \geqq (A) + (A_1A_2) - 1.$$

But

$$(A_1A_2) \geqq (A_1) + (A_2) - 1,$$

and so

$$(AA_1A_2) \geqq (A) + (A_1) + (A_2) - 2$$

for three component events. Proceeding in the same manner, we can establish the following general inequality:

$$(AA_1A_2 \cdots A_{n-1}) \geqq (A) + (A_1) + (A_2) + \cdots + (A_{n-1}) - (n - 1).$$

Applying this inequality to events $\bar{A}, \bar{A}_1, \ldots \bar{A}_{n-1}$ respectively opposite to $A, A_1, \ldots A_{n-1}$, we get

$$(\bar{A}\bar{A}_1 \cdots \bar{A}_{n-1}) \geqq (\bar{A}) + (\bar{A}_1) + \cdots + (\bar{A}_{n-1}) - (n - 1),$$

or, since $(\bar{A}_i) = 1 - (A_i)$,

$$(A) + (A_1) + \cdots + (A_{n-1}) \geqq 1 - (\bar{A}\bar{A}_1 \cdots \bar{A}_{n-1}).$$

Now the compound event $\bar{A}\bar{A}_1 \ldots \bar{A}_{n-1}$ means that neither A nor $A_1, \ldots$ nor A_{n-1} occurs. The event opposite to this is that at least one of the events $A, A_1, \ldots A_{n-1}$ occurs. Hence,

$$1 - (\bar{A}\bar{A}_1 \cdots \bar{A}_{n-1}) = (A + A_1 + \cdots + A_{n-1}),$$

and we reach the following important inequality:

$$(A + A_1 + \cdots + A_{n-1}) \leqq (A) + (A_1) + \cdots + (A_{n-1}).$$

5. Equation (1) can be extended to the case of more than two events. Let B mean the occurrence of at least one of the events A_1 or A_2. Then by (1)

$$(A + A_1 + A_2) = (A) + (A_1 + A_2) - (AB).$$

As to $(A_1 + A_2)$, its expression is given by (1). The compound event AB means the occurrence of one at least of the events AA_1 or AA_2. Hence, applying equation (1) once more, we find

$$(AB) = (AA_1 + AA_2) = (AA_1) + (AA_2) - (AA_1A_2)$$

and after due substitutions

$$(A + A_1 + A_2) = (A) + (A_1) + (A_2) - (AA_1) - (AA_2) - (A_1A_2) + \\ + (AA_1A_2).$$

Proceeding in the same way and using mathematical induction, the following general formula can be established:

$$(A + A_1 + \cdots + A_{n-1}) = \sum(A_i) - \sum_{i,j}(A_iA_j) + \sum_{i,j,k}(A_iA_jA_k) - \cdots$$

where summations refer to all combinations of subscripts taken from numbers 0, 1, 2, . . . $n - 1$, one, two, three, . . . , and n at a time.

6. Let A and B be two events whose probabilities are (A) and (B). It is understood that the probability (A) is determined without any regard to B when nothing is known about the occurrence or nonoccurrence of B. When it *is* known that B occurred, A may have a different probability, which we shall denote by the symbol (A, B) and call "conditional probability of A, given that B has actually happened."

Now we can state the second fundamental theorem, called the "theorem of compound probability" or "theorem of multiplication of probabilities," as follows:

Theorem of Compound Probability. *The probability of simultaneous occurrence of A and B is given by the product of the unconditional probability of the event A by the conditional probability of B, supposing that A actually occurred.* In other words,

$$(AB) = (A) \cdot (B, A).$$

or $P(A \cap B) = P($

Proof. Let N denote the total number of equally likely cases among which m cases are favorable to the event A. The cases favorable to A *and* B are to be found among the m cases favorable to A. Let their number be m_1. Then, by the definition of probability,

$$(AB) = \frac{m_1}{N},$$

which also can be written thus:

$$(AB) = \frac{m}{N} \cdot \frac{m_1}{m}.$$

Now the ratio m/N represents the probability of A. To find the meaning of the second factor, we observe that, assuming the occurrence of A, there are only m equally likely cases left (the remaining $N - m$ cases becoming impossible) out of which m_1 are favorable to B. Hence the ratio m_1/m represents the conditional probability (B, A) of B supposing that A has actually happened.

Now since

$$\frac{m}{N} = (A), \qquad \frac{m_1}{m} = (B, A),$$

the probability of the compound event AB is expressed by the product

$$(AB) = (A) \cdot (B, A).$$

Since the compound event AB involves A and B symmetrically, we shall have also

$$(AB) = (B) \cdot (A, B).$$

The theorem of compound probability can easily be extended to several events. For example, let us consider three events, A, B, C. The occurrence of A *and* B *and* C is evidently equivalent to the occurrence of the compound event AB *and* C. We have, therefore,

$$(ABC) = (AB) \cdot (C, AB)$$

by the theorem of compound probability. By the same theorem

$$(AB) = (A) \cdot (B, A),$$

so that

$$(ABC) = (A) \cdot (B, A) \cdot (C, AB).$$

Obviously this formula can be extended to compound events consisting of more than three components.

In one particular but very important case, the expression for the compound probability can be simplified; namely, in the case of so-called "independent events." Several events are "independent" by definition if the probability of any one of them is not affected by supplementary knowledge concerning the materialization of any number of the remaining events. For instance, if A and B represent white balls drawn from two different urns, the probability of A is the same whether the color of the ball drawn from the other urn is known or not. Similarly, granted that a coin is unbiased, heads at the first throw and heads at the second throw are independent events. In such theoretical cases the independence of events can be reasonably assumed or agreed upon. In other cases, and especially in practical applications, it is not easy to decide whether events should be considered as independent or not.

If A and B are independent, the conditional probability (B, A) is the same as the probability (B) found without any reference to A; this follows from the definition of independence. Hence, the expression of compound probability (AB) for two independent events becomes

$$(AB) = (A) \cdot (B)$$

so that the probability of a compound event with independent components is simply equal to the product of the probabilities of component events. This rule extends to any number of component events if they are independent. Let us consider three independent events, A, B, and C. The independence of these events implies

$$(B, A) = (B); \qquad (C, AB) = (C)$$

and hence

$$(ABC) = (A) \cdot (B) \cdot (C)$$

in accordance with the rule.

To illustrate the theorem of compound probability, let us consider two simple examples. An urn contains 2 white balls and 3 black ones. Two balls are drawn, and it is required to find the probability that they are both white. Let A be the event consisting in the white color of the first ball, and B the event consisting in the white color of the second ball. The probability (A) of extracting a white ball in the first place is

$$(A) = \tfrac{2}{5}.$$

To find the conditional probability (B, A) we observe, after drawing one white ball, that 1 white and 3 black balls remain in the urn. The probability of drawing a white ball under such circumstances is

$$(B, A) = \tfrac{1}{4}.$$

Now, by the theorem of compound probability, we shall have

$$(AB) = \tfrac{2}{5} \cdot \tfrac{1}{4} = \tfrac{1}{10}.$$

Evidently, in this example we dealt with dependent events.

As an example of independent events, let a coin be tossed any given number of times; say, n times. What is the probability of having only heads? The compound event in this example consists of n independent components; namely, heads at every trial. Now the probability of heads in any trial is $\frac{1}{2}$, and so the required probability will be $1/2^n$.

NOTE: Two events A and B are independent by definition, if

$$(A, B) = (A) \quad \text{and} \quad (B, A) = (B).$$

However, one of these conditions follows from the other. Suppose the condition

$$(A, B) = (A)$$

is fulfilled, so that A is independent of B. We have then

$$(AB) = (B) \cdot (A).$$

On the other hand,

$$(AB) = (A) \cdot (B, A),$$

whence

$$(B, A) = (B),$$

so that B is independent of A.

Three events A, B, C are independent if the following four conditions are fulfilled:

$$(A, B) = (A); \qquad (A, C) = (A); \qquad (B, C) = (B); \qquad (C, AB) = (C).$$

From the first three conditions it follows that

$$(B, A) = (B); \qquad (C, A) = (C); \qquad (C, B) = (C).$$

To show that the other requirements

$$(B, AC) = (B); \qquad (A, BC) = (A)$$

are also fulfilled, we notice that

$$(ABC) = (A) \cdot (B, A) \cdot (C, AB) = (A) \cdot (B) \cdot (C)$$

because $(C, AB) = (C)$ by hypothesis and $(B, A) = (B)$ as proved. On the other hand,

$$(ABC) = (A) \cdot (C, A) \cdot (B, AC)$$

and $(C, A) = (C)$. Hence, comparing with the preceding expression,

$$(B, AC) = (B).$$

Similarly, it can be shown that

$$(A, BC) = (A).$$

The independence of four events A, B, C, D is assured if the following 11 conditions are fulfilled:

$$(A, B) = (A, C) = (A, D) = (A); \qquad (B, C) = (B, D) = (B); \qquad (C, D) = (C);$$
$$(C, AB) = (C); \qquad (D, AB) = (D, AC) = (D, BC) = (D); \qquad (D, ABC) = (D).$$

And in general, independence of n events is assured if $2^n - n - 1$ conditions of similar type are fulfilled.

If several events are independent, every two of them are independent; but this does not suffice for the independence of *all* events, as can be shown by a simple example. An urn contains four tickets with numbers 112, 121, 211, 222, and one ticket is drawn. What are the probabilities that the first, second, or third digits in its number are 1? Let a unit such as the first, second, or third digit, be represented, respectively by A, B, or C. Then

$$(A) = (B) = (C) = \tfrac{2}{4} = \tfrac{1}{2}.$$

Compound probabilities (AB), (AC), (BC) are

$$(AB) = (AC) = (BC) = \tfrac{1}{4},$$

since among four tickets there is only one whose number has first and second, or first and third, or second and third digits of 1. Now, for instance,

$$(AB) = \tfrac{1}{4} = \tfrac{1}{2} \cdot \tfrac{1}{2} = (A) \cdot (B),$$

whence A and B are independent. Similarly, A and C; C and B are independent. Thus, any two of the events A, B, C are independent, but *not all three* events are. For, if they were, we should have

$$(ABC) = \tfrac{1}{8}.$$

But $(ABC) = 0$ since in no ticket are all three digits equal to 1.

7. The theorems of total and compound probability form the foundation of the theory of probability as it represents a separate branch of mathematical science. They serve the purpose of finding probabilities in more complicated cases, either by being directly applied or by enabling us to form equations from which the required probabilities can be found. A few selected problems will illustrate the various ways of using these theorems.

Problem 14. An urn contains a white balls and b black balls; another contains c white and d black balls. One ball is transferred from the first urn into the second, and then a ball is drawn from the latter. What is the probability that it will be a white ball?

Solution. The event consisting in the white color of the ball drawn from the second urn, can materialize under two mutually exclusive forms: when the transferred ball is a white one, and when it is black. By the theorem of total probability, we must find the probabilities corresponding to these two forms. To find the probability of the first form, we observe that it represents a compound event consisting in the white color of the transferred ball, combined with the white color of the extracted ball. The probability that the transferred ball is white is given by the fraction

$$\frac{a}{a+b}$$

and the probability that the ball removed from the second urn is white, is

$$\frac{c+1}{c+d+1}$$

because before the drawing there were $c + 1$ white balls and d black balls in the second urn. Hence, by the theorem of compound probability, the probability of the first form is

$$\frac{a(c+1)}{(a+b)(c+d+1)}.$$

In the same way, we find that the probability of the second form is

$$\frac{bc}{(a+b)(c+d+1)},$$

and the sum of these two numbers

$$\frac{ac+bc+a}{(a+b)(c+d+1)}$$

gives the probability of extracting a white ball from the second urn, after one ball of unknown color has been transferred from the first urn.

8. Problem 15. Two players agree to play under the following conditions: Taking turns, they draw the balls out of an urn containing a white balls and b black balls, one ball at a time. He who extracts the first white one wins the game. What is the probability that the player who starts will win the game?

Solution. Let A be the player who draws the first ball, and let B be the other player. The game can be won by A, first, if he extracts a white ball at the start; second, if A and B alternately extract 2 black balls and then A draws a white one; third, if A and B alternately extract 4 black balls and the fifth ball drawn by A is white; and so on. By the theorem of total probability, the probability for A to win the game, is the sum of the probabilities of the mutually exclusive ways (described above) in which he can win the game. The probability of extracting a white ball at first is

$$\frac{a}{a+b}.$$

The probability of extracting 2 black balls and then 1 white ball is found by direct application of the theorem of compound probabilities. Its expression is

$$\frac{b(b-1)a}{(a+b)(a+b-1)(a+b-2)}.$$

The probability of extracting 4 black balls and then 1 white ball is given by

$$\frac{b(b-1)(b-2)(b-3)a}{(a+b)(a+b-1)(a+b-2)(a+b-3)(a+b-4)},$$

using the same theorem of compound probability.

In the same way we deal with all the possible and mutually exclusive ways which would allow A to win the game. Then, by adding the above given expressions of partial probabilities, we obtain the expression for the required probability in the form of the sum

$$P = \frac{a}{a+b}\left[1 + \frac{b(b-1)}{(a+b-1)(a+b-2)} + \right.$$
$$\left. + \frac{b(b-1)(b-2)(b-3)}{(a+b-1)(a+b-2)(a+b-3)(a+b-4)} + \cdots\right].$$

The law of formation of different terms in this sum is obvious; and the sum automatically ends as soon as we arrive at a term which is equal to zero.

In the same way, we can find that the probability for the player B to win is expressed by an analogous sum:

$$Q = \frac{a}{a+b}\left[\frac{b}{a+b-1} + \frac{b(b-1)(b-2)}{(a+b-1)(a+b-2)(a+b-3)} + \cdots\right].$$

But one of the players, A or B, *must* win the game, and the winning of the game by A and B are opposite events. Hence,

$$P + Q = 1$$

or, after substituting the above expressions for P and Q and after obvious simplifications,

$$1 + \frac{b}{a+b-1} + \frac{b(b-1)}{(a+b-1)(a+b-2)} + \cdots = \frac{a+b}{a}.$$

This is a noteworthy identity, obtained, as we see, by the principles of the theory of probability. Of course, it can be proved in a direct way, and it would be a good problem for students to attempt a direct proof. There are many cases in which, by means of considerations belonging to the theory of probability, several identities or inequalities can be established whose direct proof sometimes involves considerable difficulty.

9. Problem 16. Each of k urns contains n identical balls numbered from 1 to n. One ball is drawn from every urn. What is the probability that m is the greatest number drawn?

Solution. Let us denote by P_m the required probability. It is not apparent how we can find the explicit expression for this probability, but using the theorems of total and compound probability, we can form equations which yield the desired expression for P_m without any difficulty. To this end, let us first find the probability P that the greatest number drawn does not exceed m. It is obvious that this may happen in m mutually exclusive ways; namely, when the greatest number drawn is 1, 2, 3, and so on up to m. The probabilities of these different hypotheses being $P_1, P_2, \ldots P_m,$ their sum gives the following first expression for P:

(1) $$P = P_1 + P_2 + \cdots + P_m.$$

We can find the second expression for P using the theorem of compound probability; namely, the greatest number drawn does not exceed m if balls drawn from all urns have numbers from 1 to m. The probability of drawing a ball with the number 1, 2, 3, . . . m from any urn is m/n. And the probability that this will happen for every urn is a compound event consisting of k independent events with the same probability m/n. Therefore, by the theorem of compound probability

$$P = \frac{m^k}{n^k}.$$

And this compared with (1) gives the equation

$$P_1 + P_2 + \cdots + P_m = \frac{m^k}{n^k}. \tag{2}$$

Substituting $m - 1$ for m in this equation, we get

$$P_1 + P_2 + \cdots + P_{m-1} = \frac{(m-1)^k}{n^k},$$

and it suffices to subtract this from (2) to have the required expression for P_m:

$$P_m = \frac{m^k - (m-1)^k}{n^k}.$$

10. Problem 17. Two persons, A and B, have respectively $n + 1$ and n coins, which they toss simultaneously. What is the probability that A will have more heads than B?

Solution. Let μ, μ' and ν, ν' be numbers of heads and tails thrown by A and B, respectively, so that $\mu + \nu = n + 1$, $\mu' + \nu' = n$. The required probability P is the probability of the inequality $\mu > \mu'$. The probability $1 - P$ of the opposite event $\mu \leqq \mu'$ is at the same time the probability of the inequality $\nu > \nu'$; that is, $1 - P$ is the probability that A will throw more tails than B. By reason of symmetry $1 - P = P$, $P = \frac{1}{2}$.

11. Problem 18. Three players A, B, and C agree to play a series of games observing the following rules: two players participate in each game, while the third is idle, and the game is to be won by one of them. The loser in each game quits, and his place in the next game is taken by the player who was idle. The player who succeeds in winning over both of his opponents without interruption, wins the whole series of games. Supposing that the probability for each player to win a single game is $\frac{1}{2}$ and that the first game is played by A and B, find the probability for A, B, and C, respectively, to win the whole series, if (a) the number of games to be played is limited and may not exceed a given number n; if (b) the number of games is unlimited.

Solution. Let P_n, Q_n, R_n be the probabilities for A, B, and C, respectively, to win a series of games when their number cannot exceed n. By reason of symmetry, $P_n = Q_n$ so that it remains to find P_n and R_n. The player A can win the whole series of games in two mutually exclusive ways: if he wins the first game, or if he loses the first game. Let the probability of the first case be p_n and that of the second r_n. Then

$$P_n = p_n + r_n.$$

A can win the whole series after winning the first game, in two mutually exclusive ways: (*a*) if he wins over *B* and *C* in succession; (*b*) if he wins the first game from *B* and loses the second game to *C*; then, if in the third game *C* loses to *B*, and in the fourth game *A* wins over *B* and later wins the whole series of not more than $n - 3$ games. Now, the probability of case (*a*) is $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ by the theorem of compound probability; that of case (*b*) by the same theorem is $\frac{1}{8}p_{n-3}$; and the total probability is

$$p_n = \tfrac{1}{4} + \tfrac{1}{8}p_{n-3}. \tag{1}$$

If *A* loses the first game to *B*, but wins the whole series, then in the second game *C* wins over *B* while the third game is won by *A*, and not more than $n - 2$ games are left to play. Hence,

$$r_n = \tfrac{1}{4}p_{n-2}. \tag{2}$$

Since evidently $p_2 = p_3 = p_4 = \frac{1}{4}$, equation (1) by successive substitutions yields

$$p_{3k} = \frac{1}{4}\left(1 + \frac{1}{8} + \frac{1}{8^2} + \cdots + \frac{1}{8^{k-1}}\right)$$
$$p_{3k+1} = \frac{1}{4}\left(1 + \frac{1}{8} + \frac{1}{8^2} + \cdots + \frac{1}{8^{k-1}}\right)$$
$$p_{3k+2} = \frac{1}{4}\left(1 + \frac{1}{8} + \frac{1}{8^2} + \cdots + \frac{1}{8^{k}}\right)$$

or, in condensed form for an arbitrary n

$$p_n = \tfrac{2}{7}\left(1 - 8^{-\left[\frac{n+1}{3}\right]}\right),$$

denoting by $[x]$ the greatest integer contained in x. Hence, by virtue of (2) the general expression of r_n will be

$$r_n = \tfrac{1}{14}\left(1 - 8^{-\left[\frac{n-1}{3}\right]}\right)$$

and that of P_n, Q_n,

$$P_n = Q_n = \tfrac{5}{14} - \tfrac{4}{14}8^{-\left[\frac{n+1}{3}\right]} - \tfrac{1}{14}8^{-\left[\frac{n-1}{3}\right]}.$$

Finally, to find the probability for *C* to win, we observe that this can happen only if *C* wins the second game; hence,

$$R_n = p_{n-1} = \tfrac{4}{14} - \tfrac{4}{14}8^{-\left[\frac{n}{3}\right]}.$$

Since $P_n + Q_n + R_n < 1$, the difference

$$1 - P_n - Q_n - R_n = \tfrac{4}{14}8^{-\left[\frac{n+1}{3}\right]} + \tfrac{4}{14}8^{-\left[\frac{n}{3}\right]} + \tfrac{1}{14}8^{-\left[\frac{n-1}{3}\right]}$$

represents the probability of a tie in n games. This probability decreases rapidly when n increases, so that in a long series of games a tie is practically impossible. If the number of games is not limited, the probabilities P, Q, R for A, B, C, respectively, to win are obtained as limits of P_n, Q_n, R_n, when n increases indefinitely. Thus

$$P = Q = \tfrac{5}{14}, \qquad R = \tfrac{4}{14}.$$

Problems for Solution

1. Three urns contain respectively 1 white and 2 black balls; 3 white and 1 black ball; 2 white and 3 black balls. One ball is taken from each urn. What is the probability that among the balls drawn there are 2 white and 1 black? *Ans.* $\frac{23}{60}$.

2. Cards are drawn one by one from a full deck. What is the probability that 10 cards will precede the first ace? *Ans.* $\frac{164}{4165} = 0.03938$.

3. Urn 1 contains 10 white and 3 black balls; urn 2 contains 3 white and 5 black balls. Two balls are transferred from No. 1 and placed in No. 2 and then one ball is taken from the latter. What is the probability that it is a white ball? *Ans.* $\frac{59}{130}$.

4. Two urns identical in appearance contain respectively 3 white and 2 black balls; 2 white and 5 black balls. One urn is selected and a ball taken from it. What is the probability that this ball is white? *Ans.* $\frac{31}{70}$.

5. What is the probability that 5 tickets drawn in the French lottery all have one-digit numbers? *Ans.* $\frac{7}{2441626} = 29.10^{-7}$

6. What is the probability that each of the four players in a bridge game will get a complete suit of cards? *Ans.* $24\dfrac{(1 \cdot 2 \cdots 13)^4}{1 \cdot 2 \cdots 52} = 4.474.10^{-28}$.

7. What is the probability that at least one of the players in a bridge game will get a complete suit of cards?

$$\textit{Ans.}\ \frac{16 \cdot 13! \cdot 39! - 72 \cdot (13!)^2 \cdot 26! + 72 \cdot (13!)^4}{52!} = 2.52 \cdot 10^{-11}.$$

See Sec. 5, page 31.

8. From an urn with a white and b black balls n balls are taken. Find the probability of drawing at least one white ball. *Ans.* The required probability can be expressed in two ways. First expression:

$$1 - \frac{b(b-1)\cdots(b-n+1)}{(a+b)(a+b-1)\cdots(a+b-n+1)}.$$

Second expression:

$$\frac{a}{a+b}\left[1 + \frac{b}{a+b-1} + \cdots + \frac{b(b-1)\cdots(b-n+2)}{(a+b-1)(a+b-2)\cdots(a+b-n+1)}\right].$$

Equating them, we have an identity

$$1 + \frac{b}{a+b-1} + \cdots + \frac{b(b-1)\cdots(b-n+2)}{(a+b-1)(a+b-2)\cdots(a+b-n+1)} =$$
$$= \frac{a+b}{a}\left[1 - \frac{b(b-1)\cdots(b-n+1)}{(a+b)(a+b-1)\cdots(a+b-n+1)}\right].$$

9. Three players A, B, C in turn draw balls from an urn with 10 white and 10 black balls, taking one ball at a time. He who extracts the first white ball wins the game. Supposing that they start in the order A, B, C, find the probabilities for each of them to win the game. *Ans.* For A, 0.56584; for B, 0.29144; for C, 0.14271.

10. If n dice are thrown at a time, what is the probability of having each of the points 1, 2, . . . 6, appear at least once? Find the numerical value of this probability for $n = 10$. *Ans.*

$$p_n = 1 - 6(\tfrac{5}{6})^n + 15(\tfrac{4}{6})^n - 20(\tfrac{3}{6})^n + 15(\tfrac{2}{6})^n - 6 \cdot (\tfrac{1}{6})^n$$
$$p_{10} = 0.2718.$$

HINT: Use the formula in Sec. 5, page 31.

11. In a lottery m tickets are drawn at a time out of the total number of n tickets, and returned before the next drawing is made. What is the probability that in k drawings each of the numbers 1, 2, . . . n will appear at least once? *Ans.*

$$p_k = 1 - \frac{n}{1}\left(\frac{n-m}{n}\right)^k + \frac{n(n-1)}{1 \cdot 2}\left(\frac{n-m}{n}\right)^k\left(\frac{n-m-1}{n-1}\right)^k - \cdots .$$

12. We have k varieties of objects, each variety consisting of the same number of objects. These objects are drawn one at a time and replaced before the next drawing. Find the probability that n and no less drawings will be required to produce objects of all varieties. *Ans.*

$$k^{n-1}p_n = (k-1)^{n-1} - \frac{k-1}{1}(k-2)^{n-1} + \frac{(k-1)(k-2)}{1 \cdot 2}(k-3)^{n-1} - \cdots .$$

13. Three urns contain respectively 1 white, 2 black balls; 2 white, 1 black balls; 2 white, 2 black balls. One ball is transferred from the first urn into the second; then one from the latter is transferred into the third; finally, one ball is drawn from the third urn. What is the probability of its being white? *Ans.* $^{31}\!/_{60}$.

14. Each of n urns contains a white and b black balls. One ball is transferred from the first urn into the second, then one ball from the latter into the third, and so on. Finally, one ball is taken from the last urn. What is the probability of its being white? *Ans.* Denote by p_k the probability of drawing a white ball from the kth urn. Then

$$p_{k+1} = \frac{a+1}{a+b+1}p_k + \frac{a}{a+b+1}(1 - p_k)$$

for $k = 1, 2, \ldots n - 1$. Hence,

$$p_n = \frac{a}{a+b}.$$

15. Two players A and B toss two dice, A starting the game. The game is won by A if he casts 6 points before B casts 7 points; and it is won by B if he casts 7 points before A casts 6 points. What are the probabilities for A and B to win the game if they agree to cast dice not more than n times? What is the probability of a tie? *Ans.* Probability for A:

$$p_n = \tfrac{30}{61}[1 - (\tfrac{155}{216})^m] \quad \text{if} \quad n = 2m$$
$$p_n = \tfrac{30}{61}[1 - (\tfrac{155}{216})^{m+1}] \quad \text{if} \quad n = 2m + 1.$$

Probability for B:

$$q_n = \tfrac{31}{61}[1 - (\tfrac{155}{216})^m] \quad \text{if} \quad n = 2m$$
$$q_n = \tfrac{31}{61}[1 - (\tfrac{155}{216})^m] \quad \text{if} \quad n = 2m + 1.$$

Probability of a tie:

$$r_n = \left(\tfrac{155}{216}\right)^m \quad \text{if} \quad n = 2m; \qquad r_n = \tfrac{31}{36}\left(\tfrac{155}{216}\right)^m \quad \text{if} \quad n = 2m + 1.$$

If n increases indefinitely, r_n converges to 0 and p_n, q_n converge to the limits

$$p = \tfrac{30}{61}, \qquad q = \tfrac{31}{61},$$

which may be considered as the probabilities for A and B to win if the number of throws is unlimited.

16. The game known as "craps" is played with two dice, and the caster wins unconditionally if he produces 7 or 11 points (which are called "naturals"); he loses the game in case of 2, 3, or 12 points (called "craps"). But if he produces 4, 5, 6, 8, 9, or 10 points, he has the right to cast the dice steadily until he throws the same number of points he had before or until he throws a 7. If he rolls 7 before obtaining his point, he loses the game; otherwise, he wins. What is the probability to win?

Ans. $244/495 = 0.493$.

17. Prove directly the identity in Prob. 15, page 37.

Solution 1. Let

$$\varphi(c, b) = \frac{b}{c} + \frac{b(b-1)}{c(c-1)} + \frac{b(b-1)(b-2)}{c(c-1)(c-2)} + \cdots$$

where b is a positive integer and $c > b$. Then

$$\varphi(c, b) = \frac{b}{c}[1 + \varphi(c - 1, b - 1)]$$

whence

$$\varphi(c, 1) = \frac{1}{c}; \qquad \varphi(c, 2) = \frac{2}{c-1}; \qquad \varphi(c, 3) = \frac{3}{c-2}$$

and in general

$$\varphi(c, b) = \frac{b}{c - b + 1}.$$

Taking $c = a + b - 1$, we have

$$1 + \varphi(a + b - 1, b) = \frac{a + b}{a}.$$

Solution 2. The polynomial

$$S(x) = 1 + \frac{b}{c}x + \frac{b(b-1)}{c(c-1)}x^2 + \cdots$$

can be presented in the form of a definite integral

$$S(x) = (c + 1)\int_0^1 (1 - \xi(1 - x))^b(1 - \xi)^{c-b}d\xi$$

whence

$$S(1) = (c + 1)\int_0^1 (1 - \xi)^{c-b}d\xi = \frac{c + 1}{c - b + 1} = \frac{a + b}{a}$$

if $c = a + b - 1$.

18. Find the approximate expressions for the probabilities P and Q in Prob. 15, page 36, when b is a large number. Take for numerical application $a = b = 50$.

Solution. Since $P + Q = 1$, it suffices to seek the approximate expression for $P - Q$. Now

$$P - Q = a\int_0^1 (1 - 2\xi)^b(1 - \xi)^{a-1}d\xi$$

whence

$$P - Q = \frac{a}{2}\int_0^1 (1 - u)^b\left(1 - \frac{u}{2}\right)^{a-1} du + \frac{(-1)^b}{2^a C_{a+b}^a}.$$

To find the approximate expression of this integral, we set

$$(1 - u)^b\left(1 - \frac{u}{2}\right)^{a-1} = e^{-v},$$

whence u can be expressed as a power series in v:

$$u = \frac{2}{2b + a - 1}v - \frac{4b + a - 1}{(2b + a - 1)^3}v^2 + \frac{12b^2 + (2b + a - 1)^2}{3(2b + a - 1)^5}v^3 - \cdots$$

Substituting the resulting expression of du/dv and integrating with respect to v between limits 0 and ∞, we obtain for $P - Q$ an asymptotic expansion whose first terms are

$$P - Q = \frac{a}{2b + a - 1}\left[1 - \frac{4b + a - 1}{(2b + a - 1)^2}\right] + \frac{a[12b^2 + (2b + a - 1)^2]}{(2b + a - 1)^5} + \frac{(-1)^b}{2^a C_{a+b}^a}.$$

A more detailed discussion reveals that the error of this approximate formula is less than $a(\frac{1}{2})^{b+1}(\frac{3}{4})^{a-1}$ and greater than $-\frac{a[40(a - 1)^2 - 6b(a - 1) + 32b^2]}{(2b + a - 1)^6}$ provided $b \geqq 12$. For $a = b = 50$ the formula yields

$$P - Q = 0.3318; \qquad P = 0.6659; \qquad Q = 0.3341.$$

References

Jacob Bernoulli: "Ars conjectandi," 1713.
Abr. de Moivre: "Doctrine of Chances," 3d ed., 1756.
Laplace: "Théorie analytique des probabilités," Oeuvres VII, 1886.
J. Bertrand: "Calcul des probabilités," Paris, 1889.
E. Czuber: "Wahrscheinlichkeitsrechnung," 1, Leipzig, 1908.
Whitworth: "Choice and Chance," 5th ed., 1901.
Castelnuovo: "Calcolo delle probabilità," vol. 1, Bologna, 1925.

CHAPTER III

REPEATED TRIALS

1. In the theory of probability the word "trial" means an attempt to produce, in a manner precisely described, an event E which is not certain. The outcome of a trial is called a "success" if E occurs, and a "failure" if E fails to occur. For instance, if E represents the drawing of two cards of the same denomination from a full pack of cards, the "trial" consists in taking any two cards from the full pack, and we have a success or failure in this trial according to whether both cards are of the same denomination or not.

If trials can be repeated, they form a "series" of trials. Regarding series of trials, the following two problems naturally arise:

a. What is the probability of a given number of successes in a given series of trials? And as a generalization of this problem:

b. What is the probability that the number of successes will be contained between two given limits in a given series of trials?

Problems of this kind are among the most important in the theory of probability.

2. Trials are said to be "independent" in regard to an event E if the probability of this event in any trial remains the same, whether the results of any number of other trials are known or not. On the other hand, trials are "dependent" if the probability of E in a certain trial varies according to the information we have about the outcome of one or more of the other trials.

As an example of independent trials, imagine that several times in succession we draw one ball from an urn containing white and black balls in given proportion, after each trial returning the ball that has been drawn, and thoroughly mixing the balls before proceeding to the next trial. With respect to the color of the balls taken, we may reasonably assume that these trials are independent. On the other hand, if the balls already extracted are not returned to the urn, the above described trials are no longer independent. To illustrate, suppose that the urn from which the balls are drawn, originally contained 2 white and 3 black balls, and that 4 balls are drawn. What is the probability that the third ball is white? If nothing is known about the color of the three other balls, the probability is $\frac{2}{5}$. If we know that the first ball is white, but the colors of the second and fourth balls are unknown, this probability is $\frac{1}{4}$. In general, the probability for any ball to be white (or black) depends essentially on the amount of information we possess about the color of the other balls. Since the urn contains a limited number of balls, series of trials of this kind cannot be continued indefinitely.

As an example of an indefinite series of dependent trials, suppose that we have two urns, the first containing 1 white and 2 black balls, and the second, 1 black and 2 white balls, and the trials consist in taking one ball at a time from either urn, observing the following rules: (*a*) the first ball is taken from the first urn; (*b*) after a white ball, the next is taken from the first urn; after a black one, the next is taken from the second urn; (*c*) balls are returned to the same urns from which they were taken.

Following these rules, we evidently have a definite series of trials, which can be extended indefinitely, and these trials are dependent. For if we know that a certain ball was white or black, the probability of the next ball being white is $\frac{1}{3}$ or $\frac{2}{3}$, respectively.

Assuming the independence of trials, the probability of an event E may remain constant or may vary from one trial to another. If an unbiased coin is tossed several times, we have a series of independent trials each with the same probability, $\frac{1}{2}$, for heads. It is easy to give an example of a series of independent trials with variable probability for the same event. Imagine, for instance, that we have an unlimited number of urns with white and black balls, but that the proportion of white and black balls varies from urn to urn. One ball is drawn successively from each of these urns. Evidently, here we have a series of trials independent in regard to the white color of the ball drawn, but with the probability of drawing a ball of this color varying from trial to trial.

In this chapter we shall discuss the simplest case of series of independent trials with constant probability. They are often called "Bernoullian series of trials" in honor of Jacob Bernoulli who, in his classical book, "Ars conjectandi" (1713) made a profound study of such series and was led to the discovery of one of the most important theorems in the theory of probability.

3. Considering a series of n independent trials in which the probability of an event E is p in every trial (that of the opposite event F being $q = 1 - p$), the first problem which presents itself is to find the probability that E will occur exactly m times, where m is one of the numbers $0, 1, 2, \ldots n$. In what follows, we shall denote this probability by T_m. In the extreme cases $m = n$ and $m = 0$ it is easy to find T_n and T_0. When $m = n$, the event E must occur n times in succession, so that T_n represents the probability of the compound event $EEE \ldots E$ with n identical components. These components are independent events, since the trials are independent, and the probability of each of them is p.

Hence, the compound probability is

$$T_n = p \cdot p \cdot p \cdots p \ (n \text{ times})$$

or

$$T_n = p^n.$$

The symbol T_0 denotes the probability that E will never occur in n trials, which is the same as to say that F will occur n times in succession. Hence, for the same reasons as before,

$$T_0 = q^n = (1 - p)^n.$$

When m is neither 0 nor n, the event consisting in m occurrences of E can materialize in several mutually exclusive forms, each of which may be represented by a definite succession of m letters E and $n - m$ letters F. For example, if $n = 4$ and $m = 2$, we can distinguish the following mutually exclusive forms corresponding to two occurrences of E:

EEFF, EFEF, EFFE, FEEF, FEFE, FFEE.

To find the number of all the different successions consisting of m letters E and $n - m$ letters F, we observe that any such succession is determined as soon as we know the places occupied by the letter E. Now the number of ways to select m places out of the total number of n places is evidently the number of combinations out of n objects taken m at a time. Hence, the number of mutually exclusive ways to have m successes in n trials is

$$C_n^m = \frac{n(n-1)\cdots(n-m+1)}{1 \cdot 2 \cdot 3 \cdots m}.$$

The probability of each succession of m letters E and $n - m$ letters F, by reason of independence of trials, is represented by the product of m factors p and $n - m$ factors q, and since the product does not depend upon the order of factors, this probability will be

$$p^m q^{n-m}$$

for each succession. Hence, the total probability of m successes in n trials is given by this simple formula:

(1) $$T_m = \frac{n(n-1)\cdots(n-m+1)}{1 \cdot 2 \cdot 3 \cdots m} p^m q^{n-m}$$

which can also be presented thus:

(2) $$T_m = \frac{n!}{m!(n-m)!} p^m q^{n-m}.$$

This second form can be used even for $m = 0$ or $m = n$ if, as usual, we assume $0! = 1$. Either of the expressions (1) or (2) shows that T_m may be considered as the coefficient of t^m in the expansion of

$$(q + pt)^n$$

according to ascending powers of an arbitrary variable t. In other words, we have identically

$$(q + pt)^n = T_0 + T_1 t + T_2 t^2 + \cdots + T_n t^n.$$

For this reason the function

$$(q + pt)^n$$

is called the "generating function" of probabilities $T_0, T_1, T_2, \ldots T_n$. By setting $t = 1$ we naturally obtain

$$T_0 + T_1 + T_2 + \cdots + T_n = 1.$$

4. The probability $P(k, l)$ that the number of successes m will satisfy the inequalities (or, simply, the probability of these inequalities)

$$k \leqq m \leqq l$$

where k and l are two given integers, can easily be found by distinguishing the following mutually exclusive events:

$$m = k \qquad \text{or} \qquad m = k + 1, \ldots \qquad \text{or} \qquad m = l.$$

Accordingly, by the theorem of total probability,

$$P(k, l) = T_k + T_{k+1} + \cdots + T_l$$

or, using expression (2),

$$P(k, l) = \sum_{m=k}^{l} \frac{n!}{m!(n - m)!} p^m q^{n-m}.$$

In particular, the probability that the number of successes will not be greater than l is represented by the sum

$$P(0, l) = q^n + \frac{n}{1} pq^{n-1} + \frac{n(n-1)}{1 \cdot 2} p^2 q^{n-2} + \cdots + \frac{n(n-1) \cdots (n-l+1)}{1 \cdot 2 \cdots l} p^l q^{n-l}.$$

Similarly, the probability that the number of successes in n trials will not be less than l can be presented thus:

$$P(l, n) = \frac{n(n-1)\cdots(n-l+1)}{1\cdot 2\cdots l}p^l q^{n-l}\left[1 + \frac{n-l}{l+1}\frac{p}{q} + \right.$$
$$\left. + \frac{(n-l)(n-l-1)}{(l+1)(l+2)}\left(\frac{p}{q}\right)^2 + \cdots\right]$$

where the series in the brackets ends by itself.

5. The application of the above established formulas to numerical examples does not present any difficulty so long as the numbers with which we have to deal are not large.

Example 1. In tossing 10 coins, what is the probability of having exactly 5 heads? Tossing 10 different coins at once is the same thing as tossing one coin 10 times, if all the coins are unbiased, which is assumed. Hence, the required probability is given by formula (1), where we must take $n = 10$, $m = 5$, $p = q = \frac{1}{2}$ and it is

$$\frac{10\cdot 9\cdot 8\cdot 7\cdot 6}{1\cdot 2\cdot 3\cdot 4\cdot 5}\cdot\frac{1}{2^{10}} = \frac{252}{1024} = 0.24609.$$

Example 2. If a person playing a certain game can win \$1 with the probability $\frac{1}{3}$, and lose twenty-five cents with the probability $\frac{2}{3}$, what is the probability of winning at least \$3 in 20 games? Let m be the number of times the game is won. The total gain (considering a loss as a negative gain) will be

$$m - \tfrac{1}{4}(20 - m) = \tfrac{5}{4}m - 5 \text{ dollars}$$

and the condition of the problem requires that it should not be less than \$3. Hence

$$\tfrac{5}{4}m - 5 \geqq 3,$$

whence $m \geqq 6\frac{2}{5}$ or, since m is an integer, $m \geqq 7$. That is, in 20 trials an event with the probability $\frac{1}{3}$ must happen at least 7 times and the probability for that is:

$$\sum_{m=7}^{20}\frac{20!}{m!(20-m)!}\left(\frac{1}{3}\right)^m\left(\frac{2}{3}\right)^{20-m}.$$

This sum contains 14 terms; but it can be expressed through another sum containing only 7 terms, because

$$\sum_{m=7}^{20}\frac{20!}{m!(20-m)!}\left(\frac{1}{3}\right)^m\left(\frac{2}{3}\right)^{20-m} = 1 - \sum_{m=0}^{6}\frac{20!}{m!(20-m)!}\left(\frac{1}{3}\right)^m\left(\frac{2}{3}\right)^{20-m}.$$

Using the last expression, one easily gets 0.5207 for the required probability.

***6.** In the series of probabilities

$$T_0,\ T_1,\ T_2,\ \ldots\ T_n$$

for 0, 1, 2, . . . n successes in n trials, the terms generally increase till the greatest term T_μ is reached, and then they steadily decrease. For instance, if $n = 10$, $p = q = \frac{1}{2}$ the values of the expression

$$2^{10}T_m$$

for $m = 0, 1, 2, \ldots 10$ are

1, 10, 45, 120, 210, 252, 210, 120, 45, 10, 1

so that T_5 is the greatest term. For obvious reasons the number μ (to which the greatest term T_μ in the series of probabilities $T_0, T_1, \ldots T_n$ corresponds) is called the "most probable" number of successes.

To prove this observation in general, and to find the rule for obtaining μ, we observe first that the quotient

$$\frac{T_{m+1}}{T_m} = \frac{n - m}{m + 1}\frac{p}{q}$$

decreases with increasing m, so that

$$(a) \qquad \frac{T_1}{T_0} > \frac{T_2}{T_1} > \frac{T_3}{T_2} > \cdots > \frac{T_n}{T_{n-1}}.$$

The two extreme terms in (a) are

$$\frac{T_1}{T_0} = \frac{np}{q}, \qquad \frac{T_n}{T_{n-1}} = \frac{p}{nq},$$

and if n is large enough, the first of them is > 1 and the last < 1. To find exactly how large n must be, we notice that

$$\frac{T_1}{T_0} > 1$$

if

$$np > q = 1 - p$$

whence

$$n + 1 > \frac{1}{p}.$$

Similarly,

$$\frac{T_n}{T_{n-1}} < 1$$

if

$$p < nq \qquad \text{or} \qquad 1 - q < nq$$

whence

$$n + 1 > \frac{1}{q}.$$

Consequently, if $n + 1$ is greater than both $1/p$ and $1/q$, the first term in (a) is >1 and the last term is <1. As the terms of (a) form a decreasing sequence, there must be a last term which is $\geqq 1$. Let it be

$$\frac{T_\mu}{T_{\mu-1}}.$$

Then

$$\frac{T_1}{T_0} > \frac{T_2}{T_1} > \cdots > \frac{T_\mu}{T_{\mu-1}} \geqq 1$$

and

$$1 > \frac{T_{\mu+1}}{T_\mu} > \frac{T_{\mu+2}}{T_{\mu+1}} > \cdots > \frac{T_n}{T_{n-1}}$$

or, which is the same,

$$T_0 < T_1 < T_2 < \cdots < T_{\mu-1} \leqq T_\mu$$
$$T_\mu > T_{\mu+1} > T_{\mu+2} > \cdots > T_n.$$

In other words, the sequence of probabilities increases till the greatest term T_μ is reached and steadily decreases from then on. Besides T_μ, there may be another greatest term $T_{\mu-1}$; namely, when $T_{\mu-1} = T_\mu$; but all the other terms are certainly *less* than T_μ. The number μ is perfectly determined by the conditions

$$\frac{T_\mu}{T_{\mu-1}} = \frac{n-\mu+1}{\mu}\frac{p}{q} \geqq 1, \qquad \frac{T_{\mu+1}}{T_\mu} = \frac{n-\mu}{\mu+1}\frac{p}{q} < 1$$

which are equivalent to the two inequalities

$$(n+1)p \geqq \mu(p+q), \qquad np - q < \mu(p+q).$$

These in turn can be presented thus:

$$\mu \leqq (n+1)p < \mu + 1$$

and show that μ is uniquely determined as the *greatest integer contained in* $(n+1)p$. If $(n+1)p$ is an integer, then $\mu = (n+1)p$ and $T_\mu = T_{\mu-1}$. That is, there are two greatest terms if, and only if, $(n+1)p$ is an integer.

Let us consider now what happens if

$$n + 1 \leqq \frac{1}{p} \qquad \text{or} \qquad n + 1 \leqq \frac{1}{q}.$$

In the first case, all the terms in (*a*) are less than 1 with the single exception of the first term T_1/T_0 which may be equal to 1; namely, when $n + 1 = \frac{1}{p}$. Consequently,

$$T_0 \geqq T_1 > T_2 > \cdots > T_n$$

so that T_0 is the greatest term. If $(n+1)p < 1$ the greatest integer contained in $(n+1)p$ is 0, and there is only one greatest term T_0. If, however, $(n+1)p = 1$, there are two terms $T_0 = T_1$ greater than others.

If $(n+1)q \leqq 1$, all the terms in series (*a*) are >1 with the exception of the last term, which may be equal to 1; namely, when $(n+1)q = 1$. Hence,

$$T_0 < T_1 < \cdots < T_{n-1} \leqq T_n$$

so that T_n is the greatest term, and the preceding term T_{n-1} can be equal to it only if $(n+1)q = 1$. Now the condition

$$(n+1)q \leqq 1$$

is equivalent to

$$(n+1)p \geqq n.$$

On the other hand, because $p < 1$,

$$(n+1)p < n+1.$$

Therefore n is the greatest integer contained in $(n+1)p$.

Comparing the results obtained in the last two cases (excluded at first) with the general rule, we see that in all cases the greatest term T_μ corresponds to

$$\mu = [(n+1)p].$$

If $(n+1)p$ is an integer, then there are two greatest terms T_μ and $T_{\mu-1}$. This rule for determining the most probable number of successes is very simple and easy of application to numerical examples.

Example 1. Let $n = 20$, $p = 2/5$, $q = 3/5$. Then $(n+1)p = 8.4$, and the greatest integer contained in this number is $\mu = 8$. Hence, there is only one most probable number of successes $\mu = 8$ with the corresponding probability

$$T_8 = \frac{20!}{8!12!}\left(\frac{2}{5}\right)^8\left(\frac{3}{5}\right)^{12} = 0.1797.$$

Example 2. Let $n = 110$, $p = 1/3$, $q = 2/3$, and $(n+1)p = 37$, an integer. Consequently, 36 and 37 are the most probable numbers of successes with the corresponding probability

$$T_{36} = T_{37} = \frac{110!}{37!73!}\left(\frac{1}{3}\right)^{37}\left(\frac{2}{3}\right)^{73} = 0.0801.$$

7. When n, m, and $n - m$ are large numbers, the evaluation of probability T_m by the exact formula

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

becomes impracticable and it is necessary to resort to approximations. For approximate evaluation of large factorials we possess precious means in the famous "Stirling formula." Referring the reader to Appendix I where this formula is established, we shall use it here in the following form:

$$\log x! = \log\sqrt{2\pi x} + x\log x - x + \omega(x)$$

where

$$\frac{1}{12(x+\frac{1}{2})} < \omega(x) < \frac{1}{12x}.$$

In the same appendix the following double inequality is proved:

$$\frac{1}{12n} - \frac{1}{12m} - \frac{1}{12l} < \omega(n) - \omega(m) - \omega(l) < \frac{1}{12n+6} - \frac{1}{12m+6} - \frac{1}{12l+6}.$$

Now from Stirling's formula

$$n! = \sqrt{2\pi n}\, n^n e^{-n+\omega(n)}$$

and two similar expressions for $m!$ and $(n-m)!$ follow. Substituting them into T_m, we get two limits

$$T_m < k\sqrt{\frac{n}{2\pi m(n-m)}}\left(\frac{np}{m}\right)^m\left(\frac{nq}{n-m}\right)^{n-m} \tag{3}$$

$$T_m > l\sqrt{\frac{n}{2\pi m(n-m)}}\left(\frac{np}{m}\right)^m\left(\frac{nq}{n-m}\right)^{n-m} \tag{4}$$

where

$$k = e^{\frac{1}{12n+6} - \frac{1}{12m+6} - \frac{1}{12(n-m)+6}}$$

$$l = e^{\frac{1}{12n} - \frac{1}{12m} - \frac{1}{12(n-m)}}.$$

When n, m, $n - m$ are even moderately large k and l differ little from each other.

Inequalities (3) and (4) then give very close upper and lower limits for T_m. To evaluate powers

$$\left(\frac{np}{m}\right)^m, \left(\frac{nq}{n-m}\right)^{n-m}$$

with large exponents, sufficiently extensive logarithmic tables must be available. If such tables are lacking, then in cases which ordinarily occur when ratios np/m and $nq/(n-m)$ are close to 1, we can use special short tables to evaluate logarithms of these ratios or else resort to series.

8. Another problem requiring the probability that the number of successes will be contained between two given limits is much more complex in case the number of trials as well as the difference between given limits is a large number. Ordinarily for approximate evaluation of probability under such circumstances simple and convenient formulas are used. These formulas are derived in Chap. VII. Less known is the ingenious use by Markoff of continued fractions for that purpose.

It suffices to devise a method for approximate evaluation of the probability that the number of successes will be greater than a given integer l which can be supposed $> np$. We shall denote this probability by

$P(l)$. A similar notation $Q(l)$ will be used to denote the probability that the number of failures is $>l$ where again $l > nq$. The probability $P(k, l)$ of the inequalities $k \leqq m \leqq l$ can be expressed as follows:

$$P(k, l) = 1 - P(l) - Q(n - k)$$

if $l > np$ and $k < np$;

$$P(k, l) = P(k - 1) - P(l)$$

if both k and l are $> np$; and finally

$$P(k, l) = Q(n - l - 1) - Q(n - k)$$

if both k and l are $< np$.

For $P(l)$ we have the expression

$$P(l) = \frac{n!}{(l+1)!(n-l-1)!}p^{l+1}q^{n-l-1}\left[1 + \frac{n-l-1}{l+2}\frac{p}{q} + \right. \\ \left. + \frac{(n-l-1)(n-l-2)}{(l+2)(l+3)}\left(\frac{p}{q}\right)^2 + \cdot \quad \right].$$

The first factor

$$\frac{n!}{(l+1)!(n-l-1)!}p^{l+1}q^{n-l-1}$$

can be approximately evaluated by the method of the preceding section. The whole difficulty resides in the evaluation of the sum

$$S = 1 + \frac{n-l-1}{l+2}\frac{p}{q} + \frac{(n-l-1)(n-l-2)}{(l+2)(l+3)}\left(\frac{p}{q}\right)^2 + \cdots$$

which is a particular case of the hypergeometric series

$$F(\alpha, \beta, \gamma, x) = 1 + \frac{\alpha\beta}{1 \cdot \gamma}x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{1 \cdot 2\gamma(\gamma+1)}x^2 + \cdots.$$

In fact

$$F\left(-n + l + 1, 1, l + 2, -\frac{p}{q}\right) = S.$$

Now, owing to this connection between S and hypergeometric series, S can be represented in the form of a continued fraction. First, it is easy to establish the following relations:

$$F(\alpha, \beta + 1, \gamma + 1, x) = F(\alpha, \beta, \gamma, x) + \\ + x\frac{\alpha(\gamma - \beta)}{\gamma(\gamma + 1)}F(\alpha + 1, \beta + 1, \gamma + 2, x)$$

$$F(\alpha + 1, \beta, \gamma + 1, x) = F(\alpha, \beta, \gamma, x) + \\ + x\frac{\beta(\gamma - \alpha)}{\gamma(\gamma + 1)}F(\alpha + 1, \beta + 1, \gamma + 2, x).$$

Substituting $\alpha + n, \beta + n, \gamma + 2n$ and $\alpha + n, \beta + n + 1, \gamma + 2n + 1$, respectively, for α, β, γ in these relations and setting

$$X_{2n} = F(\alpha + n, \beta + n, \gamma + 2n, x);$$
$$X_{2n+1} = F(\alpha + n, \beta + n + 1, \gamma + 2n + 1, x)$$
$$a_{2n} = \frac{(\beta + n)(\gamma - \alpha + n)}{(\gamma + 2n)(\gamma + 2n - 1)}; \qquad a_{2n+1} = \frac{(\alpha + n)(\gamma - \beta + n)}{(\gamma + 2n)(\gamma + 2n + 1)}$$

for brevity, we have

$$X_0 = X_1 - a_1 x X_2$$
$$X_1 = X_2 - a_2 x X_3$$
$$\cdots\cdots$$
$$X_{m-1} = X_m - a_m x X_{m+1}$$

whence

$$\frac{X_1}{X_0} = \cfrac{1}{1 - \cfrac{a_1 x}{1} - \cfrac{a_2 x}{1} - \cdots - \cfrac{a_{m-1}x}{1} - \cfrac{a_m x}{\cfrac{X_m}{X_{m+1}}}}.$$

In our particular case

$$X_1 = F(-n + l + 1, 1, l + 2, x), \qquad X_0 = 1$$

and $a_{2n-2l-1} = 0$.

Hence, taking $x = -\frac{p}{q}$ and introducing new notations, we have a finite continued fraction

$$S = \cfrac{1}{1 - \cfrac{c_1}{1} + \cfrac{d_1}{1} - \cfrac{c_2}{1} + \cdots - \cfrac{c_{n-l-1}}{1 + \cfrac{d_{n-l-1}}{1}}} \tag{5}$$

where

$$c_k = \frac{(n - k - l)(l + k)p}{(l + 2k - 1)(l + 2k)q}; \qquad d_k = \frac{k(n + k)p}{(l + 2k)(l + 2k + 1)q}. \tag{6}$$

Every one of the numbers c_k will be positive and <1 if this is true for c_1. Now

$$c_1 = \frac{(n - l - 1)p}{(l + 2)q} < 1$$

if $l > np$, and that is exactly what we suppose. The above continued fraction can be used to obtain approximate values of S in excess or in defect, as we please. Let us denote the continued fraction

$$\frac{c_k}{1} + \frac{d_k}{1} - \frac{c_{k+1}}{1} + \cdots$$

by ω_k. Then

$$0 < \omega_k < c_k,$$

which can be easily verified. Furthermore,

$$S = \frac{1}{1 - \omega_1}; \qquad \omega_1 = \frac{c_1}{1} + \frac{d_1}{1 - \omega_2}; \qquad \omega_2 = \frac{c_2}{1} + \frac{d_2}{1 - \omega_3}$$

and in general

$$\omega_k = \frac{c_k}{1} + \frac{d_k}{1 - \omega_{k+1}}.$$

Having selected k, depending on the degree of approximation we desire in the final result (but never too large; $k = 5$ or less generally suffices), we use the inequality

$$0 < \omega_{k+1} < c_{k+1}$$

to obtain two limits in defect and in excess for ω_k. Using these limits, we obtain similar limits for $\omega_{k-1}, \omega_{k-2}, \omega_{k-3}, \ldots$ and, finally, for ω_1 and S.

The series of operations will be better illustrated by an example.

9. Let us find approximately the probability that in 9,000 trials an event with the probability $p = 1/3$ will occur not more than 3,090 times and not less than 2,910 times. To this end we must first seek the probability of more than 3,090 occurrences, which involves, in the first place, the evaluation of

$$T_{3091} = \frac{9000!}{3091!\,5909!}\left(\frac{1}{3}\right)^{3091}\left(\frac{2}{3}\right)^{5909}.$$

By using inequalities (3) and (4) of Sec. 7, we find

$$0.011286 < T_{3091} < 0.011287.$$

Next we turn to the continued fraction to evaluate the sum S. The following table gives approximate values of $c_1, c_2, \ldots c_6$ and $d_1, d_2 \ldots d_5$ to 5 decimals and less than the exact numbers

n	c_n	d_n
1	0.95553	0.00047
2	0.95444	0.00094
3	0.95335	0.00140
4	0.95227	0.00187
5	0.95119	0.00234
6	0.95010	

We start with the inequalities

$$0 < \omega_6 < 0.95011$$

and then proceed as follows:

$$1.00234 < 1 + \frac{d_5}{1-\omega_6} < 1.04711; \qquad 0.90839 < \omega_5 < 0.94898$$

$$1.02041 < 1 + \frac{d_4}{1-\omega_5} < 1.03685; \qquad 0.91842 < \omega_4 < 0.93324$$

$$1.01716 < 1 + \frac{d_3}{1-\omega_4} < 1.02113; \qquad 0.93362 < \omega_3 < 0.93728$$

$$1.01416 < 1 + \frac{d_2}{1-\omega_3} < 1.01514; \qquad 0.94020 < \omega_2 < 0.94113$$

$$1.00785 < 1 + \frac{d_1}{1-\omega_2} < 1.00816; \qquad 0.94779 < \omega_1 < 0.94810$$

$$\frac{1}{0.05221} < S < \frac{1}{0.05190}$$

$$0.02161 < ST_{3091} < 0.02175.$$

Hence, we know for certain that

$$0.02161 < P(3{,}090) < 0.02175.$$

By a similar calculation it was found that

$$0.02129 < Q(6{,}090) < 0.02142,$$

so that

$$0.04290 < P(3{,}090) + Q(6{,}090) < 0.04317.$$

The required probability P that the number of successes will be contained between 2,910 and 3,090 (limits included) lies between 0.95683 and 0.95710 so that, taking $P = 0.9570$, the error in absolute value will be less than 1.7×10^{-4}.

Problems for Solution

1. What is the probability of having 12 three times in 100 tosses of 2 dice?

Ans. $C_{100}^{3}(\frac{1}{36})^3(\frac{35}{36})^{97} = 0.2257.$

2. What is the probability for an event E to occur at least once, or twice, or three times, in a series of n independent trials with the probability p? *Ans.*
(a) $1 - (1 - p)^n$; (b) $1 - (1 - p)^{n-1}[1 + (n - 1)p]$;

$$(c)\quad 1 - (1 - p)^{n-2}\left[1 + (n - 2)p + \frac{(n - 1)(n - 2)}{2}p^2\right].$$

3. What is the probability of having 12 points with 2 dice at least three times in 100 throws? *Ans.* 0.528.

4. In a series of 100 independent trials with the probability $\frac{1}{3}$, what is the most probable number of successes and its probability? *Ans.* $\mu = 33$; $T_{33} = 0.0844$.

NOTE: Log 100! = 157.97000; Log 67! = 94.56195; Log 33! = 36.93869.

5. A player wins \$1 if he throws heads two times in succession; otherwise he loses 25 cents. If this game is repeated 100 times, what is the probability that neither his gain nor loss will exceed \$1? Or \$5? *Ans.*

$$(a)\quad \frac{100!}{20!\,80!}\left(\frac{1}{4}\right)^{20}\left(\frac{3}{4}\right)^{80} = 0.0493;$$

$$(b)\quad \frac{100!}{20!\,80!}\left(\frac{1}{4}\right)^{20}\left(\frac{3}{4}\right)^{80}\left[1 + \frac{80}{63} + \frac{80\cdot 79}{63\cdot 66} + \frac{80\cdot 79\cdot 78}{63\cdot 66\cdot 69} + \frac{80\cdot 79\cdot 78\cdot 77}{63\cdot 66\cdot 69\cdot 72} + \right.$$
$$\left. + \frac{60}{81} + \frac{60\cdot 57}{81\cdot 82} + \frac{60\cdot 57\cdot 54}{81\cdot 82\cdot 83} + \frac{60\cdot 57\cdot 54\cdot 51}{81\cdot 82\cdot 83\cdot 84}\right] = 0.4506.$$

NOTE: Log 20! = 18.38612; Log 80! = 118.85473.

6. Show that in a series of $2s$ trials with the probability $\frac{1}{2}$ the most probable number of successes is s and the corresponding probability

$$T_s = \frac{1\cdot 3\cdot 5\cdots(2s - 1)}{2\cdot 4\cdot 6\cdots 2s}.$$

Show also that

$$T_s < \frac{1}{\sqrt{2s + 1}}.$$

HINT:

$$T_s < \frac{2\cdot 4\cdot 6\cdots 2s}{3\cdot 5\cdot 7\cdots(2s + 1)}.$$

7. Prove the following theorem: If P and P' are probabilities of the most probable number of successes, respectively, in n and $n + 1$ trials, then $P' \leqq P$, the equality sign being excluded unless $(n + 1)p$ is an integer.

8. Show that the probability T_μ corresponding to the most probable number of successes in n trials, is asymptotic to $(2\pi npq)^{-\frac{1}{2}}$, that is,

$$\lim T_\mu\sqrt{2\pi npq} = 1 \quad \text{as} \quad n \to \infty.$$

9. When $p = \frac{1}{2}$, the following inequality holds for every m:

$$T_m < \sqrt{\frac{2}{\pi n}}\,\frac{e^{-t^2}}{\sqrt{1 - \frac{2t^2}{n}}}.$$

if

$$m = \frac{n}{2} + t\sqrt{\frac{n}{2}}.$$

10. What is the probability of 215 successes in 1,000 trials if $p = \frac{1}{5}$?

Ans. 0.0154.

11. What is the probability that in 2,000 trials the number of successes will be contained between 460 and 540 (limits included) if $p = \frac{1}{4}$. *Ans.* 0.964.

12. Two players A and B agree to play until one of them wins a certain number of games, the probabilities for A and B to win a single game being p and $q = 1 - p$. However, they are forced to quit when A has a games still to win, and B has b games. How should they divide their total stake to be fair?

This problem is known as "problème de parties," one of the first problems on probability discussed and solved by Fermat and Pascal in their correspondence.

Solution 1. Let P denote the probability that A will win a remaining games before B can win b games, and let $Q = 1 - P$ denote the probability for B to win b games before A wins a games. To be fair, the players must divide their common stake M in the ratio $P:Q$ and leave the sum MP to A and the sum MQ to B.

To find P, notice that A wins in the following mutually exclusive ways:

a. If he wins in exactly a games; probability p^a.

b. If he wins in exactly $a + 1$ games; probability $\frac{a}{1}p^aq$.

c. If he wins in exactly $a + 2$ games; probability $\frac{a(a+1)}{1 \cdot 2}p^aq^2$.

. . . .

n. If he wins in exactly $a + b - 1$ games; probability

$$\frac{a(a+1)\cdots(a+b-2)}{1 \cdot 2 \cdot 3 \cdots (b-1)}p^aq^{b-1}.$$

Consequently

$$P = p^a\left[1 + \frac{a}{1}q + \frac{a(a+1)}{1 \cdot 2}q^2 + \cdots + \frac{a(a+1)\cdots(a+b-2)}{1 \cdot 2 \cdots (b-1)}q^{b-1}\right]$$

and similarly

$$Q = q^b\left[1 + \frac{b}{1}p + \frac{b(b+1)}{1 \cdot 2}p^2 + \cdots + \frac{b(b+1)\cdots(b+a-2)}{1 \cdot 2 \cdots (a-1)}p^{a-1}\right].$$

Show directly that $P + Q = 1$.

HINT: $\frac{dP}{dp} + \frac{dQ}{dp} = 0.$

Solution 2. The same problem can be solved in a different way. Whether A or B wins will be decided in not more than $a + b - 1$ games. Now if the players continue to play until the number of games reaches the limit $a + b - 1$, the number of games won by A must be not less than a. And conversely, if this number is not less than a, A will win a games before B wins b games. Therefore, P is the probability that in $a + b - 1$ games A wins not less than a times, or

$$P = \frac{(a+b-1)!}{a!(b-1)!}p^aq^{b-1}\left[1 + \frac{b-1}{a+1}\frac{p}{q} + \frac{(b-1)(b-2)}{(a+1)(a+2)}\left(\frac{p}{q}\right)^2 + \cdots + \frac{(b-1)(b-2)\cdots 2 \cdot 1}{(a+1)(a+2)\cdots(a+b-1)}\left(\frac{p}{q}\right)^{b-1}\right].$$

Show directly that both expressions for P are identical.

HINT: Proceed as before.

13. Prove the identity

$$p^n + \frac{n}{1}p^{n-1}q + \frac{n(n-1)}{1\cdot 2}p^{n-2}q^2 + \cdots + \frac{n(n-1)\cdots(n-k+1)}{1\cdot 2\cdot 3\cdots k}p^{n-k}q^k = \frac{\int_0^p x^{n-k-1}(1-x)^k dx}{\int_0^1 x^{n-k-1}(1-x)^k dx}$$

HINT: Take derivatives with respect to p.

14. A and B have, respectively, $n+1$ and n coins. If they toss their coins simultaneously, what is the probability that (a) A will have more heads than B? (b) A and B will have an equal number of heads? (c) B will have more heads than A?

Solution. *a.* Let P_n be the probability for A to have more heads than B. This probability can be expressed as the double sum

$$P_n = \frac{1}{2^{2n+1}}\sum_{x=1}^{n+1}\sum_{\alpha=0}^{n} C_{n+1}^{\alpha+x}C_n^{\alpha}.$$

Considering the coefficient of t^x in

$$(1+t)^{n+1}\left(1+\frac{1}{t}\right)^n = \frac{(1+t)^{2n+1}}{t^n},$$

we have

$$\sum_{\alpha=0}^{n} C_{n+1}^{\alpha+x}C_n^{\alpha} = C_{2n+1}^{n+x}.$$

Hence

$$P_n = \frac{1}{2^{2n+1}}\sum_{x=1}^{n+1} C_{2n+1}^{n+x} = \frac{2^{2n}}{2^{2n+1}} = \frac{1}{2}.$$

b. The probability Q_n for A and B to have an equal number of heads is

$$Q_n = \frac{1}{2^{2n+1}}\sum_{\alpha=0}^{n} C_{n+1}^{\alpha}C_n^{\alpha} = \frac{C_{2n+1}^{n}}{2^{2n+1}} < \frac{1}{\sqrt{\pi n}}.$$

c. The probability R_n for B to have more heads than A is

$$R_n = \frac{1}{2} - \frac{C_{2n+1}^{n}}{2^{2n+1}}.$$

15. If each of n independent trials can result in one of the m incompatible events $E_1, E_2, \ldots E_m$ with the respective probabilities

$$p_1, p_2, \ldots p_m; \qquad (p_1 + p_2 + \cdots + p_m = 1),$$

show that the probability to have l_1 events E_1, l_2 events E_2, ... l_m events E_m where $l_1 + l_2 + \cdots + l_m = n$, is given by

$$P_{l_1, l_2, \ldots l_m} = \frac{n!}{l_1!\,l_2!\cdots l_m!}p_1^{l_1}p_2^{l_2}\cdots p_m^{l_m}.$$

CHAPTER IV

PROBABILITIES OF HYPOTHESES AND BAYES' THEOREM

1. The nature of the problems with which we deal in this chapter may be illustrated by the following simple example: Urns 1 and 2 contain, respectively, 2 white and 3 black balls, and 4 white and 1 black balls. One of the urns is selected at random and one ball is drawn. It happens to be white. What is the probability that it came from the first urn? Before the ball was drawn and its color revealed, the probability that the first urn would be chosen had been 1/2; but the indication of the color of the ball that was drawn altered this probability. To find this new probability, the following artifice can be used:

Imagine that balls from both urns are put together in a third urn. To distinguish their origin, balls from the first urn are marked with 1 and those from the second urn are marked with 2. Since there are 5 balls marked with 1 and the same number marked with 2, in taking one ball from the third urn we have equal chances to take one coming from either the first or the second urn, and the situation is exactly the same as if we chose one of the urns at random and drew one ball from it. If the ball drawn from the third urn happens to be white, this can happen in $2 + 4 = 6$ equally likely cases. Only in 2 of these cases will the extracted ball have the mark 1. Hence, the probability that the white ball came from the first urn is $2/6 = 1/3$.

The success of this artifice depends on the equality of the number of balls in both urns. It can be applied to the case of an unequal number of balls in the urns, but with some modifications; however, it seems preferable to follow a regular method for solving problems like the preceding one.

2. The problem just solved is a particular case of the following fundamental:

Problem 1. An event A can occur only if one of the set of exhaustive and incompatible events

$$B_1,\ B_2,\ \ldots\ B_n$$

occurs. The probabilities of these events

$$(B_1),\ (B_2),\ \ldots\ (B_n)$$

corresponding to the total absence of any knowledge as to the occurrence or nonoccurrence of A, are known. Known also, are the conditional probabilities

$$(A, B_i); \qquad i = 1, 2, \ldots n$$

for A to occur, assuming the occurrence of B_i. How does the probability of B_i change with the additional information that A has actually happened?

Solution. The question amounts to finding the conditional probability (B_i, A). The probability of the compound event AB_i can be presented in two forms

$$(AB_i) = (B_i)(A, B_i)$$

or

$$(AB_i) = (A)(B_i, A).$$

Equating the right-hand members, we derive the following expression for the unknown probability (B_i, A):

$$(B_i, A) = \frac{(B_i)(A, B_i)}{(A)}.$$

Since the event A can materialize in the mutually exclusive forms

$$AB_1, AB_2, \ldots AB_n,$$

by applying the theorem of total probability, we get

$$(A) = (B_1)(A, B_1) + (B_2)(A, B_2) + \cdots + (B_n)(A, B_n).$$

It suffices now to introduce this expression into the preceding formula for (B_i, A) to get the final expression

$$(1) \quad (B_i, A) = \frac{(B_i)(A, B_i)}{(B_1)(A, B_1) + (B_2)(A, B_2) + \cdots + (B_n)(A, B_n)}.$$

This formula, when described in words, constitutes the so-called "Bayes' theorem." However, it is hardly necessary to describe its content in words; symbols speak better for themselves. For that reason, we prefer to speak of *Bayes' formula* rather than of Bayes' theorem. Bayes' formula is also known as the "formula for probabilities of hypotheses." The reason for that name is that the events $B_1, B_2, \ldots B_n$ may be considered as hypotheses to account for the occurrence of A.

It is customary to speak of probabilities

$$(B_1), (B_2), \ldots (B_n)$$

as a priori probabilities of hypotheses

$$B_1, B_2, \ldots B_n,$$

while probabilities

$$(B_i, A); \qquad i = 1, 2, \ldots n$$

are called a posteriori probabilities of the same hypotheses.

3. A few examples will help us to understand the meaning and the use of Bayes' formula.

Example 1. The contents of urns 1, 2, 3, are as follows:

1 white, 2 black, 3 red balls
2 white, 1 black, 1 red balls
4 white, 5 black, 3 red balls

One urn is chosen at random and two balls drawn. They happen to be white and red. What is the probability that they came from urn 2 or 3?

Solution. The event A represents the fact that two balls taken from the selected urn were of white and red color, respectively. To account for this fact, we have three hypotheses: The selected urn was 1 or 2 or 3. We shall represent these hypotheses in the order indicated by B_1, B_2, B_3. Since nothing distinguishes the urns, the probabilities of these hypotheses before anything was known about A are

$$(B_1) = (B_2) = (B_3) = \tfrac{1}{3}.$$

The probabilities of A, assuming these hypotheses, are

$$(A, B_1) = \tfrac{1}{5}, \qquad (A, B_2) = \tfrac{1}{3}, \qquad (A, B_3) = \tfrac{2}{11}.$$

It remains now to introduce these values into formula (1) to have a posteriori probabilities

$$(B_2, A) = \frac{\frac{1}{3}\cdot\frac{1}{3}}{\frac{1}{3}\cdot\frac{1}{5}+\frac{1}{3}\cdot\frac{1}{3}+\frac{1}{3}\cdot\frac{2}{11}} = \frac{55}{118}$$
$$(B_3, A) = \frac{\frac{1}{3}\cdot\frac{2}{11}}{\frac{1}{3}\cdot\frac{1}{5}+\frac{1}{3}\cdot\frac{1}{3}+\frac{1}{3}\cdot\frac{2}{11}} = \frac{30}{118}$$

and also, naturally,

$$(B_1, A) = 1 - (B_2, A) - (B_3, A) = \tfrac{33}{118}.$$

Example 2. It is known that an urn containing altogether 10 balls was filled in the following manner: A coin was tossed 10 times, and according as it showed heads or tails, one white or one black ball was put into the urn. Balls are drawn from this urn one at a time, 10 times in succession (always being returned before the next drawing) and every one turns out to be white. What is the probability that the urn contains nothing but white balls?

Solution. The event A consists in the fact that in 10 independent trials with a definite but unknown probability, only white balls appear. To account for this fact, we have 10 hypotheses regarding the number of white balls in the urn; namely, that this number is either 1, or 2, or 3, . . . or 10. The a priori probability of the hypothesis B_i that there are exactly i white balls in the urn, according to the manner in which the urn was filled, is the same as the probability of having i heads in 10 throws; that is,

$$(B_i) = \frac{10!}{i!(10-i)!}\frac{1}{2^{10}}; \qquad i = 1, 2, \ldots 10.$$

Granted the hypothesis B_i, the probability of A is

$$(A, B_i) \quad \left(\frac{i}{10}\right)^{10}.$$

The problem requires us to find (B_{10}, A). The expression of this probability immediately results from Bayes' formula:

$$(B_{10}, A) = \frac{1}{\sum_{i=1}^{10} C_{10}^i \left(\frac{i}{10}\right)^{10}}.$$

The denominator of this fraction is

$$14.247.$$

Hence

$$(B_{10}, A) = 0.0702.$$

This probability, although still small, is much greater than $1/1024$, the a priori probability of having only white balls in the urn.

If, instead of 10 drawings, m drawings have been made and at each drawing white balls appeared, the probability (B_{10}, A) would be given by

$$(B_{10}, A) = \frac{1}{\sum_{i=1}^{10} C_{10}^i \left(\frac{i}{10}\right)^{m}}.$$

The denominator of this formula can be presented thus:

$$\sum_{i=0}^{10} C_{10}^i \left(1 - \frac{i}{10}\right)^{m}.$$

Now

$$\left(1 - \frac{i}{10}\right)^{m} < e^{-\frac{mi}{10}}$$

and so

$$\sum_{i=0}^{10} C_{10}^i \left(1 - \frac{i}{10}\right)^{m} < \sum_{i=0}^{10} C_{10}^i e^{-\frac{mi}{10}} = \left(1 + e^{-\frac{m}{10}}\right)^{10}.$$

Hence

$$(B_{10}, A) > \left(1 + e^{-\frac{m}{10}}\right)^{-10}.$$

This shows that with increasing m the probability (B_{10}, A) rapidly approaches 1. For instance, if $m = 100$

$$(B_{10}, A) > (1 + e^{-10})^{-10} > (1.0000454)^{-10} > 0.99954.$$

Thus, after 100 drawings producing only white balls, it is almost certain that the urn contains nothing but white balls—a conclusion which mere common sense would dictate.

Example 3. Two urns, 1 and 2, contain respectively 2 white and 1 black ball, and 1 white and 5 black balls. One ball is transferred from urn 1 to urn 2 and then one ball is drawn from the latter. It happens to be white. What is the probability that the transferred ball was black?

Solution. Here we have two hypotheses: B_1, that the transferred ball was black, and B_2, that it was white. The a priori probabilities of these hypotheses are

$$(B_1) = \tfrac{1}{3}, \qquad (B_2) = \tfrac{2}{3}.$$

The probabilities of drawing a white ball from urn 2, granted that B_1 or B_2 is true, are:

$$(A, B_1) = \tfrac{1}{7}, \qquad (A, B_2) = \tfrac{2}{7}.$$

The probability of B_1, after a white ball has been drawn from the second urn, results from Bayes' formula:

$$(B_1, A) = \frac{\frac{1}{3} \cdot \frac{1}{7}}{\frac{1}{3} \cdot \frac{1}{7} + \frac{2}{3} \cdot \frac{2}{7}} = \frac{1}{5}.$$

4. Problem 2. Retaining the notations, conditions, and data of Prob. 1, find the probability of materialization of another event C granted that A has actually occurred. Conditional probabilities

$$(C, AB_i); \qquad i = 1, 2, \ldots n$$

are supposed to be known.

Solution. Since the fact of the occurrence of A involves that of one, and only one, of the events

$$B_1, B_2, \ldots B_n,$$

the event C (granted the occurrence of A) can materialize in the following mutually exclusive forms

$$CB_1, CB_2, \ldots CB_n.$$

Consequently, the probability (C, A) which we are seeking is given by

$$(C, A) = (CB_1, A) + (CB_2, A) + \cdots + (CB_n, A).$$

Applying the theorem of compound probability, we have

$$(CB_i, A) = (B_i, A)(C, B_iA)$$

and

$$(C, A) = (B_1, A)(C, AB_1) + (B_2, A)(C, AB_2) + \cdots + (B_n, A)(C, AB_n).$$

It suffices now to substitute for

$$(B_i, A)$$

its expression given by Bayes' formula, to find the final expression

$$(2) \qquad (C, A) = \frac{\sum_{i=1}^{n} (B_i)(A, B_i)(C, AB_i)}{\sum_{i=1}^{n} (B_i)(A, B_i)}.$$

It may happen that the materialization of hypothesis B_i makes C independent of A; then we have simply

$$(C, AB_i) = (C, B_i)$$

and instead of formula (2), we have a simplified formula

$$(3) \qquad (C, A) = \frac{\sum_{i=1}^{n} (B_i)(A, B_i)(C, B_i)}{\sum_{i=1}^{n} (B_i)(A, B_i)}.$$

The event C can be considered in regard to A as a *future* event. For that reason formulas (2) and (3) express probabilities of future events. For better understanding of these commonly used technical terms, we shall consider a simple example.

Example 4. From an urn containing 3 white and 5 black balls, 4 balls are transferred into an empty urn. From this urn 2 balls are taken and they both happen to be white. What is the probability that the third ball taken from the same urn, will be white?

Solution. (*a*) Let us suppose that the two balls drawn in the first place are returned to the second urn. Analyzing this problem, we distinguish first the following hypotheses concerning colors of the 4 balls transferred from the first urn. Among them, there are necessarily 2 white balls. Hence, there are only two possible hypotheses:

B_1: 2 white and 2 black balls;
B_2: 3 white and 1 black ball.

A priori probabilities of these hypotheses are

$$(B_1) = \frac{C_3^2 \cdot C_5^2}{C_8^4} = \frac{3}{7}.$$

$$(B_2) = \frac{C_3^3 \cdot C_5^1}{C_8^4} = \frac{1}{14}.$$

The event A consists in the white color of both balls drawn from the second urn The conditional probabilities (A, B_1) and (A, B_2) are

$$(A, B_1) = \tfrac{1}{6}; \quad (A, B_2) = \tfrac{3}{6}.$$

The future event C consists in the white color of the third ball. Since the 2 balls drawn at first are returned, C becomes independent of A as soon as it is known which one of the hypotheses has materialized. Hence

$$(C, AB_1) = (C, B_1) = \tfrac{1}{2}$$
$$(C, AB_2) = (C, B_2) = \tfrac{3}{4}.$$

Substituting these various numbers in formula (3), we find that

$$(C, A) = \frac{\frac{3}{7} \cdot \frac{1}{6} \cdot \frac{1}{2} + \frac{1}{14} \cdot \frac{3}{6} \cdot \frac{3}{4}}{\frac{3}{7} \cdot \frac{1}{6} + \frac{1}{14} \cdot \frac{3}{6}} = \frac{7}{12}.$$

(*b*) If the two balls drawn in the first place are not returned, we have

$$(C, AB_1) = 0, \qquad (C, AB_2) = \frac{1}{2}.$$

Then, making use of formula (2),

$$(C, A) = \frac{\frac{1}{14} \cdot \frac{3}{8} \cdot \frac{1}{2}}{\frac{3}{7} \cdot \frac{1}{8} + \frac{1}{14} \cdot \frac{3}{8}} = \frac{1}{6}.$$

5. The following problem can easily be solved by direct application of Bayes' formula.

Problem 3. A series of trials is performed, which, with certain additional data, would appear as independent trials in regard to an event E with a constant probability p.

Lacking these data, all we *know* is that the unknown probability p must be one of the numbers

$$p_1, p_2, \ldots p_k$$

and we can assume these values with the respective probabilities

$$\alpha_1, \alpha_2, \ldots \alpha_k.$$

In n trials the event E actually occurred m times. What is the probability that p lies between the two given limits α and β $(0 \leqq \alpha < \beta \leqq 1)$, or else, what is the probability of the following inequalities:

$$\alpha \leqq p \leqq \beta?$$

A particular case may illustrate the meaning of this problem. In a set of N urns, $N\alpha_1$ urns have white balls in proportion p_1 to the total number of balls; $N\alpha_2$ urns have white balls in proportion p_2; . . . $N\alpha_k$ urns have white balls in proportion p_k. An urn is chosen at random and n drawings of one ball at a time are performed, the ball being returned each time before the next drawing so as to keep a constant proportion of white balls. It is found that altogether m white balls have appeared. What is the probability that one of the $N\alpha_i$ urns with the proportion p_i of white balls was chosen? Evidently this is a particular case of the general problem, and here we possess knowledge of the necessary data, provided that the probability of selecting any one of the urns is the same.

Solution. We distinguish k exhaustive and mutually exclusive hypotheses that the unknown probability is p_1, or p_2, . . . or p_k. The a priori probabilities of these hypotheses are, respectively, $\alpha_1, \alpha_2, \ldots \alpha_k$. Assuming the hypothesis $p = p_i$, the probability of the event E occurring m times in n trials is

$$C_n^m p_i^m (1 - p_i)^{n-m}.$$

Now, after E has actually happened m times in n trials, the a posteriori probability of the hypothesis $p = p_i$, by virtue of Bayes' formula, will be

$$\frac{C_n^m \alpha_i p_i^m (1 - p_i)^{n-m}}{\sum_{i=1}^{k} C_n^m \alpha_i p_i^m (1 - p_i)^{n-m}}$$

or, canceling C_n^m,

$$\frac{\alpha_i p_i^m (1 - p_i)^{n-m}}{\sum_{i=1}^{k} \alpha_i p_i^m (1 - p_i)^{n-m}}.$$

Now, applying the theorem of total probability, the probability P of the inequalities

$$\alpha \leqq p \leqq \beta$$

will be given by

(4) $$P = \frac{\Sigma \alpha_i p_i^m (1 - p_i)^{n-m}}{\sum_{i=1}^{k} \alpha_i p_i^m (1 - p_i)^{n-m}}$$

where the summation in the numerator refers to all values of p_i lying between α and β, limits included.

An important particular case arises when the set of hypothetical probabilities is

$$p_1 = \frac{1}{k}, \quad p_2 = \frac{2}{k}, \quad \cdots \quad p_k = 1$$

and the a priori probabilities of these hypotheses are equal:

$$\alpha_1 = \alpha_2 = \cdots = \alpha_k = \frac{1}{k}.$$

Then the fraction $1/k$ can be canceled in both numerator and denominator. The final formula for the probability of the inequalities

$$\alpha \leqq p \leqq \beta$$

will be

(5) $$P = \frac{\Sigma p_i^m (1 - p_i)^{n-m}}{\sum_{i=1}^{k} p_i^m (1 - p_i)^{n-m}}$$

summation in numerator being extended over all positive integers i satisfying the inequalities

$$k\alpha \leqq i \leqq k\beta.$$

In the limit, when k tends to infinity, the a priori probability of the inequalities

$$\alpha \leqq p \leqq \beta$$

is given simply by the length $\beta - \alpha$ of the interval (α, β). The a posteriori probability of the same inequalities is obtained as the limit of expression (5). Now, as $k \to \infty$, the sums

$$\sum_{i \geqq k\alpha}^{i \leqq k\beta} \left(\frac{i}{k}\right)^m \left(1 - \frac{i}{k}\right)^{n-m} \frac{1}{k} \quad \text{and} \quad \sum_{i=1}^{k} \left(\frac{i}{k}\right)^m \left(1 - \frac{i}{k}\right)^{n-m} \frac{1}{k}$$

tend to the definite integrals

$$\int_\alpha^\beta x^m(1 - x)^{n-m}dx \quad \text{and} \quad \int_0^1 x^m(1 - x)^{n-m}dx.$$

Therefore, in the limit, the a posteriori probability of the inequalities

$$\alpha \leqq p \leqq \beta$$

is expressed by the ratio of two definite integrals

$$P = \frac{\int_\alpha^\beta x^m(1 - x)^{n-m}dx}{\int_0^1 x^m(1 - x)^{n-m}dx}. \tag{6}$$

This formula leads to the following conclusion: *When the unknown probability p of an event E may have any value between* 0 *and* 1 *and the a priori probability of its being contained between limits α and β is $\beta - \alpha$, then after n trials in which E occurred m times, the a posteriori probability of p being contained between α and β is given by formula* (6).

6. Problem 4. Assumptions and data being the same as in Prob. 3, find the probability that in n_1 trials, following n trials, which produced E m times, the same event will occur m_1 times.

Solution. It suffices to take in formula (3)

$$(B_i) = \alpha_i; \qquad (A, B_i) = C_n^m p_i^m(1 - p_i)^{n-m}$$

and

$$(C, B_i) = C_{n_1}^{m_1} p_i^{m_1}(1 - p_i)^{n_1-m_1}$$

to find for the required probability this expression:

$$Q = C_{n_1}^{m_1} \frac{\sum_{i=1}^{k} \alpha_i p_i{}^{m+m_1}(1 - p_i)^{n+n_1-m-m_1}}{\sum_{i=1}^{k} \alpha_i p_i^m(1 - p_i)^{n-m}}. \tag{7}$$

Supposing again

$$p_1 = \frac{1}{k}, \qquad p_2 = \frac{2}{k}, \quad \cdots \quad p_k = 1$$

$$\alpha_1 = \alpha_2 = \quad \cdots \quad = \alpha_k = \frac{1}{k}$$

and letting $k \to \infty$ formula (7) in the limit becomes

$$Q = C_{n_1}^{m_1} \frac{\int_0^1 x^{m+m_1}(1 - x)^{n+n_1-m-m_1}dx}{\int_0^1 x^m(1 - x)^{n-m}dx}. \tag{8}$$

This formula leads to the following conclusion: *When the unknown probability p of an event E may have any value between limits* 0 *and* 1 *and the a priori probability of its being contained between* α *and* β *is* $\beta - \alpha$ (*so that equal probabilities correspond to intervals of equal length*), *the probability that the event E will happen* m_1 *times in* n_1 *trials following n trials which produced E m times is given by formula* (8).

In particular, for $n_1 = m_1 = 1$ (evaluating integrals by the known formula), we have

$$Q = \frac{m + 1}{n + 2}.$$

This is the much disputed "law of succession" established by Laplace.

7. Bayes' formula, and other conclusions derived from it, are necessary consequences of fundamental concepts and theorems of the theory of probability. Once we admit these fundamentals, we must admit Bayes' formula and all that follows from it.

But the question arises: When may the various results established in this chapter be legitimately applied? In general, they may be applied whenever *all* the conditions of their validity are fulfilled; and in some artificial theoretical problems like those considered in this chapter, they unquestionably are legitimately applied. But in the case of practical applications it is not easy to make sure that all the conditions of validity are fulfilled, though there are some practical problems in which the use of Bayes' formula is perfectly legitimate.[1] In the history of probability it has happened that even the most illustrious men, like Laplace and Poisson, went farther than they were entitled to go and made free use principally of formulas (6) and (8) in various important practical problems. Against the indiscriminate use of these formulas sharp objections have been raised by a number of authors, especially in modern times.

The first objection is of a general nature and hits the very existence of a priori probabilities. If an urn is given to us and we know only that it contains white and black balls, it is evident that no means are available to estimate a priori probabilities of various hypotheses as to the proportion of white balls. Hence, critics say, a priori probabilities do not exist at all, and it is futile to attempt to apply Bayes' formula to an urn with an unknown proportion of balls. At first this objection may appear

[1] One such problem can be found in an excellent book by Thornton C. Fry, "Probability and Its Engineering Uses," New York, 1928.

very convincing, but its force is somewhat lessened by considering the peculiar mode of existence of mathematical objects.

Some property of integers, unknown to me, is not present in my mind, but it is hardly permissible to say that it does not exist; for it does exist in the minds of those who discover this property and know how to prove it.

Similarly, our urn might have been filled by some person, or selected from among urns with known contents. To this person the a priori probabilities of various proportions of white and black balls might have been known. To us they are unknown, but this should not prevent us from attributing to them some potential mode of existence at least as a sort of belief.

To admit a belief in the existence of certain unknown numbers is common to all sciences where mathematical analysis is applied to the world of reality. If we are allowed to introduce the element of belief into such "exact" sciences as astronomy and physics, it would be only fair to admit it in practical applications of probability.

The second and very serious objection is directed against the use of formula (6), and for similar reasons against formula (8). Imagine, again, that we are provided with an urn containing an enormous number of white and black balls in completely unknown proportion. Our aim is to find the probability that the proportion of white balls to the total number of balls is contained between two given limits. To that end, we make a long series of trials as described in Prob. 5 and find that actually in n trials, white balls appeared m times. The probability we seek would result from Bayes' formula, provided numerical values of a priori probabilities, assumed on belief to be existent, were known. Lacking such knowledge, *an arbitrary assumption is made*, namely, that all the a priori probabilities have the same value. Then, on account of the enormous number of balls in our urn, formula (6) can be used as an approximate expression of P. It can be shown that, given an arbitrary positive number ϵ, however small, the probability of the inequalities

$$\frac{m}{n} - \epsilon < p < \frac{m}{n} + \epsilon$$

can be made as near to 1 as we please by taking the number of trials greater than a certain number $N(\epsilon)$ depending upon ϵ alone. In other words, with practical certainty we can expect the proportion of white balls to the total number of balls in our urn to be contained within arbitrarily narrow limits

$$\frac{m}{n} - \epsilon \quad \text{and} \quad \frac{m}{n} + \epsilon.$$

A conclusion like this would certainly be of the greatest importance. But it is vitiated by the *arbitrary* assumption made at the beginning. The same is true of formula (8) and of Laplace's "law of succession." The objection against using formulas (6) and (8) in circumstances where we are not entitled to use them appears to us as irrefutable, and the numerical applications made by Laplace and others cannot inspire much confidence.

As an example of the extremes to which the illegitimate use of formulas (6) and (8) may lead, we quote from Laplace:

> En faisant, par exemple, remonter la plus ancienne époque de l'histoire à cinq mille ans, ou à 1,826,213 jours, et le Soleil s'étant levé constamment, dans cet intervalle, à chaque révolution de vingt-quatre heures, il y a 1,826,214 à parier contre un qu'il se levera encore demain.

It appears strange that as great a man as Laplace could make such a statement in earnest. However, under proper conditions, it would not be so objectionable. If, from the enormous number $N + 1$ of urns containing each N black and white balls in all possible proportions, one urn is taken and 1,826,213 balls are drawn and returned, and they all turn out to be white, then nobody can deny that there are very nearly 1,826,214 chances against one that the next ball will also be white.

Problems for Solution

1. Three urns of the same appearance have the following proportions of white and black balls:

Urn 1: 1 white, 2 black balls
Urn 2: 2 white, 1 black ball
Urn 3: 2 white, 2 black balls

One of the urns is selected and one ball is drawn. It turns out to be white. What is the probability that the third urn was chosen? *Ans.* $\frac{1}{3}$.

2. Under the same conditions, what is the probability of drawing a white ball again, the first one not having been returned? *Ans.* $\frac{1}{3}$.

3. An urn containing 5 balls has been filled up by taking 5 balls from another urn, which originally had 5 white and 5 black balls. A ball is taken from the first urn, and it happens to be black. What is the probability of drawing a white ball from among the remaining 4? *Ans.* $\frac{5}{9}$.

4. From an urn containing 5 white and 5 black balls, 5 balls are transferred into an empty second urn. From there, 3 balls are transferred into an empty third urn and, finally, one ball is drawn from the latter. It turns out to be white. What is the probability that all 5 balls transferred from the first urn are white? *Ans.* $\frac{1}{126}$.

5. Conditions and notations being the same as in Prob. 3 (page 66), show that the probability for an event to occur in the $(n + 1)$st trial, granted that it has occurred in all the preceding n trials, is never less than the probability for the same event to occur in the nth trial, granting that it has occurred in the preceding $n - 1$ trials.

HINT: it must be proved that

$$\sum_{i=1}^{k} \alpha_i p_i^{n+1} \cdot \sum_{i=1}^{k} \alpha_i p_i^{n-1} \geq \left(\sum_{i=1}^{k} \alpha_i p_i^{n}\right)^2.$$

For that purpose, use Cauchy's inequality

$$\left(\sum_{i=1}^{k}\beta_i\gamma_i\right)^2 \leqq \sum_{i=1}^{k}\beta_i^2 \cdot \sum_{i=1}^{k}\gamma_i^2.$$

6. Assuming that the unknown probability p of an event E can have any value between 0 and 1 and that the a priori probability of its being contained in the interval (α, β) is equal to the length of this interval, prove the following theorem: The probability a posteriori of the inequality

$$p \leqq \sigma$$

after E has occurred m times in n trials is equal to the probability of at least $m + 1$ successes in $n + 1$ independent trials with constant probability σ. (See Prob. 13, page 59.)

7. Assumptions being the same as in the preceding problem, find approximately the probability a posteriori of the inequalities

$$\tfrac{95}{200} \leqq p \leqq \tfrac{115}{200},$$

it being known that in 200 trials an event with the probability p has occurred 105 times. *Ans.* Using the preceding problem and applying Markoff's method, we find $P = 0.846$.

8. An urn contains N white and black balls in unknown proportion. The number of white balls hypothetically may be

$$0, 1, 2, \ldots N$$

and all these hypotheses are considered as equally likely. Altogether n balls are taken from the urn, m of which turned out to be white. Without returning these balls, a new group of n_1 balls is taken, and it is required to find the probability that among them there are m_1 white balls. Naturally, the total number of balls is so large as to have $n + n_1 < N$. *Ans.* The required probability has the same expression

$$C_{n_1}^{m_1}\frac{\int_0^1 x^{m+m_1}(1-x)^{n+n_1-m-m_1}dx}{\int_0^1 x^m(1-x)^{n-m}dx},$$

as in Prob. 4, page 69.

Polynomials ordinarily called "Hermite's polynomials," although they were discovered by Laplace, are defined by

$$H_n(y) = e^{\frac{y^2}{2}}\frac{d^n e^{-\frac{y^2}{2}}}{dy^n}.$$

The first four of them are

$$H_1(y) = -y; \qquad H_2(y) = y^2 - 1; \qquad H_3(y) = -y^3 + 3y; \qquad H_4(y) = y^4 - 6y^2 + 3.$$

They possess the remarkable property of orthogonality:

$$\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} H_m(y)H_n(y)dy = 0 \qquad \text{when} \qquad m \neq n$$

while

$$\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} H_n(y)^2 dy = \sqrt{2\pi}n!.$$

Under very general conditions, a function $f(y)$ defined in the interval $(-\infty, +\infty)$ can be represented by a series

$$f(y) = a_0 + a_1H_1(y) + a_2H_2(y) + \cdots$$

where in general

$$a_k = \frac{1}{k!\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}f(y)H_k(y)dy.$$

Let

$$\alpha = \frac{m}{n} \quad \text{and} \quad h^2 = \frac{n}{\alpha(1-\alpha)}$$

provided $0 < \alpha < 1$.

9. Prove the validity of the following expansion indicated by Ch. Jordan:

$$\frac{(n+1)!}{m!(n-m)!}x^m(1-x)^{n-m} = \frac{h}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\left[1 - \frac{1-2\alpha}{n+2}hH_1(y) + \frac{2n - (11n+6)\alpha(1-\alpha)}{2n(n+2)(n+3)}h^2H_2(y) + \cdots\right]$$

for $0 \leqq x \leqq 1$ where y is a new variable connected to x by the equation

$$x = \alpha + \frac{y}{h}.$$

HINT: Consider the development in a series of Hermite's polynomials of the function

$$f(y) = e^{\frac{y^2}{2}}\left(\alpha + \frac{y}{h}\right)^m\left(1 - \alpha - \frac{y}{h}\right)^{n-m} \quad \text{for} \quad -h\alpha \leqq y \leqq h(1-\alpha)$$

$$f(y) = 0 \quad \text{if either} \quad y < -h\alpha \quad \text{or} \quad y > h(1-\alpha).$$

10. Assuming that the conditions of validity of formula (6) are fulfilled, show that the a posteriori probability of the inequalities

$$\frac{m}{n} - t\sqrt{\frac{\alpha(1-\alpha)}{n}} < p < \frac{m}{n} + t\sqrt{\frac{\alpha(1-\alpha)}{n}}; \qquad \alpha = \frac{m}{n}$$

can be expanded into a convergent series

$$P = \frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{y^2}{2}}dy - \frac{te^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\frac{2n - (11n+6)\alpha(1-\alpha)}{(n+2)(n+3)\alpha(1-\alpha)} + \cdots.$$

When n is large and α is not near either to 0 nor to 1, two terms of this series suffice to give a good approximation to P (Ch. Jordan). Apply this to Prob. 7.

Ans. 0.84585.

References

BAYES: An Essay toward Solving a Problem in the Doctrine of Chances, *Philos. Trans.*, 1764, 1765.

LAPLACE: "Théorie analytique des probabilités," Oeuvres VII, pp. 370–409.

CHAPTER V

USE OF DIFFERENCE EQUATIONS IN SOLVING PROBLEMS OF PROBABILITY

1. The combined use of the theorems of total and compound probability very often leads to an equation in finite differences which, together with the initial conditions supplied by a problem itself, serves to determine an unknown probability. This method of attack is very powerful, and it is often resorted to, especially in the more difficult cases. In this chapter the use of equations in finite differences, applied to a few selected and comparatively easy examples, will be shown; but in Chap. VIII we shall apply the method to a class of interesting and historically important problems.

Certain preliminary explanations are necessary at this point. Again we consider a series of trials resulting in an event E or its opposite, F, but this time we suppose that the trials are dependent, so that the probability of E at a certain trial may vary according to the available information concerning the results of some of the other trials.

A simple and interesting case of dependent trials arises if we suppose that the probability of E in the $(n + 1)$st trial receives a definite value α if E has happened in the preceding nth trial, and this value does not change whatever further information we may possess concerning the results of trials preceding the nth. Also, the probability of E in the $(n + 1)$st trial receives another determined value β if E failed in the nth trial, no matter what happened in the trials preceding the nth.

We have a simple illustration of this kind of dependence, if we suppose that drawings are made from an urn containing black and white balls in a known proportion, and that each ball drawn is returned to the urn, but only after the next drawing has been made. It is obvious that the probability that the $(n + 1)$st ball drawn will be white, becomes perfectly definite if we know what was the color of the ball immediately preceding, and it remains the same no matter what we know about the colors of the $1, 2, \ldots (n - 1)$st balls.

If the trials depend on each other in the above-defined manner, we say that they constitute a "simple chain," to use the terminology of the late A. A. Markoff, who was the first to make a profound study of dependent trials of this and similar, but more complicated, types. It is implied in the definition of a simple chain that it breaks into two separate parts as soon as the result of a certain trial becomes known. For instance, if the result of the fifth trial is known, trials 6, 7, 8, . . . become independent of trials 1, 2, 3, 4, and the chain breaks into two distinct parts: the trials preceding the fifth, and those following it. If the results of trials 1, 2, 3, . . . $(n-1)$ remain unknown, the event E in the following nth trial has a certain probability which we shall denote by p_n. Also, if it becomes known that E happened at trial k, where $k < n - 1$, the probability of E happening in the nth trial receives a different value, $p_n^{(k)}$. It is important to find means to determine the probability p_n, the a priori probability of E in the nth trial when the results of the preceding trials remain unknown; as well as to determine the probability $p_n^{(k)}$ of E in the nth trial when we possess the positive information that E has materialized in the kth$(k < n - 1)$ trial.

2. Thus we are led to the following problem concerning simple chains of dependent trials:

Problem 1. The initial probability p_1 of the event E in a simple chain of trials being known, find the probability p_n of E in the nth trial when the results of the preceding trials remain completely unknown. Also, find the probability $p_n^{(k)}$ of E in the nth trial when it is known that E has happened in the kth trial where $k < n - 1$.

Solution. In the nth trial the event E can happen either preceded by E in the $(n-1)$st trial, the probability of which is p_{n-1}, or preceded by F in the $(n-1)$st trial, the probability of which is $1 - p_{n-1}$. By the theorem of compound probability, the probability of the succession EE is $p_{n-1}\alpha$, while the probability of the succession FE is $(1 - p_{n-1})\beta$. Hence, the total probability p_n is

$$(1) \qquad p_n = \alpha p_{n-1} + \beta(1 - p_{n-1}) = (\alpha - \beta)p_{n-1} + \beta.$$

This is an ordinary equation in finite differences. It has a particular solution

$$p_n = c = \text{const.}$$

where c is determined by the equation

$$c = (\alpha - \beta)c + \beta,$$

whence

$$c = \frac{\beta}{1 + \beta - \alpha}$$

provided $1 + \beta - \alpha \neq 0.$[1] On the other hand, the corresponding

[1] If $1 + \beta - \alpha = 0$ or $\alpha - \beta = 1$, we necessarily have $\alpha = 1$, $\beta = 0$, which means that E *must* occur in all the trials if it actually occurs in the first trial, and *never* occurs if it does not actually occur at the outset. This case, as well as the other extreme case in which $\alpha - \beta = -1$ can therefore be excluded as not possessing real interest.

homogeneous equation

$$y_n = (\alpha - \beta)y_{n-1}$$

has a general solution

$$y_n = C(\alpha - \beta)^{n-1}$$

involving an arbitrary constant C. Adding to it the previously found particular solution, we obtain the general solution of (1) in the form

$$p_n = C(\alpha - \beta)^{n-1} + \frac{\beta}{1 + \beta - \alpha}.$$

The arbitrary constant C is determined by the initial condition

$$C + \frac{\beta}{1 + \beta - \alpha} = p_1$$

so that finally

$$p_n = \frac{\beta}{1 + \beta - \alpha} + \left(p_1 - \frac{\beta}{1 + \beta - \alpha}\right)(\alpha - \beta)^{n-1}.$$

If

$$p_1 = \frac{\beta}{1 + \beta - \alpha}$$

we see that p_n does not depend on n and is constantly equal to p_1. Because we may exclude the cases $\alpha - \beta = 1$ or $\alpha - \beta = -1$, so that $\alpha - \beta$ is contained between -1 and 1, we may conclude from the above expression that p_n, if not a constant, at any rate tends to the limit

$$\frac{\beta}{1 + \beta - \alpha}$$

as n increases indefinitely.

As to $p_n^{(k)}$ we find in a similar way that it satisfies the equation

$$(2) \qquad p_n^{(k)} = \alpha p_{n-1}^{(k)} + \beta(1 - p_{n-1}^{(k)})$$

of the same form as equation (1). But the initial condition in this case is $p_{k+1}^{(k)} = \alpha$ because the probability of E happening in the $(k + 1)$st trial is α when it is known that E occurred in the preceding trial. The solution of (2) satisfying this initial condition is

$$p_n^{(k)} = \frac{\beta}{1 + \beta - \alpha} + \frac{1 - \alpha}{1 + \beta - \alpha}(\alpha - \beta)^{n-k}.$$

As the second term in the right-hand member decreases with increasing n and finally becomes less than any given number, we see that the positive information concerning the result of the kth trial has less and less influence on the probability of E in the following trials, and in remote trials this influence becomes quite insignificant.

Example. An urn contains a white and b black balls, and a series of drawings of one ball at a time is made, the ball removed being returned to the urn immediately after the taking of the next following ball. What is the probability that the nth ball drawn is white when: (a) nothing is known about the preceding drawings; (b) the kth ball drawn is white?

In this particular problem we have $\alpha = \frac{a-1}{a+b-1}$, $\beta = \frac{a}{a+b-1}$, $p_1 = \frac{a}{a+b}$ and

$$\frac{\beta}{1+\beta-\alpha} = \frac{a}{a+b} = p_1.$$

Thus

$$p_n = p_1 = \frac{a}{a+b}.$$

That is, the probability for any ball drawn to be white is the same as that for the first ball, nothing being known about the results of the previous drawings. The expression for $p_n^{(k)}$ is, in this example,

$$p_n^{(k)} = \frac{a}{a+b} + (-1)^{n-k}\frac{b}{(a+b)(a+b-1)^{n-k}}.$$

So, for instance, if $a = 1$, $b = 2$, $n = 5$, $k = 3$,

$$p_5^{(3)} = \frac{1}{3} + \frac{2}{3 \cdot 2^2} = \frac{1}{2};$$

the information that the third ball was white raises to $\frac{1}{2}$ the probability that the fifth ball will be white; it would be $\frac{1}{3}$ without such information.

3. The next problem chosen to illustrate the use of difference equations is interesting in several respects. It was first propounded and solved by de Moivre.

Problem 2. In a series of independent trials, an event E has the constant probability p. If, in this series, E occurs at least r times in succession, we say that there is a run of r successes. What is the probability of having a run of r successes in n trials, where naturally $n > r$?

Solution. Let us denote by y_n the unknown probability of a run of r in n trials. In $n + 1$ trials the probability of a run of r will then be y_{n+1}. Now, a run of r in $n + 1$ trials can happen in two mutually exclusive ways: first, if there is a run of r in the first n trials, and second, if such a run can be obtained only in $n + 1$ trials. The probability of the first hypothesis is y_n. To find the probability of the second hypothesis, we observe that it requires the simultaneous realization of the following conditions:

(a) There is no run of r in the first $n - r$ trials, the probability of which is $1 - y_{n-r}$. (b) In the $(n - r + 1)$st trial, E does not occur, the probability of which is $q = 1 - p$. (c) Finally, E occurs in the remaining r trials, the probability of which is p^r.

As (a), (b), (c) are independent events, their simultaneous materialization has the probability

$$(1 - y_{n-r})qp^r.$$

At the same time, this is the probability of the second hypothesis. Adding it to y_n, we must obtain the total probability y_{n+1}. Thus

$$y_{n+1} = y_n + (1 - y_{n-r})p^r q \tag{3}$$

and this is an ordinary linear difference equation of the order $r + 1$. Together with the obvious initial conditions

$$y_0 = y_1 = \cdot \cdot \cdot = y_{r-1} = 0, \qquad y_r = p^r$$

it serves to determine y_n completely for $n = r + 1, r + 2, \ldots$. For instance, taking $n = r$, we derive from (3)

$$y_{r+1} = p^r + p^r q.$$

Again, taking $n = r + 1$, we obtain

$$y_{r+2} = p^r + 2p^r q$$

and so forth. Although, proceeding thus, step by step, we can find the required probability y_n for any given n, this method becomes very laborious for large n and does not supply us with information as to the behavior of y_n for large n. It is preferable, therefore, to apply known methods of solution to equation (3). First we can obtain a homogeneous equation by introducing $z_n = 1 - y_n$ instead of y_n. The resulting equation in z_n is

$$z_{n+1} - z_n + qp^r z_{n-r} = 0 \tag{4}$$

and the corresponding initial conditions are:

$$z_0 = z_1 = \cdot \cdot \cdot = z_{r-1} = 1; \qquad z_r = 1 - p^r.$$

We could use the method of particular solutions as in the preceding problem, but it is more convenient to use the method of generating functions. The power series in ξ

$$\varphi(\xi) = z_0 + z_1\xi + z_2\xi^2 + \cdot \cdot \cdot$$

is the so-called generating function of the sequence $z_0, z_1, z_2, \ldots$. If we succeed in finding its sum as a definite function of ξ, the development of this function into power series will have precisely z_n as the coefficient of ξ^n. To obtain $\varphi(\xi)$ let us multiply both members of the preceding series by the polynomial

$$1 - \xi + qp^r\xi^{r+1}.$$

The multiplication performed, we have

$$(1 - \xi + qp^r\xi^{r+1})\varphi(\xi) = z_0 + (z_1 - z_0)\xi + \cdots + (z_{r-1} - z_{r-2})\xi^{r-1} + \\ + (z_r - z_{r-1})\xi^r + (z_{r+1} - z_r + qp^r z_0)\xi^{r+1} + \cdots .$$

In the right-hand member the terms involving ξ^{r+1}, ξ^{r+2}, . . . have vanishing coefficients by virtue of equation (4); also $z_k - z_{k-1} = 0$ for $k = 1, 2, 3, \ldots r - 1$, while

$$z_0 = 1 \quad \text{and} \quad z_r - z_{r-1} = -p^r$$

so that

$$(1 - \xi + qp^r\xi^{r+1})\varphi(\xi) = 1 - p^r\xi^r$$

and

$$\varphi(\xi) = \frac{1 - p^r\xi^r}{1 - \xi + qp^r\xi^{r+1}}.$$

The generating function $\varphi(\xi)$ thus is a rational function and can be developed into a power series of ξ according to the known rules. The coefficient of ξ^n gives the general expression for z_n. Without any difficulty, we find the following expression for z_n:

(5)
$$z_n = \beta_{n,r} - p^r\beta_{n-r,r}$$

where

$$\beta_{n,r} = \sum_{l=0}^{\frac{n}{r+1}} (-1)^l C_{n-lr}^l (qp^r)^l$$

and $\beta_{n-r,r}$ is obtained by substituting $n - r$ instead of n. If n is not very large compared with r, formula (5) can be used to compute z_n and

$$y_n = 1 - z_n.$$

For instance, if $n = 20$, $r = 5$, and $p = q = \frac{1}{2}$, we easily find

$$z_{20} = 1 - \frac{15}{64} + \frac{45}{64^2} - \frac{10}{64^3} - \frac{1}{32}\left(1 - \frac{10}{64} + \frac{10}{64^2}\right)$$

and hence

$$z_{20} = 0.75013$$

correct to five decimals; $y_{20} = 0.24987$ is the probability of a run of 5 heads in 20 tossings of a coin.

4. But if n is large in comparison with r, formula (5) would require so much labor that it is preferable to seek for an approximate expression for z_n which will be useful for large values of n. It often happens, and in many branches of mathematics, but especially so in the theory of probability, that exact solutions of problems in certain cases are not of any use. That raises the question of how to supplant them by convenient approximate formulas that readily yield the required numbers. Therefore, it is an important problem to find approximate formulas where exact ones cease to work. Owing to the general importance of approximations, it will not be out of order to enter into a somewhat long and complicated investigation to obtain a workable approximate solution of our problem in the interesting case of a large n.

Since $\varphi(\xi)$ is a rational function, the natural way to get an appropriate expression of z_n would be to resolve $\varphi(\xi)$ into simple fractions, corresponding to various roots of the denominator, and expand those fractions in power series of ξ. However, to attain definite conclusions following this method, we must first seek information concerning roots of the equation

$$1 - \xi + qp^r\xi^{r+1} = 0.$$

5. Let

$$f(\xi) = \xi - 1 - \alpha\xi^{r+1}$$

where

$$\alpha = p^r(1 - p).$$

When p varies from 0 to 1, the maximum of $p^r(1 - p)$ is attained for $p = \frac{r}{r+1}$ and is $r^r/(r + 1)^{r+1}$ so that $\alpha \leqq r^r/(r + 1)^{r+1}$ in all cases. To deal with the most interesting case, we shall assume

$$p < \frac{r}{r+1} \tag{6}$$

which involves

$$\alpha < \frac{r^r}{(r+1)^{r+1}}$$

and we leave it to the reader to discover how the following discussion should be modified if $p \geqq \frac{r}{r+1}$.

When ξ starts to increase from 0, the function $f(\xi)$ steadily increases and attains a positive maximum for $\xi = \xi_0$ where

$$(r + 1)\alpha\xi_0^r = 1$$

after which $f(\xi)$ decreases steadily to negative infinity. Hence, there are two positive roots of the equation $f(\xi) = 0$: ξ_1, which is less than $\frac{r+1}{r}$, and another root greater than this number. This root is $1/p$ if condition (6) is fulfilled.

The remaining roots are all imaginary if r is *odd* and there is one negative root among them if r is *even*.

Now we shall prove that the absolute value of every imaginary or negative root is $>1/p$. Let ρ be the absolute value of any such root.

We have first

$$f(\rho) = \rho - 1 - \alpha\rho^{r+1} < 0$$

so that ρ belongs either to the interval $(0, \xi_1)$ or to the interval $(1/p, +\infty)$, and if we can show that $\rho > \xi_0$ then ρ can be only $> 1/p$. If the root we consider is negative, ρ satisfies the equation

$$F(\rho) = 1 + \rho - \alpha\rho^{r+1} = 0$$

and since $F(\rho)$ increases till a positive maximum for $\rho = \xi_0$ is reached, and then decreases, the root of $F(\rho) = 0$ is necessarily $> \xi_0$. If $\xi = \rho e^{i\theta}$ is an imaginary root of $f(\xi) = 0$ we have, equating imaginary parts,

$$\alpha\rho^r \frac{\sin (r + 1)\theta}{\sin \theta} = 1. \tag{7}$$

But, whatever θ may be

$$\left|\frac{\sin (r + 1)\theta}{\sin \theta}\right| \leqq r + 1$$

the equality sign being excluded if $\sin \theta \neq 0$.[1] Hence,

$$(r + 1)\alpha\rho^r > 1$$

which implies $\rho > \xi_0$. The statement is thus completely proved.

6. The equation

$$\xi - 1 - \alpha\xi^{r+1} = 0$$

can be exhibited in the form

$$\frac{1}{\xi} + \alpha\xi^r = 1.$$

Substituting $\xi = \rho e^{i\theta}$ here, and again equating imaginary parts, we get

$$\alpha\rho^{r+1} \sin r\theta = \sin \theta$$

and, combining this with (7),

$$\rho = \frac{\sin (r + 1)\theta}{\sin r\theta}; \qquad \alpha = \frac{(\sin r\theta)^r \sin \theta}{[\sin (r + 1)\theta]^{r+1}}.$$

[1] The extreme values of the ratio $\frac{\sin m\theta}{\sin \theta}$ (m integer > 1) correspond to certain roots of the equation $m \sin \theta \cos m\theta = \sin m\theta \cos \theta$, but for every root of this equation

$$\left|\frac{\sin m\theta}{\sin \theta}\right| = \frac{m}{\sqrt{1 + (m^2 - 1) \sin^2 \theta}} \leqq m$$

The equality sign is excluded if $\sin \theta$ differs from 0.

If the imaginary part of ξ is positive, the argument θ is contained between 0 and π. In this case, it cannot be less than $\frac{\pi}{r+1}$ or greater than $\pi - \frac{\pi}{r+1}$. For, if $0 < \theta < \frac{\pi}{r+1}$

$$\frac{\sin r\theta}{r\theta} > \frac{\sin (r+1)\theta}{(r+1)\theta}$$

or

$$\frac{\sin r\theta}{\sin (r+1)\theta} > \frac{r}{r+1}.$$

At the same time

$$\frac{\sin \theta}{\sin (r+1)\theta} > \frac{1}{r+1}$$

and hence

$$\alpha = \left\{\frac{\sin r\theta}{\sin (r+1)\theta}\right\}^r \frac{\sin \theta}{\sin (r+1)\theta} > \frac{r^r}{(r+1)^{r+1}},$$

which is impossible. That θ cannot be greater than $\pi - \frac{\pi}{r+1}$ follows simply, because in this case, $\sin (r+1)\theta$ and $\sin r\theta$ would be of opposite signs and ρ would be negative.

As $\frac{\pi}{r+1} \leqq \theta \leqq \pi - \frac{\pi}{r+1}$, we have

$$\rho \sin \theta > \rho \sin \frac{\pi}{r+1}.$$

On the other hand, $\sin x > 2x/\pi$ if $0 < x < \pi/2$ and $\rho > 1/p$. Hence,

$$\rho \sin \theta > \frac{2}{(r+1)p}.$$

Thus, imaginary parts of all complex roots have the same lower bound

$$\frac{2}{(r+1)p}$$

of their absolute values.

7. Denoting the roots of the equation $f(\xi) = 0$ by

$$\xi_k;\ (k = 1, 2, \ldots r+1)$$

we have

$$\varphi(\xi) = \sum_{k=1}^{r+1} \frac{1 - p\xi_k}{(1-p)\xi_k(r+1-r\xi_k)}\left(1 - \frac{\xi}{\xi_k}\right)^{-1}.$$

Hence, expanding each term into power series of ξ and collecting coefficients of ξ^n, we find

$$z_n = \sum_{k=1}^{r+1} \frac{1 - p\xi_k}{(1-p)\xi_k} \cdot \frac{\xi_k^{-n}}{r + 1 - r\xi_k}.$$

For every imaginary root, we have

$$\left|\frac{(1 - p\xi_k)\xi_k^{-n}}{(1-p)\xi_k(r + 1 - r\xi_k)}\right| < \frac{r+1}{r(1-p)}p^{n+2}$$

since

$$|\xi_k^{-1}| < p; \qquad \left|\frac{1}{\xi_k} - p\right| < 2p; \qquad \frac{1}{|r + 1 - r\xi_k|} < \frac{(r+1)p}{2r}.$$

If r is *odd*, there are $r - 1$ imaginary roots and the part in the expression of z_n due to them in absolute value is less than

$$\frac{(r+1)(r-1)}{r(1-p)}p^{n+2} < \frac{r}{1-p}p^{n+2}.$$

The term corresponding to the root $1/p$ vanishes, so that finally

$$z_n = \frac{1 - p\xi_1}{(1-p)\xi_1} \cdot \frac{\xi_1^{-n}}{r + 1 - r\xi_1} + \theta\frac{r}{1-p}p^{n+2}$$

where $|\theta| < 1$ and ξ_1 denotes the least positive root of the equation

$$1 - \xi + qp^r\xi^{r+1} = 0.$$

If r is *even*, there is one negative root. The part of z_n corresponding to this root is less than

$$\frac{2p^{n+2}}{(1-p)r}.$$

The whole contribution due to imaginary and negative roots is less than

$$\frac{r^2 - r}{r(1-p)}p^{n+2} < \frac{r}{1-p}p^{n+2}$$

in absolute value. Thus, no matter whether r is odd or even, we have

$$(8) \qquad z_n = \frac{1 - p\xi_1}{(1-p)\xi_1} \cdot \frac{\xi_1^{-n}}{r + 1 - r\xi_1} + \theta\frac{r}{1-p}p^{n+2}; \qquad -1 < \theta < 1.$$

This is the required expression for z_n, excellently adapted to the case of a large value for n, since then the remainder term involving θ is completely negligible in comparison with the first principal term.

The root ξ_1 can be found either by direct solution of the trinomial equation following Gauss' method, or by application of Lagrange's series. Applying Lagrange's series, we have

$$\xi_1 = 1 + \alpha + \sum_{l=2}^{\infty} \frac{(lr + 2)(lr + 3) \cdots (lr + l)}{l!} \alpha^l$$

$$\log \xi_1 = \alpha + \sum_{l=2}^{\infty} \frac{(lr + 1)(lr + 2) \cdots (lr + l - 1)}{l!} \alpha^l$$

both series being convergent if $|\alpha| < r^r/(r + 1)^{r+1}$ and this condition is satisfied.

8. Let us apply the approximate formula (8) to the case $p = q = \frac{1}{2}$ and $r = 10$. Using Lagrange's series, we find that

$$\xi_1 = 1.0004909$$

and

$$z_n = 1.003947 \cdot (1.0004909)^{-n} + \frac{5\theta}{2^n}.$$

Hence, for $n = 100$, 1,000, 10,000, respectively,

$$z_n = 0.9559;\ 0.6146;\ 0.0074$$

so that, for instance, the probabilities of a run of at least 10 heads in 100, 1,000, or 10,000 throws of a coin are, respectively,

$$0.0441;\ 0.3854;\ 0.9926.$$

Thus, in 10,000 throws, it is quite likely that heads would turn up 10 or more times in succession.

In general, for a given r and increasing n, the probability y_n tends to 1, so that in a very long series of trials, runs of any length are extremely likely to occur, a conclusion which at first sight seems paradoxical.

9. In the preceding examples, an unknown probability was determined by an ordinary equation in finite differences. Very often, however, probability as a function of two or more independent variables is defined by a partial difference equation in two or more independent variables, together with a set of initial conditions suggested by the problem itself. A few examples will suffice to illustrate the use of partial equations in finite differences and to give an idea of the two principal methods for their solution; namely, Laplace's method of generating functions, and the less well known, but elegant, method proposed by Lagrange.

We start with an analytical solution of the problem which was discussed in detail in Chap. III.

Problem 3. Find the probability of exactly x successes in t independent trials with the constant probability p.

Solution by Laplace's Method. Let us denote the required probability by $y_{x,t}$. To obtain x successes in t trials can be possible only in two mutually exclusive ways: (a) by obtaining x successes in $t - 1$ trials and a failure at the last trial; (b) by obtaining success at the last trial and $x - 1$ successes in the preceding $t - 1$ trials. The probability of case (a) is $qy_{x,t-1}$ and that of case (b) is $py_{x-1,t-1}$. The total probability $y_{x,t}$ satisfies the equation

$$y_{x,t} = py_{x-1,t-1} + qy_{x,t-1} \tag{9}$$

for all positive x and t. This equation alone does not determine $y_{x,t}$ completely, but it does so in connection with certain initial conditions. These conditions are

$$\begin{aligned} y_{x,0} &= 0 \qquad \text{if} \qquad x > 0, \\ y_{0,t} &= q^t \qquad \text{if} \qquad t \geqq 0. \end{aligned} \tag{10}$$

The first set of equations is obvious; the second set is the expression of the fact that if there are no successes in t trials, the failures occur t times in succession, and the probability for that is q^t.

Following Laplace, we introduce for a given t the generating function of $y_{0,t}$; $y_{1,t}$; $y_{2,t}$, . . . , that is, the power series

$$\varphi_t(\xi) = y_{0,t} + y_{1,t}\xi + y_{2,t}\xi^2 + \cdots = \sum_{x=0}^{\infty} y_{x,t}\xi^x.$$

Taking $t - 1$ instead of t, separating the first term and multiplying by q, we have

$$q\varphi_{t-1}(\xi) = qy_{0,t-1} + \sum_{x=1}^{\infty} qy_{x,t-1}\xi^x;$$

and similarly

$$p\xi\varphi_{t-1}(\xi) = \sum_{x=1}^{\infty} py_{x-1,t-1}\xi^x.$$

Adding and noting equation (9) we obtain

$$(p\xi + q)\varphi_{t-1}(\xi) = \varphi_t(\xi) + qy_{0,t-1} - y_{0,t},$$

but because of (10)

$$qy_{0,t-1} - y_{0,t} = q^t - q^t = 0$$

and hence,

$$\varphi_t(\xi) = (p\xi + q)\varphi_{t-1}(\xi)$$

for every positive t. Taking $t = 1, 2, 3, \ldots$ and performing successive substitutions, we get

$$\varphi_t(\xi) = (p\xi + q)^t\varphi_0(\xi)$$

and it remains only to find

$$\varphi_0(\xi) = y_{0,0} + y_{1,0}\xi + y_{2,0}\xi^2 + \cdots.$$

But on account of (10), $y_{x,0} = 0$ for $x > 0$, while $y_{0,0} = 1$. Thus,

$$\varphi_0(\xi) = 1$$

and

$$\varphi_t(\xi) = (p\xi + q)^t.$$

To find $y_{x,t}$ it remains to develop the right-hand member in a power series of ξ and to find the coefficient of ξ^x. The binomial theorem readily gives

$$y_{x,t} = \frac{t(t-1)\cdots(t-x+1)}{1\cdot 2\cdots x}p^xq^{t-x}.$$

10. Poisson's Series of Trials. The analytical method thus enables us to find the same expression for probabilities in a Bernoullian series of trials as that obtained in Chap. III by elementary means. Considering how simple it is to arrive at this expression, it may appear that a new deduction of a known result is not a great gain. But one must bear in mind that a little modification of the problem may bring new difficulties which may be more easily overcome by the new method than by a generalization of the old one. Poisson substituted for the Bernoullian series another series of independent trials with probability varying from trial to trial, so that in trials $1, 2, 3, 4, \ldots$ the same event E has different probabilities $p_1, p_2, p_3, p_4, \ldots$ and correspondingly, the opposite event has probabilities $q_1, q_2, q_3, q_4, \ldots$ where $q_k = 1 - p_k$, in general. Now, for the Poisson series, the same question may be asked: what is the probability $y_{x,t}$ of obtaining x successes in t trials? The solution of this generalized problem is easier and more elegant if we make use of difference equations.

First, in the same manner as before, we can establish the equation in finite differences

$$y_{x,t} = p_ty_{x-1,t-1} + q_ty_{x,t-1}. \tag{11}$$

The corresponding set of initial conditions is

$$\begin{aligned} y_{x,0} &= 0 \quad \text{if} \quad x > 0 \\ y_{0,t} &= q_1q_2\cdots q_t \quad \text{if} \quad t > 0 \\ y_{0,0} &= 1. \end{aligned} \tag{12}$$

Giving $\varphi_t(\xi)$ the same meaning as above, we have

$$q_t\varphi_{t-1}(\xi) = q_t y_{0,t-1} + \sum_{x=1}^{\infty} q_t y_{x,t-1}\xi^x$$

$$p_t\xi\varphi_{t-1}(\xi) = \sum_{x=1}^{\infty} p_t y_{x-1,t-1}\xi^x,$$

whence

$$(p_t\xi + q_t)\varphi_{t-1}(\xi) = \varphi_t(\xi) + q_t y_{0,t-1} - y_{0,t};$$

but because of (12)

$$q_t y_{0,t-1} - y_{0,t} = q_1 q_2 \cdots q_t - q_1 q_2 \cdots q_t = 0,$$

and thus

$$\varphi_t(\xi) = (p_t\xi + q_t)\varphi_{t-1}(\xi)$$

whence again

$$\varphi_t(\xi) = (p_1\xi + q_1)(p_2\xi + q_2) \cdots (p_t\xi + q_t)\varphi_0(\xi).$$

However, by virtue of (12), $\varphi_0(\xi) = 1$ so that finally

$$\varphi_t(\xi) = (p_1\xi + q_1)(p_2\xi + q_2) \cdots (p_t\xi + q_t).$$

To find the probability of x successes in t trials in Poisson's case, one needs only to develop the product

$$(p_1\xi + q_1)(p_2\xi + q_2) \cdots (p_t\xi + q_t)$$

according to ascending powers of ξ and to find the coefficient of ξ^x.

11. Solution by Lagrange's Method. We shall now apply to equation (9) the ingenious method devised by Lagrange, with a slight modification intended to bring into full light the fundamental idea underlying this method. Equation (9) possesses particular solutions of the form

$$\alpha^x\beta^t$$

if α and β are connected by the equation

$$\alpha\beta = p + q\alpha.$$

Solving this equation for β, we find infinitely many particular solutions

$$\alpha^x(q + p\alpha^{-1})^t$$

where α is absolutely arbitrary. Multiplying this expression by an arbitrary function $\varphi(\alpha)$ and integrating between arbitrary limits, we obtain other solutions of equation (9). Now the question arises of how to choose $\varphi(\alpha)$ and the path of integration to satisfy not only equation (9) but also initial conditions (10). We shall assume that $\varphi(\alpha)$ is a regular function of a complex variable α in a ring between two concentric circles, with their center at the origin, and that it can therefore be represented in this ring by Laurent's series

$$\varphi(\alpha) = \sum_{n=-\infty}^{\infty} c_n\alpha^n.$$

If c is a circle concentric with the regularity ring of $\varphi(\alpha)$ and situated inside it, the integral

$$y_{x,t} = \frac{1}{2\pi i}\int_c \alpha^{x-1}(q + p\alpha^{-1})^t\varphi(\alpha)d\alpha$$

is perfectly determined and represents a solution of (9). To satisfy the initial conditions, we have first the set of equations

$$\frac{1}{2\pi i}\int_c \alpha^{x-1}\varphi(\alpha)d\alpha = 0 \qquad \text{for} \qquad x = 1, 2, 3, \ldots$$

which show that all the coefficients c_n with negative subscripts vanish, and that $\varphi(\alpha)$ is regular about the origin. The second set of equations obtained by setting $x = 0$

$$\frac{1}{2\pi i}\int_c (q + p\alpha^{-1})^t\frac{\varphi(\alpha)}{\alpha}d\alpha = q^t \qquad \text{for} \qquad t = 0, 1, 2, \ldots$$

serves to determine $\varphi(\alpha)$. If ϵ is a sufficiently small complex parameter, this set of equations is entirely equivalent to a single equation:

$$\frac{1}{2\pi i}\int_c \frac{\varphi(\alpha)d\alpha}{\alpha - \epsilon(p + q\alpha)} = \frac{1}{1 - \epsilon q}.$$

Now the integrand within the circle c has a single pole α_0 determined by the equation

$$\alpha_0 = \epsilon(p + q\alpha_0)$$

and the corresponding residue is

$$\frac{\varphi(\alpha_0)}{1 - q\epsilon}.$$

At the same time, this is the value of the left-hand member of the above equation, so that

$$\frac{\varphi(\alpha_0)}{1 - q\epsilon} = \frac{1}{1 - q\epsilon}$$

or

$$\varphi(\alpha_0) = 1$$

for all sufficiently small ϵ or α_0. That is, $\varphi(\alpha) = 1$ and

$$y_{x,t} = \frac{1}{2\pi i}\int_c \alpha^{x-1}\left(q + \frac{p}{\alpha}\right)^t d\alpha$$

is the required solution. It remains to find the residue of the integrand; that is, the coefficient of $1/\alpha$ in the development of

$$\alpha^{x-1}\left(q + \frac{p}{\alpha}\right)^t$$

in series of ascending powers of α. That can be easily done, using the binomial development, and we obtain

$$y_{x,t} = C_t^x p^x q^{t-x}$$

as it should be.

12. Problem 4. Two players, A and B, agree to play a series of games on the condition that A wins the series if he succeeds in winning a games before B wins b games. The probability of winning a single game is p for A and $q = 1 - p$ for B, so that each game must be won by either A or B. What is the probability that A will win the series?

Solution. This historically important problem was proposed as an exercise (Prob. 12, page 58) with a brief indication of its solution based on elementary principles. To solve it analytically, let us denote by $y_{x,t}$ the probability that A will win when x games remain for him to win, while his adversary B has t games left to win. Considering the result of the game immediately following, we distinguish two alternatives: (a) A wins the next game (probability p) and has to win $x - 1$ games before B wins t games (probability $y_{x-1,t}$); (b) A loses the next game (probability q) and has to win x games before B can win $t - 1$ games (probability $y_{x,t-1}$). The probabilities of these two alternatives being $py_{x-1,t}$ and $qy_{x,t-1}$ their sum is the total probability $y_{x,t}$. Thus, $y_{x,t}$ satisfies the equation

$$y_{x,t} = py_{x-1,t} + qy_{x,t-1}. \tag{13}$$

Now, $y_{x,0} = 0$ for $x > 0$, which means that A cannot win, B having won all his games. Also, $y_{0,t} = 1$ for $t > 0$, which means that A surely wins when he has no more games to win. The initial conditions in our problem are, therefore,

$$(14) \qquad \begin{aligned} y_{x,0} &= 0 \quad \text{if} \quad x > 0; \\ y_{0,t} &= 1 \quad \text{if} \quad t > 0. \end{aligned}$$

The symbol $y_{0,0}$ has no meaning as a probability, and remains undefined. For the sake of simplicity we shall assume, however, that $y_{0,0} = 0$.

Application of Laplace's Method. Again, let

$$\varphi_x(\xi) = y_{x,0} + y_{x,1}\xi + y_{x,2}\xi^2 + \cdots$$

be the generating function of the sequence $y_{x,0}$; $y_{x,1}$; $y_{x,2}$, . . . corresponding to an arbitrary $x > 0$. We have

$$q\xi\varphi_x(\xi) = \sum_{t=1}^{\infty} qy_{x,t-1}\xi^t$$

$$p\varphi_{x-1}(\xi) = py_{x-1,0} + \sum_{t=1}^{\infty} py_{x-1,t}\xi^t$$

and

$$q\xi\varphi_x(\xi) + p\varphi_{x-1}(\xi) = py_{x-1,0} + \sum_{t=1}^{\infty}(py_{x-1,t} + qy_{x,t-1})\xi^t$$

or, because of (13),

$$q\xi\varphi_x(\xi) + p\varphi_{x-1}(\xi) = py_{x-1,0} - y_{x,0} + \varphi_x(\xi).$$

Now, for every $x > 0$

$$y_{x,0} = y_{x-1,0} = 0$$

in conformity with the first set of initial conditions, which allows us to present the preceding relation as follows:

$$\varphi_x(\xi) = \frac{p}{1 - q\xi}\varphi_{x-1}(\xi),$$

whence

$$\varphi_x(\xi) = \frac{p^x}{(1 - q\xi)^x}\varphi_0(\xi).$$

But

$$\varphi_0(\xi) = y_{0,0} + y_{0,1}\xi + y_{0,2}\xi^2 + \cdots = \xi + \xi^2 + \xi^3 + \cdots = \frac{\xi}{1 - \xi}$$

and finally

$$\varphi_x(\xi) = \frac{\xi p^x}{(1 - \xi)(1 - q\xi)^x}.$$

It remains to develop the right-hand member in a power series of ξ and find the coefficient of ξ^t. As

$$\frac{\xi}{1 - \xi} = \xi + \xi^2 + \xi^3 + \cdots$$

and

$$\frac{1}{(1 - q\xi)^x} = 1 + \frac{x}{1}q\xi + \frac{x(x + 1)}{1 \cdot 2}q^2\xi^2 + \cdots$$

we readily get, multiplying these series according to the ordinary rules,

$$y_{x,t} = p^x\left[1 + \frac{x}{1}q + \frac{x(x + 1)}{1 \cdot 2}q^2 + \cdots + \frac{x(x + 1) \cdots (x + t - 2)}{1 \cdot 2 \cdots (t - 1)}q^{t-1}\right]$$

which coincides with the elementary solution indicated on page 58.

Application of Lagrange's Method. Equation (13) has particular solutions of the form

$$\alpha^x\beta^t$$

where

$$\alpha\beta = p\beta + q\alpha.$$

Hence, we can either express α by β or β by α. Leaving it to the reader to follow the second alternative, we shall express α as a function of β and seek the required solution in the form

$$y_{x,t} = \frac{p^x}{2\pi i}\int_c \frac{\beta^t}{(1 - q\beta^{-1})^x}\varphi(\beta)d\beta$$

where $\varphi(\beta)$ is again supposed to be developable in Laurent's series in a certain ring; c is a circle described about the origin and entirely within that ring. Setting $x = 0$, we must have

$$\frac{1}{2\pi i}\int_c \beta^t\varphi(\beta)d\beta = 1 \qquad \text{for} \qquad t = 1, 2, 3, \ldots$$

and this set of equations is satisfied if we take

$$\varphi(\beta) = \frac{1}{\beta^2} + \frac{1}{\beta^3} + \cdots = \frac{1}{\beta(\beta - 1)}; \qquad |\beta| > 1.$$

Now we have

$$y_{x,t} = \frac{p^x}{2\pi i}\int_c \frac{\beta^{t-1}d\beta}{(1 - q\beta^{-1})^x(\beta - 1)}$$

and for $t = 0$

$$y_{x,0} = \frac{p^x}{2\pi i}\int_c \frac{d\beta}{(1 - q\beta^{-1})^x\beta(\beta - 1)} = 0$$

as it should be, because for $|\beta| > 1$ the integrand can be developed into a power series of $1/\beta$, the term with $1/\beta$ being absent. Thus, the required solution is given by

$$y_{x,t} = \frac{p^x}{2\pi i}\int \frac{\beta^{t-1}d\beta}{(1 - q\beta^{-1})^x(\beta - 1)}$$

where c is a circle of radius >1 described about the origin. The final expression for $y_{x,t}$ is obtained as the coefficient of $1/\beta$ in the development of

$$\frac{p^x\beta^{t-1}}{(1 - q\beta^{-1})^x(\beta - 1)}$$

into power series of $1/\beta$. We obtain the same expression as before.

Problems for Solution

1. Each of n urns contains a white and b black balls. One ball is transferred from the first urn into the second, another one from the second into the third, and so on. Finally, a ball is drawn from the nth urn. What is the probability that it is white, when it is known that the first ball transferred was white?

Ans. $\dfrac{a}{a+b} + \dfrac{b}{a+b}(a + b + 1)^{1-n}.$

2. Two urns contain, respectively, a white and b black, and b white and a black balls. A series of drawings is made, according to the following rules:

a. Each time only one ball is drawn and immediately returned to the same urn it came from.

b. If the ball drawn is white, the next drawing is made from the first urn.

c. If it is black, the next drawing is made from the second urn.

d. The first ball drawn comes from the first urn.

What is the probability that the nth ball drawn will be white?

$$\textit{Ans. } p_n = \frac{1}{2} + \frac{1}{2}\left(\frac{a-b}{a+b}\right)^n.$$

3. Find the probability of a run of 5 in a series of 15 trials with constant probability $p = \frac{1}{3}$. *Ans.* $y_{15} = 23 \cdot 3^{-6} - 70 \cdot 3^{-12} = 0.0314184.$

4. How many throws of a coin suffice to give a probability of more than 0.999 for a run of at least 100 heads? *Ans.* $1.76 \cdot 10^{31}$ throws suffice.

5. What is the least number of trials assuring a probability of $\geqq \frac{1}{2}$ for a run of at least 10 successes if $p = q = \frac{1}{2}$? *Ans.* 1,420.

6. Seven urns contain black and white balls in the following proportions:

Urns	1	2	3	4	5	6	7
White	1	2	2	3	2	3	4
Black	2	1	2	1	5	2	5

One ball is drawn from each urn. What is the probability that there will be among them exactly 3 white balls? *Ans.* Coefficient of ξ^3 in.

$$(\tfrac{1}{3}\xi + \tfrac{2}{3})(\tfrac{2}{3}\xi + \tfrac{1}{3})(\tfrac{1}{2}\xi + \tfrac{1}{2})(\tfrac{3}{4}\xi + \tfrac{1}{4})(\tfrac{2}{7}\xi + \tfrac{5}{7})(\tfrac{3}{5}\xi + \tfrac{2}{5})(\tfrac{4}{9}\xi + \tfrac{5}{9})$$

or

$$\tfrac{227}{810} = 0.28025.$$

7. Two players, each possessing \$2, agree to play a series of games. The probability of winning a single game is $\frac{1}{2}$ for both, and the loser pays \$1 to his adversary after each game. Find the probability for each one of them to be ruined at or before the nth game?

Solution. Let y_m be the probability that after playing $2m$ games, neither of the players is ruined. We have

$$y_{m+1} = \tfrac{1}{2}y_m$$

and hence

$$y_m = \frac{1}{2^m}$$

The probability for one of the players to be ruined at or before the nth game is $\dfrac{1}{2} - \dfrac{1}{2^{m+1}}$ if $n = 2m$ or $n = 2m + 1$.

8. Solve the same problem if each player enters the game with \$3.

Ans. $\frac{1}{2} - \frac{1}{2}(\frac{3}{4})^{m-1}$ if $n = 2m - 1$ or $n = 2m$.

9. Players $A_1, A_2, \ldots A_{n+1}$ play a series of games in the following order: first A_1 plays with A_2; the loser is out and the winner plays with the following player, A_3; the loser is out again and the next game is played with A_4, and so on; the loser always being out and his place taken by the next following player. The probability of winning a single game is ½ for each player and the series is won by the player who succeeds in winning over all his adversaries in succession. What is the probability that the series will stop exactly at the xth game? What is the probability that the series will stop before or at the xth game?

Solution. Let y_x be the probability that the series terminates exactly at the xth game. That means that the player who won the game entered at the $(x - n + 1)$st game and won successively the n following games. Now, there are $n - 1$ cases to be distinguished according as the player beaten at the $(x - n + 1)$st game has already won $1, 2, 3, \ldots n - 1$ games. Let p_k be the probability that the loser in the $(x - n + 1)$st game previously has won k games. The probability of ending the series in this case is $p_k/2^n$. On the other hand,

$$\frac{p_k}{2^{n-k}} = y_{x-k}$$

so that

$$\frac{p_k}{2^n} = \frac{y_{x-k}}{2^k}.$$

Hence, for $x > n$

$$y_x = \frac{1}{2}y_{x-1} + \frac{1}{4}y_{x-2} + \cdots + \frac{1}{2^{n-1}}y_{x-n+1}.$$

Initial conditions:

$$y_1 = y_2 = \cdots = y_{n-1} = 0; \qquad y_n = \frac{1}{2^{n-1}}.$$

The generating function of y_x:

$$y_1 + y_2\xi + y_3\xi^2 + \cdots = \frac{\xi^{n-1}\left(1 - \frac{\xi}{2}\right)}{2^{n-1}\left(1 - \xi + \frac{\xi^n}{2^n}\right)}$$

and the generating function of the probability that the series will end before or at the xth game is

$$\frac{\xi^{n-1}\left(1 - \frac{\xi}{2}\right)}{2^{n-1}(1 - \xi)\left(1 - \xi + \frac{\xi^n}{2^n}\right)}.$$

10. Three players, A, B, C, play a series of games, each game being won by one of them. If the probabilities for A, B, C to win a single game are p, q, r, find the probability of A winning a games before B and C win b and c games, respectively.

Solution. Let $A_{x,y,z}$ denote the probability for A to win the series when he has still to win x games, while B and C have to win y and z games, respectively. First, we can establish the equation

$$A_{x,y,z} = pA_{x-1,y,z} + qA_{x,y-1,z} + rA_{x,y,z-1}.$$

Next, $A_{0,y,z} = 1$ for positive y, z, and $A_{x,0,z} = 0$ for positive x, z; $A_{x,y,0} = 0$ for positive x, y. Besides, although this is only a formal simplification, we shall assume

$A_{x,0,z} = 0$, $A_{x,y,0} = 0$ when x or y or z vanishes. For the generating function of $A_{x,y,z}$

$$\phi_x(\xi, \eta) = \sum_{y,z=0}^{\infty} A_{x,y,z}\xi^y\eta^z$$

we find the equation

$$\phi_x(\xi, \eta) = \frac{p}{1 - q\xi - r\eta}\phi_{x-1}(\xi, \eta)$$

whence

$$\phi_x(\xi, \eta) = \frac{p^x}{(1 - q\xi - r\eta)^x} \cdot \frac{\xi\eta}{(1 - \xi)(1 - \eta)}.$$

The final answer is

$$A_{a,b,c} = p^a\left[1 + \frac{a}{1}\overline{(q + r)} + \frac{a(a + 1)}{1 \cdot 2}\overline{(q + r)^2} + \frac{a(a + 1)(a + 2)}{1 \cdot 2 \cdot 3}\overline{(q + r)^3} + \cdots\right],$$

the dash indicating that powers of q and r with the exponents $\geqq b$ and $\geqq c$ are omitted.

Obviously, the same method can be extended to any number of players, and leads to a perfectly analogous expression of probability.

11. An urn contains n balls altogether, and among them a white balls. In a series of drawings, each time one ball is drawn, whatever its color may be, it is replaced by a white ball. Find the probability $y_{x,r}$ that after r drawings there are x white balls in the urn.

Solution. The required probability satisfies the equation

$$y_{x,r+1} = \frac{n - x + 1}{n}y_{x-1,r} + \frac{x}{n}y_{x,r}.$$

Besides,

$$y_{a,0} = 1, \qquad y_{x,0} = 0 \quad \text{if} \quad x \neq a, \qquad y_{x,r} = 0 \quad \text{if} \quad x < a.$$

From the preceding equation, combined with the initial conditions, we find successively

$$y_{a,r} = \left(\frac{a}{n}\right)^r$$

$$y_{a+1,r} = (n - a)\left[\left(\frac{a + 1}{n}\right)^r - \left(\frac{a}{n}\right)^r\right]$$

$$y_{a+2,r} = \frac{(n - a)(n - a - 1)}{1 \cdot 2}\left[\left(\frac{a + 2}{n}\right)^r - 2\left(\frac{a + 1}{n}\right)^r + \left(\frac{a}{n}\right)^r\right]$$

and so on.

12. If, in the problem of runs, p is supposed to be $> \frac{r}{r + 1}$, prove that the probability of a run of r in n trials is greater than

$$1 - \left(\frac{p - p_1}{r - (r + 1)p_1} + \frac{r(p + p_1)}{2}\right)\frac{p_1^{n+1}}{1 - p_1}$$

where $p_1 < \frac{r}{r + 1}$ is a root of the equation

$$p_1^r(1 - p_1) = p^r(1 - p).$$

13. To find an asymptotic expression of probability for a run of r in n independent trials, if $p \geqq \frac{r}{r+1}$, the following proposition is of importance: Imaginary and negative roots of the equation

$$(1-s)x^n - x + s = 0; \qquad 0 < s \leqq \frac{n}{n-1}$$

are, in absolute value, greater than the root $R > 1$ of the equation

$$(1-s)R^n - R + s\cos\frac{2\pi}{n} = 0.$$

Prove the truth of this statement.

14. Given s urns containing the same number n of black and white balls in known proportions, drawings are made in the following manner: first, a single ball is drawn out of every urn; second, the ball drawn from the first urn is placed into the second; that drawn from the second is placed in the third, and so on; finally, the ball drawn from the last urn is placed in the first, so that again every urn contains n balls. Supposing that this operation is repeated t times, find the probability of drawing a white ball from the xth urn.

Solution. Let $y_{x,t}$ be the required probability. First, it can be shown that it satisfies the equation

$$y_{x,t} = \left(1-\frac{1}{n}\right)y_{x,t-1} + \frac{1}{n}y_{x-1,t-1}.$$

The initial probabilities $y_{1,0}, y_{2,0}, \ldots y_{s,0}$ are known; and, moreover, the function $y_{x,t}$ must satisfy a boundary condition of the periodic type, $y_{0,t} = y_{s,t}$. Hence, applying Lagrange's method, the following solution is found

$$y_{x,t} = \left(1-\frac{1}{n}\right)^t\left[f(x) + \frac{t}{1\cdot(n-1)}f(x-1) + \frac{t(t-1)}{1\cdot 2(n-1)^2}f(x-2) + \cdots\right]$$

where

$$f(x) = y_{x,0} \qquad \text{when} \qquad x > 0$$

and the definition is extended to $x \leqq 0$ by setting

$$f(-x) = f(s-x).$$

If, to begin with, all urns contain the same number of white and black balls, so that $f(x) = \text{const.} = p$, we shall have, no matter what t is,

$$y_{x,t} = p\left(1-\frac{1}{n}\right)^t\left(1+\frac{1}{n-1}\right)^t = p.$$

References

DEMOIVRE: "Doctrine of Chances," 3d ed., 1756.
LAPLACE: "Théorie analytique des probabilités," Oeuvres VII, pp. 49 *ff.*
LAGRANGE: "Mémoire sur les suites récurrentes dont les termes varient de plusieurs manières différentes, etc.," Oeuvres IV, pp. 151 *ff.*

CHAPTER VI

BERNOULLI'S THEOREM

1. This chapter will be devoted to one of the most important and beautiful theorems in the theory of probability, discovered by Jacob Bernoulli and published with a proof remarkably rigorous (save for some irrelevant limitations assumed in the proof) in his admirable posthumous book "Ars conjectandi" (1713). This book is the first attempt at scientific exposition of the theory of probability as a separate branch of mathematical science.

If, in n trials, an event E occurs m times, the number m is called the "frequency" of E in n trials, and the ratio m/n receives the name of "relative frequency." Bernoulli's theorem reveals an important probability relation between the relative frequency of E and its probability p.

Bernoulli's Theorem. *With the probability approaching 1 or certainty as near as we please, we may expect that the relative frequency of an event E in a series of independent trials with constant probability p will differ from that probability by less than any given number* $\epsilon > 0$, *provided the number of trials is taken sufficiently large.*

In other words, given two positive numbers ϵ and η, the probability P of the inequality

$$\left|\frac{m}{n} - p\right| < \epsilon$$

will be greater than $1 - \eta$ if the number of trials is above a certain limit depending upon ϵ and η.

Proof. Several proofs of this important theorem are known which are shorter and simpler but less natural than Bernoulli's original proof. It is his remarkable proof that we shall reproduce here in modernized form.

a. Denoting by T_m, as usual, the probability of m successes in n trials, we shall show first that

$$(1) \qquad \frac{T_{b+k}}{T_b} < \frac{T_{a+k}}{T_a}$$

if $b > a$ and $k > 0$. Since the ratio

$$\frac{T_{x+1}}{T_x} = \frac{n-x}{x+1}\frac{p}{q}$$

decreases as x increases we have for $b > a$

$$\frac{T_{b+1}}{T_b} < \frac{T_{a+1}}{T_a} \quad \text{or} \quad \frac{T_{b+1}}{T_{a+1}} < \frac{T_b}{T_a}.$$

Changing b, a, respectively, into $b + 1, a + 1; b + 2, a + 2; \cdots b + k, a + k$, it follows from the last inequality that

$$\frac{T_{b+k}}{T_{a+k}} < \frac{T_{b+k-1}}{T_{a+k-1}} < \cdots < \frac{T_{b+1}}{T_{a+1}} < \frac{T_b}{T_a},$$

that is,

$$\frac{T_{b+k}}{T_b} < \frac{T_{a+k}}{T_a}.$$

b. Integers λ and μ being determined by the inequalities

$$\lambda - 1 < np \leqq \lambda, \qquad \mu - 1 < np + n\epsilon \leqq \mu$$

the probabilities A and C of the inequalities

$$0 \leqq \frac{m}{n} - p < \epsilon; \qquad \frac{m}{n} - p \geqq \epsilon$$

are represented, respectively, by the sums

$$\begin{aligned} A &= T_\lambda + T_{\lambda+1} + \cdots + T_{\mu-1} \\ C &= T_\mu + T_{\mu+1} + \cdots + T_n \end{aligned}$$

the first of which contains $\mu - \lambda = g$ terms. Combining terms of the second sum into groups of g terms (the last group may consist of less than g terms) and setting for brevity

$$\begin{aligned} A_1 &= T_\mu + T_{\mu+1} + \cdots + T_{\mu+g-1} \\ A_2 &= T_{\mu+g} + T_{\mu+g+1} + \cdots + T_{\mu+2g-1} \\ A_3 &= T_{\mu+2g} + T_{\mu+2g+1} + \cdots + T_{\mu+3g-1} \\ &\cdot \end{aligned}$$

we shall have

$$C = A_1 + A_2 + A_3 + \cdots$$

and at the same time

(2) $$\frac{A_1}{A} < \frac{T_\mu}{T_\lambda}, \quad \frac{A_2}{A_1} < \frac{T_\mu}{T_\lambda}, \quad \cdots.$$

The ratio

$$\frac{A_1}{A} = \frac{T_{\lambda+g} + T_{\lambda+g+1} + \cdots + T_{\lambda+2g-1}}{T_\lambda + T_{\lambda+1} + \cdots + T_{\lambda+g-1}}$$

is less than the greatest of numbers

$$\frac{T_{\lambda+g}}{T_\lambda}, \frac{T_{\lambda+g+1}}{T_{\lambda+1}}, \cdots \frac{T_{\lambda+2g-1}}{T_{\lambda+g-1}}.$$

But by inequality (1)

$$\frac{T_{\lambda+g}}{T_\lambda} > \frac{T_{\lambda+g+1}}{T_{\lambda+1}} > \cdots > \frac{T_{\lambda+2g-1}}{T_{\lambda+g-1}}$$

hence

$$\frac{A_1}{A} < \frac{T_\mu}{T_\lambda}.$$

Similarly,

$$\frac{A_2}{A_1} < \frac{T_{\mu+g}}{T_\mu}, \qquad \frac{A_3}{A_2} < \frac{T_{\mu+2g}}{T_{\mu+g}}, \cdots$$

and again by inequality (1)

$$\frac{T_{\mu+g}}{T_\mu} < \frac{T_{\lambda+g}}{T_\lambda}, \qquad \frac{T_{\mu+2g}}{T_{\mu+g}} < \frac{T_{\mu+g}}{T_\mu}, \cdots.$$

Consequently

$$\frac{A_2}{A_1} < \frac{T_\mu}{T_\lambda}, \qquad \frac{A_3}{A_2} < \frac{T_\mu}{T_\lambda}, \cdots$$

and inequalities (2) are established.

c. For $x \geqq \lambda$

$$\frac{T_{x+1}}{T_x} < 1.$$

It suffices to show that

$$\frac{T_{\lambda+1}}{T_\lambda} = \frac{n-\lambda}{\lambda+1}\frac{p}{q} < 1.$$

As $\lambda \geqq np$

$$\frac{n-\lambda}{\lambda+1}\frac{p}{q} \leqq \frac{npq}{npq+q} < 1$$

which shows that $\dfrac{T_{\lambda+1}}{T_\lambda} < 1.$

The inequality just established shows that in the following expression:

$$\frac{T_\mu}{T_\lambda} = \frac{T_\mu}{T_{\mu-1}} \cdot \frac{T_{\mu-1}}{T_{\mu-2}} \cdots \frac{T_{\mu-\alpha+1}}{T_{\mu-\alpha}} \cdot \frac{T_{\mu-\alpha}}{T_{\mu-\alpha-1}} \cdots \frac{T_{\lambda+1}}{T_\lambda}$$

all the factors are <1. Consequently, if we retain $\alpha \leqq g$ first factors only, replacing the others by 1, we get

$$\frac{T_\mu}{T_\lambda} \leqq \frac{T_\mu}{T_{\mu-1}} \cdot \frac{T_{\mu-1}}{T_{\mu-2}} \cdots \frac{T_{\mu-\alpha+1}}{T_{\mu-\alpha}}.$$

Moreover,

$$\frac{T_\mu}{T_{\mu-1}} < \frac{T_{\mu-1}}{T_{\mu-2}} < \cdots < \frac{T_{\mu-\alpha+1}}{T_\mu}$$

whence the following important inequality results:

$$\frac{T_\mu}{T_\lambda} < \left(\frac{n - \mu + \alpha}{\mu - \alpha + 1}\frac{p}{q}\right)^\alpha. \tag{3}$$

Here α is an arbitrary positive integer $\leqq g$.

Now, let ϵ be an arbitrary positive number. Then we can show that for

$$n \geqq \frac{\alpha(1 + \epsilon) - q}{\epsilon(p + \epsilon)} \tag{4}$$

we have both

$$\text{(i)}\ \frac{n - \mu + \alpha}{\mu - \alpha + 1}\frac{p}{q} \leqq \frac{p}{p + \epsilon} \qquad \text{and} \qquad \text{(ii)}\ \alpha \leqq g.$$

Since $\mu \geqq np + n\epsilon$, it suffices to show that (i) is satisfied for $\mu = np + n\epsilon$. If $\mu = np + n\epsilon$ inequality (i) is equivalent to

$$\frac{nq - n\epsilon + \alpha}{np + n\epsilon - \alpha + 1} \leqq \frac{q}{p + \epsilon}$$

or, after obvious simplifications,

$$n\epsilon(p + \epsilon) \geqq \alpha(1 + \epsilon) - q.$$

But this inequality follows from (4). To establish (ii), since α and g are integers, it suffices to show that $\alpha < g + 1$. But $\mu \geqq np + n\epsilon$, $\lambda < np + 1$ and consequently $g + 1 > n\epsilon$. Hence (ii) will be established if we can show that $n\epsilon \geqq \alpha$ which by virtue of (4) will be true if

$$\frac{\alpha(1 + \epsilon) - q}{p + \epsilon} \geqq \alpha$$

that is, if

$$\alpha(1 + \epsilon) - q \geqq \alpha p + \alpha\epsilon$$

or $\alpha q - q \geqq 0$ which is obviously true, α being a positive integer.

d. The auxiliary integer α is still at our disposal. Given an arbitrary positive number $\eta < 1$ we shall determine α as the least integer satisfying the inequality

$$\left(\frac{p}{p + \epsilon}\right)^\alpha \leqq \eta \qquad \text{or} \qquad \alpha \geqq \frac{\log \frac{1}{\eta}}{\log\left(1 + \frac{\epsilon}{p}\right)}.$$

At the same time

$$\frac{\log \frac{1}{\eta}}{\log\left(1 + \frac{\epsilon}{p}\right)} > \alpha - 1$$

and since $\log\left(1 + \frac{\epsilon}{p}\right) > \frac{\epsilon}{p + \epsilon}$, we shall have

$$\alpha < 1 + \frac{p + \epsilon}{\epsilon} \log \frac{1}{\eta}$$

and

$$\frac{\alpha(1 + \epsilon) - q}{\epsilon(p + \epsilon)} < \frac{1 + \epsilon}{\epsilon^2} \log \frac{1}{\eta} + \frac{1}{\epsilon}.$$

Consequently, if

$$n \geqq \frac{1 + \epsilon}{\epsilon^2} \log \frac{1}{\eta} + \frac{1}{\epsilon} \tag{5}$$

then by virtue of (i) and (3)

$$\frac{T_\mu}{T_\lambda} < \eta,$$

and by virtue of (2)

$$A_1 < A\eta,\ A_2 < A_1\eta < A\eta^2,\ A_3 < A_2\eta < A\eta^3,\ \cdots,$$

whence

$$C < A\eta + A\eta^2 + A\eta^3 + \cdots = \frac{A\eta}{1 - \eta}. \tag{6}$$

This inequality holds if n satisfies (5). No trace of the auxiliary integer α is left.

e. Let us now consider the inequalities

$$-\epsilon < \frac{m}{n} - p < 0 \qquad \text{and} \qquad \frac{m}{n} - p \leqq -\epsilon$$

and introduce their respective probabilities B and D. These inequalities are equivalent to

$$0 < \frac{n - m}{n} - q < \epsilon \qquad \text{and} \qquad \frac{n - m}{n} - q \geqq \epsilon.$$

It is apparent that we can interpret B or D as probabilities that the number of occurrences $m' = n - m$ of the event F opposite to E in n trials will satisfy either the inequality $0 < \frac{m'}{n} - q < \epsilon$ or $\frac{m'}{n} - q \geqq \epsilon$. Since the right-hand side of (5) contains only given numbers ϵ, η it is clear that

$$D < \frac{B\eta}{1 - \eta} \tag{7}$$

if (5) is satisfied.

Now $A + B = P$ is the probability of the inequality

$$\left|\frac{m}{n} - p\right| < \epsilon$$

and $C + D = Q$ is the probability of the opposite inequality

$$\left|\frac{m}{n} - p\right| \geqq \epsilon.$$

Hence $P + Q = 1$. Moreover, by (6) and (7)

$$Q < \frac{P\eta}{1 - \eta}.$$

Consequently,

$$P + \frac{P\eta}{1 - \eta} > 1$$

or

$$P > 1 - \eta$$

if only

$$n \geqq \frac{1 + \epsilon}{\epsilon^2} \log \frac{1}{\eta} + \frac{1}{\epsilon}.$$

This completes the proof of Bernoulli's theorem.

For example, if $p = q = \frac{1}{2}$ and $\epsilon = 0.01$, $\eta = 0.001$ we get from (5)

$$n \geqq 69{,}869$$

which shows that in 69,869 trials or more there are at least 999 chances against 1 that the relative frequency will differ from $\frac{1}{2}$ by less than $\frac{1}{100}$. The number 69,869 found as a lower limit of the number of trials is much too large. A much smaller number of trials would suffice to fulfill all the requirements. From a practical standpoint, it is important to find as low a limit as possible for the necessary number of trials (given ϵ and η). With this problem we shall deal in the next chapter.

2. Bernoulli's theorem states that for arbitrarily given ϵ and η there exists a number $n_0(\epsilon, \eta)$ such that for any single value $n > n_0(\epsilon, \eta)$ the probability of the inequality

$$\left|\frac{m}{n} - p\right| < \epsilon$$

will be greater than $1 - \eta$. The question naturally arises, whether for given ϵ and η a number $N(\epsilon, \eta)$ depending upon ϵ and η can be found such that the probability of *simultaneous* inequalities

$$\left|\frac{m}{n} - p\right| < \epsilon$$

for *all* $n > N(\epsilon, \eta)$ will still be greater than $1 - \eta$. The following theorem due to Cantelli shows that this question can be answered positively.

Cantelli's Theorem. *For given $\epsilon < 1$, $\eta < 1$ let N be an integer satisfying the inequality*

$$N > \frac{2}{\epsilon^2} \log \frac{4}{\epsilon^2 \eta} + 2$$

The probability that the relative frequencies of an event E will differ from p by less than ϵ in the Nth and all the following trials is greater than $1 - \eta$.

Proof. We shall prove first that the probability Q_n of the inequality

$$\left|\frac{m}{n} - p\right| \geqq \epsilon$$

will always be less than $2e^{-\frac{1}{4}n\epsilon^2}$. According to results proved in the preceding section for any $\eta > 0$

$$Q_n < \eta$$

if

$$n > \frac{1+\epsilon}{\epsilon^2}\log\frac{1}{\eta} + \frac{1}{\epsilon}.$$

This inequality, if we take $\eta = 2e^{-\frac{1}{4}n\epsilon^2}$ becomes

$$n > \frac{1+\epsilon}{2}n + \frac{1}{\epsilon} - \frac{1+\epsilon}{\epsilon^2}\log 2$$

and in this form it is evident, since for $\epsilon < 1$

$$1 - \frac{1+\epsilon}{\epsilon}\log 2 < 1 - 2\log 2 < 0.$$

Hence, as stated,

$$Q_n < 2e^{-\frac{1}{4}n\epsilon^2}. \tag{8}$$

The event A, in which we are interested, consists in simultaneous fulfillment of *all* the inequalities

$$\left|\frac{m}{n} - p\right| < \epsilon$$

for $n = N, N + 1, N + 2, \ldots$. The opposite event B consists in the fulfillment of at *least* one of the inequalities

$$\left|\frac{m}{n} - p\right| \geqq \epsilon$$

where n can coincide either with N, or with $N + 1$, or with $N + 2, \ldots$. The probability of B, which we shall denote by R, certainly does not exceed the sum of the probabilities of *all* the inequalities

$$\left|\frac{m}{n} - p\right| \geqq \epsilon$$

for $n = N, N + 1, N + 2, \ldots$.

Consequently, referring to (8),

$$R < 2\sum_{n=N}^{\infty} e^{-\frac{1}{4}n\epsilon^2} = \frac{2e^{-\frac{1}{4}N\epsilon^2}}{1 - e^{-\frac{1}{4}\epsilon^2}}.$$

To satisfy the inequality

$$\frac{2e^{-\frac{1}{2}N\epsilon^2}}{1 - e^{-\frac{1}{2}\epsilon^2}} < \eta$$

it suffices to take

$$N > \frac{2}{\epsilon^2} \log \frac{2}{\eta} + \frac{2}{\epsilon^2} \log \frac{1}{1 - e^{-\frac{1}{2}\epsilon^2}}.$$

Now

$$\frac{2}{\epsilon^2} \log \frac{1}{1 - e^{-\frac{1}{2}\epsilon^2}} < \frac{2}{\epsilon^2} \log \frac{2}{\epsilon^2} + 2.$$

Consequently, if

$$N \geqq \frac{2}{\epsilon^2} \log \frac{4}{\epsilon^2 \eta} + 2$$

we shall have $R < \eta$ and at the same time the probability of A will be greater than $1 - \eta$, which proves Cantelli's theorem.

SIGNIFICANCE OF BERNOULLI'S THEOREM

3. As was indicated in the Introduction, one of the most important problems in the theory of probability consists in the discovery of cases where the probability is very near to 0 or, on the contrary, very near to 1, because cases with very small or very "great" probability may have real practical interest. In Bernoulli's theorem we have a case of this kind; the theorem shows that with the probability approaching as near to 1 or certainty as we please, we may expect that in a sufficiently long series of independent trials with constant probability, the relative frequency of an event will differ from that probability by less than any specified number, no matter how small. But it lies in the nature of the idea of mathematical probability, that when it is near 1, or, on the contrary, very small, we may consider an event with such probability as practically certain in the first case, and almost impossible in the second. The reason is purely empirical.

To illustrate what we mean, let us consider an indefinite series of independent trials, in which the probability of a certain event remains constantly equal to $\frac{1}{2}$. It can be shown that if the number of trials is, for instance, 40,000 or more, we may expect with a probability > 0.999 that the relative frequency of the event will differ from $\frac{1}{2}$ by less than 0.01. In other words, we are entitled to bet at least 999 against 1 that the actual number of occurrences will lie between the limits $0.49n$ and $0.51n$ if $n \geqq 40{,}000$. If we could make a positive statement of this kind without any mention of probability, we should be offering an ideal scientific prediction. However, our knowledge in this case is incomplete and all we are entitled to state is this: we are more sure to be right in predicting the above limits for the number of occurrences than in expecting to draw a white ball from an urn containing 999 white and only 1 black ball.

In practical matters, where our actions almost never can be directed with perfect confidence, even incomplete knowledge may be taken as a sure guide. Whoever has tried to win on a single ticket out of 10,000 knows from experience that it is virtually impossible. Now the conviction of impossibility would be still greater if one tried to win on a single ticket out of 1,000,000.

In the light of such examples, we understand what value may be attached to statements derived from Bernoulli's theorem: Although the fact we expect is not bound to happen, the probability of its happening is so great that it may really be considered as certain. Once in a great while facts may happen contrary to our expectations, but such rare exceptions cannot outweigh the advantages in everyday life of following the indications of Bernoulli's theorem. And herein lies its immense practical value and the justification of a science like the theory of probability.

It should, however, be borne in mind that little, if any, value can be attached to practical applications of Bernoulli's theorem, unless the conditions presupposed in this theorem are at least approximately fulfilled: independence of trials and constant probability of an event for every trial. And in questions of application it is not easy to be sure whether one is entitled to make use of Bernoulli's theorem; consequently, it is too often used illegitimately.

It is easy to understand how essential it is to discover propositions of the same character under more general conditions, paying especial attention to the possible dependence of trials. There have been valuable achievements in this direction. In the proper place, we shall discuss the more important generalizations of Bernoulli's theorem.

4. When the probability of an event in a single experiment is known, Bernoulli's theorem may serve as a guide to indicate approximately how often this event can be expected to occur if the same experiments are repeated a considerable number of times under nearly the same conditions. When, on the contrary, the probability of an event is unknown and the number of experiments is very large, the relative frequency of that event may be taken as an approximate value of its probability. Bernoulli himself, in establishing his theorem, had in mind the approximate evaluation of unknown probabilities from repeated experiments. That is evident from his explanations preceding the statement of the theorem itself and its proof. Inasmuch as these explanations are interesting in themselves, and present the original thoughts of the great discoverer, we deem it advisable here to give a free translation from Bernoulli's book. After calling attention to the fact that only in a few cases can probabilities be found a priori, Bernoulli proceeds as follows:

So, for example, the number of cases for dice is known. Evidently there are as many cases for each die as there are faces, and all these cases have an equal chance to materialize. For, by virtue of the similitude of faces and the uniform distribution of weight in a die, there is no reason why one face should show up more readily than another, as there would be if the faces had a different shape or if one part of a die were made of heavier material than another. So one knows the number of cases when a white or a black ticket can be drawn from an urn, and besides, it is known that all these cases are equally possible, because the numbers of tickets of both kinds are determined and known, and there is no apparent reason why one of these tickets could be drawn more readily than any other. But, I ask you, who among mortals will ever be able to define as so many cases, the number, e.g., of the diseases which invade innumerable parts of the human body at any age and can cause our death? And who can say how much more easily one disease than another--plague than dropsy, dropsy than fever— can kill a man, to enable us to make conjectures about the future state of life or death? Who, again, can register the innumerable cases of changes to which the air is subject daily, to derive therefrom conjectures as to what will be its state after a month or even after a year? Again, who has sufficient knowledge of the nature of the human mind or of the admirable structure of our body to be able, in games depending on acuteness of mind or agility of body, to enumerate cases in which one or another of the participants will win? Since such and similar things depend upon completely hidden causes, which, besides, by reason of the innumerable variety of combinations will forever escape our efforts to detect them, it would plainly be an insane attempt to get any knowledge in this fashion.

However, there is another way to obtain what we want. And what is impossible to get a priori, at least can be found a posteriori; that is, by registering the results of observations performed a great many times. Because it must be presumed that something may occur or not occur as many times as it had previously been observed to occur or not occur under similar conditions. For instance, if, in the past, 300 men of the same age and physical build as Titus is now, were investigated, and it were found that 200 of them had died within a decade, the others continuing to enjoy life past this term, one could pretty safely conclude that there are twice as many cases for Titus to pay his debt to nature within the next decade than to survive beyond this term. So it is, if somebody for many preceding years had observed the weather and noticed how many times it was fair or rainy; or if somebody attended games played by two persons a great many times and noticed how often one or the other won; by these very observations he would be able to discover the ratio of cases which in the future might favor the occurrence or failure of the same event under similar circumstances.

And this empirical way of determining the number of cases by experiments is neither new nor unusual. For the author of the book "Ars cogitandi," a man of great acumen and ingenuity, in Chap. 12 recommends a similar procedure, and everybody does the same in daily practice. Moreover, it cannot be concealed that for reasoning in this fashion about some event, it is not sufficient to make a few experiments, but a great quantity of experiments is required; because even the most stupid ones by some natural instinct and without any previous instruction (which is rather remarkable) know that the more experiments are made, the less is the danger to miss the scope.

Although this is naturally known to anyone, the proof based on scientific principles is by no means trivial, and it is our duty now to explain it. However, I would consider it a small achievement if I could only prove what everybody knows anyway. There remains something else to be considered, which perhaps nobody has even thought of. Namely, it remains to inquire, whether by thus augmenting the number of experiments the probability of getting a genuine ratio between numbers of cases, in which some event may occur or fail, also augments itself in such a manner as finally to surpass any given degree of certitude; or whether the problem, so to speak, has its own asymptote; that is, there exists a degree of certitude which never can be surpassed no matter how the observations are multiplied; for instance, that it never is possible to have a probability greater than $1/2$, $2/3$, or $3/4$ that the real ratio has been attained. To illustrate this by an example, suppose that, without your knowledge, 3,000 white stones and 2,000 black stones are concealed in a certain urn, and you try to discover their numbers by drawing one stone after another (each time putting back the stone drawn before taking the next one, in order not to change the number of stones in the urn) and notice how often a white or a black stone appears. The question is, can you make so many drawings as to make it 10, or 100, or 1,000, etc., times more probable (that is, morally certain) that the ratio of frequencies of white and black stones will be 3 to 2, as is the case with the number of stones in the urn, than any other ratio different from that? If this were not true, I confess nothing would be left of our attempt to explore the number of cases by experiments. But if this can be attained and moral certitude can finally be acquired (how that can be done I shall show in the next chapter), we shall have cases enumerated a posteriori with almost the same confidence as if they were known a priori. And that, for practical purposes, where "morally certain" is taken for "absolutely certain" by Axiom 9, Chap. II, is abundantly sufficient to direct our conjectures in any contingent matter not less scientifically than in games of chance.

For if instead of an urn we take the air or the human body, that contain in themselves sources of various changes or diseases as the urn contains stones, we shall be able in the same manner to determine by observations how much more likely one event is to happen than another in these subjects.

To avoid misunderstanding, one must bear in mind that the ratio of cases which we want to determine by experiments should not be taken in the sense of a precise and indivisible ratio (for then just the contrary would happen, and the probability of attaining a true ratio would diminish with the increasing number of observations) but as an approximate one; that is, within two limits, which, however, can be taken as near as we wish to each other. For instance, if, in the case of the stones, we take pairs of ratios $301/200$ and $299/200$ or $3001/2000$ and $2999/2000$, etc., it can be shown that it will be more probable than any degree of probability that the ratio found in experiments will fall within these limits than outside of them. Such, therefore, is the problem which we have decided to publish here, now that we have struggled with it for about twenty years. The novelty of this problem as well as its great utility, combined with equal difficulty, may add to the weight and value of other parts of this doctrine.—"Ars Conjectandi," pars quarta, Cap. IV, pp. 224–227.

Application to Games of Chance

5. One of the cases in which the conditions for application of Bernoulli's theorem are fulfilled is that of games of chance. It is not out of place to discuss the question of the commercial values of games from the standpoint of Bernoulli's theorem. "Game of chance" is the term we apply to any enterprise which may give us profit or may cause us loss, depending on chance, the probabilities of gain or loss being known. The following considerations can be applied, therefore, to more serious questions and not only to games played for pastime or for the sake of gaining money, as in gambling.

Suppose that, by the conditions of the game, a player can win a certain sum a of money, with the probability p; or can lose another sum b with the probability $q = 1 - p$.

If this game can be repeated any number of times under the same conditions, the question arises as to the probability for a player to gain or lose a sum of money not below a given limit. Let us denote by n the total number of games, and by m the number of times the player wins. Considering a loss as a negative gain, his total gain will be

$$K = ma - (n - m)b.$$

It is convenient to introduce instead of m another number α defined by

$$\alpha = m - np$$

and called "discrepancy." Expressed in terms of α the preceding expression for the gain becomes

$$K = n(pa - qb) + (a + b)\alpha.$$

The expression

$$E = pa - qb$$

entering as the coefficient of n has, as we shall see, an important bearing on the conclusion as to the commercial value of the game. It is called the "mathematical expectation" of the player. Suppose at first that this expectation is *positive*. By Bernoulli's theorem the probability for a discrepancy less than $-n\epsilon$, ϵ being an arbitrary positive number, is smaller than any given number, provided, of course, the number of games is sufficiently large. At the same time, with the probability approaching 1 as near as we please, we may expect the discrepancy to be $\geqq -n\epsilon$. However, if this is the case, the total gain will surpass the number

$$n[E - \epsilon(a + b)]$$

which, for sufficiently large n, itself is greater than any specified positive number. It is supposed, of course, that ϵ is small enough to make the difference

$$E - \epsilon(a + b)$$

positive. And that means that the player whose mathematical expectation is positive may expect with a probability approaching certainty as near as we please to gain an arbitrarily large amount of money if nothing prevents him from playing a sufficient number of games.

On the contrary, by a similar argument, we can see that in case of a negative mathematical expectation, the player has an arbitrarily small probability to escape a loss of an arbitrarily large amount of money, again under the condition that he plays a sufficiently large number of games.

Finally, if the mathematical expectation is 0, it is impossible to make any definite statement concerning the gain or loss by the player, except that it is very unlikely that the amount of gain or loss will be considerable compared with the number of games.

It follows from this discussion that the game is certainly favorable for the player if his mathematical expectation is positive, and unfavorable if it is negative. In case the mathematical expectation is 0, neither of the parties participating in the game has a decided advantage and then the game is called equitable. Usually, games serving as amusements are equitable. On the contrary, all of the games operated for commercial purposes by individuals or corporations are expressly made to be profitable for the administration; that is, the mathematical expectation of the administration of a game operated for lucrative purposes is positive at each single turn of the game and, correspondingly, the expectation of any gambler is negative. This confirms the common observation that those gamblers who extend their gambling over large numbers of games are almost inevitably ruined. At the same time, the theory agrees with the fact that great profits are derived by the administrations of gaming places.

A good illustration is afforded by the French lottery mentioned on page 19, which, as is well known, was a very profitable enterprise operated by the French government. Now, if we consider the mathematical expectation of ticket holders in that lottery, we find that it was negative in all cases; namely, denoting by M the sum paid for tickets, we find the following expectations:

On 1 ticket $\left(\frac{15}{18} - 1\right)M = -\frac{1}{6}M,$

On 2 tickets $\left(\frac{540}{801} - 1\right)M = -\frac{29}{89}M,$

On 3 tickets $\left(\frac{5500}{11748} - 1\right)M = -\frac{1562}{2937}M,$

and so forth.

On the other hand, the expectation of the administration was always positive, and because of the great number of persons taking part in this lottery, the number of games played by the administration was enormous, and it was assured of a steady and considerable income. This was an enterprise avowedly operated for the purpose of gambling, but the same principles underlie the operations of institutions having great public value, such as insurance companies, which, to secure their income, always reserve certain advantages for themselves.

Experimental Verification of Bernoulli's Theorem

6. Bernoulli's theorem, like any other mathematical proposition, is a deduction from ideal premises. To what extent these premises may be considered as a good approximation to reality can be decided only by experiments. Several experiments established for the purpose of testing various theoretical statements derived from general propositions of the theory of probability, are reported by different authors. Here we shall discuss those purporting to test Bernoulli's theorem.

I. Buffon, the French naturalist of the eighteenth century, tossed a coin 4,040 times and obtained 2,048 heads and 1,992 tails. Assuming that his coin was ideal, we have a probability of $\frac{1}{2}$ for either heads or tails. Now, the relative frequencies obtained by his experiments are:

$$\tfrac{2048}{4040} = 0.507 \text{ for heads}$$
$$\tfrac{1992}{4040} = 0.493 \text{ for tails}$$

and they differ very little from the corresponding probabilities, 0.500. [illegible] the conclusions one might derive from Bernoulli's theorem are verified in a very satisfactory manner.

II. De Morgan, in his book "Budget of Paradoxes" (1872), reports the results of four similar experiments. In each of them a coin was tossed 2,048 times and the observed frequencies of heads were, respectively, 1,061, 1,048, 1,017, 1,039. The relative frequencies corresponding to these numbers are

$$\tfrac{1061}{2048} = 0.518; \qquad \tfrac{1048}{2048} = 0.512; \qquad \tfrac{1017}{2048} = 0.497; \qquad \tfrac{1039}{2048} = 0.507.$$

The agreement with the theory again is satisfactory.

III. Charlier, in his book "Grundzüge der mathematischen Statistik," reports the results of 10,000 drawings of one playing card out of a full deck. Each card drawn was returned to the deck before the next drawing. The actual result of these experiments was that black cards appeared 4,933 times, and consequently the frequency of red cards was 5,067. The relative frequencies in this instance are:

$$\tfrac{4933}{10000} = 0.4933 \text{ for a black card}$$
$$\tfrac{5067}{10000} = 0.5067 \text{ for a red card}$$

and they differ but slightly from the probability, 0.5000, that the card drawn will be black or white. The agreement between theory and experiment in this case, too, is satisfactory.

IV. The author of this book made the following experiment with playing cards: After excluding the 12 face cards from the pack, 4 cards were drawn at a time from the remaining 40, and the number of trials was carried to 7,000. The number of times in each thousand that the four cards belonged to different suits, was:

I	II	III	IV	V	VI	VII
113	113	103	105	105	118	108

Altogether the frequency of such cases was 765 in 7,000 trials, whence we find for the relative frequency

$$\frac{765}{7000} = 0.1093$$

while the probability for taking 4 cards belonging to different suits is

$$\frac{1000}{9139} = 0.1094.$$

V. In J. L. Coolidge's "Introduction to Mathematical Probability," one finds a reference to an experiment made by Lieutenant R. S. Hoar, U.S.A., but the reported results are incomplete. The author of this book repeated the same experiment which consisted in 1,000 drawings of 5 cards at a time, from a full pack of 52 cards. The results were: 503 times the 5 cards were each of different denominations; 436 times 2 were of the same denomination with 3 scattered; 45 times there were 2 pairs of 2 different denominations and 1 odd card; 14 times 3 were of the same with 2 scattered; 2 times there were 2 of one denomination and 3 of another. The remaining possible combination, 4 cards of the same denomination with 1 odd, never appeared. The probabilities of these different cases are, respectively,

$$\frac{2112}{4165} = 0.507; \qquad \frac{1760}{4165} = 0.423; \qquad \frac{198}{4165} = 0.048;$$
$$\frac{88}{4165} = 0.021; \qquad \frac{6}{4165} = 0.001; \qquad \frac{1}{4165} = 0.000.$$

The corresponding theoretical frequencies are 507, 423, 48, 21, 1, 0, while the observed frequencies were 503, 436, 45, 14, 2, 0. The discrepancies are generally small and the greatest of them, 13, is still within reasonable limits. Deeper investigation shows that the probability that a discrepancy will not exceed 13 is about ½; hence, the observed deviation of 13 units cannot be considered abnormal.

VI. Bancroft H. Brown published, in the *American Mathematical Monthly*, vol. 26, page 351, the results of a series of 9,900 games of craps. This game is played with two dice, and the caster wins unconditionally if he produces 7 or 11 points, which are called "naturals"; he loses the game in case of 2, 3, or 12 points, called "craps." But if he produces 4, 5, 6, 8, 9, or 10 "points," he does not win, but has the right to cast the dice an unlimited number of times until he throws the same number of points that he had before, or until he throws a 7. If he throws 7 before obtaining his point, he loses the game; otherwise he wins.

It is a good exercise to find the probability of winning this game. It is

$$\tfrac{244}{495} = 0.493$$

that is, a little less than ½. Multiplying the number of games, in our case 9,900, by this probability, we find that the theoretical number of successes is 4,880 and of failures, 5,020. Now, according to Bancroft H. Brown, the actual numbers of successes and losses are, respectively, 4,871 and 5,029. The discrepancy

$$4871 - 4880 = -9$$

is extremely small, even smaller than could reasonably be expected. The same article gives the number of times "craps" were produced; namely, 2 appeared 259 times, 3 appeared 508 times, and 12 appeared 293 times, making the total number of craps 1,060. The probability of obtaining craps is

$$\tfrac{1}{36} + \tfrac{1}{36} + \tfrac{2}{36} = \tfrac{1}{9}$$

hence, the theoretical number of craps should be 1,100. The discrepancy, $1060 - 1100 = -40$, is more considerable this time but still lies within reasonable limits.

VII. E. Czuber made a complete investigation of lotteries operated on the same plan as the French lottery, in Prague between 1754 and 1886, and in Brünn between 1771 and 1886. The number of drawings was 2,854 in Prague and 2,703 in Brünn. The probability that in each drawing the sequence of numbers is either increasing or decreasing, is

$$\tfrac{1}{60} = 0.01667$$

while the observed relative frequency of such cases was

Prague: 0.01612; Brünn: 0.01739

and in both places combined

0.01674.

The probabilities that among five numbers in each drawing there is none or only one of the numbers 1, 2, 3, . . . 9, are, respectively,

0.58298 and 0.34070.

The corresponding relative frequencies were

Prague: 0.58655 and 0.32656
Brünn: 0.57899 and 0.34591

and in both places combined

0.58183 and 0.33587, respectively.

The probability of drawing a determined number is $\frac{1}{18}$. Now, according to Czuber, for the lottery in Prague the actual number of occurrences for single tickets varied from 138 (for No. 6) to 189 (for No. 83), so that for all tickets the discrepancy varied from -20 to 31. Besides, there were only 16 numbers with a discrepancy greater than 15 in absolute value. All these results stand in good accord with the theory.

VIII. One of the most striking experimental tests of Bernoulli's theorem was made in connection with a problem considered for the first time by Buffon. A board is ruled with a series of equidistant parallel lines, and a very fine needle, which is shorter than the distance between lines, is thrown at random on the board. Denoting by l the length of the needle and by h the distance between lines, the probability that the needle will intersect one of the lines (the other possibility is that the needle will be completely contained within the strip between two lines) is found to be

$$p = \frac{2l}{\pi h}.$$

The remarkable thing about this expression is that it contains the number $\pi = 3.14159 \cdots$ expressing the ratio of the circumference of a circle to its diameter. In the appendix we shall indicate how this expression can be obtained, because in this problem we deal with a different concept of probability.

Suppose we throw the needle a great many times and count the number of times it cuts the lines. By Bernoulli's theorem we may expect that the relative frequency of intersections will not differ greatly from the theoretical probability, so that, equating them, we have the means of finding an approximate value of π.

One series of experiments of this kind was performed by R. Wolf, astronomer in Zurich, between 1849 and 1853. In his experiments the width of the strips was 45 mm., and the length of the needle was 36 mm. Thus the theoretical probability of intersections is

$$\frac{72}{45\pi} = 0.5093.$$

The needle was thrown 5,000 times and it cut the lines 2,532 times; whence, the relative frequency

$$\frac{2532}{5000} = 0.5064.$$

The agreement between the two numbers is very satisfactory. If, relying on Bernoulli's theorem, we set the approximate equation

$$\frac{72}{45\pi} = 0.5064,$$

we should find the number 3.1596 for π, which differs from the known value of π by less than 0.02.

In another experiment of the same kind reported by De Morgan in the aforementioned book, Ambrose Smith in 1855 made 3,204 trials with a needle the length of which was $\frac{3}{5}$ of the distance between lines. There were 1,213 clear intersections, and 11 contacts on which it was difficult to decide. If on this ground, we should consider half of them as intersections, we should obtain about 1,218 intersections in 3,204 trials, which would give the number 3.155 for π. If all of the contacts had been treated as intersections the result would have been 3.1412—very close to the real value of π.

In an excellent book "Calcolo delle Probabilità," vol. 1, page 183, 1925, by G. Castelnuovo, reference is made to experiments performed by Professor Reina under whose direction a needle of 3 cm. in length was thrown 2,520 times, the distance between lines being 6 cm. Taking into account the thickness of the needle, the probability of intersection was found to be 0.345, while actual experiments gave the relative frequency of intersections as 0.341.

APPENDIX

Buffon's Needle Problem. Let h be the width of the strip between two lines and $l < h$ the length of the needle. The position of the needle can be determined by the distance x of its middle point from the nearest line and the acute angle φ formed by the needle and a perpendicular dropped from the middle point to the line. It is apparent that x may vary from 0 to $h/2$ and φ varies within the limits 0 and $\pi/2$. We cannot define in the usual way the probability of the needle cutting the line, for there are infinitely many cases with respect to the position of the needle. However, it is possible to treat this problem as the limiting case of another problem with a finite number of possible cases, where the usual definition of probability can be applied.

Suppose that $h/2$ is divided into an arbitrary number m of equal parts $\delta = h/2m$ and the right angle $\pi/2$ into n equal parts $\omega = \pi/2n$. Suppose, further, that the distance x may have only the values

$$0, \delta, 2\delta, \ldots m\delta$$

and the angle φ the values

$$0,\ \omega,\ 2\omega,\ \ldots\ n\omega.$$

This gives

$$N = (m + 1)(n + 1)$$

cases as to the position of the needle, and it is reasonable to assume that these cases are equally likely. To find the number of favorable cases, we notice that the needle cuts one of the lines if x and φ satisfy the inequality

$$x < \frac{l}{2}\cos\varphi.$$

The number of favorable cases therefore, is equal to the number of systems of integers i, j satisfying the inequality

$$(A) \qquad i\delta < \frac{l}{2}\cos j\omega$$

supposing that i may assume only the values $0, 1, 2, \ldots m$ and j only the values $0, 1, 2, \ldots n$. Because we suppose $l < h$ the greatest value of i satisfying condition (A) is less than m and we can disregard the requirement that i should be $\leqq m$. Now for given j there are $k + 1$ values of i satisfying (A) if k denotes the greatest integer which is *less* than

$$\frac{l}{2\delta}\cos j\omega.$$

In other words, k is an integer determined by the conditions

$$k < \frac{l}{2\delta}\cos j\omega \leqq k + 1.$$

The number of possible values for i corresponding to a given j can therefore be represented thus

$$m_j = \frac{l}{2\delta}\cos j\omega + \vartheta_j$$

where ϑ_j may depend on j but for all j is $\geqq 0$ and <1. Taking the sum of all the m_j corresponding to $j = 0, 1, 2, \ldots n$, we obtain the number of favorable cases

$$M = \frac{l}{2\delta}(1 + \cos\omega + \cos 2\omega + \cdots + \cos n\omega) + n\Theta$$

where Θ again is a number satisfying the inequalities

$$0 \leqq \Theta < 1.$$

But, as is well known,

$$1 + \cos \omega + \cos 2\omega + \cdots + \cos n\omega = \frac{1}{2} + \frac{\sin (n + \frac{1}{2})\omega}{2 \sin \frac{\omega}{2}}$$

or, because $\omega = \dfrac{\pi}{2n}$

$$1 + \cos \omega + \cos 2\omega + \cdots + \cos n\omega = \frac{1}{2} + \frac{1}{2} \cot \frac{\omega}{2};$$

therefore

$$M = \frac{l}{4\delta} \cot \frac{\omega}{2} + \frac{l}{4\delta} + n\Theta.$$

Dividing this by $N = (m + 1)(n + 1)$ and substituting for δ and ω their expressions

$$\delta = \frac{h}{2m}, \qquad \omega = \frac{\pi}{2n}$$

we obtain the probability in the problem with a finite number of cases

$$\frac{M}{N} = \frac{l}{2h} \cdot \frac{m}{m + 1} \cdot \frac{\cot \frac{\pi}{4n}}{n + 1} + \frac{l}{2h} \cdot \frac{m}{m + 1} \cdot \frac{1}{n + 1} + \frac{n\Theta}{(n + 1)(m + 1)}.$$

The probability in Buffon's problem will be obtained by making m and n increase indefinitely in the above expression. Now, since

$$\lim \frac{m}{m + 1} = 1,$$

$$\lim \frac{m}{(m + 1)(n + 1)} = \lim \frac{n}{(n + 1)(m + 1)} = 0 \qquad (m, n \to \infty)$$

and

$$\lim \frac{\cot \frac{\pi}{4n}}{n + 1} = \frac{4}{\pi},$$

we have

$$\lim \frac{M}{N} = \frac{2l}{h\pi}.$$

Thus we arrive at the expression of probability

$$p = \frac{2l}{h\pi}$$

in Buffon's needle problem.

Problems for Solution

Another very simple proof of Bernoulli's theorem, due to Tshebysheff (1821–1894), is based upon the following considerations:

1. Prove the following identities:

$$\sum_{m=0}^{n} T_m(m - np) = 0, \qquad \sum_{m=0}^{n} T_m(m - np)^2 = npq.$$

Indication of the Proof. Differentiate the identity

$$e^{-npu}(pe^u + q)^n = \sum_{m=0}^{n} T_m e^{(m-np)u}$$

twice with respect to u and set $u = 0$.

2. If Q is the probability of the inequality $|m - np| \geqq n\epsilon$ prove that

$$Q < \frac{pq}{n\epsilon^2}.$$

Indication of the Proof. In the identity

$$\sum_{m=0}^{n} T_m(m - np)^2 = npq$$

drop all the terms in which $|m - np| < n\epsilon$ and in the remaining terms replace

$$(m - np)^2$$

by $n^2\epsilon^2$. The resulting inequality

$$\sum_{|m-np|\geqq n\epsilon} T_m < \frac{pq}{n\epsilon^2}$$

is equivalent to the statement.

3. Prove that

$$P > 1 - \eta$$

if $n > pq/\eta\epsilon^2$.

Indication of the Proof. $P = 1 - Q$, $Q < pq/n\epsilon^2$ and $pq/n\epsilon^2 < \eta$ if $n > pq/\eta\epsilon^2$.

The following two problems show how probability considerations can be used in proving purely analytical propositions.

4. *S. Bernstein's Proof of Weierstrass' Theorem.* The famous theorem due to Weierstrass states that for any continuous function $f(x)$ in a closed interval $a \leqq x \leqq b$ there exists a polynomial $P(x)$ such that

$$|f(x) - P(x)| < \sigma$$

for $a \leqq x \leqq b$ where σ is an arbitrary positive number. By a proper linear transformation the interval (a, b) can be transformed into the interval $(0, 1)$. According to S. Bernstein, the polynomial

$$P(x) = \sum_{m=0}^{n} C_n^m x^m (1 - x)^{n-m} f\left(\frac{m}{n}\right)$$

for sufficiently large n satisfies the inequality

$$|f(x) - P(x)| < \sigma$$

uniformly in the interval $0 \leqq x \leqq 1$.

Indication of the Proof. For $x = 0$ and $x = 1$ we have $f(0) = P(0)$ and

$$f(1) = P(1).$$

It suffices to prove the statement for $0 < x < 1$. Let x be a constant probability in n independent trials. We have

$$(a) \qquad f(x) - P(x) = \sum_{m=0}^{n} C_n^m x^m (1 - x)^{n-m}\left[f(x) - f\left(\frac{m}{n}\right)\right].$$

By the property of continuous functions, there is a number ϵ corresponding to any positive number σ such that

$$|f(x') - f(x)| < \frac{\sigma}{2}$$

whenever

$$|x' - x| < \epsilon \qquad (0 \leqq x', x \leqq 1).$$

Also, there exists a number M such that $|f(x)| \leqq M$ for $0 \leqq x \leqq 1$. From equation (a) we get

$$|f(x) - P(x)| \leqq \frac{\sigma}{2}P + 2MR$$

where P and R are, respectively, the probabilities of the inequalities

$$\left|\frac{m}{n} - x\right| < \epsilon \quad \text{and} \quad \left|\frac{m}{n} - x\right| \geqq \epsilon.$$

Now $P < 1$ and

$$R < \eta$$

if $n > 1/4\epsilon^2\eta$. Take $\eta = \sigma/4M$; then

$$|f(x) - P(x)| < \sigma$$

if

$$n > \frac{M}{\epsilon^2\sigma}.$$

5. Show that

$$\frac{\int_{\frac{m}{n}-\epsilon}^{\frac{m}{n}+\epsilon} x^m(1 - x)^{n-m}dx}{\int_0^1 x^m(1 - x)^{n-m}dx} > 1 - \frac{1}{2(n + 1)\epsilon^2}$$

provided $0 < m < n$ and $\frac{m}{n} - \epsilon > 0$, $\frac{m}{n} + \epsilon < 1$ (Castelnuovo).

Indication of the Proof. By Prob. 6, Chap. IV, page 72, the ratio

$$\frac{\int_0^{\frac{m}{n}-\epsilon} x^m(1-x)^{n-m}dx}{\int_0^1 x^m(1-x)^{n-m}dx}$$

represents the probability Q of at least $m + 1$ successes in a series of $n + 1$ independent trials with constant probability

$$p = \frac{m}{n} - \epsilon.$$

Set

$$m + 1 = (n + 1)p + (n + 1)\sigma$$

whence

$$\sigma = \frac{n - m}{n(n + 1)} + \epsilon > \epsilon.$$

But

$$Q < \frac{p(1 - p)}{(n + 1)\sigma^2} < \frac{1}{4(n + 1)\epsilon^2}.$$

Hence

$$\frac{\int_0^{\frac{m}{n}-\epsilon} x^m(1-x)^{n-m}dx}{\int_0^1 x^m(1-x)^{n-m}dx} < \frac{1}{4(n+1)\epsilon^2}$$

and by a similar argument

$$\frac{\int_{\frac{m}{n}+\epsilon}^1 x^m(1-x)^{n-m}dx}{\int_0^1 x^m(1-x)^{n-m}dx} < \frac{1}{4(n+1)\epsilon^2}.$$

References

JACOB BERNOULLI: "Ars Conjectandi," pars quarta, 1713.

P. L. TSHEBYSHEFF: "Sur les valeurs moyennes," Oeuvres, I, pp. 687–694.

F. P. CANTELLI: "Sulla probabilità come limite di frequenza," *Rend. d. R. Accad. Naz. dei Lincei*, 26, 1917.

A MARKOFF: "Calculus of Probability," 4th Russian ed., Leningrad, 1924.

CHAPTER VII

APPROXIMATE EVALUATION OF PROBABILITIES IN BERNOULLIAN CASE

1. In connection with Bernoulli's theorem, the following important question arises: when the number of trials is large, how can one find, at least approximately, the probability of the inequality

$$\left|\frac{m}{n} - p\right| \leqq \epsilon$$

where ϵ is a given number? Or, in a more general form: How can one find, approximately, the probability of the inequalities

$$l \leqq m \leqq l'$$

where l and l' are given integers, the number of trials n being large?

The exact formula for this probability is

$$P = \sum_{s=l}^{s=l'} T_s$$

where T_s, as before, represents the probability of s successes in n trials. While this formula cannot be of any practical use when n and $l' - l$ are large numbers, yet it is precisely such cases that present the greatest theoretical and practical interest. Hence, the problem naturally arises of substituting for the exact expression of P an approximate formula which will be easy to use in practice and which, for large n, will give a sufficiently close approximation to P. De Moivre was the first successfully to attack this difficult problem. After him, in essentially the same way, but using more powerful analytical tools, Laplace succeeded in establishing a simple approximate formula which is given in all books on probability.

When we use an approximate formula instead of an exact one, there is always this question to consider: How large is the committed error? If, as is usually done, this question is left unanswered, the derivation of Laplace's formula becomes an easy matter. However, to estimate the error comparatively long and detailed investigation is required. Except for its length, this investigation is not very difficult.

2. First we shall present the probability T_s in a convenient analytical form. The identity

$$F(t) = (pt + q)^n = T_0 + T_1t + T_2t^2 + \cdots + T_nt^n$$

after substituting $t = e^{i\varphi}$ becomes

$$F(e^{i\varphi}) = T_0 + T_1e^{i\varphi} + T_2e^{2i\varphi} + \cdots + T_ne^{ni\varphi}.$$

Multiplying it by $e^{-is\varphi}$ and integrating between $-\pi$ and π, we get

$$\int_{-\pi}^{\pi} e^{-is\varphi}F(e^{i\varphi})d\varphi = 2\pi T_s$$

because for an integral exponent k

$$\int_{-\pi}^{\pi} e^{ki\varphi}d\varphi = \begin{matrix} 0 & \text{if} & k \neq 0 \\ 2\pi & \text{if} & k = 0. \end{matrix}$$

Thus

$$T_s = \frac{1}{2\pi}\int_{-\pi}^{\pi} F(e^{i\varphi})e^{-si\varphi}d\varphi$$

and this is the expression for T_s suitable for our purposes. To find the sum

$$P = \sum_{s=l}^{s=l'} T_s$$

we observe first that

$$\sum_{s=l}^{s=l'} e^{-si\varphi} = \frac{e^{-il\varphi} - e^{-i(l'+1)\varphi}}{1 - e^{-i\varphi}} = e^{-\frac{l+l'}{2}i\varphi} \cdot \frac{\sin\left(\frac{l'-l+1}{2}\right)\varphi}{\sin\frac{\varphi}{2}}.$$

On the other hand, the complex number $F(e^{i\varphi})$ can be presented in trigonometrical form, thus:

$$F(e^{i\varphi}) = Re^{i\Theta}$$

whence

$$P = \frac{1}{2\pi}\int_{-\pi}^{\pi} Re^{i\left(\Theta - \frac{l+l'}{2}\varphi\right)} \frac{\sin\left(\frac{l'-l+1}{2}\right)\varphi}{\sin\frac{\varphi}{2}}d\varphi$$

or, because P is real,

$$P = \frac{1}{2\pi}\int_{-\pi}^{\pi} R\cos\left(\Theta - \frac{l+l'}{2}\varphi\right)\frac{\sin\left(\frac{l'-l+1}{2}\right)\varphi}{\sin\frac{\varphi}{2}}d\varphi.$$

Finally, because R is an even function of φ and Θ is an odd one, we can extend the integration over the interval $0, \pi$ on the condition that we double the result. Thus we obtain

$$P = \frac{1}{\pi}\int_0^\pi R\cos\left(\theta - \frac{l+l'}{2}\varphi\right)\frac{\sin\left(\frac{l'-l+1}{2}\right)\varphi}{\sin\frac{\varphi}{2}}d\varphi.$$

It is convenient to introduce instead of l and l' two numbers ζ_1 and ζ_2 defined by

$$l = np + \tfrac{1}{2} + \zeta_1\sqrt{B_n}, \qquad l' = np - \tfrac{1}{2} + \zeta_2\sqrt{B_n}$$

where $B_n = npq$. Setting further

$$\theta = np\varphi + \chi,$$

P can be presented as

$$P = P_2 - P_1$$

where P_1 and P_2 are obtained by taking $\zeta = \zeta_1$ and $\zeta = \zeta_2$ in the integral

(1) $$J = \frac{1}{2\pi}\int_0^\pi R\frac{\sin(\zeta\sqrt{B_n}\varphi - \chi)}{\sin\frac{1}{2}\varphi}d\varphi.$$

3. Our next aim is to establish upper and lower limits for R. Evidently

$$R = (p^2 + q^2 + 2pq\cos\varphi)^{\frac{n}{2}} = \left(1 - 4pq\sin^2\frac{\varphi}{2}\right)^{\frac{n}{2}} = \rho^n.$$

Now

$$\log\rho = \frac{1}{2}\log\left(1 - 4pq\sin^2\frac{\varphi}{2}\right) = -2pq\sin^2\frac{\varphi}{2} - \frac{1}{4}(4pq)^2\sin^4\frac{\varphi}{2} - \\ - \frac{1}{6}(4pq)^3\sin^6\frac{\varphi}{2} - \cdots$$

whence

$$\log\rho < -2pq\sin^2\frac{\varphi}{2}.$$

Since $\frac{1}{2}\varphi < \pi/2$, we have

$$\sin\frac{\varphi}{2} > \frac{\varphi}{\pi}$$

and consequently

$$\log\rho < -\frac{2pq}{\pi^2}\varphi^2$$

or

(2) $$\rho < e^{-\frac{2pq}{\pi^2}\varphi^2}$$

for all values of φ in the interval of integration. On the other hand, we have

$$\sin\frac{\varphi}{2} > \frac{\varphi}{2} - \frac{\varphi^3}{48} > 0 \qquad \text{for} \qquad \varphi^2 < 24$$

and

$$\sin^2\frac{\varphi}{2} > \frac{\varphi^2}{4} - \frac{\varphi^4}{48},$$

which gives another upper bound for ρ:

$$\rho < e^{-\frac{pq}{2}\varphi^2 + \frac{pq}{24}\varphi^4}. \tag{3}$$

The corresponding upper bounds for R will be

$$R < e^{-\frac{2B_n}{\pi^2}\varphi^2} \tag{4}$$

$$R < e^{-\frac{B_n}{2}\varphi^2 + \frac{B_n}{24}\varphi^4} \tag{5}$$

To find a lower bound for R we shall assume $\varphi \leqq \pi/2$. We can present $\log \rho$ thus:

$$\log \rho = -\frac{pq}{2}\varphi^2 - \frac{1}{4}(4pq)^2 \sin^4\frac{\varphi}{2} + 2pq\left\{\left(\frac{\varphi}{2}\right)^2 - \sin^2\frac{\varphi}{2}\right\} - \\ - \frac{1}{6}(4pq)^3 \sin^6\frac{\varphi}{2} - \cdots .$$

On the other hand,

$$\frac{1}{6}(4pq)^3 \sin^6\frac{\varphi}{2} + \frac{1}{8}(4pq)^4 \sin^8\frac{\varphi}{2} + \cdots < \frac{\frac{1}{6}(4pq)^3 \sin^6\frac{\varphi}{2}}{1 - 4pq\sin^2\frac{\varphi}{2}} < \frac{1}{3}(4pq)^3 \sin^6\frac{\varphi}{2}$$

and

$$\left(\frac{\varphi}{2}\right)^2 - \sin^2\frac{\varphi}{2} > \frac{1}{3}\sin^4\frac{\varphi}{2}$$

so that

$$2pq\left\{\left(\frac{\varphi}{2}\right)^2 - \sin^2\frac{\varphi}{2}\right\} - \frac{1}{6}(4pq)^3 \sin^6\frac{\varphi}{2} - \cdots > \frac{2pq}{3}\sin^4\frac{\varphi}{2} - \\ - \frac{1}{3}(4pq)^3 \sin^6\frac{\varphi}{2} = \frac{2pq}{3}\sin^4\frac{\varphi}{2}\left\{1 - 32p^2q^2\sin^2\frac{\varphi}{2}\right\} > 0$$

and consequently

$$\log \rho > -\frac{pq}{2}\varphi^2 - \frac{1}{4}(4pq)^2 \sin^4\frac{\varphi}{2} > -\frac{pq}{2}\varphi^2 - \frac{p^2q^2}{4}\varphi^4$$

if $\varphi \leqq \frac{\pi}{2}$. Hence,

(6) $$R > e^{-\frac{1}{2}B_n\varphi^2 - \frac{1}{4}pqB_n\varphi^4}$$

and this is valid for $\varphi \leqq \pi/2$.

4. Let τ be defined by

$$\tau^2 = 3B_n^{-\frac{1}{2}}.$$

Assuming $B_n \geqq 25$ from now on, we shall have,

$$\tau^2 \leqq \tfrac{3}{5}$$

and a fortiori $\tau < \pi/2$. Let us suppose now that φ varies in the interval $0 \leqq \varphi \leqq \tau$. By inequality (6) we shall have

$$R - e^{-\frac{1}{2}B_n\varphi^2} > e^{-\frac{1}{2}B_n\varphi^2}\left(e^{-\frac{1}{4}pqB_n\varphi^4} - 1\right) > -\frac{pq}{4}B_n\varphi^4 e^{-\frac{1}{2}B_n\varphi^2} > -\frac{1}{16}B_n\varphi^4 e^{-\frac{1}{2}B_n\varphi^2}$$

because $e^{-x} - 1 > -x$ for $x > 0$ and $pq \leqq \frac{1}{4}$.

On the other hand, using inequality (5), we find that

$$R - e^{-\frac{1}{2}B_n\varphi^2} < e^{-\frac{1}{2}B_n\varphi^2}\left\{e^{\frac{B_n}{24}\varphi^4} - 1\right\} < \frac{B_n}{24}\varphi^4 e^{-\frac{1}{2}B_n\varphi^2} \cdot e^{\frac{B_n}{24}\tau^4} < \frac{1}{16}B_n\varphi^4 e^{-\frac{1}{2}B_n\varphi^2}$$

since

$$\tfrac{1}{6}e^{\frac{B_n\tau^4}{24}} = \tfrac{1}{6}e^{\frac{3}{8}} < \tfrac{1}{4}.$$

From the two inequalities just established it follows that

(7) $$\left|R - e^{-\frac{1}{2}B_n\varphi^2}\right| < \tfrac{1}{16}B_n\varphi^4 e^{-\frac{1}{2}B_n\varphi^2}$$

in the interval

$$0 \leqq \varphi \leqq \tau.$$

5. We turn now to the angle Θ. Evidently

$$\Theta = n \text{ arc tg } \frac{p \sin \varphi}{q + p \cos \varphi} = n\omega$$

where

$$\omega = \text{arc tg } \frac{p \sin \varphi}{q + p \cos \varphi}.$$

By successive derivations with respect to φ we find

$$\frac{d\omega}{d\varphi} = \frac{p^2 + pq \cos \varphi}{p^2 + 2pq \cos \varphi + q^2}; \qquad \frac{d^2\omega}{d\varphi^2} = \frac{pq(p - q) \sin \varphi}{(p^2 + 2pq \cos \varphi + q^2)^2}$$

$$\frac{d^3\omega}{d\varphi^3} = pq(p - q)\frac{4pq + (1 - 2pq) \cos \varphi - 2pq \cos^2 \varphi}{(p^2 + 2pq \cos \varphi + q^2)^3}$$

$$\frac{d^4\omega}{d\varphi^4} = pq(p-q)\frac{\sin \varphi[-1 + 4pq+20p^2q^2+8pq(1-2pq) \cos \varphi - 4p^2q^2 \cos^2 \varphi]}{(p^2+2pq \cos \varphi+q^2)^4}$$

and for $\varphi = 0$

$$\left(\frac{d\omega}{d\varphi}\right)_0 = p, \qquad \left(\frac{d^2\omega}{d\varphi^2}\right)_0 = 0, \qquad \left(\frac{d^3\omega}{d\varphi^3}\right)_0 = pq(p - q).$$

Furthermore, one easily verifies that in the interval $0 \leqq \varphi \leqq \pi/2$

$$\left|\frac{d^3\omega}{d\varphi^3}\right| \leqq \frac{9}{8}pq|p - q|\left(1 - 4pq \sin^2 \frac{\varphi}{2}\right)^{-3}$$

$$\left|\frac{d^4\omega}{d\varphi^4}\right| \leqq 2pq|p - q|\left(1 - 4pq \sin^2 \frac{\varphi}{2}\right)^{-4} \varphi.$$

Hence, applying Taylor's formula and supposing $0 \leqq \varphi \leqq \tau$, we get for χ

(8) $$\chi = \tfrac{1}{6}B_n(p - q)\varphi^3 + M\varphi^5$$

where

(9) $$|M| < \tfrac{1}{12}B_n|p - q|(1 - pq\tau^2)^{-4},$$

or

(10) $$\chi = L\varphi^3$$

where

(11) $$|L| < \tfrac{3}{16}B_n|p - q|(1 - pq\tau^2)^{-3}.$$

Using inequalities (9) and (11), we easily find

(12) $$\sin (\zeta\sqrt{B_n}\varphi - \chi) = \sin (\zeta\sqrt{B_n}\varphi) - \tfrac{1}{6}B_n(p - q)\varphi^3 \cos (\zeta\sqrt{B_n}\varphi) + r$$

where

(13) $$|r| < \tfrac{1}{12}B_n|p - q|(1 - pq\tau^2)^{-4}\varphi^5 + \tfrac{9}{512}B_n^2(p - q)^2(1 - pq\tau^2)^{-6}\varphi^6,$$

provided $0 \leqq \varphi \leqq \tau$.

6. To find an appropriate expression of the integral J we split it into two integrals, J_1 and J_2, taken respectively between limits $0, \tau$ and τ, π. We have

$$|J_2| \leqq \frac{1}{2\pi}\int_\tau^\pi R\frac{d\varphi}{\sin \frac{\varphi}{2}} < \frac{1}{2}\int_\tau^\pi R\frac{d\varphi}{\varphi}$$

because $\sin\frac{\varphi}{2} > \frac{\varphi}{\pi}$. Let $\tau_1 = \tau\frac{\pi}{2}$, then by inequality (4)

$$\int_{\tau_1}^{\pi} R\frac{d\varphi}{\varphi} < \int_{\tau_1}^{\infty} \frac{e^{-\frac{2B_n}{\pi^2}\varphi^2}d\varphi}{\varphi} = \int_{\tau\sqrt{\frac{B_n}{2}}}^{\infty} \frac{e^{-u^2}du}{u}.$$

But for positive x the following inequality holds:

$$\int_x^{\infty} \frac{e^{-u^2}du}{u} < \frac{e^{-x^2}}{2x^2}; \tag{14}$$

consequently

$$\int_{\tau_1}^{\pi} R\frac{d\varphi}{\varphi} < \frac{e^{-\frac{1}{2}B_n\tau^2}}{B_n\tau^2} = \frac{e^{-\frac{3}{2}\sqrt{B_n}}}{3B_n^{\frac{1}{2}}}.$$

Noting that $R(\varphi)$ is a decreasing function of φ we have for $\tau \leqq \varphi \leqq \tau_1$

$$R(\varphi) \leqq R(\tau) < \tfrac{3}{2}e^{-\frac{3}{2}\sqrt{B_n}}.$$

Hence,

$$\int_{\tau}^{\tau_1} R\frac{d\varphi}{\varphi} < \frac{3}{2}\log\frac{\pi}{2}\,e^{-\frac{3}{2}\sqrt{B_n}},$$

and combining this inequality with the one previously established, we have finally

$$|J_2| < \left(\frac{3}{4}\log\frac{\pi}{2} + \frac{B_n^{-\frac{1}{2}}}{6}\right)e^{-\frac{3}{2}\sqrt{B_n}}. \tag{15}$$

7. More elaborate considerations are necessary to separate the principal term and to estimate the error term in J_1. Making use of the inequality

$$\left|\frac{1}{\sin x} - \frac{1}{x}\right| < \frac{x^2}{6\sin x}$$

we can present J_1 thus:

$$J_1 = \frac{2}{2\pi}\int_0^{\tau} R\frac{\sin(\zeta\sqrt{B_n}\varphi - \chi)}{\varphi}d\varphi + \Delta$$

where

$$|\Delta| < \frac{\tau}{48\pi\sin\frac{\tau}{2}}\int_0^{\tau} R\varphi d\varphi,$$

and, because $R < \frac{3}{2}e^{-\frac{1}{3}B_n\varphi^2}$ in the interval $0 < \varphi < \tau$

$$|\Delta| < \frac{\tau}{32\pi\sin\frac{\tau}{2}}B_n^{-1}.$$

Since $\tau^2 \leqq 3/5$ we find by direct numerical calculation

$$\frac{\tau}{32\pi \sin \frac{\tau}{2}} < 0.0205,$$

and so, finally,

$$|\Delta| < 0.0205 B_n^{-1}.$$

8. Referring now to inequality (7), we can write

$$\frac{2}{2\pi}\int_0^\tau R\,\frac{\sin(\zeta\sqrt{B_n}\varphi - \chi)}{\varphi}d\varphi = \frac{2}{2\pi}\int_0^\tau e^{-\frac{1}{2}B_n\varphi^2}\frac{\sin(\zeta\sqrt{B_n}\varphi - \chi)}{\varphi}d\varphi + \Delta_1$$

where

$$|\Delta_1| < \frac{B_n}{16\pi}\int_0^\infty e^{-\frac{1}{2}B_n\varphi^2}\varphi^3 d\varphi = \frac{B_n^{-1}}{8\pi} < 0.04 B_n^{-1}.$$

Combining this with the result of the preceding section, we can present J_1 thus

$$J_1 = \frac{2}{2\pi}\int_0^\tau e^{-\frac{1}{2}B_n\varphi^2}\frac{\sin(\zeta\sqrt{B_n}\varphi - \chi)}{\varphi}d\varphi + \Delta_2 \tag{16}$$

and

$$|\Delta_2| < 0.0605 B_n^{-1}.$$

9. To simplify the integral in the right member of (16), we substitute for $\sin(\zeta\sqrt{B_n}\varphi - \chi)$ its expression (12). Taking into account inequality (13), we get (17):

$$\frac{2}{2\pi}\int_0^\tau e^{-\frac{1}{2}B_n\varphi^2}\frac{\sin(\zeta\sqrt{B_n}\varphi - \chi)}{\varphi}d\varphi = \frac{2}{2\pi}\int_0^\tau e^{-\frac{1}{2}B_n\varphi^2}\frac{\sin(\zeta\sqrt{B_n}\varphi)}{\varphi}d\varphi -$$
$$- \frac{B_n}{6\pi}(p-q)\int_0^\tau e^{-\frac{B_n\varphi^2}{2}}\varphi^2\cos(\zeta\sqrt{B_n}\varphi)d\varphi + \Delta_3$$

where

$$|\Delta_3| < \frac{1}{12\pi}B_n|p-q|(1-pq\tau^2)^{-4}\int_0^\infty e^{-\frac{1}{2}B_n\varphi^2}\varphi^4 d\varphi +$$
$$+ \frac{9}{512\pi}B_n^2(p-q)^2(1-pq\tau^2)^{-6}\int_0^\infty e^{-\frac{1}{2}B_n\varphi^2}\varphi^6 d\varphi.$$

But

$$\int_0^\infty e^{-\frac{1}{2}B_n\varphi^2}\varphi^5 d\varphi = 8B_n^{-3}, \qquad \int_0^\infty e^{-\frac{1}{2}B_n\varphi^2}\varphi^4 d\varphi = 3\left(\frac{\pi}{2}\right)^{\frac{1}{2}} B_n^{-\frac{5}{2}}$$

and so

$$|\Delta_3| < \frac{1}{4\sqrt{2\pi}} B_n^{-\frac{3}{2}} |p - q| (1 - pq\tau^2)^{-4} + \frac{9}{64\pi}(p - q)^2(1 - pq\tau^2)^{-6} B_n^{-1}.$$

Now $pq \leqq \frac{1}{4}$, $\tau^2 \leqq \frac{3}{5}$, $B_n \geqq 25$, consequently

$$\frac{1}{4\sqrt{2\pi}} B_n^{-\frac{1}{2}}(1 - pq\tau^2)^{-4} \leqq \frac{1}{20\sqrt{2\pi}}\left(\frac{20}{17}\right)^4 < 0.0385.$$

On the other hand,

$$1 - pq\tau^2 \geqq 1 - \frac{3}{5}\left\{\left(\frac{p + q}{2}\right)^2 - \left(\frac{p - q}{2}\right)^2\right\} = \frac{17}{20} + \frac{3}{20}(p - q)^2,$$

and for positive x the maximum of

$$x\left(\tfrac{17}{20} + \tfrac{3}{20}x^2\right)^{-6}$$

is attained for $x^2 = \frac{17}{33}$, whence it follows that

$$\frac{9}{64\pi}|p - q|(1 - pq\tau^2)^{-6} \leqq \frac{9}{64\pi}\left(\frac{17}{33}\right)^{\frac{1}{2}}\left(\frac{55}{51}\right)^6 < 0.051.$$

Taking into account all this, we have

$$|\Delta_3| < 0.09|p - q|B_n^{-1}.$$

10. As to integrals in the right-hand member of (17) we can write

$$(18)\quad \frac{2}{2\pi}\int_0^\tau e^{-\frac{1}{2}B_n\varphi^2}\frac{\sin(\zeta\sqrt{B_n}\varphi)}{\varphi}d\varphi = \frac{2}{2\pi}\int_0^\infty e^{-\frac{1}{2}B_n\varphi^2}\frac{\sin(\zeta\sqrt{B_n}\varphi)}{\varphi}d\varphi + \Delta_4$$

$$(19)\quad \frac{B_n(p - q)}{6\pi}\int_0^\tau e^{-\frac{1}{2}B_n\varphi^2}\varphi^2\cos(\zeta\sqrt{B_n}\varphi)d\varphi =$$

$$= \frac{B_n(p - q)}{6\pi}\int_0^\infty e^{-\frac{1}{2}B_n\varphi^2}\varphi^2\cos(\zeta\sqrt{B_n}\varphi)d\varphi + \Delta_5$$

where

$$|\Delta_4| < \frac{1}{\pi}\int_\tau^\infty e^{-\frac{1}{2}B_n\varphi^2}\frac{d\varphi}{\varphi} < \frac{1}{3\pi}B^{-\frac{1}{2}}e^{-\frac{3}{2}\sqrt{B_n}}$$

and

$$|\Delta_5| < \frac{B_n}{6\pi}\cdot\frac{2^{\frac{3}{2}}}{B_n^{\frac{3}{2}}}\int_{\tau\sqrt{\frac{B_n}{2}}}^\infty e^{-u^2}u^2du < \frac{B_n^{-\frac{1}{2}}}{\pi\sqrt{3}}e^{-\frac{3}{2}\sqrt{B_n}}$$

because

$$\int_x^\infty e^{-u^2}u^2du < xe^{-x^2}$$

for $x > 1$, as can easily be proved. Finally, taking into account (15), (16), (17), (18), (19), we get

$$(20)\quad \left|J - \frac{2}{2\pi}\int_0^\infty e^{-\frac{1}{2}B_n\varphi^2}\frac{\sin(\zeta\sqrt{B_n}\varphi)}{\varphi}d\varphi + \right.$$
$$\left. + \frac{B_n(p-q)}{6\pi}\int_0^\infty e^{-\frac{1}{2}B_n\varphi^2}\varphi^2\cos(\zeta\sqrt{B_n}\varphi)d\varphi\right| < \frac{0.065 + 0.09|p-q|}{B_n} +$$
$$+ \left(\frac{3}{4}\log\frac{\pi}{2} + \frac{B_n^{-\frac{1}{2}}}{6} + \frac{B_n^{-\frac{1}{2}}}{3\pi} + \frac{B_n^{-\frac{1}{2}}}{\pi\sqrt{3}}\right)e^{-\frac{1}{2}\sqrt{B_n}} < \frac{0.065 + 0.09|p-q|}{B_n} + \frac{1}{2}e^{-\frac{1}{2}\sqrt{B_n}}$$

since for $B_n \geqq 25$

$$\frac{3}{4}\log\frac{\pi}{2} + \frac{B_n^{-\frac{1}{2}}}{6} + \frac{B_n^{-\frac{1}{2}}}{3\pi} + \frac{B_n^{-\frac{1}{2}}}{\pi\sqrt{3}} < \frac{1}{2}.$$

It now remains to evaluate definite integrals in (20). We have

$$(21)\quad \frac{2}{2\pi}\int_0^\infty e^{-\frac{1}{2}B_n\varphi^2}\frac{\sin(\zeta\sqrt{B_n}\varphi)}{\varphi}d\varphi = \frac{2}{2\pi}\int_0^\infty e^{-\frac{1}{2}u^2}\frac{\sin\zeta u}{u}du$$

$$(22)\quad \frac{B_n(p-q)}{6\pi}\int_0^\infty e^{-\frac{1}{2}B_n\varphi^2}\varphi^2\cos(\zeta\sqrt{B_n}\varphi)d\varphi =$$
$$= \frac{p-q}{6\pi\sqrt{B_n}}\int_0^\infty e^{-\frac{u^2}{2}}u^2\cos\zeta u\,du.$$

Differentiating the well-known integral

$$\int_0^\infty e^{-ax^2}\cos bx\,dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}e^{-\frac{b^2}{4a}} \qquad (a > 0)$$

twice with respect to b, and after that substituting $a = \frac{1}{2}$, $b = \zeta$, we find for (22) this expression:

$$\frac{p-q}{6\sqrt{2\pi B_n}}(1 - \zeta^2)e^{-\frac{1}{2}\zeta^2}.$$

On the other hand, an integral of the type

$$L(a) = \int_0^\infty e^{-\frac{1}{2}u^2}\frac{\sin au}{u}du$$

can be reduced to a so-called "probability integral." In fact, the derivation with respect to a gives

$$L'(a) = \int_0^\infty e^{-\frac{1}{2}u^2}\cos au\,du = \frac{1}{2}\sqrt{2\pi}e^{-\frac{a^2}{2}}$$

and since $L(0) = 0$,

$$L(a) = \frac{1}{2}\sqrt{2\pi}\int_0^a e^{-\frac{1}{2}u^2}du.$$

Consequently, integral (21) can be reduced to

$$\frac{1}{\sqrt{2\pi}}\int_0^{\zeta} e^{-\frac{1}{2}u^2}du.$$

Having found an approximate expression of the integral J after substituting in it ζ_2 and ζ_1 for ζ and taking the difference of the results, we find the desired expression of P.

11. The result of this long and detailed investigation can be summarized as follows:

Theorem. *Let m be the number of occurrences of an event in a series of n independent trials with the constant probability p. The probability P of the inequalities*

$$np + \tfrac{1}{2} + \zeta_1\sqrt{npq} \leqq m \leqq np - \tfrac{1}{2} + \zeta_2\sqrt{npq}$$

where extreme members are integers, can be represented in the form

$$(23)\quad P = \frac{1}{\sqrt{2\pi}}\int_{\zeta_1}^{\zeta_2} e^{-\frac{u^2}{2}}du + \frac{q-p}{6\sqrt{2\pi npq}}\left[(1-\zeta_2^2)e^{-\frac{\zeta_2^2}{2}} - (1-\zeta_1^2)e^{-\frac{\zeta_1^2}{2}}\right] + \omega.$$

The error term ω satisfies the inequality

$$|\omega| < \frac{0.13 + 0.18|p-q|}{npq} + e^{-\frac{3}{2}\sqrt{npq}}$$

provided $npq \geqq 25$.

By slightly increasing the limit of the error term, this theorem can be put into more convenient form. Let t_1 and t_2 be two arbitrary real numbers and let P denote the probability of the inequalities

$$np + t_1\sqrt{npq} \leqq m \leqq np + t_2\sqrt{npq}.$$

If the greatest integers contained in

$$np + t_2\sqrt{npq} \qquad \text{and} \qquad nq - t_1\sqrt{npq}$$

are respectively, A_2 and A_1, the preceding inequalities are equivalent to

$$n - A_1 \leqq m \leqq A_2.$$

To apply the theorem, we set

$$np - \tfrac{1}{2} + \zeta_2\sqrt{npq} = A_2 = np + t_2\sqrt{npq} - \theta_2$$
$$np + \tfrac{1}{2} + \zeta_1\sqrt{npq} = n - A_1 = np + t_1\sqrt{npq} + \theta_1$$

θ_2 and θ_1 being, respectively, the fractional parts of $np + t_2\sqrt{npq}$ and $nq - t_1\sqrt{npq}$. Hence,

$$\zeta_2 = t_2 + \frac{\frac{1}{2} - \theta_2}{\sqrt{npq}}$$
$$\zeta_1 = t_1 - \frac{\frac{1}{2} - \theta_1}{\sqrt{npq}}.$$

Applying Taylor's formula, it is easy to verify that

$$\left|\frac{1}{\sqrt{2\pi}}\int_{\zeta_1}^{\zeta_2} e^{-\frac{u^2}{2}}du - \frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2} e^{-\frac{u^2}{2}}du - \frac{(\frac{1}{2}-\theta_1)e^{-\frac{t_1^2}{2}} + (\frac{1}{2}-\theta_2)e^{-\frac{t_2^2}{2}}}{\sqrt{2\pi npq}}\right| < \frac{0.061}{npq}$$

$$\left|\frac{q-p}{6\sqrt{2\pi npq}}\left[(1-\zeta_2^2)e^{-\frac{\zeta_2^2}{2}} - (1-\zeta_1^2)e^{-\frac{\zeta_1^2}{2}}\right] - \frac{q-p}{6\sqrt{2\pi npq}}\left[(1-t_2^2)e^{-\frac{t_2^2}{2}} - \right.\right.$$
$$\left.\left. - (1-t_1^2)e^{-\frac{t_1^2}{2}}\right]\right| < \frac{0.069|p-q|}{npq}$$

whence, finally, we can draw the following conclusion: For any two real numbers t_1, t_2, the probability of the inequalities

$$t_1\sqrt{npq} \leqq m - np \leqq t_2\sqrt{npq}$$

can be expressed as follows:

$$P = \frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2} e^{-\frac{1}{2}u^2}du + \frac{(\frac{1}{2}-\theta_1)e^{-\frac{1}{2}t_1^2} + (\frac{1}{2}-\theta_2)e^{-\frac{1}{2}t_2^2}}{\sqrt{2\pi npq}} +$$
$$+ \frac{q-p}{6\sqrt{2\pi npq}}[(1-t_2^2)e^{-\frac{1}{2}t_2^2} - (1-t_1^2)e^{-\frac{1}{2}t_1^2}] + \Omega$$

where θ_2 and θ_1 are the respective fractional parts of

$$np + t_2\sqrt{npq} \qquad \text{and} \qquad nq - t_1\sqrt{npq}$$

and

$$|\Omega| < \frac{0.20 + 0.25|p-q|}{npq} + e^{-\frac{3}{2}\sqrt{npq}}$$

provided $npq \geqq 25$.

In particular, if $t_2 = -t_1 = t$, the probability of the inequality

$$|m - np| \leqq t\sqrt{npq}$$

is expressed by

$$P = \frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{1}{2}u^2}du + \frac{1-\theta_1-\theta_2}{\sqrt{2\pi npq}}e^{-\frac{1}{2}t^2} + \Omega$$

with the same upper limit for Ω. Laplace, supposing that $np + t\sqrt{npq}$ is an integer in which case $\theta_2 = 0$ and θ_1 is a fraction less than $(npq)^{-\frac{1}{2}}$, gives for P the approximate expression

$$P = \frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{1}{2}u^2}du + \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi npq}}$$

without indicating the limit of the error. Evidently Laplace's formula coincides with the formula obtained here by a rigorous analysis, save for terms of the same order as the error term Ω.

To find an approximate expression for the probability P of the inequality

$$\left|\frac{m}{n} - p\right| \leqq \epsilon$$

it suffices to take

$$t = \epsilon\sqrt{\frac{n}{pq}}.$$

Then

$$P = \frac{2}{\sqrt{2\pi}}\int_0^{\epsilon\sqrt{\frac{n}{pq}}} e^{-\frac{1}{2}u^2}du + \frac{1 - \theta_1 - \theta_2}{\sqrt{2\pi npq}}e^{-\frac{n\epsilon^2}{2pq}} + \Omega$$

and evidently P tends to 1 as n increases indefinitely. This is the second proof of Bernoulli's theorem.

Referring to the above expression for the probability of the inequalities

$$t_1\sqrt{npq} \leqq m - np \leqq t_2\sqrt{npq}$$

and supposing that the number of trials n increases indefinitely while t_1 and t_2 remain fixed, we immediately perceive the truth of the following limit theorem: *The probability of the inequalities*

$$t_1 \leqq \frac{m - np}{\sqrt{npq}} \leqq t_2$$

tends to the limit

$$\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2} e^{-\frac{1}{2}u^2}du$$

as n tends to infinity.

This limit theorem is a very particular case of an extremely general theorem which we shall consider in Chap. XIV.

12. To form an idea of the accuracy to be expected by using the foregoing approximate formulas, it is worth while to take up a few numerical examples. Let $n = 200$, $p = q = \frac{1}{2}$ and

$$95 \leqq m \leqq 105.$$

The exact expression of the probability that m will satisfy these inequalities is

$$P = \frac{200!}{100!100!}2^{-200}\left[1 + 2\left(\frac{100}{101} + \frac{100 \cdot 99}{101 \cdot 102} + \frac{100 \cdot 99 \cdot 98}{101 \cdot 102 \cdot 103} + \right.\right.$$
$$\left.\left. + \frac{100 \cdot 99 \cdot 98 \cdot 97}{101 \cdot 102 \cdot 103 \cdot 104} + \frac{100 \cdot 99 \cdot 98 \cdot 97 \cdot 96}{101 \cdot 102 \cdot 103 \cdot 104 \cdot 105}\right)\right].$$

The number in the brackets is found to be 9.995776 and its logarithm to five decimals

$$0.99982.$$

The logarithm of the first factor, again to five decimals, is

$$\bar{2}.75088,$$

whence

$$\log P = \bar{1}.75070; \qquad P = 0.56325,$$

and this value may be regarded as correct to five decimals. Let us see now what result is obtained by using approximate formulas. In our example

$$t\sqrt{npq} = t\sqrt{50} = 5; \qquad t = \frac{1}{\sqrt{2}} = 0.707107$$

and

$$\frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{u^2}{2}}du = 0.52050.$$

The additional term

$$\frac{e^{-0.25}}{\sqrt{100\pi}} = 0.04394$$

and by Laplace's formula

$$P = 0.56444.$$

This is greater than the true value of P by 0.00119. Now, the theoretical limit of the error is nearly

$$\tfrac{1}{250} = 0.004$$

so that, actually, Laplace's formula gives an even closer approximation than can be expected theoretically.

When npq is large, the second term in Laplace's formula ordinarily is omitted and the probability is computed by using a simpler expression:

$$P = \frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{u^2}{2}}du.$$

In our case this expression would give

$$P = 0.52050$$

instead of 0.56325 with the error about 0.043, which amounts to about 8 per cent of the exact number. Such a comparatively large error is explained by the fact that in our example $npq = 50$ is not large enough. In practice, when npq attains a few hundreds, the simplified expression for

P can be used when an accuracy of about two or three decimals is considered as satisfactory. In general, the larger t is, the better approximation can be expected.

For the second example, let us evaluate the probability that in 6,520 trials the relative frequency of an event with the probability $p = 3/5$ will differ from that probability by less than $\epsilon = 1/50$. To find t, we have the equation

$$t\sqrt{npq} = \epsilon n$$

where

$$n = 6520, \qquad p = \tfrac{3}{5}, \qquad q = \tfrac{2}{5}, \qquad \epsilon = \tfrac{1}{50},$$

which gives

$$t = \frac{130.4}{\sqrt{1564.8}} = 3.2965,$$

and, correspondingly,

$$\frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{u^2}{2}}du = 0.999021.$$

Since m satisfies the inequalities

$$3912 - 130.4 \leqq m \leqq 3912 + 130.4$$

the fractions θ_1 and θ_2 are $\theta_1 = \theta_2 = 0.4$ and the additional term is

$$\frac{0.2}{\sqrt{3129.6\pi}}e^{-5.4334} = 0.000009.$$

Hence, the approximate value of P is

$$P = 0.999030.$$

To judge what is the error, we can apply Markoff's method of continued fractions to find the limits between which P lies. These limits are

$$0.999028 \quad \text{and} \quad 0.999044.$$

The result obtained by using an approximate formula is unusually good, which can be explained by the fact that in our example t is a rather large number. Even the simplified formula gives 0.999021, very near the true value.

Finally, let us apply our formulas to the solution of the inverse problem: How large should the number of trials be to secure a probability larger than a given fraction for the inequality

$$\left|\frac{m}{n} - p\right| \leqq \epsilon?$$

Let us take, for example, $p = \frac{1}{3}$, $\epsilon = 0.01$ and the lower limit of probability 0.999. To find n approximately, we first determine t by the equation

$$\frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{u^2}{2}}du = 0.999,$$

which gives

$$t = 3.291.$$

Hence,

$$n = \frac{pqt^2}{\epsilon^2} = \frac{20000}{9}(3.291)^2 = 24{,}066, \text{ approximately.}$$

We cannot be sure that this limit is precise, since an approximate formula was used. But it can serve as an indication that for n exceeding this limit by a comparatively small amount, the probability in question will be >0.999. For instance, let us take $n = 24{,}300$. The limits for m being

$$8{,}100 - 243 \leqq m \leqq 8{,}100 + 243,$$

we find t from the equation

$$t = \epsilon\sqrt{\frac{n}{pq}} = 3.3068$$

and correspondingly

$$\frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{u^2}{2}}du = 0.999057.$$

The additional term in Laplace's formula being 0.000023, we find

$$P > 0.99908 - 0.00006 > 0.999.$$

Thus, 24,300 trials surely satisfy all the requirements.

Problems for Solution

1. Find approximately the probability that the number of successes will be contained between 2,910 and 3,090 in 9,000 independent trials with constant probability $\frac{1}{3}$. *Ans.* 0.9570 with an error in absolute value $<10^{-4}$ [using (23)].

2. In Buffon's experiment a coin was tossed 4,040 times, with the result that heads turned up 2,048 times. What would be the probability of having more than 2,050 or less than 1,990 heads? *Ans.* 0.337.

3. R. Wolf threw a pair of dice 100,000 times and noted that 83,533 times the numbers of points on the two dice were different. What is the probability of having such an event occur not less than 83,533 and not more than 83,133 times? Does the result suggest a doubt that for each die the probability of any number of points was $\frac{1}{6}$? *Ans.* This probability is approximately 0.0898 and on account of its smallness some doubt may exist.

4. If the probability of an event E is $\frac{1}{2}$, what number of trials guarantees a probability of more than 0.999 that the difference between the relative frequency of E and $\frac{1}{2}$ will be in absolute value less than 0.01? *Ans.* 27,500.

5. If a man plays 10,000 equitable games, staking \$1 in each game, what is the probability that the increase or decrease in his fortune will not exceed \$20 or \$50?
Ans. (a) 0.166; (b) 0.390.

6. If a man plays 100,000 games of craps and stakes 50 cents in each game, what is the probability that he will lose less than \$300? *Ans.* About $\frac{1}{200}$.

7. Following the method developed in this chapter, prove the following formula for the probability of exactly m successes in n independent trials with constant probability p:

$$T_m = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{t^2}{2}} \left[1 + \frac{(q-p)(t^3 - 3t)}{6\sqrt{npq}} \right] + \Delta$$

where t is determined by the equation

$$m = np + t\sqrt{npq}$$

and

$$|\Delta| < \frac{0.15 + 0.25|p-q|}{(npq)^{\frac{3}{2}}} + e^{-\frac{3}{2}\sqrt{npq}}$$

provided $npq \geqq 25$.

8. Developments of this chapter can be greatly simplified if $p = q = \frac{1}{2}$ (symmetrical case). In this case one can prove the following statement: The probability of the inequalities

$$\frac{n}{2} + \frac{1}{2} + \zeta_1 \sqrt{\frac{n}{4}} \leqq m \leqq \frac{n}{2} - \frac{1}{2} + \zeta_2 \sqrt{\frac{n}{4}}$$

can be expressed as follows:

$$P = \frac{1}{\sqrt{2\pi}} \int_{\zeta_1}^{\zeta_2} e^{-\frac{1}{2}u^2} du + \frac{(\zeta_2^3 - \zeta_2)e^{-\frac{1}{2}\zeta_2^2} - (\zeta_1^3 - \zeta_1)e^{-\frac{1}{2}\zeta_1^2}}{12\sqrt{2\pi n}} + \Delta$$

where $|\Delta| < 1/2n^2$ for $n > 16$.

9. In case of "rare" events, the probability p may be so small that even for a large number of trials the quantity $\lambda = np$ may be small; for example, 10 or less. In cases of this kind, approximation formulas of the type of Laplace's cannot be used with confidence. To meet such cases, Poisson proposed approximate formulas of a different character. Let P_m represent the probability that in n trials an event with the probability p will occur not more than m times. Show that

$$P_m = e^{-\lambda} \left[1 + \frac{\lambda}{1} + \frac{\lambda^2}{1 \cdot 2} + \cdots + \frac{\lambda^m}{1 \cdot 2 \cdot 3 \cdots m} \right] + \Delta = Q_m + \Delta$$

where

$$|\Delta| < (e^{\chi} - 1)Q_m \quad \text{if} \quad Q_m \geqq \tfrac{1}{2}$$
$$|\Delta| < (e^{\chi} - 1)(1 - Q_m) \quad \text{if} \quad Q_m < \tfrac{1}{2}$$

and

$$\chi = \frac{\lambda + \frac{1}{4} + \frac{\lambda^3}{n}}{2(n - \lambda)}.$$

Indication of the Proof. We have

$$P_m = q^n\left[1 + \frac{\lambda}{q} + \frac{1 - \frac{1}{n}}{1 \cdot 2}\frac{\lambda^2}{q^2} + \cdots + \frac{\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{m-1}{n}\right)}{1 \cdot 2 \cdot 3 \cdots m}\frac{\lambda^m}{q^m}\right].$$

Now, since $q = 1 - \frac{\lambda}{n}$

$$\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{i-1}{n}\right)q^{-i} = \frac{\left(1 - \frac{0}{n}\right)\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{i-1}{n}\right)}{\left(1 - \frac{\lambda}{n}\right)\left(1 - \frac{\lambda}{n}\right)\cdots\left(1 - \frac{\lambda}{n}\right)} =$$

$$= \prod_{k=0}^{i-1}\left(1 + \frac{\lambda - k}{n - \lambda}\right)$$

and

$$\prod_{k=0}^{i-1}\left(1 + \frac{\lambda - k}{n - \lambda}\right) \leqq \prod_{k=0}^{[\lambda]}\left(1 + \frac{\lambda - k}{n - \lambda}\right) < e^{\sum_{k=0}^{[\lambda]}\frac{\lambda - k}{n - \lambda}} \leqq e^{\frac{(\lambda + \frac{1}{2})^2}{2(n-\lambda)}}.$$

Consequently

$$P_m < \left(1 - \frac{\lambda}{n}\right)^n e^{\frac{(\lambda+\frac{1}{2})^2}{2(n-\lambda)}}\left[1 + \frac{\lambda}{1} + \frac{\lambda^2}{1 \cdot 2} + \cdots + \frac{\lambda^m}{1 \cdot 2 \cdot 3 \cdots m}\right].$$

But

$$\left(1 - \frac{\lambda}{n}\right)^n < e^{-\lambda - \frac{\lambda^2}{2n}},$$

whence

$$P_m < e^{\chi}Q_m; \qquad Q_m = e^{-\lambda}\left[1 + \frac{\lambda}{1} + \frac{\lambda^2}{1 \cdot 2} + \cdots + \frac{\lambda^m}{1 \cdot 2 \cdot 3 \cdots m}\right].$$

On the other hand,

$$1 - P_m = \sum_{\mu=m+1}^{n}\frac{n(n-1)\cdots(n-\mu+1)}{\mu!}q^{n-\mu}p^{\mu} =$$

$$= q^n\sum_{\mu=m+1}^{n}\frac{\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{\mu-1}{n}\right)}{q^{\mu}}\frac{\lambda^{\mu}}{1 \cdot 2 \cdot 3 \cdots \mu},$$

whence

$$1 - P_m < e^{\chi}(1 - Q_m)$$

and

$$P_m > e^{\chi}Q_m + 1 - e^{\chi}.$$

The final statement follows immediately from both inequalities obtained for P_m.

10. With the usual notation, show that

$$T_m = e^{-\lambda}\frac{\lambda^m}{m!}Q$$

where

$$Q = e^{\frac{m\lambda}{n} - \frac{(n-m)\lambda^2}{2n^2} - \frac{m(m-1)}{2n}}\left[1 - \theta\left(\frac{(n-m)\lambda^3}{3(n-\lambda)^3} + \frac{m^3}{2n(n-m)}\right)\right]; \qquad 0 < \theta < 1.$$

Indication of the Proof. Referring to Chap. I, page 23, we have

$$T_m < \frac{\lambda^m}{m!}\left(1 - \frac{\lambda}{n}\right)^{n-m}\left(1 - \frac{m}{2n}\right)^{m-1}$$

$$T_m > \frac{\lambda^m}{m!}\left(1 - \frac{\lambda}{n}\right)^{n-m}\left(1 - \frac{m}{n}\right)^{\frac{m-1}{2}}.$$

But

$$\left(1 - \frac{\lambda}{n}\right)^{n-m} < e^{-\lambda + \frac{m}{n}\lambda - \frac{(n-m)\lambda^2}{2n^2}}, \qquad \left(1 - \frac{m}{2n}\right)^{m-1} < e^{-\frac{m(m-1)}{2n}},$$

whence

$$T_m < \frac{\lambda^m}{m!}e^{-\lambda}\cdot e^{\frac{m\lambda}{n} - \frac{(n-m)\lambda^2}{2n^2} - \frac{m(m-1)}{2n}}.$$

On the other hand,

$$\left(1 - \frac{\lambda}{n}\right)^{n-m} = \left(1 + \frac{\lambda}{n-\lambda}\right)^{-(n-m)} > e^{-\lambda + \lambda\frac{m-\lambda}{n-\lambda} + \frac{(n-m)\lambda^2}{2(n-\lambda)^2} - \frac{(n-m)\lambda^3}{3(n-\lambda)^3}}$$

$$\left(1 - \frac{m}{n}\right)^{\frac{m-1}{2}} = \left(1 + \frac{m}{n-m}\right)^{-\frac{m-1}{2}} > e^{-\frac{m(m-1)}{2(n-m)}}.$$

Hence

$$\left(1 - \frac{\lambda}{n}\right)^{n-m}\left(1 - \frac{m}{n}\right)^{\frac{m-1}{2}} > e^{-\lambda + \frac{m\lambda}{n} - \frac{(n-m)\lambda^2}{2n^2} - \frac{m(m-1)}{2n}}\cdot e^{-\frac{(n-m)\lambda^3}{3(n-\lambda)^3} - \frac{m^3}{2n(n-m)}},$$

and a fortiori

$$\left(1 - \frac{\lambda}{n}\right)^{n-m}\left(1 - \frac{m}{n}\right)^{\frac{m-1}{2}} > e^{-\lambda + \frac{m\lambda}{n} - \frac{(n-m)\lambda^2}{2n^2} - \frac{m(m-1)}{2n}}\left[1 - \frac{(n-m)\lambda^3}{3(n-\lambda)^3} - \frac{m^3}{2n(n-m)}\right]$$

If λ and m are both small in comparison to n the above-introduced factor Q will be near 1. Under such circumstances we may be entitled to use an approximate formula due to Poisson

$$T_m = \frac{\lambda^m}{m!}e^{-\lambda}.$$

The preceding elementary analysis gives means to estimate the error incurred by using this formula.

11. Apply the preceding considerations to the case $n = 1{,}000$, $p = 1/100$, $\lambda = 10$ and $m = 10$. *Ans.* $0.1256 < T_{10} < 0.1258$. Poisson's formula gives 0.1251—a very good approximation. Alo, $0.5807 < P_{10} < 0.5863$. Taking $P_{10} = 0.583$, the error in absolute value will be less than $3.3 \cdot 10^{-3}$. By a more elaborate method it is found $P_{10} = 0.5830$.

References

De Moivre: "Miscellanea analytica," Lib. V, 1730.

De Moivre: "Doctrine of Chances," 3d ed., 1756.

Laplace: "Théorie analytique des probabilités," Oeuvres VII, pp. 280–285.

Ch. de la Vallée-Poussin: Demonstration nouvelle du théorème de Bernoulli, *Ann. Soc. Sci. Bruxelles*, 31.

D. Mirimanoff: Le jeu de pile ou face et les formules de Laplace et de J. Eggenberger, *Commentarii Mathematici Helvetici* 2, pp. 133–168.

CHAPTER VIII

FURTHER CONSIDERATIONS ON GAMES OF CHANCE

1. When a person undertakes to play a very large number of games under theoretically identical conditions, the inference to be drawn from Bernoulli's theorem is that that person will almost certainly be ruined if the mathematical expectation of his gain in a single game is negative. In case of a positive expectation, on the other hand, he is very likely to win as large a sum as he likes in a sufficiently long series of games. Finally, in an equitable game when the mathematical expectation of a gain is zero, the only inference to be drawn from Bernoulli's theorem is that his gain or loss will likely be small in comparison with the number of games played.

These conclusions are appropriate however, only if it is possible to continue the series of games indefinitely, with an agreement to postpone the final settling of accounts until the end of the series. But if the settlement, as in ordinary gambling, is made at the end of each game, it may happen that even playing a profitable game one will lose all his money and will have to discontinue playing long before the number of games becomes large enough to enable him to realize the advantages which continuation of the games would bring to him.

A whole series of new problems arises in this connection, known as problems on the duration of play or ruin of gamblers. Since the science of probability had its humble origin in computing chances of players in different games, the important question of the ruin of gamblers was discussed at a very early stage in the historical development of the theory of probability. The simplest problem of this kind was solved by Huygens, who in this field had such great successors as de Moivre, Lagrange, and Laplace.

2. It is natural to attack the problem first in its simplest aspect, and then to proceed to more involved and difficult questions.

Problem 1. Two players A and B play a series of games, the probability of winning a single game being p for A and q for B, and each game ends with a loss for one of them. If the loser after each game gives his adversary an amount representing a unit of money and the fortunes of A and B are measured by the whole numbers a and b, what is the probability that A (or B) will be ruined if no limit is set for the number of games?

Solution. It is necessary first to show how we can attach a definite numerical value to the probability of the ruin of A if no limit is set for the number of games. As in many similar cases (see, for instance, Prob. 15, page 41) we start by supposing that a limit *is* set. Let n be this limit. There is only a finite number of mutually exclusive ways in which A can be ruined in n games; either he can be ruined just after the first game, or just after the second, and so on. Denoting by $p_1, p_2, \dots p_n$ the probabilities for A to be ruined just after the first, second, . . . nth game, the probability of his ruin before or at the nth game is

$$p_1 + p_2 + \cdots + p_n.$$

Now, this sum being a probability, must remain <1 whatever n is. On the other hand, each term of this sum is $\geqq 0$ for the same reason. Both remarks combined, show that the series

$$p_1 + p_2 + p_3 + \cdots$$

is convergent. We take its sum as the probability for A to be ruined when nothing limits the number of games played. So it is clear that this probability, although unknown, possesses a perfectly determined numerical value. Let us denote by y_x the probability for A to be ruined when his fortune is x. The probability we seek is y_a. Obviously,

$$(1) \qquad y_0 = 1,$$

for A is certainly ruined if he has no money left. Similarly

$$(2) \qquad y_{a+b} = 0$$

because if the fortune of A is $a + b$, it means that B has no money wherewith to play, and certainly the ruin of A is then impossible. Further, considering the result of the game immediately following the situation in which the fortune of A amounted to x it is possible to establish an equation in finite differences which y_x satisfies. For, if A wins this game (the probability of which case is p), his fortune becomes $x + 1$ and the probability of being ruined later is y_{x+1}. By the theorem of compound probability, the probability of this case is py_{x+1}. But if A loses (the probability of which is q), his fortune becomes $x - 1$ and the probability that the one possessing this fortune will be ruined is y_{x-1}. The probability of this case is qy_{x-1}. Now, applying the theorem of total probability, we arrive at the equation

$$(3) \qquad y_x = py_{x+1} + qy_{x-1}.$$

This equation has a particular solution of the form a^x where a is a root of the equation

$$a = pa^2 + q.$$

If $p \neq q$ there are two roots

$$1, \frac{q}{p}$$

and, correspondingly, there are two distinct particular solutions of equation (3):

$$1 \text{ and } \left(\frac{q}{p}\right)^x.$$

Obviously,

$$y_x = C + D\left(\frac{q}{p}\right)^x$$

is also a solution of (3) for arbitrary C and D. Now, we can dispose of C and D so as to satisfy conditions (1) and (2). To this end we have the equations

$$C + D = 1$$
$$p^{a+b}C + q^{a+b}D = 0,$$

whence

$$C = \frac{q^{a+b}}{q^{a+b} - p^{a+b}}; \qquad D = -\frac{p^{a+b}}{q^{a+b} - p^{a+b}},$$

and

$$y_x = \frac{q^{a+b}p^x - p^{a+b}q^x}{p^x(q^{a+b} - p^{a+b})}.$$

It remains to take $x = a$ to obtain the required probability

$$y_a = \frac{q^a(q^b - p^b)}{q^{a+b} - p^{a+b}} = \frac{q^a(p^b - q^b)}{p^{a+b} - q^{a+b}}$$

that the player A possessing the fortune a will be ruined. Similarly, the probability of the ruin of B is

$$z_b = \frac{p^b(p^a - q^a)}{p^{a+b} - q^{a+b}}.$$

It turns out that

$$y_a + z_b = 1,$$

so that the probability that the series of games will continue indefinitely without A or B being ruined, is 0. The probability 0 does not show the impossibility of an eternal game, because this number was obtained, not by direct enumeration of cases, but by passage to the limit. Theoretically, an eternal game is not excluded. Actually, of course, this possibility can be disregarded.

If $p = q = \frac{1}{2}$, so that each single game is equitable, the preceding solution must be modified. In this case, the above quadratic equation in α has two coincident roots $= 1$, and we have only one particular solution of (3), $y_x = 1$. But another particular solution in this case is x, so that we can assume

$$y_x = C + Dx$$

and determine C and D from the equations

$$C = 1; \qquad C + D(a + b) = 0.$$

Thus, we find that

$$y_x = 1 - \frac{x}{a + b}$$

and for $x = a$

$$y_a = \frac{b}{a + b}.$$

Similarly, giving z_b the same meaning as above,

$$z_b = \frac{a}{a + b}.$$

If, therefore, each single game is equitable, the probabilities of ruin are inversely proportional to the fortunes of the players. The practical conclusion to be derived from this theoretical result is sheer common sense: It is unwise to play indefinitely with an adversary whose fortune is very large without submitting oneself to the great risk of losing all one's money in the course of the games, even if each single game is equitable. Gamblers who gamble at an even game with any willing individual are in the same condition as if they were gambling with an infinitely rich adversary. Their ruin in the long run is practically certain.

If single games of the series are not equitable, that is, $p \neq q$ the conclusion may be different. Supposing $p > q$, we have a case when the expectation of A is positive; in each single game, A has an advantage over his adversary. The above expression for y_a may be written in the form

$$y_a = \left(\frac{q}{p}\right)^a \frac{1 - \left(\frac{q}{p}\right)^b}{1 - \left(\frac{q}{p}\right)^{a+b}}$$

and, because $q/p < 1$, it is easy to see that y_a remains always less than

$$\left(\frac{q}{p}\right)^a$$

and converges to this number when b becomes infinite. Thus, playing a series of advantageous games even against an infinitely rich adversary, the probability of escaping ruin is

$$1 - \left(\frac{q}{p}\right)^a.$$

If a is large enough, this can be made as near 1 as we please, so that a player with a large fortune has good reason to believe that in the course of the games he will never be ruined, but that actually he is very likely to win a large sum of money.

This conclusion again is confirmed by experience. Big gambling institutions, like the Casino at Monte Carlo, always reserve certain advantages to themselves, and, although they are willing to play with practically everybody (as if they played against an infinitely rich adversary) the chance of their being ruined is slight because of the large capital in their possession.

3. In the problem solved above the stakes of both players were supposed to be equal, and we took them as units to measure the fortunes of both players. Next it would be interesting to investigate the case in which the stakes of A and B are unequal. An exact solution of this modified problem, since it depends on a difference equation of higher order, would be too complicated to be of practical use. It is therefore extremely interesting that, following an ingenious method developed by A. A. Markoff, one can establish simple inequalities for the required probabilities which give a good approximation if the fortunes of the players are large in comparison with their stakes.

Problem 2. If the conditions presupposed in Prob. 1 are modified, in that the stakes of A and B measured in a convenient unit are α and β and their respective fortunes are a and b, find the probabilities for A or B to be ruined in the sense that at a certain stage the capital of A will become less than α or that of B less than β.

Solution. Let y_x be the probability for A to be forced out of the game by the lack of sufficient money to set a full stake α when his fortune amounts to x and consequently that of his adversary is $a + b - x$. In the same way as before, we find that y_x is a solution of the equation in finite differences:

$$y_x = py_{x+\beta} + qy_{x-\alpha}. \tag{4}$$

To determine y_x completely, in addition to (4), we have two sets of supplementary conditions:

$$y_0 = y_1 = \cdots = y_{\alpha-1} = 1 \tag{5}$$

$$y_{a+b} = y_{a+b-1} = \cdots = y_{a+b-(\beta-1)} = 0. \tag{6}$$

Equation (5) expresses the fact that if the fortune of A becomes less than his stake, it is certain that A must quit. On the contrary, equation (6) indicates the impossibility for A to be ruined if the other player B does not have enough money to continue gaming. Equation (4) is an ordinary equation in finite differences of the order $\alpha + \beta$. It has particular solutions of the form θ^x where θ is a root of the equation

$$p\theta^{\alpha+\beta} - \theta^\alpha + q = 0. \tag{7}$$

The left-hand member for $\theta = 0$ is positive and with increasing θ decreases and attains a minimum when

$$p\theta^\beta = \frac{\alpha}{\alpha + \beta}$$

and then steadily increases and assumes positive values for large θ. This minimum must be negative or zero because $\theta = 1$ is a root of (7). Now, if it is negative, there are two positive roots of (7). One of them is $\theta = 1$ and another $>$ or <1 according as

$$p < \frac{\alpha}{\alpha + \beta} \qquad \text{or} \qquad p > \frac{\alpha}{\alpha + \beta}$$

or else

$$p\beta - q\alpha < 0 \qquad \text{or} \qquad > 0.$$

That is, the positive root of (7) different from 1 is >1 when single games are favorable to B and <1 if they are favorable to A. In case of equitable games, both positive roots coincide and $\theta = 1$ is a double root of (7). All the other roots of (7) are negative or imaginary.

The regular way to solve the problem would be to write down the general solution of (4) involving $\alpha + \beta$ arbitrary constants to be determined by conditions (5) and (6). As this method would lead to a complicated expression for y_x, we shall refrain from seeking the exact solution of our problem, and instead, following A. A. Markoff's ingenious remark, we shall establish simple lower and upper limits for y_x which are close enough if the fortunes of the players are large in comparison with their stakes.

Lemma. *If y_x is a solution of equation* (4) *and none of the numbers*

$$y_0, y_1, \ldots y_{\alpha-1}$$
$$y_{a+b}, y_{a+b-1}, \ldots y_{a+b-\beta+1}$$

is negative, then $y_x \geqq 0$ for $x = 0, 1, 2, \ldots a + b$.

Proof. Let $u_x^{(k)}$ $(k = 0, 1, 2, \ldots \alpha - 1)$ represent the probability that the player A whose actual fortune is x (and that of his adversary $a + b - x$) will be forced to quit when his fortune becomes exactly $= k$. Evidently $u_x^{(k)}$ is a solution of equation (4) satisfying the conditions

$u_x^{(k)} = 0$ for $x = 0, 1, \ldots k-1, k+1, \ldots \alpha - 1; a+b, a+b-1, \ldots a+b-\beta+1;$ $u_k^{(k)} = 1.$

Similarly, if $v_x^{(l)}(l = 0, 1, 2, \ldots \beta - 1)$ represents the probability that the player B will be forced to quit when the fortune of A becomes exactly $= a+b-l$, $v_x^{(l)}$ will be a solution of (4) satisfying the conditions

$v_x^{(l)} = 0$ for $x = 0, 1, 2, \ldots \alpha - 1; a+b, \ldots a+b-l+1, a+b-l-1, \ldots a+b-\beta+1;$ $v_{a+b-l}^{(l)} = 1.$

Thus we get $\alpha + \beta$ particular solutions of (4), and it is almost evident that these solutions are independent. Moreover, since they represent probabilities, $u_x^{(k)} \geqq 0$, $v_x^{(l)} \geqq 0$ for $x = 0, 1, 2, \ldots a+b$. Now, any solution y_x of (4) with given values of

$$y_0, y_1, \ldots y_{\alpha-1}$$
$$y_{a+b}, y_{a+b-1}, \ldots y_{a+b-\beta+1}$$

can be represented thus

$$y_x = \sum_{k=0}^{\alpha-1} y_k u_x^{(k)} + \sum_{l=0}^{\beta-1} y_{a+b-l} v_x^{(l)}.$$

Hence, $y_x \geqq 0$ for $x = 0, 1, 2, \ldots a+b$ if none of the numbers

$$y_0, y_1, \ldots y_{\alpha-1}$$
$$y_{a+b}, y_{a+b-1}, \ldots y_{a+b-\beta+1}$$

is negative. This interesting property of the solutions of equation (4) derived almost intuitively from the consideration of probabilities can be established directly. (See Prob. 9, page 160.)

The lemma just proved yields almost immediately the following proposition: If for any two solutions y_x' and y_x'' of equation (4) the inequality

$$y_x'' \geqq y_x'$$

holds for

$$x = 0, 1, 2, \ldots \alpha - 1; a+b, a+b-1, \ldots a+b-\beta+1,$$

the same inequality will be true for all $x = 0, 1, 2, \ldots a+b$. It suffices to notice that $y_x = y_x'' - y_x'$ is a solution of the linear equation (4) and, by hypothesis, $y_x \geqq 0$ for $x = 0, 1, 2, \ldots \alpha - 1;\ a+b, a+b-1, \ldots a+b-\beta+1$.

Now we can come back to our problem. First, if the mathematical expectation of A

$$p\beta - q\alpha$$

is different from 0, equation (7) has two positive roots: 1 and θ. With arbitrary constants C and D

$$y'_x = C + D\theta^x$$

is a solution of (4). Whatever C and D may be, y'_x as a function of x varies monotonically. Therefore, if C and D are determined by the conditions

$$y'_0 = 1, \qquad y'_{a+b-\beta+1} = 0$$

we shall have

$$\begin{aligned} y'_x &\leqq 1 \quad \text{if} \quad x = 0, 1, 2, \ldots \alpha - 1 \\ y'_x &\leqq 0 \quad \text{if} \quad x = a + b - \beta + 1, \ldots a + b \end{aligned}$$

and by the above established lemma, taking into account conditions (5) and (6), we shall have for the required probability the following inequality

$$y_x \geqq y'_x;$$

or, substituting the explicit expression for y'_x,

$$y_x \geqq \frac{\theta^{a+b-\beta+1} - \theta^x}{\theta^{a+b-\beta+1} - 1}.$$

If, on the contrary, C and D are determined by

$$y'_{\alpha-1} = 1, \qquad y'_{a+b} = 0$$

we shall have

$$\begin{aligned} y'_x &\geqq 1 \quad \text{if} \quad x = 0, 1, 2, \ldots \alpha - 1 \\ y'_x &\geqq 0 \quad \text{if} \quad x = a + b - \beta + 1, \ldots a + b \end{aligned}$$

and

$$y_x \leqq \frac{\theta^{a+b-\alpha+1} - \theta^{x-\alpha+1}}{\theta^{a+b-\alpha+1} - 1}.$$

Finally, taking $x = a$, we obtain the following limits for the initial probability y_a:

$$\theta^a \frac{\theta^{b-\beta+1} - 1}{\theta^{a+b-\beta+1} - 1} \leqq y_a \leqq \theta^{a-\alpha+1} \frac{\theta^b - 1}{\theta^{a+b-\alpha+1} - 1}.$$

They give a sufficient approximation to y_a if a and b are large compared with α and β.

If each single game is equitable, equation (4) has a solution with two arbitrary constants:

$$y'_x = C + Dx.$$

Proceeding in the same way as before, we obtain the inequalities

$$\frac{b - \beta + 1}{a + b - \beta + 1} \leqq y_a \leqq \frac{b}{a + b - \alpha + 1}.$$

4. To simplify the analysis, it was supposed that nothing limited the number of games played by A and B so that an eternal game, although extremely improbable, was theoretically possible. We now turn to problems in which the number of games is limited.

Problem 3. Players A and B agree to play not more than n games. The probabilities of winning a single game are p and q, respectively, and the stakes are equal. Taking these stakes as monetary units, the fortune of A is measured by the whole number a and that of B is infinite or at least so large that he cannot be ruined in n games. What is the probability for A to be ruined in the course of n games?

Solution. Let $y_{x,t}$ represent the probability for A to be ruined when his fortune is measured by the number x and he cannot play more than t games. The reasoning we have used several times shows that $y_{x,t}$ satisfies a partial equation in finite differences:

$$y_{x,t} = py_{x+1,t-1} + qy_{x-1,t-1}. \tag{8}$$

Moreover, if A has no money left, his ruin is certain, which gives the condition

$$y_{0,t} = 1 \qquad \text{if} \qquad t \geqq 0. \tag{9}$$

On the other hand, if A still possesses money and cannot play any more, his ruin is impossible, so that

$$y_{x,0} = 0 \qquad \text{if} \qquad x > 0. \tag{10}$$

Conditions (9) and (10) together with equation (8) determine $y_{x,t}$ completely for all positive values of x and t. To find an explicit expression for $y_{x,t}$ we shall use Lagrange's method. Equation (8) has particular solutions of the form

$$\alpha^x\beta^t$$

where α and β satisfy the relation

$$\alpha\beta = p\alpha^2 + q.$$

We can solve this equation either for β or for α which leads to two different expressions of $y_{x,t}$. Solving for β we have infinitely many particular solutions

$$\alpha^x(p\alpha + q\alpha^{-1})^t$$

with an arbitrary α and we can seek to obtain the required solution in the form

$$y_{x,t} = \frac{1}{2\pi i}\int_c \alpha^{x-1}(p\alpha + q\alpha^{-1})^t f(\alpha)d\alpha$$

where $f(\alpha)$ is supposed to be developable in Laurent's series on a certain circle c. To satisfy (10) we must have

$$\frac{1}{2\pi i}\int_c \alpha^{x-1}f(\alpha)d\alpha = 0 \qquad \text{for} \qquad x = 1, 2, 3, \ldots$$

which shows that $f(\alpha)$ is regular within the circle c. To determine $f(\alpha)$ completely, we must have, according to (9)

$$\frac{1}{2\pi i}\int_c (p\alpha + q\alpha^{-1})^t \frac{f(\alpha)}{\alpha}d\alpha = 1 \qquad \text{for} \qquad t = 0, 1, 2, \ldots .$$

All these equations are equivalent to a single equation

$$\frac{1}{2\pi i}\int_c \frac{f(\alpha)d\alpha}{\alpha - p\epsilon\alpha^2 - q\epsilon} = \frac{1}{1-\epsilon}$$

holding good for all sufficiently small ϵ. The integrand has a single pole α_0 within c defined by

$$\alpha_0 - p\epsilon\alpha_0^2 - q\epsilon = 0,$$

and the corresponding residue is

$$\frac{q + p\alpha_0^2}{q - p\alpha_0^2}f(\alpha_0).$$

But this must be equal to

$$\frac{1}{1-\epsilon}$$

or, substituting for ϵ its expression in α_0

$$\frac{q + p\alpha_0^2}{p\alpha_0^2 - \alpha_0 + q},$$

and hence for all sufficiently small α_0

$$f(\alpha_0) = \frac{q - p\alpha_0^2}{p\alpha_0^2 - \alpha_0 + q};$$

that is, if

$$f(\alpha) = \frac{q - p\alpha^2}{p\alpha^2 - \alpha + q}$$

all the requirements are satisfied. Taking into account that $p + q = 1$, we have

$$f(\alpha) = \frac{1}{1-\alpha} + \frac{p\alpha}{q - p\alpha},$$

and also

$$f(\alpha) = 1 + \sum_{n=1}^{\infty}\left[1 + \left(\frac{p}{q}\right)^n\right]\alpha^n.$$

The expression for $y_{x,t}$ is therefore

$$y_{x,t} = \frac{1}{2\pi i}\int_c \alpha^{x-1}(p\alpha + q\alpha^{-1})^t \sum_{n=0}^{\infty} c_n\alpha^n d\alpha$$

where $c_0 = 1$ and $c_n = 1 + (p/q)^n$ if $n \geqq 1$.

It remains to find the coefficient of $1/\alpha$ in the development of the integrand in a series of descending powers of α. Since

$$\alpha^{x-1}(p\alpha + q\alpha^{-1})^t = \sum_{l=0}^{t} C_t^l p^l q^{t-l} \alpha^{2l+x-t-1}$$

this coefficient is given by the sum

$$\sum_{l=0}^{\frac{t-x}{2}} C_t^l p^l q^{t-l} c_{t-x-2l}$$

extended over all integers l from 0 up to the greatest integer not exceeding $\frac{t-x}{2}$. Hence, the final expression for the probability $y_{a,n}$ is

(11) $$y_{a,n} = q^a \sum_{l=0}^{\frac{n-a}{2}} C_n^l (pq)^l [p^{n-a-2l} + q^{n-a-2l}]$$

with the agreement, in case of an even $n - a$, to replace the sum

$$p^0 + q^0$$

corresponding to $l = \frac{n-a}{2}$ by 1. It is natural that the right-hand member of the preceding expression should be replaced by 0 if $n < a$, which is in perfect agreement with the fact that A cannot be ruined in less than a games.

The second form of solution is obtained if we express α as a function of β. The equation

$$p\alpha^2 - \alpha\beta + q = 0$$

having two roots, we shall take for α the root

$$\alpha = \frac{\beta - \sqrt{\beta^2 - 4pq}}{2p}$$

determined by the condition that it vanishes for infinitely large positive β and can be developed in power series of $1/\beta$ when $|\beta| > 2\sqrt{pq}$. Using

α in this perfectly determined sense, it is easy to verify that

$$y_{x,t} = \frac{1}{2\pi i}\int_c\left(\frac{\beta - \sqrt{\beta^2 - 4pq}}{2p}\right)^x \frac{\beta^t}{\beta - 1}d\beta$$

where c is a circle of radius >1 described from 0 as its center, satisfies all the requirements. For it is a solution of equation (8). Next, for $x = 0$ and $t \geqq 0$,

$$y_{0,t} = \frac{1}{2\pi i}\int_c \beta^t\left(\frac{1}{\beta} + \frac{1}{\beta^2} + \cdots\right)d\beta = 1$$

and, finally, for $t = 0$ and $x > 0$

$$y_{x,0} = \frac{1}{2\pi i}\int_c\left(\frac{\beta - \sqrt{\beta^2 - 4pq}}{2p}\right)^x \frac{d\beta}{\beta - 1} = 0$$

because the development of the integrand into power series of $1/\beta$ starts at least with the second power of $1/\beta$.

To find $y_{x,t}$ in explicit form, it remains to find the coefficient of $1/\beta$ in the development of

$$\left(\frac{\beta - \sqrt{\beta^2 - 4pq}}{2p}\right)^x \frac{\beta^t}{\beta - 1}$$

in a series of descending powers of β. Let

$$\left(\frac{\beta - \sqrt{\beta^2 - 4pq}}{2p}\right)^x = \frac{l_x}{\beta^x} + \frac{l_{x+1}}{\beta^{x+1}} + \cdots;$$

multiplying this series by

$$\frac{\beta^t}{\beta - 1} = \beta^{t-1} + \beta^{t-2} + \cdots + \frac{1}{\beta} + \frac{1}{\beta^2} + \cdots$$

we find that the coefficient of $1/\beta$ in the product is

$$l_x + l_{x+1} + \cdots + l_t,$$

and hence

$$y_{x,t} = l_x + l_{x+1} + \cdots + l_t$$

provided $t \geqq x$, for otherwise $y_{x,t} = 0$. The quadratic equation in α can be written in the form

$$\alpha = \frac{1}{\beta}(q + p\alpha^2)$$

and the development of any power of its root vanishing for $\beta = \infty$ into power series of $1/\beta$ can be obtained by application of Lagrange's series. We have

$$\alpha^x = \sum_{n=x}^{\infty}\frac{x\beta^{-n}}{n!}\left[\frac{d^{n-1}(q + p\xi^2)^n\xi^{x-1}}{d\xi^{n-1}}\right]_{\xi=0},$$

but

$$\frac{1}{n!}\left[\frac{d^{n-1}(q+p\xi^2)^n\xi^{x-1}}{d\xi^{n-1}}\right]_{\xi=0}=\frac{(x+2i-1)!}{i!(x+i)!}q^{x+i}p^i$$

if $n = x + 2i$, and $= 0$ if $n = x + 2i + 1$. Hence,

$$l_{x+2i}=\frac{x(x+2i-1)!}{i!(x+i)!}q^{x+i}p^i$$
$$l_{x+2i+1}=0,$$

and finally

$$(12)\quad y_{a,n}=q^a\left[1+\frac{a}{1}pq+\frac{a(a+3)}{1\cdot 2}(pq)^2+\frac{a(a+4)(a+5)}{1\cdot 2\cdot 3}(pq)^3+\right.$$
$$\left.+\cdots+\frac{a(a+k+1)\cdots(a+2k-1)}{1\cdot 2\cdots k}(pq)^k\right]$$

where $k = \dfrac{n-a}{2}$ or $k = \dfrac{n-1-a}{2}$ according as n and a are of the same parity or not.

5. The difference $y_{a,n} - y_{a,n-1}$ gives the probability for the player A to be ruined at exactly the nth game and not before. Now, this difference is 0 if n differs from a by an odd number, so that the probability of ruin at the $(a + 2i - 1)$st game is 0. That is almost evident because after every game the fortune of A is increased or diminished by 1 and therefore can be reduced to 0 only if the number of games played is of the same parity as a. If $n = a + 2i$, the difference $y_{a,n} - y_{a,n-1}$ is

$$\frac{a(a+i+1)\cdots(a+2i-1)}{1\cdot 2\cdot 3\cdots i}q^{a+i}p^i.$$

Such, therefore, is the probability for A to be ruined at exactly the $(a + 2i)$th game. The remarkable simplicity of this expression obtained by means which are not quite elementary leads to a suspicion that it might also be obtained in a simple way. And, indeed, there is a simple way to arrive at this expression and thus to have a third, elementary, solution of Prob. 3.

Considering the possible results of a series of $a + 2i$ games, let A stand for a game won by A, and B for a game lost by A. The result of every series will thus be represented by a succession of letters A and B. We are interested in finding all the sequences which ruin A at exactly the last game. Because the fortune of A sinks from a to 0 there must be i letters A and $i + a$ letters B in every sequence we consider. Besides, there is another important condition. Let us imagine that the sequence is divided into two arbitrary parts, one containing the first letter and another the last letter of the sequence. Let x be the number of letters B, and y that of letters A in the second or right part of the sequence. There will be $a + i - x$ letters B and $i - y$ letters A in the first or left part. It means that the fortune of A after a game corresponding to the last letter in the left part, becomes

$$a + i - y - (a + i - x) = x - y$$

and since A cannot be ruined before the $(a + 2i)$th game, x must always be $>y$. That is, counting letters A and B from the right end of the sequence, the number of letters B must surpass the number of letters A at every stage. Conversely, if this condition is satisfied the succession represents a series of games resulting in the ruin of A at the end of the series and not before.

To find directly the number of sequences satisfying this requirement is not so easy, and it is much easier, following an ingenious method proposed by D. André, to find the number of all the remaining sequences of i letters A and $i + a$ letters B. These can be divided into two classes: those ending with A and those ending with B. Now, it is easy to show that there exists a one-to-one correspondence between successions of these two classes, so that both classes contain the same number of sequences. For, in a sequence of the second class (ending with B) starting from the right end, we necessarily find a shortest group of letters containing A and B in equal numbers. This group must end with A. Writing letters of this group in reverse order without changing the preceding letters, we obtain a sequence of the first class ending with A. Conversely, in a sequence of the first class there is a shortest group at the right end ending with B and containing an equal number of letters A and B. Writing letters of this group in reverse order, we obtain a sequence of the second class.

An example will illustrate the described manner of establishing the one-to-one correspondence between sequences of the first and of the second class. Consider a sequence of the first kind

$$B|BBABAA.$$

The vertical bar separates the shortest group from the right containing letters A and B in equal numbers. Reversing the order of letters in this group, we obtain a sequence of the second class

$$B|AABABB$$

and this sequence, by application of the above rule, is transformed again into the original sequence of the first class. The number of sequences of the first class can now be easily found. It is the same as the number of all possible sequences of $i - 1$ letters A and $a + i$ letters B, that is,

$$\frac{(a + 2i - 1)!}{(i - 1)!(a + i)!} = \frac{(a + i + 1)(a + i + 2) \cdots (a + 2i - 1)}{1 \cdot 2 \cdots (i - 1)}.$$

The total number of sequences in both classes is

$$2\frac{(a+i+1)(a+i+2)\cdots(a+2i-1)}{1\cdot 2\cdots(i-1)}.$$

Hence, the number of sequences leading to ruin of A in exactly $a + 2i$ games is

$$\frac{(a+i+1)(a+i+2)\cdots(a+2i)}{1\cdot 2\cdots i} - \\ - 2\frac{(a+i+1)(a+i+2)\cdots(a+2i-1)}{1\cdot 2\cdots(i-1)} = \\ = \frac{a(a+i+1)\cdots(a+2i-1)}{1\cdot 2\cdots i}.$$

As the probability of gains and losses indicated by every such sequence is the same, namely, $q^{a+i}p^i$ the probability of the ruin of A in exactly $a + 2i$ games is

$$\frac{a(a+i+1)\cdots(a+2i-1)}{1\cdot 2\cdot 3\cdots i}q^{a+i}p^i$$

and hence the second expression found for $y_{a,n}$ follows immediately.

The problem concerning the probability of ruin in the course of a prescribed number of games for a player playing against an infinitely rich adversary was first considered by de Moivre, who gave both the preceding solutions without proof; it was later solved completely by Lagrange and Laplace. The elementary treatment can be found in Bertrand's "Calcul des probabilités."

6. Formulas (11) and (12), though elegant and useful when n is not large, become impracticable when n is somewhat large, and that is precisely the most interesting case. Since the question of the risk of ruin incurred in playing equitable games possesses special interest, it would not be out of place at least to indicate here, though without proof, a convenient approximate expression for the probability $y_{a,n}$ in case of a large n and $p = q = \frac{1}{2}$. Let t be defined by

$$t = \frac{a}{\sqrt{2(n+\frac{2}{3})}};$$

then for $n \geqq 50$ it is possible to establish the approximate formula

$$y_{a,n} = 1 - \frac{2}{\sqrt{\pi}}\int_0^t e^{-z^2}dz + \frac{\theta}{6n}$$

where $-1 < \theta < 1$. Suppose, for instance, that the fortune of a player amounts to \$100, each stake being \$1, and he decides to play 1,000, 5,000, 10,000, 100,000, 1,000,000 games. Corresponding to these cases, we find

$$t = 2.2354, 0.9999, 0.7071, 0.2236, 0.0707$$

and hence

$$\frac{2}{\sqrt{\pi}}\int_0^t e^{-z^2}dz = 0.9984, 0.8427, 0.6827, 0.2482, 0.0796.$$

The corresponding approximate values of $y_{100,n}$ are

$$0.0016, 0.1573, 0.3173, 0.7518, 0.9204.$$

Thus, for a player possessing $100 there is very little risk of being ruined in the course of 1,000 games even if he stakes $1 at each game. The risk is considerably larger, but still fairly small, when 5,000 games are played. In 10,000 games we can bet 2 to 1 that the player will still be able to continue. But when the limit set for the number of games becomes 100,000, we can bet 3 to 1 that the player will be ruined somewhere in the course of those 100,000 games. Finally, there is little chance to escape ruin in a series of 1,000,000 games. The risk of ruin naturally increases with the number of games, but not so fast as might appear at first sight.

7. We conclude this chapter by solving the following problem, where the fortunes of both players are finite.

Problem 4. Players A and B agree to play not more than n games, the probabilities of winning a single game being p and q, respectively. Assuming that the fortunes of A and B amount to a and b single stakes which are equal for both, find the probability for A to be ruined in the course of n games.

Solution. Let $z_{x,t}$ be the probability for the player A to be ruined when his fortune is x (and that of his adversary $a + b - x$) and he can play only t games. Evidently $z_{x,t}$ satisfies the equation

$$z_{x,t} = pz_{x+1,t-1} + qz_{x-1,t-1} \tag{13}$$

perfectly similar to equation (8), but the complementary conditions serving to determine $z_{x,t}$ completely are different. First we have

$$z_{0,t} = 1 \quad \text{for} \quad t \geqq 0. \tag{14}$$

Next,

$$z_{a+b,t} = 0 \quad \text{for} \quad t \geqq 0, \tag{15}$$

because if A gets all the money from B, the games stop and A cannot be ruined. Finally,

$$z_{x,0} = 0 \quad \text{for} \quad x = 1, 2, 3, \ldots a + b - 1, \tag{16}$$

because A, having money left at the end of play, naturally cannot be ruined.

Since (13) has two series of particular solutions

$$\alpha^x\beta^t \text{ and } \alpha'^x\beta^t$$

where α and α' are roots of the equation

$$p\alpha^2 - \beta\alpha + q = 0$$

both developable into series of descending powers of β for $|\beta| > 1$, we shall seek $z_{x,t}$ in the form

$$z_{x,t} = \frac{1}{2\pi i}\int_c [f(\beta)\alpha^x + \varphi(\beta)\alpha'^x]\beta^t d\beta.$$

Here the integration is made along a circle of sufficiently large radius and $f(\beta)$ and $\varphi(\beta)$ are two unknown functions which can be developed into series of descending powers of β. Obviously $z_{x,t}$ satisfies (13) identically in x and t. For $x = 0$ and $t \geqq 0$ we have the condition

$$\frac{1}{2\pi i}\int_c [f(\beta) + \varphi(\beta)]\beta^t d\beta = 1; \qquad t = 0, 1, 2, \ldots$$

which is satisfied if

(17) $$f(\beta) + \varphi(\beta) = \frac{1}{\beta - 1}.$$

Condition (15) will be satisfied if

(18) $$\alpha^{a+b}f(\beta) + \alpha'^{a+b}\varphi(\beta) = 0$$

and it remains to show that at the same time (16) is satisfied. Solving (17) and (18), we have

$$f(\beta) = \frac{\alpha'^{a+b}}{\alpha'^{a+b} - \alpha^{a+b}} \cdot \frac{1}{\beta - 1}$$

$$\varphi(\beta) = \frac{-\alpha^{a+b}}{\alpha'^{a+b} - \alpha^{a+b}} \cdot \frac{1}{\beta - 1}$$

and

(19) $$f(\beta)\alpha^x + \varphi(\beta)\alpha'^x = \frac{\alpha'^{a+b}\alpha^x - \alpha^{a+b}\alpha'^x}{(\beta - 1)(\alpha'^{a+b} - \alpha^{a+b})} = $$

$$= \left(\frac{q}{p}\right)^x \frac{\alpha'^{a+b-x} - \alpha^{a+b-x}}{(\beta - 1)(\alpha'^{a+b} - \alpha^{a+b})}.$$

Now let α be the root vanishing for $\beta = \infty$ and α' the other root whose development in series of descending powers of β starts with the term containing β. Evidently the development of (19) for

$$x = 1, 2, 3, \ldots a + b - 1$$

does not contain terms involving the first power of $1/\beta$, and hence $z_{x,0} = 0$ if $x = 1, 2, 3, \ldots a + b - 1$ as it should be. The solution of (13) satisfying (14), (15), (16) being unique, its analytical expression is therefore

$$z_{x,t} = \frac{\left(\frac{q}{p}\right)^x}{2\pi i}\int_c \frac{\alpha'^{a+b-x} - \alpha^{a+b-x}}{\alpha'^{a+b} - \alpha^{a+b}}\frac{\beta^t d\beta}{\beta - 1},$$

whence for $x = a$ and $t = n$

$$z_{a,n} = \frac{\left(\frac{q}{p}\right)^a}{2\pi i}\int_c \frac{\alpha'^{b} - \alpha^{b}}{\alpha'^{a+b} - \alpha^{a+b}}\frac{\beta^n d\beta}{\beta - 1}.$$

To find an explicit expression for $z_{a,n}$ it remains to find the coefficient of $1/\beta$ in the development of

$$P = \left(\frac{q}{p}\right)^a \frac{\alpha'^{b} - \alpha^{b}}{\alpha'^{a+b} - \alpha^{a+b}}\frac{\beta^n}{\beta - 1}$$

in series of descending powers of β. This can be done in two different ways. First we can substitute for α' its expression in α:

$$\alpha' = \frac{q}{p}\alpha^{-1}$$

and present P in the form

$$P = \frac{\alpha^a - \left(\frac{p}{q}\right)^b \alpha^{a+2b}}{1 - \left(\frac{p}{q}\right)^{a+b}\alpha^{2a+2b}}\frac{\beta^n}{\beta - 1},$$

or developing into series

$$P = \left[\alpha^a - \left(\frac{p}{q}\right)^b \alpha^{a+2b} + \left(\frac{p}{q}\right)^{a+b}\alpha^{3a+2b} - \left(\frac{p}{q}\right)^{a+2b}\alpha^{3a+4b} + \cdots\right]\frac{\beta^n}{\beta - 1}.$$

But the coefficient of $1/\beta$ in

$$\frac{\alpha^m \beta^n}{\beta - 1}$$

by the second solution of Prob. 3 is the probability $y_{m,n}$ for a player with a fortune m to be ruined by an infinitely rich player in the course of n games. Hence, the final expression for $z_{a,n}$ is

$$z_{a,n} = y_{a,n} - \left(\frac{p}{q}\right)^b y_{a+2b,n} + \left(\frac{p}{q}\right)^{a+b} y_{3a+2b,n} - \left(\frac{p}{q}\right)^{a+2b} y_{3a+4b,n} + \cdots,$$

the terms of this series being alternately of the form

$$\left(\frac{p}{q}\right)^{ka+kb} y_{(2k+1)a+2kb,n}$$

and

$$-\left(\frac{p}{q}\right)^{ka+(k+1)b} y_{(2k+1)a+(2k+2)b,n}$$

for $k = 0, 1, 2, \ldots$. The series stops by itself as soon as the first subscript of $y_{x,n}$ becomes greater than n.

To obtain a second expression of $z_{a,n}$ we notice that

$$\frac{\alpha'^b - \alpha^b}{\alpha'^{a+b} - \alpha^{a+b}} = \frac{\alpha'^b - \alpha^b}{\alpha' - \alpha} \div \frac{\alpha'^{a+b} - \alpha^{a+b}}{\alpha' - \alpha} = Q \div R$$

is a rational function of β whose denominator

$$R = \frac{\alpha'^{a+b} - \alpha^{a+b}}{\alpha' - \alpha}$$

is a polynomial in β of the degree $a + b - 1$. To find the roots of $R = 0$, we set $\beta = 2\sqrt{pq} \cos \varphi$. Since, then,

$$\alpha' = \sqrt{\frac{q}{p}} e^{i\varphi}, \qquad \alpha = \sqrt{\frac{q}{p}} e^{-i\varphi},$$

we have

$$R = \left(\frac{q}{p}\right)^{\frac{a+b-1}{2}} \frac{\sin (a + b)\varphi}{\sin \varphi}.$$

The equation

$$\frac{\sin (a + b)\varphi}{\sin \varphi} = 0$$

having roots

$$\varphi_h = \frac{h\pi}{a + b}; \qquad h = 1, 2, \ldots a + b - 1,$$

the $a + b - 1$ roots of R are

$$\beta_h = 2\sqrt{pq} \cos \varphi_h.$$

Now we can resolve the rational function P into a sum of simple elements as follows:

$$P = E(\beta) + \frac{A_0}{\beta - 1} + \sum_{h=1}^{a+b-1} \frac{A_h}{\beta - \beta_h}$$

where

$$A_0 = \frac{q^a(p^b - q^b)}{p^{a+b} - q^{a+b}}$$

and for $h > 0$

$$A_h = -(2\sqrt{pq})^{n+1}\left(\frac{q}{p}\right)^{\frac{a}{2}}\frac{\sin\varphi_h}{(a+b)(1-2\sqrt{pq}\cos\varphi_h)}\sin a\varphi_h(\cos\varphi_h)^n$$

while $E(\beta)$ is the integral part of P. The coefficient of $1/\beta$ in the development of P being

$$A_0 + \sum_{h=1}^{a+b-1} A_h,$$

we have a new explicit expression for $z_{a,n}$:

$$(20)\qquad z_{a,n} = \frac{q^a(p^b - q^b)}{p^{a+b} - q^{a+b}} -$$

$$-\frac{(2\sqrt{pq})^{n+1}(qp^{-1})^{\frac{a}{2}}}{a+b}\sum_{h=1}^{a+b-1}\frac{\sin\dfrac{\pi h}{a+b}}{1 - 2\sqrt{pq}\cos\dfrac{\pi h}{a+b}}\sin\frac{\pi a h}{a+b}\left(\cos\frac{\pi h}{a+b}\right)^n.$$

This expression shows clearly that $z_{a,n}$, with increasing n, approaches the limit

$$z_{a,\infty} = \frac{q^a(p^b - q^b)}{p^{a+b} - q^{a+b}}$$

representing the probability of ruin when the number of games is unlimited, in complete accord with the solution of Prob. 1.

The first term in (20) naturally must be replaced by $\dfrac{b}{a+b}$ in case $p = q = \frac{1}{2}$. This form of solution was given first by Lagrange.

Problems for Solution

1. Players A and B with fortunes of \$50 and \$100, respectively, agree to play until one of them is ruined. The probabilities of winning a single game are $\frac{2}{3}$ and $\frac{1}{3}$, respectively, for A and B, and they stake \$1 at each game. What is the probability of ruin for the player A? *Ans.* Very nearly $2^{-50} = 8.88 \cdot 10^{-16}$.

2. If A and B at each single game stake \$3 and \$2, respectively, and have fortunes of \$30 and \$20 at the beginning, what is the approximate value of the probability that A will be ruined if the probability of his winning a single game is (*a*) $p = \frac{3}{5}$; (*b*) $p = \frac{1}{2}$?

Ans. (*a*) $0.40 + \Delta$; $|\Delta| < 1.7 \times 10^{-2}$; (*b*) $0.96 + \Delta$; $|\Delta| < 4.6 \times 10^{-3}$.

3. A player A with the fortune \$$a$ plays an unlimited number of games against an infinitely rich adversary with the probability p of winning a single game. He stakes \$1 at each game, while his rich adversary risks staking such a sum β as to make the game favorable to A. What is the probability that A will be ruined in the course of the games? Give numerical results if (a) $a = 10$, $p = \frac{1}{2}$, $\beta = 3$; (b) $a = 100$, $p = \frac{1}{2}$, $\beta = 3$. *Ans.* Let $\theta < 1$ be a positive root of the equation $p\theta^{\beta+1} - \theta + q = 0$. The required probability P is: $P = \theta^a$.

In case (a) $P = 0.002257$; in case (b) $P = 3.43 \cdot 10^{-27}$.

4. A player A whose fortune is \$10 agrees to play not more than 20 games against an infinitely rich adversary, both staking \$1 with an equal probability of winning a single game. What is the probability that A will not be ruined in the course of 20 games? *Ans.* 0.9734.

5. Players A and B with \$1 and \$2, respectively, agree to play not more than n equitable games, staking \$1 at each game. What are the probabilities of their ruin?

Ans. For A: $\frac{2}{3} - \frac{3 + (-1)^n}{3 \cdot 2^{n+1}}$; for B: $\frac{1}{3} - \frac{3 - (-1)^n}{3 \cdot 2^{n+1}}$.

6. Players A and B with \$2 and \$3, respectively, play a series of equitable games, both staking \$1 at each game. What are the probabilities of their ruin in n games? Give the numerical result if $n = 20$. *Ans.*

For A: $\frac{3}{5} - \frac{4}{5}\left\{\left(\frac{\sqrt{5}+1}{4}\right)^{n+\epsilon} + \left(\frac{\sqrt{5}-1}{4}\right)^{n+\epsilon}\right\}$; $\epsilon = 1$ if n is odd, $\epsilon = 2$ if n is even.

For B: $\frac{2}{5} - \frac{4}{5}\left\{\left(\frac{\sqrt{5}+1}{4}\right)^{n+\eta} - \left(\frac{\sqrt{5}-1}{4}\right)^{n+\eta}\right\}$; $\eta = 1$ if n is even, $\eta = 2$ if n is odd.

7. Find the expression of $y_{a,n}$, the probability of the ruin of A when his adversary B is infinitely rich, corresponding to formula (20). *Ans.* From the definition of a definite integral it follows that

$$y_{a,n} = y_{a,\infty} - \frac{(2\sqrt{pq})^{n+1}\left(\frac{q}{p}\right)^{\frac{a}{2}}}{\pi}\int_0^{\pi} \frac{\sin\varphi \sin a\varphi}{1 - 2\sqrt{pq}\cos\varphi}(\cos\varphi)^n d\varphi$$

where

$$y_{a,\infty} = 1 \quad \text{if} \quad p \leqq q$$

$$y_{a,\infty} = \left(\frac{q}{p}\right)^a \quad \text{if} \quad p > q.$$

If the games are equitable and n differs from a by an even number, then

$$y_{a,n} = 1 - \frac{2}{\pi}\int_0^{\frac{\pi}{2}} \frac{\sin a\varphi}{\sin\varphi}(\cos\varphi)^{n+1} d\varphi.$$

This formula was given by Laplace.

8. Referring to the last formula in the preceding problem, show that

$$y_{a,n} = 1 - \frac{2}{\sqrt{\pi}}\int_0^t e^{-u^2} du + \Delta$$

where

$$t = \frac{a}{\sqrt{2(n + \frac{4}{3})}}; \qquad |\Delta| < \frac{1}{2\pi n} + \frac{2}{n} e^{-\frac{\pi^2 n}{32}}.$$

Indication of the Proof. It is important to prove the following inequalities first

$$\frac{\varphi\,(\cos\varphi)^{n+1}}{\sin\varphi} < e^{-\frac{n+\frac{4}{3}}{2}\varphi^2} \quad \text{for} \quad 0 < \varphi \leqq \frac{\pi}{2}$$

$$\frac{\varphi(\cos\varphi)^{n+1}}{\sin\varphi} > e^{-\frac{n+\frac{4}{3}}{2}\varphi^2 - \frac{(n+1)\varphi^4}{8}} \quad \text{for} \quad 0 < \varphi \leqq \frac{\pi}{4}$$

whence

$$\frac{\varphi\,(\cos\varphi)^{n+1}}{\sin\varphi} = e^{-\frac{n+\frac{4}{3}}{2}\varphi^2}\left[1 - \theta\frac{n+1}{8}\varphi^4\right]; \qquad 0 < \theta < 1$$

provided $0 < \varphi \leqq \pi/4$. The rest of the proof is easy.

9. Attempt a direct proof of the important lemma (page 144) used in the discussion of Prob. 2.

HINT: The proof can be based upon the following proposition[1] generalizing an important theorem on determinants due to Minkowski: Let

$$f_i = a_{1i}x_1 + a_{2i}x_2 + \cdots + a_{ni}x_n; \qquad i = 1, 2, 3, \ldots n$$

be a system of linear forms whose coefficients satisfy the following conditions:

(1) $a_{ii} > 0$; $a_{ki} \leqq 0$ *if* $k \neq i$; $a_{1i} + a_{2i} + \cdots + a_{ni} \geqq 0$.

(2) One of these sums is positive.

If these forms assume nonnegative values, then every $x_i \geqq 0 (i = 1, 2, \ldots n)$. *Proof by induction:* Express x_n through $x_1, x_2, \ldots x_{n-1}$, thus:

$$x_n = \frac{f_n - a_{1n}x_1 - a_{2n}x_2 - \cdots - a_{n-1,n}x_{n-1}}{a_{nn}}$$

and substitute into the remaining forms. Show that the resulting forms in $x_1, x_2, \ldots x_{n-1}$ satisfy the same conditions (1) and (2). Hence, it remains to prove the proposition for two forms, which can easily be done

References

CHR. HUYGENS: "De ratiociniis in ludo aleae," 1654.
DE MOIVRE: "Doctrine of Chances," 3d ed., 1756.
LAGRANGE: Mémoire sur les suites récurrentes dont les termes varient de plusieurs manières différentes, etc., Oeuvres IV, pp. 151*ff*.
LAPLACE: "Théorie analytique des probabilités," Oeuvres VII, pp. 228–242.
BERTRAND: "Calcul des probabilités," pp. 104–141, Paris, 1889.
MARKOFF: "Wahrscheinlichkeitsrechnung," pp. 142–146, Leipzig, 1912.

[1] The author is indebted to Professor Besikovitch of Cambridge, England, for the communication of this direct proof.

CHAPTER IX

MATHEMATICAL EXPECTATION

1. Bernoulli's theorem, important though it is, is but the first link in a chain of theorems of the same character, all contained in an extremely general proposition with which we shall deal in the next chapter. But before proceeding to this task, it is necessary to extend the definition of "mathematical expectation"—an important concept originating in connection with games of chance.

If, according to the conditions of the game, the player can win a sum a with probability p, and lose a sum b with probability $q = 1 - p$, the mathematical expectation of his gain is by definition

$$pa - qb.$$

Considering the loss as a negative gain, we may say that the gain of the player may have only two values, a and $-b$, with the corresponding probabilities p and q, so that the expectation of his gain is the sum of the products of two possible values of the gain by their probabilities. In this case, the gain appears as a variable quantity possessing two values.

Variable quantities with a definite range of values each one of which, depending on chance, can be attained with a definite probability, are called "chance variables," or, using a Greek term, "stochastic" variables. They play an important part in the theory of probability. A stochastic variable is defined (a) if the set of its possible values is given, and (b) if the probability to attain each particular value is also given.

It is easy to give examples of stochastic variables. The gain in a game of chance is a stochastic variable with two values. The number of points on a die that is tossed, is a stochastic variable with six values, 1, 2, . . . 6, each of which has the same probability $\frac{1}{6}$. A number on a ticket drawn from an urn containing 20 tickets numbered from 1 to 20, is a stochastic variable with 20 values, and the probability to attain any one of them is $\frac{1}{20}$. Each of two urns contains 2 white and 2 black balls. Simultaneously, one ball is transferred from the first urn into the second, while one ball from the latter is transferred into the first. After this exchange, the number of white balls in one of the urns may be regarded as a stochastic variable with three values, 1, 2, 3, whose corresponding probabilities are, respectively, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$. It is natural to extend the concept of mathematical expectation to stochastic variables in general.

Suppose that a stochastic variable x possesses n values:

$$x_1, x_2, \ldots x_n,$$

and

$$p_1, p_2, \ldots p_n$$

denote the respective probabilities for x to assume values $x_1, x_2, \ldots x_n$. By definition the mathematical expectation of x is

$$E(x) = p_1x_1 + p_2x_2 + \cdots + p_nx_n.$$

It is understood in this definition that the possible values of the variable x are *numerically* different. For instance, if the variable is a number of points on a die, its numerically different values are 1, 2, 3, 4, 5, 6, each having the same probability, $1/6$. By definition, the mathematical expectation of the number of points on a die is

$$\tfrac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5.$$

If the variable is the number on a ticket drawn from an urn containing 20 tickets numbered from 1 to 20, its numerically different values are represented by numbers from 1 to 20, and the probability of each of these values is $1/20$, so that the mathematical expectation of the number on a ticket is

$$\tfrac{1}{20}(1 + 2 + \cdots + 20) = 10.5.$$

2. It is obvious that the computation of mathematical expectation requires only the knowledge of the numerically different values of the variables with their respective probabilities. But in some cases this computation is greatly simplified by extending the definition of mathematical expectation. Suppose that, corresponding to mutually exclusive and exhaustive cases $A_1, A_2, \ldots A_m$, the variable x assumes the values $x_1, x_2, \ldots x_m$, with the corresponding probabilities $p_1, p_2, \ldots p_m$; we can define the mathematical expectation of x by

$$E(x) = p_1x_1 + p_2x_2 + \cdots + p_mx_m.$$

What distinguishes this extended definition from the original one is that in the second definition the values $x_1, x_2, \ldots x_m$ need not be numerically different; the only condition is that they are determined by mutually exclusive and exhaustive cases.

To make this distinction clear, suppose that the variable x is the number of points on two dice. Numerically different values of this variable are

$$2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$$

and their respective probabilities

$$\tfrac{1}{36}, \tfrac{2}{36}, \tfrac{3}{36}, \tfrac{4}{36}, \tfrac{5}{36}, \tfrac{6}{36}, \tfrac{5}{36}, \tfrac{4}{36}, \tfrac{3}{36}, \tfrac{2}{36}, \tfrac{1}{36}.$$

Therefore, by original definition, the expectation of x is

$$\frac{2}{36} + \frac{6}{36} + \frac{12}{36} + \frac{20}{36} + \frac{30}{36} + \frac{42}{36} + \frac{40}{36} + \frac{36}{36} + \frac{30}{36} + \frac{22}{36} + \frac{12}{36} = \frac{252}{36} = 7.$$

But we can distinguish 36 exhaustive and mutually exclusive cases according to the number of points on each die and, correspondingly, 36 values of the variable x, as shown in the following table:

First die	Second die	x	First die	Second die	x
1	1	2	4	1	5
1	2	3	4	2	6
1	3	4	4	3	7
1	4	5	4	4	8
1	5	6	4	5	9
1	6	7	4	6	10
2	1	3	5	1	6
2	2	4	5	2	7
2	3	5	5	3	8
2	4	6	5	4	9
2	5	7	5	5	10
2	6	8	5	6	11
3	1	4	6	1	7
3	2	5	6	2	8
3	3	6	6	3	9
3	4	7	6	4	10
3	5	8	6	5	11
3	6	9	6	6	12

The probability of each of these 36 cases being $1/36$, by the extended definition the mathematical expectation of x is

$$\frac{2 + 2\cdot 3 + 4\cdot 3 + 5\cdot 4 + 6\cdot 5 + 7\cdot 6 + 8\cdot 5 + 9\cdot 4 + 10\cdot 3 + 11\cdot 2 + 12}{36} = 7$$

as it should be.

It is important to show that both definitions always give the same value for the mathematical expectation.

Let $x_1, x_2, \ldots x_m$ be the values of the variable x corresponding to mutually exclusive and exhaustive cases $A_1, A_2, \ldots A_m$, and, $p_1, p_2, \ldots p_m$, their respective probabilities. By the extended definition of mathematical expectation, we have

$$E(x) = p_1x_1 + p_2x_2 + \cdots + p_mx_m. \tag{1}$$

The values $x_1, x_2, \ldots x_m$ are not necessarily numerically different, the numerically different values being

$$\xi, \eta, \zeta, \ldots \lambda.$$

We can suppose that the notation is chosen in such a way that

$$\begin{array}{l} x_1, x_2, \ldots x_a \text{ are equal to } \xi; \\ x_{a+1}, x_{a+2}, \ldots x_b \text{ are equal to } \eta; \\ x_{b+1}, x_{b+2}, \ldots x_c \text{ are equal to } \zeta; \\ \quad \cdots\cdots\cdots \\ x_{l+1}, x_{l+2}, \ldots x_m \text{ are equal to } \lambda. \end{array}$$

Hence, the right-hand member of (1) can be represented thus:

$$(p_1 + p_2 + \cdots + p_a)\xi + (p_{a+1} + p_{a+2} + \cdots + p_b)\eta + \cdots + \\ + (p_{l+1} + p_{l+2} + \cdots + p_m)\lambda.$$

But by the theorem of total probabilities, the sum

$$p_1 + p_2 + \cdots + p_a$$

represents the probability P for the variable x to assume a determined value ξ, because this can happen in a mutually exclusive ways; namely, when $x = x_1$, or $x = x_2, \ldots$ or $x = x_a$. By a similar argument we see that the sums

$$\begin{array}{l} p_{a+1} + p_{a+2} + \cdots + p_b \\ p_{b+1} + p_{b+2} + \cdots + p_c \\ \quad \cdots\cdots\cdots \\ p_{l+1} + p_{l+2} + \cdots + p_m \end{array}$$

represent the probabilities $Q, R, \ldots T$ for the variable x to assume values $\eta, \zeta, \ldots \lambda$. Therefore, the right-hand member of (1) reduces to the sum

$$P\xi + Q\eta + R\zeta + \cdots + T\lambda$$

which, by the original definition, is the mathematical expectation of x.

If, corresponding to mutually exclusive and exhaustive cases, a variable x assumes the same value a—in other words, remains constant—it is almost evident that its mathematical expectation is a, because the sum of the probabilities of mutually exclusive and exhaustive cases is 1. It is also evident that the expectation of ax where a is a constant, is equal to a times the expectation of x.

NOTE: Very often the mathematical expectation of a stochastic variable is called its "mean value."

MATHEMATICAL EXPECTATION OF A SUM

3. In many cases the computation of mathematical expectation is greatly facilitated by means of the following very general theorem:

Theorem. *The mathematical expectation of the sum of several variables is equal to the sum of their expectations; or, in symbols,*

$$E(x + y + z + \cdots + w) = E(x) + E(y) + E(z) + \cdots + E(w).$$

Proof. We shall prove this theorem first in the case of a sum of two variables. Let x assume numerically different values $x_1, x_2, \ldots x_m$, while numerically different values of y are $y_1, y_2, \ldots y_n$. In regard to the sum $x + y$ we can distinguish mn mutually exclusive cases; namely, when x assumes a definite value x_i and y another definite value y_j, while i and j range respectively over numbers $1, 2, 3, \ldots m$ and $1, 2, 3, \ldots n$. If p_{ij} denotes the probability of coexistence of the equalities

$$x = x_i, \qquad y = y_j$$

we have by the extended definition of mathematical expectation

$$E(x + y) = \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}(x_i + y_j),$$

or

$$E(x + y) = \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}x_i + \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}y_j. \tag{2}$$

As the variable x assumes a definite value x_i in n mutually exclusive ways (namely, when the value x_i of x is accompanied by the values $y_1, y_2, \ldots y_n$ of y) it is obvious that the sum

$$\sum_{j=1}^{n} p_{ij}$$

represents the probability p_i of the equality $x = x_i$. In a similar manner we see that the sum

$$\sum_{i=1}^{m} p_{ij}$$

represents the probability q_j of the equality $y = y_j$. Therefore

$$\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}x_i = \sum_{i=1}^{m} x_i \sum_{j=1}^{n} p_{ij} = \sum_{i=1}^{m} p_i x_i = E(x),$$

and similarly

$$\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}y_j = \sum_{j=1}^{n}\sum_{i=1}^{m} p_{ij}y_j = \sum_{j=1}^{n} q_j y_j = E(y);$$

that is, by (2)

$$E(x + y) = E(x) + E(y)$$

which proves the theorem for the sum of two variables.

If we deal with the sum of three variables $x + y + z$, we may consider it at first as the sum of $x + y$ and z and, applying the foregoing result, we get

$$E(x + y + z) = E(x + y) + E(z);$$

and again, by substituting $E(x) + E(y)$ for $E(x + y)$,

$$E(x + y + z) = E(x) + E(y) + E(z).$$

In a similar way we may proceed farther and prove the theorem for the sum of any number of variables.

4. The theorem concerning mathematical expectation of sums, simple though it is, is of fundamental importance on account of its very general nature and will be used frequently. At present, we shall use it in the solution of a few selected problems.

Problem 1. What is the mathematical expectation of the sum of points on n dice?

Solution. Denoting by x_i the number of points on the ith die, the sum of the points on n dice will be

$$s = x_1 + x_2 + \cdots + x_n,$$

and by the preceding theorem

$$E(s) = E(x_1) + E(x_2) + \cdots + E(x_n).$$

But for every single die

$$E(x_i) = \tfrac{7}{2}; \qquad i = 1, 2, \ldots n;$$

therefore

$$E(s) = \frac{7n}{2}.$$

Problem 2. What is the mathematical expectation of the number of successes in n trials with constant probability p?

Solution. Suppose that we attach to every trial a variable which has the value 1 in case of a success and the value 0 in case of failure. If the variables attached to trials $1, 2, 3, \ldots n$ are denoted by $x_1, x_2, \ldots x_n$, their sum

$$m = x_1 + x_2 + \cdots + x_n$$

obviously gives the number of successes in n trials. Therefore, the required expectation is

$$E(m) = E(x_1) + E(x_2) + \cdots + E(x_n).$$

But for every $i = 1, 2, 3, \ldots n$

$$E(x_i) = p \cdot 1 + (1 - p) \cdot 0 = p,$$

because x_i may have values 1 and 0 with the probabilities p and $1 - p$ which are the same as the probabilities of a success or a failure in the ith trial. Hence,

$$E(m) = np$$

or

$$E(m - np) = 0,$$

which may also be written in the form

$$\sum_{m=0}^{n} T_m(m - np) = 0.$$

This result was obtained on page 116 in a totally different and more complicated way. The new deduction is preferable in that it is more elementary and can easily be extended to more complicated cases, as we shall see in the next problem.

Problem 3. Suppose that we have a series of n trials independent or not, the probability of an event being p_i in the ith trial when nothing is known about the results of other trials. What is the mathematical expectation of the number of successes m in n trials?

Solution. Again let us introduce the variable x_i connected with the ith trial in such a way that $x_i = 1$ when the trial results in a success and $x_i = 0$ when it results in failure. Obviously,

$$m = x_1 + x_2 + \cdots + x_n$$

and

$$E(m) = E(x_1) + E(x_2) + \cdots + E(x_n).$$

But

$$E(x_i) = 1 \cdot p_i + 0 \cdot (1 - p_i) = p_i$$

and therefore

$$E(m) = p_1 + p_2 + \cdots + p_n.$$

For instance, if we have 5 urns containing 1 white, 9 black; 2 white, 8 black; 3 white, 7 black; 4 white, 6 black; 5 white, 5 black balls, and we draw one ball out of every urn, the mathematical expectation of the number of white balls taken will be:

$$E(m) = \tfrac{1}{10} + \tfrac{2}{10} + \tfrac{3}{10} + \tfrac{4}{10} + \tfrac{5}{10} = 1.5.$$

Problem 4. An urn contains a white and b black balls, and c balls are drawn. What is the mathematical expectation of the number of the white balls drawn?

Solution. To every ball taken we attach a variable which has the value 1 if the extracted ball is white, and the value 0 otherwise. The number of white balls drawn will then be

$$s = x_1 + x_2 + \cdots + x_c.$$

But the probability that the ith ball removed will be white when nothing is known of the other balls is $\frac{a}{a+b}$; therefore

$$E(x_i) = \frac{a}{a+b} \cdot 1 + \frac{b}{a+b} \cdot 0 = \frac{a}{a+b}$$

for every i, and the required expectation is

$$E(s) = \frac{ca}{a+b}.$$

Problem 5. An urn contains n tickets numbered from 1 to n, and m tickets are drawn at a time. What is the mathematical expectation of the sum of numbers on the tickets drawn?

Solution. Suppose that m tickets drawn from the urn are disposed in a certain order, and a variable is attached to every ticket expressing its number. Denoting the variable attached to the ith ticket by x_i, the sum of the numbers on all m tickets apparently is

$$s = x_1 + x_2 + \cdots + x_m.$$

But when taken singly, the variable x_i may represent any of the numbers $1, 2, 3, \ldots n$, the probability of its being equal to any one of these numbers being $1/n$. By the definition of mathematical expectation, we have

$$E(x_i) = \frac{1 + 2 + 3 + \cdots + n}{n} = \frac{n+1}{2},$$

and therefore

$$E(s) = \frac{m(n+1)}{2}.$$

For example, taking the French lottery where $n = 90$ and $m = 5$, we find for the mathematical expectation of the sum of numbers on all 5 tickets

$$E(s) = \frac{5 \cdot 91}{2} = 227.5.$$

Problem 6. An urn contains n tickets numbered from 1 to n. These tickets are drawn one by one, so that a certain number appears in the first place, another number in the second place, and so on. We shall say that there is a "coincidence" when the number on a ticket corresponds to the place it occupies. For instance, there is a coincidence when the first ticket has number 1 or the second ticket has number 2, etc. Find the mathematical expectation of the number of coincidences. Also, find the probability that there will be none, or one, or two, etc., coincidences.

Solution. Let x_i denote a variable which has the value 1 if there is coincidence in the ith place, otherwise $x_i = 0$. The sum

$$s = x_1 + x_2 + \cdots + x_n$$

gives the total number of coincidences and

$$E(s) = E(x_1) + E(x_2) + \cdots + E(x_n).$$

But

$$E(x_i) = \frac{1}{n} \cdot 1 = \frac{1}{n}$$

because the probability of drawing a ticket with the number i in the ith place without any regard to other tickets obviously is $1/n$; therefore,

$$E(s) = n \cdot \frac{1}{n} = 1.$$

On the other hand, denoting the probability of exactly i coincidences by p_i, we have by definition

$$E(s) = p_1 + 2p_2 + \cdots + np_n,$$

and, comparing with the preceding result, we obtain

$$(3) \qquad p_1 + 2p_2 + \cdots + np_n = 1.$$

Let us denote by $\varphi(n)$ the probability that in drawing n tickets, we shall have no coincidences. It is easy to express p_i by means of $\varphi(n - i)$. In fact, we have exactly i coincidences in

$$C_n^i = \frac{n(n-1) \cdots (n-i+1)}{1 \cdot 2 \cdot 3 \cdots i}$$

mutually exclusive cases; namely, when the tickets of one of the

$$C_n^i$$

specified groups of i tickets have numbers corresponding to their places while the remaining $n - i$ tickets do not present coincidences at all. By the theorem of compound probability, the probability of i coincidences in i specified places is

$$\frac{1}{n} \cdot \frac{1}{n-1} \cdots \frac{1}{n-i+1}$$

and the probability of the absence of coincidences in the remaining $n - i$ places is $\varphi(n - i)$. The probability of exactly i coincidences in i specified places is therefore

$$\frac{\varphi(n-i)}{n(n-1)\cdots(n-i+1)},$$

and the total probability p_i of exactly i coincidences without specification of places is

$$p_i = \frac{n(n-1)\cdots(n-i+1)}{1\cdot 2\cdot 3\cdots i}\cdot\frac{\varphi(n-i)}{n(n-1)\cdots(n-i+1)},$$

or

$$p_i = \frac{\varphi(n-i)}{1\cdot 2\cdot 3\cdots i}. \tag{4}$$

The symbol $\varphi(0)$ has no meaning, but the preceding formula holds good even for $i = n$ if we assume $\varphi(0) = 1$.

Substituting expression (4) for p_i into (3), we reach the relation

$$\varphi(n-1) + \frac{\varphi(n-2)}{1!} + \frac{\varphi(n-3)}{2!} + \cdots + \frac{\varphi(0)}{(n-1)!} = 1;$$

or changing n into $n + 1$

$$\varphi(n) + \frac{\varphi(n-1)}{1!} + \frac{\varphi(n-2)}{2!} + \cdots + \frac{\varphi(0)}{n!} = 1,$$

which gives successively $\varphi(1)$, $\varphi(2)$, $\varphi(3)$, . . . by taking

$$n = 1, 2, 3, \ldots .$$

The general result, which can easily be verified, is

$$\varphi(n) = \sum_{k=0}^{n}\frac{(-1)^k}{k!},$$

or, in an explicit form,

$$\varphi(n) = 1 - \frac{1}{1} + \frac{1}{1\cdot 2} - \frac{1}{1\cdot 2\cdot 3} + \cdots + \frac{(-1)^n}{1\cdot 2\cdot 3\cdots n}.$$

Even for moderate n this is very near to

$$\frac{1}{e} = 1 - \frac{1}{1} + \frac{1}{1\cdot 2} - \frac{1}{1\cdot 2\cdot 3} + \cdots \textit{ ad inf.} = 0.36787944.$$

MATHEMATICAL EXPECTATION OF A PRODUCT

5. For the product of two or more stochastic variables we do not possess anything so general as the foregoing theorem concerning the mathematical expectation of sums. An analogous theorem with respect to the product of stochastic variables can be established only under certain restrictive conditions.

Several stochastic variables are called "independent" if the probability for any one of them to assume a determined value does not depend on the values assumed by the remaining variables. For instance, if the variables are the numbers of points on dice, they may be considered as independent.

On the other hand, we have a case of dependent variables in numbers on tickets drawn in a lottery. For, in this case the fact that certain tickets have determined numbers precludes the possibility of any one of these numbers appearing on other tickets drawn at the same time.

If more than two variables are independent according to the above definition, it is clear that any two of them are independent. But the converse is not true: It is easy to imagine cases when any two of the variables are independent and yet they are not independent when taken in their totality. Therefore, when speaking of independence of variables, we must always specify whether they are independent in their totality or only in pairs.

For two independent variables we have the following simple theorem:

Theorem. *The mathematical expectation of the product xy of two independent variables x and y is equal to the product of their expectations; or, in symbols*

$$E(xy) = E(x)E(y).$$

Proof. Let $x_1, x_2, \ldots x_m$ be the complete set of values for x, and $y_1, y_2, \ldots y_n$ the analogous set for y. Denoting the probability of x being equal to x_i by p_i, and similarly, the probability of y being equal to y_j by q_j, the events

$$x = x_i \qquad \text{and} \qquad y = y_j$$

are *independent* by definition of independence—because the probability of x being equal to x_i is not affected by the fact that y has assumed any one of its possible values, and it remains p_i.

By the theorem of compound probability the simultaneous occurrence of the events

$$x = x_i \qquad \text{and} \qquad y = y_j$$

has the probability p_iq_j. Again, by the extended definition of mathematical expectation

$$E(xy) = \sum_{i=1}^{m}\sum_{j=1}^{n} p_iq_jx_iy_j$$

because the values of the product xy are determined by mn exhaustive and mutually exclusive cases

$$x = x_i, \qquad y = y_j$$
$$i = 1, 2, \ldots m; \qquad j = 1, 2, \ldots n.$$

Now, performing the summation with respect to j first, while i remains constant, we have

$$\sum_{j=1}^{n} p_i q_j x_i y_j = p_i x_i \sum_{j=1}^{n} q_j y_j = p_i x_i E(y),$$

and again

$$E(xy) = \sum_{i=1}^{m} p_i x_i E(y) = E(y) \sum_{i=1}^{m} p_i x_i,$$

or

$$E(xy) = E(x)E(y).$$

This theorem can be extended to the case of several factors *independent in their totality*. For instance, if x, y, z are independent, it is obvious that xy and z are also independent. Hence

$$E(xyz) = E(xy)E(z),$$

and again

$$E(xyz) = E(x)E(y)E(z).$$

In a similar way we can extend this theorem to any number of independent factors.

As an important application, let us consider two independent variables x and y with the respective expectations a and b. The variables $x - a$ and $y - b$ being independent also, we have

$$E(x - a)(y - b) = E(x - a)E(y - b);$$

but

$$E(x - a) = E(x) - a = a - a = 0;$$

therefore

(5) $$E(x - a)(y - b) = 0.$$

DISPERSION AND STANDARD DEVIATION

6. Let x be a variable and a its mathematical expectation. The expectation of

$$(x - a)^2$$

is called "dispersion" of the variable, and the square root of dispersion is usually called "standard deviation." As

$$(x - a)^2 = x^2 - 2ax + a^2$$

we can apply the theorem on the expectation of sums to the right-hand member of this identity and find

$$E(x - a)^2 = E(x^2) - 2aE(x) + a^2 = E(x^2) - a^2$$

or, denoting by b the expectation of x^2,

$$E(x - a)^2 = b - a^2. \tag{6}$$

Thus, the computation of dispersion can be reduced to the computation of the expectation of the variable itself and its square. Also, denoting by σ the standard deviation of x, we have the formula

$$\sigma^2 = b - a^2.$$

For instance, if the variable is the number of points on a die, we have

$$a = \frac{7}{2}, \qquad b = \frac{1^2 + 2^2 + \cdots + 6^2}{6} = \frac{91}{6}$$

and

$$\sigma^2 = \tfrac{91}{6} - \tfrac{49}{4} = 2.917; \qquad \sigma = 1.708.$$

Dispersion of Sums

7. It is important to have a convenient formula to find the dispersion of a sum

$$s = x_1 + x_2 + \cdots + x_n$$

of several stochastic variables. The expectation of s is given by

$$E(s) = E(x_1) + E(x_2) + \cdots + E(x_n)$$

or

$$E(s) = a_1 + a_2 + \cdots + a_n,$$

denoting by a_i the expectation of x_i. The deviation of s from its expectation is, therefore,

$$x_1 + x_2 + \cdots + x_n - (a_1 + a_2 + \cdots + a_n),$$

and we have to find the expectation of

$$(x_1 + x_2 + \cdots + x_n - a_1 - a_2 - \cdots - a_n)^2.$$

Now we have identically

$$(x_1 + x_2 + \cdots + x_n - a_1 - a_2 - \cdots - a_n)^2 = \sum_{i=1}^{n}(x_i - a_i)^2 + $$
$$+ 2\sum_{i,j}(x_i - a_i)(x_j - a_j),$$

the last sum being extended over all the different combinations of subscripts i and j for which $i \neq j$ and consisting of $n(n-1)/2$ terms. The mathematical expectation of a sum being equal to the sum of the expectations of its terms, we must find the expectations of the terms

$$(x_i - a_i)^2 \quad \text{and} \quad (x_i - a_i)(x_j - a_j).$$

The first is the dispersion of x_i and can be found from (6); namely,

$$E(x_i - a_i)^2 = b_i - a_i^2 = \sigma_i^2$$

if b_i is the expectation of x_i^2.

As to

$$E(x_i - a_i)(x_j - a_j),$$

instead of it we introduce the so-called "correlation coefficient" of x_i and x_j

$$R_{i,j} = \frac{E(x_i - a_i)(x_j - a_j)}{\sigma_i \sigma_j}.$$

Denoting the required dispersion by D, we obtain

$$(7) \quad D = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2 + 2R_{1,2}\sigma_1\sigma_2 + 2R_{1,3}\sigma_1\sigma_3 + \cdots + \\ + 2R_{n-1,n}\sigma_{n-1}\sigma_n$$

so that the dispersion of a sum can be obtained as soon as we know the dispersion of its terms and their correlation coefficients.

In an important case, expression (7) for dispersion can be greatly simplified. If the variables $x_1, x_2, \ldots x_n$ are *independent in pairs*, we see from (5) that all the correlation coefficients are $= 0$, so that in this case simply

$$(8) \quad D = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2 = b_1 - a_1^2 + b_2 - a_2^2 + \cdots + b_n - a_n^2.$$

In other words, the dispersion of a sum of variables, any two of which are independent, is equal to the sum of dispersions of its terms.

8. A few examples will serve to illustrate the use of these formulas.

Problem 7. Find the dispersion of the number of successes in series of n independent trials with probabilities $p_1, p_2, \ldots p_n$ corresponding to first, second, . . . nth trial.

Solution. As in Prob. 2 we associate with every trial a variable which assumes the value 1 or 0, according as the trial resulted in success or failure. These variables $x_1, x_2, \ldots x_n$ are independent because the trials are supposed to be independent. The number of successes

$$m = x_1 + x_2 + \cdots + x_n$$

is thus the sum of the independent variables. To find the dispersion of any one of these variables x_i we notice that

$$E(x_i) = 1 \cdot p_i + 0 \cdot q_i = p_i$$
$$E(x_i^2) = 1 \cdot p_i + 0 \cdot q_i = p_i;$$

therefore the dispersion of x_i is

$$\sigma_i^2 = p_i - p_i^2 = p_i q_i$$

and by (8)

$$D = E(m - p_1 - p_2 - \cdots - p_n)^2 = p_1 q_1 + p_2 q_2 + \cdots + p_n q_n.$$

In the Bernoullian case of independent trials with the same probability p, we have $p_1 = p_2 = \cdots = p_n = p$ and

$$E(m - np)^2 = npq.$$

This formula is equivalent to the relation

$$\sum_{m=0}^{n} T_m(m - np)^2 = npq$$

established on page 116.

Problem 8. In a lottery m tickets are drawn at a time out of n tickets numbered from 1 to n. Find the dispersion of the sum s of the numbers on the tickets drawn.

Solution. Let $x_1, x_2, \ldots x_m$ be the variables representing the numbers on the first, second, . . . mth tickets. By Prob. 5 we know that

$$E(x_i) = \frac{n+1}{2},$$

and in a similar way we find

$$E(x_i^2) = \frac{1^2 + 2^2 + \cdots + n^2}{n} = \frac{(n+1)(2n+1)}{6},$$

whence the dispersion of x_i is

$$E\left(x_i - \frac{n+1}{2}\right)^2 = \frac{n^2 - 1}{12}.$$

Since we deal in the present case with dependent variables, we must find the correlation coefficients, or, which is the same,

$$E\left(x_i - \frac{n+1}{2}\right)\left(x_j - \frac{n+1}{2}\right)$$

for every pair of subscripts i and j. The variable x_i may have any of the values 1, 2, 3, . . . n, with the same probability $1/n$; and x_j may have any of the same values with the exception of that assumed by x_i

with the probability $\frac{1}{n-1}$, so that the preceding expression consists of terms

$$\frac{1}{n(n-1)}\left(x_i - \frac{n+1}{2}\right)\left(x_j - \frac{n+1}{2}\right)$$

where x_j for given $x_i = 1, 2, \ldots n$, ranges over all numbers $1, 2, 3, \ldots n$ with the exception of x_i. As

$$\sum_{\xi=1}^{n}\left(\xi - \frac{n+1}{2}\right) = 0$$

it is obvious that

$$\sum\left(x_j - \frac{n+1}{2}\right) = -\left(x_i - \frac{n+1}{2}\right),$$

and

$$E\left(x_i - \frac{n+1}{2}\right)\left(x_j - \frac{n+1}{2}\right) = -\frac{1}{n(n-1)}\sum_{x_i=1}^{n}\left(x_i - \frac{n+1}{2}\right)^2 = $$
$$= -\frac{n+1}{12}.$$

Everything now is ready for the application of (7). All simplifications performed, we get the following expression of the required dispersion

$$D = \frac{m(n^2-1)}{12}\left(1 - \frac{m-1}{n-1}\right).$$

If the variables were independent, the dispersion would be

$$\frac{m(n^2-1)}{12}.$$

The dependence diminishes it, but the influence of dependence is not great if the ratio m/n is small.

Problems for Solution

1. Find the mathematical expectation M of the absolute value of the discrepancy $m - np$ in a series of n independent trials with constant probability p. *Ans.* By definition

$$M = \sum_{m=0}^{n} T_m|m - np|$$

where, as usual,

$$T_m = \frac{n!}{m!(n-m)!}p^m q^{n-m}.$$

But since

$$\sum_{m=0}^{n} T_m(m - np) = 0,$$

we have also

$$M = 2\sum_{m>np} T_m(m - np),$$

the sum being extended over all integers m which are $>np$. Denoting by $F(x, y)$ the sum

$$F(x, y) = \sum_{m>np} C_n^m x^m y^{n-m}$$

we have

$$\sum_{m>np} T_m(m - np) = p\frac{\partial F(p, q)}{\partial p} - npF(p, q).$$

On the other hand, by Euler's theorem on homogeneous functions

$$nF(p, q) = p\frac{\partial F}{\partial p} + q\frac{\partial F}{\partial q},$$

whence

$$\sum_{m>np} T_m(m - np) = pq\left(\frac{\partial F}{\partial p} - \frac{\partial F}{\partial q}\right) = npqC_{n-1}^{\mu-1}p^{\mu-1}q^{n-\mu}.$$

Here μ represents an integer determined by

$$\mu \leqq np + 1 < \mu + 1.$$

The answer is therefore given by the simple formula

$$M = 2npqC_{n-1}^{\mu-1}p^{\mu-1}q^{n-\mu}.$$

2. By applying Stirling's formula (Appendix 1, page 347) prove the following result:

$$M = \sqrt{\frac{2npq}{\pi}}\left(1 + \frac{1}{2}\theta c\right); \qquad |\theta| < 1$$

where

$$c = \max.\left(\frac{1}{np - 1}, \frac{1}{nq - 1}\right)$$

and n is so large as to make $c \leqq 1/10$.

Hint:

$$\log\left(M:\sqrt{\frac{2npq}{\pi}}\right) < \frac{\vartheta}{2(np - \vartheta)} + \frac{\vartheta'}{2(nq - \vartheta')} - \frac{1}{24}\max.\left(\frac{1}{np - \vartheta}, \frac{1}{nq - \vartheta'}\right)$$

$$\log\left(M:\sqrt{\frac{2npq}{\pi}}\right) > -\frac{1}{12(np - \vartheta)} - \frac{1}{12(nq - \vartheta')} - \frac{\vartheta^2}{4(np - \vartheta)^2} - \frac{\vartheta'^2}{4(nq - \vartheta')^2}$$

$$0 \leqq \vartheta \leqq 1; \qquad \vartheta' + \vartheta = 1$$

3. What is the expectation of the number of failures preceding the first success in an indefinite series of independent trials with the probability p?

$$Ans. \quad qp + 2q^2p + 3q^3p + \cdots = \frac{pq}{(1-q)^2} = \frac{q}{p}.$$

4. Balls are taken one by one out of an urn containing a white and b black balls until the first white ball is drawn. What is the expectation of the number of black balls preceding the first white ball?

Ans. 1. By direct application of definition the following first expression for the required expectation M is obtained:

$$M = \frac{a}{a+b}\left[\frac{b}{a+b-1} + 2\frac{b(b-1)}{(a+b-1)(a+b-2)} + \right.$$
$$\left. + 3\frac{b(b-1)(b-2)}{(a+b-1)(a+b-2)(a+b-3)} + \cdots\right].$$

Ans. 2. However, it is possible to find a simpler expression for M. Denote by x_1 the number of black balls preceding the first white ball, by x_2 the number of black balls between the first and second white ball, and so on; finally, by x_{a+1} the number of black balls following the last white ball. We have

$$x_1 + x_2 + \cdots + x_{a+1} = b$$

and

$$E(x_1) + E(x_2) + \cdots + E(x_{a+1}) = b.$$

But as the probability of every sequence of balls (that is, of every system of numbers $x_1, x_2, \ldots x_{a+1}$) is the same, namely,

$$\frac{a!b!}{(a+b)!}$$

it is easy to see that

$$E(x_1) = E(x_2) = \cdots = E(x_{a+1}) = M.$$

That is,

$$(a+1)M = b$$

or

$$M = \frac{b}{a+1}.$$

Equating this to the preceding expression for M, an interesting identity can be obtained, whose direct proof is left to the student.

5. In Prob. 6, page 168, to determine the probability $\varphi(n)$, we had an equation

$$\varphi(n) + \frac{\varphi(n-1)}{1!} + \frac{\varphi(n-2)}{2!} + \cdots + \frac{\varphi(0)}{n!} = 1; \qquad \varphi(0) = 1.$$

Find the general expression for $\varphi(n)$ using the method of generating functions. *Ans.* Let

$$F(x) = \varphi(0) + \varphi(1)x + \varphi(2)x^2 + \cdots$$

be the generating function of $\varphi(n)$. Multiplying this series by

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots,$$

we find

$$e^x F(x) = 1 + x + x^2 + \cdots = \frac{1}{1 - x}$$

or

$$F(x) = \frac{e^{-x}}{1 - x},$$

whence

$$\varphi(n) = 1 - \frac{1}{1!} + \frac{1}{2!} - \cdots + \frac{(-1)^n}{n!}.$$

6. The total number of balls in an urn is known, but the number of white balls depends on chance and only its mathematical expectation is known. Find the probability of drawing a white ball. *Ans.* Let N be the total number of balls and M the expectation of the number of white balls. The required probability is M/N.

7. Two urns contain, respectively, a white and b black and α white and β black balls. A certain number c (naturally not exceeding $a + b$) of balls is transferred from the first urn into the second. What is the probability of drawing a white ball from the second urn after the transfers? *Ans.* The required probability is

$$\frac{\alpha + \dfrac{ca}{a + b}}{\alpha + \beta + c}.$$

8. An urn contains a white and b black balls. After a ball is drawn, it is to be returned to the urn if it is white; but if it is black, it is to be replaced by a white ball from another urn. What is the probability of drawing a white ball after the foregoing operation has been repeated x times? *Ans.* Denote by M_x the expectation of the number of white balls after x operations. From the equation

$$M_{x+1} = \left(1 - \frac{1}{a + b}\right)M_x + 1$$

the following expression for M_x can be derived:

$$M_x = a + b - b\left(1 - \frac{1}{a + b}\right)^x.$$

It follows that the required probability is

$$p = 1 - \frac{b}{a + b}\left(1 - \frac{1}{a + b}\right)^x.$$

9. Urns 1 and 2 contain, respectively, a white and b black and c white and d black balls. One ball is taken from the first urn and transferred into the second, while simultaneously one ball taken from the second urn is transferred into the first. What is the probability of drawing a white ball from the first urn after such an exchange has been repeated x times? *Ans.* Let M_x and P_x represent the mathematical expectations of the number of white balls in the first and second urn after x exchanges. Then

$$M_{x+1} = M_x + \frac{P_x}{c + d} - \frac{M_x}{a + b}; \qquad M_x + P_x = a + c$$

whence

$$M_x = \frac{(a+c)(a+b)}{a+b+c+d} + \frac{ad-bc}{a+b+c+d}\left(1 - \frac{1}{a+b} - \frac{1}{c+d}\right)^x.$$

10. An urn contains pN white and qN black balls, the total number of balls being N. Balls are drawn one by one (without being returned to the urn) until a certain number n of balls is reached. What is the dispersion of the number m of white balls drawn? *Ans.* Let $x_i = 1$ if the ith ball drawn is white and $x_i = 0$ if it is black. We have

$$E(x_i) = p, \qquad E(m) = np, \qquad E(x_i^2) = p$$

and

$$E(x_i - p)(x_j - p) = E(x_i x_j) - p^2 = -\frac{pq}{N-1}.$$

The required dispersion is

$$D = E(m - np)^2 = npq\frac{N-n}{N-1}.$$

11. In a lottery containing n numbers $(1, 2, 3, \ldots n)$ m numbers are drawn at a time. Let x_i represent the frequency of a specified number i in N drawings. Prove that

$$E(x_i) = Np, \qquad E(x_i - Np)^2 = Npq$$
$$E(x_i - Np)(x_j - Np) = Np(p' - p); \qquad (i \neq j)$$

where

$$p = \frac{m}{n}, \qquad q = 1 - p, \qquad p' = \frac{m-1}{n-1}.$$

12. Let

$$z_i = (x_i - Np)^2 - Npq.$$

Show that the dispersion of the sum

$$z_1 + z_2 + \cdots + z_n$$

is

$$D = \frac{2N(N-1)}{n-1}(npq)^2.$$

Indication of the Proof. Let N variables $\xi_1, \xi_2, \ldots \xi_N$ be defined as follows:

$\xi_k = -p$ if in the kth drawing the number i fails to appear
$\xi_k = q$ if in the kth drawing the number i appears.

In a similar way, we can define N variables $\eta_1, \eta_2, \ldots \eta_N$ associated with the number $j \neq i$. Since

$$x_i - Np = \xi_1 + \xi_2 + \cdots + \xi_N$$
$$x_j - Np = \eta_1 + \eta_2 + \cdots + \eta_N$$

we have

$$e^{u(x_i - Np)} \cdot e^{v(x_j - Np)} = e^{u\xi_1 + v\eta_1} \cdot e^{u\xi_2 + v\eta_2} \cdots e^{u\xi_N + v\eta_N}.$$

The variables

$$e^{u\xi_1 + v\eta_1}, e^{u\xi_2 + v\eta_2}, \ldots e^{u\xi_N + v\eta_N}$$

being independent, we have

$$E(e^{u(x_i-Np)+v(x_j-Np)}) = E(e^{u\xi_1+v\eta_1}) \cdot E(e^{u\xi_2+v\eta_2}) \cdot \cdot \cdot E(e^{u\xi_N+v\eta_N}).$$

But

$$\begin{aligned} E(e^{u\xi_1+v\eta_1}) &= E(e^{u\xi_2+v\eta_2}) = \cdot \cdot \cdot = E(e^{u\xi_N+v\eta_N}) = \\ &= pp'e^{qu+qv} + p(1-p')e^{qu-pv} + p(1-p')e^{qv-pu} + (q-p+pp')e^{-pu-pv} = \\ &= F(u, v). \end{aligned}$$

Hence

$$E(e^{u(x_i-Np)+v(x_j-Np)}) = F(u, v)^N.$$

It suffices to expand both members into power series in u and v and compare terms involving u^2v^2 to find

$$E(z_i z_j); \qquad i \neq j.$$

The rest does not present serious difficulties except for somewhat complicated calculations.

13. A box contains 2^n tickets among which C_n^i tickets bear the number i (i = 0, 1, 2, . . . n). A group of m tickets is drawn; denoting by s the sum of their numbers, it is required to find the expectation E and the dispersion D of s.

Ans. $E = \frac{1}{2}mn$; $D = \frac{1}{4}mn - \frac{m(m-1)n}{4(2^n-1)}$.

14. A box contains k varieties of objects, the number of objects of each variety being the same. These objects are drawn one at a time and put back before the next drawing. Denoting by n the smallest number of drawings which produce objects of all varieties, find $E(n)$ and $E(n^2)$. *Ans.*

$$E(n) = k\left(1 + \frac{1}{2} + \frac{1}{3} + \cdot \cdot \cdot + \frac{1}{k}\right)$$

$$E(n^2) - E(n)^2 = k^2\left(1 + \frac{1}{2^2} + \cdot \cdot \cdot + \frac{1}{k^2}\right) - k\left(1 + \frac{1}{2} + \cdot \cdot \cdot + \frac{1}{k}\right).$$

Use the result of Prob. 12, p. 41.

References

A. Markoff: "Wahrscheinlichkeitsrechnung," pp. 45*ff*., Leipzig, 1912.
G. Castelnuovo: "Calcolo delle probabilità," vol. 1, pp. 35–62, Bologna, 1925.
J. Bertrand: "Calcul des probabilités," Paris, 1889.

CHAPTER X

THE LAW OF LARGE NUMBERS

1. The developments of the preceding chapter, combined with a simple lemma due to Tshebysheff, lead in a natural and easy way to a far reaching generalization of Bernoulli's theorem, known under the name of the "law of large numbers."

Tshebysheff's Lemma. *Let u be a variable which does not assume negative values, and a its mathematical expectation. The probability of the inequality*

$$u \leqq at^2$$

is always greater than

$$1 - \frac{1}{t^2}$$

whatever t may be.

Proof. Let

$$u_1, u_2, \ldots u_n$$

be all the possible values of the variable u and

$$p_1, p_2, \ldots p_n$$

their respective probabilities. By the definition of mathematical expectation, we have

$$p_1u_1 + p_2u_2 + \cdots + p_nu_n = a. \tag{1}$$

We may suppose the notations so chosen that

$$u_1, u_2, \ldots u_\alpha$$

are all the values of u which are $\leqq at^2$, the remaining values

$$u_{\alpha+1}, u_{\alpha+2}, \ldots u_n$$

being $> at^2$. If all the terms in (1) with subscripts $1, 2, \ldots \alpha$ are dropped, the left-hand members can only be diminished, since these terms are positive or at least nonnegative by hypothesis. We have, therefore,

$$p_{\alpha+1}u_{\alpha+1} + \cdots + p_nu_n \leqq a.$$

But as

$$u_i > at^2$$

for $i = \alpha + 1, \alpha + 2, \ldots n$ a still stronger inequality,

$$at^2(p_{\alpha+1} + \cdots + p_n) < a$$

or

$$p_{\alpha+1} + \cdots + p_n < \frac{1}{t^2}$$

will hold.

Here the left-hand member represents the probability Q of the inequality

$$u > at^2$$

because this inequality can materialize only in the following mutually exclusive forms: either $u = u_{\alpha+1}$, or $u = u_{\alpha+2}, \ldots$ or $u = u_n$ whose probabilities are, respectively, $p_{\alpha+1}, p_{\alpha+2}, \ldots p_n$. Thus

$$Q < \frac{1}{t^2}.$$

But if P is the probability of the opposite event

$$u \leqq at^2,$$

we must have

$$P + Q = 1,$$

whence

$$P > 1 - \frac{1}{t^2}$$

which proves the lemma.

2. Let $x_1, x_2, \ldots x_n$ be a set of stochastic variables and $a_1, a_2, \ldots a_n$ their respective expectations. The dispersion of the sum

$$x_1 + x_2 + \cdots + x_n$$

which we shall denote by B_n is, by definition, the mathematical expectation of the variable

$$u = (x_1 + x_2 + \cdots + x_n - a_1 - a_2 - \cdots - a_n)^2.$$

Tshebysheff's lemma, applied to this variable u, shows that the probability of the inequality

$$(x_1 + x_2 + \cdots + x_n - a_1 - a_2 - \cdots - a_n)^2 \leqq B_n t^2$$

is greater than

$$1 - \frac{1}{t^2}.$$

But the preceding inequality is equivalent to two inequalities

$$-t\sqrt{B_n} \leqq x_1 + x_2 + \cdots + x_n - a_1 - a_2 - \cdots - a_n \leqq t\sqrt{B_n}$$

or, dividing through by n,

$$-t\sqrt{\frac{B_n}{n^2}} \leqq \frac{x_1 + x_2 + \cdots + x_n}{n} - \frac{a_1 + a_2 + \cdots + a_n}{n} \leqq t\sqrt{\frac{B_n}{n^2}}.$$

Hence, the probability of these inequalities for an arbitrary positive t is greater than

$$1 - \frac{1}{t^2}.$$

Let ϵ be an arbitrary positive number. Defining t by the equation

$$t\sqrt{\frac{B_n}{n^2}} = \epsilon,$$

whence

$$t^2 = \frac{n^2\epsilon^2}{B_n},$$

we arrive at the following conclusion: The probability P of the inequalities

$$-\epsilon \leqq \frac{x_1 + x_2 + \cdots + x_n}{n} - \frac{a_1 + a_2 + \cdots + a_n}{n} \leqq \epsilon$$

equivalent to a single inequality

$$\left|\frac{x_1 + x_2 + \cdots + x_n}{n} - \frac{a_1 + a_2 + \cdots + a_n}{n}\right| \leqq \epsilon$$

is greater than

$$1 - \frac{B_n}{n^2\epsilon^2}.$$

Thus far nothing has been supposed about the behavior of B_n for indefinitely increasing n. We shall now suppose that the quotient B_n/n^2 tends to 0 as n increases indefinitely. Then, having chosen two arbitrarily small positive numbers ϵ and η, a number n_0 can be found so that the inequality

$$\frac{B_n}{n^2\epsilon^2} < \eta$$

will hold for $n > n_0$. Consequently, we shall have

$$P > 1 - \eta$$

for all $n > n_0$. This conclusion leads to the following important theorem due, in the main, to Tshebysheff:

The Law of Large Numbers. *With the probability approaching* 1 *or certainty as near as we please, we may expect that the arithmetic mean of values actually assumed by n stochastic variables will differ from the arithmetic mean of their expectations by less than any given number, however small, provided the number of variables can be taken sufficiently large and provided the condition*

$$\frac{B_n}{n^2} \to 0 \qquad \text{as} \qquad n \to \infty$$

is fulfilled.

If, instead of variables x_i, we consider new variables $z_i = x_i - a_i$ with their means $= 0$, the same theorem can be stated as follows:

For a fixed $\epsilon > 0$, however small, the probability of the inequality

$$\left|\frac{z_1 + z_2 + \cdots + z_n}{n}\right| \leqq \epsilon$$

tends to 1 as a limit when n increases indefinitely, provided

$$\frac{B_n}{n^2} \to 0.$$

This theorem is very general. It holds for independent or dependent variables indifferently if the sufficient condition for its validity, namely, that

$$\frac{B_n}{n^2} \to 0 \qquad \text{as} \qquad n \to \infty$$

is fulfilled.

3. This condition, which is recognized as sufficient, is at the same time necessary, if the variables $z_1, z_2, \ldots z_n$ are uniformly bounded; that is, if a constant number (one independent of n), C, can be found so that all particular values of $z_i (i = 1, 2, \ldots n)$ are numerically less than C. Let P, as before, denote the probability of the inequality

$$|z_1 + z_2 + \cdots + z_n| \leqq n\epsilon.$$

Then the probability of the opposite inequality

$$|z_1 + z_2 + \cdots + z_n| > n\epsilon$$

will be $1 - P$.

Now, by definition,

$$B_n = E(z_1 + z_2 + \cdots + z_n)^2$$

whence one can easily derive the inequality

$$B_n < n^2C^2(1 - P) + n^2\epsilon^2P$$

from which it follows that

$$\frac{B_n}{n^2} < C^2(1 - P) + \epsilon^2 P < \epsilon^2 + C^2(1 - P).$$

If the law of large numbers holds, $1 - P$ converges to 0 when n increases indefinitely, so that the right-hand member for sufficiently large n becomes less than *any* given number, and that implies

$$\frac{B_n}{n^2} \to 0,$$

which proves the statement.

4. There is an important case in which the law of large numbers certainly holds; namely, when variables $x_1, x_2, \ldots x_n$ are independent and the expectations of their squares are bounded. Then a constant number C exists such that

$$b_i = E(x_i^2) < C \qquad \text{for} \qquad i = 1, 2, 3, \ldots.$$

On the other hand, for independent variables

$$B_n = \sum_{i=1}^{n}(b_i - a_i^2) \leqq \sum_{i=1}^{n} b_i < nC$$

and

$$\frac{B_n}{n^2} < \frac{C}{n} \to 0 \qquad \text{as} \qquad n \to \infty.$$

The expectations of squares are bounded, for instance, when all the variables are uniformly bounded, which is true, for instance, for "identical" or "equal" variables. Variables are said to be identical if they possess the same set of values with the same corresponding probabilities.

5. E. Czuber made a complete investigation of the results of 2,854 drawings in a lottery operated in Prague between 1754 and 1886. It consisted of 90 numbers, of which 5 were taken in each drawing. From Czuber's book "Wahrscheinlichkeitsrechnung," vol. 1, p. 141 (2d ed., 1908), we reprint the table shown on page 187.

With the 2,854 drawings, we associate 2,854 variables, $x_1, x_2, \ldots x_{2854}$ representing the sum of five numbers appearing in each of the 2,854 drawings. These variables are identical and independent with the common mathematical expectation 227.5. Hence, by the law of large numbers, we can expect that the arithmetic mean of actually observed values of these variables will not notably differ from 227.5. To form the sum

$$S = \sum_{i=1}^{2854} x_i$$

Numbers	Their frequency m	Difference $m - 158$
6	138	−20
39, 65	139	−19
16, 41, 76, 87	142	−16
2, 14, 56, 79, 86	143	−15
18, 44, 47	144	−14
72, 80	145	−13
12	146	−12
21, 53	147	−11
70	149	− 9
24, 32, 55, 69	150	− 8
27, 64, 75	151	− 7
81	152	− 6
23, 29, 85	153	− 5
19, 35, 42, 74	154	− 4
7, 20, 59	155	− 3
13, 34, 40, 67, 88	156	− 2
11, 52, 68	157	− 1
17, 82	158	0
15, 90	159	1
58	160	2
8, 25, 36	161	3
22	162	4
33, 57	163	5
51	164	6
3, 43, 45, 48	165	7
10, 26, 66	166	8
1, 5, 60, 84	167	9
50, 62	168	10
9, 61, 63	170	12
54, 73	171	13
49, 71, 78	172	14
28	173	15
37	176	18
30, 46	177	19
89	178	20
31	179	21
38	184	26
4	185	27
77	186	28
83	189	31

we must multiply the frequencies given in the preceding table by the sum of corresponding numbers. To simplify the task we notice that all numbers from 1 to 90, actually appeared. Hence, we multiply the sum of these numbers, 4,095, by 158, which gives:

$$4095 \cdot 158 = 647{,}010,$$

and then add to this number the sum of the differences $m - 158$ multiplied by the sum of the numbers in the same line. The results are:

Sum of positive products	Sum of negative products
22,336	−19,587.

Hence

$$S = 647{,}010 + 22{,}336 - 19{,}587 = 649{,}759$$

and

$$\frac{S}{2854} = 227.67,$$

which differs very little from the expected value 227.5. An even larger difference would be in perfect agreement with the law of large numbers since 2,854, the number of variables, is not very great.

6. The two experiments reported in this section were made by the author in spare moments. In the first experiment 64 tickets bearing numbers 0, 1, 2, 3, 4, 5, 6 and occurring in the following proportions:

Number	0	1	2	3	4	5	6
Frequency	1	6	15	20	15	6	1

were vigorously agitated in a tin can and then 10 tickets were drawn at a time and their numbers added. Altogether 2,500 such drawings were made and their results carefully recorded. From these records we derive Tables I and II.

TABLE I

Number	Frequency observed	Expected frequency	Discrepancy
0	404	390 625	+13.375
1	2,321	2,343.75	−22.75
2	5,850	5,859.375	− 9.375
3	7,863	7,812.5	+50.5
4	5,821	5,859.375	−38 375
5	2,344	2,343.75	+ 0.25
6	397	390.625	+ 6.375

The next table gives the absolute values of differences $s - 30$ where s is the sum of the numbers on 10 tickets drawn at one time, and their respective frequencies.

From Table I it is easy to find that the arithmetic mean of all 2,500 sums observed is:

$$\frac{74996}{2500} = 29.9984$$

TABLE II

$\lvert s - 30\rvert$	Frequency observed	$\lvert s - 30\rvert$	Frequency observed
0	246	7	71
1	549	8	44
2	479	9	25
3	379	10	8
4	324	11	4
5	241	12	1
6	129		

whereas the expectation of each of the 2,500 identical variables under consideration by Prob. 13, page 181, is 30. By the same problem the dispersion of s, that is, $E(s - 30)^2$ is 12.857. On the other hand, from Table II we find that

$$\Sigma(s - 30)^2 = 31477$$

and

$$\frac{\Sigma(s - 30)^2}{2500} = 12.5908$$

fairly close to 12.857.

In the second experiment we tried to produce cards of every suit in n drawings (n being the smallest number required) of one card at a time, each card taken being returned before the next drawing. By Prob. 14, page 181, we find that the expectation and the dispersion of this number n are, respectively, $8\frac{1}{3}$ and 14.44. Altogether 3,000 values of n were recorded, of which 33 was the largest. Values of the difference $n - 8$ are given in Table III.

TABLE III

$n - 8$	Frequency	$n - 8$	Frequency	$n - 8$	Frequency
−4	282	6	77	16	3
−3	420	7	50	17	5
−2	426	8	40	18	2
−1	407	9	31	19	1
0	348	10	17	20	3
1	247	11	15	21	1
2	228	12	13	22	1
3	156	13	6	23	1
4	116	14	9	24	0
5	88	15	6	25	1

From this table we find

$$\Sigma(n - 8) = 965, \qquad \Sigma(n - 8)^2 = 43{,}395,$$

whence

$$\Sigma(n - 8\tfrac{1}{3})^2 = \Sigma(n - 8)^2 - \tfrac{2}{3}\Sigma(n - 8) + \tfrac{3000}{9} = 43{,}085$$
$$\Sigma n = 24{,}965.$$

By the law of large numbers we may expect that the quotients

$$\frac{\Sigma n}{3000} \quad \text{and} \quad \frac{\Sigma(n - 8\frac{1}{3})^2}{3000}$$

will not considerably differ from $8\frac{1}{3}$ and 14.44, respectively. As a matter of fact,

$$\frac{\Sigma n}{3000} = 8.322, \qquad \frac{\Sigma(n - 8\frac{1}{3})^2}{3000} = 14.362.$$

There is a very satisfactory agreement between the theory and this experiment in another respect. Of 24,965 cards drawn there were

6,304 hearts
6,236 diamonds
6,131 clubs
6,294 spades

whereas the expected number for each suit is 6241.25.

7. So far, we have dealt with stochastic variables having only a finite number of values. However, the notion of mathematical expectation, and the propositions essentially based on this notion, can be extended to variables with infinitely many values. Here we shall consider the simplest case of variables with a countable set of values, that can be arranged in a sequence

$$\cdots < \alpha_{-2} < \alpha_{-1} < \alpha_0 < \alpha_1 < \alpha_2 < \cdots$$

in the order of their magnitude.

With this sequence is associated the sequence of probabilities

$$\ldots, p_{-2}, p_{-1}, p_0, p_1, p_2, \ldots$$

so that in general p_i is the probability for x to assume the value α_i. These probabilities are subject to the condition that the series

$$\Sigma p_i = \cdots + p_{-2} + p_{-1} + p_0 + p_1 + p_2 + \cdots$$

must be convergent with the sum 1.

The definition of mathematical expectation is essentially the same as that for variables with a finite number of values, but instead of a finite sum, we have an infinite series

$$E(x) = \Sigma p_i \alpha_i$$

provided this series is convergent (it is absolutely convergent, if convergent at all). If this series is divergent, it is meaningless to speak of the mathematical expectation of x. Likewise, the mathematical expectation of any function $\varphi(x)$ is defined as being the sum of the series

$$E\{\varphi(x)\} = \Sigma p_i \varphi(\alpha_i),$$

provided the latter is convergent.

It can easily be seen that various theorems established in Chap. IX, as well as Tshebysheff's lemma, continue to hold when the various mathematical expectations involved exist.

The law of large numbers follows, as a simple corollary, from Tshebysheff's lemma if the following requirements are fulfilled:

a. Mathematical expectations of all variables $x_1, x_2, x_3, \ldots$ exist.

b. The dispersion B_n of the sum $x_1 + x_2 + \cdots + x_n$ exists.

c. The quotient B_n/n^2 tends to 0 as n tends to infinity.

The first requirement is absolutely indispensable. Without it the theorem itself cannot be stated. The second requirement (not to speak of the third) need not be fulfilled; and still the law of large numbers may hold, as Markoff pointed out.

8. Let $x_1, x_2, x_3, \ldots$ be independent variables. If for every i the mathematical expectation

$$E(x_i^2)$$

exists, the quantity B_n exists also. But if at least one of these expectations does not exist, the quantity B_n has no meaning. However, the following theorem, due to Markoff, holds:

Theorem. *The law of large numbers holds, provided that for some* $\delta > 0$ *all the mathematical expectations*

$$E(|x_i|^{1+\delta}); \qquad i = 1, 2, 3, \ldots$$

exist and are bounded.

Proof. For the sake of simplicity we may assume that

$$E(x_i) = 0; \qquad i = 1, 2, 3, \ldots.$$

For, supposing

$$E(x_i) = a_i; \qquad i = 1, 2, 3, \ldots$$

instead of x_i, we may consider new variables

$$z_i = x_i - a_i.$$

Then

$$E(z_i) = 0$$

and it remains to prove the existence and boundedness of

$$E(|z_i|^{1+\delta}); \qquad i = 1, 2, 3, \ldots.$$

The proof follows immediately from the inequalities

$$|x_i - a_i|^{1+\delta} \leqq 2^\delta\{|x_i|^{1+\delta} + |a_i|^{1+\delta}\}$$
$$|a_i|^{1+\delta} \leqq E(|x_i|^{1+\delta})$$

the first of which is well known; the second is a particular case of Liapounoff's inequality, established in Chap. XIII, page 265.

Thus, from the outset we are entitled to assume that

$$E(x_i) = 0.$$

The proof of the theorem is based on a very ingenious and useful device due to Markoff. Let N be a positive number which later we shall increase indefinitely. Together with x_i we shall consider two new variables, u_i and v_i, defined as follows: α being a particular value of x_i, the corresponding values of u_i and v_i are

$$u_i = \alpha, \qquad v_i = 0$$

if $|\alpha| \leqq N$ and

$$u_i = 0, \qquad v_i = \alpha$$

if $|\alpha| > N$. Thus, stochastic variables u_i and v_i are completely defined. Evidently

$$x_i = u_i + v_i$$

whence

$$0 = E(u_i) + E(v_i)$$

and

$$\beta_i = E(u_i) = -E(v_i).$$

Now

$$E(|v_i|^{1+\delta}) \leqq E(|x_i|^{1+\delta}) < c$$

by hypothesis. Since v_i is either 0 or its absolute value is $> N$, we have

$$N^\delta E(|v_i|) \leqq E(|v_i|^{1+\delta}) < c,$$

whence

(2) $$|\beta_i| = |E(v_i)| < \frac{c}{N^\delta}.$$

Likewise, the probability q_i for $v_i \neq 0$ satisfies the inequality

$$N^{1+\delta} q_i \leqq E(|v_i|^{1+\delta}) < c,$$

whence

(3) $$q_i < \frac{c}{N^{1+\delta}}.$$

Now, let us consider two inequalities

(4) $$\left|\frac{u_1 + u_2 + \cdots + u_n}{n}\right| < \sigma$$

(5) $$\left|\frac{x_1 + x_2 + \cdots + x_n}{n}\right| < \sigma$$

where σ is an arbitrary positive number and let P_0 and P be their respective probabilities. The inequalities (4) and (5) coincide when

$$v_1 = v_2 = \cdots = v_n = 0.$$

With this supplementary condition they have the same probability Q. But they can hold also when at least one of the numbers

$$v_1, v_2, \ldots v_n$$

is different from 0. Let the probabilities of (4) and (5) under such circumstances be R_0 and R. Then

$$P_0 = Q + R_0, \qquad P = Q + R.$$

But evidently neither R_0 nor R can exceed the probability that in the series

$$v_1, v_2, \ldots v_n$$

at least one number is different from 0; this probability in turn does not exceed (see Chap. II, page 30)

$$q_1 + q_2 + \cdots + q_n < \frac{nc}{N^{1+\delta}}.$$

Hence

$$R_0 < \frac{nc}{N^{1+\delta}}, \qquad R < \frac{nc}{N^{1+\delta}}$$

and

(6) $$|P - P_0| < \frac{nc}{N^{1+\delta}}.$$

On the other hand, since none of the values of $u_i (i = 1, 2, \ldots n)$ exceeds N, we have

$$E(u_i^2) \leqq N^{1-\delta}E(|u_i|^{1+\delta}) \leqq N^{1-\delta}E(|x_i|^{1+\delta}) < cN^{1-\delta}.$$

Accordingly, the dispersion of the sum $u_1 + u_2 + \cdots + u_n$ will be less than

$$cnN^{1-\delta}.$$

Hence, by what has been proved in Sec. 2, the probability of the inequality

(7) $$\left|\frac{u_1 + u_2 + \cdots + u_n}{n} - \frac{\beta_1 + \beta_2 + \cdots + \beta_n}{n}\right| \leqq \frac{\epsilon}{2}$$

is greater than

$$1 - \frac{4cN^{1-\delta}}{\epsilon^2 n}.$$

But whenever (7) is satisfied, the inequality

$$\left|\frac{u_1 + u_2 + \cdots + u_n}{n}\right| \leqq \frac{\epsilon}{2} + \frac{|\beta_1 + \beta_2 + \cdots + \beta_n|}{n} \tag{8}$$

is also satisfied. Hence, the probability of this inequality is a fortiori greater than

$$1 - \frac{4cN^{1-\delta}}{\epsilon^2 n}.$$

Owing to inequalities (2), the following inequality follows from (8):

$$\left|\frac{u_1 + u_2 + \cdots + u_n}{n}\right| < \frac{\epsilon}{2} + \frac{c}{N^\delta} = \sigma.$$

Hence

$$P_0 > 1 - \frac{4cN^{1-\delta}}{\epsilon^2 n},$$

and on account of (6)

$$P > 1 - \frac{4cN^{1-\delta}}{\epsilon^2 n} - \frac{nc}{N^{1+\delta}}.$$

Now we can dispose of the arbitrary number N by taking

$$N = \frac{n\epsilon}{2}.$$

Then

$$P > 1 - 2c\left(\frac{2}{\epsilon}\right)^{1+\delta} n^{-\delta}.$$

Now N tends to infinity with n and as soon as n surpasses a certain limit n_0, the fraction

$$\frac{c}{N^\delta}$$

will become and remain less than $\epsilon/2$. The probability of the inequality

$$\left|\frac{x_1 + x_2 + \cdots + x_n}{n}\right| < \epsilon$$

for $n > n_0$ will be greater than P and consequently greater than

$$1 - 2c\left(\frac{2}{\epsilon}\right)^{1+\delta} n^{-\delta}.$$

It tends, therefore, to 1 as n tends to infinity, and that proves Markoff's theorem.

Example. Let the possible values of the variable $x_p (p = 1, 2, 3, \ldots)$ be

$$p^{-1}(p+1)^{\frac{1}{2}},\ p^{-1}(p+1)^{\frac{2}{2}},\ p^{-1}(p+1)^{\frac{3}{2}},\ \ldots$$

with the corresponding probabilities

$$\frac{p}{p+1},\ \frac{p}{(p+1)^2},\ \frac{p}{(p+1)^3},\ \ldots.$$

Since the series

$$\frac{1}{p} + \frac{1}{p} + \frac{1}{p} + \cdots$$

is divergent, the mathematical expectation

$$E(x_p^2)$$

does not exist. Yet the law of large numbers holds. For

$$E(|x_p|^{1+\delta}) = p^{-\delta} \sum_{n=1}^{\infty} \frac{1}{(p+1)^{\frac{n}{2}(1-\delta)}}$$

is a convergent series for any $0 < \delta < 1$. Moreover,

$$p^{-\delta} \cdot \sum_{n=1}^{\infty} \frac{1}{(p+1)^{\frac{n}{2}(1-\delta)}} \leqq \frac{1}{2^{\frac{1-\delta}{2}} - 1},$$

and consequently the conditions of Markoff's theorem are satisfied for any $0 < \delta < 1$. Hence, the law of large numbers holds in this example.

9. If variables $x_1, x_2, x_3, \ldots$ are identical, the law of large numbers holds without any other restrictions, except that for these variables mathematical expectations exist. In fact, Khintchine proved the following theorem:

Theorem. *If, as we may naturally suppose, $E(x_i) = 0$, the probability of the inequality*

$$\left|\frac{x_1 + x_2 + \cdots + x_n}{n}\right| \leqq \epsilon$$

tends to 1 as n increases indefinitely.

Proof. The proof is quite similar to that of Markoff's theorem and is based on the same ingenious artifice. Let

$$\cdots < \alpha_{-2} < \alpha_{-1} < \alpha_0 < \alpha_1 < \alpha_2 < \cdots$$

be different values of any one of the identical variables $x_1, x_2, x_3, \ldots$ and

$$\ldots, p_{-2}, p_{-1}, p_0, p_1, p_2, \ldots$$

their probabilities. By hypothesis

$$\Sigma p_i \alpha_i$$

is a convergent series with the sum 0. The series

$$\Sigma p_i |\alpha_i|$$

is also convergent; let $c > 0$ be its sum.

Keeping the same notations as before, we have

$$|\beta_i| \leqq E(|v_i|) = \sum_{|\alpha_i| > N} p_i |\alpha_i| = \psi(N)$$

where $\psi(N)$ is a decreasing function tending to 0 as $N \rightarrow \infty$. Also

$$E(u_i^2) \leqq NE|x_i| = cN$$

so that the dispersion of the sum

$$u_1 + u_2 + \cdot \cdot \cdot + u_n$$

is less than

$$cNn.$$

Consequently the probability of the inequality

$$\left|\frac{u_1 + u_2 + \cdot \cdot \cdot + u_n}{n} - \frac{\beta_1 + \beta_2 + \cdot \cdot \cdot + \beta_n}{n}\right| \leqq \frac{\epsilon}{2} \tag{9}$$

is greater than

$$1 - \frac{4cN}{\epsilon^2 n}.$$

On the other hand, the probability q_s of the inequality $v_s \gtrless 0$ is less than

$$\frac{\psi(N)}{N}$$

because

$$N \sum_{|\alpha_i| > N} p_i < \psi(N)$$

and

$$q_s = \sum_{|\alpha_i| > N} p_i.$$

Hence, the difference between the probability of the inequality

$$\left|\frac{u_1 + u_2 + \cdot \cdot \cdot + u_n}{n}\right| < \sigma$$

and that of the inequality

$$\left|\frac{x_1 + x_2 + \cdots + x_n}{n}\right| < \sigma$$

is numerically less than

$$\frac{n\psi(N)}{N}.$$

As in the preceding section we conclude that the probability of the inequality

$$\left|\frac{u_1 + u_2 + \cdots + u_n}{n}\right| \leqq \frac{\epsilon}{2} + \psi(N)$$

is greater than

$$1 - \frac{4cN}{\epsilon^2 n}.$$

Finally, the probability of the inequality

$$\left|\frac{x_1 + x_2 + \cdots + x_n}{n}\right| \leqq \frac{\epsilon}{2} + \psi(N) \tag{10}$$

is greater than

$$1 - \frac{4cN}{\epsilon^2 n} - \frac{n\psi(N)}{N}.$$

To dispose of N we observe that the ratio

$$\frac{\sqrt{\psi(N)}}{N}$$

is a decreasing function of N and tends to 0 as $N \to \infty$. Hence, at least for large n, there exists an integer N such that

$$\frac{\sqrt{\psi(N)}}{N} < \frac{\sqrt{4c}}{\epsilon n} \leqq \frac{\sqrt{\psi(N-1)}}{N-1}.$$

Then

$$\frac{n\psi(N)}{N} < \frac{\sqrt{4c}}{\epsilon}\sqrt{\psi(N)}; \qquad \frac{4cN}{\epsilon^2 n} \leqq \frac{\sqrt{4c}}{\epsilon} \cdot \frac{N}{N-1}\sqrt{\psi(N-1)}$$

whence it follows that the probability of inequality (10) is greater than

$$1 - \frac{\sqrt{4c}}{\epsilon}\left[\sqrt{\psi(N)} + \frac{N}{N-1}\sqrt{\psi(N-1)}\right].$$

Now N increases indefinitely together with n; therefore, for all n above a certain limit n_0,

$$\psi(N) < \frac{\epsilon}{2}$$

so that for $n > n_0$ the probability of the inequality

$$\left|\frac{x_1 + x_2 + \cdots + x_n}{n}\right| < \epsilon$$

will be greater than

$$1 - \frac{\sqrt{4c}}{\epsilon}\left[\sqrt{\psi(N)} + \frac{N}{N-1}\sqrt{\psi(N-1)}\right]$$

and with indefinitely increasing n will approach the limit 1. Thus Khintchine's theorem is completely proved.

Example. Let

$$2^{1-2\log 1},\ 2^{2-2\log 2},\ 2^{3-2\log 3},\ \ldots\ 2^{n-2\log n},\ \ldots$$

be all possible values of identical variables $x_1, x_2, x_3, \ldots$ and

$$\frac{1}{2}, \frac{1}{2^2}, \frac{1}{2^3}, \ldots \frac{1}{2^n}, \ldots$$

their corresponding probabilities. Since the series

$$\frac{1}{2^{2\log 1}} + \frac{1}{2^{2\log 2}} + \frac{1}{2^{2\log 3}} + \cdots = 1 + \frac{1}{2^{\log 4}} + \frac{1}{3^{\log 4}} + \cdots$$

is convergent, mathematical expectations of the variables $x_1, x_2, x_3, \ldots$ exist. Hence, the law of large numbers holds in this case.

Markoff's theorem cannot be applied here, because for any positve δ the series

$$\sum_1^\infty \frac{2^{n\delta}}{n^{(1+\delta)\log 4}}$$

is divergent.

Problems for Solution

1. Let x be a stochastic variable with the mean $= 0$ and the standard deviation σ. Denoting by $P(t)$ the probability of the inequality

$$x \leqq t$$

show that

$$P(t) \leqq \frac{\sigma^2}{\sigma^2 + t^2} \qquad \text{for} \qquad t < 0$$

$$1 - P(t) \leqq \frac{\sigma^2}{\sigma^2 + t^2} \qquad \text{for} \qquad t > 0.$$

Show also that the right-hand members cannot be replaced by smaller numbers.

Indication of the Proof. Since

$$\Sigma p_i x_i = 0, \qquad \Sigma p_i x_i^2 = \sigma^2,$$

we have also

$$\Sigma p_i(x_i - t) = -t, \qquad \Sigma p_i(x_i - t)^2 = \sigma^2 + t^2$$

whence, supposing that $x_i > t$ for $i = 1, 2, \ldots s$ and first taking t negative,

$$t^2 \leqq \left\{ \sum_{i=1}^{s} p_i(x_i - t) \right\}^2 \leqq \sum_{i=1}^{s} p_i \sum_{i=1}^{s} p_i(x_i - t)^2 \leqq (1 - P(t))(\sigma^2 + t^2)$$

$$1 - P(t) \geqq \frac{t^2}{\sigma^2 + t^2}; \qquad P(t) \leqq \frac{\sigma^2}{\sigma^2 + t^2}.$$

For positive t the proof is quite similar. Considering a stochastic variable with two values:

$$x_1 = t, \qquad p_1 = \frac{\sigma^2}{\sigma^2 + t^2}$$

$$x_2 = -\frac{\sigma^2}{t}, \qquad p_2 = \frac{t^2}{\sigma^2 + t^2}$$

one can easily prove the last part of our statement.

2. *Tshebysheff's Problem.*[1] If x is a positive stochastic variable with given

$$E(x) = \sigma^2, \qquad E(x^2) = \tau^4$$

then the probability P of the inequality

$$x \geqq v$$

has the following precise upper bounds:

$$P \leqq 1 \qquad \text{for} \qquad v < \sigma^2$$

$$P \leqq \frac{\sigma^2}{v} \qquad \text{for} \qquad \sigma^2 \leqq v < \frac{\tau^4}{\sigma^2}$$

$$P \leqq \frac{\tau^4 - \sigma^4}{\tau^4 + v^2 - 2\sigma^2 v} \qquad \text{for} \qquad v \geqq \frac{\tau^4}{\sigma^2}.$$

Indication of the Proof. Let

$$\xi = \frac{\sigma^2 v - \tau^4}{v - \sigma^2}.$$

Then $\xi < v$ if $v \geqq \tau^4/\sigma^2$ and

$$P \leqq E\left(\frac{x - \xi}{v - \xi}\right)^2$$

since

$$\left(\frac{x - \xi}{v - \xi}\right)^2 \geqq 1$$

for $x \geqq v$. On the other hand,

$$E\left(\frac{x - \xi}{v - \xi}\right)^2 = \frac{\tau^4 - 2\sigma^2\xi + \xi^2}{(v - \xi)^2} = \frac{\tau^4 - \sigma^4}{\tau^4 + v^2 - 2\sigma^2 v}$$

whence

$$P \leqq \frac{\tau^4 - \sigma^4}{\tau^4 + v^2 - 2\sigma^2 v}.$$

[1] Sur les valeurs limites des integrales, *Jour. Liouville*, Ser. 2, T. XIX, 1874.

The equality sign is reached for the stochastic variable with two values:

$$x_1 = \xi, \qquad p_1 = \frac{(v - \sigma^2)^2}{\tau^4 + v^2 - 2\sigma^2 v}$$

$$x_2 = v, \qquad p_2 = \frac{\tau^4 - \sigma^4}{\tau^4 + v^2 - 2\sigma^2 v}.$$

If $\sigma^2 \leqq v < \tau^4/\sigma^2$ we have an obvious inequality

$$P \leqq E\left(\frac{x}{v}\right) = \frac{\sigma^2}{v}.$$

To show that the right-hand member cannot be replaced by a smaller number, consider the following stochastic variable with three values:

$$x_1 = 0, \qquad p_1 = \frac{(l - \sigma^2)v - l\sigma^2 + \tau^4}{lv}$$

$$x_2 = v, \qquad p_2 = \frac{l\sigma^2 - \tau^4}{v(l - v)}$$

$$x_3 = l, \qquad p_3 = \frac{\tau^4 - \sigma^2 v}{l(l - v)}$$

where $l > v$ is an arbitrary number. For this variable

$$P = p_2 + p_3 = \frac{\sigma^2}{v} - \frac{\tau^4 - \sigma^2 v}{lv}$$

is arbitrarily near to σ^2/v for sufficiently large l.

3. If x is an arbitrary stochastic variable with given

$$E(x^2) = \sigma^2, \qquad E(x^4) = \tau^4$$

and P denotes the probability of the inequality

$$|x| \geqq k\sigma,$$

then

$$P \leqq \frac{1}{k^2} \qquad \text{if} \qquad 1 \leqq k < \left(\frac{\tau}{\sigma}\right)^2$$

$$P \leqq \frac{\left(\frac{\tau}{\sigma}\right)^4 - 1}{\left(\frac{\tau}{\sigma}\right)^4 + k^4 - 2k^2} \qquad \text{if} \qquad k \geqq \left(\frac{\tau}{\sigma}\right)^2.$$

These inequalities cannot be improved.

HINT: Follows from Tshebysheff's problem.

4. Let x_i assume two values, i and $-i$ with equal probabilities. Show that the law of large numbers cannot be applied to variables $x_1, x_2, x_3, \ldots$.

5. Variables $x_1, x_2, x_3, \ldots$ each assume two values:
$\log a$ or $-\log a$; $\log(a + 1)$ or $-\log(a + 1)$; $\log(a + 2)$ or $-\log(a + 2)$; $\cdots$ with equal probabilities. Show that the law of large numbers holds for these variables.

HINT: $E(x_i) = 0; i = 1, 2, 3, \ldots$

$$B_n = E(x_1 + x_2 + \cdots + x_n)^2 =$$
$$= \sum_{i=0}^{n-1} \{\log (a + i)\}^2 \sim (a + n - 1)\{\log (a + n - 1)\}^2$$

as can easily be established by using Euler's summation formula (Appendix 1, page 347). Hence

$$\frac{B_n}{n^2} \to 0 \quad \text{as} \quad n \to \infty.$$

6. If x_i can have only two values with equal probabilities, i^α and $-i^\alpha$, show that the law of large numbers can be applied to $x_1, x_2, x_3, \ldots$ if $\alpha < \frac{1}{2}$.

HINT:

$$B_n = 1^{2\alpha} + 2^{2\alpha} + \cdots + n^{2\alpha} \sim \frac{n^{2\alpha+1}}{2\alpha + 1}, \qquad \frac{B_n}{n^2} \to 0 \quad \text{if} \quad \alpha < \frac{1}{2}.$$

It can be shown that the law of large numbers does not hold if $\alpha \geqq \frac{1}{2}$.

7. In an indefinite Bernoullian series of trials with the constant probability p, let m_i denote the number of successes in the first i trials. Show that the law of large numbers holds for variables

$$x_i = \frac{m_i - ip}{(ipq)^\alpha}; \qquad i = 1, 2, 3, \ldots$$

if $\alpha > \frac{1}{2}$.

HINT: Evidently $E(x_i) = 0$, $E(x_i^2) = (ipq)^{1-2\alpha}$ and

$$B_n = \sum_{i=1}^{n} (ipq)^{1-2\alpha} + 2\sum_{j>i} E(x_i x_j).$$

Now

$$E(x_i x_j) = (ij)^{-\alpha}(pq)^{-2\alpha}E(m_i - ip)^2 + (ij)^{-\alpha}(pq)^{-2\alpha}E\{(m_i - ip)(m_j - m_i - (j-i)p)\} =$$
$$= (pq)^{1-2\alpha}i^{1-\alpha}j^{-\alpha}$$

since $m_i - ip$ and $m_j - m_i - (j - i)p$ are independent variables. Thus

$$B_n = (pq)^{1-2\alpha}\left[\sum_{i=1}^{n} i^{1-2\alpha} + 2\sum_{j>i} i^{1-\alpha}j^{-\alpha}\right]$$

and it is easy to show that

$$\frac{B_n}{n^2} \to 0 \quad \text{as} \quad n \to \infty$$

provided $\alpha > \frac{1}{2}$. But the law of large numbers no longer holds if $\alpha \leqq \frac{1}{2}$. The proof of this is more difficult.

8. The following extension of Tshebysheff's lemma was indicated by Kolmogoroff. Let $x_1, x_2 \ldots x_n$ be independent variables; $E(x_i) = 0$, $E(x_i^2) = b_i$,

$$B_n = b_1 + b_2 + \cdots + b_n$$

and

$$s_k = x_1 + x_2 + \cdots + x_k; \qquad k = 1, 2, \ldots n.$$

Denoting by P the probability of the inequality

$$(A) \qquad \max.\ (s_1^2, s_2^2, \ldots s_n^2) > B_n t^2,$$

we shall have $P < 1/t^2$.

Indication of the Proof. The inequality (A) can materialize if and only if one of the following mutually exclusive events occurs:

event e_1: $s_1^2 > B_n t^2$;
event e_2: $s_1^2 \leqq B_n t^2$; $\quad s_2^2 > B_n t^2$;
event e_3: $s_1^2 \leqq B_n t^2$; $\quad s_2^2 \leqq B_n t^2$; $\quad s_3^2 > B_n t^2$;
. . . .
event e_n: $s_1^2 \leqq B_n t^2$; $\quad s_2^2 \leqq B_n t^2$; $\cdots s_{n-1}^2 \leqq B_n t^2$; $\quad s_n^2 > B_n t^2$.

If (e_i) represents the probability of $e_i (i = 1, 2, \ldots n)$ then

$$P = (e_1) + (e_2) + \cdots + (e_n).$$

Now consider the conditional mathematical expectation $E(s_n^2|e_k)$ of s_n^2 given that e_k has occurred. Since the indication of e_k does not affect variables $x_{k+1}, x_{k+2}, \ldots x_n$, these variables and s_k are independent. Hence

$$E(s_n^2|e_k) = E(s_k^2|e_k) + b_{k+1} + \cdots + b_n > B_n t^2.$$

On the other hand

$$B_n = E(s_n^2) = \sum_{k=1}^{n} (e_k) E(s_n^2|e_k) > B_n t^2 \{(e_1) + (e_2) + \cdots + (e_n)\}$$

whence $P < 1/t^2$.

9. *The Strong Law of Large Numbers (Kolmogoroff).* Using the same notations as in the preceding problem, show that the probability of the simultaneous inequalities

$$\left|\frac{s_n}{n}\right| \leqq \epsilon, \qquad \left|\frac{s_{n+1}}{n+1}\right| \leqq \epsilon, \qquad \left|\frac{s_{n+2}}{n+2}\right| \leqq \epsilon, \ldots$$

will be greater than $1 - \eta$, provided n exceeds a certain limit depending on the choice of ϵ and η, and granted the convergence of the series

$$\sum_{1}^{\infty} \frac{b_n}{n^2}.$$

Indication of the Proof. Consider variables

$$\tau_i = \max. \left|\frac{s_m - s_{\nu_{i-1}}}{\nu_i}\right| \text{ for } \nu_i = 2^{i-1}n \leqq m < 2^i n; \quad i = 1, 2, 3, \ldots$$

and denote by q_i the probability of the inequality $\tau_i > \frac{1}{2}\epsilon$. By Kolmogoroff's lemma

$$q_i < \frac{4 \sum_{l=2^{i-1}n}^{l=2^i n - 1} b_l}{2^{2i-2} n^2 \epsilon^2}$$

and

$$q_1 + q_2 + q_3 + \cdots < 4\epsilon^{-2}\sum_{i=1}^{\infty}\frac{1}{2^{2i-2}n^2}\sum_{l=2^{i-1}n}^{l=2^in-1} b_l < 16\epsilon^{-2}\sum_{i=1}^{\infty}\sum_{l=2^{i-1}n}^{l=2^in-1}\frac{b_l}{l^2}$$

or

$$q_1 + q_2 + q_3 + \cdots < 16\epsilon^{-2}\sum_{k=n}^{\infty}\frac{b_k}{k^2}.$$

Hence, the probability of fulfillment of all the inequalities $\tau_i \leqq \frac{1}{2}\epsilon;\ i = 1, 2, 3, \ldots$ is greater than

$$1 - 16\epsilon^{-2}\sum_{k=n}^{\infty}\frac{b_k}{k^2}.$$

The inequalities $|s_k/k| \leqq \epsilon;\ k = n, n + 1, n + 2, \ldots$ are satisfied when simultaneously

$$\tau_i \leqq \tfrac{1}{2}\epsilon; \qquad i = 1, 2, 3, \ldots$$

and

$$\left|\frac{s_{n-1}}{n}\right| \leqq \frac{1}{2}\epsilon.$$

The probability of the last inequality being greater than $1 - \frac{4B_n}{n^2\epsilon^2}$, the probability of simultaneous inequalities

$$\left|\frac{s_k}{k}\right| \leqq \epsilon; \qquad k = n, n + 1, n + 2, \ldots$$

a fortiori will be greater than

$$1 - 16\epsilon^{-2}\sum_{k=n}^{\infty}\frac{b_k}{k^2} - \frac{4B_n}{n^2\epsilon^2}.$$

This inequality suffices to complete the proof if we notice that B_n/n^2 tends to 0 when the series

$$\sum_{k=1}^{\infty}\frac{b_k}{k^2}$$

is convergent.

10. Let $x_1, x_2, \ldots x_n$ be identical stochastic variables and $E(x_1) = 0$. Denoting by $\overset{+}{P}_n(\epsilon)$ and $\overset{-}{P}_n(\epsilon)$, respectively, the probabilities of the inequalities

$$\frac{x_1 + x_2 + \cdots + x_n}{n} > \epsilon \qquad \text{and} \qquad \frac{x_1 + x_2 + \cdots + x_n}{n} < -\epsilon$$

show that

$$\lim_{n=\infty}\frac{\overset{+}{P}_n(\epsilon)}{\overset{-}{P}_n(\epsilon)} = 0 \text{ or } +\infty$$

according as $E(x_1^3) >$ or <0.

For the proof see Khintchine's paper in *Mathematische Annalen* (vol. 101, pp. 381–385).

11. *The Law of the Repeated Logarithm* (*Khintchine, Kolmogoroff*). Let $x_1, x_2, \ldots x_n$ be bounded independent variables, $E(x_i) = 0, i = 1, 2, \ldots n$ and $B_n \to \infty$ as $n \to \infty$. For an arbitrarily small $\delta > 0$ and $\epsilon > 0$ and for an arbitrarily large N one can choose $n_0 > N$ so that:

a. The probability of the fulfillment of the inequality

$$|s_n| > (1 + \delta)\sqrt{2B_n \log \log B_n}$$

for at least one $n \geqq n_0$ is less than ϵ.

b. The probability of the fulfillment of the inequality

$$|s_n| > (1 - \delta)\sqrt{2B_n \log \log B_n}$$

for at least one $n \geqq n_0$ is greater than $1 - \epsilon$.

For the proof see Kolmogoroff's paper in *Mathematische Annalen* (vol. 101, pp. 126–135).

If $x_1, x_2, \ldots x_n$ are variables independent in pairs and B_n the dispersion of their sum $s = x_1 + x_2 + \cdots + x_n$, then the probability P that

$$|s| \leqq t\sqrt{B_n}$$

satisfies the inequality

$$P > 1 - \frac{1}{t^2} \qquad \text{(Tshebysheff's inequality)}$$

provided $E(x_i) = 0, i = 1, 2, \ldots n$, which can be assumed without loss of generality. In case variables are *totally* independent and are subject to certain limitations of comparatively mild character, S. Bernstein has shown that Tshebysheff's inequality can be considerably improved.

12. Let $x_1, x_2, \ldots x_n$ be totally independent variables. We suppose $E(x_i) = 0$, $E(x_i^2) = b_i$ and

$$E(|x_i|^h) \leqq \frac{b_i}{2}h!c^{h-2}$$

for $i = 1, 2, \ldots n$ and $h > 2$, c being a certain constant. Show that

$$A = E\{e^{\epsilon(x_1+x_2+\cdots+x_n)}\} < e^{\frac{B_n\epsilon^2}{2(1-\sigma)}}$$

where σ is an arbitrary positive number <1 and ϵ is a positive number so small that $\epsilon c \leqq \sigma$.

Indication of the Proof. We have

$$e^{\epsilon x_i} \leqq 1 + \epsilon x_i + \sum_{n=2}^{\infty} \frac{\epsilon^n |x_i|^n}{n!}$$

whence

$$E(e^{\epsilon x_i}) \leqq 1 + \frac{b_i}{2}\epsilon^2 \sum_{n=0}^{\infty} (\epsilon c)^n < e^{\frac{b_i\epsilon^2}{2(1-\sigma)}}.$$

13. If Q denotes the probability of the inequality

$$x_1 + x_2 + \cdots + x_n > \frac{B_n\epsilon}{2(1-\sigma)} + \frac{t^2}{\epsilon}$$

show that $Q < e^{-t^2}$.

Indication of the Proof. If $\overline{Q}$ is the probability of the inequality

$$e^{\epsilon(x_1+x_2+\cdots+x_n)} > Ae^{t^2},$$

then, by Tshebysheff's lemma, $\overline{Q} < e^{-t^2}$ and $Q < \overline{Q}$ by Prob. 12.

14. *S. Bernstein's Inequality.* Denoting by P the probability of the inequality

$$|x_1 + x_2 + \cdots + x_n| \leqq \omega,$$

ω being a given positive number, show that

$$P > 1 - 2e^{-\frac{\omega^2}{2B_n+2c\omega}}.$$

Indication of the Proof. To make $\frac{B_n\epsilon}{2(1-\sigma)} + \frac{t^2}{\epsilon} = F$ minimum take $\epsilon = \frac{t\sqrt{2(1-\sigma)}}{\sqrt{B_n}}$; then $F = t\sqrt{\frac{2B_n}{1-\sigma}}$ and t is determined by equating F to ω. The resulting value of ϵ,

$$\epsilon = \frac{\omega}{B_n}(1-\sigma)$$

is admissible only if $\epsilon c \leqq \sigma$ or $\frac{c\omega}{B_n}(1-\sigma) \leqq \sigma$. The best choice for σ is $\sigma = \frac{c\omega}{B_n + c\omega}$ and correspondingly $t = \frac{\omega}{\sqrt{2B_n + 2c\omega}}$. By Prob. 13 the probability of the inequality

$$x_1 + x_2 + \cdots + x_n > \omega$$

is less than $e^{-\frac{\omega^2}{2B_n+2c\omega}}$. The same is true of the probability of the inequality

$$x_1 + x_2 + \cdots + x_n < -\omega \quad \text{or} \quad -x_1 - x_2 - \cdots - x_n > \omega.$$

15. If variables $x_1, x_2, \ldots x_n$ are uniformly bounded and M is an upper bound of their numerical values, then we may take $c = M/3$.

Indication of the Proof. Note that

$$E(|x_i|^h) \leqq M^{h-2}b_i, \qquad b_iM^{h-2} \leqq \frac{b_i}{2}h!\left(\frac{M}{3}\right)^{h-2}.$$

16. Consider a Poisson's series of trials with probabilities $p_1, p_2, \ldots p_n$ for an event E to occur. Let m be the frequency of E in n trials, $p = \frac{p_1 + p_2 + \cdots + p_n}{n}$, $\lambda = \frac{1}{n}(p_1q_1 + p_2q_2 + \cdots + p_nq_n)$. Show that the probability P of the inequality $\left|\frac{m}{n} - p\right| \leqq \epsilon$ has the following lower limit:

$$P > 1 - 2e^{-\frac{n\epsilon^2}{2\lambda+\frac{2}{3}\epsilon}}.$$

In the Bernoullian case $p_1 = p_2 = \cdots = p_n$, $\lambda = pq$ and consequently

$$P > 1 - 2e^{-\frac{n\epsilon^2}{2pq+\frac{2}{3}\epsilon}}.$$

17. An indefinite series of totally independent variables $x_1, x_2, x_3, \ldots$ has the property that the mathematical expectations of any odd power of these variables is rigorously $= 0$ while

$$E(x_i^{2k}) \leqq \left(\frac{b_i}{2}\right)^k \frac{(2k)!}{k!}; \qquad b_i = E(x_i^2)$$

for $i = 1, 2, 3, \ldots$. Prove that the probability of either one of the inequalities

$$x_1 + x_2 + \cdots + x_n > t\sqrt{2B_n} \qquad \text{or} \qquad x_1 + x_2 + \cdots + x_n < -t\sqrt{2B_n}$$

where $B_n = b_1 + b_2 + \cdots + b_n$ is less than e^{-t^2} (S. Bernstein). Prove first that

$$E(e^{\epsilon x_i}) \leqq e^{\frac{\epsilon^2 b_i}{2}}.$$

18. Positive and negative proper decimal fractions limited to, say, five decimals, are obtained in the following manner: From an urn containing tickets with numbers 0, 1, 2, . . . 9 in equal proportion, five tickets are drawn in succession (the ticket drawn in a previous trial being returned before the next) and their respective numbers are written in succession as five decimals of a proper fraction. This fraction, if not equal to 0, is preceded by the sign + or −, according as a coin tossed at the same time shows heads or tails. Thus, repeating this process several times, we may obtain as many positive or negative proper fractions with five decimals as we desire. What can be said about the probability that the sum of n such fractions will be contained between prescribed limits $-\omega$ and ω? *Ans.* These n fractions may be considered as so many identical stochastic variables for each of which

$$E(x^{2k+1}) = 0, \qquad \beta = E(x^2) = \frac{(1 - 10^{-5})(2 - 10^{-5})}{6} < \frac{1}{3}.$$

Besides,

$$E(x^{2k}) = \frac{\sum_{l=1}^{10^5-1} l^{2k}}{10^{10k+5}} < \frac{1}{2k+1},$$

since in general

$$1^{2k} + 2^{2k} + \cdots + (s-1)^{2k} < \frac{s^{2k+1}}{2k+1}.$$

Again, the inequality

$$E(x^{2k}) \leqq \left(\frac{\beta}{2}\right)^k \frac{(2k)!}{k!}$$

can easily be verified and we can apply the result of Prob. 17. For the required probability P the following lower limit can be obtained:

$$P > 1 - 2e^{-\frac{\omega^2}{2n\beta}} > 1 - 2e^{-\frac{3\omega^2}{2n}};$$

or, if $\omega = n\epsilon$

$$P > 1 - 2e^{-\frac{3}{2}n\epsilon^2}.$$

For example, if $\epsilon = \frac{1}{10}$ and $n \geqq 814$,

$$P > 0.99999,$$

that is, almost certainly the sum of 814 fractions formed in the above described manner will be contained between -82 and 82.

References

P. L. Tshebysheff: Sur les valeurs moyennes, Oeuvres I, pp. 687–694.
A. Markoff: "Calculus of Probability," 4th Russian edition, Leningrad, 1924.
S. Bernstein: "Theory of Probability" (Russian), Moscow, 1927.
A. Khintchine: Sur la loi des grands nombres, *C.R.* 188, 1929.

CHAPTER XI

APPLICATIONS OF THE LAW OF LARGE NUMBERS

1. A theorem of such wide generality as the law of large numbers is a source of a great many important particular theorems. We shall begin with a generalization of Bernoulli's theorem due to Poisson.

Let us consider a series of independent trials with the respective probabilities $p_1, p_2, p_3, \ldots$, varying from one trial to another. Considering n trials, we shall denote by m the number of successes. The arithmetic mean of probabilities in n trials

$$p = \frac{p_1 + p_2 + \cdots + p_n}{n}$$

will be called the "mean probability in n trials." With such conditions and notations adopted, we can state Poisson's theorem as follows:

Poisson's Theorem. *The probability of the inequality*

$$\left|\frac{m}{n} - p\right| \leqq \epsilon$$

for fixed $\epsilon > 0$, no matter how small, can be made as near to 1 (certainty) as we please, provided the number of trials n is sufficiently large.

Proof. To show that this theorem is but a particular case of the law of large numbers, we use an artifice often applied in similar circumstances, namely, we associate with trials 1, 2, 3, . . . n variables $x_1, x_2, x_3, \ldots x_n$ defined as follows:

$$x_i = 1 \text{ in case of success in the } i\text{th trial,}$$
$$x_i = 0 \text{ in case of failure in the } i\text{th trial.}$$

Since the trials are independent, these variables are also independent. Moreover

$$E(x_i) = E(x_i^2) = p_i$$

and the dispersion of x_i is

$$p_i - p_i^2 = p_i q_i.$$

The dispersion B_n of the sum

$$x_1 + x_2 + \cdots + x_n$$

is the sum of the dispersions of its terms, that is,

$$B_n = p_1q_1 + p_2q_2 + \cdots + p_nq_n \leqq \frac{n}{4}.$$

At the same time, the former sum represents the number of successes m.

Now, applying the results established in Chap. X, Sec. 2, we arrive at this conclusion: Denoting by P the probability of the inequality

$$\left|\frac{m}{n} - p\right| \leqq \epsilon,$$

we shall have

$$P > 1 - \frac{B_n}{n^2\epsilon^2} \geqq 1 - \frac{1}{4n\epsilon^2}.$$

It now suffices to take

$$n > \frac{1}{4\epsilon^2\eta}$$

to have

$$P > 1 - \eta$$

where η is an arbitrary positive number no matter how small. That completes the proof of Poisson's theorem.

Evidently Bernoulli's theorem is contained in Poisson's theorem as a particular case when

$$p_1 = p_2 = \cdot \cdot \cdot = p_n = p.$$

Poisson himself attached great importance to his theorem and adopted for it the name of the "law of large numbers," which is still used by many authors. However, it appears more proper to reserve this name to the theorem established in Chap. X, Sec. 2, which is due to Tshebysheff.

2. Let us consider n series each consisting of s independent trials with the constant probability p. Also, let

$$m_1, m_2, \ldots m_n$$

represent the number of successes in each of these s series. Stochastic variables

$$x_1 = (m_1 - sp)^2, \qquad x_2 = (m_2 - sp)^2, \cdot \cdot \cdot x_n = (m_n - sp)^2$$

are independent and identical. Their common mathematical expectation is spq. The law of large numbers can be applied to these variables and leads immediately to the conclusion: The probability of the inequality

$$\left|\frac{\sum_{i=1}^{n}(m_i - sp)^2}{n} - spq\right| < \epsilon$$

can be brought as near as we please to 1 (or certainty) if the number of series n is sufficiently large.

Substituting ϵspq for ϵ and dividing through by spq, we may state the same proposition as follows: The probability of the inequalities

$$1 - \epsilon < \frac{\sum_{i=1}^{n}(m_i - sp)^2}{Npq} < 1 + \epsilon,$$

where $N = ns$ is the total number of trials in all n series, can be brought as near to 1 as we please if the number of series is sufficiently large.

The law of large numbers can be legitimately applied to the variables

$$x_i = |m_i - sp|; \qquad i = 1, 2, 3, \ldots$$

with the common mathematical expectation

$$M_s = 2spqC_{s-1}^{\mu-1}p^{\mu-1}q^{s-\mu}$$

where $\mu = [sp + 1]$, and leads to the following proposition: The probability of the inequalities

$$1 - \epsilon < \frac{\sum_{i=1}^{n}|m_i - sp|}{nM_s} < 1 + \epsilon$$

can be brought as near to 1 as we please if the number of series is sufficiently large.

For the sake of simplicity, let us use the notations

$$A^2 = \frac{\sum_{i=1}^{n}(m_i - sp)^2}{n}$$

$$B = \frac{\sum_{i=1}^{n}|m_i - sp|}{n}.$$

The probabilities P and P' of the inequalities

(1) $$\sqrt{spq}(1 - \sigma) < A < \sqrt{spq}(1 + \sigma)$$

(2) $$M_s(1 - \sigma) < B < M_s(1 + \sigma)$$

which are equivalent to

$$(1 - \sigma)^2 < \frac{\sum_{i=1}^{n}(m_i - sp)^2}{nspq} < (1 + \sigma)^2$$

$$1 - \sigma < \frac{\sum_{i=1}^{n}|m_i - sp|}{nM_s} < 1 + \sigma$$

can both be made greater than $1 - \eta$, where η is an arbitrarily small positive number. The probability of simultaneous materialization of (1) and (2) is not less than

$$P + P' - 1 > 1 - 2\eta.$$

But whenever (1) and (2) hold simultaneously, we have

$$\frac{\sqrt{spq}}{M_s}\frac{1-\sigma}{1+\sigma} < \frac{A}{B} < \frac{\sqrt{spq}}{M_s}\frac{1+\sigma}{1-\sigma}. \tag{3}$$

Therefore the probability of these inequalities is again $>1 - 2\eta$. Now let us take

$$\sigma = \frac{\tau}{2+\tau}$$

where τ is another positive number arbitrarily chosen. Then

$$\frac{1+\sigma}{1-\sigma} = 1 + \tau; \qquad \frac{1-\sigma}{1+\sigma} > 1 - \tau.$$

Hence, the inequalities

$$\frac{\sqrt{spq}}{M_s}(1-\tau) < \frac{A}{B} < \frac{\sqrt{spq}}{M_s}(1+\tau)$$

follow from inequalities (3) and their probability is a fortiori $> 1 - 2\eta$. It suffices to take

$$\tau = \frac{M_s}{\sqrt{spq}}\epsilon$$

to arrive at the following proposition:

The probability of the inequality

$$\left|\frac{A}{B} - \frac{\sqrt{spq}}{M_s}\right| < \epsilon$$

for a fixed ϵ and sufficiently large number of series can be made as near to 1 as we please.

If spq is somewhat large, the quotient

$$\frac{\sqrt{spq}}{M_s}$$

differs but little from $\sqrt{\pi/2}$ (see Chap. IX, Prob. 2, page 177). Hence, when the number of series is large and the series themselves sufficiently long, we may expect with great probability that the quotient

$$\frac{A}{B}$$

will not differ much from $\sqrt{\pi/2}$.

DIVERGENCE COEFFICIENT

3. The considerations of the preceding section can be generalized. Let us consider again n series containing s trials each, and let

$$m_1, m_2, \ldots m_n$$

represent the numbers of successes in each of these series. Without specifying the nature of the trials (which can be independent or dependent) we shall denote by p the mean probability in all $N = ns$ trials and by $q = 1 - p$ its complement. Again considering the quotient

$$Q = \frac{\sum_{i=1}^{n}(m_i - sp)^2}{Npq},$$

we seek its mathematical expectation

$$E(Q) = D.$$

When all the N trials are of the Bernoullian type, $D = 1$. But it is also possible to imagine cases when $D > 1$ or $D < 1$. Lexis calls $\sqrt{D}$ the "coefficient of dispersion." We shall call D itself the "theoretical divergence coefficient." If $m_1, m_2, \ldots m_n$ are actually observed frequencies in n series, the quotient

$$D' = \frac{\sum_{i=1}^{n}(m_i - sp)^2}{Npq}$$

may be called "empirical divergence coefficient." Then, if the law of large numbers can be applied to variables

$$x_i = \frac{(m_i - sp)^2}{spq}; \qquad i = 1, 2, 3, \ldots,$$

we can expect with probability, approaching certainty as near as we please, that the inequality

$$|D' - D| < \epsilon$$

will be fulfilled for an adequately large number of series.

Thus far we have not specified the nature of the trials. Now we shall suppose that all $N = ns$ trials, distributed in n series, are independent but with probabilities varying in general from trial to trial. Let

$$p_{1i}, p_{2i}, \ldots p_{si} \qquad (i = 1, 2, \ldots n)$$

be the probabilities in successive trials of the ith series. Their mean

$$p_i = \frac{p_{1i} + p_{2i} + \cdots + p_{si}}{s}$$

is the mean probability in the ith series. Finally

$$p = \frac{p_1 + p_2 + \cdots + p_n}{n}$$

is the mean probability in all $N = ns$ trials. As to the expectation of $(m_i - sp)^2$, we find

$$E(m_i - sp)^2 = E(m_i - sp_i + s(p_i - p))^2 = E(m_i - sp_i)^2 + s^2(p_i - p)^2$$

since

$$E(m_i - sp_i) = 0.$$

On the other hand,

$$E(m_i - sp_i)^2 = \sum_{j=1}^{s} p_{ji} - \sum_{j=1}^{s} p_{ji}^2 = sp_i - \sum_{i=1}^{s} p_{ji}^2$$

and

$$\sum_{j=1}^{s} (p_i - p_{ji})^2 = -sp_i^2 + \sum_{j=1}^{s} p_{ji}^2,$$

whence

$$E(m_i - sp_i)^2 = sp_i - sp_i^2 - \sum_{j=1}^{s} (p_i - p_{ji})^2.$$

Now, letting i take values $1, 2, \ldots n$ and taking the sum of the results, we get

$$\sum_{i=1}^{n} E(m_i - sp_i)^2 = nsp - s\sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n}\sum_{j=1}^{s} (p_i - p_{ji})^2.$$

But

$$s\sum_{i=1}^{n} (p - p_i)^2 = -nsp^2 + s\sum_{i=1}^{n} p_i^2$$

whence finally

$$D = 1 + \frac{s-1}{npq}\sum_{i=1}^{n} (p - p_i)^2 - \frac{1}{Npq}\sum_{i=1}^{n}\sum_{j=1}^{s} (p_i - p_{ji})^2.$$

Two particular cases deserve special attention.

Lexis' Case. Probabilities remain the same within each series, but vary from series to series. In this case $p_{ji} = p_i$ and the expression of D becomes:

$$D = 1 + \frac{s-1}{npq}\sum_{i=1}^{n}(p - p_i)^2.$$

The theoretical divergence coefficient in this case is always greater than 1 and may be arbitrarily large.

Poisson's Case. The probabilities of the corresponding trials in all series are the same, so that

$$p_{ji} = \pi_j$$

and

$$p = p_i = \frac{\pi_1 + \pi_2 + \cdots + \pi_s}{s}.$$

In this case the divergence coefficient

$$D = 1 - \frac{\sum_{i=1}^{s}(p - \pi_i)^2}{spq}$$

is always less than 1.

Since the law of large numbers evidently is applicable to variables

$$x_i = \frac{(m_i - sp)^2}{spq},$$

we may expect that the empirical divergence coefficient D' will not differ much from D if the number of series is sufficiently large.

For numerical illustration let us consider 100 series each containing 100 trials, such that in 50 series the probability is $\frac{2}{5}$ and in the remaining 50 series it is $\frac{3}{5}$. Here we evidently have Lexis' case. The mean probability in all trials is

$$p = \tfrac{1}{2}$$

and

$$\sum_{i=1}^{100}(\tfrac{1}{2} - p_i)^2 = 50 \cdot \tfrac{1}{100} + 50 \cdot \tfrac{1}{100} = 1.$$

Finally,

$$D = 1 + \tfrac{99}{25} = 4.96.$$

Now, suppose that we combine in pairs series of 100 trials with probability $\frac{2}{5}$ and series of 100 trials with probability $\frac{3}{5}$, to form 50 series each of 200 trials. Evidently we have here Poisson's case. The mean probability in each series again is $p = \frac{1}{2}$ and

$$\sum_{i=1}^{200} (\tfrac{1}{2} - \pi_i)^2 = 100 \cdot \tfrac{1}{100} + 100 \cdot \tfrac{1}{100} = 2.$$

Finally,

$$D = 1 - \tfrac{2}{50} = 0.96.$$

The consideration of the divergence coefficient may be useful in testing the assumed independence of trials and values of probabilities attached to these trials. In the simplest case of Bernoullian trials with a constant and known probability, the theoretical divergence coefficient is 1. Now, if the number of series is sufficiently large and the empirical divergence coefficient turns out to be considerably different from 1, we must admit with great probability that the trials we deal with are not of the supposed type. If, however, the empirical divergence coefficient turns out to be near 1, that does not conclusively prove the hypothesis concerning the independence of trials and the assumed value of the probability. It only makes this hypothesis plausible.

There are cases of dependent trials (complex chains considered by Markoff) in which the theoretical divergence coefficient is exactly 1 and the probability of an event has the same constant value in each trial, insofar as the results of other trials remain unknown. Cases like that may easily be mistaken for Bernoullian trials without further detailed study of the entire course of trials.

4. When there is good reason to believe that the trials are independent with a constant but unknown probability, we cannot in all rigor find the value of the empirical divergence coefficient

$$D' = \frac{\sum_{i=1}^{n} (m_i - sp)^2}{Npq}$$

to compare it with the theoretical divergence coefficient $D = 1$, since p remains unknown.

But, relying on Bernoulli's theorem, we can take the quotient

$$\frac{M}{N}$$

where

$$M = m_1 + m_2 + \cdots + m_n$$

as an approximate value of p. By taking $p = M/N$ in the preceding expression for D' we get another number

$$D'' = \frac{N\sum_{i=1}^{n}\left(m_i - s\frac{M}{N}\right)^2}{M(N - M)}$$

which in general is close to D'. However, considering $m_1, m_2, \dots m_n$ not as observed but as eventual numbers of successes in n series, the mathematical expectation of D'' is different from 1. To avoid this difficulty, it is better to consider a slightly different quotient

$$Q = \frac{n(N - 1)\sum_{i=1}^{n}\left(m_i - s\frac{M}{N}\right)^2}{(n - 1)M(N - M)}.$$

For this quotient there exists a theorem discovered and proved for the first time by the eminent Russian statistician Tschuprow.

Theorem. *The mathematical expectation of Q is rigorously equal to* 1.[1]

Proof. Here we shall develop the proof given by Markoff. The above given expression of Q presents itself in the form $0/0$ and therefore has no meaning in two cases: $M = 0$ or $M = N$. For these exceptional cases we set $Q = 1$ by definition. If neither $M = 0$ nor $M = N$, we can present Q in the form

$$(4) \qquad Q = \frac{n(N - 1)}{n - 1}\,\frac{\sum m_i^2 - \frac{M^2}{n}}{M(N - M)}.$$

Considering $m_1, m_2, \dots m_n$ as stochastic variables assuming integral values from 0 to s, the probability of a definite system of values

$$m_1, m_2, \dots m_n$$

is

$$P = \frac{s!}{m_1!(s - m_1)!}\cdot\frac{s!}{m_2!(s - m_2)!}\cdots\frac{s!}{m_n!(s - m_n)!}p^M q^{N-M}.$$

To get the expectation of Q we must multiply it by P and take the sum

$$E(Q) = \Sigma PQ$$

extended over all non-negative integers $m_1, m_2, \dots m_n$, each of them not exceeding s. To perform this multiple summation we first collect all terms with a given sum

$$m_1 + m_2 + \cdots + m_n = M.$$

[1] The theorem itself and its proof given by Markoff can be extended to the case of series of unequal length.

Let the result of this summation be S_M. Then it remains to take the sum

$$\sum_{M=0}^{N} S_M$$

to have the desired expression $E(Q)$. To this end we first separate two terms corresponding to $M = 0$ and $M = N$. In the former case

$$m_1 = m_2 = \cdots = m_n = 0$$

and the probability of such an event is q^N while $Q = 1$. In the latter case

$$m_1 = m_2 = \cdots = m_n = s$$

the probability of which is p^N, while again $Q = 1$. Thus

$$E(Q) = p^N + q^N + \sum_{M=1}^{N-1} S_M.$$

To find S_M we observe that the denominator of Q has a constant value when summation is performed over variable integers $m_1, m_2, \ldots m_n$ connected by the relation

$$m_1 + m_2 + \cdots + m_n = M.$$

Hence, it suffices to find two sums

$$\Sigma P \qquad \text{and} \qquad \Sigma P m_i^2$$

extended over integers $m_1, m_2, \ldots m_n$ varying within limits 0 and s and having the sum M. To this end consider the function

$$V = (pte^{\xi_1} + q)^s(pte^{\xi_2} + q)^s \cdots (pte^{\xi_n} + q)^s$$

involving $n + 1$ arbitrary variables $t, \xi_1, \xi_2, \ldots \xi_n$. When developed, V consists of terms of the form

$$Pt^{m_1+m_2+\cdots+m_n}e^{m_1\xi_1+m_2\xi_2+\cdots+m_n\xi_n}.$$

Evidently we obtain the sum ΣP by setting $\xi_1 = \xi_2 = \cdots = \xi_n = 0$ and taking the coefficient of t^M in the expansion

$$(V)_{\xi_1=\xi_2=\cdots\xi_n=0} = (pt + q)^N.$$

Thus

$$\sum P = \frac{N!}{M!(N-M)!}p^M q^{N-M}. \tag{5}$$

To find $\Sigma P m_i^2$ take the second derivative

$$\frac{\partial^2 V}{\partial \xi_i^2}$$

and after setting $\xi_1 = \xi_2 = \cdots = \xi_n = 0$, expand

$$\left(\frac{\partial^2 V}{\partial \xi_i^2}\right)_{\xi_1=\xi_2=\cdots=\xi_n=0}$$

and take the coefficient of t^M. Thus we find

$$\sum P m_i^2 = \left[s\frac{(N-1)!}{(M-1)!(N-M)!} + s(s-1)\frac{(N-2)!}{(M-2)!(N-M)!}\right] p^M q^{N-M}. \tag{6}$$

Referring to (4), (5), and (6), we easily get

$$S_M = \frac{n(N-1)}{(n-1)M(N-M)} \cdot \frac{(N-2)!N}{n(M-1)!(N-M)!}[nN - n + (N-n)(M-1) - M(N-1)]p^M q^{N-M};$$

or, after obvious simplifications,

$$S_M = \frac{N!}{M!(N-M)!} p^M q^{N-M}.$$

Hence

$$\sum_{M=1}^{N-1} S_M = (p+q)^N - p^N - q^N = 1 - p^N - q^N,$$

and finally

$$E(Q) = 1.$$

Markoff, using the same method, succeeded in finding the explicit expression of the expectation

$$E(Q-1)^2.$$

Since there is no difficulty in finding this expression except for somewhat tedious calculations, we give it here without entering into details of the proof:

$$E(Q-1)^2 = \frac{2N(N-n)}{(n-1)(N-2)(N-3)} \sum_{M=1}^{N-1} \frac{M-1}{M} \cdot \frac{N-M-1}{N-M} C_N^M p^M q^{N-M},$$

whence the following inequality immediately follows:

$$E(Q-1)^2 < \frac{2N(N-n)}{(n-1)(N-2)(N-3)}.$$

In case $n \geqq 5$ a still simpler inequality holds:

$$E(Q-1)^2 < \frac{2}{n-1}. \tag{7}$$

Let R be the probability of the inequality

$$Q \geqq 1 + \epsilon,$$

where ϵ is a positive number. Applying the same reasoning to inequality (7) as was used in establishing Tshebysheff's lemma, we find that

$$R < \frac{2}{(n-1)\epsilon^2}.$$

Likewise, denoting by R' the probability of the inequality

$$Q \leqq 1 - \epsilon,$$

we have

$$R' < \frac{2}{(n-1)\epsilon^2}.$$

Thus, in a large number of series it becomes very unlikely that the value of Q found in actual experiment would lie outside of the interval $1 - \epsilon$, $1 + \epsilon$. For instance, the probability for $Q \geqq 2$ in 100 series is surely less than

$$\frac{2}{99}$$

or nearly 0.02. However, this limit is much too high. It would be greatly desirable to have a good approximate expression for the probability of either one of the inequalities

$$Q \geqq 1 + \epsilon \qquad \text{or} \qquad Q \leqq 1 - \epsilon.$$

But this important and difficult problem has not yet been solved.

5. In order to illustrate the foregoing theoretical considerations we turn to experiments reported by Charlier in his book "Vorlesungen über die Grundzüge der mathematischen Statistik" (Lund, 1920). He made 10,000 drawings of single cards from a complete deck of 52 cards (each card taken being returned before the next drawing), and noted the frequency of black cards. The drawings were divided into 1,000 series of 10 cards, or into 200 series of 50 cards. The results are given in the tables on page 220.

Assuming the independence of trials and the constant probability $p = \frac{1}{2}$, the theoretical divergence coefficient must be 1. Let us compare it with the empirical divergence coefficient derived from Tables I and II. To this end we multiply the squares of numbers in the second column by the numbers given in the third column. The results are:

For 200 series of 50 cards	For 1,000 series of 10 cards
$\Sigma(m_i - ps)^2 = 2{,}487$	$\Sigma(m_i - ps)^2 = 2{,}419$

TABLE I.—NUMBER OF BLACK CARDS IN 200 GROUPS OF 50 CARDS EACH

Frequency	Difference $m - 25$	Number of groups with these frequencies
14	−11	1
15	−10	0
16	− 9	2
17	− 8	2
18	− 7	4
19	− 6	8
20	− 5	6
21	− 4	15
22	− 3	13
23	− 2	15
24	− 1	34
25	0	14
26	1	21
27	2	26
28	3	14
29	4	10
30	5	5
31	6	5
32	7	3
33	8	2

TABLE II.—NUMBER OF BLACK CARDS IN 1,000 GROUPS OF 10 CARDS EACH

Frequency	Difference $m - 5$	Number of groups with these frequencies
0	−5	3
1	−4	10
2	−3	43
3	−2	116
4	−1	221
5	0	247
6	1	202
7	2	115
8	3	34
9	4	9
10	5	0

Dividing these numbers by $10{,}000 \cdot \frac{1}{4} = 2{,}500$, we get the following empirical divergence coefficients:

$$D' = 0.9948; \qquad D'' = 0.9676.$$

Both are close to 1, so that the hypotheses of independence of trials and constant probability $\frac{1}{2}$ for each of them, are in good agreement with empirical results. The second divergence coefficient, corresponding to more numerous groups, differs from 1 more than the first, corresponding to only 200 groups. But such a difference can be accounted for by fluctuations due to chance.

Series of 50 trials are long enough to test the theorem established in Sec. 2 of this chapter. The quantities denoted there by A and B are here correspondingly:

$$A^2 = \frac{2487}{200}; \qquad A = 3.5263$$
$$B = \frac{561}{200}; \qquad B = 2.805$$

whence

$$\frac{A}{B} = 1.2571$$

while

$$\sqrt{\frac{\pi}{2}} = 1.2533.$$

Again the difference, only about 4.10^{-3}, is rather small.

In this example, the probability of drawing a black card was assumed to be $\frac{1}{2}$. In case we do not know the probability, but suppose it to be constant throughout 10,000 independent trials, we must consider the coefficient

$$Q = \frac{n(N-1)}{(n-1)M(N-M)}\sum_{s=1}^{n}\left(m_i - s\frac{M}{N}\right)^2.$$

In our example

$$n = 1{,}000; \qquad N = 10{,}000; \qquad M = 4{,}933$$
$$s = 10; \qquad s\frac{M}{N} = 4.933.$$

To evaluate the sum

$$S = \sum_{i=1}^{1,000}(m_i - 4.933)^2$$

we write it in the form

$$S = \sum_{i=1}^{1,000}(m_i - 5)^2 + 0.134\sum_{i=1}^{1,000}(m_i - 5) + 1{,}000\cdot(0.067)^2.$$

Now

$$\begin{aligned}
\sum_{1}^{1,000}(m_i - 5)^2 &= 2{,}419 \\
1{,}000\cdot(0.067)^2 &= \quad 4.489 \\
0.134\sum_{1}^{1,000}(m_i - 5) &= \quad -8.978 \\
\hline
S &= 2{,}414.51
\end{aligned}$$

This is to be multiplied by the number

$$\frac{n(N-1)}{(n-1)M(N-M)} = \frac{1}{2497.3}.$$

The result is

$$0.9668,$$

near enough to 1 for us to consider the hypothesis of independence of trials and the constant value of probability as in agreement with experimental data.

EXAMPLES OF DEPENDENT TRIALS

6. So far we have dealt only with independent variables. But the law of large numbers holds, under certain conditions, even in the case of dependent variables. Leaving aside generalities, we shall show the application of the law of large numbers to a few interesting problems involving dependent variables.

Let us consider first a Bernoullian series consisting of $n + 1$ independent trials with the same probability p for an event E, the opposite event being denoted by F. We associate with trials $1, 2, \dots n$ variables $x_1, x_2, \dots x_n$ defined as follows:

$$x_i = 1 \text{ if } E \text{ occurs in trials } i \text{ and } i + 1,$$
$$x_i = 0 \text{ in all other cases.}$$

The probability of $x_i = 1$ evidently is p^2 when nothing is known about the values of other variables. But if we know that $x_{i-1} = 1$, which implies the occurrence of E in the ith trial, then the probability of $x_i = 1$ is p. Thus, consecutive variables are dependent. However, x_i and x_k are independent if $|k - i| > 1$, as we can easily see. Since

$$E(x_i) = E(x_i^2) = p^2 \cdot 1 + (1 - p^2) \cdot 0 = p^2$$

the expectation of the sum $x_1 + x_2 + \cdots + x_n$ will be

$$E(x_1 + x_2 + \cdots + x_n) = np^2.$$

As to the dispersion of this sum, it can be expressed as follows:

$$B_n = \sum_{i=1}^{n} E(x_i - p^2)^2 + 2\sum_{j>i} E(x_i - p^2)(x_j - p^2).$$

Now

$$E(x_i - p^2)^2 = E(x_i^2) - 2p^2E(x_i) + p^4 = p^2(1 - p^2) \tag{8}$$

and

$$E(x_i - p^2)(x_j - p^2) = E(x_i - p^2) \cdot E(x_j - p^2) = 0 \tag{9}$$

for $j > i + 1$ because then x_i and x_j are independent. But

$$E(x_i - p^2)(x_{i+1} - p^2) = E(x_i x_{i+1}) - p^4 = p^3 - p^4 \tag{10}$$

since the probability of simultaneous events

$$x_i = 1, \qquad x_{i+1} = 1$$

is p^3. Taking into account (8), (9), and (10), we find

$$B_n = np^2q(3p + 1) - 2p^3q$$

and the condition

$$\frac{B_n}{n^2} \to 0 \qquad \text{as} \qquad n \to \infty$$

is satisfied. Hence, the law of large numbers holds for variables $x_1, x_2, \ldots x_n$. To express it in the simplest form, it suffices to notice that the sum

$$x_1 + x_2 + \cdots + x_n$$

represents the number of pairs EE occurring in consecutive trials of the Bernoullian series of $n + 1$ trials. Let us denote the frequency of such pairs by m. Then, referring to the law of large numbers, we get the following proposition:

If in n consecutive pairs of Bernoullian trials the frequency of double successes EE is m, then the probability of the inequality

$$\left|\frac{m}{n} - p^2\right| < \epsilon$$

will approach 1 *as near as we please, when n becomes sufficiently large.*

7. Simple chains of trials, described in Chap. V, Sec. 1, offer a good example of dependent trials to which the law of large numbers can be applied. Let p_1 be the given probability of an event E in the first trial. According to the definition of a simple chain, the probability of E in any subsequent trial is α or β according as E occurred or failed to occur in the preceding trial. By p_n we denote the probability for E to occur in the nth trial when the results of other trials are unknown. Let

$$\delta = \alpha - \beta, \qquad p = \frac{\beta}{1 - \delta}.$$

Then, according to the developments in Chap. V, Sec. 2,

$$p_n = p + (p_1 - p)\delta^{n-1},$$

whence

$$\frac{p_1 + p_2 + \cdots + p_n}{n} = p + \frac{p_1 - p}{n}\frac{1 - \delta^n}{1 - \delta},$$

barring the trivial cases $\delta = 1$ or $\delta = -1$. It follows that p represents the limit of the mean probability in n trials when n increases indefinitely, and for that reason p may be called the mean probability in an infinite chain of trials. When it is known that E has occurred in the ith trial, its probability of occurring in some subsequent jth trial is given by

$$p_j^{(i)} = p + q\delta^{j-i}, \qquad q = 1 - p.$$

In the usual way we associate with trials 1, 2, 3, . . . variables $x_1, x_2, x_3, \ldots$ so that in general

$$x_i = 1 \text{ when } E \text{ occurs in the } i\text{th trial}$$
$$x_i = 0 \text{ when } E \text{ fails to occur in the } i\text{th trial.}$$

Evidently

$$E(x_i) = E(x_i^2) = p_i.$$

In order to prove that the law of large numbers can be applied to variables $x_1, x_2, x_3, \ldots$, we must have an idea of the behavior of B_n for large n. By definition

$$B_n = E(x_1 - p_1 + x_2 - p_2 + \cdots + x_n - p_n)^2 = \sum_{i=1}^{n} E(x_i - p_i)^2 + $$
$$+ 2\sum_{j>i} E[(x_i - p_i)(x_j - p_j)].$$

The first sum can easily be found. We have

$$E(x_i - p_i)^2 = p_i - p_i^2 = pq + (q - p)(p_1 - p)\delta^{i-1} - (p_1 - p)^2\delta^{2i-2}$$

whence

$$A = \sum_{i=1}^{n} E(x_i - p_i)^2 \sim npq$$

neglecting terms which remain bounded. As to the second sum, we observe first that

$$E(x_i - p_i)(x_j - p_j) = E(x_i x_j) - p_i p_j.$$

Again, since the probability of

$$x_i x_j = 1$$

is evidently $p_i p_j^{(i)}$ we have

$$E(x_i x_j) = p_i p_j^{(i)},$$

and

$$E(x_i - p_i)(x_j - p_j) = p_i(p_j^{(i)} - p_j) = pq\delta^{j-i} +$$
$$+ (p_1 - p)(q - p)\delta^{j-1} - (p_1 - p)^2\delta^{i+j-2}.$$

Now, for a fixed $i = 1, 2, \ldots n - 1$, we must take the sum of these expressions letting j run over $i + 1, i + 2, \ldots n$. The result of this summation is

$$pq\frac{\delta - \delta^{n-i+1}}{1 - \delta} + (p_1 - p)(q - p)\frac{\delta^i - \delta^n}{1 - \delta} - (p_1 - p)^2\delta^i\frac{\delta^{i-1} - \delta^{n-1}}{1 - \delta}.$$

Taking $i = 1, 2, 3, \ldots n - 1$ and neglecting in the sum the terms which remain bounded, we get

$$B = \sum_{j>i} E(x_i - p_i)(x_j - p_j) \sim npq\frac{\delta}{1 - \delta}$$

whence

$$B_n = A + 2B \sim npq\frac{1 + \delta}{1 - \delta}.$$

This asymptotic equality suffices to show that

$$\frac{B_n}{n^2} \to 0 \qquad \text{as} \qquad n \to \infty.$$

Therefore the law of large numbers can be applied to variables x_1, x_2, x_3, ... Since the sum

$$x_1 + x_2 + \cdots + x_n = m$$

represents the frequency of E in n trials, the law of large numbers in this particular case can be stated as follows: For a fixed $\epsilon > 0$, no matter how small, the probability of the inequality

$$\left|\frac{m}{n} - \frac{p_1 + p_2 + \cdots + p_n}{n}\right| < \epsilon$$

tends to 1 as $n \to \infty$.

The arithmetic mean

$$\frac{p_1 + p_2 + \cdots + p_n}{n}$$

itself approaches the limit p. It is easy then to express the preceding theorem thus: *The probability of the inequality*

$$\left|\frac{m}{n} - p\right| < \epsilon$$

tends to 1 *as* $n \to \infty$.

This proposition is of exactly the same type as Bernoulli's theorem, but applies to series of dependent trials.

8. Let a simple chain of $N = ns$ trials be divided into n consecutive series each consisting of s trials; also, let $m_1, m_2, \ldots m_n$ be the frequencies of E in each of these series. When N is a large number, the mean probability in N trials differs little from the quantity denoted by p. It is natural to modify the definition of the divergence coefficient given in Sec. 3 by taking p instead of the variable mean probability in N trials. Thus we define

$$D = E\frac{\sum_{i=1}^{n}(m_i - sp)^2}{Npq}.$$

In our case, the variables

$$X_1 = (m_1 - sp)^2, \qquad X_2 = (m_2 - sp)^2, \quad \cdots \quad X_n = (m_n - sp)^2$$

are neither identical nor independent, although the degree of dependence is evidently very slight. These variables can also be presented in the form

$$(11) \qquad (x_a - p + x_{a+1} - p + \cdots + x_{a+s-1} - p)^2$$

taking successively $a = 1, s + 1, 2s + 1, \ldots (n - 1)s + 1$.

To find the mathematical expectation of (11) it suffices to notice that

$$E(x_i - p)^2 = E(x_i - p_i)^2 + (p_i - p)^2 = pq + (q - p)(p_1 - p)\delta^{i-1}$$
$$E(x_i - p)(x_j - p) = E(x_i - p_i)(x_j - p_j) + (p_i - p)(p_j - p)$$
$$= pq\delta^{j-i} + (p_1 - p)(q - p)\delta^{j-1}$$

and then proceed exactly as in the approximate evaluation of B_n in Sec. 7. The final result is

$$E(x_a - p + x_{a+1} - p + \cdots + x_{a+s-1} - p)^2 =$$
$$= spq\frac{1 + \delta}{1 - \delta} - \frac{2pq\delta}{(1 - \delta)^2} + \frac{(q - p)(p_1 - p)(1 + \delta)}{(1 - \delta)^2}\delta^{a-1} +$$
$$+ \frac{2pq}{(1 - \delta)^2}\delta^{s+1} - \frac{(q - p)(p_1 - p)}{(1 - \delta)^2}[2s(1 - \delta) + 1 + \delta]\delta^{a+s-1}.$$

For somewhat large s the two last terms in the right member are completely negligible; so is the third term if $a \geqq s + 1$. Hence, with a good approximation,

$$E(X_1) = spq\frac{1 + \delta}{1 - \delta} - \frac{2pq\delta}{(1 - \delta)^2} + \frac{(q - p)(p_1 - p)(1 + \delta)}{(1 - \delta)^2}$$
$$E(X_i) = spq\frac{1 + \delta}{1 - \delta} - \frac{2pq\delta}{(1 - \delta)^2} \qquad \text{if} \qquad i > 1$$

and

$$D = \frac{1 + \delta}{1 - \delta} - \frac{2\delta}{s(1 - \delta)^2} + \frac{(q - p)(p_1 - p)(1 + \delta)}{Npq(1 - \delta)^2}.$$

Again, when N is large, the last term can be dropped and as a good approximation to D we can take

$$(12) \qquad D = \frac{1 + \delta}{1 - \delta} - \frac{2\delta}{s(1 - \delta)^2}.$$

It can be shown that the law of large numbers holds for variables X_1, $X_2, \ldots X_n$ and therefore when n (or the number of series) is large, the empirical divergence coefficient is not likely to differ considerably from D as given by the above approximate formula.

9. In order to see how far the theory of simple chains agrees with actual experiments, the author of this book himself has done extensive experimental work. To form a chain of trials, one can take two sets of cards containing red and black cards in different proportions, and proceed to draw one card at a time (returning it to the pack in which it belongs after each drawing) according to the following rules: At the outset one card is taken from a pack which we shall call the *first* set; then, whenever a red card is drawn, the next card is taken from the first set; but after a black card, the next one is taken from the *second* set. Evidently, these rules completely determine a series of trials possessing properties of a simple chain. In the first experiment the first pack contained 10 red and 10 black cards, while the second pack contained 5 red and 15 black cards. Altogether, 10,000 drawings were made, and following their natural order, they were divided into 400 series of 25 drawings each. The results are given in Table III.

TABLE III.—DISTRIBUTION OF RED CARDS IN 400 SERIES OF 25 CARDS

Frequency of red cards, m	Difference, $m - 8$	Number of series with these frequencies
1	−7	2
2	−6	4
3	−5	8
4	−4	27
5	−3	29
6	−2	54
7	−1	37
8	0	52
9	1	47
10	2	44
11	3	41
12	4	20
13	5	20
14	6	7
15	7	4
16	8	3
17	9	1

The sum of the numbers in column 3 is 400, as it should be. Taking the sum of the products of numbers in columns 1 and 3, we get 3,323, which is the total number of red cards. The relative frequency of red cards in 10,000 trials is, therefore,

$$0.3323.$$

In our case

$$\alpha = \tfrac{1}{2}, \qquad \beta = \tfrac{1}{4}, \qquad \delta = \tfrac{1}{4}$$

and the mean probability p in an infinite series of trials

$$p = \frac{\beta}{1 - \delta} = \frac{1}{3} = 0.3333.$$

Thus, the relative frequency observed differs from p only by 10^{-3} and in this respect the agreement between theory and experiment is very satisfactory. Now let us consider the theoretical divergence coefficient for which we have the approximate expression

$$D = \frac{1 + \delta}{1 - \delta} - \frac{2\delta}{s(1 - \delta)^2}.$$

Here we must substitute $\delta = 1/4$ and $s = 25$. The result is

$$D = 1.631, \text{ approximately.}$$

To find the empirical divergence coefficient we must first evaluate the sum

$$S = \Sigma(m - \tfrac{25}{3})^2$$

extended over all 400 series. For the sake of easier calculation, we present S thus:

$$S = \Sigma(m - 8)^2 - \tfrac{2}{3}\Sigma(m - 8) + \tfrac{400}{9}.$$

Now from Table III we get

$$\Sigma(m - 8)^2 = 3{,}521; \qquad \Sigma(m - 8) = 123$$

whence

$$S = 3{,}483.4.$$

Dividing this number by $20000/9 = 2{,}222.2$, we find the empirical divergence coefficient

$$D' = 1.568$$

which differs from $D = 1.631$ by only about 0.06, well within reasonable limits.

10. In two other experiments two packs were used: one containing 13 red and 7 black cards, and another 7 red and 13 black cards. In one experiment the pack with 13 red cards was considered as the *first* deck, and in the other experiment it became the *second* deck. The new experiments were conducted in the same way as that described in Sec. 9, but they were both carried to 20,000 trials divided into 1,000 series of 20 trials each. In the first experiment, we have

$$\alpha = \tfrac{13}{20}, \qquad \beta = \tfrac{7}{20}, \qquad \delta = \tfrac{8}{20}, \qquad p = \tfrac{1}{2}$$

and

$$D = 1.796, \text{ approximately,}$$

while the same quantities for the second experiment are

$$\alpha = \frac{7}{20}, \qquad \beta = \frac{13}{20}, \qquad \delta = -\frac{8}{10}, \qquad p = \frac{1}{2}$$

and

$$D = 0.556, \text{ approximately.}$$

The results of these experiments are recorded in the following two tables:

TABLE IV.—CONCERNING THE FIRST EXPERIMENT

Frequency of red cards, m	Difference, $m - 10$	Number of series with these frequencies
2	−8	3
3	−7	5
4	−6	18
5	−5	36
6	−4	59
7	−3	93
8	−2	103
9	−1	117
10	0	128
11	1	121
12	2	101
13	3	93
14	4	48
15	5	39
16	6	26
17	7	7
18	8	1
19	9	1
20	10	1

TABLE V.—CONCERNING THE SECOND EXPERIMENT

Frequency of red cards, m	Difference, $m - 10$	Number of series with these frequencies
5	−5	2
6	−4	10
7	−3	48
8	−2	112
9	−1	193
10	0	251
11	1	201
12	2	113
13	3	56
14	4	9
15	5	5

Taking the sum of the products of numbers in columns 1 and 3, we find

$$10{,}036 \qquad \text{and} \qquad 10{,}045$$

as the total number of red cards in the first and second experiments. Dividing these numbers by 20,000, we have the following relative frequencies of red cards:

$$0.50018 \qquad \text{and} \qquad 0.500225$$

extremely near to $p = 0.5$. From the first table we find that

$$\Sigma(m - 10)^2 = 8{,}924$$

summation being extended over all 1,000 series. Dividing this number by $20{,}000 \cdot \frac{1}{4} = 5{,}000$, we find the empirical divergence coefficient in the first experiment

$$D' = 1.785$$

which comes close to

$$D = 1.796.$$

Likewise, from the second table we find

$$\Sigma(m - 10)^2 = 2{,}709,$$

whence, dividing by 5,000,

$$D'' = 0.5418$$

again close to

$$D = 0.5562.$$

Thus, all the essential circumstances foreseen theoretically, for simple chains of trials, are in excellent agreement with our experiments.

Problems for Solution

1. From an urn originally containing a white and b black balls, n balls are drawn in succession, each ball drawn being replaced by $1 + c(c > 0)$ balls of the same color before the next drawing. If m is the frequency of white balls, show that the probability of the inequality

$$\left|\frac{m}{n} - \frac{a}{a+b}\right| < \epsilon$$

does not tend to 1 as n increases indefinitely (Markoff, G. Pólya).

Indication of the Proof. If $x_i = 1$ or $x_i = 0$, according as a white or a black ball appears in the ith drawing, we have

$$E(x_i) = E(x_i^2) = \frac{a}{a+b}, \qquad E(x_ix_j) = \frac{a}{a+b}\,\frac{a+c}{a+b+c}$$

Hence

$$B_n = E\left(x_1 + x_2 + \cdots + x_n - \frac{na}{a+b}\right)^2 = \frac{n^2abc}{(a+b)^2(a+b+c)} + \\ + \frac{nab}{(a+b)(a+b+c)}.$$

2. *Marbe's Problem.* A group of exactly m uninterrupted successes E or failures F in a Bernoullian series of trials with the probability p for a success is called an "m sequence." If N is the frequency of m sequences in n trials. show that the probability of the inequality

$$\left|\frac{N}{n} - (p^mq^2 + p^2q^m)\right| < \epsilon$$

for a fixed ϵ converges to 1 as n becomes infinite.

Indication of the Proof. Associate with each of the $\mu = n - m + 1$ first trials variables $x_1, x_2, \ldots x_\mu$ assuming only two values, 0 and 1. For $1 < i < \mu$ we set $x_i = 1$ if, beginning with the ith trial, a succession of m letters E or F is preceded and followed by F or E. In all other cases $x_i = 0$. We set $x_1 = 1$ if, beginning with the first trial, there is a succession of m letters E or F ended by F or E; otherwise $x_1 = 0$. Finally, $x_\mu = 1$ if, beginning with the μth trial there is a succession of m letters E or F preceded by F or E, otherwise $x_\mu = 0$. Show that

$$E(x_1 + x_2 + \cdots + x_\mu) = (n - m - 1)(p^mq^2 + p^2q^m) + 2(p^mq + pq^m)$$
$$E(x_1 + x_2 + \cdots + x_\mu)^2 = n^2(p^mq^2 + p^2q^m)^2 + nP$$

where P remains bounded.

3. The following interesting series of dependent trials has been suggested by S. Bernstein: Two urns contain white and black balls. The probabilities of drawing white balls from the first and second urns are, respectively, p and p'. The probabilities of drawing black balls from the same urns are $q = 1 - p$ and $q' = 1 - p'$. Finally, the probability of taking a ball from the first urn at the outset of the trials is α. A series of trials is uniquely defined by the following rule: Whenever a white ball is drawn (and returned), the next ball is drawn from the same urn; but when a black ball is drawn, the next ball is taken from the other urn. Let α_n be the probability that the nth ball will be drawn from the first urn when the results of other drawings remain unknown. Under the same assumption, let p_n be the probability of the nth ball being white. Find general expressions of α_n and p_n.

HINT:

$$\alpha_{n+1} = \alpha_n(p + p' - 1) + 1 - p'$$

whence

$$\alpha_n = \frac{1 - p'}{2 - p - p'} + \left(\alpha - \frac{1 - p'}{2 - p - p'}\right)(p + p' - 1)^{n-1}.$$

Also

$$p_n = \alpha_n p + (1 - \alpha_n)p'$$

whence

$$p_n = \frac{p + p' - 2pp'}{2 - p - p'} + \left(\alpha - \frac{1 - p'}{2 - p - p'}\right)(p - p')(p' + p - 1)^{n-1}.$$

4. When it becomes known that in the ith trial a white ball was drawn, what are the probabilities $\alpha_j^{(i)}$ and $p_j^{(i)}$ of taking a ball from the first urn in the jth ($j > i$) trial and of drawing a white ball in the same trial?

HINT: The probability $\alpha_i^{(i)}$ that it was the first urn from which a white ball was drawn in the ith trial is determined by Bayes' formula:

$$\alpha_{i+1}^{(i)} = \alpha_i^{(i)} = \frac{\alpha_i p}{p_i}.$$

For $n \geqq i + 1$

$$\alpha_{n+1}^{(i)} = \alpha_n^{(i)}(p + p' - 1) + 1 - p'$$

whence

$$\alpha_j^{(i)} = \frac{1 - p'}{2 - p - p'} + \left(\frac{\alpha_i p}{p_i} - \frac{1 - p'}{2 - p - p'}\right)(p + p' - 1)^{j-i-1}$$

for $j > i + 1$. Furthermore

$$p_j^{(i)} = \alpha_j^{(i)}p + (1 - \alpha_j^{(i)})p'$$

for $j \geqq i + 1$.

5. From now on we shall assume $p + p' = 1$ or $p' = q$, $q' = p$. Show that the law of large numbers can be applied to variables $x_1, x_2, x_3, \ldots$ which are defined in the usual way:

$$x_i = 1 \text{ if a white ball is drawn in the } i\text{th trial},$$
$$x_i = 0 \text{ if a black ball is drawn in the } i\text{th trial}.$$

Indication of the Proof. Evidently $E(x_i) = E(x_i^2) = p_i$. Furthermore

$$B_n = \sum_{i=1}^{n} E(x_i - p_i)^2 + 2\sum_{j>i} E(x_i - p_i)(x_j - p_j).$$

Now

$$E(x_i - p_i)^2 = 2pq(1 - 2pq); \qquad i > 1$$
$$E(x_1 - p_1)^2 = pq + \alpha(1 - \alpha)(p - q)^2.$$

For $j > i > 1$

$$E(x_i - p_i)(x_j - p_j) = 0 \qquad \text{if} \qquad j > i + 1$$
$$E(x_i - p_i)(x_{i+1} - p_{i+1}) = pq(1 - 4pq).$$

For $i = 1$ and $j > 1$

$$E(x_1 - p_1)(x_j - p_j) = 0 \qquad \text{if} \qquad j > 2$$
$$E(x_1 - p_1)(x_2 - p_2) = \alpha p^2 + (1 - \alpha)q^2 - (1 - 2pq)(q + (p - q)\alpha).$$

Hence

$$B_n \sim 4pq(1 - 3pq)n$$

and the law of large numbers holds. It can be stated as follows: If in n trials the frequency of white balls is m, then the probability of the inequality

$$\left|\frac{m}{n} - (p^2 + q^2)\right| < \epsilon$$

tends to 1 as n tends to infinity for any given positive number ϵ.

6. Let $r = p^2 + q^2$ be the mean probability in infinitely many trials. Find the divergence coefficient

$$D = E\frac{\sum_{i=1}^{n}(m_i' - sr)^2}{Nr(1 - r)}$$

when $N = ns$ trials are divided in n consecutive groups containing s trials each.

Indication of Solution. From the foregoing formulas it follows that

$$E(x_a - r + x_{a+1} - r + \cdots + x_{a+s-1} - r)^2 = 4spq(1 - 3pq) - 2pq(1 - 4pq)$$

if $a > 1$. Hence

$$E\sum_{i=2}^{n}(m_i - sr)^2 = 4Npq(1 - 3pq) - 4spq(1 - 3pq) - 2(n - 1)pq(1 - 4pq).$$

Again

$$E(m_1 - sr)^2 = 4spq(1 - 3pq) - 2pq(3 - 10pq) + p(1 - 6q + 12q^2 - 4q^3) - \\ - \alpha(p - q)(1 - 8pq)$$

so that finally

$$D = \frac{2 - 6pq}{1 - 2pq} - \frac{1 - 4pq}{s(1 - 2pq)} + \frac{(p - q)(p - \alpha)(1 - 8pq)}{2Npq(1 - 2pq)}.$$

For large N with a good approximation

$$D = \frac{2 - 6pq}{1 - 2pq} - \frac{1 - 4pq}{s(1 - 2pq)}.$$

7. Two sets of cards containing respectively 12 red and 4 black cards (the *first* deck) and 4 red and 12 black cards (the *second* deck) were used in the following experiment: The first card was taken from the first deck, and in the following trials, after a red card the next one was taken from the *same* deck, but after a black one the next card was taken from the *other* deck. Altogether 25,000 cards were drawn, and in their natural order were divided in 1,000 series of 25 cards each. The results are recorded in Table VI. How close is the agreement between this experiment and the theory?

TABLE VI.—DISTRIBUTION OF RED CARDS IN 1,000 SERIES OF 25 CARDS

Frequency of red cards, m	Difference, $m - 16$	Number of series with these frequencies
6	−10	1
7	− 9	1
8	− 8	1
9	− 7	12
10	− 6	13
11	− 5	43
12	− 4	65
13	− 3	92
14	− 2	101
15	− 1	162
16	0	94
17	1	164
18	2	68
19	3	110
20	4	26
21	5	28
22	6	10
23	7	7
24	8	1
25	9	1

Ans. In the present case $p = q' = 3/4$, $p' = q = 1/4$. Mean probability in infinitely many trials:

$$p^2 + q^2 = \tfrac{5}{8} = 0.625.$$

Theoretical divergence coefficient: $D = 1.384$. Frequency of red cards: 15,696. Relative frequency:

$$\tfrac{15696}{25000} = 0.62784,$$

close to 0.625.

Empirical divergence coefficient: $D' = 1.3845$, very close to 1.384.

The probability of taking a card from the second deck is 0.25. Now, by actual counting, it was found that in 7,500 trials a card was taken from the second deck 1,856 times. Hence, the relative frequency of this event in 7,500 trials is

$$\tfrac{1856}{7500} = 0.2475,$$

again very close to 0.25.

References

Poisson: "Recherches sur la probabilité des jugements en matière criminelle et en matière civile," Paris, 1837.

Tshebysheff: Sur les valeurs moyennes, Oeuvres I, pp. 687–694.

W. Lexis: "Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft," 1877.

A. Tschuprow: "Grundbegriffe und Grundprobleme der Korrelationstheorie," Leipzig, 1925.

A. Markoff: "Calculus of Probability," 4th Russian edition, Leningrad, 1924.

S. Bernstein: "Theory of Probability" (Russian), Moscow, 1927.

CHAPTER XII

PROBABILITIES IN CONTINUUM

1. In the preceding parts of this book, whenever we dealt with stochastic variables, it was understood that their range of variation was represented by a finite set of numbers. Although, for the sake of better understanding of the subject, it was natural to begin with this simplest case, there are many reasons why it is necessary to introduce into the calculus of probability stochastic variables with infinitely many values. Such variables present themselves naturally in many cases of the type of Buffon's needle problem which we had occasion to mention in Chap. VI.

On the other hand, even in dealing with stochastic variables with a finite, but very large number of values, it is often profitable for the sake of approximate evaluations, to substitute for them fictitious variables with infinitely many values. Among these the most important ones by far are *continuous variables.*

Case of One Variable

2. Beginning with the case of a single continuous variable x, we must assume that its range of variation is known and represented by a given interval (a, b), finite or infinite. The knowledge only of the range of variation of x would not enable us to consider x as a stochastic variable; to be able to do so, we must introduce in some form or other the considerations of probability. For a continuous variable it is as unnatural to speak of the probability of any selected single value, as it is to speak of the dimension of a single selected point on a line. But just as we speak of the length of a segment of a line, we may introduce the notion of the probability that x will be confined to a given interval (c, d), part of (a, b).

In introducing this new notion of probability in any manner whatsoever, we must be careful not to fall into contradiction with the laws of probability which are assumed as fundamental. To this end, if $P\ (c, d)$ is the probability for x to lie in the interval (c, d), we are led to assume

$$1^\circ\ P(c, d) \geqq 0$$
$$2^\circ\ P(a, b) = 1.$$

The first assumption is an expression of the fact that probability can never be negative. The second assumption corresponds to the fact that x certainly assumes one out of the totality of its possible values.

Next, if the interval (c, d) is divided into two adjoining intervals (c, e) and (e, d), we assume

$$3° \; P(c, d) = P(c, e) + P(e, d)$$

in conformity with the theorem of total probability.

For continuous variables it is furthermore assumed: 4° for an infinitesimal interval (c, d), $P(c, d)$ is also infinitesimal.

Properties 3° and 4° show that $P(c, d)$ is a continuous function of c and d and that

$$P(c, c) = 0.$$

In other words, the probability that x will assume any given value is 0. At the same time $P(c, d)$ represents the probability of any one of the four inequalities

$$c < x < d; \qquad c \leqq x < d; \qquad c < x \leqq d; \qquad c \leqq x \leqq d.$$

3. A simple example will serve to clarify these general considerations. A small ball of negligible dimensions is made to move on the rim of a circular disk. It is set in motion by a vehement impulse and after many complete revolutions, retarded by friction and the resistance of the air, comes to rest. The variety and complexity of causes influencing the motion of the ball make it impossible to foresee the final position of the ball when it comes to rest and the whole phenomenon bears characteristic features of a play of chance. The stochastic variable associated with this chance phenomenon is the distance from a certain definite point on the rim (origin) to the final position of the ball, counted in a definite direction, for example, clockwise. This variable, when we consider the ball as a mere point, may have any value between 0 and the length of the rim. The question now arises, how to define the probability that the ball will stop in a specified portion of the rim, or else that the variable we consider will have a value belonging to a definite interval, part of its total range of variation. In trying to define this probability, we must observe the fundamental requirements set forth in Sec. 2. Besides that, we must of necessity resort to considerations which are not mathematical in their nature but are based partly on aprioristic and partly on experimental grounds. Suppose we take two equal arcs on the rim. There is nothing perceptible a priori that would make the ball stop in one arc rather than in another. Besides, actual experiments show that the ball stops in one arc approximately the same number of times as in another, and this experimental knowledge together with aprioristic considerations suggests the assumption that we must attribute equal probabilities to equal arcs, irrespective of the position of the arcs on the rim. As soon as we agree on this assumption or hypothesis, the problem becomes mathematical and can easily be solved.

Before proceeding to the solution, a remark on the meaning of zero probability in connection with continuous variables is not out of place. Zero probability in this case does not mean logical impossibility. We attribute zero probability to the event that the ball will stop precisely at the origin. However, that possibility is not altogether excluded so far as we consider the origin and the ball as mere points. The question lacks sense if we deal with a material ball and a material rim, no matter how small the former and how fine the latter.

4. A stochastic variable is said to have uniform distribution of probability if probabilities attached to two equal intervals are equal. This means that $P(c, d)$ depends only upon the length $d - c = s$ of the interval (c, d) and accordingly can be denoted simply by $P(s)$. Combining two adjoining intervals of the respective lengths s and s' into a single interval of length $s + s'$, according to requirement 3°, we must have

$$P(s + s') = P(s) + P(s'). \tag{1}$$

Suppose now that the interval (a, b) of the length $b - a = l$, representing the whole range of variation of x, is divided into n equal intervals of the length l/n. The repeated application of equation (1) gives

$$P(l) = nP\left(\frac{l}{n}\right).$$

But by requirement 2° $P(l) = 1$ and hence

$$P\left(\frac{l}{n}\right) = \frac{1}{n}.$$

Again, repeated application of (1) gives

$$P\left(\frac{m}{n}l\right) = \frac{m}{n}$$

for any integer $m < n$. Now let us take any interval of length s. For an appropriate m it will contain the interval $\frac{m}{n}l$ and be contained in the interval $\frac{m+1}{n}l$; hence, referring to requirements 1° and 3°, we shall have

$$\frac{m}{n} \leqq P(s) \leqq \frac{m+1}{n}$$

while

$$\frac{m}{n}l \leqq s < \frac{m+1}{n}l,$$

or

$$\frac{m}{n} \leqq \frac{s}{l} < \frac{m+1}{n}.$$

Since $P(s)$ and s/l are contained in the same interval of length $1/n$,

$$\left|P(s) - \frac{s}{l}\right| < \frac{1}{n}$$

and this being true for an arbitrary n, no matter how large, it follows that

$$P(s) = \frac{s}{l}.$$

Thus for a variable x with uniform distribution of probability, the probability of assuming a value belonging to an interval of length s is given by the ratio of s to the length l of the whole range of variation of x.

5. In the general case, when we cannot assume the uniform distribution of probability throughout the whole range of variation of x, we let ourselves be guided by an analogy with a mass distributed continuously over a line. In fact, the distribution of a mass satisfies all the requirements set forth for probability. In particular, the mass Δm contained in an infinitesimal interval $(z, z + \Delta z)$ is also infinitesimal and the mean density

$$\frac{\Delta m}{\Delta z}$$

is generally supposed to tend, with Δz converging to 0, to a limit called "density at the point z." If this density $\rho(z)$ is known, the mass contained in any interval (c, d) is represented by an integral

$$\int_c^d \rho(z)dz.$$

Following this analogy we admit that the *mean density of probability*

$$\frac{P(z, z + \Delta z)}{\Delta z}$$

tends to a limit $f(z)$: density of probability at the point z when the length of the interval Δz tends to 0. Hence, again the probability corresponding to an interval (c, d) will be represented by the integral

$$P(c, d) = \int_c^d f(z)dz.$$

This expression satisfies all the requirements of Sec. 2 if the density of the probability $f(z)$ is subject to two conditions:

(a) $$f(z) \geqq 0 \text{ for all } z \text{ in } (a, b).$$

(b) $$\int_a^b f(z)dz = 1.$$

The second condition implies, of course, the existence of the integral itself. But in all cases of any importance the density is continuous, save for discontinuities of the simplest kind which do not cause any doubts as to the existence of the above integral.

From the general expression of $P(c, d)$ it follows that for an infinitesimal interval $(z, z + dz)$ the probability is given by $f(z)dz$ neglecting infinitesimals of a higher order. For the uniform distribution of probability over an interval of length l the density is constant and $= 1/l$.

In other cases we cannot expect to obtain a definite expression for density unless the variable itself is sufficiently characterized by additional conditions, either hypothetical or implied by the problem. Thus, for instance, in applications of probability to problems of theoretical physics, the physicists have succeeded in obtaining definite probability distributions by invoking physical laws of admitted universal validity together with some plausible hypotheses.

6. The interval containing all possible values of a stochastic variable may be finite or infinite according to the nature of that variable. However, in all cases we may take the largest possible interval from $-\infty$ to $+\infty$; to this end it suffices to *define* the density outside of the originally given interval as being $= 0$. Then the density will be defined for all real values of z and will satisfy the conditions:

(a) $$f(z) \geqq 0 \text{ for all } z$$

(b) $$\int_{-\infty}^{\infty} f(z)dz = 1$$

Furthermore, the probability for x to be in any interval (c, d) will be given by

$$\int_c^d f(z)dz.$$

In particular, taking $c = -\infty$ and writing t instead of d,

$$F(t) = \int_{-\infty}^{t} f(z)dz$$

represents the probability that x will not exceed or will be less than t. Considered as a function of t, $F(t)$ is never decreasing and varies between $F(-\infty) = 0$ and $F(+\infty) = 1$. It is called the "distribution function of probability." In case x has uniform distribution of probability over an interval (a, b) its distribution function is evidently defined as follows:

$$F(t) = 0 \quad \text{for} \quad t < a$$

$$F(t) = \frac{t-a}{b-a} \quad \text{for} \quad a \leqq t \leqq b$$

$$F(t) = 1 \quad \text{for} \quad t > b.$$

Its graph is shown in Fig. 1 on page 240.

7. The definition of mathematical expectation can easily be extended to continuous variables; namely, the expectation of x or the mean value of x is defined by

$$E(x) = \int_{-\infty}^{\infty} zf(z)dz$$

provided this integral exists. Similarly, the mathematical expectation of any function $\varphi(x)$ is given by

$$E[\varphi(x)] = \int_{-\infty}^{\infty} \varphi(z)f(z)dz.$$

Of course, the existence of the integral in the right member is presupposed again. When this integral does not exist, it is meaningless to speak of the mathematical expectation of $\varphi(x)$. The mathematical expectation of the power x^n with positive integer exponent is called the *moment of the order n* or *n*th *moment*. We shall denote it by m_n so that

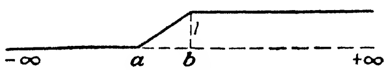

FIG. 1.

$$m_n = \int_{-\infty}^{\infty} z^n f(z)dz.$$

The dispersion D and the standard deviation of x are defined in the same way as in Chap. IX; namely,

$$D = \sigma^2 = E(x - m_1)^2 = \int_{-\infty}^{\infty} (z - m_1)^2 f(z)dz = m_2 - m_1^2.$$

Often it is advisable to consider the mathematical expectation of $|x|^\alpha$ where α may be any real number, ordinarily positive. This expectation is called the "absolute moment of the order α." Its expression is

$$\mu_\alpha = \int_{-\infty}^{\infty} |z|^\alpha f(z)dz,$$

and it is evident that

$$m_{2k} = \mu_{2k}; \qquad |m_{2k+1}| \leqq \mu_{2k+1}.$$

The mathematical expectation of the function

$$e^{itx}$$

where t is a real variable, is of the utmost importance. It is called the "characteristic function" of distribution and is defined by

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itz} f(z)dz.$$

Since $f(z) \geqq 0$ and

$$\int_{-\infty}^{\infty} f(z)dz = 1$$

the integral defining $\varphi(t)$ is always convergent and

$$|\varphi(t)| \leqq 1.$$

The distribution is completely determined by its characteristic function. Because by the Fourier theorem

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} e^{it(z-x)}f(z)dz = f(x)$$

at all points of continuity of $f(x)$. But the left-hand member is

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi(t)e^{-itx}dt$$

by the definition of $\varphi(t)$ and so

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi(t)e^{-itx}dt.$$

8. To illustrate the preceding general explanations we shall now consider a few examples.

Example 1. Let x be a variable with uniform distribution of probability over the interval $(0, l)$. The density of this distribution being constant

$$f(z) = \frac{1}{l}$$

the mean value of x is

$$m_1 = \int_0^l z\frac{dz}{l} = \frac{l}{2}$$

and the second moment

$$m_2 = \int_0^l z^2\frac{dz}{l} = \frac{l^2}{3}.$$

Hence, the square of the standard deviation

$$\sigma^2 = m_2 - m_1^2 = \frac{l^2}{12}.$$

This simple example may be used to illustrate a remark made at the beginning of this chapter, that sometimes it is profitable to substitute for a variable with a finite but large number of values a fictitious continuous variable. Suppose that in flipping a coin n times, we mark heads by 1 and tails by 0, thus obtaining a sequence comprising n units and zeros altogether, disposed in the order of trials. This sequence may be considered as successive digits in the binary representation of a fraction:

$$X = \frac{\alpha_1}{2} + \frac{\alpha_2}{4} + \cdots + \frac{\alpha_n}{2^n}$$

contained between 0 and 1. X may be considered as a stochastic variable with 2^n values each having the probability $1/2^n$. The probability $\Pi(\alpha, \beta)$ that X will be contained in the interval (α, β), or more definitely that X will satisfy the inequalities

$$\alpha < X \leqq \beta$$

is obviously obtained by multiplying the number of integers N contained in the limits

$$2^n\alpha < N \leqq 2^n\beta$$

by $1/2^n$. Now there are exactly

$$[2^n\beta] - [2^n\alpha] = 2^n(\beta - \alpha) + \theta; \qquad -1 < \theta < 1$$

such integers; hence

$$\Pi(\alpha, \beta) = \beta - \alpha + \frac{\theta}{2^n}.$$

If n is even moderately large, this probability is very near to the probability

$$P(\alpha, \beta) = \beta - \alpha$$

that a fictitious variable x with uniform distribution over the interval (0, 1) will assume a value in the interval (α, β). The first two moments of the variable X are, respectively

$$M_1 = \frac{0 + 1 + 2 + \cdots + 2^n - 1}{2^{2n}} = \frac{1}{2} - \frac{1}{2^{n+1}}$$

$$M_2 = \frac{0^2 + 1^2 + 2^2 + \cdots + (2^n - 1)^2}{2^{3n}} = \frac{1}{3} - \frac{1}{2^{n+1}} + \frac{1}{3 \cdot 2^{2n+1}}$$

and differ little from the respective moments $\frac{1}{2}$ and $\frac{1}{3}$ of the fictitious continuous variable. Without losing anything essential, we here gain considerably in simplicity by substituting a fictitious continuous variable for the discontinuous variable X.

Example 2. A thin bar can rotate freely about its middle point P. It is set in motion and after several revolutions comes to a stop pointing toward a point X on a line l. The position of the bar is determined by an angle θ formed by itself and the perpendicular PO dropped from P on l; θ varies between the limits $-\pi/2$ and $\pi/2$ and its distribution is supposed to be uniform. The position of X is determined by its distance $OX = x$ from O, this distance being positive or negative according as X is to the right or to the left of the point O.

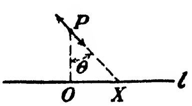

FIG. 2.

It is required to find the distribution of the probability of x. The relation between θ and x is

$$x = a \operatorname{tg} \theta$$

if $OP = a$ or, conversely,

$$\theta = \operatorname{arc} \operatorname{tg} \frac{x}{a}.$$

By differentiation we find the relation between $d\theta$ and dx:

$$d\theta = \frac{adx}{a^2 + x^2}.$$

Now, by hypothesis, the probability that $\sphericalangle OPX$ will be contained between θ and $\theta + d\theta$ is

$$\frac{d\theta}{\pi} = \frac{1}{\pi}\frac{adx}{a^2 + x^2}.$$

And the probability that the distance of X from O will be contained between x and $x + dx$ is the same. Hence, the density of probability for the variable x is

$$f(z) = \frac{1}{\pi}\frac{a}{a^2 + z^2}$$

and the probability corresponding to a finite interval (c, d) is given by

$$P(c, d) = \frac{1}{\pi}\int_c^d \frac{adz}{a^2 + z^2}.$$

For the whole range of variation of x

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{adz}{a^2 + z^2} = 1$$

as it should be. However, we cannot speak of the mean value of x or of moments of higher order, since the integrals

$$\int_{-\infty}^{\infty} \frac{xdx}{a^2 + x^2}, \int_{-\infty}^{\infty} \frac{x^2dx}{a^2 + x^2}, \text{ etc.}$$

have no meaning. But the characteristic function $\varphi(t)$ exists and is given by

$$\varphi(t) = \frac{a}{\pi}\int_{-\infty}^{\infty} \frac{e^{ixt}dx}{a^2 + x^2} = e^{-a|t|}.$$

Example 3. One of the most important distributions (theoretically and practically) is the so-called "Gaussian" or "normal" distribution. The density of this distribution is given by

$$f(z) = Ke^{-h^2(z-a)^2}$$

with three parameters K, h, a. However, only two of these parameters are independent, since we must have

$$\int_{-\infty}^{\infty} f(z)dz = K\int_{-\infty}^{\infty} e^{-h^2(z-a)^2}dz = K\int_{-\infty}^{\infty} e^{-h^2u^2}du = \frac{K\sqrt{\pi}}{h} = 1;$$

whence

$$K = \frac{h}{\sqrt{\pi}}$$

and finally

$$f(z) = \frac{h}{\sqrt{\pi}}e^{-h^2(z-a)^2}.$$

To find the meaning of a and h we observe that the mean value of our variable is

$$\frac{h}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-h^2(z-a)^2}zdz = \frac{h}{\sqrt{\pi}}\int_{-\infty}^{\infty} (z - a)e^{-h^2(z-a)^2}dz + \frac{ah}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-h^2(z-a)^2}dz = a$$

since

$$\int_{-\infty}^{\infty} (z - a)e^{-h^2(z-a)^2}dz = \int_{-\infty}^{\infty} ue^{-h^2u^2}du = 0.$$

Thus a has the meaning of the mean value of the normally distributed variable x. The square of the standard deviation is given by

$$\sigma^2 = \frac{h}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-h^2(z-a)^2}(z - a)^2 dz = \frac{h}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-h^2u^2}u^2du = \frac{1}{2h^2}$$

whence

$$h = \frac{1}{\sigma\sqrt{2}}.$$

Thus for the normally distributed variable with the mean a and standard deviation σ the density of probability is

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(z-a)^2}{2\sigma^2}}.$$

Finally, for the variable $u = x - a$ with the mean value 0 and the same standard deviation, the expression of density takes the simplest form

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{z^2}{2\sigma^2}}$$

and the distribution function of probability is represented by the integral

$$F(t) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{z^2}{2\sigma^2}}dz.$$

The curve of density

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

or the probability curve has a bell-shaped form as shown in the figure corresponding to $\sigma = 1$. It has a single maximum corresponding to $x = 0$ and on both sides of this maximum it rapidly approaches the x axis.

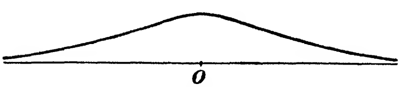

FIG. 3.

The characteristic function of normal distribution has a very simple form. By definition

$$\varphi(t) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2\sigma^2}}e^{izt}dz.$$

But as

$$\int_{-\infty}^{\infty} e^{-\alpha x^2}\cos\beta x dx = \sqrt{\frac{\pi}{\alpha}}e^{-\frac{\beta^2}{4\alpha}} \qquad (\alpha > 0)$$

we find that

$$\varphi(t) = e^{-\frac{\sigma^2t^2}{2}}.$$

The moments of normal distribution (with the mean = 0) can now be easily found. From the definition of the characteristic function it follows that

$$i^n m_n = \left(\frac{d^n\varphi(t)}{dt^n}\right)_{t=0}.$$

In our case

$$\varphi(t) = 1 - \frac{\sigma^2}{2}t^2 + \frac{1}{1 \cdot 2}\left(\frac{\sigma^2}{2}\right)^2 t^4 - \frac{1}{1 \cdot 2 \cdot 3}\left(\frac{\sigma^2}{2}\right)^3 t^6 + \cdots,$$

whence

$$\left(\frac{d^{2k+1}\varphi(t)}{dt^{2k+1}}\right)_{t=0} = 0; \qquad (-1)^k\left(\frac{d^{2k}\varphi(t)}{dt^{2k}}\right)_{t=0} = 1 \cdot 3 \cdot 5 \cdots (2k - 1)\sigma^{2k}.$$

Thus

$$m_{2k+1} = 0$$
$$m_{2k} = 1 \cdot 3 \cdot 5 \cdots (2k - 1)\sigma^{2k}.$$

CASE OF TWO OR MORE VARIABLES

9. By analogy it is easy now to extend the notion of probability to two or more variables considered simultaneously. A pair of special values x, y of two stochastic variables X, Y will be represented geometrically by a point with the coordinates x, y referred to a rectangular system of axes. The domain S of all the possible values of X and Y will be represented by a portion (finite or infinite) of a plane with a definite boundary unless this domain coincides with the whole plane. The probability that the point x, y should belong to an infinitesimal area $dxdy$ will be expressed by the product $\varphi(x, y)\,dxdy$ where the function $\varphi(x, y)$ is again called the density of probability at the point x, y. The density of probability must satisfy two requirements: it is non-negative in the whole domain S and

$$\iint_S \varphi(x, y)dxdy = 1$$

where the double integral is extended over all the domain S. The probability for the point x, y to be located in a given domain σ is then given by the integral

$$\iint_\sigma \varphi(x, y)dxdy$$

extended over σ.

If $\varphi(x, y)$ is a constant in S, the distribution of probability is called uniform. The domain S in this case must be finite and if its area is denoted by the same letter, then

$$\varphi(x, y) = \frac{1}{S}.$$

The probability for the point x, y to be within the domain σ will be given by the ratio

$$\frac{\sigma}{S}$$

denoting the area of the domain σ by σ again.

10. We can always substitute the whole plane for the domain S. To that end it suffices to set

$$\varphi(x, y) = 0$$

in all points not belonging to S. We shall then have

$$\varphi(x, y) \geqq 0$$

everywhere and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x, y)dxdy = 1.$$

By doing so we have the advantage of stating results in a perfectly general form without mentioning the domain S. However, in dealing with particular problems, it is more convenient to consider only those points which can actually represent simultaneous values of the variables. The probability of simultaneous inequalities

$$a < x < b; \qquad c < y < d$$

according to the general definition is represented by the double integral

$$\int_a^b\int_c^d \varphi(x, y)dxdy.$$

This corresponds to the compound probability of two events and we must see that the fundamental theorem of compound probabilities continues to hold. Taking $c = -\infty$, $d = +\infty$ the repeated integral

$$\int_a^b dx\int_{-\infty}^{\infty}\varphi(x, y)dy$$

represents the probability $P(a, b)$ for the variable X (as if it were considered alone without any reference to Y) to have its value in (a, b). The function

$$f(x) = \int_{-\infty}^{\infty}\varphi(x, y)dy$$

represents the density of probability of X. Thus

$$P(a, b) = \int_a^b f(x)dx.$$

In a similar way

$$F(y) = \int_{-\infty}^{\infty}\varphi(x, y)dx$$

represents the density of the probability of Y; and the probability $Q(c, d)$ that this variable has its value in (c, d) is given by

$$Q(c, d) = \int_c^d F(y)dy.$$

Now the double integral

$$\int_a^b \int_c^d \varphi(x, y)dxdy$$

can be written in either of the forms

$$\int_a^b \int_c^d \varphi(x, y)dxdy = \int_a^b f(x)dx \cdot \int_c^d F_1(y)dy$$
$$\int_a^b \int_c^d \varphi(x, y)dxdy = \int_c^d F(y)dy \cdot \int_a^b f_1(x)dx$$

where

$$F_1(y) = \frac{\int_a^b \varphi(x, y)dx}{\int_a^b f(x)dx}; \qquad f_1(x) = \frac{\int_c^d \varphi(x, y)dy}{\int_c^d F(y)dy}$$

may be considered as densities of conditional probabilities, respectively, for Y when it is known that X has a value in (a, b) and for X when it is known that Y has value in (c, d). The preceding expressions for the probability of the simultaneous inequalities

$$a < x < b, \qquad c < y < d$$

have the same form as the theorem of compound probability and may be considered as its extension. The conditional probability for Y to have its value in (c, d) when it is known that X has its value in (a, b) is given by

$$\int_c^d F_1(y)dy.$$

Now, we define variables X and Y as independent when the probability for Y to be in (c, d) is not affected by the knowledge that X belongs to (a, b), which means that

$$\int_c^d F_1(y)dy = \int_c^d F(y)dy$$

or

$$\int_a^b \int_c^d \varphi(x, y)dxdy = \int_c^d F(y)dy \cdot \int_a^b f(x)dx$$

and, since intervals (a, b) and (c, d) are arbitrary,

$$\varphi(x, y) = f(x) \cdot F(y)$$

at points of continuity. Hence, the density of probability for two independent variables is a product of a function of x alone by a function of y alone. Conversely, when this condition is satisfied the variables are independent. For independent variables the probability of the simultaneous inequalities

$$a < x < b$$
$$c < y < d$$

has a simple expression

$$\int_a^b f(x)dx \cdot \int_c^d F(y)dy$$

which is the product of the probability for X to have its value in the interval (a, b) by the probability for Y to have its value in the interval (c, d), in perfect analogy with the compound probability of two independent events.

Finally, the mathematical expectation of any function $\psi(x, y)$ can be defined by

$$E(\psi(x, y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x, y)\varphi(x, y)dxdy$$

provided the integral in the right member exists.

11. It is hardly necessary to dwell at length upon the case of several stochastic variables. A system of particular values $x_1, x_2, \ldots x_n$ of n stochastic variables $X_1, X_2, \ldots X_n$ may be considered as a point in n-dimensional space. The density of probability is a non-negative function $\varphi(x_1, x_2, \ldots x_n)$ defined in the whole space and satisfying the condition

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi(x_1, x_2, \ldots x_n)dx_1dx_2 \cdots dx_n = 1.$$

The probability for a point representing $X_1, X_2, \ldots X_n$ to be located in a given domain σ is given by the integral

$$\int\int \ldots \int \varphi(x_1, x_2, \ldots x_n)dx_1dx_2 \ldots dx_n$$

extended over σ. In the case of uniform distribution of probability, $\varphi(x_1, x_2, \ldots x_n)$ is by definition a constant in a certain finite region of space and $=0$ outside of that region. If V is the volume of that region and v the volume of the domain σ, the ratio v/V gives the probability that a point belongs to σ.

The probability of the simultaneous inequalities

$$a_1 < x_1 < b_1; \qquad a_2 < x_2 < b_2; \ldots a_n < x_n < b_n$$

is given by the integral

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2} \ldots \int_{a_n}^{b_n} \varphi(x_1, x_2, \ldots x_n)dx_1dx_2 \ldots dx_n$$

which, by introduction of the conditional probabilities as in the case of two variables, can be put into the form of a product of n integrals in a manner perfectly analogous to the expression of the probability of a compound event with n components. Finally, the variables are independent if the density $\varphi(x_1, x_2, \ldots x_n)$ is a product of n functions depending only upon $x_1, x_2, \ldots x_n$, respectively, and conversely.

The expression

$$E[\psi(x_1, x_2, \cdots x_n)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \psi\varphi dx_1 dx_2 \cdots dx_n$$

serves to define the mathematical expectation of any function $\psi(x_1, x_2, \ldots x_n)$ of $x_1, x_2, \ldots x_n$.

12. Since in introducing the extended idea of probability we took care to preserve the fundamental theorems of the calculus of probability, we may be sure that other theorems derived from them will hold for continuous variables. In particular, theorems concerning mathematical expectation and the fundamental lemma in Chap. X, Sec. 1, hold for continuous variables. Upon this basis as we have seen was built the proof of the law of large numbers. Hence, this important theorem applies equally to continuous variables.

GEOMETRICAL PROBLEMS

13. A few geometrical problems will afford a good illustration of the foregoing general principles.

Problem 1. A rectilinear segment AB is divided by a point C into two parts $AC = a$, $CB = b$. Points X and Y are taken at random on AC and CB, respectively. What is the probability that AX, XY, BY can form a triangle?

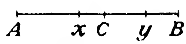

FIG. 4.

Solution. We must first agree upon the meaning of the expression "at random." The idea suggested by this expression implies that the way of selecting points X and Y gives no preference to any point of AC and CB, respectively. Consequently, variables $x = AX$ and $y = BY$ may be assumed to have uniform distribution of probability. The domain of the point x, y is a rectangle $OMPN$ with the sides $OM = a$, $ON = b$. In order that AX, XY, BY can form a triangle the following inequalities must be fulfilled:

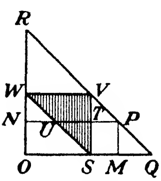

FIG. 5.

$$x < (a + b - x - y) + y \quad \text{or} \quad x < a + b - x$$
$$y < (a + b - x - y) + x \quad \text{or} \quad y < a + b - y$$
$$a + b - x - y < x + y.$$

These inequalities are equivalent to

$$x < \frac{a+b}{2}, \qquad y < \frac{a+b}{2}, \qquad x + y > \frac{a+b}{2}.$$

To interpret them geometrically through P draw a line QPR making $\sphericalangle RQO = 45°$. From the mid-point of QR drop the perpendiculars VS, VW on OX, OY. Then the preceding inequalities limit the position of the point x, y to the shaded area SVW, whose part TSU is contained in the rectangle $OMPN$. Variables x and y are independent and have uniform distribution. Hence, the density of probability of the pair x, y is constant and the probability that the point x, y is in the triangle TSU will be

$$\frac{\text{Area } TSU}{\text{Area } OMPN} = \frac{\frac{1}{2}bb}{ab} = \frac{1}{2}\frac{b}{a}.$$

At the same time this is the probability for AX, XY, BY to form a triangle.

Problem 2. On a line AB two points X_1, X_2 are taken at random. What is the probability that AX_1, X_1X_2, X_2B can form a triangle?

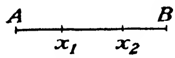

FIG. 6.

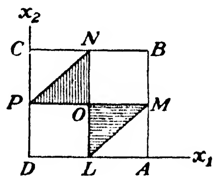

FIG. 7.

Solution. Variables $AX_1 = x_1$, $AX_2 = x_2$ are independent and have uniform distribution of probability. The domain of all possible positions of the point x_1, x_2 is a square with the side $AB = l$. Positions of this point when AX_1, X_1X_2, X_2B form a triangle can be characterized as follows. First, if X_1 precedes X_2, we have

$$x_2 - x_1 < x_1 + l - x_2 \quad \text{or} \quad x_2 - x_1 < \frac{l}{2}$$

$$x_1 < x_2 - x_1 + l - x_2 \quad \text{or} \quad x_1 < \frac{l}{2}$$

$$l - x_2 < x_2 - x_1 + x_1 \quad \text{or} \quad x_2 > \frac{l}{2}$$

which means that x_1, x_2 belongs to the triangle OPN, the definition of which is evident if L, M, N, P are mid-points of the sides of the square $ABCD$. Second, if X_1 follows X_2, we have

$$x_1 - x_2 < \frac{l}{2}; \qquad x_2 < \frac{l}{2}; \qquad x_1 > \frac{l}{2}$$

and these inequalities define the area OLM. Since the distribution of x_1, x_2 is uniform, the required probability is

$$\frac{\text{Area } OLM + \text{Area } ONP}{\text{Area } ABCD} = \frac{\frac{1}{4}ll}{ll} = \frac{1}{4}.$$

Problem 3. A chord is drawn at random in a given circle. What is the probability that it is greater than the side of the equilateral triangle inscribed in that circle?

Solution 1. The position of the chord drawn at random can be determined by its distance from the center of the circle. This distance may vary between 0 and R, the radius of the circle. The chord is greater than the side of the equilateral triangle inscribed in the circle if its distance from the center is less than $\frac{1}{2}R$. Hence, the required probability

$$p_1 = \frac{\frac{1}{2}R}{R} = \frac{1}{2}.$$

Solution 2. Through one end of the chord, draw a tangent AT. The angle φ varying from 0° to 180° determines the position of the chord. If it is greater than the side of the inscribed equilateral triangle, the angle φ must lie between 60° and 120°. Hence the answer

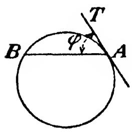

FIG. 8.

$$p_2 = \frac{120° - 60°}{180°} = \frac{1}{3}.$$

The fact that we obtain two different numbers for the same probability seems paradoxical, and the problem itself is known as "Bertrand's paradox." However, going attentively over both solutions, we discover that we are really dealing with two different problems. In the first solution it was assumed that the distance of the chord from the center has uniform distribution, while in the second solution the distribution of the angle φ was taken as uniform. The second solution may be considered reasonable if a thin bar or a needle can rotate freely about A and if, being set in motion, it determines the chord AB by its ultimate position. On the other hand, the first solution is acceptable if a circular disk is thrown upon a board ruled with parallel lines distant from one another by the diameter of the disk. The intersection of the disk with one of the lines determines a chord, and the probability that it is greater than the side of the inscribed equilateral triangle can reasonably be assumed to be $\frac{1}{2}$.

A general remark applies to all problems of this kind. When a certain geometrical element, such as a point or a line, is supposed to be taken at random, it should be clearly indicated by what kind of mechanism this is to be done. Only then the hypothetically assumed distribution can be put to an experimental test and either confirmed (approximately) or rejected.

14. Buffon's Needle Problem. A board is ruled with equidistant parallel lines, the width of the strip between two consecutive lines being d. A needle so fine that it can be likened to a rectilinear segment of the length $l < d$ is thrown on the board. What is the probability that the needle will intersect one of the lines (naturally not more than one)?

Solution. This is the oldest problem dealing with geometrical probabilities. It was mentioned by Buffon, the celebrated French naturalist of the eighteenth century, in the Proceedings of the Paris Academy of Sciences (1733) and later reproduced with its solution in Buffon's book "Essai d'arithmétique morale," published in 1777.

Let us determine the position of the needle by the distance $OP = x$ of its middle point from the nearest line, and the acute angle φ between OP and the needle. Variables x and φ may be considered as independent. Furthermore, x and φ vary respectively between 0 and $\frac{1}{2}d$, and 0 and $\pi/2$. As a hypothesis we assume the distribution of probability for

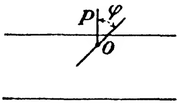

FIG. 9.

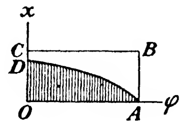

FIG. 10.

x and φ as uniform. The domain of x, φ is a rectangle $OABC$ with $OA = \pi/2$, $OC = d/2$. Now, the needle intersects one of the lines if

$$x < \frac{l}{2} \cos \varphi$$

and then the point x, φ lies in the shaded area below the curve

$$x = \frac{l}{2} \cos \varphi.$$

Since the distribution of x, φ is uniform, the required probability will be

$$p = \frac{\text{Area } OAD}{\text{Area } OABC}.$$

But

$$\text{Area } OAD = \frac{l}{2}\int_0^{\frac{\pi}{2}} \cos \varphi d\varphi = \frac{l}{2}$$

$$\text{Area } OABC = \frac{\pi}{2} \cdot \frac{d}{2}$$

and consequently

$$p = \frac{2l}{\pi d}.$$

On pages 112–113 an account was given of experiments made by several authors in connection with Buffon's problem. They all show good agreement with the theory and indirectly confirm the hypothesis assumed in deriving the above expression for probability.

15. Extension of Buffon's Problem. A thin plate in the shape of a convex polygon, of dimensions so small that it cannot intersect two of the lines simultaneously, is thrown on a board ruled, as in Buffon's needle problem. What is the probability that the boundary of the plate will intersect one of the lines?

Solution. Suppose that the polygonal boundary has five sides. Let these sides (and their lengths) be denoted by

$$\alpha,\ \beta,\ \gamma,\ \delta,\ \epsilon.$$

Each of them is shorter than the distance d between two consecutive lines. On account of convexity, a line can intersect either none or two (and only two) sides. Accordingly, combining sides in pairs, we can distinguish 10 mutually exclusive cases and denote their probabilities by

$$(\alpha\beta),\ (\alpha\gamma),\ (\alpha\delta),\ (\alpha\epsilon),\ (\beta\gamma),\ (\beta\delta),\ (\beta\epsilon),\ (\gamma\delta),\ (\gamma\epsilon),\ (\delta\epsilon).$$

The required probability will be given by the sum

$$p = (\alpha\beta) + (\alpha\gamma) + (\alpha\delta) + (\alpha\epsilon) + (\beta\gamma) + (\beta\delta) + (\beta\epsilon) + (\gamma\delta) + (\gamma\epsilon) + (\delta\epsilon).$$

On the other hand, the side α can be intersected by a line in four mutually exclusive ways; namely, together with β or γ, or δ, or ϵ. Hence, if (α) is the probability of intersection

$$(\alpha) = (\alpha\beta) + (\alpha\gamma) + (\alpha\delta) + (\alpha\epsilon),$$

and similarly

$$\begin{aligned}(\beta) &= (\beta\alpha) + (\beta\gamma) + (\beta\delta) + (\beta\epsilon)\\ (\gamma) &= (\gamma\alpha) + (\gamma\beta) + (\gamma\delta) + (\gamma\epsilon)\\ (\delta) &= (\delta\alpha) + (\delta\beta) + (\delta\gamma) + (\delta\epsilon)\\ (\epsilon) &= (\epsilon\alpha) + (\epsilon\beta) + (\epsilon\gamma) + (\epsilon\delta),\end{aligned}$$

whence

$$(\alpha) + (\beta) + (\gamma) + (\delta) + (\epsilon) = 2p.$$

But

$$(\alpha) = \frac{2\alpha}{\pi d}, \qquad (\beta) = \frac{2\beta}{\pi d}, \qquad (\gamma) = \frac{2\gamma}{\pi d}, \qquad (\delta) = \frac{2\delta}{\pi d}, \qquad (\epsilon) = \frac{2\epsilon}{\pi d},$$

and consequently

$$p = \frac{\alpha + \beta + \gamma + \delta + \epsilon}{\pi d} = \frac{P}{\pi d}$$

where P is the perimeter of the polygonal boundary. Evidently this result is perfectly general. Since it does not depend upon the number of sides, by passage to the limit, it can be extended to the case of a plate bounded by any convex curve.

16. Second Solution of Buffon's Problem. Barbier has given another extremely ingenious solution of Buffon's problem and of its extension. Let $f(l)$ be an unknown probability that the needle will intersect a line.

Imagine that the needle is divided into two parts l' and l''. Evidently a line intersects the needle if, and only if, it intersects either the first or the second part. Hence, by the theorem of total probabilities

$$f(l) = f(l') + f(l''),$$

whence, as in Sec. 4, we conclude

$$f(l) = Cl$$

where C is a constant independent of l. The whole question is how to determine this constant. Barbier's ingenious idea was to let this problem depend on the solution of another one: A polygonal line (convex or not) is thrown upon the board; what is the mathematical expectation of the number of points of intersection? The perimeter of the polygonal line can be subdivided into n rectilinear parts $a_1, a_2, \ldots a_n$ all less than d. With these n parts we can associate n variables $x_1, x_2, \ldots x_n$, such that

$x_i = 1$ if one of the lines intersects a_i
$x_i = 0$ otherwise.

The sum

$$s = x_1 + x_2 + \cdots + x_n$$

evidently gives the total number of the points of intersection. Hence

$$E(s) = E(x_1) + E(x_2) + \cdots + E(x_n)$$

and, if p_i is the probability of intersection of a_i with one (and only one) line,

$$E(x_i) = p_i.$$

But, according to the previous result,

$$p_i = Ca_i.$$

Hence, we have a perfectly general formula

$$E(s) = C(a_1 + a_2 + \cdots + a_n) = CP$$

where P is the perimeter of the polygonal line. The result holds for any curvilinear arc (closed or not) as can be seen by the method of limits.

This formula applied to a circle with the diameter d gives

$$C \cdot \pi d = 2$$

since such a circle has always exactly two points of intersection with the lines of the system. Thus we find that

$$C = \frac{2}{\pi d}$$

and

$$f(l) = \frac{2l}{\pi d}$$

as obtained before. For a closed convex line of sufficiently small dimensions only two cases are possible: two intersections (probability p), or none (probability $1 - p$), whence $E(s) = 2p$ and

$$2p = \frac{2P}{\pi d}$$

or

$$p = \frac{P}{\pi d}$$

in agreement with the result obtained in Sec. 15.

17. Laplace's Problem. A board is covered with a set of congruent rectangles as shown in the figure, and a thin needle is thrown on the board. Supposing that the needle is shorter than the smaller sides of the rectangles, find the probability that the needle will be entirely contained in one of the rectangles of the set.

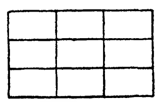

FIG. 11.

Solution. Let $AB = a$, $AD = b$ be the sides of the rectangle which contains the middle point of the needle, the length of which is

$$l \ (l < a, l < b).$$

Taking AB and AD for coordinate axes, the position of the needle is determined by two coordinates x, y of its middle point and the angle φ formed by the needle with the x axis. We may consider x, y, φ as three independent variables with uniform distribution of probability. The domain filled up with all possible points x, y, φ is a parallelepipedon

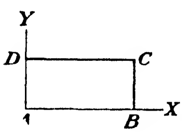

FIG. 12.

$$0 < x < a; \qquad 0 < y < b; \qquad -\frac{\pi}{2} < \varphi < \frac{\pi}{2}$$

and the distribution of probability throughout this domain is uniform. To characterize the domain of points representing positions of the middle point of the needle when it is located entirely within $ABCD$ we consider the sections of that domain by planes $\varphi =$ constant and their

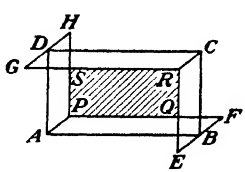

FIG. 13.

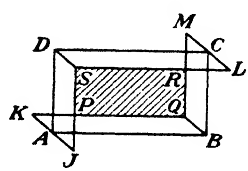

FIG. 14.

projections on the plane xy. These projections are represented by the shaded areas in Figs. 13 and 14 corresponding to positive and negative φ, respectively.

In Fig. 13

$$\measuredangle PAB = \varphi; \qquad AP \| BF \| CR \| DG$$

and $AP = BE = BF = CR = DG = DH = \frac{1}{2}l$.

Similarly, in the second figure

$$\measuredangle JAB = \varphi; \qquad AJ \| BQ \| CL \| DS$$

and $AJ = AK = BQ = CL = CM = DS = \frac{1}{2}l$.

The area of the rectangle $PQRS$ corresponding to these two cases can be expressed as follows:

$$\text{Area } PQRS = (a - l\cos\varphi)(b - l\sin\varphi) = ab - l(b\cos\varphi + a\sin\varphi) + \\ + l^2\sin\varphi\cos\varphi,$$

$$\text{Area } PQRS = (a - l\cos\varphi)(b + l\sin\varphi) = ab - l(b\cos\varphi - a\sin\varphi) - \\ - l^2\sin\varphi\cos\varphi.$$

Without distinguishing positive and negative values of φ, we may write

$$F(\varphi) = \text{area } PQRS = ab - bl\cos\varphi - la|\sin\varphi| + \tfrac{1}{2}l^2|\sin 2\varphi|.$$

The volume of the domain representing positions of the needle entirely within $ABCD$ is:

$$v = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} F(\varphi)d\varphi = \pi ab - 2bl - 2al + l^2$$

while

$$V = \pi ab$$

is the volume of the domain

$$0 < x < a, \qquad 0 < y < b, \qquad -\frac{\pi}{2} < \varphi < \frac{\pi}{2}.$$

Hence, the required probability is:

$$p = 1 - \frac{2l(a + b) - l^2}{\pi ab}$$

and the complementary probability for the needle to intersect the boundary of one of the rectangles is:

$$q = \frac{2l(a + b) - l^2}{\pi ab}.$$

Buffon's problem may be considered as a limiting case when $a = \infty$ and, indeed, by setting $a = \infty$, we find that

$$q = \frac{2l}{\pi b}$$

in conformity with the result in Sec. 14.

These examples may suffice to give an idea of problems in geometric probabilities. Sylvester, Crofton, and others have enriched this field by extremely ingenious methods of evaluating, or rather of avoiding evaluations, of very complicated multiple integrals. However, from the standpoint of principles, these investigations, ingenious as they are, do not contribute much to the general theory of probability.

Problems for Solution

1. A point X is taken at random on a rectilinear segment $AB = l$ whose middle point is O. What is the probability that AX, BX, and AO can form a triangle? The distribution of $AX = x$ is assumed to be uniform. *Ans.* $\frac{1}{2}$.

2. Two points X_1, X_2 are taken at random on $AB = l$. Assuming uniform distribution of probability, what is the mathematical expectation of any power n of the distance between X_1 and X_2?

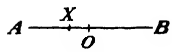

FIG. 15.

$$\textit{Ans.} \int_0^l \int_0^l |x_1 - x_2|^n \frac{dx_1 dx_2}{l^2} = \frac{2l^n}{(n+1)(n+2)}.$$

3. Three points X_1, X_2, X_3 are taken at random on AB. What is the probability that X_3 lies between X_1 and X_2?

Ans. $\frac{1}{3}$, assuming uniform distribution of probability.

4. A rectilinear segment AB is divided into four equal parts

$$AC = CO = OD = DB.$$

Supposing that the distribution of probability is symmetric with respect to O, let P be the probability that a point selected at random on AB will be between C and D. Also, let Q be the probability that the middle point between two points selected at random will be between C and D. Prove that $Q > \frac{1 + P^2}{2}$.

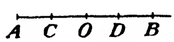

FIG. 16.

HINT: The middle point of a segment X_1X_2 is surely between C and D if : (i) X_1 and X_2 are in CO; or (ii) X_1 and X_2 are in OD; or (iii) X_1 and X_2 are on opposite sides of O.

5. Two points X_1, X_2 are chosen at random in a circle of radius r. Assuming uniform distribution of probability, what is the mathematical expectation of their distance? *Ans.* Denoting the required mathematical expectation by M, we have

$$\pi^2 r^4 M = \int_0^{2\pi} \int_0^{2\pi} F(r, \theta, \theta') d\theta d\theta'$$

where

$$F(r, \theta, \theta') = \int_0^r \int_0^r \sqrt{\rho^2 + \rho'^2 - 2\rho\rho' \cos(\theta - \theta')}\, \rho\rho' d\rho d\rho'.$$

Hence, varying r by dr

$$dF = 2r dr \int_0^r \sqrt{r^2 + \rho^2 - 2r\rho \cos(\theta - \theta')}\, \rho d\rho$$

and

$$d(\pi^2r^4M) = 4\pi r dr\int_0^{2\pi}\int_0^r\sqrt{r^2+\rho^2-2r\rho\cos\omega}\,\rho d\rho d\omega.$$

By introduction of new polar coordinates the integral in the right member can be exhibited as

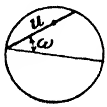

FIG. 17.

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}d\omega\int_0^{2r\cos\omega}u^2du = \frac{16}{3}r^3\int_0^{\frac{\pi}{2}}\cos^3\omega d\omega = \frac{32}{9}r^3.$$

Thus

$$d(\pi r^4M) = \tfrac{128}{9}r^4dr$$

whence

$$M = \frac{128r}{45\pi}.$$

6. A board is covered with congruent rectangles as in Laplace's problem. A coin the diameter of which is less than the smaller side of the rectangles is thrown on the board. What is the probability that it will be partly in one rectangle and partly in another? *Ans.* a, b, r being respectively the sides of the rectangles and radius of the coin, the required probability is

$$\frac{2r(a+b-2r)}{ab}.$$

7. Solve Buffon's problem when the needle is longer than the distance between two consecutive lines. *Ans.* The probability for the needle to intersect at least one line is

$$p = \frac{2l}{\pi d}(1-\sin\varphi_0) + \frac{2\varphi_0}{\pi}$$

where φ_0 is determined by $\cos\varphi_0 = d/l$.

8. A board is covered with congruent triangles whose sides are a, b, c. A needle whose length is less than the shortest altitude of any one of these triangles is thrown on the board. What is the probability that the needle will be contained entirely within one of the triangles? *Ans.* The required probability is

$$1 + \frac{(Aa^2+Bb^2+Cc^2)l^2}{2\pi Q^2} - \frac{(4a+4b+4c-3l)l}{2\pi Q}$$

where A, B, C are angles opposite to sides a, b, c and Q is double the area of the triangle. For equilateral triangles

$$1 + \frac{2}{3}\left(\frac{l}{a}\right)^2 - \frac{l\sqrt{3}}{\pi a}\left(4 - \frac{l}{a}\right).$$

9. On each of the circles O_1, O_2, O_3, . . . with respective radii r_1, r_2, r_3, . . . points M_1, M_2, M_3, . . . are taken at random. Supposing that the series

$$r_1 + r_2 + r_3 + \cdots$$

is divergent, while the series

$$r_1^2 + r_2^2 + r_3^2 + \cdots$$

is convergent, prove that the probability that the length of the vector

$$\overline{OM} = \overline{O_1M_1} + \overline{O_2M_2} + \overline{O_3M_3} + \cdots + \overline{O_nM_n}$$

will be $> R$ tends to 0 as $R \to \infty$ no matter how large n is.

Indication of Solution. Let $x_1, x_2, \ldots x_n$; $y_1, y_2, \ldots y_n$ be components of $OM_1, OM_2, \ldots OM_n$ on two rectangular axes OX, OY. Then

FIG. 18.

$$E(x_i) = E(y_i) = 0$$

$$E(x_i^2) = E(y_i^2) = \frac{r_i^2}{2}.$$

By Tshebysheff's lemma (Chap. X, Sec. 1) the probabilities Q and Q' of the inequalities

$$|x_1 + x_2 + \cdots + x_n| > t\sqrt{\frac{r_1^2 + r_2^2 + r_3^2 + \cdots}{2}} = t\sqrt{\frac{G}{2}}$$

$$|y_1 + y_2 + \cdots + y_n| > t\sqrt{\frac{r_1^2 + r_2^2 + r_3^2 + \cdots}{2}} = t\sqrt{\frac{G}{2}}$$

are both less than $1/t^2$. Now, if the length $OM > R$ then either

$$|x_1 + x_2 + \cdots + x_n| > \frac{R}{\sqrt{2}} = t\sqrt{\frac{G}{2}}$$

or

$$|y_1 + y_2 + \cdots + y_n| > \frac{R}{\sqrt{2}} = t\sqrt{\frac{G}{2}}.$$

Hence, the probability P for the length of OM to be $> R$ is less than $Q + Q'$; that is,

$$P < Q + Q' < \frac{2G}{R^2}.$$

10. Prove that

$$\lim_{n=\infty} \int_0^1 \int_0^1 \cdots \int_0^1 \frac{x_1^2 + x_2^2 + \cdots + x_n^2}{x_1 + x_2 + \cdots + x_n} dx_1 dx_2 \cdots dx_n = \frac{2}{3}.$$

HINT: Considering $x_1, x_2, \ldots x_n$ as continuous stochastic variables with uniform distribution over the interval (0, 1) prove with the help of Tshebysheff's inequality that the probability of

$$\frac{2}{3} - \epsilon < \frac{x_1^2 + x_2^2 + \cdots + x_n^2}{x_1 + x_2 + \cdots + x_n} < \frac{2}{3} + \epsilon$$

for any $\epsilon > 0$ tends to 1 as $n \to \infty$.

References

E. CZUBER: "Geometrische Wahrscheinlichkeiten und Mittelwerte," Leipzig, 1884.

E. CZUBER: "Wahrscheinlichkeitsrechnung," 1, pp. 75–109, Leipzig, 1908.

H. POINCARÉ: "Calcul des probabilités," pp. 118–152, Paris, 1912.

W. CROFTON: On the Theory of Local Probability Applied to Straight Lines Drawn at Random in a Plane, the Method Used Being Also Extended to the Proof of Certain New Theorems in Integral Calculus, *Philos. Trans.*, vol. 158, 1868.

W. CROFTON: Probability, "Encyclopaedia Britannica," 9th ed.

CHAPTER XIII

THE GENERAL CONCEPT OF DISTRIBUTION

1. In dealing with continuous stochastic variables we have introduced the important concept of the function of distribution. Denoting the density of probability by $f(z)$, this function was defined by

$$F(t) = \int_{-\infty}^{t} f(z)dz$$

and it represents the probability of the inequality

$$x < t.$$

For a variable with a finite number of values the function of distribution can be defined as the sum

$$F(t) = \sum_{x_i < t} p_i$$

where $p_1, p_2, \ldots p_n$ are respective probabilities of all possible values $x_1, x_2, \ldots x_n$ of the variable x. The notation $x_i < t$ is intended to show that the summation is extended over all values of x less than t. Again, $F(t)$ for any real t represents the probability of the inequality

$$x < t.$$

In this case $F(t)$ is a discontinuous function, never decreasing and varying between $F(-\infty) = 0$ and $F(+\infty) = 1$. Its discontinuities are located at the points $x_1, x_2, \ldots x_n$ and are such that

$$F(x_i + 0) - F(x_i - 0) = p_i,$$

denoting, in the customary way,

$$F(x_i + 0) = \lim F(x_i + \epsilon)$$
$$F(x_i - 0) = \lim F(x_i - \epsilon)$$

when ϵ, through positive values, converges to 0. To represent $F(t)$ graphically we note that

$$\begin{aligned} F(t) &= 0 && \text{for} \quad t < x_1 \\ F(t) &= p_1 && \text{for} \quad x_1 < t < x_2 \\ F(t) &= p_1 + p_2 && \text{for} \quad x_2 < t < x_3 \\ &\cdots\cdots \\ F(t) &= p_1 + p_2 + \cdots + p_n && \text{for} \quad x_n < t. \end{aligned}$$

As for the value of $F(t)$ at the point $t = x_i$, it is $F(x_i - 0)$. Hence, the graph of $F(t)$ consists of rectilinear segments as shown in the figure (for $n = 4$; $x_1 = -2$; $x_2 = 0$; $x_3 = 1$; $x_4 = 3$; $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$) and belongs to the so-called step lines.

Thus, in case of a continuous variable the distribution function is given by an integral, and in case of a discontinuous variable, by a sum. In stating theorems equally true for continuous and discontinuous variables, it would be tedious always to distinguish these two cases. The question naturally arises whether it is possible to represent distribution functions, moments, and similar quantities by using new symbols equally applicable to continuous and discontinuous variables. In a similar kind of investigation Stieltjes was confronted with the same

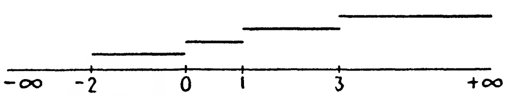

FIG. 19.

difficulties and he succeeded in overcoming them by introducing a new kind of integrals known as "Stieltjes' integrals."

STIELTJES' INTEGRALS

2. Let $\varphi(x)$ be a never decreasing function defined in the interval $a \leqq x \leqq b$. For any particular value of the argument both the limits (for ϵ converging to 0 through positive values)

$$\lim \varphi(x_0 + \epsilon) = \varphi(x_0 + 0)$$
$$\lim \varphi(x_0 - \epsilon) = \varphi(x_0 - 0)$$

exist. Since evidently

$$\varphi(x_0 - 0) \leqq \varphi(x_0) \leqq \varphi(x_0 + 0),$$

x_0 is a point of continuity of $\varphi(x)$ if

$$\varphi(x_0 - 0) = \varphi(x_0 + 0).$$

If, however,

$$\varphi(x_0 - 0) < \varphi(x_0 + 0)$$

$\varphi(x)$ is discontinuous at x_0, and the difference

$$m_0 = \varphi(x_0 + 0) - \varphi(x_0 - 0)$$

gives the measure of discontinuity or simply discontinuity. Since for any number of points of discontinuity $x_0, x_1, \ldots x_n$ the sum of discontinuities

$$m_0 + m_1 + \cdots + m_n \leqq \varphi(b) - \varphi(a)$$

the points of discontinuity form a countable set. For there are only a finite number of discontinuities above any given number, so that, considering the sequence

$$\delta > \delta_1 > \delta_2 > \cdots$$

tending to 0, there is only a finite number of points with discontinuities $>\delta$; also a finite number of points with discontinuities $\leqq \delta$ and $>\delta_1$, and so on. It follows that points of discontinuity can be arranged into a single sequence and hence form a countable set.

It may happen, however, that $\varphi(x)$ may have discontinuities in any interval, no matter how small; but at any rate there are points of continuity in any interval. If $\varphi(x_0 + \epsilon) > \varphi(x_0 - \epsilon)$ for all sufficiently small $\epsilon > 0$ the point x_0 is called a "point of increase" of $\varphi(x)$. In particular, any point of discontinuity is a point of increase.

3. Let $f(x)$ be a continuous function in the interval $a \leqq x \leqq b$. By inserting points $x_1 < x_2 < \cdots < x_n$ this interval is subdivided into $n + 1$ partial intervals. In each of these we arbitrarily select points $\xi_0, \xi_1, \ldots \xi_n$ and form the sum

$$S = f(\xi_0)[\varphi(x_1) - \varphi(a)] + f(\xi_1)[\varphi(x_2) - \varphi(x_1)] + \cdots + \\ + f(\xi_n)[\varphi(b) - \varphi(x_n)].$$

It can be proved in the same way as for ordinary integrals that when all intervals

$$x_1 - a, x_2 - x_1, \ldots b - x_n$$

tend to zero uniformly, the sum S tends to a definite limit. This limit, called Stieltjes' integral, does not depend upon the manner of subdividing the interval (a, b) or upon the choice of points $\xi_0, \xi_1, \ldots \xi_n$. It has a perfectly definite value as soon as $f(x)$ and $\varphi(x)$ (together with a, b) are given, and accordingly is denoted by

$$\int_a^b f(x)d\varphi(x).$$

In case $\varphi(x)$ has a continuous derivative, $d\varphi(x)$ can be interpreted as the ordinary differential; Stieltjes' integral then coincides with the ordinary one. In other cases $d\varphi(x)$ is a new symbol introduced as a reminder of the origin of Stieltjes' integral. In particular, if $\varphi(x)$ is a step function with discontinuities $p_1, p_2, p_3, \ldots$ at the points $x_1, x_2, x_3, \ldots$, Stieltjes' integral coincides with the sum

$$\Sigma p_i f(x_i)$$

which is a finite sum or an absolutely convergent infinite series according as the set of points of discontinuity is finite or infinite.

Stieltjes' integrals possess many properties of ordinary integrals. For instance, the mean-value theorem holds for them in the form:

$$\int_a^b f(x)d\varphi(x) = f(\xi)[\varphi(b) - \varphi(a)]$$

where $a \leqq \xi \leqq b$. Also, if $f(x)$ has a continuous derivative, we have an analogue for the integration by parts

$$\int_a^b f(x)d\varphi(x) = f(b)\varphi(b) - f(a)\varphi(a) - \int_a^b \varphi(x)df(x)$$

where $df(x)$ means an ordinary differential and the integral in the right member is an ordinary integral. However, some important properties of ordinary integrals do not hold universally for Stieltjes' integrals. For instance, considered as functions of b or a, they may have discontinuities.

In the definition of Stieltjes' integral it was assumed that a and b were finite numbers. Stieltjes' integral over the interval $-\infty, +\infty$ is defined in an ordinary way as being the limit of

$$\int_a^b f(x)d\varphi(x)$$

when a and b tend independently to $-\infty$ and $+\infty$, respectively. In other words,

$$\int_{-\infty}^{\infty} f(x)d\varphi(x) = \lim \int_a^b f(x)d\varphi(x) \qquad \text{when} \qquad a \to -\infty, \qquad b \to +\infty,$$

provided this limit exists. If it does not exist, the symbol

$$\int_{-\infty}^{\infty} f(x)d\varphi(x)$$

has no meaning.

THE GENERAL CONCEPT OF DISTRIBUTION

4. The most general type of distribution function of probability, covering all imaginable cases, is given by a never decreasing function $F(t)$ defined for all real values of t and varying from $F(-\infty) = 0$ to $F(+\infty) = 1$. If at points of discontinuity we set

$$F(t) = F(t - 0),$$

then for any t the probability of the inequality

$$x < t$$

will be given by $F(t)$. Also, the probability of the inequalities

$$t_1 \leqq x < t_2$$

will be

$$F(t_2) - F(t_1).$$

The case of continuous $F(t)$, having a continuous derivative $f(t)$ (save for a finite set of points of discontinuity), corresponds to a continuous variable distributed with the density $f(t)$, since

$$F(t) = \int_{-\infty}^{t} f(x)dx.$$

If $F(t)$ is a step function with a finite number of discontinuities, it characterizes the distribution of probability of a variable with a finite number of values. Finally, if $F(t)$ is a step function with an infinite set of discontinuities distributed without density, it corresponds to a variable whose values can be arranged in a sequence according to their magnitude. These are the most important types of variables considered in the calculus of probability, and for all of them the distribution function can be represented by Stieltjes' integral

$$F(t) = \int_{-\infty}^{t} dF(x).$$

The mathematical expectation of any continuous function $f(t)$ is defined by Stieltjes' integral

$$E(f(t)) = \int_{-\infty}^{\infty} f(t)dF(t)$$

provided it has a meaning. In particular, moments of the order n (n positive integer) and absolute moments of the order α (α real) are defined, respectively, by

$$m_n = \int_{-\infty}^{\infty} t^n dF(t)$$
$$\mu_\alpha = \int_{-\infty}^{\infty} |t|^\alpha dF(t)$$

and we always have

$$|m_n| \leqq \mu_n.$$

Finally,

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x)$$

is the characteristic function of distribution. Since the integral exists for any real t, this function is defined for all real values t and satisfies the inequality

$$|\varphi(t)| \leqq 1.$$

INEQUALITIES FOR MOMENTS

5. Moments of any distribution satisfy certain inequalities, which it is important to know. They all are particular cases of the following very general inequality due to Liapounoff.

Liapounoff's Inequality. Let a, b, c be three real numbers satisfying the inequalities

$$a \geqq b \geqq c \geqq 0$$

and μ_a, μ_b, μ_c absolute moments of orders a, b, c for an arbitrary distribution. Then the following inequality holds:

$$\mu_b^{a-c} \leqq \mu_c^{a-b}\mu_a^{b-c}.$$

Proof. *a.* Let $p_1, p_2, \ldots p_n$; $x_1, x_2, \ldots x_n$ be positive numbers and

$$\varphi(\alpha) = p_1x_1^\alpha + p_2x_2^\alpha + \cdots + p_nx_n^\alpha.$$

Then for arbitrary real numbers $s_1, s_2, \ldots s_p$ the following inequality holds:

$$\varphi\left(\frac{s_1 + s_2 + \cdots + s_p}{p}\right)^p \leqq \varphi(s_1)\varphi(s_2) \cdots \varphi(s_p). \tag{1}$$

For $p = 2$ this inequality follows immediately from the known inequality due to Cauchy:

$$\left(\sum_1^n a_ib_i\right)^2 \leqq \sum_1^n a_i^2 \cdot \sum_1^n b_i^2$$

by taking in it

$$a_i = \sqrt{p_i}x_i^{\frac{s_1}{2}}, \qquad b_i = \sqrt{p_i}x_i^{\frac{s_2}{2}}.$$

For $p = 4$ we have

$$\varphi\left(\frac{s_1 + s_2 + s_3 + s_4}{4}\right)^4 \leqq \varphi\left(\frac{s_1 + s_2}{2}\right)^2\varphi\left(\frac{s_3 + s_4}{2}\right)^2 \leqq \varphi(s_1)\varphi(s_2)\varphi(s_3)\varphi(s_4)$$

and continuing in the same manner we find in general that

$$\varphi\left(\frac{s_1 + s_2 + \cdots + s_{2^m}}{2^m}\right)^{2^m} \leqq \varphi(s_1)\varphi(s_2) \cdots \varphi(s_{2^m}).$$

Let m be taken so that $2^m > p$ and let us take in the last inequality

$$s_{p+1} = s_{p+2} = \cdots = s_{2^m} = s = \frac{s_1 + s_2 + \cdots + s_p}{p}.$$

Since

$$\frac{s_1 + s_2 + \cdots + s_{2^m}}{2^m} = \frac{ps + (2^m - p)s}{2^m} = s$$

we shall have

$$\varphi(s)^{2^m} \leqq \varphi(s_1)\varphi(s_2) \cdots \varphi(s_p)\varphi(s)^{2^m-p},$$

whence

$$\varphi(s)^p \leqq \varphi(s_1)\varphi(s_2) \cdot \cdot \cdot \varphi(s_p),$$

which is inequality (1).

b. Let $a \geqq b \geqq c \geqq 0$ be integers. Taking $p = a - c$; $s_1 = s_2 = \cdot \cdot \cdot = s_{a-b} = c$; $s_{a-b+1} = \cdot \cdot \cdot = s_{a-c} = a$, we have

$$\frac{s_1 + s_2 + \cdot \cdot \cdot + s_{a-c}}{a - c} = \frac{(a - b)c + (b - c)a}{a - c} = b$$

and consequently, by virtue of (1),

$$\left(\sum_1^n p_i x_i^b\right)^{a-c} \leqq \left(\sum_1^n p_i x_i^c\right)^{a-b} \left(\sum_1^n p_i x_i^a\right)^{b-c}. \tag{2}$$

If $a = p/s$, $b = q/s$, $c = r/s$ are rational numbers $(a \geqq b \geqq c \geqq 0)$, it suffices to take, in (2), p, q, r instead of a, b, c, replace x_i by $x_i^{\frac{1}{s}}$, and raise both members to the power $1/s$ to ascertain that (2) holds for rational a, b, c. Finally, the passage to the limit makes it clear that (2) holds for real a, b, c, provided $a \geqq b \geqq c \geqq 0$.

c. Let the interval A to B be subdivided into partial intervals by inserting numbers $t_1 < t_2 < \cdot \cdot \cdot < t_n$ between A and B and let

$$p_0 = F(t_1) - F(A), \quad p_1 = F(t_2) - F(t_1), \ . \ . \ . \ p_n = F(B) - F(t_n)$$
$$x_0 = |A|, \quad x_1 = |t_1|, \ . \ . \ . \ x_n = |t_n|.$$

Then the three sums

$$\sum_0^n p_i x_i^b, \ \sum_0^n p_i x_i^c, \ \sum_0^n p_i x_i^a$$

will tend to the respective limits

$$\int_A^B |t|^b dF(t), \ \int_A^B |t|^c dF(t), \ \int_A^B |t|^a dF(t)$$

when all differences $A - t_1$, $t_2 - t_1$, . . . $B - t_n$ tend to 0 uniformly. Hence, passing to the limit in (2), we get

$$\left(\int_A^B |t|^b dF(t)\right)^{a-c} \leqq \left(\int_A^B |t|^c dF(t)\right)^{a-b} \cdot \left(\int_A^B |t|^a dF(t)\right)^{b-c};$$

and finally, letting A tend to $-\infty$ and B to $+\infty$,

$$\left(\int_{-\infty}^{\infty} |t|^b dF(t)\right)^{a-c} \leqq \left(\int_{-\infty}^{\infty} |t|^c dF(t)\right)^{a-b} \cdot \left(\int_{-\infty}^{\infty} |t|^a dF(t)\right)^{b-c}$$

or

$$\mu_b^{a-c} \leqq \mu_c^{a-b} \mu_a^{b-c}$$

as stated.

Taking $b = \frac{a + c}{2}$, Liapounoff's inequality becomes

$$\mu_{\frac{a+c}{2}}^{a-c} \leqq \mu_c^{\frac{a-c}{2}} \mu_a^{\frac{a-c}{2}},$$

whence

$$\mu_{\frac{a+c}{2}}^2 \leqq \mu_c \mu_a$$

for any two real positive numbers a and c. If k and l are two positive integers and we take $c = 2k$, $a = 2l$, then

$$\mu_{k+l}^2 \leqq \mu_{2k}\mu_{2l}$$

or

$$m_{k+l}^2 \leqq m_{2k}m_{2l}$$

since

$$|m_{k+l}| \leqq \mu_{k+l} \qquad \text{and} \qquad \mu_{2k} = m_{2k}, \qquad \mu_{2l} = m_{2l}.$$

Another important inequality results if we take $c = 0$. Then, since $\mu_0 = 1$,

$$\mu_b^a \leqq \mu_a^b$$

or

$$\mu_b^{\frac{1}{b}} \leqq \mu_a^{\frac{1}{a}}$$

if $a > b > 0$. This amounts to

$$\frac{\log \mu_b}{b} \leqq \frac{\log \mu_a}{a} \qquad \text{if} \qquad a > b$$

which is equivalent to the statement that

$$\frac{\log \mu_x}{x}$$

is an increasing function of x for positive x.

Composition of Distribution Functions

6. An important problem in the calculus of probability is to find the distribution function of the sum of several independent variables when distribution functions of these variables are known. It suffices to show how this problem can be solved for the sum of two independent variables.

Let x and y be two independent variables with the corresponding distribution functions $F(t)$ and $G(t)$. To find the distribution function $H(t)$ of their sum

$$z = x + y$$

is the same as to find the probability of the inequality

$$x + y < t$$

for an arbitrary real number t. Here, for the sake of simplicity and in view of the applications we propose to consider later, we shall assume that one, at least, of the variables x, y has continuous distribution with generally continuous density.

At first, let both x and y have continuous distributions so that

$$F(t) = \int_{-\infty}^{t} f(x)dx; \qquad G(t) = \int_{-\infty}^{t} g(x)dx.$$

The probability of the inequality

$$x + y < t$$

according to the general principles stated in Chap. XII is expressed by the double integral

$$H(t) = \int\int f(x)g(y)dxdy$$

extended over the domain

$$x + y < t.$$

Now, following ordinary rules, we can reduce this double integral to a repeated integral. To this end, for any fixed x we integrate $g(y)$ between limits $-\infty$ and $t - x$, thus obtaining

$$\int_{-\infty}^{t-x} g(y)dy = G(t - x).$$

Then, after multiplying by $f(x)$, we integrate the resulting expression between limits $-\infty$ and $+\infty$ for x. The final result will be

$$H(t) = \int_{-\infty}^{\infty} G(t - x)f(x)dx$$

or, written as Stieltjes' integral,

$$H(t) = \int_{-\infty}^{\infty} G(t - x)dF(x).$$

In the second place, let x be a discontinuous variable with different values $x_1, x_2, x_3, \ldots$ and corresponding probabilities $p_1, p_2, p_3, \ldots$. For $x = x_i$ the inequality

$$x + y < t$$

is equivalent to

$$y < t - x_i$$

and the probability of this inequality is $G(t - x_i)$. Since the probability of $x = x_i$ is p_i, the compound probability of the two events

$$x = x_i$$
$$x + y < t$$

will be

$$p_i G(t - x_i).$$

The total probability $H(t)$ of the inequality

$$x + y < t$$

will be expressed by the sum

$$H(t) = \Sigma p_i G(t - x_i)$$

extended over all possible values of x. But this sum can again be written as Stieltjes' integral:

$$(1) \qquad H(t) = \int_{-\infty}^{\infty} G(t - x)dF(x).$$

In both cases we obtain the same expression for $H(t)$. Evidently $H(t)$ can also be defined as the mathematical expectation of $G(t - x)$:

$$H(t) = E\{G(t - x)\}$$

taken with respect to the variable x. The important formula (1) is known as the formula for composition of distribution functions $F(t)$ and $G(t)$.

Example. Let x and y be two normally distributed variables with means $= 0$ and respective standard deviations σ_1 and σ_2. Instead of using (1), it is better to write $H(t)$ as a double integral

$$H(t) = \frac{1}{2\pi\sigma_1\sigma_2}\int\int e^{-\frac{x^2}{2\sigma_1^2}-\frac{y^2}{2\sigma_2^2}}dxdy$$

extended over the domain

$$x + y < t.$$

To evaluate this integral, it is natural to introduce $x + y = z$ as a new variable and find constants C, D, α, β so as to have identically

$$\frac{x^2}{2\sigma_1^2} + \frac{y^2}{2\sigma_2^2} = C(x + y)^2 + D(\alpha x + \beta y)^2,$$

whence one easily finds

$$C = \frac{1}{2(\sigma_1^2 + \sigma_2^2)}, \qquad D = \frac{1}{2\sigma_1^2\sigma_2^2(\sigma_1^2 + \sigma_2^2)}$$
$$\alpha = \sigma_2^2, \qquad \beta = -\sigma_1^2$$

and

$$\frac{x^2}{2\sigma_1^2} + \frac{y^2}{2\sigma_2^2} = \frac{1}{2(\sigma_1^2 + \sigma_2^2)}\left\{(x + y)^2 + \left(\frac{\sigma_2}{\sigma_1}x - \frac{\sigma_1}{\sigma_2}y\right)^2\right\}.$$

The Jacobian of

$$z = x + y, \qquad u = \frac{\sigma_2}{\sigma_1}x - \frac{\sigma_1}{\sigma_2}y$$

with respect to x, y being

$$\begin{vmatrix} 1 & 1 \\ \frac{\sigma_2}{\sigma_1} & -\frac{\sigma_1}{\sigma_2} \end{vmatrix} = -\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1\sigma_2}$$

$H(t)$ can be presented as the double integral

$$H(t) = \frac{1}{2\pi(\sigma_1^2 + \sigma_2^2)}\int\int e^{-\frac{z^2+u^2}{2(\sigma_1^2+\sigma_2^2)}}dzdu$$

with the domain of integration defined by a single inequality:

$$z < t.$$

Hence,

$$H(t) = \frac{1}{2\pi(\sigma_1^2 + \sigma_2^2)}\int_{-\infty}^{t} e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}}dz\int_{-\infty}^{\infty} e^{-\frac{u^2}{2(\sigma_1^2+\sigma_2^2)}}du$$

or

$$H(t) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}}\int_{-\infty}^{t} e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}}dz,$$

since

$$\int_{-\infty}^{\infty} e^{-\frac{u^2}{2(\sigma_1^2+\sigma_2^2)}}du = \sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}.$$

The expression obtained for $H(t)$ leads to a remarkable conclusion: The sum of two normally distributed variables with means $= 0$ and standard deviations σ_1 and σ_2 is also a normally distributed variable with the mean $= 0$ and the standard deviation $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. If the means of x and y are a_1 and a_2, then evidently z will be normally distributed with the mean $a = a_1 + a_2$ and the standard deviation $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$.

Repeated application of this result leads to the following important theorem:

If $x_1, x_2, \ldots x_n$ are normally distributed independent variables with means $a_1, a_2, \ldots a_n$ and standard deviations $\sigma_1, \sigma_2, \ldots \sigma_n$, then their sum

$$z = x_1 + x_2 + \cdots + x_n$$

is again normally distributed with the mean $a = a_1 + a_2 + \cdots + a_n$ and the standard deviation $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2}$.

Finally, any linear function

$$u = c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

is normally distributed with the mean $a = c_1a_1 + c_2a_2 + \cdots + c_na_n$

and the standard deviation $\sigma = \sqrt{c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \cdots + c_n^2\sigma_n^2}$. In particular, the arithmetic mean

$$\frac{x_1 + x_2 + \cdots + x_n}{n}$$

of identical normally distributed variables with the mean a and the standard deviation σ is normally distributed about the mean a and with the standard deviation $\sigma/\sqrt{n}$. Hence, the conclusion may be drawn that the probability P of the inequality

$$\left|\frac{x_1 + x_2 + \cdots + x_n}{n} - a\right| < \epsilon$$

is given by

$$P = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}}\int_{-\epsilon}^{\epsilon} e^{-\frac{nx^2}{2\sigma^2}}dx = \frac{2}{\sqrt{2\pi}}\int_0^{\frac{\epsilon\sqrt{n}}{\sigma}} e^{-\frac{t^2}{2}}dt$$

and rapidly approaches 1 as n increases. This is a more definite form of the law of large numbers applied to normally distributed (identical or equal) variables.

DETERMINATION OF DISTRIBUTION WHEN ITS CHARACTERISTIC FUNCTION IS GIVEN

7. One of the most important conclusions to be drawn from the preceding considerations is that the distribution function of probability is uniquely determined by the characteristic function. The known proofs of this fact are rather subtle, owing to the use of conditionally convergent integrals. However, such integrals can be avoided by resorting to an ingenious device due to Liapounoff. In the general case, the distribution function of a variable x has discontinuities. To avoid the bad effect of these discontinuities, Liapounoff introduces a continuous variable y that, with reasonable probability, can have values only in the vicinity of 0. It may be surmised, therefore, that the continuous distribution function of the sum $x + y$ will approximately represent that of x and, by disposing of a parameter involved in the distribution function of y, will tend to it as a limit. To make these explanations more definite, let y be a normally distributed variable whose distribution function is

$$G(t) = \frac{1}{h\sqrt{\pi}}\int_{-\infty}^{t} e^{-\frac{z^2}{h^2}}dz.$$

When h is small, the probabilities of any one of the inequalities

$$y > \epsilon, \qquad y < -\epsilon$$

will be extremely small and even will tend to 0 when h tends to 0. Hence, the distribution function $H(t)$ of the sum $x + y$ is likely to tend to $F(t)$ as a limit when h tends to 0.

To prove this in all rigor, we apply the composition formula (Sec. 6) to our case. We obtain the following expression for $H(t)$:

$$H(t) = \frac{1}{h\sqrt{\pi}}\int_{-\infty}^{\infty} dF(x) \int_{-\infty}^{t-x} e^{-\frac{z^2}{h^2}}dz$$

or, in more convenient form

$$H(t) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} dF(x) \int_{-\infty}^{\frac{t-x}{h}} e^{-u^2}du;$$

and furthermore, integrating by parts,

$$H(t) = \frac{1}{h\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-\left(\frac{t-x}{h}\right)^2} F(x)dx.$$

The integral in the right member can be split into three parts

$$\frac{1}{h\sqrt{\pi}}\int_{t-\epsilon}^{t+\epsilon} e^{-\left(\frac{t-x}{h}\right)^2} F(x)dx + \frac{1}{h\sqrt{\pi}}\int_{t+\epsilon}^{\infty} e^{-\left(\frac{t-x}{h}\right)^2} F(x)dx +$$
$$+ \frac{1}{h\sqrt{\pi}}\int_{-\infty}^{t-\epsilon} e^{-\left(\frac{t-x}{h}\right)^2} F(x)dx.$$

Now, for positive T

$$\frac{1}{\sqrt{\pi}}\int_{T}^{\infty} e^{-u^2}du < \frac{1}{2}e^{-T^2}.$$

Making use of this inequality, we find that

$$\frac{1}{h\sqrt{\pi}}\int_{t+\epsilon}^{\infty} e^{-\left(\frac{t-x}{h}\right)^2} F(x)dx < \frac{1}{h\sqrt{\pi}}\int_{t+\epsilon}^{\infty} e^{-\left(\frac{t-x}{h}\right)^2} dx = \frac{1}{\sqrt{\pi}}\int_{\frac{\epsilon}{h}}^{\infty} e^{-u^2}du < \frac{1}{2}e^{-\frac{\epsilon^2}{h^2}}$$

and similarly

$$\frac{1}{h\sqrt{\pi}}\int_{-\infty}^{t-\epsilon} e^{-\left(\frac{t-x}{h}\right)^2} F(x)dx < \frac{1}{2}e^{-\frac{\epsilon^2}{h^2}},$$

so that

$$H(t) = \frac{1}{h\sqrt{\pi}}\int_{0}^{\epsilon} e^{-\frac{u^2}{h^2}}F(t+u)du + \frac{1}{h\sqrt{\pi}}\int_{0}^{\epsilon} e^{-\frac{u^2}{h^2}}F(t-u)du + \theta e^{-\frac{\epsilon^2}{h^2}};$$
$$0 < \theta < 1.$$

Given an arbitrary $\sigma > 0$, the number ϵ can be taken so small that

$$0 \leqq F(t+u) - F(t+0) < \sigma$$
$$0 \leqq F(t-0) - F(t-u) < \sigma$$

for $0 < u < \epsilon$, whence

$$\left|\frac{1}{h\sqrt{\pi}}\int_0^{\epsilon} e^{-\frac{u^2}{h^2}}F(t+u)du - \frac{F(t+0)}{\sqrt{\pi}}\int_0^{\frac{\epsilon}{h}} e^{-u^2}du\right| < \sigma$$

$$\left|\frac{1}{h\sqrt{\pi}}\int_0^{\epsilon} e^{-\frac{u^2}{h^2}}F(t-u)du - \frac{F(t-0)}{\sqrt{\pi}}\int_0^{\frac{\epsilon}{h}} e^{-u^2}du\right| < \sigma$$

and

$$H(t) = \frac{F(t+0)+F(t-0)}{\sqrt{\pi}}\int_0^{\frac{\epsilon}{h}} e^{-u^2}du + \theta'\left(2\sigma + e^{-\frac{\epsilon^2}{h^2}}\right); \qquad |\theta'| < 1.$$

On the other hand,

$$\frac{1}{\sqrt{\pi}}\int_0^{\frac{\epsilon}{h}} e^{-u^2}du = \frac{1}{2} - \frac{1}{\sqrt{\pi}}\int_{\frac{\epsilon}{h}}^{\infty} e^{-u^2}du = \frac{1}{2} - \frac{\theta''}{2}e^{-\frac{\epsilon^2}{h^2}}; \qquad 0 < \theta'' < 1,$$

so that finally

$$\left|H(t) - \frac{F(t+0)+F(t-0)}{2}\right| < 2\sigma + 2e^{-\frac{\epsilon^2}{h^2}},$$

and for all sufficiently small h (ϵ being kept fixed)

$$\left|H(t) - \frac{F(t+0)+F(t-0)}{2}\right| < 4\sigma;$$

that is,

$$\lim_{h\to 0} H(t) = \frac{F(t+0)+F(t-0)}{2}$$

or, if t is a point of continuity,

$$\lim_{h\to 0} H(t) = F(t).$$

Now we must find another analytical representation for $H(t)$. To this end we consider the difference

$$H(t) - H(0) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} dF(x)\int_{-\frac{x}{h}}^{\frac{t-x}{h}} e^{-u^2}du,$$

and, to represent in a convenient way the inner integral, we make use of the known integral

$$\frac{1}{2\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{4}v^2}e^{-ivu}dv = e^{-u^2}.$$

Multiplying both sides by du and integrating between $-\frac{x}{h}$ and $\frac{t-x}{h}$ we find

$$\frac{1}{\sqrt{\pi}}\int_{-\frac{x}{h}}^{\frac{t-x}{h}} e^{-u^2}du = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{h^2v^2}{4}} e^{ivx}\frac{1-e^{-ivt}}{iv}dv$$

and

$$H(t) - H(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dF(x)\int_{-\infty}^{\infty} e^{-\frac{h^2v^2}{4}} e^{ivx}\frac{1-e^{-ivt}}{iv}dv.$$

The next step is to reverse the order of integrations, an operation which can be easily justified in this case. The result will be:

$$H(t) - H(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{h^2v^2}{4}}\frac{1-e^{-ivt}}{iv}dv\int_{-\infty}^{\infty} e^{ivx}dF(x)$$

or

$$H(t) - H(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{h^2v^2}{4}}\varphi(v)\frac{1-e^{-ivt}}{iv}dv$$

since

$$\varphi(v) = \int_{-\infty}^{\infty} e^{ivx}dF(x).$$

Now, taking the limit of $H(t)$ for h converging to 0, we have at any point of continuity of $F(t)$

$$F(t) = C + \frac{1}{2\pi}\lim_{h=0}\int_{-\infty}^{\infty} e^{-\frac{h^2v^2}{4}}\varphi(v)\frac{1-e^{-ivt}}{iv}dv \tag{2}$$

where the constant

$$C = \frac{F(+0)+F(-0)}{2}$$

is determined by the condition $F(-\infty) = 0$. Thus, the distribution function is completely determined by (2) at all points of continuity when the characteristic function $\varphi(v)$ is given.

Example 1. Let us apply (2) to find the distribution corresponding to the characteristic function

$$\varphi(v) = e^{-\frac{\sigma^2v^2}{2}}.$$

Since in this case the integral whose limit we seek is uniformly convergent with respect to h, we find simply

$$\begin{aligned} F(t) &= C + \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{\sigma^2v^2}{2}}\frac{1-e^{-ivt}}{iv}dv \\ &= C + \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{\sigma^2v^2}{2}}\frac{\sin tv}{v}dv. \end{aligned}$$

On the other hand (Chap. VII, page 128),

$$\int_{-\infty}^{\infty} e^{-\frac{\sigma^2 v^2}{2}} \frac{\sin tv}{v} dv = \frac{\sqrt{2\pi}}{\sigma} \int_0^t e^{-\frac{u^2}{2\sigma^2}} du,$$

so that

$$F(t) = C - \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{u^2}{2\sigma^2}} du + \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2\sigma^2}} du.$$

Taking $t = -\infty$, the condition $F(-\infty) = 0$ gives

$$C = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{u^2}{2\sigma^2}} du,$$

and so finally

$$F(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2\sigma^2}} du.$$

Naturally, we find a normal distribution with the standard deviation σ (compare page 270).

Example 2. What is the distribution determined by the characteristic function

$$\varphi(v) = e^{-a|v|}, \qquad a > 0?$$

As in the preceding example we find that

$$F(t) = C + \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-a|v|} \frac{\sin tv}{v} dv = C + \frac{1}{\pi} \int_0^{\infty} e^{-av} \frac{\sin tv}{v} dv.$$

But

$$\frac{d}{dt} \int_0^{\infty} e^{-av} \frac{\sin tv}{v} dv = \int_0^{\infty} e^{-av} \cos tv dv = \frac{a}{a^2 + t^2},$$

whence

$$\frac{1}{\pi} \int_0^{\infty} e^{-av} \frac{\sin tv}{v} dv = \frac{a}{\pi} \int_0^t \frac{dx}{a^2 + x^2} = \frac{a}{\pi} \int_{-\infty}^{t} \frac{dx}{a^2 + x^2} - \frac{1}{2}.$$

Thus

$$F(t) = C - \frac{1}{2} + \frac{a}{\pi} \int_{-\infty}^{t} \frac{dx}{a^2 + x^2}$$

and the condition $F(-\infty) = 0$ gives $C = \frac{1}{2}$, so that finally

$$F(t) = \frac{a}{\pi} \int_{-\infty}^{t} \frac{dx}{a^2 + x^2}.$$

Naturally we find the same distribution as that considered in Example 2, page 243. Sometimes it is called "Cauchy's distribution" with the parameter a.

Composition of Characteristic Functions

8. Having n independent variables $x_1, x_2, \ldots x_n$ whose characteristic functions are $\varphi_1(t), \varphi_2(t), \ldots \varphi_n(t)$, the product

$$\varphi(t) = \varphi_1(t)\varphi_2(t) \cdots \varphi_n(t)$$

is the characteristic function of their sum

$$s = x_1 + x_2 + \cdots + x_n.$$

In fact, the characteristic function of s is by definition

$$\varphi(t) = E(e^{ist}) = E(e^{ix_1t} \cdot e^{ix_2t} \cdots e^{ix_nt}).$$

Since $x_1, x_2, \ldots x_n$ are independent variables, the expectation of the product

$$e^{ix_1t} \cdot e^{ix_2t} \cdots e^{ix_nt}$$

is equal to the product of the expectations of the factors, whence

$$\varphi(t) = \varphi_1(t)\varphi_2(t) \cdots \varphi_n(t).$$

This simple theorem is of great importance since it determines the characteristic function of the sum of independent variables and indirectly its function of distribution.

9. A few examples will illustrate the preceding remark.

Example 1. Consider n independent normally distributed variables $x_1, x_2, \ldots x_n$ with means $= 0$ and standard deviations $\sigma_1, \sigma_2, \ldots \sigma_n$. Their characteristic functions are

$$\varphi_k(t) = e^{-\frac{\sigma_k^2t^2}{2}}; \qquad k = 1, 2, \ldots n$$

and the characteristic function of their sum

$$s = x_1 + x_2 + \cdots + x_n$$

will be

$$\varphi(t) = e^{-\frac{\sigma^2t^2}{2}}$$

where

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2.$$

Hence s is a normally distributed variable with the mean 0 and the standard deviation

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2}$$

as we found previously by a method involving a considerable amount of calculation.

Example 2. Independent variables $x_1, x_2, \ldots x_n$ have Cauchy's distributions with parameters $a_1, a_2, \ldots a_n$. Since the characteristic function of x_k is

$$e^{-a_k|t|},$$

the characteristic function of the sum

$$s = x_1 + x_2 + \cdots + x_n$$

will be

$$\varphi(t) = e^{-a|t|}$$

where

$$a = a_1 + a_2 + \cdots + a_n.$$

Hence, s again has Cauchy's distribution with the parameter $a_1 + a_2 + \cdots + a_n$.

Example 3. Let $x_1, x_2, \ldots x_n$ be independent variables with uniform distribution of probability in the interval $(0, l)$. The characteristic function of any one of them is

$$\frac{1}{l}\int_0^l e^{itx}dx = \frac{e^{ilt}-1}{ilt}.$$

Hence, the characteristic function of their sum s will be

$$\varphi(t) = \left(\frac{e^{ilt}-1}{ilt}\right)^n.$$

The distribution function of s is given by

$$F(t) = C + \frac{1}{2\pi}\lim_{h=0}\int_{-\infty}^{\infty} e^{-\frac{h^2v^2}{4}}\left(\frac{e^{ilv}-1}{ilv}\right)^n\frac{1-e^{-ivt}}{iv}dv$$

and, since the integral again is uniformly convergent,

$$F(t) = C + \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\frac{e^{ilv}-1}{ilv}\right)^n\frac{1-e^{-ivt}}{iv}dv.$$

The evaluation of this integral presents certain difficulties. To avoid them we notice that the integrand considered as a function of a complex variable v is holomorphic everywhere. Hence, we can substitute for the rectilinear path of integration the path Γ as shown in Fig. 20.

FIG. 20.

Now it is easy to show that integrating over the path Γ we have

$$f(g) = \int_\Gamma \frac{e^{igz}}{z^{n+1}}dz = \begin{cases} 0 & \text{if} \quad g > 0 \\ -2\pi i^{n+1}\frac{g^n}{n!} & \text{if} \quad g \leqq 0 \end{cases}$$

The integral

$$\int_\Gamma\left(\frac{e^{ilz}-1}{ilz}\right)^n\frac{dz}{iz}$$

being a linear combination of integrals of the type $f(g)$ with $g \geqq 0$ reduces to 0. Similarly,

$$J = (-1)^{n+1}\int_\Gamma\left(\frac{1-e^{ilz}}{ilz}\right)^n\frac{e^{-izt}}{iz}dz = i^{n+1}l^{-n}\sum_0^n(-1)^kC_n^kf(kl-t)$$

or, in explicit form,

$$J = \frac{2\pi}{n!}\sum_{k\leqq\frac{t}{l}}(-1)^kC_n^k\left(\frac{t}{l}-k\right)^n.$$

Referring to the above expression of $F(t)$, we find that

$$F(t) = C + \frac{1}{n!}\sum_{k\leqq\frac{t}{l}}(-1)^kC_n^k\left(\frac{t}{l}-k\right)^n.$$

The constant $C = 0$ since $F(t)$ and the sum in the right member both vanish for $t = 0$. The final expression of $F(t)$ is, therefore:

$$F(t) = \frac{1}{1 \cdot 2 \cdot 3 \cdots n}\left[\left(\frac{t}{l}\right)^n - \frac{n}{1}\left(\frac{t}{l} - 1\right)^n + \frac{n(n-1)}{1 \cdot 2}\left(\frac{t}{l} - 2\right)^n - \cdots \right].$$

The series in the right member is continued as long as arguments remain positive. Such is the probability that the sum

$$x_1 + x_2 + \cdots + x_n$$

of n independent variables, uniformly distributed throughout the interval $(0, l)$, will be less than t. The above expression is due to Laplace, who, however, obtained it in quite a different manner.

Problems for Solution

1. Prove directly the inequality

$$\mu^2_{\frac{a+c}{2}} \leqq \mu_a\mu_c$$

for absolute moments.

HINT: The quadratic form in λ, μ

$$\int_{-\infty}^{\infty} (\lambda|x|^{\frac{a}{2}} + \mu|x|^{\frac{c}{2}})^2 d\varphi(x)$$

is definite or semidefinite. Show that the equality sign cannot hold if $\varphi(x)$ has at least two points of increase α, β such that $\alpha:\beta$ is neither 0 nor ± 1.

2. Let $x_1, x_2, \ldots x_n$ be n variables. Denoting the absolute moment of the order α for x_i by $\mu_\alpha^{(i)}$, and by ω_δ the quotient

$$\omega_\delta = \frac{\mu_{2+\delta}^{(1)} + \mu_{2+\delta}^{(2)} + \cdots + \mu_{2+\delta}^{(n)}}{(\mu_2^{(1)} + \mu_2^{(2)} + \cdots + \mu_2^{(n)})^{1+\frac{\delta}{2}}}$$

prove that

$$\omega_\delta^{\frac{1}{\delta}} \leqq \omega_{\delta'}^{\frac{1}{\delta'}}$$

if $\delta' > \delta > 0$.

HINT: Use Liapounoff's inequality.

3. A variable is distributed over the interval $(0, +\infty)$ with a decreasing density of probability. Show that in this case moments M_2 and M_4 satisfy the inequality

$$M_2^2 \leqq \tfrac{5}{9}M_4 \text{ (Gauss)}$$

and that in general

$$[(\mu + 1)M_\mu]^{\frac{1}{\mu}} \leqq [(\nu + 1)M_\nu]^{\frac{1}{\nu}}$$

if $\nu > \mu > 0$.

Indication of the Proof. Show first that the existence of the integral

$$\int_0^\infty x^s f(x)dx$$

in case $f(x)$ is a positive and decreasing function implies the existence of the limit

$$\lim a^{s+1}f(a) = 0; \qquad a \to +\infty.$$

Hence, deduce that

$$\int_0^\infty x\,d\varphi(x) = 1, \qquad \int_0^\infty x^{\mu+1}d\varphi(x) = (\mu+1)M_\mu, \qquad \int_0^\infty x^{\nu+1}d\varphi(x) = (\nu+1)M_\nu$$

where $\varphi(x) = f(0) - f(x)$ and, finally, apply the inequality

$$\left[\int_0^\infty x^{\mu+1}d\varphi(x)\right]^\nu \leqq \left[\int_0^\infty x^{\nu+1}d\varphi(x)\right]^\mu \cdot \left[\int_0^\infty x\,d\varphi(x)\right]^{\nu-\mu}.$$

4. Using the composition formula (1), page 269, prove Laplace's formula on page 278 by mathematical induction.

5. Prove that the distribution function of probability for a variable whose characteristic function $\varphi(t)$ is given can be determined by the formula

$$F(t) = C + \lim_{h=0}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\varphi(v)}{1+h^2v^2}\,\frac{1-e^{-ivt}}{iv}\,dv.$$

HINT: In carrying out Liapounoff's idea, take an auxiliary variable with the distribution

$$G(y) = \frac{1}{2h}\int_{-\infty}^{y} e^{-\left|\frac{z}{h}\right|}dz.$$

Also make use of the integral

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{e^{-ibx}dx}{1+x^2} = e^{-|b|}.$$

Many definite integrals can be evaluated using the relation between characteristic and distribution functions, as the following example shows.

6. Let x be distributed over $(-\infty, +\infty)$ with the density $\frac{1}{2}e^{-|x|}$. The characteristic function being in this case

$$\varphi(t) = \frac{1}{1+t^2}$$

we find

$$F(t) = C + \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1-e^{-ivt}}{iv(1+v^2)}dv = \frac{1}{2}\int_{-\infty}^{t} e^{-|x|}dx,$$

whence

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{e^{-ivt}}{1+v^2}dv = e^{-|t|},$$

an integral due to Laplace.

7. A variable is said to have Poisson's distribution if it can have only integral values 0, 1, 2, . . . and the probability of $x = k$ is

$$\frac{a^k e^{-a}}{k!};$$

the quantity a is called "parameter" of distribution. If n variables have Poisson's distribution with parameters $a_1, a_2, \ldots a_n$, show that their sum has also Poisson's distribution, the parameter of which is $a_1 + a_2 + \cdots + a_n$.

8. Prove the following result:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\frac{\sin v}{v}\right)^n\frac{\sin tv}{v}dv = -\frac{1}{2} + \frac{1}{2\cdot 4\cdot 6\cdots 2n}\left[(t+n)^n - \frac{n}{1}(t+n-2)^n + \frac{n(n-1)}{1\cdot 2}(t+n-4)^n - \cdots\right]$$

the series being continued as long as arguments remain positive.

HINT: Consider the sum of n uniformly distributed variables in the interval $(-1, +1)$ and express its distribution function in two different ways.

9. Establish the expression for the mathematical expectation of the absolute value of the sum of n uniformly distributed variables in the interval $(-\frac{1}{2}, +\frac{1}{2})$. *Ans.*

$$E|x_1 + x_2 + \cdots + x_n| = \frac{2}{2\cdot 4\cdot 6\cdots(2n+2)}\left[n^{n+1} - \frac{n}{1}(n-2)^{n+1} + \frac{n(n-1)}{1\cdot 2}(n-4)^{n+1} - \cdots\right],$$

the series being continued as long as the arguments remain positive.

HINT: Apply Laplace's formula on page 278, conveniently modified, to express the expectation of $x_1 + x_2 + \cdots + x_n$ and that of $|x_1 + x_2 + \cdots + x_n|$.

10. Show that under the same conditions as in Prob. 9

$$E|x_1 + x_2 + \cdots + x_n| = \frac{n}{2\pi}\int_{-\infty}^{\infty}\left(\frac{\sin t}{t}\right)^{n-1}\frac{\sin t - t\cos t}{t^3}dt.$$

HINT: Prove and use the following formula

$$\lim_{T=\infty}\int_{-T}^{T}\frac{e^{iwx} - 1 - iwx}{x^2}dx = -\pi|w|.$$

11. Let x_1 and x_2 be two identical and normally distributed variables with the mean $= 0$ and the standard deviation σ. If x is defined as the greater of the values $|x_1|$, $|x_2|$, that is,

$$x = \max.\ (|x_1|, |x_2|)$$

find the mean value of x as well as that of x^2. *Ans.*

$$E(x) = \frac{2\sigma}{\sqrt{\pi}}, \qquad E(x^2) = \left(1 + \frac{2}{\pi}\right)\sigma^2.$$

12. Let

$$x = \min.\ (|x_1|, |x_2|, \ldots |x_n|)$$

where $x_1, x_2, \ldots x_n$ are identical normally distributed variables with the mean $= 0$ and the standard deviation σ. Find the mean value of x. *Ans.* Setting for brevity

$$\frac{\sqrt{2}}{\sigma\sqrt{\pi}}\int_0^t e^{-\frac{u^2}{2\sigma^2}}du = \theta(t),$$

we have

$$E(x) = \int_0^{\infty}\{1 - \theta(t)\}^n dt.$$

In particular for $n = 2$

$$E(x) = \frac{2\sigma}{\sqrt{\pi}}(\sqrt{2} - 1).$$

For large n asymptotically

$$E(x) \sim \frac{\sigma\sqrt{\pi/2}}{n + 1}.$$

13. A variable with the mean $= 0$ and the standard deviation $= 1$ is called a "reduced variable." By changing the origin and the unit of measurement any variable can be made reduced. For, if x has the mean a and the standard deviation σ the variable

$$u = \frac{x - a}{\sigma}$$

is reduced. The distribution function of the reduced variable u can be called the "reduced law of distribution."

As we have seen, variables x_1 and x_2 with normal distribution have the same reduced law of distribution, as does their sum. The question may be raised: Is the normal law of distribution a unique law possessing this property? (G. Pólya.)

Solution. Let x_1, x_2 be two variables for which the second moment of the distribution exists, so that we can speak of their means and standard deviations. Let x_1 have its mean a_1 and its standard deviation σ_1; likewise, let a_2 and σ_2 be the mean and the standard deviation of x_2. Three reduced variables

$$u_1 = \frac{x_1 - a_1}{\sigma_1}, \qquad u_2 = \frac{x_2 - a_2}{\sigma_2}, \qquad u_3 = \frac{x_1 + x_2 - a_1 - a_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$$

have by hypothesis the same law of distribution. Hence, they have the same characteristic function $\varphi(t)$ whence we can draw the conclusion that the characteristic functions of x_1, x_2, $x_1 + x_2$ are, respectively,

$$\varphi_1(t) = e^{ita_1}\varphi(\sigma_1 t); \qquad \varphi_2(t) = e^{ita_2}\varphi(\sigma_2 t); \qquad \varphi_3(t) = e^{it(a_1+a_2)}\varphi(\sqrt{\sigma_1^2 + \sigma_2^2}t).$$

Since

$$\varphi_3(t) = \varphi_1(t)\varphi_2(t),$$

we must have for an arbitrary real t

$$\varphi(\sigma_1 t)\varphi(\sigma_2 t) = \varphi(\sqrt{\sigma_1^2 + \sigma_2^2}t),$$

or

(1) $$\varphi(\alpha t)\varphi(\beta t) = \varphi(t)$$

where

$$\alpha = \frac{\sigma_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}, \qquad \beta = \frac{\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}; \qquad \alpha^2 + \beta^2 = 1.$$

Since (1) holds for every real t, we shall have

$$\varphi(\alpha t) = \varphi(\alpha^2 t)\varphi(\alpha\beta t); \qquad \varphi(\beta t) = \varphi(\alpha\beta t)\varphi(\beta^2 t)$$

and

(2) $$\varphi(t) = \varphi(\alpha^2 t)\varphi(\alpha\beta t)^2\varphi(\beta^2 t).$$

Applying (1) again to each of these factors in the right member of (2), we find that

(3) $$\varphi(t) = \varphi(\alpha^3 t)\varphi(\alpha^2\beta t)^3\varphi(\alpha\beta^2 t)^3\varphi(\beta^3 t)$$

and proceeding in the same way, we arrive at the general formula

$$\varphi(t) = \varphi(\alpha^n t)^{p_0}\varphi(\alpha^{n-1}\beta t)^{p_1} \cdots \varphi(\beta^n t)^{p_n} \tag{4}$$

where $p_0, p_1, \ldots p_n$ are coefficients in the expansion

$$(1 + z)^n = p_0 + p_1 z + \cdots + p_n z^n.$$

The arguments

$$v_0 = \alpha^n t, \qquad v_1 = \alpha^{n-1}\beta t, \ldots v_n = \beta^n t$$

tend uniformly to 0 since $\alpha < 1$, $\beta < 1$. The quotient

$$\frac{\varphi(v) - 1}{v^2} = -\int_{-\infty}^{\infty} t^2 dF(t) \int_0^1 (1 - x)e^{ivtx}dx$$

is represented by a *uniformly* convergent integral; hence

$$\lim \frac{\varphi(v) - 1}{v^2} = -\frac{1}{2}\int_{-\infty}^{\infty} t^2 dF(t) = -\frac{1}{2}$$

or

$$\varphi(v) = 1 + [-\tfrac{1}{2} + \epsilon(v)]v^2$$

where

$$\epsilon(v) \to 0 \qquad \text{as} \qquad v \to 0.$$

At the same time

$$\log \varphi(v) = [-\tfrac{1}{2} + \delta(v)]v^2 \text{ (principal branch of log)}$$

where again

$$\delta(v) \to 0 \qquad \text{as} \qquad v \to 0.$$

Now, taking logarithms of both members of (4)

$$\log \varphi(t) = -\tfrac{1}{2}t^2(p_0\alpha^{2n} + p_1\alpha^{2n-2}\beta^2 + \cdots + p_n\beta^{2n}) + \Omega = -\tfrac{1}{2}t^2 + \Omega$$

where

$$\Omega = t^2[p_0\delta(v_0)\alpha^{2n} + p_1\delta(v_1)\alpha^{2n-2}\beta^2 + \cdots + p_n\delta(v_n)\beta^{2n}].$$

Given $\epsilon > 0$, we can take n so large that

$$|\delta(v_i)| < \epsilon; \qquad i = 0, 1, \ldots n$$

whence

$$|\Omega| < \epsilon t^2.$$

Thus

$$|\log \varphi(t) + \tfrac{1}{2}t^2| < \epsilon t^2$$

and since ϵ can be taken arbitrarily small,

$$\log \varphi(t) + \tfrac{1}{2}t^2 = 0$$

or

$$\varphi(t) = e^{-\frac{1}{2}t^2},$$

which shows that the normal law is the only one with the required properties, among all laws with finite second moments.

References

P. LAPLACE: Théorie analytique des probabilités, Oeuvres VII, pp. 260–261.

T. J. STIELTJES: Recherches sur les fractions continues, Oeuvres II, pp. 402–566.

A. LIAPOUNOFF: Nouvelle forme du théorème sur la limite de probabilité, *Mém. Acad. Sci. St. Petersbourg*, 12, 1901.

P. LEVY: "Calcul des probabilités" Paris, 1925.

CHAPTER XIV

FUNDAMENTAL LIMIT THEOREMS

1. Bernoulli's theorem, as we have seen in Chap. VII, follows from a more general one known as Laplace's limit theorem. In terms already familiar to us, this theorem can be stated as follows: Let an event E occur m times in a series of n independent trials with constant probability p. As n becomes infinite, the distribution function of the quotient

$$\frac{m - np}{\sqrt{npq}}$$

approaches

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}u^2}du$$

as a limit; or, to state it in a less precise form, the distribution of the above quotient tends to normal.

Just as Bernoulli's theorem itself is a very particular case of the general law of large numbers, so Laplace's limit theorem is a special case of another extremely general theorem, the discovery of which by Laplace may be considered as the crowning achievement of his persistent efforts, extending over a period of more than twenty years, to find the approximate distribution of probability for sums consisting of a great many independent components with almost arbitrary distributions. The result at which Laplace finally arrived is as astonishing as it is simple: if $x_1, x_2, \ldots x_n$ ($E(x_i) = 0, i = 1, 2, \ldots n$) are independent variables (subject to some very mild limitations not stated, however, by Laplace) and B_n is the dispersion of their sum, then for large n the distribution of the quotient

$$\frac{x_1 + x_2 + \cdots + x_n}{\sqrt{B_n}}$$

is nearly normal. To put it more precisely, the distribution function of this quotient tends to the limit

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}u^2}du$$

as n becomes infinite.

Laplace's attempt to prove this important proposition does not stand the test of modern rigor and, besides, cannot easily be made rigorous.

The same is true of the attempts made by later investigators, notably Poisson, Cauchy, and many others. Only after a lapse of many years were truly rigorous proofs of Laplace's theorem given. This important achievement is the result of the work of three great Russian mathematicians: Tshebysheff (1887), Markoff (1898), and Liapounoff (1900–1901). An account of Tshebysheff's and Markoff's ingenious investigations is given in Appendix II. Here we shall follow Liapounoff; for his method of proof has the advantage of simplicity even compared with more recent proofs, of which that given by J. W. Lindeberg deserves special mention.[1]

2. Before going into details of analysis, we shall state the limit theorem in a very general form due to Liapounoff.

Laplace-Liapounoff's Theorem. *Let $x_1, x_2, \ldots x_n$ be independent variables with their means = 0, possessing absolute moments of the order $2+\delta$ (where δ is some number > 0):*

$$\mu_{2+\delta}^{(1)}, \mu_{2+\delta}^{(2)}, \ldots \mu_{2+\delta}^{(n)}.$$

If, denoting by B_n the dispersion of the sum $x_1 + x_2 + \cdots + x_n$, the quotient

$$\omega_n = \frac{\mu_{2+\delta}^{(1)} + \mu_{2+\delta}^{(2)} + \cdots + \mu_{2+\delta}^{(n)}}{B_n^{1+\frac{\delta}{2}}}$$

tends to 0 as $n \to \infty$, the probability of the inequality

$$\frac{x_1 + x_2 + \cdots + x_n}{\sqrt{B_n}} < t$$

tends uniformly to the limit

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}u^2}du.$$

It is natural that the complete proof of a theorem of such character cannot be too short, and to make the proof clearer it is advisable to divide it into logically separated parts.

3. The Fundamental Lemma. *Let s_n be a variable, depending on an integer n, with the mean = 0 and the standard deviation = 1. If its characteristic function*

$$\varphi_n(v) = E(e^{ivs_n})$$

tends to

$$e^{-\frac{v^2}{2}}$$

[1] Lindeberg's proof, as well as later proofs by P. Levy and others, make use of an ingenious artifice due to Liapounoff. Lindeberg explicitly acknowledges his indebtedness to Liapounoff, while Levy and other French writers fail to give due credit to the great Russian mathematician.

uniformly in any given finite interval $(-l, l)$, *then the distribution function* $F_n(t)$ *of* s_n *tends uniformly (in the domain of all real values of* t*) to the limit*

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}u^2}du.$$

Proof. *a.* Together with the variable s_n, whose distribution function is $F_n(t)$, Liapounoff considers another variable

$$\tau_n = s_n + y$$

where y is a normally distributed variable with the distribution function

$$G(y) = \frac{1}{h\sqrt{\pi}}\int_{-\infty}^{y} e^{-\frac{x^2}{h^2}}dx.$$

Denoting the distribution function of τ_n by $H_n(t)$, we have (Chap. XIII, Sec. 7)

$$H_n(t) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} dF_n(x)\int_{-\infty}^{\frac{t-x}{h}} e^{-u^2}du. \tag{1}$$

On account of the inequality

$$\frac{1}{\sqrt{\pi}}\int_{T}^{\infty} e^{-u^2}du \leqq \frac{1}{2}e^{-T^2}; \qquad T \geqq 0$$

we have:

For $t - x < 0$: $\displaystyle\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\frac{t-x}{h}} e^{-u^2}du = \frac{\theta'}{2}e^{-\left(\frac{t-x}{h}\right)^2}; \qquad 0 < \theta' \leqq 1.$

For $t - x \geqq 0$: $\displaystyle\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\frac{t-x}{h}} e^{-u^2}du = 1 - \frac{1}{\sqrt{\pi}}\int_{\frac{t-x}{h}}^{\infty} e^{-u^2}du = 1 - \frac{\theta''}{2}e^{-\left(\frac{t-x}{h}\right)^2};$

$$0 < \theta'' \leqq 1.$$

Hence, introducing these expressions into (1),

$$H_n(t) = \int_{-\infty}^{t} dF_n(x) + \frac{\theta_0}{2}\int_{t}^{\infty} e^{-\left(\frac{t-x}{h}\right)^2}dF_n(x) - \frac{\theta_1}{2}\int_{-\infty}^{t} e^{-\left(\frac{t-x}{h}\right)^2}dF_n(x)$$

where again $0 < \theta_0 < 1$; $0 < \theta_1 < 1$. This leads to the following inequality:

$$|H_n(t) - F_n(t)| < \tfrac{1}{2}\int_{-\infty}^{\infty} e^{-\left(\frac{t-x}{h}\right)^2}dF_n(x).$$

But

$$e^{-\left(\frac{t-x}{h}\right)^2} = \frac{h}{2\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-\frac{h^2v^2}{4}+iv(x-t)}dv$$

and consequently

$$|H_n(t) - F_n(t)| < \frac{h}{4\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-\frac{h^2v^2}{4}} e^{-ivt}\varphi_n(v)dv$$

or

$$(2)\quad |H_n(t) - F_n(t)| < \frac{h}{4\sqrt{\pi}}\left\{\int_{-\infty}^{\infty} e^{-\frac{h^2v^2}{4}-ivt}[\varphi_n(v) - e^{-\frac{v^2}{2}}]dv + \right.$$
$$\left. + \int_{-\infty}^{\infty} e^{-\frac{h^2v^2}{4}-\frac{v^2}{2}}e^{-ivt}dv\right\}.$$

Here we split the first integral into three J_1, J_2, J_3, taken respectively between limits $-\infty, -l$; $-l, l$; $l, +\infty$ and denote the second integral by J_4. Since $|\varphi_n(v) - e^{-\frac{v^2}{2}}| \leqq 2$, we shall have

$$(3)\qquad \frac{h}{4\sqrt{\pi}}|J_1 + J_3| < \frac{h}{\sqrt{\pi}}\int_l^{\infty} e^{-\frac{h^2v^2}{4}}dv < \frac{2}{\sqrt{\pi}}\frac{e^{-\frac{(hl)^2}{4}}}{hl}$$

because

$$\int_x^{\infty} e^{-u^2}du < \frac{e^{-x^2}}{2x}$$

for positive x. Also

$$(4)\qquad \frac{h}{4\sqrt{\pi}}|J_4| < \frac{h}{4\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-\frac{v^2}{2}}dv = \frac{h}{\sqrt{8}}.$$

To estimate J_2 we shall denote by $\epsilon_n(l)$ the maximum of $|\varphi_n(v) - e^{-\frac{v^2}{2}}|$ in the interval $-l \leqq v \leqq l$. Then

$$(5)\qquad \frac{h}{4\sqrt{\pi}}|J_2| < \frac{h\epsilon_n(l)}{4\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-\frac{h^2v^2}{4}}dv = \frac{1}{2}\epsilon_n(l).$$

Finally, taking into account (2), (3), (4), and (5), we find

$$(6)\qquad |H_n(t) - F_n(t)| < \frac{1}{2}\epsilon_n(l) + \frac{h}{\sqrt{8}} + \frac{2}{\sqrt{\pi}}\frac{e^{-\frac{(hl)^2}{4}}}{hl}.$$

b. Expression (1) of $H_n(t)$ can be transformed in a manner similar to that employed in Chap. XIII, Sec. 7, if we first write

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\frac{t-x}{h}} e^{-u^2}du = \frac{1}{2} + \frac{1}{\sqrt{\pi}}\int_0^{\frac{t-x}{h}} e^{-u^2}du.$$

Thus we get

$$H_n(t) = \frac{1}{2} + \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{h^2v^2}{4}} \frac{1 - e^{-itv}\varphi_n(v)}{iv}dv$$

or

$$H_n(t) = \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} e^{-\frac{h^2v^2}{4}-\frac{v^2}{2}} \frac{\sin tv}{v}dv + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-\frac{h^2v^2}{4}-itv}}{iv}(e^{-\frac{v^2}{2}} - \varphi_n(v))dv.$$

Now

$$\left|\frac{1}{\pi}\int_0^{\infty} e^{-\frac{v^2}{2}}\frac{\sin tv}{v}dv - \frac{1}{\pi}\int_0^{\infty} e^{-\frac{h^2v^2}{4}-\frac{v^2}{2}}\frac{\sin tv}{v}dv\right| < \frac{h^2}{4\pi}\int_0^{\infty} ve^{-\frac{v^2}{2}}dv = \frac{h^2}{4\pi}$$

since

$$0 < 1 - e^{-\frac{h^2v^2}{4}} < \frac{h^2v^2}{4}$$

and consequently

$$\left|H_n(t) - \frac{1}{2} - \frac{1}{\pi}\int_0^{\infty} e^{-\frac{v^2}{2}}\frac{\sin tv}{v}dv\right| < \frac{h^2}{4\pi} + \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{h^2v^2}{4}}\frac{|\varphi_n(v) - e^{-\frac{v^2}{2}}|}{|v|}dv. \tag{7}$$

To find an upper bound of the integral in the right member, we split it into five integrals I_1, I_2, I_3, I_4, I_5 taken respectively between limits $-\infty$, $-l$; $-l$, $-\lambda$; $-\lambda$, λ; λ, l; l, $+\infty$. To estimate I_3, we notice that

$$|\varphi_n(v) - 1| \leqq \frac{v^2}{2}\int_{-\infty}^{\infty} x^2dF_n(x) = \frac{v^2}{2}$$

$$\left|e^{-\frac{v^2}{2}} - 1\right| \leqq \frac{v^2}{2}$$

and

$$|\varphi_n(v) - e^{-\frac{v^2}{2}}| \leqq v^2.$$

Hence

$$\frac{1}{2\pi}|I_3| \leqq \frac{1}{\pi}\int_0^{\lambda} vdv = \frac{\lambda^2}{2\pi}. \tag{8}$$

To estimate $I_2 + I_4$, we use the inequality $|\varphi_n(v) - e^{-\frac{v^2}{2}}| \leqq \epsilon_n(l)$ and we get

$$\frac{1}{2\pi}|I_2 + I_4| \leqq \frac{\epsilon_n(l)}{\pi}\int_{\lambda}^{l} e^{-\frac{h^2v^2}{4}}\frac{dv}{v} < \frac{\epsilon_n(l)}{\sqrt{\pi}h\lambda}. \tag{9}$$

Finally, dealing with I_1 and I_5, we use the obvious inequality

$$|\varphi_n(v) - e^{-\frac{v^2}{2}}| \leqq 2$$

and we obtain

$$(10) \qquad \frac{1}{2\pi}|I_1 + I_5| \leqq \frac{2}{\pi}\int_l^\infty e^{-\frac{h^2v^2}{4}}\frac{dv}{v} < \frac{4}{\pi}\frac{e^{-\frac{(hl)^2}{4}}}{(hl)^2}.$$

Taking into account (7), (8), (9), and (10), the following inequality results:

$$\left|H_n(t) - \frac{1}{2} - \frac{1}{\pi}\int_0^\infty e^{-\frac{v^2}{2}}\frac{\sin tv}{v}dv\right| < \frac{h^2}{4\pi} + \frac{\lambda^2}{2\pi} + \frac{\epsilon_n(l)}{\sqrt{\pi}h\lambda} + \frac{4}{\pi}\frac{e^{-\frac{(hl)^2}{4}}}{(hl)^2}.$$

In it, since λ is still at our disposal, we can take

$$\lambda = \epsilon_n(l)^{\frac{1}{3}}h^{-\frac{1}{3}}.$$

The inequality thus obtained when combined with (6) gives ($\alpha = hl$)

$$(11) \quad \left|F_n(t) - \frac{1}{2} - \frac{1}{\pi}\int_0^\infty e^{-\frac{v^2}{2}}\frac{\sin tv}{v}dv\right| < \frac{4}{\pi}\frac{e^{-\frac{\alpha^2}{4}}}{\alpha^2} + \frac{2}{\sqrt{\pi}}\frac{e^{-\frac{\alpha^2}{4}}}{\alpha} + \frac{\alpha}{\sqrt{8}l} +$$
$$+ \frac{\alpha^2}{4\pi l^2} + \left(\frac{1}{2\pi} + \frac{1}{\sqrt{\pi}}\right)(l\alpha^{-1})^{\frac{2}{3}}\epsilon_n(l)^{\frac{2}{3}} + \frac{1}{2}\epsilon_n(l).$$

Here α and l are arbitrary positive numbers. We dispose of them in the following manner: Given an arbitrary positive number ϵ, we take α so large as to have

$$\frac{4}{\pi}\frac{e^{-\frac{\alpha^2}{4}}}{\alpha^2} + \frac{2}{\sqrt{\pi}}\frac{e^{-\frac{\alpha^2}{4}}}{\alpha} < \frac{1}{3}\epsilon$$

and after that we select l large enough to make

$$\frac{\alpha}{\sqrt{8}l} + \frac{\alpha^2}{4\pi l^2} < \frac{1}{3}\epsilon.$$

Finally, since for a fixed l, $\epsilon_n(l)$ by hypothesis, tends to 0 when $n \to \infty$, there exists a number n_0 such that

$$\left(\frac{1}{2\pi} + \frac{1}{\sqrt{\pi}}\right)(l\alpha^{-1})^{\frac{2}{3}}\epsilon_n(l)^{\frac{2}{3}} + \frac{1}{2}\epsilon_n(l) < \frac{1}{3}\epsilon$$

for all $n > n_0$. The inequality (11) then shows that

$$\left|F_n(t) - \frac{1}{2} - \frac{1}{\pi}\int_0^\infty e^{-\frac{v^2}{2}}\frac{\sin tv}{v}dv\right| < \epsilon$$

for $n > n_0$ and this means that

$$\lim_{n\to\infty} F_n(t) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty e^{-\frac{v^2}{2}}\frac{\sin tv}{v}dv = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^t e^{-\frac{1}{2}u^2}du$$

uniformly in t because the number n_0, as clearly follows from the preceding analysis, depends upon ϵ only and not upon t.

Remark 1. Without changing anything in the proof, we can state the fundamental lemma in a slightly generalized form as follows: *If t_n tends to the limit t, the probability of the inequality*

$$s_n < t_n$$

tends to

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}u^2}du.$$

Remark 2. The fundamental lemma, although not explicitly stated by Liapounoff, is implicitly contained in his proof. More general propositions of the same nature have been published by Pólya and Lévy. The very elegant result due to the latter can be stated as follows: *If the characteristic function of the variable s_n tends to the characteristic function*

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx}dF(x)$$

of a fixed distribution uniformly in any finite interval, then

$$\lim F_n(t) = F(t)$$

at any point of continuity of $F(t)$.

The above proof, corresponding to the particular case

$$F(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}u^2}du,$$

can be used, almost without any changes, in proving the general proposition of Lévy.

4. Proof of Liapounoff's Theorem. *a.* If Liapounoff's condition

$$\frac{\mu_{2+\delta}^{(1)} + \mu_{2+\delta}^{(2)} + \cdots + \mu_{2+\delta}^{(n)}}{B_n^{1+\frac{\delta}{2}}} \to 0$$

is satisfied for a certain $\delta > 0$, it will be satisfied for all smaller δ.

Let $f_i(t)$ be the distribution function of $x_i (i = 1, 2, \ldots n)$. The sum

$$f(t) = f_1(t) + f_2(t) + \cdots + f_n(t)$$

being a nondecreasing function of t, the following inequality holds (Chap. XIII, Sec. 5):

$$\left(\int_{-\infty}^{\infty}|t|^b df(t)\right)^{a-c} \leq \left(\int_{-\infty}^{\infty}|t|^c df(t)\right)^{a-b} \cdot \left(\int_{-\infty}^{\infty}|t|^a df(t)\right)^{b-c},$$

provided $a > b > c > 0$. We take here

$$a = 2 + \delta, \qquad b = 2 + \delta', \qquad c = 2$$

supposing $0 < \delta' < \delta$. Then

$$\int_{-\infty}^{\infty}|t|^b df(t) = \sum_1^n \mu_{2+\delta'}^{(k)}; \qquad \int_{-\infty}^{\infty}|t|^a df(t) = \sum_1^n \mu_{2+\delta}^{(k)}; \qquad \int_{-\infty}^{\infty}|t|^c df(t) = B_n$$

and

$$\left(\sum_1^n \mu_{2+\delta'}^{(k)}\right)^\delta \leqq B_n^{\delta-\delta'}\left(\sum_1^n \mu_{2+\delta}^{(k)}\right)^{\delta'}.$$

But this inequality is equivalent to

$$\frac{\sum_1^n \mu_{2+\delta'}^{(k)}}{B_n^{1+\frac{\delta'}{2}}} \leqq \left(\frac{\sum_1^n \mu_{2+\delta}^{(k)}}{B_n^{1+\frac{\delta}{2}}}\right)^{\frac{\delta'}{\delta}}$$

and it shows that

$$\frac{\sum_1^n \mu_{2+\delta'}^{(k)}}{B_n^{1+\frac{\delta'}{2}}} \to 0$$

if

$$\frac{\sum_1^n \mu_{2+\delta}^{(k)}}{B_n^{1+\frac{\delta}{2}}} \to 0,$$

provided $0 < \delta' < \delta$. Hence, in the proof we can assume that the fundamental condition is satisfied for some positive $\delta \leqq 1$.

b. Liapounoff's inequality (Chap. XIII, Sec. 5) with $c = 0$, $b = 2$, $a = 2 + \delta$ when applied to x_i gives

$$b_i^{2+\delta} \leqq (\mu_{2+\delta}^{(i)})^2; \qquad b_i = E(x_i^2).$$

Hence,

$$(12) \qquad \frac{b_i}{B_n} \leqq \left(\frac{\mu_{2+\delta}^{(i)}}{B_n^{1+\frac{\delta}{2}}}\right)^{\frac{2}{2+\delta}} < \omega_n^{\frac{2}{2+\delta}}$$

and, since it is assumed that $\omega_n \to 0$, all the quotients

$$\frac{b_i}{B_n} = \frac{b_i}{b_1 + b_2 + \cdots + b_n} \qquad (i = 1, 2, \ldots n)$$

will converge to 0 uniformly as $n \to \infty$.

c. The following formula can easily be obtained by means of integration by parts:

$$e^{ix} = 1 + ix - \frac{x^2}{2} - x^2 \int_0^1 (e^{ixt} - 1)(1 - t)dt.$$

If x is real and in absolute value >2, we have

$$\left| x^2 \int_0^1 (e^{ixt} - 1)(1 - t)dt \right| \leqq x^2 < \frac{|x|^{2+\delta}}{2^\delta}$$

since

$$|e^{ixt} - 1| \leqq 2.$$

If $|x| \leqq 2$, we can use the inequality

$$|e^{ixt} - 1| \leqq 2\left|\frac{x}{2}\right| t \leqq 2\frac{|x|^\delta}{2^\delta} t$$

and find

$$\left| x^2 \int_0^1 (e^{ixt} - 1)(1 - t)dt \right| \leqq \frac{1}{3} \frac{|x|^{2+\delta}}{2^\delta} < \frac{|x|^{2+\delta}}{2^\delta}$$

Thus, for every real x

$$e^{ix} = 1 + ix - \frac{x^2}{2} + \theta \frac{|x|^{2+\delta}}{2^\delta}; \qquad |\theta| \leqq 1.$$

Substituting here

$$x = t\frac{x_k}{\sqrt{B_n}} = t\xi_k$$

and taking the mathematical expectation of both members, we have

$$(13) \qquad \varphi_k(t) = E(e^{it\xi_k}) = 1 - \frac{b_k}{2B_n} t^2 + \theta_k \frac{\mu_{2+\delta}^{(k)}}{2^\delta B_n^{1+\frac{\delta}{2}}} |t|^{2+\delta}; \qquad |\theta_k| \leqq 1.$$

Furthermore, since

$$1 - x = e^{-x} - \frac{\theta}{2} x^2; \qquad x > 0; \qquad 0 < \theta < 1,$$

we can write

$$(14) \qquad 1 - \frac{b_k}{2B_n} t^2 = e^{-\frac{b_k}{2B_n} t^2} - \frac{\theta}{2}\left(\frac{t^2 b_k}{2B_n}\right)^2.$$

If $\omega_n |t|^{2+\delta} < 1$, we shall have, by virtue of (12),

$$\frac{b_k}{B_n} t^2 < 1$$

and consequently

$$\left(\frac{b_k t^2}{2B_n}\right)^2 = \left(\frac{b_k t^2}{2B_n}\right)^{1+\frac{\delta}{2}}\left(\frac{b_k t^2}{2B_n}\right)^{1-\frac{\delta}{2}} < \frac{1}{2^{1-\frac{\delta}{2}}}\left(\frac{b_k t^2}{2B_n}\right)^{1+\frac{\delta}{2}} < \frac{\mu_{2+\delta}^{(k)}}{4B_n^{1+\frac{\delta}{2}}}|t|^{2+\delta}.$$

This inequality, together with (13) and (14), leads to the following expression of $\varphi_k(t)$:

$$\varphi_k(t) = e^{-\frac{b_k}{2B_n}t^2}(1 + \sigma_k) \tag{15}$$

where

$$|\sigma_k| < \frac{9}{8}e^{\frac{1}{4}}\frac{\mu_{2+\delta}^{(k)}}{B_n^{1+\frac{\delta}{2}}}|t|^{2+\delta} < 3\frac{\mu_{2+\delta}^{(k)}}{B_n^{1+\frac{\delta}{2}}}|t|^{2+\delta}. \tag{16}$$

d. The characteristic function of the variable

$$s_n = \frac{x_1 + x_2 + \cdots + x_n}{\sqrt{B_n}}$$

is

$$\varphi(t) = \varphi_1(t)\varphi_2(t) \cdots \varphi_n(t)$$

because $x_1, x_2, \ldots x_n$ are independent variables. Hence, by (15)

$$\varphi(t) = e^{-\frac{1}{2}t^2}(1 + \sigma_1)(1 + \sigma_2) \cdots (1 + \sigma_n)$$

$$|\varphi(t) - e^{-\frac{1}{2}t^2}| < (1 + |\sigma_1|)(1 + |\sigma_2|) \cdots (1 + |\sigma_n|) - 1 < e^{|\sigma_1| + |\sigma_2| + \cdots + |\sigma_n|} - 1$$

and

$$|\varphi(t) - e^{-\frac{1}{2}t^2}| < e^{3\omega_n|t|^{2+\delta}} - 1 \tag{17}$$

taking into account inequalities (16). Inequality (17) holds if

$$\omega_n|t|^{2+\delta} < 1.$$

Suppose, now, that t is confined to an arbitrary finite interval

$$-l \leqq t \leqq l.$$

Because ω_n, by hypothesis, tends to 0, the difference

$$e^{3\omega_n l^{2+\delta}} - 1$$

will tend to 0 as $n \to \infty$. In connection with (17) this shows that

$$\varphi(t) \to e^{-\frac{1}{2}t^2}$$

uniformly in any finite interval. It suffices now to invoke the fundamental lemma to complete the proof of Liapounoff's theorem.

5. Particular Cases. This theorem is extremely general and it is hardly possible to find cases of any practical importance to which it could not be applied. Two particularly significant cases deserve special mention.

First Case. Let us suppose that variables $x_1, x_2, \ldots x_n$ are bounded, so that any possible value of any one of them is absolutely less than a constant C. Evidently

$$\mu_{2+\delta}^{(k)} \leqq C^\delta E(x_i^2) = C^\delta b_i$$

and hence

$$\omega_n \leqq \frac{C^\delta}{B_n^{\frac{\delta}{2}}}.$$

It suffices to assume that

$$B_n = b_1 + b_2 + \cdots + b_n$$

tends to infinity to be sure that $\omega_n \to 0$. Hence, dealing with bounded independent variables, the condition for the validity of the limit theorem is

$$B_n \to \infty \qquad \text{as} \qquad n \to \infty,$$

which is equivalent to the statement that the series

$$b_1 + b_2 + b_3 + \cdots$$

is divergent.

Poisson's series of trials affords a good illustration of this case. In the usual way, we attach to each of the trials a variable which assumes two values, 1 and 0, according as an event E occurs or fails in that trial. Let p_i and $q_i = 1 - p_i$ be the respective probabilities of the occurrence and failure of E in the ith trial. The variable z_i attached to this trial is defined by

$$z_i = 1 \text{ if } E \text{ occurs,}$$
$$z_i = 0 \text{ if } E \text{ fails.}$$

Noticing that

$$E(z_i) = p_i,$$

we introduce new variables

$$x_i = z_i - p_i \qquad (i = 1, 2, \ldots n)$$

with the mean 0, whose sum is given by

$$m - np$$

where m is the number of occurrences of E in n trials and p the mean probability

$$p = \frac{p_1 + p_2 + \cdots + p_n}{n}.$$

In our case

$$E(x_i^2) = p_i q_i$$

and

$$B_n = \sum_1^n p_i q_i.$$

Hence, we can formulate the following theorem:

Theorem. *The probability of the inequality*

$$m - np < t\sqrt{B_n}$$

tends uniformly to the limit

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}} du$$

as $n \to \infty$, provided the series

$$\sum_1^\infty p_i q_i$$

is divergent. At the same time the probability of the inequalities

$$t_1\sqrt{B_n} < m - np < t_2\sqrt{B_n}$$

tends uniformly (in t_1, t_2) to the limit

$$\frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{u^2}{2}} du.$$

Second Case. Let $z_1, z_2, \ldots z_n$ be identical variables with the common mean a and dispersion b. Supposing that for some positive δ

$$E|z_i - a|^{2+\delta} = c$$

exists, we have

$$\omega_n = \frac{nc}{(nb)^{1+\frac{\delta}{2}}} = \frac{c}{b^{1+\frac{\delta}{2}}} \cdot n^{-\frac{\delta}{2}},$$

and hence $\omega_n \to 0$ as $n \to \infty$. The limit theorem applied to this case can be stated as follows:

The probability of the inequality

$$z_1 + z_2 + \cdots + z_n - na < t\sqrt{nb}$$

tends uniformly to

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}} du,$$

provided

$$E|z_i - a|^{2+\delta}$$

exists for some positive δ. As a corollary we have: *The probability of the inequalities*

$$-t\sqrt{\frac{b}{n}} < \frac{z_1 + z_2 + \cdots + z_n}{n} - a < t\sqrt{\frac{b}{n}}$$

tends to

$$\frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{u^2}{2}}du.$$

This proposition is regarded as justification of the ordinary procedure of taking a mean of several observed measurements of the same quantity, made under the same conditions, to approximate its "true value." Barring systematical errors which should be eliminated by a careful study of the tools used for measurements, the true value of the unknown quantity is regarded as coinciding with the expectation of a set of potentially possible values each having a certain probability of materializing in actual measurement. Since for comparatively small t the above integral comes very near to 1 and

$$t\sqrt{\frac{b}{n}}$$

for large n becomes as small as we please, the probability of the mean of a very large number of observations deviating very little from the true value of the quantity to be measured, will be close to 1 and herein lies the justification of the rule of mean mentioned above.

Estimation of the Error Term

6. The limit theorem is a proposition of an essentially asymptotic character. It states merely that the distribution function $F_n(t)$ of the variable

$$s_n = \frac{x_1 + x_2 + \cdots + x_n}{\sqrt{B_n}}$$

approaches the limit

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^t e^{-\frac{u^2}{2}}du$$

as n becomes infinite when a certain condition is fulfilled. For practical purposes it is very important to estimate the error committed by replacing $F_n(t)$ by its limit when n is a finite but very large number. In his original paper Liapounoff had this important problem in his mind and for that reason entered into more detailed elaboration of various parts of his proof than was strictly necessary to establish an asymptotic theorem.

We do not intend to reproduce here this part of Liapounoff's investigation; it suffices to indicate the final result. Assuming the existence of absolute moments of the third order $E|x_i|^3$; $i = 1, 2, \ldots n$, we shall suppose n so large that

$$\omega_n = \frac{\mu_3^{(1)} + \mu_3^{(2)} + \cdots + \mu_3^{(n)}}{B_n^{3/2}} < \frac{1}{20}.$$

Then, setting

$$F_n(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}u^2}du + R,$$

we shall have

$$|R| < \frac{8}{5}\omega_n\left[\left(\log\frac{1}{3\omega_n}\right)^{\frac{1}{2}} + 1.1\right] + \omega_n^2\log\frac{1}{3\omega_n} + \frac{5}{3}\omega_n^{\frac{1}{3}}e^{-\frac{1}{8}\omega_n^{-\frac{2}{3}}}.$$

Although this limit for the error term is probably too high, it seems to be the best available. However, it is greatly desirable to have a more genuine estimation of R.

7. Hypothesis of Elementary Errors. It is considered as an experimental fact that accidental errors of observations (or measurements) follow closely the law of normal distribution. In the sphere of biology, similar phenomena have been observed as to the size of the bodies and various organs of living organisms. What can be suggested as an explanation of these observed facts? In regard to errors of observations, Laplace proposed a hypothesis which may sound plausible. He considers the total error as a sum of numerous very small elementary errors due to independent causes.

It can hardly be doubted that various independent or nearly independent causes contribute to the total error. In astronomical observations, for instance, slight changes in the temperature, irregular currents of air, vibrations of buildings, and even the state of the organs of perception of an observer may be considered as but a small part of such causes. One can easily understand that the growth of the organs of living organisms is also dependent on many factors of accidental character which independently tend to increase or decrease the size of the organs. If, on the ground of such evidence, we accept Laplace's hypothesis, we can try the explanation of the normal law of distribution on the basis of the general theorems established above.

Suppose that elementary errors do not exceed in absolute value a certain number l, very small compared with the standard deviation σ of their sum. The quantity denoted by ω_n in the preceding section will be less than the ratio l/σ and hence will be a small number; and the same will be true of the error term R. Hence, the distribution of the total error will be nearly normal.

Laplace's explanation of the observed prevalence of normal distributions may be accepted as plausible, at least. But the question may be raised whether elementary errors are small enough and numerous enough to make the difference between the true distribution function of the total error and that of a normal distribution small. Besides, Laplace's hypothesis is based on the principle of superposition of small effects and thus introduces another assumption of an arbitrary character.

Finally, the experimental data quoted in support of the normal distribution of errors of observations and biological measurements are not numerous enough for one to place full confidence in them. Hence, the widely accepted statistical theories based on the normal law of distribution cannot be fully relied on and may be considered merely as substitutes for more accurate knowledge which we do not yet possess in dealing with problems of vital importance in the sphere of human activities.

LIMIT THEOREMS FOR DEPENDENT VARIABLES

8. The fundamental limit theorem can be extended to sums of dependent variables as, under special assumptions, was shown first by Markoff and later by S. Bernstein, whose work may be considered an outstanding recent contribution to the theory of probability. However, the conditions for the validity of the theorems established by Bernstein are rather complicated, and the whole subject seems to lack ultimate simplicity. For that reason we confine ourselves here to a few special cases.

Example 1. Let us consider a simple chain in which probabilities for an event E to occur in any trial are p' and p'', respectively, according as E occurred or failed in the preceding trial. The probability for E to occur at the nth trial when the results of other trials are unknown is

$$p_n = p + (p_1 - p)\delta^{n-1}$$

where p_1 is the initial probability, $\delta = p' - p''$ and

$$p = \frac{p''}{1 - \delta}.$$

The mean probability for n trials is given by

$$\bar{p}_n = p + \frac{p_1 - p}{n}\,\frac{1 - \delta^n}{1 - \delta}$$

so that p may be considered as the mean probability in infinitely many trials.

In the usual way, to trials 1, 2, 3, . . . we attach variables $x_1, x_2, x_3, \ldots$ so that in general

$$x_i = 1 - p_i \qquad \text{or} \qquad x_i = -p_i$$

according as E occurs or fails in the ith trial. If m is the number of occurrences of E in n trials, the sum

$$x_1 + x_2 + \cdots + x_n$$

of dependent variables represents

$$m - n\bar{p}_n.$$

Evidently

$$E(m - n\bar{p}_n) = 0$$

and, as we have seen in Chap. XI, Sec. 7,

$$B_n = E(m - n\bar{p}_n)^2 \sim npq\frac{1 + \delta}{1 - \delta};$$

that is, the ratio of B_n: $npq\frac{1+\delta}{1-\delta}$ tends to 1 as n becomes infinite.

In order to find an appropriate expression of the characteristic function of the quotient

$$\frac{m - n\bar{p}_n}{\sqrt{B_n}}$$

we shall endeavor first to find the generating function $\omega_n(t)$ for probabilities

$$P_{m,n}(m = 0, 1, 2, \ldots n)$$

to have exactly m occurrences of E in n trials. Let $A_{m,n}$ be the probability of m occurrences when the whole series ends with E and similarly $B_{m,n}$ the probability of m occurrences when this series ends with F, the event opposite to E. The following relations follow immediately from the definition of a chain

$$(18)\qquad \begin{aligned} A_{m,n+1} &= A_{m-1,n}p' + B_{m-1,n}p'' \\ B_{m,n+1} &= A_{m,n}q' + B_{m,n}q''. \end{aligned}$$

Let

$$\theta_n(t) = \sum_{m=0}^{\infty} A_{m,n}t^m, \qquad \psi_n(t) = \sum_{m=0}^{\infty} B_{m,n}t^m$$

be the generating function of $A_{m,n}$ and $B_{m,n}$. From relations (18) it follows that

$$(19)\qquad \begin{aligned} \theta_{n+1}(t) &= p't\theta_n(t) + p''t\psi_n(t) \\ \psi_{n+1}(t) &= q'\theta_n(t) + q''\psi_n(t). \end{aligned}$$

These relations established for $n \geqq 1$ will hold even for $n = 0$ if we define $\theta_0(t)$ and $\psi_0(t)$ by

$$\begin{aligned} p'\theta_0 + p''\psi_0 &= p_1 \\ q'\theta_0 + q''\psi_0 &= 1 - p_1 \end{aligned}$$

whence

$$\theta_0 + \psi_0 = 1.$$

From (19) one can easily conclude that both $\theta_n(t)$ and $\psi_n(t)$ satisfy the same equation in finite differences of the second order

$$\begin{aligned} \theta_{n+2} - (p't + q'')\theta_{n+1} + \delta t\theta_n &= 0 \\ \psi_{n+2} - (p't + q'')\psi_{n+1} + \delta t\psi_n &= 0. \end{aligned}$$

Evidently

$$P_{m,n} = A_{m,n} + B_{m,n};$$

hence

$$\omega_n(t) = \theta_n(t) + \psi_n(t)$$

satisfies the equation

$$\omega_{n+2} - (p't + q'')\omega_{n+1} + \delta t\omega_n = 0 \tag{20}$$

and is completely determined by it and the initial conditions

$$\omega_0 = 1, \qquad \omega_1 = q_1 + p_1 t.$$

Since

$$p' = p + q\delta, \qquad q'' = q + p\delta$$

the characteristic equation corresponding to (20) can be written

$$(\zeta - 1)(\zeta - \delta) = (t - 1)[(p + q\delta)\zeta - \delta]$$

and for small $t - 1$ its roots can be expanded into power series

$$\zeta_1 = 1 + c_1(t - 1) + c_2(t - 1)^2 + \cdots$$
$$\zeta_2 = \delta + d_1(t - 1) + d_2(t - 1)^2 + \cdots.$$

The general expression of $\omega_n(t)$ will be

$$\omega_n(t) = A\zeta_1^n + B\zeta_2^n = A\zeta_1^n + B\delta^n t^n \zeta_1^{-n}$$

where to satisfy the initial conditions we must take

$$A = \frac{\zeta_2 - q_1 - p_1 t}{\zeta_2 - \zeta_1}; \qquad B = \frac{-\zeta_1 + q_1 + p_1 t}{\zeta_2 - \zeta_1}.$$

Having found $\omega_n(t)$, the characteristic function of

$$s_n = \frac{m - n\bar{p}_n}{\sqrt{B_n}}$$

will be given by

$$\varphi_n(v) = e^{-n\bar{p}_n \cdot \frac{vi}{\sqrt{B_n}}}\omega_n\left(e^{i\frac{v}{\sqrt{B_n}}}\right).$$

To study the asymptotic behavior of $\varphi_n(v)$ when v is confined to a finite fixed interval $-l \leq v \leq l$, we notice that then

$$u = \frac{v}{\sqrt{B_n}}$$

will be well within the convergence region of the series we are going to consider now. By means of Lagrange's series or otherwise, we find the following expansion of $\log \zeta_1$ in power series of $t - 1$

$$\log \zeta_1 = p(t - 1) - \left(\frac{p^2}{2} - \frac{pq\delta}{1 - \delta}\right)(t - 1)^2 + \cdots$$

convergent for sufficiently small values of $t - 1$. By setting $t = e^{iu}$ we obtain another power series in u

$$\log \zeta_1 = piu - \frac{pq}{2}\frac{1 + \delta}{1 - \delta}u^2 + \cdots$$

convergent for sufficiently small u. Hence

$$\zeta_1^n = e^{npui - npq\frac{1+\delta}{1-\delta}\frac{u^2}{2} + nu^3g(u)}$$
$$\zeta_1^{-n} = e^{-npui + npq\frac{1+\delta}{1-\delta}\frac{u^2}{2} - nu^3g(u)}$$

where $g(u)$ is a bounded function of u, u being contained in a certain interval $(-r, r)$. By substituting

$$u = \frac{v}{\sqrt{B_n}}$$

here, we easily conclude that

$$e^{-n\bar{p}_n\frac{vi}{\sqrt{B_n}}}\zeta_1^n$$

tends uniformly to the limit

$$e^{-\frac{v^2}{2}}$$

in the interval $-l \leqq v \leqq l$ while

$$e^{-n\bar{p}_n\frac{vi}{\sqrt{B_n}}}\zeta_1^{-n}$$

remains there uniformly bounded. Since, as can easily be seen, A and B can be represented by power series

$$\begin{aligned} A &= 1 + a_1u + a_2u^2 + \cdots \\ B &= -a_1u - a_2u^2 - \cdots \end{aligned}$$

A tends uniformly to 1 and B tends uniformly to 0. Hence, finally, $\varphi_n(v)$ in any fixed interval $-l \leqq v \leqq l$ tends uniformly to $e^{-\frac{v^2}{2}}$. It suffices to apply the fundamental lemma to conclude that the probability of the inequality

$$m - n\bar{p}_n < t_n\sqrt{B_n}$$

tends uniformly to the limit

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{u^2}{2}}du$$

if t_n tends to t.

Since B_n is asymptotic to $npq\frac{1+\delta}{1-\delta}$ and $\bar{p}_n$ differs from p by a quantity of the order $1/n$, the inequality

$$m - np < t\sqrt{\frac{1+\delta}{1-\delta}npq}$$

can be written in the form

$$m - n\bar{p}_n < t_n\sqrt{B_n}$$

with t_n tending to t, whence, using the above established result, the following theorem due to Markoff can be derived:

Theorem. *For a simple chain the probability of the inequalities*

$$t_1\sqrt{\frac{1+\delta}{1-\delta}npq} < m - np < t_2\sqrt{\frac{1+\delta}{1-\delta}npq}$$

tends to the limit

$$\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2} e^{-\frac{u^2}{2}}du$$

as $n \to \infty$.

Example 2. Considering an indefinite series of Bernoullian trials with the probability p for an event A to occur, we can regard pairs of consecutive trials 1 and 2, 2 and 3, 3 and 4, and so on, as forming a new series of trials which may produce an event E consisting of two successive occurrences of $A (E = AA)$ or an event F opposite to E $(F = AB, BA, BB)$. With respect to E the trials of the new series are no longer independent. Let m be the number of occurrences of E in n trials. Then

$$E(m - np^2) = 0$$

and

$$B_n = E(m - np^2)^2 = np^2q(1 + 3p) - 2p^3q$$

as was shown in Chap. XI, Sec. 6.

Let $P_{m,n}$ be the probability of exactly m occurrences of E in a series of n trials. Evidently

$$P_{m,n} = A_{m,n} + B_{m,n}$$

where $A_{m,n}$ and $B_{m,n}$ are the probabilities of m occurrences of E when the Bernoullian series of $n + 1$ trials ends with A or B, respectively. By an easy application of the theorems of total and compound probabilities we get

$$A_{m,n+1} = A_{m-1,n}p + B_{m,n}p$$
$$B_{m,n+1} = A_{m,n}q + B_{m,n}q.$$

Corresponding to these relations the generating functions

$$\theta_n(t) = \sum_{m=0}^{\infty} A_{m,n}t^m, \qquad \psi_n(t) = \sum_{m=0}^{\infty} B_{m,n}t^m$$

satisfy the following equations in finite differences:

$$\theta_{n+1} = pt\theta_n + p\psi_n$$
$$\psi_{n+1} = q\theta_n + q\psi_n$$

holding even for $n = 0$ if we set $\theta_0 = p$, $\psi_0 = q$. Hence, it follows that $\theta_n(t)$ and $\psi_n(t)$ satisfy the same equations of the second order

$$\theta_{n+2} - (pt + q)\theta_{n+1} + pq(t - 1)\theta_n = 0$$
$$\psi_{n+2} - (pt + q)\psi_{n+1} + pq(t - 1)\psi_n = 0$$

and so does their sum

$$\omega_n(t) = \theta_n(t) + \psi_n(t) = \sum_{m=0}^{\infty} P_{m,n}t^m.$$

Thus, to determine $\omega_n(t)$ we have the equation

$$\omega_{n+2} - (pt + q)\omega_{n+1} + pq(t - 1)\omega_n = 0$$

and the initial conditions

$$\omega_0 = 1, \qquad \omega_1 = 1 - p^2 + p^2t.$$

The general expression of $\omega_n(t)$ is

$$\omega_n(t) = A\zeta_1^n + B\zeta_2^n = A\zeta_1^n + Bp^nq^n(t - 1)^n\zeta_1^{-n}$$

where ζ_1 and ζ_2 are roots of the equation

$$\zeta^2 - \zeta = p(t - 1)(\zeta - q)$$

and

$$A = \frac{-\zeta_2 + 1 + p^2(t - 1)}{\zeta_1 - \zeta_2}; \qquad B = \frac{\zeta_1 - 1 - p^2(t - 1)}{\zeta_1 - \zeta_2}.$$

If ζ_1 is the root which for $t = 1$ reduces to 1, we easily find the following series

$$\log \zeta_1 = p^2(t - 1) + \frac{p^2(-p^2 + 2pq)}{2}(t - 1)^2 + \cdots$$

or, setting $t = e^{iu}$ and supposing u sufficiently small,

$$\log \zeta_1 = ip^2u - \frac{p^2q(1 + 3p)}{2}u^2 + \cdots.$$

As to A and B, they can be developed into series of the form

$$A = 1 + cu^2 + \cdots$$
$$B = -cu^2 + \cdots$$

Hence, reasoning in the same manner as in Example 1, we can conclude that the characteristic function

$$\varphi_n(v) = e^{-\frac{np^2v}{\sqrt{B_n}}i}\omega_n\left(e^{\frac{iv}{\sqrt{B_n}}}\right)$$

of the variable

$$\frac{m - np^2}{\sqrt{B_n}}$$

tends to the limit $e^{-\frac{v^2}{2}}$ uniformly in any finite and fixed interval $-l \leqq v \leqq l$. Referring, finally, to the fundamental lemma, we reach the following conclusion: The probability of the inequalities

$$t_1\sqrt{np^2q(1 + 3p)} < m - np^2 < t_2\sqrt{np^2q(1 + 3p)}$$

tends uniformly (with respect to t_1 and t_2) to the limit

$$\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2} e^{-\frac{u^2}{2}}du$$

as $n \to \infty$.

Problems for Solution

1. Consider a series of independent variables $x_1, x_2, x_3, \ldots$ where in general x_k ($k = 1, 2, 3, \ldots$) can have only two values k^α and $-k^\alpha$ each with the probability $\frac{1}{2}$. Show that the limit theorem holds for the variables thus defined if $\alpha > -\frac{1}{2}$, but the law of large numbers holds only if $\alpha < \frac{1}{2}$.

Solution. Evidently

$$E(x_k) = 0, \qquad E(x_k^2) = k^{2\alpha}, \qquad E|x_k|^3 = k^{3\alpha}.$$

From Euler's formula (Appendix I) we derive two asymptotic expressions

$$B_n = 1^{2\alpha} + 2^{2\alpha} + \cdots + n^{2\alpha} \sim \frac{n^{2\alpha+1}}{2\alpha + 1}$$

$$1^{3\alpha} + 2^{3\alpha} + \cdots + n^{3\alpha} \sim \frac{n^{3\alpha+1}}{3\alpha + 1}.$$

Hence

$$\omega_n \sim \frac{(2\alpha + 1)^{\frac{3}{2}}}{3\alpha + 1} n^{-\frac{1}{2}}, \qquad \omega_n \to 0$$

so that the limit theorem holds. For $\alpha = \frac{1}{2}$ the probability of the inequalities

$$-\epsilon < \frac{x_1 + x_2 + \cdots + x_n}{n} < \epsilon$$

tends to the limit

$$\sqrt{\frac{2}{\pi}} \int_0^{\epsilon\sqrt{2}} e^{-\frac{1}{2}u^2} du = \frac{2}{\sqrt{\pi}} \int_0^{\epsilon} e^{-u^2} du$$

and the law of large numbers does not hold.

2. Let m_i be the number of successes in i Bernoullian trials with the probability p. Show that the limit theorem holds for variables

$$s_i = \frac{m_i - sp}{\sqrt{ipq}}; \qquad i = 1, 2, \ldots n$$

but the law of large numbers does not hold (Bernstein).

HINT:

$$s_1 + s_2 + \cdots + s_n = (pq)^{-\frac{1}{2}} \left[\left(1 + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n}}\right) x_1 + \right.$$
$$\left. + \left(\frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n}}\right) x_2 + \cdots + \frac{1}{\sqrt{n}} x_n \right]$$

where $x_1, x_2, \ldots x_n$ are independent variables with two values q and $-p$ associated in the customary way with trials $1, 2, \ldots n$.

3. Consider an infinite sequence of independent variables $x_1, x_2, x_3, \ldots$ where x_k can have three values

$$0, \ (\log k)^\mu, \ -(\log k)^\mu$$

with the corresponding probabilities

$$1 - \frac{2}{(k+\alpha)\{\log(k+\alpha)\}^\rho}, \quad \frac{1}{(k+\alpha)\{\log(k+\alpha)\}^\rho}, \quad \frac{1}{(k+\alpha)\{\log(k+\alpha)\}^\rho}$$

α being a sufficiently large constant. Moreover, μ and ρ satisfy the inequality

$$2\mu - \rho + 1 > 0.$$

Show (*a*) that Liapounoff's condition is satisfied when $\rho < 1$ and hence the limit theorem holds; (*b*) that this condition is not satisfied if $\rho \geqq 1$ and at the same time the limit theorem fails at least for $\rho > 1$.

Solution. *a.* By using Euler's formula we find

$$\omega_n \sim \frac{(2\mu + 1 - \rho)^{1+\frac{\delta}{2}}}{(2+\delta)\mu + 1 - \rho}\{\log(n+\alpha)\}^{\frac{\delta}{2}(\rho-1)}.$$

Hence the first part is answered.

b. The probability of the inequality

$$x_1 + x_2 + \cdots + x_n \gtrless 0$$

is less than

$$2\sum_{k=1}^{n} \frac{1}{(k+\alpha)\{\log(k+\alpha)\}^\rho}$$

and this, in case $\rho > 1$, is less than

$$\frac{2}{\rho - 1}(\log \alpha)^{1-\rho}.$$

Hence, the probability of the equality

$$x_1 + x_2 + \cdots + x_n = 0$$

remains always $>1 - \dfrac{2}{\rho - 1}(\log \alpha)^{1-\rho}$ and the limit theorem cannot hold. Note that $B_n \to \infty$ because $2\mu - \rho + 1 > 0$.

4. Prove the asymptotic formula

$$1 + n + \frac{n^2}{1 \cdot 2} + \cdots + \frac{n^n}{1 \cdot 2 \cdots n} \sim \frac{1}{2}e^n,$$

n being a large integer.

HINT: Apply Liapounoff's theorem to n variables distributed according to Poisson's law with parameter 1.

5. By resorting to the fundamental lemma, prove the following theorem due to Markoff: If for a variable s_n with the mean $= 0$ and the standard deviation $= 1$

$$\lim_{n\to\infty} E(s_n^k) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t^k e^{-\frac{1}{2}t^2}dt$$

for any given $k = 3, 4, 5, \ldots$, then the probability of the inequality $s_n < t$ tends to the limit

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}u^2}du.$$

6. In many special cases the limit of the error term can be considerably lower than that given in Sec. 6. For instance, if variables $x_1, x_2, \ldots x_n$ are identical and uniformly distributed in the interval $-\frac{1}{2}, \frac{1}{2}$ the probability $F_n(t)$ of the inequality

$$x_1 + x_2 + \cdots + x_n < t\sqrt{\frac{n}{12}}$$

differs from

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{u^2}{2}}du$$

by less (in absolute value) than

$$\frac{1}{7.5n} + \frac{1}{\pi}\left(\frac{2}{\pi}\right)^n + \frac{12}{\pi^3 n}e^{-\frac{\pi^2 n}{24}}$$

the last two terms being completely negligible for somewhat large n.

Indication of the Proof. First establish the inequalities

$$\frac{\sin\varphi}{\varphi} < e^{-\frac{\varphi^2}{6}}, \qquad \frac{\sin\varphi}{\varphi} > e^{-\frac{\varphi^2}{6}-\frac{\varphi^4}{135}}$$

for $0 \leqq \varphi \leqq \pi/2$. Further, represent $F_n(t)$ by the integral

$$F_n(t) = \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty}\left(\frac{\sin v\sqrt{\frac{3}{n}}}{v\sqrt{\frac{3}{n}}}\right)^n \frac{\sin vt}{v}dv$$

and split it into two integrals taken between 0 and $\pi\sqrt{n}/\sqrt{12}$ and $\pi\sqrt{n}/\sqrt{12}$ and $+\infty$.

7. Supposing again that $x_1, x_2, \ldots x_n$ are identical and uniformly distributed in the interval $-\frac{1}{2}, \frac{1}{2}$, prove that for $n \geqq 2$

$$E|x_1 + x_2 + \cdots + x_n| = \sqrt{\frac{n}{6\pi}} + \frac{\theta}{60\sqrt{n}}; \qquad 0 < \theta < 1.$$

8. Let s_n be a variable with the mean $= 0$ and standard deviation $= 1$. If its characteristic function $\varphi_n(t)$ tends to $e^{-\frac{1}{2}t^2}$ as $n \to \infty$ uniformly in any finite interval $-l \leqq t \leqq l$, show that

$$E|s_n| \to \sqrt{\frac{2}{\pi}}.$$

HINT:

$$E|s_n| - \sqrt{\frac{2}{\pi}} = \frac{2}{\pi}R\int_0^{\infty}\frac{e^{-\frac{1}{2}t^2} - \varphi_n(t)}{t^2}dt.$$

9. If independent variables $x_1, x_2, \ldots x_n$ with means $=0$ satisfy Liapounoff's condition, prove that

$$E|x_1 + x_2 + \cdots + x_n| \sim \sqrt{\frac{2}{\pi}B_n}.$$

10. Show that for a simple chain of trials

$$E|m - np| \sim \sqrt{\frac{2npq}{\pi}\frac{1+\delta}{1-\delta}},$$

p being the mean probability in infinite series of trials and $\delta = p' - p''$.

11. A series of dependent trials can be illustrated by the following urn scheme: Two urns, 1 and 2, contain white and black balls in such proportions that the probability of drawing a white ball from 1 is p, whereas the probability of drawing a white ball from 2 is $q = 1 - p$. Whenever a ball taken from an urn is white, the next ball is taken from the same urn, but if it is black, the next ball is drawn from the other urn. The urn at the first drawing is selected by lot, the probabilities of selecting the first or the second urn being given. Evidently the course of trials is determined by these rules without any ambiguity. Let m denote the number of white balls obtained in n drawings and let

$$\alpha = p^2 + q^2.$$

Show that the probability of the inequality

$$m - n\alpha < t\sqrt{L\alpha(1-\alpha)n}; \qquad L = \frac{2(1-3pq)}{1-2pq}$$

approaches the limit

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{u^2}{2}}du.$$

Indication of the Proof. Let

$$P_{m,n}^{(1)};\ P_{m,n}^{(2)};\ P_{m,n}^{(3)};\ P_{m,n}^{(4)}$$

be the probabilities of having m white balls in n trials when (a) the last ball is white and from urn 1; (b) the last ball is white and from urn 2; (c) the last ball is black and from urn 1; and (d) the last ball is black and from urn 2. The sum

$$P_{m,n} = P_{m,n}^{(1)} + P_{m,n}^{(2)} + P_{m,n}^{(3)} + P_{m,n}^{(4)}$$

represents the probability of having exactly m white balls in n trials. The generating functions of probabilities $P_{m,n}^{(i)}$ satisfy the following equations

$$\begin{aligned}
\varphi_{n+1}^{(1)} &= pt(\varphi_n^{(1)} + \varphi_n^{(4)}) \\
\varphi_{n+1}^{(2)} &= qt(\varphi_n^{(2)} + \varphi_n^{(3)}) \\
\varphi_{n+1}^{(3)} &= q(\varphi_n^{(1)} + \varphi_n^{(4)}) \\
\varphi_{n+1}^{(4)} &= p(\varphi_n^{(2)} + \varphi_n^{(3)})
\end{aligned}$$

whence it can be shown that they all, as well as their sum—the generating function of $P_{m,n}$—satisfy the same equation of the second order

$$z_{n+2} - tz_{n+1} + pq(t^2 - 1)z_n = 0.$$

Setting $t = e^{iu}$, one of the characteristic roots will be given by

$$e^{(1-2pq)iu-4pq(1-3pq)\frac{u^2}{2}+\cdots}$$

for small u, while the other root tends to 0 as $u \to 0$. The final conclusion can now be reached in the same way as in Examples 1 and 2, pages 297 and 301.

References

P. Laplace: "Théorie analytique des probabilités," Oeuvres VII, pp. 309*ff.*

S. Poisson: "Recherches sur la probabilité des jugements en matière criminelle et en matière civile," pp. 246*ff.*, Paris, 1837.

P. Tshebysheff: "Sur deux théorèmes relatifs aux probabilités," Oeuvres 2, pp. 481–491.

A. Markoff: The Law of Large Numbers and the Method of the Least Squares (Russian), *Proc. Phys.-Math. Soc. Kazan*, 1898.

———: Sur les racines de l'equation $e^{x^2}\frac{d^n e^{-x^2}}{dx^n} = 0$, *Bull. Acad. Sci. St. Petersbourg*, 9, 1898.

———: "Démonstration du second théorème limite du calcul des probabilités par la méthode des moments," St. Petersbourg, 1913.

———: "Wahrscheinlichkeitsrechnung," Leipzig, 1912.

A. Liapounoff: Sur une proposition de la théorie des probabilités, *Bull. Acad. Sci. St. Petersbourg*, 13, 1900.

———: Nouvelle forme du théorème sur la limite de probabilité, *Mém. Acad. Sci. St. Petersbourg*, 12, 1901.

G. Pólya: Ueber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung, *Math. Z.*, 8, 1920.

J. W. Lindeberg: Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, *Math. Z.*, 15, 1922.

P. Levy: "Calcul des probabilités," Paris, 1925.

S. Bernstein: Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes, *Math. Ann.*, 97, 1927.

A. Kolmogoroff: Eine Verallgemeinerung des Laplace-Liapounoffschen Satzes, *Bull. Acad. Sci. U.S.S.R.*, p. 959, 1931.

———: Ueber die Grenzwertsätze der Wahrscheinlichkeitsrechnung, *Bull. Acad. Sci. U.S.S.R.*, p. 363, 1933.

A. Khintchine: "Asymptotische Gesetze der Wahrscheinlichkeitsrechnung," Berlin, 1933.

CHAPTER XV

NORMAL DISTRIBUTION IN TWO DIMENSIONS. LIMIT THEOREM FOR SUMS OF INDEPENDENT VECTORS. ORIGIN OF NORMAL CORRELATION

1. The concept of normal distribution can easily be extended to two and more variables. Since the extension to more than two variables does not involve new ideas, we shall confine ourselves to the case of two-dimensional normal distribution.

Two variables, x, y, are said to be normally distributed if for them the density of probability has the form

$$e^{-\varphi}$$

where

$$\varphi = ax^2 + 2bxy + cy^2 + 2dx + 2ey + f$$

is a quadratic function of x, y becoming positive and infinitely large together with $|x| + |y|$. This requirement is fulfilled if, and only if,

$$ax^2 + 2bxy + cy^2$$

is a positive quadratic form. The necessary and sufficient conditions for this are:

$$a > 0; \qquad ac - b^2 = \Delta > 0.$$

Since $\Delta > 0$ (even a milder requirement $\Delta \neq 0$ suffices), constants x_0, y_0 can be found so that

$$\varphi = a(x - x_0)^2 + 2b(x - x_0)(y - y_0) + c(y - y_0)^2 + g$$

identically in x, y. It follows that the density of probability $e^{-\varphi}$ may be presented thus:

$$e^{-\varphi} = Ke^{-a(x-x_0)^2-2b(x-x_0)(y-y_0)-c(y-y_0)^2}.$$

The expression in the right member depends on six parameters K; a, b, c; x_0, y_0. But the requirement

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\varphi}dxdy = 1$$

reduces the number of independent parameters to five. We can take a, b, c; x_0, y_0 for independent parameters and determine K by the condition

$$K\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-a(x-x_0)^2-2b(x-x_0)(y-y_0)-c(y-y_0)^2}dxdy = 1$$

which, by introducing new variables

$$\xi = x - x_0, \qquad \eta = y - y_0$$

can be exhibited thus

$$K\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-a\xi^2-2b\xi\eta-c\eta^2}d\xi d\eta = 1.$$

To evaluate this and similar double integrals we observe that the positive quadratic form

$$a\xi^2 + 2b\xi\eta + c\eta^2$$

can be presented in infinitely many ways as a sum of two squares

$$a\xi^2 + 2b\xi\eta + c\eta^2 = (\alpha\xi + \beta\eta)^2 + (\gamma\xi + \delta\eta)^2,$$

whence

$$a = \alpha^2 + \gamma^2; \qquad c = \beta^2 + \delta^2; \qquad b = \alpha\beta + \gamma\delta$$

and

$$(\alpha\delta - \beta\gamma)^2 = \Delta.$$

By changing the signs of α and β if necessary, we can always suppose

$$\alpha\delta - \beta\gamma = +\sqrt{\Delta}.$$

Now we take

$$u = \alpha\xi + \beta\eta; \qquad v = \gamma\xi + \delta\eta$$

for new variables of integration. Since the Jacobian of u, v with respect to ξ, η is $\sqrt{\Delta}$, the Jacobian of ξ, η with respect to u, v will be $1/\sqrt{\Delta}$ and, by the known rules

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-a\xi^2-2b\xi\eta-c\eta^2}d\xi d\eta = \frac{1}{\sqrt{\Delta}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-u^2-v^2}dudv = \frac{\pi}{\sqrt{\Delta}}.$$

Thus

$$\frac{K\pi}{\sqrt{\Delta}} = 1, \qquad K = \frac{\sqrt{\Delta}}{\pi}.$$

That is, the general expression for the density of probability in two-dimensional normal distribution is

$$\frac{\sqrt{ac - b^2}}{\pi}e^{-a(x-x_0)^2-2b(x-x_0)(y-y_0)-c(y-y_0)^2}.$$

2. Parameters x_0, y_0 represent the mean values of variables x, y. To prove this, let us consider

$$E(x - x_0) = \frac{\sqrt{\Delta}}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - x_0)e^{-a(x-x_0)^2-2b(x-x_0)(y-y_0)-c(y-y_0)^2}dxdy.$$

To evaluate the double integral, we can express x and y through new variables u, v introduced in the preceding section. We have

$$x - x_0 = \frac{\delta u - \beta v}{\sqrt{\Delta}}, \qquad y - y_0 = \frac{-\gamma u + \alpha v}{\sqrt{\Delta}}$$

and

$$E(x - x_0) = \frac{1}{\pi\sqrt{\Delta}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\delta u - \beta v)e^{-u^2-v^2}dudv = 0,$$

whence

$$E(x) = x_0,$$

and similarly

$$E(y) = y_0.$$

3. Having found the meaning of x_0, y_0 we may consider instead of x, y, variables $x - x_0$, $y - y_0$ whose mean values $= 0$. Denoting these new variables by x, y again the expression of the density of probability for x, y will be:

$$\frac{\sqrt{ac - b^2}}{\pi}e^{-ax^2-2bxy-cy^2}.$$

It contains only three parameters, a, b, c. To find the intrinsic meaning of a, b, c let us consider the mathematical expectation of $(x + \lambda y)^2$ where λ is an arbitrary constant. We have

$$E(x + \lambda y)^2 = \frac{\sqrt{\Delta}}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x + \lambda y)^2 e^{-ax^2-2bxy-cy^2}dxdy,$$

or, introducing u, v defined as in Sec. 1 as new variables of integration,

$$E(x + \lambda y)^2 = \frac{1}{\pi\Delta}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[(\delta - \lambda\gamma)^2u^2 + 2(\delta - \lambda\gamma)(-\beta + \lambda\alpha)uv +$$
$$+ (\beta - \lambda\alpha)^2v^2]e^{-u^2-v^2}dudv =$$
$$= \frac{1}{\pi\Delta}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[(\delta - \lambda\gamma)^2 + (\beta - \lambda\alpha)^2]u^2e^{-u^2-v^2}dudv =$$
$$= \frac{\delta^2 + \beta^2}{2\Delta} - 2\lambda\frac{\alpha\beta + \gamma\delta}{2\Delta} + \lambda^2\frac{\gamma^2 + \alpha^2}{2\Delta}.$$

But

$$\delta^2 + \beta^2 = c, \qquad \gamma^2 + \alpha^2 = a, \qquad \alpha\beta + \gamma\delta = b,$$

whence

$$E(x^2) + 2\lambda E(xy) + \lambda^2E(y^2) = \frac{c}{2\Delta} - 2\lambda\frac{b}{2\Delta} + \lambda^2\frac{a}{2\Delta},$$

and since λ is arbitrary

$$E(x^2) = \frac{c}{2\Delta}, \qquad E(xy) = -\frac{b}{2\Delta}, \qquad E(y^2) = \frac{a}{2\Delta}.$$

On the other hand, if σ_1, σ_2, and r are respectively standard deviations of x, y and their correlation coefficient, we have

$$E(x^2) = \sigma_1^2, \qquad E(xy) = r\sigma_1\sigma_2, \qquad E(y^2) = \sigma_2^2.$$

Hence

$$\frac{c}{2\Delta} = \sigma_1^2, \qquad \frac{a}{2\Delta} = \sigma_2^2, \qquad \frac{b}{2\Delta} = -r\sigma_1\sigma_2$$

and

$$\frac{ac - b^2}{4\Delta^2} = \sigma_1^2\sigma_2^2(1 - r^2)$$

or

$$2\Delta = \frac{1}{2\sigma_1^2\sigma_2^2(1 - r^2)}.$$

Finally,

$$a = \frac{1}{2\sigma_1^2(1 - r^2)}, \qquad b = -\frac{r}{2\sigma_1\sigma_2(1 - r^2)}, \qquad c = \frac{1}{2\sigma_2^2(1 - r^2)}$$

$$\sqrt{\Delta} = \frac{1}{2\sigma_1\sigma_2\sqrt{1 - r^2}}.$$

With these values for a, b, c, and $\sqrt{\Delta}$ the density of probability can be presented as follows:

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - r^2}} e^{-\frac{1}{2(1-r^2)}\left[\left(\frac{x}{\sigma_1}\right)^2 - 2r\frac{x}{\sigma_1}\frac{y}{\sigma_2} + \left(\frac{y}{\sigma_2}\right)^2\right]}$$

and the probability for a point x, y to belong to a given domain D will be expressed by the double integral

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - r^2}} \int\int_{(D)} e^{-\frac{1}{2(1-r^2)}\left[\left(\frac{x}{\sigma_1}\right)^2 - 2r\frac{x}{\sigma_1}\frac{y}{\sigma_2} + \left(\frac{y}{\sigma_2}\right)^2\right]} dxdy$$

extended over D.

4. Curves

$$\frac{1}{2(1 - r^2)}\left[\left(\frac{x}{\sigma_1}\right)^2 - 2r\frac{x}{\sigma_1}\frac{y}{\sigma_2} + \left(\frac{y}{\sigma_2}\right)^2\right] = l = \text{const.}$$

are evidently similar and similarly placed ellipses with the common center at the origin. For obvious reasons they are called ellipses of equal probability. The area of an ellipse corresponding to a given value of l (ellipse l) is

$$\frac{l\pi}{\sqrt{\Delta}} = 2\pi l\sigma_1\sigma_2\sqrt{1 - r^2}.$$

whence the area of an infinitesimal ring between ellipses l and $l + dl$ has the expression

$$2\pi\sigma_1\sigma_2\sqrt{1 - r^2}dl.$$

The infinitesimal probability for a point x, y to lie in that ring is expressed by

$$e^{-l}dl.$$

Finally, by integrating this expression between limits l_1 and $l_2 > l_1$, we find

$$e^{-l_1} - e^{-l_2}$$

as the expression of the probability for x, y to belong to the ring between two ellipses l_1 and l_2. If $l_1 = 0$ and $l_2 = l$,

$$1 - e^{-l}$$

gives the probability for x, y to belong to the ellipse l.

If n numbers $l, l_1, l_2, \ldots l_{n-1}$ are determined by the conditions

$$1 - e^{-l} = e^{-l} - e^{-l_1} = e^{-l_1} - e^{-l_2} = \cdots = e^{-l_{n-2}} - e^{-l_{n-1}} = \frac{1}{n+1},$$

the whole plane is divided into $n + 1$ regions of equal probability: namely, the interior of the ellipse l, rings between $l, l_1; l_1, l_2; \ldots l_{n-2}, l_{n-1}$ and, finally, part of the plane outside of the ellipse l_{n-1}.

5. To find the distribution function of the variable x (without any regard to y), we must take for D the domain

$$-\infty < x < t; \qquad -\infty < y < +\infty.$$

As the integral

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\int_{-\infty}^{t}\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-r^2)}\left[\left(\frac{y}{\sigma_2}-\frac{rx}{\sigma_1}\right)^2+(1-r^2)\left(\frac{x}{\sigma_1}\right)^2\right]}dxdy =$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-r^2}}\int_{-\infty}^{t} e^{-\frac{x^2}{2\sigma_1^2}}dx \cdot \int_{-\infty}^{\infty} e^{-\frac{z^2}{2(1-r^2)}}dz = \frac{1}{\sigma_1\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{x^2}{2\sigma_1^2}}dx,$$

we see that the probability of the inequality

$$x < t$$

is expressed by

$$\frac{1}{\sigma_1\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{x^2}{2\sigma_1^2}}dx.$$

Similarly, the probability of the inequality

$$y < t$$

is

$$\frac{1}{\sigma_2\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{y^2}{2\sigma_2^2}}dy.$$

Thus, if two variables x, y are normally distributed with their means $= 0$, each one of them taken separately has a normal distribution of probability with the common mean 0 and the respective standard deviations σ_1 and σ_2. Variables x and y are not independent except when $r = 0$. For if they were independent the probability of the point x, y belonging to an infinitesimal rectangle

$$t < x < t + dt; \qquad \tau < y < \tau + d\tau$$

would be

$$\frac{1}{2\pi\sigma_1\sigma_2}e^{-\frac{t^2}{2\sigma_1^2}-\frac{\tau^2}{2\sigma_2^2}}dtd\tau,$$

whereas it is

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)}\left[\left(\frac{t}{\sigma_1}\right)^2-2r\frac{t}{\sigma_1}\cdot\frac{\tau}{\sigma_2}+\left(\frac{\tau}{\sigma_2}\right)^2\right]}dtd\tau,$$

and these expressions are different unless $r = 0$. Thus, except for $r = 0$, normally distributed variables are necessarily dependent in the sense of the theory of probability. Dependent variables are often called "correlated variables." In particular, variables are said to be in "normal correlation" when they are normally distributed.

6. The probability of simultaneous inequalities

$$X < x < X', \qquad y < t$$

is represented by the repeated integral

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\int_X^{X'} e^{-\frac{x^2}{2\sigma_1^2}}dx\int_{-\infty}^{t} e^{-\frac{1}{2\sigma_2^2(1-r^2)}\left[y-r\frac{\sigma_2}{\sigma_1}x\right]^2}dy$$

while

$$\frac{1}{\sigma_1\sqrt{2\pi}}\int_X^{X'} e^{-\frac{x^2}{2\sigma_1^2}}dx$$

is the probability that x will be contained between X and X'. Hence (Chap. XII, Sec. 10) the ratio

$$\frac{1}{\sigma_2\sqrt{2\pi(1-r^2)}}\frac{\int_X^{X'} e^{-\frac{x^2}{2\sigma_1^2}}dx\int_{-\infty}^{t} e^{-\frac{1}{2\sigma_2^2(1-r^2)}\left[y-r\frac{\sigma_2}{\sigma_1}x\right]^2}dy}{\int_X^{X'} e^{-\frac{x^2}{2\sigma_1^2}}dx}$$

can be considered as the probability of the inequality

$$y < t$$

it being known that x is contained between X and X'. Considering X' as variable and converging to X the above ratio evidently tends to the limit

$$\frac{1}{\sigma_2\sqrt{1-r^2}\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2\sigma_2^2(1-r^2)}\left[y-r\frac{\sigma_2}{\sigma_1}X\right]^2}dy$$

which can be considered as the distribution function of y when x has a fixed value X. Hence, y for $x = X$ has a normal distribution with the standard deviation

$$\sigma_2\sqrt{1-r^2}$$

and the mean

$$Y = r\frac{\sigma_2}{\sigma_1}X.$$

Interpreted geometrically, this equation represents the so-called "line of regression" of y on x.

In a similar way, we conclude that for $y = Y$ the distribution of x is normal with the standard deviation

$$\sigma_1\sqrt{1-r^2}$$

and the mean

$$X = r\frac{\sigma_1}{\sigma_2}Y.$$

This equation represents the line of regression of x on y.

LIMIT THEOREM FOR SUMS OF INDEPENDENT VECTORS

7. So far normal distribution in two dimensions has been considered abstractly without indication of its natural origin. One-dimensional normal distribution may be considered as a limiting case of probability distributions of sums of independent variables. In the same manner two-dimensional normal distribution or normal correlation appears as a limit of probability distributions of sums of independent *vectors*.

Two series of stochastic variables

$$x_1, x_2, \ldots x_n$$
$$y_1, y_2, \ldots y_n$$

define n stochastic *vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ so that x_i, y_i represent components of $\mathbf{v}_i$ on two fixed coordinate axes. If

$$E(x_i) = a_i \qquad E(y_i) = b_i,$$

the vector $\mathbf{a}_i$ with the components a_i, b_i is called the mean value of $\mathbf{v}_i$. Evidently the mean value of

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_n$$

is represented by the vector

$$\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_n$$

and that of $\mathbf{v} - \mathbf{a}$ is a vanishing vector. Without loss of generality we may assume at the outset that

$$E(x_i) = E(y_i) = 0; \qquad i = 1, 2, \ldots n,$$

in which case $E(\mathbf{v}) = 0$. Vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ are said to be independent if variables x_i, y_i are independent of the rest of the variables x_j, y_j where $j \neq i$.

In what follows we shall deal exclusively with independent vectors.

8. As before, let x_k, y_k be components of the vector

$$\mathbf{v}_k (k = 1, 2, \ldots n).$$

Then

$$X = x_1 + x_2 + \cdots + x_n$$
$$Y = y_1 + y_2 + \cdots + y_n$$

will be the components of the sum

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_n.$$

If

$$E(x_k) = E(y_k) = 0$$
$$E(x_k^2) = b_k, \qquad E(y_k^2) = c_k, \qquad E(x_k y_k) = d_k$$

then

$$E(X) = 0, \qquad E(Y) = 0$$
$$E(X^2) = b_1 + b_2 + \cdots + b_n = B_n$$
$$E(Y^2) = c_1 + c_2 + \cdots + c_n = C_n$$
$$E(XY) = d_1 + d_2 + \cdots + d_n = r_n\sqrt{B_n}\sqrt{C_n}$$

because

$$E(x_i y_j) = 0 \quad \text{if} \quad j \neq i,$$

variables x_i and y_j being independent.

Let us introduce instead of variables x_k, $y_k (k = 1, 2, \ldots n)$ new variables

$$\xi_k = \frac{x_k}{\sqrt{B_n}}, \qquad \eta_k = \frac{y_k}{\sqrt{C_n}}$$

and correspondingly

$$s = \frac{X}{\sqrt{B_n}}, \qquad \sigma = \frac{Y}{\sqrt{C_n}}$$

instead of X, Y. We shall have:

$$E(\xi_k) = E(\eta_k) = 0$$
$$E(\xi_k^2) = \frac{b_k}{B_n}; \qquad E(\eta_k^2) = \frac{c_k}{C_n}$$

and

$$E(s) = E(\sigma) = 0$$
$$E(s^2) = E(\sigma^2) = 1$$
$$E(s\sigma) = r_n.$$

The quantity r_n, the correlation coefficient of s and σ, is in absolute value $\leqq 1$. We define

$$\phi(u, v) = E[e^{i(us+v\sigma)}]$$

as the characteristic function of the vector s, σ. Evidently $\phi(u, 0)$ and $\phi(0, v)$ are respectively the characteristic functions of s and σ. Since

$$e^{i(us+v\sigma)} = e^{i(u\xi_1+v\eta_1)} \cdot e^{i(u\xi_2+v\eta_2)} \cdot \cdot \cdot e^{i(u\xi_n+v\eta_n)}$$

and the factors in the right-hand member represent independent variables, we shall have

$$\phi(u, v) = E(e^{i(u\xi_1+v\eta_1)}) \cdot E(e^{i(u\xi_2+v\eta_2)}) \cdot \cdot \cdot E(e^{i(u\xi_n+v\eta_n)}).$$

9. For what follows it is very important to investigate the behavior of $\phi(u, v)$ when n increases indefinitely while u, v do not exceed an arbitrary but fixed number l in absolute value.

Let

$$E|x_k|^3 = f_k, \qquad E|y_k|^3 = g_k$$

and

$$\frac{f_1 + f_2 + \cdots + f_n}{B_n^{\frac{3}{2}}} = \omega_n$$
$$\frac{g_1 + g_2 + \cdots + g_n}{C_n^{\frac{3}{2}}} = \eta_n.$$

If ω_n and η_n tend to 0 as $n \to \infty$, we shall have

(1) $$|\phi(u, v) - e^{-\frac{1}{2}(u^2+2r_nuv+v^2)}| < e^{4l^3(\omega_n+\eta_n)} - 1$$

provided

$$|u| \leqq l, \qquad |v| \leqq l$$

and n is so large as to make

$$l(\omega_n^{\frac{1}{3}} + \eta_n^{\frac{1}{3}}) < 1.$$

Since

$$e^{i(u\xi_k+v\eta_k)} = 1 + i(u\xi_k + v\eta_k) - \frac{1}{2}(u\xi_k + v\eta_k)^2 + \\ + \theta \cdot \frac{(u\xi_k + v\eta_k)^3}{6}; \qquad |\theta| < 1,$$

we shall have:

$$E(e^{i(u\xi_k+v\eta_k)}) = 1 - \frac{b_k}{2B_n}u^2 - \frac{2d_k}{2\sqrt{B_nC_n}}uv - \frac{c_k}{2C_n}v^2 + \\ + \frac{\theta'}{6}E|u\xi_k + v\eta_k|^3; \qquad |\theta'| < 1.$$

On the other hand,

$$1 - \frac{b_k}{2B_n}u^2 - \frac{2d_k}{2\sqrt{B_nC_n}}uv - \frac{c_k}{2C_n}v^2 = e^{-\frac{b_k}{2B_n}u^2 - \frac{2d_k}{2\sqrt{B_nC_n}}uv - \frac{c_k}{2C_n}v^2} + \\ + \frac{\theta''}{8}[E(u\xi_k + v\eta_k)^2]^2; \qquad |\theta''| < 1$$

and so

$$E(e^{i(u\xi_k+v\eta_k)}) = e^{-\frac{b_k}{2B_n}u^2 - \frac{2d_k}{2\sqrt{B_nC_n}}uv - \frac{c_k}{2C_n}v^2} + \frac{\theta''}{8}[E(u\xi_k + v\eta_k)^2]^2 + \\ + \frac{\theta'}{6}E|u\xi_k + v\eta_k|^3.$$

Furthermore,

$$E(u\xi_k + v\eta_k)^2 \leqq l^2(\omega_n^{\frac{2}{3}} + 2\omega_n^{\frac{1}{3}}\eta_n^{\frac{1}{3}} + \eta_n^{\frac{2}{3}}) < 1$$

because

$$E(\xi_k^2) = \frac{b_k}{B_n} < \omega_n^{\frac{2}{3}}, \qquad E(\eta_k^2) = \frac{c_k}{C_n} < \eta_n^{\frac{2}{3}}, \qquad E|\xi_k\eta_k| < \omega_n^{\frac{1}{3}}\eta_n^{\frac{1}{3}}.$$

Also

$$[E(u\xi_k + v\eta_k)^2]^2 < [E(u\xi_k + v\eta_k)^2]^{\frac{3}{2}} \leqq E|u\xi_k + v\eta_k|^3$$

$$E|u\xi_k + v\eta_k|^3 \leqq 4l^3\left(\frac{f_k}{B_n^{\frac{3}{2}}} + \frac{g_k}{C_n^{\frac{3}{2}}}\right).$$

Taking into account these various inequalities, we may write

$$E(e^{i(u\xi_k+v\eta_k)}) = e^{-\frac{b_k}{2B_n}u^2 - \frac{2d_k}{2\sqrt{B_nC_n}}uv - \frac{c_k}{2C_n}v^2}(1 + \sigma_k)$$

where

$$|\sigma_k| < \frac{7}{24}\cdot 4e^{\frac{1}{3}}l^3\left(\frac{f_k}{B_n^{\frac{3}{2}}} + \frac{g_k}{C_n^{\frac{3}{2}}}\right) < 4l^3\left(\frac{f_k}{B_n^{\frac{3}{2}}} + \frac{g_k}{C_n^{\frac{3}{2}}}\right).$$

Finally,

$$\phi(u, v) = e^{-\frac{1}{2}(u^2+2r_nuv+v^2)}(1+\sigma_1)(1+\sigma_2)\cdots(1+\sigma_n)$$

and

$$\left|\phi(u, v) - e^{-\frac{1}{2}(u^2+2r_nuv+v^2)}\right| < e^{|\sigma_1|+|\sigma_2|+\cdots+|\sigma_n|} - 1 < e^{4l^3(\omega_n+\eta_n)} - 1$$

as was stated.

10. Theorem. *Let P denote the probability of simultaneous inequalities*

$$t_0 \leqq s < t_1; \qquad \tau_0 \leqq \sigma < \tau_1.$$

Provided r_n remains less than a fixed number $\alpha < 1$ in absolute value and the above introduced quantities ω_n, η_n tend to 0 as $n \to \infty$, P can be expressed as

$$P = \frac{1}{2\pi\sqrt{1-r_n^2}}\int_{t_0}^{t_1}\int_{\tau_0}^{\tau_1} e^{-\frac{1}{2(1-r_n^2)}(t^2-2r_nt\tau+\tau^2)}dtd\tau + \Delta_n$$

where Δ_n tends to 0 uniformly in t_0, t_1; τ_0, τ_1.

If, in addition, r_n itself tends to the limit $r(|r| < 1)P$ will tend uniformly to

$$\frac{1}{2\pi\sqrt{1-r^2}}\int_{t_0}^{t_1}\int_{\tau_0}^{\tau_1} e^{-\frac{1}{2(1-r^2)}(t^2-2rt\tau+\tau^2)}dtd\tau.$$

Proof. *a.* In trying to extend Liapounoff's proof to the present case we introduce an auxiliary quantity Π defined as

$$\Pi = E\left(\frac{1}{h^2\pi}\int_{t_0}^{t_1} e^{-\left(\frac{u-s}{h}\right)^2}du \cdot \int_{\tau_0}^{\tau_1} e^{-\left(\frac{v-\sigma}{h}\right)^2}dv\right).$$

Using the inequality

$$\frac{1}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt < \frac{e^{-x^2}}{2} \qquad \text{for} \qquad x > 0,$$

one can easily derive the following inequalities:

$$(2)\quad \left|\frac{1}{h^2\pi}\int_{t_0}^{t_1} e^{-\left(\frac{u-s}{h}\right)^2}du \cdot \int_{\tau_0}^{\tau_1} e^{-\left(\frac{v-\sigma}{h}\right)^2}dv - 1\right| <$$
$$< \frac{1}{2}\left(e^{-\left(\frac{t_1-s}{h}\right)^2} + e^{-\left(\frac{t_0-s}{h}\right)^2} + e^{-\left(\frac{\tau_1-\sigma}{h}\right)^2} + e^{-\left(\frac{\tau_0-\sigma}{h}\right)^2}\right)$$

if

$$(3)\qquad t_0 \leqq s < t_1; \qquad \tau_0 \leqq \sigma < \tau_1,$$

and

$$(4)\quad \frac{1}{h^2\pi}\int_{t_0}^{t_1} e^{-\left(\frac{u-s}{h}\right)^2}du\cdot\int_{\tau_0}^{\tau_1} e^{-\left(\frac{v-\sigma}{h}\right)^2}dv < \frac{1}{2}\left(e^{-\left(\frac{t_1-s}{h}\right)^2} + e^{-\left(\frac{t_0-s}{h}\right)^2} + e^{-\left(\frac{\tau_1-\sigma}{h}\right)^2} + e^{-\left(\frac{\tau_0-\sigma}{h}\right)^2}\right)$$

if at least one of the inequalities (3) is not fulfilled. From the definition of Π, P and from (2) and (4) it follows that

$$|P - \Pi| < \tfrac{1}{2}E\left(e^{-\left(\frac{t_1-s}{h}\right)^2} + e^{-\left(\frac{t_0-s}{h}\right)^2} + e^{-\left(\frac{\tau_1-\sigma}{h}\right)^2} + e^{-\left(\frac{\tau_0-\sigma}{h}\right)^2}\right).$$

But referring to (1) and setting

$$e^{4l^3(\omega_n+\eta_n)} - 1 = \alpha_n(l)$$

we have by virtue of the developments in Chap. XIV, Sec. 3,

$$(5)\qquad |P - \Pi| < 2\alpha_n(l) + h\sqrt{2} + \frac{8}{\sqrt{\pi}}\,\frac{e^{-\left(\frac{hl}{2}\right)^2}}{hl}.$$

b. Replacing t_1, τ_1 by variable quantities t, τ and taking the second derivative of Π with respect to t and τ, we get

$$\frac{d^2\Pi}{dtd\tau} = E\left(\frac{1}{h^2\pi}e^{-\left(\frac{t-s}{h}\right)^2-\left(\frac{\tau-\sigma}{h}\right)^2}\right).$$

On the other hand

$$\frac{1}{\pi}e^{-\left(\frac{t-s}{h}\right)^2-\left(\frac{\tau-\sigma}{h}\right)^2} = \frac{h^2}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{h^2}{4}(u^2+v^2)}e^{-i(tu+\tau v)}e^{i(us+v\sigma)}dudv,$$

whence

$$(6)\qquad \frac{d^2\Pi}{dtd\tau} = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{h^2}{4}(u^2+v^2)}e^{-i(tu+\tau v)}\phi(u, v)dudv.$$

Here we substitute

$$\phi(u, v) = e^{-\frac{1}{2}(u^2+2r_nuv+v^2)} + g(u, v).$$

For all real u, v

$$|g(u, v)| \leqq 2.$$

If $|u| \leqq l$, $|v| \leqq l$, where l is an arbitrarily fixed number, and n is large enough, we have

$$|g(u, v)| \leqq \alpha_n(l).$$

Hence, the double integral

$$\frac{1}{4\pi^2}\int\int e^{-\frac{h^2}{4}(u^2+v^2)}e^{-i(tu+\tau v)}g(u, v)dudv$$

extended over the region outside of the square $|u| \leqq l$, $|v| \leqq l$ is less than

$$\frac{1}{2\pi^2}\underset{u^2+v^2\geqq l^2}{\int\int} e^{-\frac{h^2}{4}(u^2+v^2)}dudv = \frac{1}{\pi}\int_l^\infty e^{-\frac{h^2r^2}{4}}rdr < \frac{e^{-\frac{h^2l^2}{4}}}{h^2}$$

in absolute value. The same double integral extended over the square $|u| \leqq l$, $|v| \leqq l$ is less than

$$\frac{l^2}{\pi^2}\alpha_n(l)$$

in absolute value. Thus, referring to (6)

$$\frac{d^2\Pi}{dtd\tau} = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{h^2}{4}(u^2+v^2)-\frac{1}{2}(u^2+2r_nuv+v^2)}e^{-i(tu+\tau v)}dudv + R$$

and

$$|R| < \frac{l^2}{\pi^2}\alpha_n(l) + \frac{e^{-\frac{h^2l^2}{4}}}{h^2}.$$

Now

$$e^{-\frac{h^2}{4}(u^2+v^2)} = 1 - \frac{\lambda h^2}{4}(u^2+v^2); \qquad |\lambda| < 1$$

and

$$\frac{h^2}{16\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(u^2+2r_nuv+v^2)}(u^2+v^2)dudv = \frac{h^2}{4\pi(1-r_n^2)^{\frac{3}{2}}} < \frac{h^2}{4\pi(1-\alpha^2)^{\frac{3}{2}}}.$$

Hence

$$\frac{d^2\Pi}{dtd\tau} = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(u^2+2r_nuv+v^2)}e^{-i(tu+\tau v)}dudv + R'$$

and

$$|R'| < \frac{l^2}{\pi^2}\alpha_n(l) + \frac{e^{-\frac{(hl)^2}{4}}}{h^2} + \frac{h^2}{4\pi(1-\alpha^2)^{\frac{3}{2}}}.$$

By transformation to new variables

$$\xi = u + r_nv; \qquad \eta = v\sqrt{1-r_n^2}$$

the foregoing double integral becomes

$$\frac{1}{4\pi^2\sqrt{1-r_n^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\xi^2-\frac{1}{2}\eta^2}\cdot e^{it\xi+i\frac{\tau-tr_n}{\sqrt{1-r_n^2}}\eta}\,d\xi d\eta =$$
$$= \frac{1}{2\pi\sqrt{1-r_n^2}}e^{-\frac{1}{2(1-r_n^2)}(t^2-2r_nt\tau+\tau^2)},$$

so that finally

$$\frac{d^2\Pi}{dtd\tau} = \frac{1}{2\pi\sqrt{1-r_n^2}}e^{-\frac{1}{2(1-r_n^2)}(t^2-2r_nt\tau+\tau^2)} + R'.$$

Integrating this expression with respect to t and τ between limits t_0, t_1 and τ_0, τ_1, we get:

$$\Pi = \frac{1}{2\pi\sqrt{1-r_n^2}}\int_{t_0}^{t_1}\int_{\tau_0}^{\tau_1} e^{-\frac{1}{2(1-r_n^2)}(t^2-2r_nt\tau+\tau^2)}dtd\tau + \rho \tag{7}$$

where

$$|\rho| < (t_1-t_0)(\tau_1-\tau_0)\left[\frac{l^2}{\pi^2}\alpha_n(l) + \frac{e^{-\frac{(hl)^2}{4}}}{h^2} + \frac{h^2}{4\pi(1-\alpha^2)^{\frac{3}{2}}}\right]. \tag{8}$$

Hence combining inequality (5) with (7) and (8),

$$P = \frac{1}{2\pi\sqrt{1-r_n^2}}\int_{t_0}^{t_1}\int_{\tau_0}^{\tau_1} e^{-\frac{1}{2(1-r_n^2)}\{t^2-2r_nt\tau+\tau^2\}}dtd\tau + \Delta_n$$

where

$$|\Delta_n| < \left[2+\frac{l^2}{\pi^2}(t_1-t_0)(\tau_1-\tau_0)\right]\alpha_n(l) + \frac{e^{-\frac{(hl)^2}{4}}}{h^2}\left\{\frac{8}{\sqrt{\pi}}\frac{h}{l} + \right.$$
$$\left. + (t_1-t_0)(\tau_1-\tau_0)\right\} + h\sqrt{2} + \frac{(t_1-t_0)(\tau_1-\tau_0)h^2}{4\pi(1-\alpha^2)^{\frac{3}{2}}}.$$

Considering t_0, t_1; τ_0, τ_1 as variable and denoting an arbitrarily large number by L, we shall assume at first that the rectangle D

$$t_0 \leqq s \leqq t_1; \qquad \tau_0 \leqq \sigma \leqq \tau_1$$

is completely contained in the square Q:

$$|s| \leqq L, \qquad |\sigma| \leqq L.$$

Then, taking $h = l^{-\frac{1}{2}}$ we shall have

$$|\Delta_n| < \left(2+\frac{4L^2l^2}{\pi^2}\right)\alpha_n(l) + le^{-\frac{l}{4}}\left(\frac{8}{\sqrt{\pi}}l^{-\frac{3}{2}} + 4L^2\right) + \sqrt{2}l^{-\frac{1}{2}} + \frac{L^2l^{-1}}{\pi(1-\alpha^2)^{\frac{3}{2}}}.$$

Given an arbitrary positive number ϵ, we take l so large as to have

$$le^{-\frac{l}{4}}\left(\frac{8}{\sqrt{\pi}}l^{-\frac{3}{2}} + 4L^2\right) + \sqrt{2}l^{-\frac{1}{2}} + \frac{L^2l^{-1}}{\pi(1 - \alpha^2)^{\frac{3}{2}}} < \frac{1}{2}\epsilon.$$

After that, since $\alpha_n(l) \to 0$ as $n \to \infty$ (for a fixed l) we can find a number $n_0(\epsilon)$ so that

$$\alpha_n(l) < \frac{\epsilon}{2}\left(2 + \frac{4L^2l^2}{\pi}\right)^{-1}$$

for $n > n_0(\epsilon)$. Finally, we shall have

$$|\Delta_n| < \epsilon$$

as soon as $n > n_0(\epsilon)$; that is, Δ_n tends to 0 uniformly in any rectangle D contained in the square Q with an arbitrarily large side $2L$.

c. To prove that Δ_n tends to 0 uniformly no matter what are t_0, t_1; τ_0, τ_1 we observe that the integral

$$\frac{1}{2\pi\sqrt{1 - r_n^2}}\int\int e^{-\frac{1}{2(1-r_n^2)}(t^2-2r_nt\tau+\tau^2)}dtd\tau$$

extended over the area outside of Q becomes infinitesimal as $L \to \infty$. Accordingly, we take L so large as to make this integral $<\epsilon/2$ (no matter what n is) and in addition to have $L^{-1} < \epsilon/4$. The number L selected according to these requirements will be kept fixed.

Let D' represent that part of D which is inside Q, the remaining part or parts (if there are any) being D''. Let P' and P'' denote the probabilities that the point s, σ shall be contained in D' or D'', respectively. Also, let J' and J'' be the integrals

$$\frac{1}{2\pi\sqrt{1 - r_n^2}}\int\int e^{-\frac{1}{2(1-r_n^2)}(t^2-2r_nt\tau+\tau^2)}dtd\tau$$

extended over D' and D'', respectively. By what has been proved, given $\epsilon > 0$ a number $n_0(\epsilon)$ can be found so that

$$|P' - J'| < \epsilon$$

for $n > n_0(\epsilon)$. Now

$$P = P' + P''; \qquad J = J' + J'',$$

whence

$$|P - J| < \epsilon + P'' + J''$$

for $n > n_0(\epsilon)$. Since by Tshebysheff's lemma (Chap. X, Sec. 1) the probability of either one of the inequalities

$$|s| > L \qquad \text{or} \qquad |\sigma| > L$$

is less than $1/L$, we shall have

$$P'' < \frac{2}{L} < \frac{\epsilon}{2}.$$

Also,

$$J'' < \frac{\epsilon}{2},$$

whence

$$|P - J| < 2\epsilon$$

for $n > n_0(\epsilon)$; that is, the difference

$$P - \frac{1}{2\pi\sqrt{1 - r_n^2}} \int_{t_0}^{t_1} \int_{\tau_0}^{\tau_1} e^{-\frac{1}{2(1-r_n^2)}(t^2 - 2r_n t\tau + \tau^2)} dt d\tau$$

tends to 0 uniformly, no matter what t_0, t_1; τ_0, τ_1 are.

Finally, the last statement of the theorem appears as almost evident and does not require an elaborate proof.

11. The theorem just proved concerns the asymptotic behavior of the probability P of simultaneous inequalities

$$t_0 \leqq s < t_1; \qquad \tau_0 \leqq \sigma < \tau_1$$

which, due to the definition of s and σ, are equivalent to the inequalities

$$t_0\sqrt{B_n} \leqq x_1 + x_2 + \cdots + x_n < t_1\sqrt{B_n}$$
$$\tau_0\sqrt{C_n} \leqq y_1 + y_2 + \cdots + y_n < \tau_1\sqrt{C_n}.$$

From the geometrical standpoint the above domain of s, σ is a rectangle. But the theorem can be extended to the case of any given domain R for the point s, σ. It is hardly necessary to enter into details of the proof based on the definition of a double integral. It suffices to state the theorem itself:

Fundamental Theorem. *The probability for the point* (s, σ) *to be located in a given domain* R *can be represented, for large* n, *by the integral*

$$\frac{1}{2\pi\sqrt{1 - r_n^2}} \int\int e^{-\frac{1}{2(1-r_n^2)}(t^2 - 2r_n t\tau + \tau^2)} dt d\tau$$

extended over R, *with an error which tends uniformly to* 0 *as* n *becomes infinite, provided*

$$\omega_n \rightarrow 0, \qquad \eta_n \rightarrow 0,$$

while for all n

$$|r_n| < \alpha < 1.$$

In less precise terms we may say that under very general conditions the probability distribution of the components of a vector which is the sum of a great many independent vectors will be nearly normal.

The first rigorous proof of the limit theorem for sums of independent vectors was published by S. Bernstein in 1926. Like the proof developed here it proceeds on the same lines as Liapounoff's proof for sums of independent variables. Moreover, Bernstein has shown that the limit theorem may hold even in case of dependent vectors when certain additional conditions are fulfilled.

12. A good illustration of the fundamental theorem is afforded by series of independent trials with three alternatives, E, F, G. For the sake of simplicity we shall assume that probabilities of E, F, G are p, q, r in all trials. Naturally

$$p + q + r = 1.$$

In the usual way, we associate with these trials triads of variables

$$x_i, y_i, z_i \qquad (i = 1, 2, 3, \ldots)$$

so that

$x_i = 1$ or 0 according as E occurs or fails at the ith trial;
$y_i = 1$ or 0 according as F occurs or fails at the ith trial;
$z_i = 1$ or 0 according as G occurs or fails at the ith trial.

Evidently

$$E(x_i) = E(x_i^2) = p$$
$$E(y_i) = E(y_i^2) = q$$

so that vectors $\mathbf{v}_i$ with components

$$\xi_i = x_i - p, \qquad \eta_i = y_i - q$$

have their means $= 0$. The independence of trials involves the independence of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$. Hence we can apply the preceding considerations to the vector

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_n$$

with the components

$$X = \xi_1 + \xi_2 + \cdots + \xi_n$$
$$Y = \eta_1 + \eta_2 + \cdots + \eta_n.$$

We have

$$B_n = E(X^2) = np(1 - p); \qquad C_n = E(Y^2) = nq(1 - q).$$

Moreover,

$$E(\xi_i\eta_i) = E(x_iy_i) - pq = -pq$$

and

$$E(XY) = r_n\sqrt{B_n}\sqrt{C_n} = -npq$$

whence

$$r_n = -\frac{pq}{\sqrt{pq(1-p)(1-q)}}.$$

The quantities denoted by f_k, g_k in Sec. 9 are in our case

$$f_k = E|\xi_k|^3 = p(1-p)^3 + (1-p)p^3$$
$$g_k = E|\eta_k|^3 = q(1-q)^3 + (1-q)q^3.$$

Hence

$$\omega_n = \frac{p(1-p)^3 + (1-p)p^3}{n^{\frac{1}{2}}p^{\frac{3}{2}}(1-p)^{\frac{3}{2}}}, \qquad \eta_n = \frac{q(1-q)^3 + (1-q)q^3}{n^{\frac{1}{2}}q^{\frac{3}{2}}(1-q)^{\frac{3}{2}}},$$

and the conditions

$$\omega_n \to 0, \qquad \eta_n \to 0$$

are satisfied. The fundamental theorem, therefore, can be applied. If k, l, m are the respective frequencies of events E, F, G in n trials, the quantities X and Y represent the discrepancies

$$\lambda = k - np, \qquad \mu = l - nq.$$

Introducing the third discrepancy

$$\nu = m - nr$$

we shall have

$$\lambda + \mu + \nu = 0$$

so that ν is determined when λ and μ are given. The last two quantities, however, may have various values depending on chance. Concerning them the following statement follows from the fundamental theorem:

Theorem. *The probability that discrepancies λ, μ in n trials shall simultaneously satisfy the inequalities*

$$\alpha_0\sqrt{n} < \lambda < \alpha_1\sqrt{n}; \qquad \beta_0\sqrt{n} < \mu < \beta_1\sqrt{n}$$

tends uniformly, with indefinitely increasing n, to the limit

$$\frac{1}{2\pi\sqrt{pqr}}\int_{\alpha_0}^{\alpha_1}\int_{\beta_0}^{\beta_1} e^{-\frac{1}{2}\left(\frac{\alpha^2}{p}+\frac{\beta^2}{q}+\frac{\gamma^2}{r}\right)}d\alpha d\beta$$

where, to have symmetrical notation, γ is a variable defined by

$$\alpha + \beta + \gamma = 0.$$

On account of symmetry, perfectly similar statements can be made in regard to any two pairs of discrepancies λ, μ, ν.

Since the fundamental theorem and its proof can be extended without any difficulty to vectors of more than two dimensions, we shall have in the case of trials with more than three alternatives a result perfectly analogous to the last theorem.

Theorem. *Each of n independent trials admits of k alternatives E_1, E_2, . . . E_k the probabilities and the frequencies of which respectively are p_1, p_2, . . . p_k and m_1, m_2, . . . m_k. The probability that the discrepancies $m_i - np_i (i = 1, 2, \dots k - 1)$ should satisfy simultaneously the inequalities*

$$\alpha_i\sqrt{n} < m_i - np_i < \beta_i\sqrt{n}$$

tends uniformly, with indefinitely increasing n, to the limit

$$\frac{1}{(2\pi)^{\frac{k-1}{2}}\sqrt{p_1p_2\cdots p_k}}\int_{\alpha_1}^{\beta_1}\cdots\int_{\alpha_{k-1}}^{\beta_{k-1}} e^{-\frac{1}{2}\sum_1^k\frac{t_i^2}{p_i}}dt_1dt_2\cdots dt_{k-1}$$

where

$$t_k = -(t_1 + t_2 + \cdots + t_{k-1}).$$

From this theorem, by resorting to the definition of a multiple integral, we may deduce an important corollary: *Let P_n denote the probability of the inequality*

$$\frac{(m_1 - np_1)^2}{np_1} + \frac{(m_2 - np_2)^2}{np_2} + \cdots + \frac{(m_k - np_k)^2}{np_k} \leqq \chi^2.$$

Then, as n tends to infinity P_n tends to the limit

$$\frac{1}{(2\pi)^{\frac{k-1}{2}}\sqrt{p_1p_2\cdots p_k}}\int\cdots\int e^{-\frac{1}{2}\left(\frac{t_1^2}{p_1}+\frac{t_2^2}{p_2}+\cdots+\frac{t_k^2}{p_k}\right)}dt_1dt_2\cdots dt_{k-1}$$

where the integration is extended over the $(k - 1)$ dimensional ellipsoid

$$\varphi = \frac{t_1^2}{p_1} + \frac{t_2^2}{p_2} + \cdots + \frac{t_k^2}{p_k} \leqq \chi^2.$$

It is easy to see that the determinant of the quadratic form φ in $(k - 1)$ variables is $(p_1p_2 \cdots p_k)^{-1}$. Hence, by a proper linear transformation the above integral reduces to

$$\frac{1}{(2\pi)^{\frac{k-1}{2}}}\int\int\cdots\int e^{-\frac{1}{2}(v_1^2+v_2^2+\cdots+v_{k-1}^2)}dv_1dv_2\cdots dv_{k-1}$$

the domain of integration being $v_1^2 + v_2^2 + \cdots + v_{k-1}^2 \leqq \chi^2$. But this multiple integral, as will be shown in Chap. XVI, Sec. 1, can be reduced to a simple integral

$$\frac{2\pi^{\frac{k-1}{2}}}{\Gamma\left(\frac{k-1}{2}\right)}\int_0^{\chi} e^{-\frac{1}{2}u^2}u^{k-2}du.$$

Thus

$$\lim P_n = \frac{1}{2^{\frac{k-3}{2}}\Gamma\left(\frac{k-1}{2}\right)}\int_0^{\chi} e^{-\frac{1}{2}u^2}u^{k-2}du.$$

The probability $Q_n = 1 - P_n$ of the opposite inequality

$$(A) \qquad \frac{(m_1 - np_1)^2}{np_1} + \frac{(m_2 - np_2)^2}{np_2} + \cdots + \frac{(m_k - np_k)^2}{np_k} > \chi^2$$

tends to the limit

$$\frac{1}{2^{\frac{k-3}{2}}\Gamma\left(\frac{k-1}{2}\right)}\int_{\chi}^{\infty} e^{-\frac{1}{2}u^2}u^{k-2}du$$

and for large n we have an approximate formula

$$Q_n = \frac{1}{2^{\frac{k-3}{2}}\Gamma\left(\frac{k-1}{2}\right)}\int_{\chi}^{\infty} e^{-\frac{1}{2}u^2}u^{k-2}du,$$

but the degree of approximation remains unknown. In practice, to test whether the observed deviations of frequencies from their expected values are significant, the value of the sum (A), say χ^2, is found; then by the above approximate formula the probability that the sum (A) will be greater than χ^2 is computed. If this probability is very small, then the obtained system of deviations is significantly different from what could be expected as a result of chance alone. The lack of information as to the error incurred by using an approximate expression of Q_n renders the application of this "χ^2-test" devised by Pearson somewhat dubious.

Hypothetical Explanation of Empirically Verified Cases of Normal Correlation

13. Normal distribution in two dimensions plays an important part in target practice. It is generally assumed on the basis of varied evidence collected in actual target practice that points of a target hit by projectiles are scattered in a manner suggesting normal distribution. By referring points hit by projectiles to a fixed coordinate system on the target, it is possible from their coordinates to find approximately (provided the number of shots is large) the elements of ellipses of equal probability. Dividing the surface of the target into regions of equal probabilities as described in Sec. 4, and counting the actual number of hits in each region, the resulting numbers in many reported instances are nearly equal. That and the agreement with other criteria are generally considered as evidence in favor of assuming the probability in target practice to be normally distributed.

Two-dimensional normal distribution or normal correlation has been found to exist between measurable attributes, such as the length of the body and weight of living organisms. Attributes like statures of parents and their descendants, according to Galton, again show evidence of normal correlation.

Facing such a variety of facts pointing to the existence of normal correlation, one is tempted to account for it by some more or less plausible hypothesis. It is generally assumed that deviations of two magnitudes from their mean values are caused by the combined action of a great many independent causes, each affecting both magnitudes in a very small degree. Clearly, the resulting deviations under such circumstances may be regarded as components of the sum of a great many independent vectors. Then, to explain the existence of normal correlation, reference is made to the fundamental theorem in Sec. 11.

Problems for Solution

1. Let p denote the probability that two normally distributed variables (with means $= 0$) will have values of opposite signs. Show that between p and the correlation coefficient r the following relation holds:

$$r = \cos p\pi.$$

2. Variables x, y (with the means $= 0$) are normally distributed. Show that the probability for the point x, y to be located in an ellipse

$$\frac{x^2}{\sigma_1^2} - 2r\frac{x}{\sigma_1}\frac{y}{\sigma_2} + \frac{y^2}{\sigma_2^2} = l$$

is greater than the probability corresponding to any other domain of the same area.

3. Three dice colored in white, red, and blue are tossed simultaneously n times. Let X and Y represent the total number of points on pairs: white, red and white, blue Show that the probability of simultaneous inequalities

$$7n + t_0\sqrt{\tfrac{35}{6}n} < X < 7n + t_1\sqrt{\tfrac{35}{6}n}; \qquad 7n + \tau_0\sqrt{\tfrac{35}{6}n} < Y < 7n + \tau_1\sqrt{\tfrac{35}{6}n}$$

tends to the limit

$$\frac{1}{\pi\sqrt{3}}\int_{t_0}^{t_1}\int_{\tau_0}^{\tau_1} e^{-\frac{2}{3}(t^2 - t\tau + \tau^2)}\,dt\,d\tau$$

as $n \to \infty$.

4. Three dice, white, red, and blue, are tossed simultaneously n times. If k and l are frequencies of 10 points on pairs: white, red; red, blue; show that the probability of simultaneous inequalities

$$\frac{n}{12}+t_0\sqrt{\frac{11}{144}n}<k<\frac{n}{12}+t_1\sqrt{\frac{11}{144}n};\qquad \frac{n}{12}+\tau_0\sqrt{\frac{11}{144}n}<l<\frac{n}{12}+\tau_1\sqrt{\frac{11}{144}n}$$

tends to the limit

$$\frac{11}{2\pi\sqrt{120}}\int_{t_0}^{t_1}\int_{\tau_0}^{\tau_1}e^{-\frac{121}{240}\left(t^2-\frac{2}{11}t\tau+\tau^2\right)}dt\,d\tau$$

as $n\to\infty$.

5. Two players, A and B, take part in a game arranged as follows: Each time one ball is taken from an urn containing 8 white, 6 black, and 1 red ball; if this ball is

white, A and B both gain \$1;
black, A loses \$2, B loses \$4;
red, A gains \$4, B gains \$16.

Let s_n and σ_n be the sums gained by A and B after n games. Show that the probability of simultaneous inequalities

$$t_0\sqrt{\tfrac{32}{5}n}<s_n<t_1\sqrt{\tfrac{32}{5}n};\qquad \tau_0\sqrt{48n}<\sigma_n<\tau_1\sqrt{48n}$$

for very large n will be approximately equal to

$$\frac{\sqrt{6}}{\pi}\int_{t_0}^{t_1}\int_{\tau_0}^{\tau_1}e^{-6\left(t^2+\tau^2-\sqrt{\frac{10}{3}}t\tau\right)}dt\,d\tau.$$

Note that the probability of the inequality $s_n\sigma_n<0$ is about 0.13—not very small—so that it is not very unlikely that the luck will be with one player and against another.

6. Concentric circles C_1, C_2, C_3, . . . in unlimited numbers are described about the origin O. Points P_1, P_2, P_3, . . . are taken at random on these circles. Let R be the end point of the vector representing the sum of vectors OP_1, OP_2, OP_3, ... If r_1, r_2, r_3, . . . are radii of C_1, C_2, C_3, . . . and the condition

$$\frac{r_1^3+r_2^3+\cdots+r_n^3}{(r_1^2+r_2^2+\cdots+r_n^2)^{\frac{3}{2}}}\to 0\qquad\text{as}\qquad n\to\infty$$

is fulfilled, show that the probability that R will lie within the circle described with the radius ρ about the origin will be very nearly equal to

$$1-e^{-\frac{\rho^2}{r_1^2+r_2^2+\cdots+r_n^2}}$$

for large n.

References

A. Bravais: Analyse mathématique sur les probabilités des erreurs de situation d'un point, *Mém. présentés par div. sav. à l'Acad. Roy. Sci. l'Inst. France*, 9, 1846.

Ch. M. Schols: Théorie des erreurs dans le plan et dans l'espace, *Ann. l'École polyt. Delft*, 2, 1886.

K. Pearson: Mathematical Contributions to the Theory of Evolution, *Philos. Trans.*, 187*A*, 1896.

S. BERNSTEIN: Sur l'extension du théorème limite du calcul des probabilités aux sommes des quantités dépendantes, *Math. Ann.*, 97, 1927.

A. KHINTCHINE: Begründung der Normalkorrelation nach der Lindebergschen Methode, *Nachr. Ver. Forschungsinstitute Univ. Moskau*, 1, 1928.

G. CASTELNUOVO: "Calcolo delle probabilità," vol. 2, Bologna, 1928.

V. ROMANOVSKY: Sur une extension du théorème de Liapounoff sur la limite de probabilité, *Bull. l'Acad. Sci. U.S.S.R.*, No. 2, 1929.

CHAPTER XVI

DISTRIBUTION OF CERTAIN FUNCTIONS OF NORMALLY DISTRIBUTED VARIABLES

1. In modern statistics much emphasis is laid upon distributions of certain functions involving normally distributed variables. Such distributions are considered as a basis for various "tests of significance" for small samples, that is, when the number of observed data is small. Some of the most important cases of this kind will be considered in this chapter.

Problem 1. Independent variables $x_1, x_2, \ldots x_n$ are normally distributed about their common mean $= 0$ with the same standard deviation σ. Find the distribution function of the sum of their squares

$$s = x_1^2 + x_2^2 + \cdots + x_n^2.$$

Solution. The inequality

$$x_i^2 < t$$

being equivalent to

$$-\sqrt{t} < x_i < \sqrt{t},$$

the distribution function of x_i^2 is

$$F_i(t) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\sqrt{t}}^{\sqrt{t}} e^{-\frac{x^2}{2\sigma^2}}dx = \frac{1}{\sigma\sqrt{2\pi}}\int_0^t e^{-\frac{u}{2\sigma^2}}u^{-\frac{1}{2}}du \qquad \text{for} \qquad t \geqq 0$$

$$F_i(t) = 0 \qquad \text{for} \qquad t < 0.$$

Hence, the characteristic function of any one of the variables $x_1^2, x_2^2, \ldots x_n^2$ is

$$\varphi_k(t) = \frac{1}{\sigma\sqrt{2\pi}}\int_0^\infty e^{-\left(\frac{1}{2\sigma^2}-it\right)u}u^{-\frac{1}{2}}du = \frac{1}{\sigma\sqrt{2}}\left(\frac{1}{2\sigma^2} - it\right)^{-\frac{1}{2}}$$

and that of their sum

$$\varphi(t) = \frac{1}{(\sigma\sqrt{2})^n}\left(\frac{1}{2\sigma^2} - it\right)^{-\frac{n}{2}}.$$

Consequently, the distribution function of s is expressed by

$$F(t) = C + \frac{(\sigma\sqrt{2})^{-n}}{2\pi}\int_{-\infty}^{\infty}\frac{1 - e^{-itv}}{iv\left(\frac{1}{2\sigma^2} - iv\right)^{\frac{n}{2}}}dv$$

and it remains to transform this integral. To this end, imagine a variable distributed over the interval $(0, +\infty)$ with the density

$$\frac{2^{-\frac{n}{2}}\sigma^{-n}}{\Gamma\left(\frac{n}{2}\right)}e^{-\frac{u}{2\sigma^2}}u^{\frac{n}{2}-1}$$

Its characteristic function is

$$(\sigma\sqrt{2})^{-n}\left(\frac{1}{2\sigma^2} - it\right)^{-\frac{n}{2}}$$

and since the distribution function is given a priori, we must have for $t \geqq 0$

$$\frac{(\sigma\sqrt{2})^{-n}}{\Gamma\left(\frac{n}{2}\right)}\int_0^t e^{-\frac{u}{2\sigma^2}}u^{\frac{n}{2}-1}du = \text{const.} + \frac{(\sigma\sqrt{2})^{-n}}{2\pi}\int_{-\infty}^{\infty}\frac{1 - e^{-itv}}{iv\left(\frac{1}{2\sigma^2} - iv\right)^{\frac{n}{2}}}dv.$$

Hence

$$F(t) = \text{const.} + \frac{1}{(\sigma\sqrt{2})^n\Gamma\left(\frac{n}{2}\right)}\int_0^t e^{-\frac{u}{2\sigma^2}}u^{\frac{n}{2}-1}du.$$

The constant must be $= 0$ since $F(t)$ as well as the integral in the right member vanishes for $t = 0$. The final expression is therefore:

$$F(t) = \frac{1}{(\sigma\sqrt{2})^n\Gamma\left(\frac{n}{2}\right)}\int_0^t e^{-\frac{u}{2\sigma^2}}u^{\frac{n}{2}-1}du \qquad \text{for} \qquad t \geqq 0$$

$$F(t) = 0 \qquad \text{for} \qquad t \leqq 0.$$

The probability of the inequality

$$x_1^2 + x_2^2 + \cdots + x_n^2 < t,$$

on the other hand, can be expressed directly as a multiple integral

$$F(t) = \frac{1}{(\sigma\sqrt{2\pi})^n}\int\int\cdots\int e^{-\frac{x_1^2+x_2^2+\cdots+x_n^2}{2\sigma^2}}dx_1dx_2\cdots dx_n$$

extended over the volume of the n-dimensional sphere S

$$x_1^2 + x_2^2 + \cdots + x_n^2 < t.$$

By equating both expressions of $F(t)$, we obtain an important transformation,

$$(1)\quad \int\int\cdots\int e^{-\frac{x_1^2+x_2^2+\cdots+x_n^2}{2\sigma^2}}dx_1dx_2\cdots dx_n =$$

$$= \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\int_0^t e^{-\frac{u}{2\sigma^2}}u^{\frac{n}{2}-1}du.$$

If $F(x_1^2 + x_2^2 + \cdots + x_n^2)$ is an arbitrary function of

$$u = x_1^2 + x_2^2 + \cdots + x_n^2,$$

the integral

$$\frac{1}{(\sigma\sqrt{2\pi})^n}\int\int\cdots\int e^{-\frac{x_1^2+x_2^2+\cdots+x_n^2}{2\sigma^2}}F(x_1^2 + \cdots + x_n^2)dx_1dx_2\cdots dx_n$$

extended over the whole n-dimensional space represents the mathematical expectation of $F(u)$. On the other hand, the distribution function of u being known the same multiple integral will be equal to

$$\frac{1}{(\sigma\sqrt{2})^n\Gamma\left(\frac{n}{2}\right)}\int_0^\infty e^{-\frac{u}{2\sigma^2}}F(u)u^{\frac{n-2}{2}}du.$$

Taking in particular $\sigma = 1$, $F(u) = e^{au^{\frac{1}{2}}}$, we get the formula

$$(2)\quad \int\int\cdots\int e^{-\frac{1}{2}(x_1^2+\cdots+x_n^2)+a\sqrt{x_1^2+\cdots+x_n^2}}dx_1dx_2\cdots dx_n =$$

$$= \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\int_0^\infty e^{-\frac{u}{2}+au^{\frac{1}{2}}}u^{\frac{n-2}{2}}du,$$

which will be used later.

2. Problem 2. Variables $x_1, x_2, \ldots x_n$ are defined as in Prob. 1. Denoting their arithmetic mean by

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

find the distribution function of the sum

$$\Sigma = (x_1 - s)^2 + (x_2 - s)^2 + \cdots + (x_n - s)^2.$$

Solution. The probability of the inequality

$$\Sigma < t$$

is expressed by the multiple integral

$$F(t) = \frac{1}{(\sigma\sqrt{2\pi})^n}\int\int\cdots\int e^{-\frac{x_1^2+x_2^2+\cdots+x_n^2}{2\sigma^2}}dx_1dx_2\cdots dx_n$$

extended over the volume of the n-dimensional ellipsoid

$$(x_1 - s)^2 + (x_2 - s)^2 + \cdots + (x_n - s)^2 < t.$$

Let

$$x_1 - s = u_1, \qquad x_2 - s = u_2, \cdots x_n - s = u_n,$$

whence

$$u_1 + u_2 + \cdots + u_n = 0$$

and

$$x_1^2 + x_2^2 + \cdots + x_n^2 = u_1^2 + u_2^2 + \cdots + u_n^2 + ns^2.$$

Taking $u_1, u_2, \ldots u_{n-1}$, and s for new variables, we must first find the Jacobian J of $x_1, x_2, \ldots x_n$ with respect to $u_1, u_2, \ldots u_{n-1}, s$. It is

$$J = \begin{vmatrix} 1 & 1 & 0 & 0 \cdots & 0 \\ 1 & 0 & 1 & 0 \cdots & 0 \\ 1 & 0 & 0 & 1 \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 0 & 0 \cdots & 1 \\ 1 & -1 & -1 & -1 \cdots & -1 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 0 & 0 \cdots & 0 \\ 1 & 0 & 1 & 0 \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 0 & 0 \cdots & 1 \\ n & 0 & 0 & 0 \cdots & 0 \end{vmatrix} = (-1)^{n-1}n.$$

In the new variables the expression for $F(t)$ will be

$$F(t) = \frac{n}{(\sigma\sqrt{2\pi})^n}\int\int\cdots\int e^{-\frac{ns^2}{2\sigma^2}-\frac{u_1^2+u_2^2+\cdots+u_n^2}{2\sigma^2}}dsdu_1du_2\cdots du_{n-1}$$

and the domain of integration in the space of the new variables is defined by

$$-\infty < s < \infty$$
$$u_1^2 + u_2^2 + \cdots + u_{n-1}^2 + (u_1 + u_2 + \cdots + u_{n-1})^2 < t.$$

After performing the integration with respect to s, we get

$$F(t) = \frac{\sqrt{n}}{(\sigma\sqrt{2\pi})^{n-1}}\int\int\cdots\int e^{-\frac{u_1^2+u_2^2+\cdots+u_n^2}{2\sigma^2}}du_1du_2\cdots du_{n-1}.$$

The quadratic form

$$\varphi = u_1^2 + u_2^2 + \cdots + u_{n-1}^2 + (u_1 + u_2 + \cdots + u_{n-1})^2$$

can be represented as a sum of the squares of $(n - 1)$ linear forms in variables $u_1, u_2, \ldots u_{n-1}$:

$$\varphi = v_1^2 + v_2^2 + \cdots + v_{n-1}^2.$$

The Jacobian

$$\frac{\partial(v_1, v_2, \ldots v_{n-1})}{\partial(u_1, u_2, \ldots u_{n-1})}$$

is the square root of the determinant of the form φ, which is the same as the determinant of linear forms

$$\frac{1}{2}\frac{\partial\varphi}{\partial u_1} = 2u_1 + u_2 + \cdots + u_{n-1}$$
$$\frac{1}{2}\frac{\partial\varphi}{\partial u_2} = u_1 + 2u_2 + \cdots + u_{n-1}$$
$$\cdots\cdots$$
$$\frac{1}{2}\frac{\partial\varphi}{\partial u_{n-1}} = u_1 + u_2 + \cdots + 2u_{n-1}.$$

Now, in general

$$\overbrace{\begin{vmatrix} \lambda 1 1 & \cdots & 1 \\ 1 \lambda 1 & \cdots & 1 \\ \cdot & \cdot & \cdot \\ 1 1 1 & \cdots & \lambda \end{vmatrix}}^{p \text{ times}} = (\lambda - 1)^{p-1}(\lambda + p - 1)$$

so that the determinant of φ is $= n$, whence

$$\frac{\partial(v_1, v_2, \cdots v_{n-1})}{\partial(u_1, u_2, \cdots u_{n-1})} = \sqrt{n}$$

and

$$\frac{\partial(u_1, u_2, \cdots u_{n-1})}{\partial(v_1, v_2, \cdots v_{n-1})} = \frac{1}{\sqrt{n}}.$$

Therefore, taking $v_1, v_2, \ldots v_{n-1}$ for new variables, $F(t)$ can be expressed as follows

$$F(t) = \frac{1}{(\sigma\sqrt{2\pi})^{n-1}}\int\int\cdots\int e^{-\frac{v_1^2+v_2^2+\cdots+v_{n-1}^2}{2\sigma^2}}dv_1dv_2\cdots dv_{n-1}$$

where the integral is extended over the volume of the sphere

$$v_1^2 + v_2^2 + \cdots + v_{n-1}^2 < t.$$

This multiple integral is exactly of the type considered in the preceding problem, and it can be reduced to a simple integral as follows

$$\int\int\cdots\int e^{-\frac{v_1^2+v_2^2+\cdots+v_{n-1}^2}{2\sigma^2}}dv_1dv_2\cdots dv_{n-1} =$$
$$= \frac{\pi^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}\int_0^t e^{-\frac{u}{2\sigma^2}}u^{\frac{n-3}{2}}du.$$

After substitution, the final expression of $F(t)$ is

$$F(t) = \frac{1}{(\sigma\sqrt{2})^{n-1}\Gamma\left(\frac{n-1}{2}\right)}\int_0^t e^{-\frac{u}{2\sigma^2}}u^{\frac{n-3}{2}}du \qquad \text{for} \qquad t > 0$$

$$F(t) = 0 \qquad \text{for} \qquad t \leqq 0.$$

3. Problem 3. Variables $x_1, x_2, \ldots x_n$ are defined as in Prob. 1. As in Prob. 2, we set

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

$$u_i = x_i - s; \qquad i = 1, 2, \ldots n$$

and introduce the quantity

$$\epsilon = \sqrt{\frac{u_1^2 + u_2^2 + \cdots + u_n^2}{n}}.$$

What is the distribution function of the ratio

$$\frac{s}{\epsilon}$$

or, which is the same, the probability $F(t)$ of the inequality

$$s < t\epsilon?$$

Solution. First, assuming t to be positive, let us find the probability $\phi(t)$ of the inequality

$$s \geqq t\epsilon$$

or

$$u_1^2 + u_2^2 + \cdots + u_n^2 \leqq \frac{ns^2}{t^2}.$$

This probability can be presented in the form

$$\phi(t) = \frac{n}{(\sigma\sqrt{2\pi})^n}\int_0^\infty e^{-\frac{ns^2}{2\sigma^2}}\Psi(s)ds$$

where the multiple integral

$$\Psi(s) = \int\int \cdots \int e^{-\frac{u_1^2+u_2^2+\cdots+u_n^2}{2\sigma^2}}du_1du_2 \cdots du_{n-1}$$

in which

$$u_n = -(u_1 + u_2 + \cdots + u_{n-1})$$

is extended over the domain

$$u_1^2 + u_2^2 + \cdots + u_{n-1}^2 + (u_1 + u_2 + \cdots + u_{n-1})^2 \leqq \frac{ns^2}{t^2}.$$

Proceeding in exactly the same manner as in Prob. 2, we can transform $\Psi(s)$ into

$$\Psi(s) = \frac{1}{\sqrt{n}} \int\int \cdots \int e^{-\frac{v_1^2 + v_2^2 + \cdots + v_{n-1}^2}{2\sigma^2}} dv_1 dv_2 \cdots dv_{n-1}$$

extended over the sphere

$$v_1^2 + v_2^2 + \cdots + v_{n-1}^2 \leqq \frac{ns^2}{t^2}$$

in the space of the variables $v_1, v_2, \ldots v_{n-1}$. For this multiple integral we can substitute a simple integral

$$\frac{\pi^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \int_0^{\frac{ns^2}{t^2}} e^{-\frac{u}{2\sigma^2}} u^{\frac{n-3}{2}} du = \frac{2\pi^{\frac{n-1}{2}} n^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \int_0^{\frac{s}{t}} e^{-\frac{n\xi^2}{2\sigma^2}} \xi^{n-2} d\xi$$

and thus reduce $\Psi(s)$ to the form

$$\Psi(s) = \frac{2\pi^{\frac{n-1}{2}} n^{\frac{n-2}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \int_0^{\frac{s}{t}} e^{-\frac{n\xi^2}{2\sigma^2}} \xi^{n-2} d\xi.$$

After substitution we can express $\phi(t)$ as a repeated integral

$$\phi(t) = \frac{2n^{\frac{n}{2}}}{\sqrt{\pi}(\sigma\sqrt{2})^n \Gamma\left(\frac{n-1}{2}\right)} \int_0^{\infty} e^{-\frac{ns^2}{2\sigma^2}} ds \int_0^{\frac{s}{t}} e^{-\frac{n\xi^2}{2\sigma^2}} \xi^{n-2} d\xi.$$

The derivative of $\phi(t)$ is

$$\phi'(t) = -\frac{2n^{\frac{n}{2}} t^{-n}}{\sqrt{\pi}(\sigma\sqrt{2})^n \Gamma\left(\frac{n-1}{2}\right)} \int_0^{\infty} e^{-\frac{ns^2}{2\sigma^2}\left(1+\frac{1}{t^2}\right)} s^{n-1} ds =$$

$$= -\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)} (1 + t^2)^{-\frac{n}{2}}$$

whence

$$\phi(t) = C - \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)} \int_{-\infty}^{t} \frac{dz}{(1+z^2)^{\frac{n}{2}}}.$$

Now

$$\phi(+\infty) = 0; \qquad \int_{-\infty}^{\infty} \frac{dz}{(1+z^2)^{\frac{n}{2}}} = \frac{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$$

so that $C = 1$ and

$$\phi(t) = 1 - \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)} \int_{-\infty}^{t} \frac{dz}{(1+z^2)^{\frac{n}{2}}}.$$

Such is the probability of the inequality

$$s \geqq t\epsilon.$$

The probability $F(t)$ of the inequality

$$s < t\epsilon$$

will be $1 - \phi(t)$ or

$$F(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)} \int_{-\infty}^{t} \frac{dz}{(1+z^2)^{\frac{n}{2}}},$$

but this is established only for positive t. However, this result holds for negative t as well. For t being negative $= -\tau$ the inequality

$$s < -\tau\epsilon$$

is entirely equivalent to

$$-s > \tau\epsilon$$

and its probability is evidently

$$F(-\tau) = \phi(\tau) = 1 - \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)} \int_{-\infty}^{\tau} (1+z^2)^{-\frac{n}{2}} dz.$$

But

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)}\int_{-\infty}^{\infty}(1+z^2)^{-\frac{n}{2}}dz = 1$$

which permits of writing the preceding expression for $F(-\tau)$ as follows

$$F(-\tau) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)}\int_{\tau}^{\infty}(1+z^2)^{-\frac{n}{2}}dz =$$

$$= \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)}\int_{-\infty}^{-\tau}(1+z^2)^{-\frac{n}{2}}dz.$$

Thus, no matter whether t is positive or negative, the distribution function of the ratio

$$\frac{s}{\epsilon}$$

or the probability of the inequality

$$s < t\epsilon$$

is given by

$$F(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)}\int_{-\infty}^{t}(1+z^2)^{-\frac{n}{2}}dz.$$

The distribution of the quotient s/ϵ was discovered by a British statistician who wrote under the pseudonym "Student," and it is commonly referred to as "Student's distribution." The first rigorous proof was published by R. A. Fisher.

4. Problem 4. Variables x, y are in normal correlation. A sample of n corresponding pairs, $x_1, y_1; x_2, y_2; \cdots x_n, y_n$ is taken and the "correlation coefficient of the sample" is found by the formula

$$\rho = \frac{\Sigma(x_i - s)(y_i - s')}{\sqrt{\Sigma(x_i - s)^2 \cdot \Sigma(y_i - s')^2}}$$

where, for the sake of abbreviation,

$$s = \frac{x_1 + x_2 + \cdots + x_n}{n}, \qquad s' = \frac{y_1 + y_2 + \cdots + y_n}{n}.$$

Find the distribution function of ρ, that is, the probability P of the inequality $\rho < R$ for a given $R(-1 < R < 1)$.

Solution. Since the expression of ρ is homogeneous of degree 0 in $x_1, x_2, \ldots x_n; y_1, y_2, \ldots y_n$ we can assume $\sigma_1 = \sigma_2 = 1$. Also without loss of generality the expectations of x and y may be supposed $=0$. Denoting by r the correlation coefficient of x and y, the density of probability in the two-dimensional distribution will be:

$$\frac{1}{2\pi(1-r^2)^{\frac{1}{2}}}e^{-\frac{1}{2(1-r^2)}(x^2+y^2-2rxy)}.$$

Hence the required probability will be expressed by the multiple integral

$$P = \frac{1}{(2\pi)^n(1-r^2)^{\frac{n}{2}}}\int\int\cdots\int e^{-\frac{\varphi}{2(1-r^2)}}dx_1\cdots dx_n dy_1\cdots dy_n$$

extended over the $2n$-dimensional domain

(3) $$\Sigma(x_i - s)(y_i - s') < R\sqrt{\Sigma(x_i - s)^2 \cdot \Sigma(y_i - s')^2}$$

and

(4) $$\varphi = \Sigma x_i^2 + \Sigma y_i^2 - 2r\Sigma x_i y_i.$$

Replacing $x_i, y_i (i = 1, 2, \ldots n)$, respectively, by $\sqrt{1-r^2}x_i$, $\sqrt{1-r^2}y_i$, we can write P thus:

$$P = \frac{(1-r^2)^{\frac{n}{2}}}{(2\pi)^n}\int\int\cdots\int e^{-\frac{1}{2}\varphi}dx_1\cdots dx_n dy_1\cdots dy_n$$

while (3) and (4) still hold but with the new notation for the variables.

Let us set now

$$x_i - s = u_i, \qquad y_i - s' = v_i,$$

then

$$u_1 + u_2 + \cdots + u_n = 0, \qquad v_1 + v_2 + \cdots + v_n = 0.$$

Introducing $s, s'; u_1, u_2, \ldots u_{n-1}; v_1, v_2, \ldots v_{n-1}$ as new variables, we find as in Sec. 2

$$P = \frac{n^2(1-r^2)^{\frac{n}{2}}}{(2\pi)^n}\int\int\cdots\int e^{-\frac{1}{2}\psi}ds ds' du_1\cdots du_{n-1}dv_1\cdots dv_{n-1}$$

where

$$\psi = ns^2 + ns'^2 - 2nrss' + \Sigma u_i^2 + \Sigma v_i^2 - 2r\Sigma u_i v_i$$

and the domain of integration is defined by the inequalities

$$-\infty < s < \infty; \qquad -\infty < s' < \infty$$
$$\Sigma u_i v_i < R\sqrt{\Sigma u_i^2 \cdot \Sigma v_i^2}.$$

Now by the same linear transformation the quadratic forms Σu_i^2, Σv_i^2 (each containing $n - 1$ independent variables) can be transformed into

$$\sum_{i=1}^{n-1} w_i^2, \qquad \sum_{i=1}^{n-1} z_i^2;$$

at the same time

$$\sum_{i=1}^{n} u_i v_i = \sum_{i=1}^{n-1} w_i z_i.$$

Proceeding as in Sec. 2 and noting that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{n}{2}(s^2+s'^2-2rss')}ds\,ds' = \frac{2\pi}{n\sqrt{1-r^2}},$$

we find that

$$P = \frac{(1-r^2)^{\frac{n-1}{2}}}{(2\pi)^{n-1}}\int\int \cdots \int e^{-\frac{1}{2}\chi}dw_1 \cdots dw_{n-1}dz_1 \cdots dz_{n-1}$$

where

$$\chi = \Sigma w_i^2 + \Sigma z_i^2 - 2r\Sigma w_i z_i$$

and the domain of integration in the space of $2n - 2$ dimensions is defined by

$$\Sigma w_i z_i < R\sqrt{\Sigma w_i^2 \cdot \Sigma z_i^2}.$$

We shall integrate now in regard to variables $z_1, z_2, \ldots z_{n-1}$ for a *fixed* system of values $w_1, w_2, \ldots w_{n-1}$. To this end we use an orthogonal transformation

$$z_1 = c_{1,1}\zeta_1 + c_{1,2}\zeta_2 + \cdots + c_{1,n-1}\zeta_{n-1}$$
$$z_2 = c_{2,1}\zeta_1 + c_{2,2}\zeta_2 + \cdots + c_{2,n-1}\zeta_{n-1}$$
$$\cdots\cdots$$
$$z_{n-1} = c_{n-1,1}\zeta_1 + c_{n-1,2}\zeta_2 + \cdots + c_{n-1,n-1}\zeta_{n-1}$$

in which the elements of the first column are

$$c_{i,1} = \frac{w_i}{\sqrt{w_1^2 + \cdots + w_{n-1}^2}} = \frac{w_i}{w}.$$

Defining $\xi_1, \xi_2, \ldots \xi_{n-1}$ by

$$\begin{aligned} w_1 &= c_{1,1}\xi_1 + c_{1,2}\xi_2 + \cdots + c_{1,n-1}\xi_{n-1} \\ w_2 &= c_{2,1}\xi_1 + c_{2,2}\xi_2 + \cdots + c_{2,n-1}\xi_{n-1} \\ &\cdots\cdots\cdots \end{aligned}$$

we shall have $\xi_1 = w,\ \xi_2 = \cdots = \xi_{n-1} = 0$. By the properties of orthogonal transformations

$$\Sigma z_i^2 = \Sigma \zeta_i^2, \qquad \Sigma z_i w_i = \Sigma \zeta_i \xi_i = w\zeta_1$$

so that for a fixed system of values $w_1, w_2, \ldots w_{n-1}$ the domain of integration in the space of variables $\zeta_1, \zeta_2, \ldots \zeta_{n-1}$ will be

(5) $$\zeta_1 < R\sqrt{\Sigma\zeta_i^2}.$$

Thus we must first evaluate the integral

$$J = \int\int \cdots \int e^{-\frac{1}{2}(\zeta_2^2 + \cdots + \zeta_{n-1}^2) - \frac{1}{2}\zeta_1^2 + rw\zeta_1} d\zeta_1 d\zeta_2 \cdots d\zeta_{n-1}.$$

If $\zeta_1 < 0$ no restriction is imposed upon $\zeta_2, \ldots \zeta_{n-1}$; if $\zeta_1 > 0$, then

$$\zeta_2^2 + \cdots + \zeta_{n-1}^2 > \left(\frac{1}{R^2} - 1\right)\zeta_1^2.$$

Consequently the result of integration in regard to $\zeta_2, \ldots \zeta_{n-1}$ can be presented thus:

$$J = ce^{\frac{r^2w^2}{2}} - \int_0^\infty e^{-\frac{1}{2}\zeta_1^2 + rw\zeta_1} d\zeta_1 \int\int \cdots \int e^{-\frac{1}{2}(\zeta_2^2 + \cdots + \zeta_{n-1}^2)} d\zeta_2 \cdots d\zeta_{n-1}$$

where the inner integral is extended over the domain

$$\zeta_2^2 + \cdots + \zeta_{n-1}^2 < \left(\frac{1}{R^2} - 1\right)\zeta_1^2$$

and c is a constant. Making use of formula (1), Sec. 1, the expression of J reduces to

$$J = ce^{\frac{r^2w^2}{2}} - \frac{2\pi^{\frac{n-2}{2}}}{\Gamma\left(\frac{n-2}{2}\right)} \int_0^\infty e^{-\frac{1}{2}\zeta_1^2 + rw\zeta_1} d\zeta_1 \int_0^{\zeta_1\left(\frac{1}{R^2} - 1\right)^{\frac{1}{2}}} e^{-\frac{v^2}{2}} v^{n-3} dv.$$

This has to be multiplied by

$$\frac{1}{(2\pi)^{n-1}}(1 - r^2)^{\frac{n-1}{2}} e^{-\frac{1}{2}\Sigma w_i^2} dw_1 \cdots dw_{n-1}$$

and integrated over the whole space of the variables $w_1, w_2, \ldots w_{n-1}$. The resulting expression for P will be

$$P = \text{const.} - \frac{2\pi^{\frac{n-2}{2}}(1-r^2)^{\frac{n-1}{2}}}{(2\pi)^{n-1}\Gamma\left(\frac{n-2}{2}\right)}\int\int\cdots\int e^{-\frac{1}{2}\Sigma w_i^2}M\,dw_1\cdots dw_{n-1}$$

where

$$M = \int_0^\infty e^{-\frac{1}{2}\zeta_1^2+rw\zeta_1}d\zeta_1\int_0^{\zeta_1\left(\frac{1}{R^2}-1\right)^{\frac{1}{2}}} e^{-\frac{v^2}{2}}v^{n-3}dv.$$

Now we differentiate in regard to R, reverse the order of integrations, and make use of formula (2), Sec. 1; the resulting value of dP/dR will then be expressed as a double integral

$$\frac{dP}{dR} = \frac{\pi^{-\frac{1}{2}}(1-r^2)^{\frac{n-1}{2}}(1-R^2)^{\frac{n-4}{2}}}{2^{n-3}\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}\int_0^\infty\int_0^\infty e^{-\frac{1}{2}(t^2+u^2)+Rrtu}(tu)^{n-2}dtdu,$$

or

$$\frac{dP}{dR} = \frac{(1-r^2)^{\frac{n-1}{2}}(1-R^2)^{\frac{n-4}{2}}}{\pi\Gamma(n-2)}\int_0^\infty\int_0^\infty e^{-\frac{1}{2}(t^2+u^2)+Rrtu}(tu)^{n-2}dtdu,$$

since

$$\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-1}{2}\right) = \frac{\sqrt{\pi}}{2^{n-3}}\Gamma(n-2).$$

In the double integral we make transformation to new variables ξ, η defined by

$$\xi = \frac{t}{u}, \qquad \eta = tu.$$

The Jacobian of t, u in regard to ξ, η, being $\frac{1}{2}\xi^{-1}$, we have

$$\int_0^\infty\int_0^\infty e^{-\frac{1}{2}(t^2+u^2)+Rrtu}(tu)^{n-2}dtdu = \frac{1}{2}\int_0^\infty\int_0^\infty e^{-\eta\left(\frac{\xi+\xi^{-1}}{2}-Rr\right)}\eta^{n-2}\frac{d\eta d\xi}{\xi} =$$
$$= \frac{1}{2}\Gamma(n-1)\int_0^\infty\frac{\xi^{-1}d\xi}{\left(\frac{\xi+\xi^{-1}}{2}-Rr\right)^{n-1}} = \Gamma(n-1)\int_0^\infty\frac{dt}{(cht-Rr)^{n-1}};$$

and so, finally,

$$\frac{dP}{dR} = \frac{n-2}{\pi}(1-r^2)^{\frac{n-1}{2}}(1-R^2)^{\frac{n-4}{2}}\int_0^\infty\frac{dt}{(cht-Rr)^{n-1}}.$$

In case $r = 0$, that is, when the variables x, y are uncorrelated, we have a very simple expression of P:

$$P = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-2}{2}\right)}\int_{-1}^{R}(1-\rho^2)^{\frac{n-4}{2}}d\rho.$$

In case $r \neq 0$ the integral

$$\int_0^\infty \frac{dt}{(cht - Rr)^{n-1}}$$

can still be found in finite form. We have, in fact,

$$\int_0^\infty \frac{dt}{cht - Rr} = \frac{1}{\sqrt{1 - R^2r^2}}\left[\frac{\pi}{2} + \text{arc sin } (Rr)\right],$$

whence

$$\int_0^\infty \frac{dt}{(cht - Rr)^{n-1}} = \frac{r^{-(n-2)}}{(n-2)!}\frac{d^{n-2}}{dR^{n-2}}\left\{[1 - R^2r^2]^{-\frac{1}{2}}\left[\frac{\pi}{2} + \text{arc sin } (Rr)\right]\right\},$$

and so

$$P = A\int_{-1}^{R}(1-\rho^2)^{\frac{n-4}{2}}\frac{d^{n-2}}{d\rho^{n-2}}\left\{[1 - \rho^2r^2]^{-\frac{1}{2}}\left[\frac{\pi}{2} + \text{arc sin } (r\rho)\right]\right\}d\rho,$$

where

$$A = \frac{r^{-(n-2)}(1-r^2)^{\frac{n-1}{2}}}{\pi(n-3)!}.$$

When n is an even number, this integral appears in a very simple finite form, but in case of an odd n certain integrals of a rather complicated type appear. Besides, the behavior of P for somewhat large n cannot be easily grasped by using this integral expression for P.

5. Fisher, who was first to discover the rigorous distribution of the correlation coefficient, called attention to the fact that, setting

$$thz = \frac{\Sigma(x_i - s)(y_i - s')}{\sqrt{\Sigma(x_i - s)^2\Sigma(y_i - s')^2}},$$

the distribution of z will be nearly normal even for comparatively small values of n. Let us set $thR = \omega$, $th\zeta = r$; then P can be expressed thus:

$$P = \frac{n-2}{\pi}\int_{-\infty}^{\omega}\int_0^\infty \frac{chzdtdz}{(chtchzch\zeta - sh\zeta shz)^{n-1}}.$$

Instead of t it is convenient to introduce a new variable τ so that

$$chtchzch\zeta - sh\zeta shz = \tau^{-1}ch(z - \zeta).$$

Then

$$P = \frac{n-2}{\pi\sqrt{2}} \int_{-\infty}^{\omega} \left(\frac{chz}{ch\zeta}\right)^{\frac{1}{2}} \frac{dz}{[ch(z-\zeta)]^{n-\frac{3}{2}}} \int_0^1 \frac{\tau^{-\frac{1}{2}}(1-\tau)^{n-2}d\tau}{\sqrt{1-p\tau}}$$

where

$$p = \frac{ch(z+\zeta)}{2chzch\zeta} \leqq \frac{ch(\omega+\zeta)}{2ch\omega ch\zeta}$$

for all values of z under consideration. Now

$$\int_0^1 \frac{\tau^{-\frac{1}{2}}(1-\tau)^{n-2}d\tau}{\sqrt{1-p\tau}} > \frac{\sqrt{\pi}\Gamma(n-1)}{\Gamma(n-\frac{1}{2})}$$

and

$$\int_0^1 \frac{\tau^{-\frac{1}{2}}(1-\tau)^{n-2}d\tau}{\sqrt{1-p\tau}} < \frac{\sqrt{\pi}\Gamma(n-1)}{\Gamma(n-\frac{1}{2})}\left(1+\frac{p}{2n-1}\right)$$

since

$$\int_0^1 \frac{\tau^{-\frac{1}{2}}(1-\tau)^{n-2}d\tau}{\sqrt{1-p\tau}} < \int_0^1 \tau^{-\frac{1}{2}}(1-\tau)^{n-2}(1+p\tau)d\tau$$

for $0 < p < 1$ as can be easily verified. Consequently

$$P = \frac{(n-2)\Gamma(n-1)}{\sqrt{2\pi}\Gamma(n-\frac{1}{2})} \int_{-\infty}^{\omega} \left(\frac{chz}{ch\zeta}\right)^{\frac{1}{2}} \frac{dz}{[ch(z-\zeta)]^{n-\frac{3}{2}}} \cdot$$

$$\cdot \left[1 + \frac{ch(\omega+\zeta)}{2ch\omega ch\zeta}\frac{\theta}{2n-1}\right]; \qquad 0 < \theta < 1.$$

As to the integral in this formula, its approximate expression, omitting terms of the higher order, is:

$$\int_{-\infty}^{\omega} e^{-\frac{2n-3}{4}(z-\zeta)^2}dz - \frac{th\zeta}{2n-3}e^{-\frac{2n-3}{4}(\omega-\zeta)^2}.$$

Thus for somewhat large n the required value of P can be found with the help of a simple approximate formula.

The various distributions dealt with in this chapter are undoubtedly of great value when applied to variables which have normal or nearly normal distribution. Whether they are always used legitimately can be doubted. At least the "onus probandi" that the "populations" with which they deal are even approximately normal rests with the statisticians.

Problems for Solution

1. Show that

$$\lim_{n\to\infty} \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \int_0^{1+t\sqrt{\frac{2}{n}}} e^{-\frac{nu}{2}} u^{\frac{n}{2}-1} du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}u^2} du$$

HINT: Liapounoff's theorem and Prob. 1, page 332.

2. With the same assumptions and notations as in Prob. 3, page 336, show that the distribution function of the quotient

$$\frac{x_i - s}{\epsilon}; \qquad i = 1, 2, \ldots n$$

is

$$F(t) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi(n-1)}\,\Gamma\left(\frac{n-2}{2}\right)} \int_{-\sqrt{n-1}}^{t} \left(1 - \frac{\tau^2}{n-1}\right)^{\frac{n-4}{2}} d\tau \quad \text{if} \quad |t| \leqq \sqrt{n-1}$$

$$F(t) = 1 \quad \text{if} \quad t > \sqrt{n-1}; \qquad F(t) = 0 \quad \text{if} \quad t < -\sqrt{n-1}.$$

It is worthy of notice that for $n = 4$ the distribution is uniform.[1]

3. In two series of observations, samples $x_1, x_2, \ldots x_n$ and $y_1, y_2, \ldots y_{n'}$ from the same normally distributed population (or of the same normally distributed variable) are obtained. Denoting for brevity

$$s = \frac{x_1 + x_2 + \cdots + x_n}{n}, \qquad s' = \frac{y_1 + y_2 + \cdots + y_{n'}}{n'}$$

$$\epsilon = \sqrt{\left(\frac{1}{n} + \frac{1}{n'}\right)[\Sigma(x_i - s)^2 + \Sigma(y_i - s')^2]},$$

find the distribution function of the quotient $\frac{s - s'}{\epsilon}$ ("Student"). *Ans.*

$$F(t) = \frac{\Gamma\left(\frac{n + n' - 1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n + n' - 2}{2}\right)} \int_{-\infty}^{t} \frac{d\tau}{(1 + \tau^2)^{\frac{n+n'-1}{2}}}.$$

References

"STUDENT": The Standard Error of a Mean, *Biometrica*, vol. 6, 1908.

R. A. FISHER: Frequency Distribution of the Values of the Correlation Coefficient in Samples from Indefinitely Large Population, *Biometrica*, vol. 10, 1915.

———: On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample, *Metron*, vol. 1, 1920.

D. JACKSON: On the Mathematical Theory of Small Samples, *Amer. Math. Monthly*, vol. 42, 1935.

[1] COMMANDER LHOSTE: Mémorial de l'artillerie française, pp. 245 and 1027, 1925.

APPENDIX I

1. Euler's Summation Formula. Let $f(x)$ be a function with a continuous derivative $f'(x)$ in an interval (a, b) where a and $b > a$ are arbitrary real numbers. The notation

$$\sum_{n>a}^{n\leqq b} f(n)$$

will be used to designate the sum extended over all integers n which are $>a$ and $\leqq b$. It is an important problem to devise means for the approximate evaluation of the above sum when it contains a considerable number of terms.

Let $[x]$, as usual, denote the largest integer contained in a real number x, so that

$$x = [x] + \theta$$

where θ, so-called "fractional part" of x, satisfies the inequalities

$$0 \leqq \theta < 1.$$

Considered as functions of a continuous variable x, both $[x]$ and θ have discontinuities for integral values of x. The function

$$\rho(x) = \tfrac{1}{2} - \theta = [x] - x + \tfrac{1}{2}$$

is likewise discontinuous for integral values of x. Besides, it is a periodic function of x with the period 1; that is, we have

$$\rho(x + 1) = \rho(x)$$

for any real x. With this notation adopted we have the following important formula:

$$(1) \qquad \sum_{n>a}^{n\leqq b} f(n) = \int_a^b f(x)dx + \rho(b)f(b) - \rho(a)f(a) - \int_a^b \rho(x)f'(x)dx$$

which is known as "Euler's summation formula."

Proof. Let k be the least integer $>a$ and l the greatest integer $\leqq b$. The sum in the left member of (1) is, by definition,

$$f(k) + f(k + 1) + \cdot\cdot\cdot + f(l)$$

and we must show that this is equal to the right member. To this end we write first

$$\int_a^b \rho(x)f'(x)dx = \int_a^k \rho(x)f'(x)dx + \int_l^b \rho(x)f'(x)dx + \sum_{j=k}^{j=l-1} \int_j^{j+1} \rho(x)f'(x)dx.$$

Next, since j is an integer,

$$\int_j^{j+1} \rho(x)f'(x)dx = \int_j^{j+1}\left(j - x + \frac{1}{2}\right)f'(x)dx = -\frac{f(j) + f(j+1)}{2} + \\ + \int_j^{j+1} f(x)dx$$

and

$$\sum_{j=k}^{j=l-1} \int_j^{j+1} \rho(x)f'(x)dx = -\frac{f(k) + f(l)}{2} - \sum_{n=k+1}^{n=l-1} f(n) + \int_k^l f(x)dx.$$

On the other hand,

$$\int_a^k \rho(x)f'(x)dx = \int_a^k\left(k - 1 - x + \frac{1}{2}\right)f'(x)dx = -\frac{f(k)}{2} - \rho(a)f(a) + \\ + \int_a^k f(x)dx$$

$$\int_l^b \rho(x)f'(x)dx = \int_l^b\left(l - x + \frac{1}{2}\right)f'(x)dx = -\frac{f(l)}{2} + \rho(b)f(b) + \int_l^b f(x)dx,$$

so that finally

$$\int_a^b \rho(x)f'(x)dx = -f(k) - f(k+1) - \cdots - f(l) + \\ + \rho(b)f(b) - \rho(a)f(a) + \int_a^b f(x)dx;$$

whence

$$\sum_{n>a}^{n\leq b} f(n) = \int_a^b f(x)dx + \rho(b)f(b) - \rho(a)f(a) - \int_a^b \rho(x)f'(x)dx,$$

which completes the proof of Euler's formula.

Corollary 1. The integral

$$\int_0^x \rho(z)dz = \sigma(x)$$

represents a continuous and periodic function of x with the period 1. For

$$\sigma(x+1) - \sigma(x) = \int_x^{x+1} \rho(z)dz = \int_0^1 \rho(z)dz = \int_0^1 (\tfrac{1}{2} - z)dz = 0.$$

If $0 \leqq x \leqq 1$,

$$\sigma(x) = \int_0^x \left(\frac{1}{2} - z\right)dz = \frac{x(1-x)}{2}$$

and in general

$$\sigma(x) = \frac{\theta(1-\theta)}{2}$$

where θ is a fractional part of x. Hence, for every real x

$$0 \leqq \sigma(x) \leqq \tfrac{1}{8}.$$

Supposing that $f''(x)$ exists and is continuous in (a, b) and integrating by parts, we get

$$\int_a^b \rho(x)f'(x)dx = \sigma(b)f'(b) - \sigma(a)f'(a) - \int_a^b \sigma(x)f''(x)dx,$$

which leads to another form of Euler's formula:

$$\sum_{n>a}^{n \leqq b} f(n) = \int_a^b f(x)dx + \rho(b)f(b) - \rho(a)f(a) - \sigma(b)f'(b) +$$

$$+ \sigma(a)f'(a) + \int_a^b \sigma(x)f''(x)dx.$$

Corollary 2. If $f(x)$ is defined for all $x \geqq a$ and possesses a continuous derivative throughout the interval $(a, +\infty)$; if, besides, the integral

$$\int_a^\infty \rho(x)f'(x)dx$$

exists, then for a variable limit b we have

$$\text{(2)} \qquad \sum_{n>a}^{n \leqq b} f(n) = C + \int f(b)db + \rho(b)f(b) + \int_b^\infty \rho(x)f'(x)dx$$

where C is a constant with respect to b.

It suffices to substitute for

$$\int_a^b \rho(x)f'(x)dx$$

the difference

$$\int_a^\infty \rho(x)f'(x)dx - \int_b^\infty \rho(x)f'(x)dx$$

and separate the terms depending upon b from those involving a.

2. Stirling's Formula. Factorials increase with extreme rapidity and their exact computation soon becomes practically impossible. The question then naturally arises of finding a convenient approximate expression for large factorials, which question is answered by a celebrated formula usually known as "Stirling's formula," although, in the main, it was established by de Moivre in connection with problems on probability. De Moivre did not establish the relation to the number

$$\pi = 3.14159 \ldots$$

of the constant involved in his formula; it was done by Stirling.

In formula (2) it suffices to take $a = \frac{1}{2}$, $f(x) = \log x$, and replace b by an arbitrary integer n to arrive at the remarkable expression

$$\log (1 \cdot 2 \cdot 3 \cdots n) = C + \left(n + \frac{1}{2}\right) \log n - n + \int_n^\infty \frac{\rho(x)dx}{x}$$

where C is a constant. For the sake of brevity we shall set

$$\omega(n) = \int_n^\infty \frac{\rho(x)dx}{x}.$$

Now

$$\int_n^\infty \frac{\rho(x)dx}{x} = \int_n^{n+1} \frac{\rho(x)dx}{x} + \int_{n+1}^{n+2} \frac{\rho(x)dx}{x} + \cdots$$

and

$$\int_k^{k+1} \frac{\rho(x)dx}{x} = \int_0^1 \frac{\rho(u)du}{u+k} = \int_0^{\frac{1}{2}} \frac{\rho(u)du}{u+k} + \int_{\frac{1}{2}}^1 \frac{\rho(u)du}{u+k} =$$

$$= \int_0^{\frac{1}{2}} \frac{(\frac{1}{2} - u)du}{u+k} + \int_{\frac{1}{2}}^1 \frac{(\frac{3}{2} - u)du}{u+k} = \frac{1}{2}\int_0^{\frac{1}{2}} \frac{(1-2u)^2du}{(k+u)(k+1-u)}.$$

Hence

$$\omega(n) = \tfrac{1}{2}\int_0^{\frac{1}{2}} (1 - 2u)^2 F_n(u)du$$

where

$$F_n(u) = \sum_{k=n}^{\infty} \frac{1}{(k+u)(k+1-u)}.$$

Since

$$(k+u)(k+1-u) = k(k+1) + u - u^2,$$

it follows that for $0 < u < \frac{1}{2}$

$$(k+u)(k+1-u) > k(k+1)$$
$$(k+u)(k+1-u) < (k+\tfrac{1}{2})^2 < (k+\tfrac{1}{3})(k+\tfrac{2}{3}).$$

Thus for $0 < u < \frac{1}{2}$

$$F_n(u) < \sum_{k=n}^{\infty} \frac{1}{k(k+1)} = \frac{1}{n}$$

$$F_n(u) > \sum_{k=n}^{\infty} \frac{1}{(k+\frac{1}{2})(k+\frac{3}{2})} = \frac{1}{n+\frac{1}{2}}.$$

Making use of these limits, we find that

$$\omega(n) < \frac{1}{2n}\int_0^{\frac{1}{2}} (1-2u)^2 du = \frac{1}{12n}$$

$$\omega(n) > \frac{1}{2n+1}\int_0^{\frac{1}{2}} (1-2u)^2 du = \frac{1}{12(n+\frac{1}{2})},$$

and consequently can set

$$\omega(n) = \frac{1}{12(n+\theta)}$$

where

$$0 < \theta < \tfrac{1}{2}.$$

Accordingly

$$\log(1 \cdot 2 \cdot 3 \cdots n) = C + \left(n + \frac{1}{2}\right)\log n - n + \frac{1}{12(n+\theta)}.$$

The constant C depends in a remarkable way on the number π. To show this we start from the well-known expression for π due to Wallis:

$$\frac{\pi}{2} = \lim\left(\frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdots\frac{2n}{2n-1}\cdot\frac{2n}{2n+1}\right), \qquad n \to \infty$$

which follows from the infinite product

$$\sin x = x\left(1 - \frac{x^2}{\pi^2}\right)\left(1 - \frac{x^2}{4\pi^2}\right)\left(1 - \frac{x^2}{9\pi^2}\right)\cdots$$

by taking $x = \pi/2$. Since

$$\frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdots\frac{2n}{2n-1}\cdot\frac{2n}{2n+1} = \left[\frac{2\cdot 4\cdot 6\cdots 2n}{1\cdot 3\cdot 5\cdots(2n-1)}\right]^2 \frac{1}{2n+1}$$

we get from Wallis' formula

$$\sqrt{\pi} = \lim\left[\frac{2\cdot 4\cdot 6\cdots 2n}{1\cdot 3\cdot 5\cdots(2n-1)}\frac{1}{\sqrt{n}}\right], \qquad n \to \infty.$$

On the other hand,

$$2\cdot 4\cdot 6\cdots 2n = 2^n\cdot 1\cdot 2\cdot 3\cdots n$$

$$1\cdot 3\cdot 5\cdots(2n-1) = \frac{1\cdot 2\cdot 3\cdots 2n}{2^n\cdot 1\cdot 2\cdot 3\cdots n},$$

so that

$$\sqrt{\pi} = \lim \left\{\frac{2^{2n}(1 \cdot 2 \cdot 3 \cdots n)^2}{1 \cdot 2 \cdot 3 \cdots 2n} \cdot \frac{1}{\sqrt{n}}\right\}, \qquad n \to \infty$$

or, taking logarithms

$$\log \sqrt{\pi} = \lim [2n \log 2 + 2 \log (1 \cdot 2 \cdot 3 \cdots n) - \\ - \log (1 \cdot 2 \cdot 3 \cdots 2n) - \tfrac{1}{2} \log n]$$

But, neglecting infinitesimals,

$$\log (1 \cdot 2 \cdot 3 \cdots n) = C + (n + \tfrac{1}{2}) \log n - n$$
$$\log (1 \cdot 2 \cdot 3 \cdots 2n) = C + (2n + \tfrac{1}{2}) \log 2n - 2n$$

whence

$$\lim [2n \log 2 + 2 \log (1 \cdot 2 \cdot 3 \cdots n) - \\ - \log (1 \cdot 2 \cdot 3 \cdots 2n) - \tfrac{1}{2} \log n] = C - \tfrac{1}{2} \log 2.$$

Thus

$$\log\sqrt{\pi} = C - \tfrac{1}{2} \log 2, \qquad C = \log \sqrt{2\pi}$$

and finally

$$\text{(3)} \quad \log (1 \cdot 2 \cdot 3 \cdots n) = \log \sqrt{2\pi} + \left(n + \frac{1}{2}\right) \log n - n + \\ + \frac{1}{12(n + \theta)}; \qquad 0 < \theta < \frac{1}{2}.$$

This is equivalent to two inequalities

$$e^{\frac{1}{12n+6}} < \frac{1 \cdot 2 \cdot 3 \cdots n}{\sqrt{2\pi n}\, n^n e^{-n}} < e^{\frac{1}{12n}}$$

which show that for indefinitely increasing n

$$\lim \frac{1 \cdot 2 \cdot 3 \cdots n}{\sqrt{2\pi n}\, n^n e^{-n}} = 1.$$

This result is commonly known as Stirling's formula.

For a finite n we have

$$1 \cdot 2 \cdot 3 \cdots n = \sqrt{2\pi n} n^n e^n \cdot e^{\omega(n)}$$

where

$$\frac{1}{12(n + \frac{1}{2})} < \omega(n) < \frac{1}{12n}.$$

The expression

$$\sqrt{2\pi n} n^n e^{-n}$$

is thus an approximate value of the factorial $1 \cdot 2 \cdot 3 \cdots n$ for large n in the sense that the ratio of both is near to 1; that is, the relative error is small. On the contrary, the absolute error will be arbitrarily large for large n, but this is irrelevant when Stirling's approximation is applied to quotients of factorials.

In this connection it is useful to derive two further inequalities.

Let $m < n$; we have, then,

$$F_m(u) - F_n(u) = \sum_{k=m}^{k=n-1} \frac{1}{(k+u)(k+1-u)};$$

and further, supposing $0 < u < \frac{1}{2}$,

$$F_m(u) - F_n(u) < \sum_{k=m}^{k=n-1} \frac{1}{k(k+1)} = \frac{1}{m} - \frac{1}{n}$$

$$F_m(u) - F_n(u) > \sum_{k=m}^{k=n-1} \frac{1}{(k+\frac{1}{2})(k+\frac{3}{2})} = \frac{1}{m+\frac{1}{2}} - \frac{1}{n+\frac{1}{2}}.$$

Hence,

$$\omega(m) - \omega(n) < \frac{1}{12m} - \frac{1}{12n}, \qquad \omega(m) - \omega(n) > \frac{1}{12(m+\frac{1}{2})} - \frac{1}{12(n+\frac{1}{2})}$$

and, if l is a third arbitrary positive integer,

$$\omega(m) + \omega(l) - \omega(n) < \frac{1}{12m} + \frac{1}{12l} - \frac{1}{12n}$$

$$\omega(m) + \omega(l) - \omega(n) > \frac{1}{12(m+\frac{1}{2})} + \frac{1}{12(l+\frac{1}{2})} - \frac{1}{12(n+\frac{1}{2})}.$$

3. Some Definite Integrals. The value of the important definite integral

$$\int_0^\infty e^{-t^2} dt$$

can be found in various ways. One of the simplest is the following: Let

$$J_n = \int_0^\infty e^{-t^2} t^n dt$$

in general where n is an arbitrary integer ≥ 0. Integrating by parts one can easily establish the recurrence relation

$$J_n = \frac{n-1}{2} J_{n-2};$$

whence

$$J_{2m} = \frac{1 \cdot 3 \cdot 5 \cdots (2m-1)}{2^m} J_0$$

$$J_{2m+1} = \frac{1 \cdot 2 \cdot 3 \cdots m}{2}.$$

On the other hand,

$$J_{n+1} + 2\lambda J_n + \lambda^2 J_{n-1} = \int_0^\infty e^{-t^2} t^{n-1}(t+\lambda)^2 dt,$$

which shows that

$$J_{n+1} + 2\lambda J_n + \lambda^2 J_{n-1} > 0$$

for all real λ. Hence, the roots of the polynomial in the left member are imaginary, and this implies

$$J_n^2 < J_{n+1} J_{n-1}.$$

Taking $n = 2m$ and $n = 2m + 1$ and using the preceding expression for J_{2m} and J_{2m+1}, we find

$$\frac{2 \cdot 4 \cdot 6 \cdots 2m}{1 \cdot 3 \cdot 5 \cdots (2m-1)} \frac{1}{\sqrt{4m+2}} < J_0 < \frac{2 \cdot 4 \cdot 6 \cdots 2m}{1 \cdot 3 \cdot 5 \cdots (2m-1)} \frac{1}{\sqrt{4m}}.$$

But

$$\lim_{m \to \infty} \frac{2 \cdot 4 \cdot 6 \cdots 2m}{1 \cdot 3 \cdot 5 \cdots (2m-1)} \frac{1}{\sqrt{m}} = \sqrt{\pi};$$

hence

$$J_0 = \int_0^\infty e^{-t^2} dt = \tfrac{1}{2}\sqrt{\pi}.$$

Here substituting $t = \sqrt{a}u$, where a is a positive parameter, we get

$$\int_0^\infty e^{-au^2} du = \frac{1}{2}\sqrt{\frac{\pi}{a}}.$$

As a generalization of the last integral we may consider the following one:

$$V = \int_0^\infty e^{-au^2} \cos bu du.$$

The simplest way to find the value of this integral is to take the derivative

$$\frac{dV}{db} = -\int_0^\infty e^{-au^2} \sin bu \cdot u du$$

and transform the right member by partial integration. The result is

$$\frac{dV}{db} = -\frac{b}{2a}V$$

or

$$d(Ve^{\frac{b^2}{4a}}) = 0,$$

whence

$$V = Ce^{-\frac{b^2}{4a}}.$$

To determine the constant C, take $b = 0$; then

$$C = (V)_{b=0} = \int_0^\infty e^{-au^2}du = \frac{1}{2}\sqrt{\frac{\pi}{a}},$$

so that finally

$$\int_0^\infty e^{-au^2} \cos bu du = \frac{1}{2}\sqrt{\frac{\pi}{a}}e^{-\frac{b^2}{4a}}.$$

The equivalent form of this integral is as follows:

$$\int_{-\infty}^\infty e^{-au^2} \cos bu du = \int_{-\infty}^\infty e^{-au^2+ibu}du = \sqrt{\frac{\pi}{a}}e^{-\frac{b^2}{4a}}.$$

APPENDIX II

METHOD OF MOMENTS AND ITS APPLICATIONS

1. Introductory Remarks. To prove the fundamental limit theorem Tshebysheff devised an ingenious method, known as the "method of moments," which later was completed and simplified by one of the most prominent among Tshebysheff's disciples, the late Markoff. The simplicity and elegance inherent in this method of moments make it advisable to present in this Appendix a brief exposition of it.

The distribution of a mass spread over a given interval (a, b) may be characterized by a never decreasing function $\varphi(x)$, defined in (a, b) and varying from $\varphi(a) = 0$ to $\varphi(b) = m_0$, where m_0 is the total mass contained in (a, b). Since $\varphi(x)$ is never decreasing, for any particular point x_0, both the limits

$$\lim \varphi(x_0 - \epsilon) = \varphi(x_0 - 0)$$
$$\lim \varphi(x_0 + \epsilon) = \varphi(x_0 + 0)$$

exist when a positive number ϵ tends to 0. Evidently

$$\varphi(x_0 - 0) \leqq \varphi(x_0) \leqq \varphi(x_0 + 0).$$

If

$$\varphi(x_0 - 0) = \varphi(x_0 + 0) = \varphi(x_0),$$

then x_0 is a "point of continuity" of $\varphi(x)$. In case

$$\varphi(x_0 + 0) > \varphi(x_0 - 0),$$

x_0 is a point of discontinuity of $\varphi(x)$, and the positive difference

$$\varphi(x_0 + 0) - \varphi(x_0 - 0)$$

may be considered as a mass concentrated at the point x_0. In all cases $\varphi(x_0 - 0)$ is the total mass on the segment (a, x_0) excluding the end point x_0, whereas $\varphi(x_0 + 0)$ is the mass spread over the same segment including the point x_0.

The points of discontinuity, if there are any, form an enumerable set, whence it follows that in any part of the interval (a, b) there are points of continuity.

If for any sufficiently small positive ϵ

$$\varphi(x_0 + \epsilon) > \varphi(x_0 - \epsilon),$$

x_0 is called a "point of increase" of $\varphi(x)$. There is at least one point of increase and there might be infinitely many. For instance, if

$$\varphi(x) = 0 \quad \text{for} \quad a \leqq x \leqq c$$
$$\varphi(x) = m_0 \quad \text{for} \quad c < x \leqq b,$$

then c is the only point of increase. On the other hand, for

$$\varphi(x) = m_0 \frac{x - a}{b - a}$$

every point of the interval (a, b) is a point of increase. In case of a finite number of points of increase the whole mass is concentrated in these points and the distribution function $\varphi(x)$ is a step function with a finite number of steps.

Stieltjes' integrals

$$\int_a^b d\varphi(x) = m_0, \qquad \int_a^b x d\varphi(x) = m_1, \cdots \int_a^b x^i d\varphi(x) = m_i$$

represent respectively the whole mass m_0 and its moments about the origin of the order 1, 2, . . . i. When the distribution function $\varphi(x)$ is given, moments $m_0, m_1, m_2, \ldots m_i$ (provided they exist) are determined. If, however, these moments are given and are known to originate in a certain distribution of a mass over (a, b), the question may be raised with what error the mass spread over an interval (a, x) can be determined by these data? In other words, given $m_0, m_1, m_2, \ldots m_i$, what are the precise upper and lower bounds of a mass spread over an interval (a, x)? Such is the question raised by Tshebysheff in a short but important article "Sur les valeurs limites des intégrales" (1874).[1] The results contained in this article, including very remarkable inequalities which indeed are of fundamental importance, are given without proof. The first proof of these results and the complete solution of the question raised by Tshebysheff was given by Markoff in his eminent thesis "On some applications of algebraic continued fractions" (St. Petersburg, 1884), written in Russian and therefore comparatively little known.

Suppose that ρ_i is the limit of the error with which we can evaluate the mass belonging to the interval (a, x) or, which is almost the same, the value of $\varphi(x)$, when moments $m_0, m_1, m_2, \ldots m_i$ are given. If, with i tending to infinity, ρ_i tends to 0 for any given x, then the distribution function $\varphi(x)$ will be completely determined by giving all the moments

$$m_0, m_1, m_2, \ldots .$$

One case of this kind, that in which

$$m_0 = 1, \qquad m_{2k} = \frac{1 \cdot 3 \cdot 5 \cdots (2k - 1)}{2^k}, \qquad m_{2k+1} = 0$$

[1] *Jour. Liouville*, Ser. 2, T. XIX, 1874.

was considered by Tshebysheff in a later paper, "Sur deux théorèmes relatifs aux probabilités" (1887)[1] devoted to the application of his method to the proof of the limit theorem under certain rather general conditions. The success of this proof is due to the fact that moments, as given above, uniquely determine the normal distribution

$$\varphi(x) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{x} e^{-u^2}du$$

of the mass 1 over the infinite interval $(-\infty, +\infty)$.

After these preliminary remarks and before proceeding to an orderly exposition of the method of moments, it is advisable to devote a few pages to continued fractions associated with power series, for continued fractions are the natural tools in questions of the kind we shall consider.

2. Continued Fractions Associated with Power Series. Let

$$\phi(z) = \frac{A_1}{z^{\alpha_1}} + \frac{A_2}{z^{\alpha_2}} + \frac{A_3}{z^{\alpha_3}} + \cdots; \qquad (A_1 \neq 0)$$

be a power series arranged according to decreasing powers of z where the smallest exponent α_1 is positive. We consider this power series from a purely formal point of view merely as a means to form a sequence of rational fractions

$$\frac{A_1}{z^{\alpha_1}}, \quad \frac{A_1}{z^{\alpha_1}} + \frac{A_2}{z^{\alpha_2}}, \quad \frac{A_1}{z^{\alpha_1}} + \frac{A_2}{z^{\alpha_2}} + \frac{A_3}{z^{\alpha_3}}, \cdots$$

and we need not be concerned about its convergence.

Evidently $1/\phi(z)$ can again be expanded into power series, arranged according to decreasing powers of z. Let its integral part, containing non-negative powers of z, be denoted by $q_1(z)$, and let the fractional part

$$\frac{B_1}{z^{\beta_1}} + \frac{B_2}{z^{\beta_2}} + \frac{B_3}{z^{\beta_3}} + \cdots$$

containing negative powers of z, be denoted by $-\phi_1(z)$, so that

$$\frac{1}{\phi(z)} = q_1(z) - \phi_1(z).$$

In the same way

$$\frac{1}{\phi_1(z)}$$

can be represented thus:

$$\frac{1}{\phi_1(z)} = q_2(z) - \phi_2(z)$$

[1] Oeuvres complètes de P. L. Tshebysheff, Tome 2, p. 482.

where $q_2(z)$ is a polynomial and

$$\phi_2(z) = \frac{C_1}{z^{\gamma_1}} + \frac{C_2}{z^{\gamma_2}} + \frac{C_3}{z^{\gamma_3}} + \cdots,$$

a power series containing only negative powers of z. Further, we shall have

$$\frac{1}{\phi_2(z)} = q_3(z) - \phi_3(z)$$

with a certain polynomial $q_3(z)$ and a power series

$$\phi_3(z) = \frac{D_1}{z^{\delta_1}} + \frac{D_2}{z^{\delta_2}} + \frac{D_3}{z^{\delta_3}} + \cdots$$

containing negative powers of z, and so on. Thus we are led to consider a continued fraction (finite or infinite)

(1)
$$\cfrac{1}{q_1 - \cfrac{1}{q_2 - \cfrac{1}{q_3 - \cdots}}}$$

associated with $\phi(z)$ in the sense that the formal expansion of

$$\cfrac{1}{q_1 - \cfrac{1}{q_2 - \cdots - \cfrac{1}{q_i - \phi_i(z)}}}$$

into a power series will reproduce exactly $\phi(z)$. The continued fraction (1) is again considered from a purely formal standpoint as a mere abbreviation of the sequence of its convergents

$$\frac{P_1}{Q_1} = \frac{1}{q_1}; \quad \frac{P_2}{Q_2} = \cfrac{1}{q_1 - \cfrac{1}{q_2}}; \quad \frac{P_3}{Q_3} = \cfrac{1}{q_1 - \cfrac{1}{q_2 - \cfrac{1}{q_3}}}; \cdots.$$

The polynomials

$$P_1, P_2, P_3, \ldots$$
$$Q_1, Q_2, Q_3, \ldots$$

can be found step by step by the recurrence relations

(2)
$$\left.\begin{aligned} P_i &= q_iP_{i-1} - P_{i-2} \\ Q_i &= q_iQ_{i-1} - Q_{i-2} \end{aligned}\right\} i = 2, 3, 4, \ldots$$
$$P_1 = 1, \qquad P_0 = 0$$
$$Q_1 = q_1, \qquad Q_0 = 1$$

from which the following identical relation follows:

$$P_i(z)Q_{i-1}(z) - Q_i(z)P_{i-1}(z) = 1, \tag{3}$$

showing that all fractions

$$\frac{P_i(z)}{Q_i(z)}$$

are irreducible. Evidently degrees of consecutive denominators of convergents form an increasing sequence and the degree of $Q_i(z)$ is at least i. Since

$$\cfrac{1}{q_1 - \cfrac{1}{q_2 - \cdots - \cfrac{1}{q_{i+1} - \phi_{i+1}(z)}}} = \frac{P_i(q_{i+1} - \phi_{i+1}(z)) - P_{i-1}}{Q_i(q_{i+1} - \phi_{i+1}(z)) - Q_{i-1}} = \frac{P_{i+1} - P_i\phi_{i+1}(z)}{Q_{i+1} - Q_i\phi_{i+1}(z)}$$

we can write

$$\phi(z) = \frac{P_{i+1} - P_i\phi_{i+1}(z)}{Q_{i+1} - Q_i\phi_{i+1}(z)}$$

in the sense that the formal development of the right-hand member is identical with $\phi(z)$. By virtue of relation (3)

$$\phi(z) - \frac{P_i}{Q_i} = \frac{1}{Q_i(Q_{i+1} - Q_i\phi_{i+1})}.$$

The degree of Q_i being λ_i and that of Q_{i+1} being λ_{i+1}, the expansion of

$$Q_i(Q_{i+1} - Q_i\phi_{i+1})$$

in a series of descending powers of z begins with the power $z^{\lambda_i+\lambda_{i+1}}$. Hence,

$$\phi(z) - \frac{P_i}{Q_i} = \frac{M}{z^{\lambda_i+\lambda_{i+1}}} + \cdots$$

and, since $\lambda_{i+1} \geqq \lambda_i + 1$, the expansion of

$$\phi(z) - \frac{P_i}{Q_i}$$

begins with a term of the order $2\lambda_i + 1$ in $1/z$ at least. This property characterizes the convergents P_i/Q_i completely. For let P/Q be a rational fraction whose denominator is of the nth degree and such that in the expansion of

$$\phi(z) - \frac{P}{Q}$$

the lowest term is of the order $2n + 1$ in $1/z$ at least. Then P/Q coincides with one of the convergents to the continued fraction (1). Let i be determined by the condition

$$\lambda_i \leqq n < \lambda_{i+1}.$$

Then

$$\phi(z) - \frac{P_i}{Q_i} = \frac{M}{z^{\lambda_i+\lambda_{i+1}}} + \cdots$$

$$\phi(z) - \frac{P}{Q} = \frac{N}{z^{2n+1}} + \cdots$$

whence in the expansion of

$$\frac{P}{Q} - \frac{P_i}{Q_i}$$

the lowest term will be of degree $2n + 1$ or $\lambda_i + \lambda_{i+1}$ in $1/z$. Hence, the degree of

$$PQ_i - P_iQ$$

in z is not greater than both the numbers

$$\lambda_i - n - 1 \qquad \text{and} \qquad n - \lambda_{i+1}$$

which are both negative while

$$PQ_i - P_iQ$$

is a polynomial. Hence, identically,

$$PQ_i - P_iQ = 0$$

or

$$\frac{P}{Q} = \frac{P_i}{Q_i}$$

which proves the statement.

3. Continued Fraction Associated with $\int_a^b \frac{d\varphi(x)}{z - x}$. Let $\varphi(x)$ be a never decreasing function characterizing the distribution of a mass over an interval (a, b). The moments of this distribution up to the moment of the order $2n$ are represented by integrals

$$m_0 = \int_a^b d\varphi(x), \qquad m_1 = \int_a^b x\,d\varphi(x),$$

$$m_2 = \int_a^b x^2 d\varphi(x), \quad \cdots \quad m_{2n} = \int_a^b x^{2n} d\varphi(x).$$

Let

$$\Delta_0 = m_0; \quad \Delta_1 = \begin{vmatrix} m_0 m_1 \\ m_1 m_2 \end{vmatrix}; \quad \Delta_2 = \begin{vmatrix} m_0 m_1 m_2 \\ m_1 m_2 m_3 \\ m_2 m_3 m_4 \end{vmatrix}; \cdots \Delta_n = \begin{vmatrix} m_0 m_1 & \cdots & m_n \\ m_1 m_2 & \cdots & m_{n+1} \\ \cdot & \cdot\ \cdot\ \cdot\ \cdot & \cdot \\ m_n m_{n+1} & \cdots & m_{2n} \end{vmatrix}.$$

If $\varphi(x)$ has not less than $n + 1$ points of increase, we must have

$$\Delta_0 > 0, \qquad \Delta_1 > 0, \cdots \Delta_n > 0,$$

and conversely, if these inequalities are satisfied, $\varphi(x)$ has at least $n + 1$ points of increase. To prove this, consider the quadratic form

$$\phi = \int_a^b (t_0 + t_1 x + \cdots + t_n x^n)^2 d\varphi(x)$$

in $n + 1$ variables $t_0, t_1, \ldots t_n$. Evidently

$$\phi = \Sigma m_{i+j} t_i t_j \qquad (i, j = 0, 1, 2, \ldots n)$$

so that Δ_n is the determinant of ϕ and $\Delta_0, \Delta_1, \ldots \Delta_{n-1}$ its principal minors. The form ϕ cannot vanish unless $t_0 = t_1 = \cdots = t_n = 0$. For if $x = \xi$ is a point of increase and $\phi = 0$, we must have also

$$\int_{\xi-\epsilon}^{\xi+\epsilon} (t_0 + t_1 x + \cdots + t_n x^n)^2 d\varphi(x) = 0$$

for an arbitrary positive ϵ, whence by the mean value theorem

$$(t_0 + t_1\eta + \cdots + t_n\eta^n)^2 \int_{\xi-\epsilon}^{\xi+\epsilon} d\varphi(x) = 0 \ (\xi - \epsilon < \eta < \xi + \epsilon)$$

or

$$t_0 + t_1\eta + \cdots + t_n\eta^n = 0$$

because

$$\int_{\xi-\epsilon}^{\xi+\epsilon} d\varphi(x) > 0.$$

Letting ϵ converge to 0, we conclude

$$t_0 + t_1 \xi + \cdots + t_n\xi^n = 0$$

at any point of increase. Since there are at least $n + 1$ points of increase the equation

$$t_0 + t_1 x + \cdots + t_n x^n = 0$$

would have at least $n + 1$ roots and that necessitates

$$t_0 = t_1 = \cdots = t_n = 0.$$

Hence, the quadratic form ϕ, which is never negative, can vanish only if all its variables vanish; that is, ϕ is a definite positive form. Its determinant Δ_n and all its principal minors $\Delta_{n-1}, \Delta_{n-2}, \ldots \Delta_0$ must be positive, which proves the first statement.

Suppose the conditions

$$\Delta_0 > 0, \qquad \Delta_1 > 0, \ldots \Delta_n > 0$$

satisfied and let $\varphi(x)$ have $s < n + 1$ points of increase. Then the integral representing ϕ reduces to a finite sum

$$\phi = p_1(t_0 + t_1\xi_1 + \cdots + t_n\xi_1^n)^2 + p_2(t_0 + t_1\xi_2 + \cdots + t_n\xi_2^n)^2 + \\ + \cdots + p_s(t_0 + t_1\xi_s + \cdots + t_n\xi_s^n)^2$$

denoting by $p_1, p_2, \ldots p_s$ masses concentrated in the s points of increase $\xi_1, \xi_2, \ldots \xi_s$. Now, since $s \leqq n$ constants $t_0, t_1, \ldots t_n$, not all zero, can be determined by the system of equations

$$\begin{aligned} t_0 + t_1\xi_1 + \cdots + t_n\xi_1^n &= 0 \\ t_0 + t_1\xi_2 + \cdots + t_n\xi_2^n &= 0 \\ \cdots\cdots & \\ t_0 + t_1\xi_s + \cdots + t_n\xi_s^n &= 0. \end{aligned}$$

Thus ϕ vanishes when not all variables vanish; hence, its determinant $\Delta_n = 0$, contrary to hypothesis.

From now on we shall assume that $\varphi(x)$ has at least $n + 1$ points of increase. The integral

$$\int_a^b \frac{d\varphi(x)}{z - x}$$

can be expanded into a formal power series of $1/z$, thus

$$\int_a^b \frac{d\varphi(x)}{z - x} = \frac{m_0}{z} + \frac{m_1}{z^2} + \frac{m_2}{z^3} + \cdots + \frac{m_{2n}}{z^{2n+1}} + \cdots$$

and this power series can be converted into a continued fraction as explained in Sec. 2. Let

$$\frac{P_1}{Q_1}, \frac{P_2}{Q_2}, \ldots \frac{P_n}{Q_n}, \frac{P_{n+1}}{Q_{n+1}}$$

be the first $n + 1$ convergents to that continued fraction. I say that the degrees of their denominators are, respectively, 1, 2, 3, . . . $n + 1$. Since these degrees form an increasing sequence, it suffices to show that there exists a convergent with the denominator of a given degree

$$s \leqq n + 1.$$

This convergent P/Q is completely determined by the condition that in a formal expansion of the difference

$$\int_a^b \frac{d\varphi(x)}{z-x} - \frac{P}{Q}$$

into a power series of $1/z$, terms involving $1/z, 1/z^2, \ldots 1/z^{2s}$ are absent. This is the same as to say that in the expansion of

$$Q(z)\int_a^b \frac{d\varphi(x)}{z-x} - P(z)$$

there are no terms involving $1/z, 1/z^2, \ldots 1/z^s$. The preceding expression can be written thus:

$$\int_a^b \frac{Q(x)d\varphi(x)}{z-x} + \int_a^b \frac{Q(z)-Q(x)}{z-x}d\varphi(x) - P(z) = \frac{A}{z^{s+1}} + \cdots.$$

Since

$$\int_a^b \frac{Q(z)-Q(x)}{z-x}d\varphi(x) - P(z)$$

is a polynomial in z, it must vanish identically. That gives

$$P(z) = \int_a^b \frac{Q(z)-Q(x)}{z-x}d\varphi(x). \tag{4}$$

To determine $Q(z)$ we must express the conditions that in the expansion of

$$\int_a^b \frac{Q(x)d\varphi(x)}{z-x}$$

terms in $1/z, 1/z^2, \ldots 1/z^s$ vanish. These conditions are equivalent to s relations

$$\int_a^b Q(x)d\varphi(x) = 0, \qquad \int_a^b xQ(x)d\varphi(x) = 0, \cdots \int_a^b x^{s-1}Q(x)d\varphi(x) = 0, \tag{5}$$

which in turn amount to the single requirement that

$$\int_a^b \theta(x)Q(x)d\varphi(x) = 0 \tag{6}$$

for an arbitrary polynomial $\theta(x)$ of degree $\leqq s - 1$.

Conversely, if there exists a polynomial $Q(z)$ of degree s satisfying conditions (5), and $P(z)$ is determined by equation (4), then $P(z)/Q(z)$ is a convergent whose denominator is of degree s. For then the expansion of

$$\int_a^b \frac{d\varphi(x)}{z-x} - \frac{P(z)}{Q(z)}$$

lacks the terms in $1/z, 1/z^2, \ldots 1/z^{2s}$.

Let

$$Q(z) = l_0 + l_1 z + l_2 z^2 + \cdots + l_{s-1} z^{s-1} + z^s.$$

Then equations (5) become

$$\begin{aligned} m_0 l_0 + m_1 l_1 + m_2 l_2 + \cdots + m_{s-1} l_{s-1} + m_s &= 0 \\ m_1 l_0 + m_2 l_1 + m_3 l_2 + \cdots + m_s l_{s-1} + m_{s+1} &= 0 \\ \cdots\cdots\cdots & \\ m_{s-1} l_0 + m_s l_1 + m_{s+1} l_2 + \cdots + m_{2s-2} l_{s-1} + m_{2s-1} &= 0. \end{aligned}$$

This system of linear equations determines completely the coefficients $l_0, l_1, \ldots l_{s-1}$ since its determinant $\Delta_{s-1} > 0$.

The existence of a convergent with the denominator of degree

$$s \leqq n + 1$$

being established, it follows that the denominator of the sth convergent P_s/Q_s is exactly of degree s. The denominator Q_s is determined, except for a constant factor, and can be presented in the form:

$$Q_s = \frac{c}{\Delta_{s-1}} \begin{vmatrix} 1 & z & z^2 & \cdots & z^s \\ m_0 & m_1 & m_2 & \cdots & m_s \\ m_1 & m_2 & m_3 & \cdots & m_{s+1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ m_{s-1} & m_s & m_{s+1} & \cdots & m_{2s-1} \end{vmatrix}.$$

A remarkable result follows from equation (6) by taking $Q = Q_s$ and $\theta = Q_{s'}$; namely,

$$(7) \qquad \int_a^b Q_s Q_{s'}\, d\varphi(x) = 0 \qquad \text{if} \qquad s \neq s'$$

while

$$\int_a^b Q_s^2 d\varphi(x) > 0 \qquad (s \leqq n).$$

In the general relation

$$Q_s = q_s Q_{s-1} - Q_{s-2}$$

the polynomial q_s must be of the first degree

$$q_s = \alpha_s z + \beta_s,$$

which shows that the continued fraction associated with

$$\int_a^b \frac{d\varphi(x)}{z - x}$$

has the form

$$\frac{1}{\alpha_1 z + \beta_1} - \frac{1}{\alpha_2 z + \beta_2} - \frac{1}{\alpha_3 z + \beta_3} - \cdots$$

The next question is, how to determine the constants α_s and β_s. Multiplying both members of the equation

$$Q_s = (\alpha_s z + \beta_s)Q_{s-1} - Q_{s-2} \qquad (s \geqq 2)$$

by $Q_{s-2}d\varphi(z)$, integrating between limits a and b, and taking into account (7), we get

$$0 = \alpha_s \int_a^b zQ_{s-1}Q_{s-2}d\varphi(z) - \int_a^b Q_{s-2}^2 d\varphi(z).$$

On the other hand, the highest terms in Q_{s-1} and Q_{s-2} are

$$\alpha_1\alpha_2 \cdots \alpha_{s-1}z^{s-1}, \qquad \alpha_1\alpha_2 \cdots \alpha_{s-2}z^{s-2}.$$

Hence,

$$zQ_{s-2} = \frac{1}{\alpha_{s-1}}Q_{s-1} + \psi$$

where ψ is a polynomial of degree $\leqq s - 2$. Referring to equation (6), we have

$$\int_a^b zQ_{s-2}Q_{s-1}d\varphi(z) = \frac{1}{\alpha_{s-1}}\int_a^b Q_{s-1}^2 d\varphi(z)$$

and consequently

$$\frac{\alpha_s}{\alpha_{s-1}} = \frac{\int_a^b Q_{s-2}^2 d\varphi(z)}{\int_a^b Q_{s-1}^2 d\varphi(z)}. \tag{8}$$

Suppose that the following moments are given: $m_0, m_1, \ldots m_{2n}$; how many of the coefficients α_s can be found? Evidently $\alpha_1 = 1/m_0$. Furthermore, $Q_0 = 1$ and Q_1 is completely determined given m_0 and m_1. Relation (8) determines α_2, and Q_2 will be completely determined given m_0, m_1, m_2, m_3. The same relation again determines α_3, and Q_3 will be determined given $m_0, m_1, \ldots m_5$. Proceeding in the same way, we conclude that, given $m_0, m_1, m_2, \ldots m_{2n}$, all the polynomials

$$Q_0, Q_1, Q_2, \ldots Q_n$$

as well as constants

$$\alpha_1, \alpha_2, \alpha_3, \ldots \alpha_{n+1}$$

can be determined. It is important to note that all these constants are positive.

Proceeding in a similar manner, the following expression can be found

$$\beta_s = -\alpha_s \frac{\int_a^b zQ_{s-1}^2 d\varphi(z)}{\int_a^b Q_{s-1}^2 d\varphi(z)}.$$

It follows that constants

$$\beta_1, \beta_2, \ . \ . \ . \ \beta_n$$

are determined by our data, but not β_{n+1}. For if $s = n + 1$, the integral

$$\int_a^b zQ_n^2 d\varphi(z)$$

can be expressed as a linear function of $m_0, m_1, \ . \ . \ . \ m_{2n+1}$ with known coefficients. But m_{2n+1} is not included among our data; hence, β_{n+1} cannot be determined.

4. Properties of Polynomials Q_s. Theorem. *Roots of the equation*

$$Q_s(z) = 0 \qquad (s \leqq n)$$

are real, simple, and contained within the interval (a, b).

Proof. Let $Q_s(z)$ change its sign $r < s$ times when z passes through points $z_1, z_2, \ . \ . \ . \ z_r$ contained strictly within (a, b). Setting

$$\theta(z) = (z - z_1)(z - z_2) \cdot \cdot \cdot (z - z_r)$$

the product

$$\theta(z)Q_s(z)$$

does not change its sign when z increases from a to b. However,

$$\int_a^b \theta(z)Q_s(z)d\varphi(z) = 0,$$

and this necessitates that

$$\theta(z)Q_s(z)$$

or $Q_s(z)$ vanishes in all points of increase of $\varphi(z)$. But this is impossible, since by hypothesis there are at least $n + 1$ points of increase, whereas the degree s of Q_s does not exceed n. Consequently, $Q_s(z)$ changes its sign in the interval (a, b) exactly s times and has all its roots real, simple, and located within (a, b).

It follows from this theorem that the convergent

$$\frac{P_n}{Q_n}$$

can be resolved into a sum of simple fractions as follows:

$$\frac{P_n(z)}{Q_n(z)} = \frac{A_1}{z - z_1} + \frac{A_2}{z - z_2} + \cdots + \frac{A_n}{z - z_n} \tag{9}$$

where $z_1, z_2, \ldots z_n$ are roots of the equation $Q_n(z) = 0$ and in general

$$A_k = \frac{P_n(z_k)}{Q'_n(z_k)}.$$

The right member of (9) can be expanded into power series of $1/z$, the coefficient of $1/z^k$ being

$$\sum_{\alpha=1}^{n} A_\alpha z_\alpha^{k-1}.$$

By the property of convergents we must have the following equations:

$$\sum_{\alpha=1}^{n} A_\alpha = m_0$$

$$\sum_{\alpha=1}^{n} A_\alpha z_\alpha = m_1$$

$$\cdots\cdots$$

$$\sum_{\alpha=1}^{n} A_\alpha z_\alpha^{2n-1} = m_{2n-1}.$$

These equations can be condensed into one,

$$\sum_{\alpha=1}^{n} A_\alpha T(z_\alpha) = \int_a^b T(z)d\varphi(z) \tag{10}$$

which should hold for any polynomial $T(z)$ of degree $\leqq 2n - 1$.

Let us take for $T(z)$ a polynomial of degree $2n - 2$:

$$T(z) = \left[\frac{Q_n(z)}{(z - z_\alpha)Q'_n(z_\alpha)}\right]^2.$$

Then

$$T(z_\alpha) = 1, \qquad T(z_\beta) = 0 \quad \text{if} \quad \beta \neq \alpha$$

and consequently, by virtue of equation (10),

$$A_\alpha = \int_a^b \left[\frac{Q_n(z)}{(z - z_\alpha)Q'_n(z_\alpha)}\right]^2 d\varphi(z) > 0.$$

Thus constants $A_1, A_2, \ldots A_n$ are all positive, which shows that $P_n(z_k)$

has the same sign as $Q'_n(z_k)$. Now in the sequence

$$Q'_n(z_1),\ Q'_n(z_2),\ .\ .\ .\ Q'_n(z_n)$$

any two consecutive terms are of opposite signs. The same being true of the sequence

$$P_n(z_1),\ P_n(z_2),\ .\ .\ .\ P_n(z_n),$$

it follows that the roots of $P_n(z)$ are all simple, real, and located in the intervals

$$(z_1,\ z_2);\ (z_2,\ z_3);\ .\ .\ .\ (z_{n-1},\ z_n).$$

Finally, we shall prove the following theorem:

Theorem. *For any real* x

$$Q'_n(x)Q_{n-1}(x) - Q'_{n-1}(x)Q_n(x)$$

is a positive number.

Proof. From the relations

$$\begin{aligned} Q_s(z) &= (\alpha_s z + \beta_s)Q_{s-1}(z) - Q_{s-2}(z) \\ Q_s(x) &= (\alpha_s x + \beta_s)Q_{s-1}(x) - Q_{s-2}(x) \end{aligned}$$

it follows that

$$\frac{Q_s(z)Q_{s-1}(x) - Q_s(x)Q_{s-1}(z)}{z - x} = \alpha_s Q_{s-1}(z)Q_{s-1}(x) + \\ + \frac{Q_{s-1}(z)Q_{s-2}(x) - Q_{s-1}(x)Q_{s-2}(z)}{z - x}$$

whence, taking $s = 1, 2, 3, \ldots n$ and adding results,

$$\frac{Q_n(z)Q_{n-1}(x) - Q_n(x)Q_{n-1}(z)}{z - x} = \sum_{s=1}^{n} \alpha_s Q_{s-1}(x)Q_{s-1}(z).$$

It suffices now to take $z = x$ to arrive at the identity

$$Q'_n(x)Q_{n-1}(x) - Q'_{n-1}(x)Q_n(x) = \sum_{s=1}^{n} \alpha_s Q_{s-1}(x)^2.$$

Since $Q_0 = 1$ and $\alpha_s > 0$, it is evident that

$$Q'_n(x)Q_{n-1}(x) - Q'_{n-1}(x)Q_n(x) > 0$$

for every real x.

5. Equivalent Point Distributions. If the whole mass can be concentrated in a finite number of points so as to produce the same l first moments as a given distribution, we have an "equivalent point distribution" in respect to the l first moments. In what follows we shall suppose that the whole mass is spread over an infinite interval $-\infty, \infty$ and that the given moments, originating in a distribution with at least $n+1$ points of increase, are

$$m_0, m_1, m_2, \dots m_{2n}.$$

The question is: Is it possible to find an equivalent point distribution where the whole mass is concentrated in $n+1$ points? Let the unknown points be

$$\xi_1, \xi_2, \dots \xi_{n+1}$$

and the masses concentrated in them

$$A_1, A_2, \dots A_{n+1}.$$

Evidently the question will be answered in the affirmative if the system of $2n+1$ equations

$$(A) \qquad \begin{aligned} \sum_{\alpha=1}^{n+1} A_\alpha &= m_0 \\ \sum_{\alpha=1}^{n+1} A_\alpha \xi_\alpha &= m_1 \\ \sum_{\alpha=1}^{n+1} A_\alpha \xi_\alpha^2 &= m_2 \\ &\cdots\cdots \\ \sum_{\alpha=1}^{n+1} A_\alpha \xi_\alpha^{2n} &= m_{2n} \end{aligned}$$

can be satisfied by real numbers $\xi_1, \xi_2, \dots \xi_{n+1}; A_1, A_2, \dots A_{n+1}$, the last $n+1$ numbers being *positive*. The number of unknowns being greater by one unit than the number of equations, we can introduce the additional requirement that one of the numbers $\xi_1, \xi_2, \dots \xi_{n+1}$ should be equal to a given real number v. The system (A) may be replaced by the single requirement that the equation

$$(11) \qquad \sum_{\alpha=1}^{n+1} A_\alpha T(\xi_\alpha) = \int_{-\infty}^{\infty} T(x)d\varphi(x)$$

shall hold for any polynomial $T(x)$ of degree $\leqq 2n$. Let $Q(x)$ be the polynomial of degree $n+1$ having roots $\xi_1, \xi_2, \dots \xi_{n+1}$ and let $\theta(x)$ be an arbitrary polynomial of degree $\leqq n-1$. Then we can apply equation (11) to

$$T(x) = \theta(x)Q(x).$$

Since $Q(\xi_\alpha) = 0$, we shall have

$$\int_{-\infty}^{\infty} \theta(x)Q(x)d\varphi(x) = 0 \tag{12}$$

for an arbitrary polynomial $\theta(x)$ of degree $\leqq n - 1$. Presently we shall see that requirement (12) together with $Q(v) = 0$ determines $Q(x)$, save for a constant factor if

$$Q_n(v) \neq 0.$$

Dividing $Q(x)$ by $Q_n(x)$, we have identically

$$Q(x) = (\lambda x + \mu)Q_n(x) + R_{n-1}(x)$$

where $R_{n-1}(x)$ is a polynomial of degree $\leqq n - 1$. If $\theta(x)$ is an arbitrary polynomial of degree $\leqq n - 2$,

$$(\lambda x + \mu)\theta(x)$$

will be of degree $\leqq n - 1$. Hence

$$\int_a^b (\lambda x + \mu)\theta(x)Q_n(x)d\varphi(x) = 0$$

by (6), and (12) shows that

$$\int_a^b \theta(x)R_{n-1}(x)d\varphi(x) = 0$$

for an arbitrary polynomial $\theta(x)$ of degree $\leqq n - 2$. The last requirement shows that $R_{n-1}(x)$ differs from $Q_{n-1}(x)$ by a constant factor. Since the highest coefficient in $Q(x)$ is arbitrary, we can set

$$R_{n-1}(x) = -Q_{n-1}(x).$$

In the equation

$$Q(x) = (\lambda x + \mu)Q_n(x) - Q_{n-1}(x)$$

it remains to determine constants λ and μ. Multiplying both members by $Q_{n-1}(x)d\varphi(x)$ and integrating between $-\infty$ and ∞, we get

$$\lambda \int_{-\infty}^{\infty} xQ_{n-1}Q_n d\varphi(x) = \int_{-\infty}^{\infty} Q_{n-1}^2 d\varphi(x)$$

or

$$\frac{\lambda}{\alpha_n}\int_{-\infty}^{\infty} Q_n^2 d\varphi(x) = \int_{-\infty}^{\infty} Q_{n-1}^2 d\varphi(x).$$

But

$$\frac{\int_{-\infty}^{\infty} Q_{n-1}^2 d\varphi(x)}{\int_{-\infty}^{\infty} Q_n^2 d\varphi(x)} = \frac{\alpha_{n+1}}{\alpha_n}$$

whence

$$\lambda = \alpha_{n+1}.$$

The equation

$$0 = Q(v) = (\alpha_{n+1}v + \mu)Q_n(v) - Q_{n-1}(v)$$

serves to determine μ if $Q_n(v) \neq 0$. The final expression of $Q(x)$ will be

$$Q(x) = \left(\alpha_{n+1}(x - v) + \frac{Q_{n-1}(v)}{Q_n(v)}\right)Q_n(x) - Q_{n-1}(x).$$

Owing to recurrence relations

$$Q_2 = (\alpha_2 x + \beta_2)Q_1 - Q_0;\ Q_3 = (\alpha_3 x + \beta_3)Q_2 - Q_1;\ \cdots$$
$$Q_n = (\alpha_n x + \beta_n)Q_{n-1} - Q_{n-2},$$

it is evident that

$$Q,\ Q_n,\ Q_{n-1},\ \ldots\ Q_1,\ Q_0 = 1$$

in a Sturm series. For $x = -\infty$, it contains $n + 1$ variations and for $x = \infty$ only permanences. It follows that the equation

$$Q(x) = 0$$

has exactly $n + 1$ distinct real roots and among them v. Thus, if the problem is solvable, the numbers $\xi_1,\ \xi_2,\ \ldots\ \xi_{n+1}$ are determined as roots of

$$Q(x) = 0.$$

Furthermore, all unknowns A_α will be positive. In fact, from equation (11) it follows that

$$A_\alpha = \int_{-\infty}^{\infty}\left[\frac{Q(x)}{(x - \xi_\alpha)Q'(\xi_\alpha)}\right]^2 d\varphi(x) > 0.$$

Now we must show that constants A_α can actually be determined so as to satisfy equations (A). To this end let

$$P(x) = \int_{-\infty}^{\infty}\frac{Q(x) - Q(z)}{x - z}d\varphi(z) = \left[\alpha_{n+1}(x - v) + \frac{Q_{n-1}(v)}{Q_n(v)}\right]P_n(x) - P_{n-1}(x).$$

Then

$$Q(x)\int_{-\infty}^{\infty}\frac{d\varphi(z)}{x - z} - P(x) = \int_{-\infty}^{\infty}\frac{Q(z)d\varphi(z)}{x - z}$$

and, on account of (12), the expansion of the right member into power series of $1/x$ lacks the terms in $1/x,\ 1/x^2,\ \ldots\ 1/x^n$. Hence, the expansion of

$$\int_{-\infty}^{\infty}\frac{d\varphi(z)}{x - z} - \frac{P(x)}{Q(x)}$$

lacks the terms in $1/x$, $1/x^2$, . . . $1/x^{2n+1}$; that is,

$$\frac{P(x)}{Q(x)} = \frac{m_0}{x} + \frac{m_1}{x^2} + \cdots + \frac{m_{2n}}{x^{2n+1}} + \cdots .$$

On the other hand, resolving in simple fractions,

$$\frac{P(x)}{Q(x)} = \frac{A_1}{x - \xi_1} + \frac{A_2}{x - \xi_2} + \cdots + \frac{A_{n+1}}{x - \xi_{n+1}}.$$

Expanding the right member into power series of $1/x$ and comparing with the preceding expansion, we obtain the system (A). By the previous remark all constants A_α are positive. Thus, there exists a point distribution in which masses concentrated in $n + 1$ points produce moments $m_0, m_1, \ldots m_{2n}$. One of these points v may be taken arbitrarily, with the condition

$$Q_n(v) \neq 0$$

being observed, however.

6. Tshebysheff's Inequalities. In a note referred to in the introduction Tshebysheff made known certain inequalities of the utmost importance for the theory we are concerned with. The first very ingenious proof of them was given by Markoff in 1884 and, by a remarkable coincidence, the same proof was rediscovered almost at the same time by Stieltjes. A few years later, Stieltjes found another totally different proof; and it is this second proof that we shall follow.

Let $\varphi(x)$ be a distribution function of a mass spread over the interval $-\infty, \infty$. Supposing that a moment of the order i,

$$\int_{-\infty}^{\infty} x^i d\varphi(x) = m_i,$$

exists, we shall show first that

$$\lim l^i(m_0 - \varphi(l)) = 0$$
$$\lim l^i\varphi(-l) = 0$$

when l tends to $+\infty$. For

$$\int_l^{\infty} x^i d\varphi(x) \geqq l^i\int_l^{\infty} d\varphi(x) = l^i[\varphi(+\infty) - \varphi(l)]$$

or

$$l^i(m_0 - \varphi(l)) \leqq \int_l^{\infty} x^i d\varphi(x).$$

Similarly

$$\left|\int_{-\infty}^{-l} x^i d\varphi(x)\right| \geqq l^i\int_{-\infty}^{-l} d\varphi(x) = l^i\varphi(-l)$$

or

$$l^i\varphi(-l) \leqq \left|\int_{-\infty}^{-l} x^i d\varphi(x)\right|.$$

Now both integrals

$$\int_l^{\infty} x^i d\varphi(x) \quad \text{and} \quad \int_{-\infty}^{-l} x^i d\varphi(x)$$

converge to 0 as l tends to $+\infty$; whence both statements follow immediately. Integrating by parts, we have

$$\int_0^l x^i d\varphi(x) = l^i[\varphi(l) - m_0] - i\int_0^l [\varphi(x) - m_0]x^{i-1}dx$$
$$\int_{-l}^0 x^i d\varphi(x) = (-1)^{i-1}l^i\varphi(-l) - i\int_{-l}^0 x^{i-1}\varphi(x)dx,$$

whence, letting l converge to $+\infty$,

$$m_i = \int_{-\infty}^{\infty} x^i d\varphi(x) = -i\int_0^{\infty}[\varphi(x) - m_0]x^{i-1}dx - i\int_{-\infty}^0 x^{i-1}\varphi(x)dx.$$

If the same mass m_0, with the same moment m_i, is spread according to the law characterized by the function $\psi(x)$, we shall have

$$m_i = \int_{-\infty}^{\infty} x^i d\psi(x) = -i\int_0^{\infty}[\psi(x) - m_0]x^{i-1}dx - i\int_{-\infty}^0 x^{i-1}\psi(x)dx,$$

whence

$$\int_{-\infty}^{\infty} x^{i-1}[\varphi(x) - \psi(x)]dx = 0. \tag{13}$$

Suppose the moments

$$m_0, m_1, m_2, \ldots m_{2n}$$

of the distribution characterized by $\varphi(x)$ are known. Provided $\varphi(x)$ has at least $n + 1$ points of increase, there exists an equivalent point distribution, defined in Sec. 5 and characterized by the step function $\psi(x)$ which can be defined as follows:

$$\begin{array}{lll}
\psi(x) = 0 & \text{for} & -\infty < x < \xi_1 \\
\psi(x) = A_1 & \text{for} & \xi_1 \leqq x < \xi_2 \\
\psi(x) = A_1 + A_2 & \text{for} & \xi_2 \leqq x < \xi_3 \\
\cdots & & \\
\psi(x) = A_1 + A_2 + \cdots + A_n & \text{for} & \xi_n \leqq x < \xi_{n+1} \\
\psi(x) = A_1 + A_2 + \cdots + A_{n+1} & \text{for} & \xi_{n+1} \leqq x < +\infty,
\end{array}$$

provided roots $\xi_1, \xi_2, \ldots \xi_{n+1}$ of the equation $Q(x) = 0$ are arranged in an increasing order of magnitude.

Equation (13) will hold for $i = 1, 2, 3, \ldots 2n$ or, which is the same, the equation

$$\int_{-\infty}^{\infty} \theta(x)[\varphi(x) - \psi(x)]dx = 0 \tag{14}$$

will hold for an arbitrary polynomial $\theta(x)$ of degree $\leqq 2n - 1$. The function

$$h(x) = \varphi(x) - \psi(x)$$

in general has ordinary discontinuities. We can prove now that $h(x)$, if not identically equal to 0 at all points of continuity, changes its sign at least $2n$ times.[1] Suppose, on the contrary, that it changes sign $r < 2n$ times; namely, at the points

$$a_1, a_2, \ldots a_r.$$

Taking

$$\theta(x) = (x - a_1)(x - a_2) \cdots (x - a_r),$$

equation (14) will be satisfied, while the integrand

$$\theta(x)h(x),$$

if not 0, will be of the same sign, for example, positive. Let ξ be any point of continuity of $h(x)$. If $\xi = a_i$ $(i = 1, 2, \ldots r)$ then $h(a_i) = 0$ since $h(x)$ changes sign at a_i. If ξ does not coincide with any one of the numbers $a_1, a_2, \ldots a_r$, then for an arbitrarily small positive ϵ we must have

$$\int_{\xi-\epsilon}^{\xi+\epsilon} \theta(x)h(x)dx = 0.$$

But by continuity

$$\theta(x)h(x)$$

remains in the interval $(\xi - \epsilon, \xi + \epsilon)$ for sufficiently small ϵ above a certain positive number unless $h(\xi) = 0$. Thus, if $h(x)$ does not vanish at all points of continuity (in which case $\varphi(x)$ and $\psi(x)$ do not differ essentially), it must change sign at least $2n$ times. Let us see now where the change of sign can occur. In the intervals

$$-\infty, \xi_1 \text{ and } \xi_{n+1}, +\infty$$

[1] A function $f(x)$ is said to change sign once in (a, b) if in this interval there exists a point or points c such that, for instance, $f(x) \geqq 0$ in (a, c) and $f(x) \leqq 0$ in (c, b), equality signs not holding throughout the respective intervals. The change of sign occurs n times if (a, b) can be divided in n intervals in which $f(x)$ changes sign once.

$\varphi(x) - \psi(x)$ evidently cannot change sign. Within each of the intervals

$$\xi_{i-1},\ \xi_i$$

there can be at most one change of sign, since $\psi(x)$ remains constant there, and $\varphi(x)$ can only increase. The sign may change also at the points of discontinuity of $\psi(x)$; that is, at the points $\xi_1, \xi_2, \ldots \xi_{n+1}$. Altogether, $\varphi(x) - \psi(x)$ cannot change sign more than $2n + 1$ times and not less than $2n$ times.

Since $\psi(x) = 0$ so far as $x < \xi_1$ and $\varphi(\xi_1 - \epsilon)$ is not negative for positive ϵ, we must have

$$\varphi(\xi_1 - \epsilon) - \psi(\xi_1 - \epsilon) \geqq 0.$$

Also $\psi(x) = m_0$ for $x > \xi_{n+1}$ and $\varphi(x) \leqq m_0$, so that

$$\varphi(\xi_{n+1} + \epsilon) - \psi(\xi_{n+1} + \epsilon) \leqq 0.$$

At first let us suppose

$$\varphi(\xi_1 - \epsilon) - \psi(\xi_1 - \epsilon) > 0, \qquad \varphi(\xi_{n+1} + \epsilon) - \psi(\xi_{n+1} + \epsilon) < 0.$$

In this case $\varphi(x) - \psi(x)$ must change sign an *odd* number of times; that is, not less than $2n + 1$ times. Since this cannot happen more than $2n + 1$ times, the number of times $\varphi(x) - \psi(x)$ changes its sign must be exactly $2n + 1$. These changes occur once within each interval

$$\xi_{i-1},\ \xi_i$$

and in each of the points $\xi_1, \xi_2, \ldots \xi_{n+1}$. When the change of sign occurs in the interval (ξ_{i-1}, ξ_i) where $\psi(x)$ remains constant, because $\varphi(x)$ never decreases, we must have for sufficiently small ϵ

$$\varphi(\xi_i - \epsilon) - \psi(\xi_i - \epsilon) > 0. \tag{15}$$

But the sign changes in passing the point ξ_i; therefore,

$$\varphi(\xi_i + \epsilon) - \psi(\xi_i + \epsilon) < 0. \tag{16}$$

The equalities

$$\varphi(\xi_1 - \epsilon) - \psi(\xi_1 - \epsilon) = 0, \qquad \varphi(\xi_{n+1} + \epsilon) - \psi(\xi_{n+1} + \epsilon) = 0$$

cannot both hold for all sufficiently small ϵ. For then there would not be a change of sign at ξ_1 and ξ_{n+1}, so that the number of changes would not be greater than $2n - 1$ which is impossible. Therefore, let

$$\varphi(\xi_1 - \epsilon) - \psi(\xi_1 - \epsilon) = 0 \quad \text{and} \quad \varphi(\xi_{n+1} + \epsilon) - \psi(\xi_{n+1} + \epsilon) < 0.$$

Then there will be exactly $2n$ changes of sign: one in each of the intervals

$$\xi_{i-1},\ \xi_i$$

and in each of the points $\xi_2, \xi_3, \ldots \xi_{n+1}$. The inequalities (15) and (16) would hold for $i \geqq 2$, but

$$\varphi(\xi_1 - \epsilon) - \psi(\xi_1 - \epsilon) = 0, \qquad \varphi(\xi_1 + \epsilon) - \psi(\xi_1 + \epsilon) < 0$$

for all sufficiently small ϵ.

Now let

$$\varphi(\xi_{n+1} + \epsilon) - \psi(\xi_{n+1} + \epsilon) = 0 \qquad \text{and} \qquad \varphi(\xi_1 - \epsilon) - \psi(\xi_1 - \epsilon) > 0$$

for all sufficiently small positive ϵ. Then there will be exactly $2n$ changes of sign: In each of the points $\xi_1, \xi_2, \ldots \xi_n$ and in each of the n intervals

$$\xi_{i-1}, \ \xi_i.$$

The inequalities (15) and (16) will again hold for $i \leqq n$, but

$$\varphi(\xi_{n+1} - \epsilon) - \psi(\xi_{n+1} - \epsilon) > 0 \qquad \text{and} \qquad \varphi(\xi_{n+1} + \epsilon) - \psi(\xi_{n+1} + \epsilon) = 0$$

for all sufficiently small ϵ. Letting ϵ converge to 0, we shall have

$$\varphi(\xi_i - 0) \geqq \psi(\xi_i - 0)$$
$$\varphi(\xi_i + 0) \leqq \psi(\xi_i + 0)$$

for $i = 1, 2, 3, \ldots n + 1$ in all cases. Then, since

$$\varphi(\xi_i) \geqq \varphi(\xi_i - 0); \qquad \varphi(\xi_i) \leqq \varphi(\xi_i + 0),$$

we shall have also

$$\varphi(\xi_i) \geqq \psi(\xi_i - 0)$$
$$\varphi(\xi_i) \leqq \psi(\xi_i + 0)$$

or, taking into consideration the definition of the function $\psi(x)$

$$\varphi(\xi_i) \geqq \sum_{l=1}^{i-1} \frac{P(\xi_l)}{Q'(\xi_l)}$$

$$\varphi(\xi_i) \leqq \sum_{l=1}^{i} \frac{P(\xi_l)}{Q'(\xi_l)}.$$

These are the inequalities to which Tshebysheff's name is justly attached. For a particular root $\xi_i = v$ they can be written thus:

$$\varphi(v) \geqq \sum_{\xi_l < v} \frac{P(\xi_l)}{Q'(\xi_l)}$$

(17)

$$\varphi(v) \leqq \sum_{\xi_l \leqq v} \frac{P(\xi_l)}{Q'(\xi_l)}$$

with the evident meaning of the extent of summations. Another, less explicit, form of the same inequalities is

$$\begin{aligned} \varphi(v) &\geqq \psi(v - 0) \\ \varphi(v) &\leqq \psi(v + 0). \end{aligned} \tag{18}$$

As to $P(x)$ and $Q(x)$, they can be taken in the form:

$$\begin{aligned} P(x) &= [\alpha_{n+1}(x - v)Q_n(v) + Q_{n-1}(v)]P_n(x) - Q_n(v)P_{n-1}(x) \\ Q(x) &= [\alpha_{n+1}(x - v)Q_n(v) + Q_{n-1}(v)]Q_n(x) - Q_n(v)Q_{n-1}(x). \end{aligned}$$

Thus far we have assumed that v was different from any root of the equation

$$Q_n(x) = 0,$$

but all the results hold, even if

$$Q_n(v) = 0.$$

To prove this, we note first that when a variable v approaches a root ξ of $Q_n(x)$, one root of $Q(x)$ (either ξ_1 or ξ_{n+1}) tends to $-\infty$ or $+\infty$, while the remaining n roots approach the n roots $x_1, x_2, \ldots x_n$ of the equation

$$Q_n(x) = 0.$$

If ξ_1 tends to negative infinity, it is easy to see that

$$\frac{P(\xi_1)}{Q'(\xi_1)}$$

tends to 0. In this case the other quotients

$$\frac{P(\xi_l)}{Q'(\xi_l)}$$

tend respectively to

$$\frac{P_n(x_1)}{Q'_n(x_1)}, \frac{P_n(x_2)}{Q'_n(x_2)}, \ldots.$$

If ξ_{n+1} tends to positive infinity the quotients

$$\frac{P(\xi_l)}{Q'(\xi_l)}; \; l = 1, 2, \ldots n$$

approach respectively

$$\frac{P_n(x_l)}{Q'_n(x_l)}; \; l = 1, 2, 3, \ldots n,$$

while

$$\frac{P(\xi_{n+1})}{Q'(\xi_{n+1})}$$

tends to 0. Now take $v = \xi - \epsilon$ and $v = \xi + \epsilon$ in (17) and let the positive number ϵ converge to 0. Taking into account the preceding remarks, we find in the limit

$$\varphi(\xi - 0) \geqq \sum_{x_l < \xi} \frac{P_n(x_l)}{Q_n'(x_l)}$$

$$\varphi(\xi + 0) \leqq \sum_{x_l \leqq \xi} \frac{P_n(x_l)}{Q_n'(x_l)},$$

whence again

$$\varphi(\xi) \geqq \sum_{x_l < \xi} \frac{P_n(x_l)}{Q_n'(x_l)}$$

$$\varphi(\xi) \leqq \sum_{x_l \leqq \xi} \frac{P_n(x_l)}{Q_n'(x_l)}.$$

But these inequalities follow directly from (17) by taking $v = \xi$.

Since

$$\psi(v + 0) - \psi(v - 0) = \frac{P(v)}{Q'(v)}$$

it follows from inequalities (18) that

$$0 \leqq \varphi(v) - \psi(v - 0) \leqq \frac{P(v)}{Q'(v)}.$$

On the other hand, one easily finds that

$$\frac{P(v)}{Q'(v)} = \frac{1}{\alpha_{n+1}Q_n(v)^2 + Q_n'(v)Q_{n-1}(v) - Q_{n-1}'(v)Q_n(v)}.$$

But referring to the end of Sec. 4,

$$Q_n'(v)Q_{n-1}(v) - Q_{n-1}'(v)Q_n(v) = \sum_{s=1}^{n} \alpha_s Q_{s-1}(v)^2,$$

whence

$$\alpha_{n+1}Q_n(v)^2 + Q_n'(v)Q_{n-1}(v) - Q_{n-1}'(v)Q_n(v) = Q_{n+1}'(v)Q_n(v) - Q_n'(v)Q_{n+1}(v).$$

Finally,

$$0 \leqq \varphi(v) - \psi(v - 0) \leqq \frac{1}{Q_{n+1}'(v)Q_n(v) - Q_n'(v)Q_{n+1}(v)}.$$

If $\varphi_1(v)$ is another distribution function with the same moments

$$m_0, m_1, m_2, \ldots m_{2n},$$

we shall have also

$$0 \leqq \varphi_1(v) - \psi(v - 0) \leqq \frac{1}{Q'_{n+1}(v)Q_n(v) - Q'_n(v)Q_{n+1}(v)},$$

and as a consequence,

(19) $$|\varphi_1(v) - \varphi(v)| \leqq \chi_n(v)$$

—a very important inequality. Here for brevity we use the notation

$$\chi_n(v) = \frac{1}{Q'_{n+1}(v)Q_n(v) - Q'_n(v)Q_{n+1}(v)}.$$

7. Application to Normal Distribution. An important particular case is that of a normal distribution characterized by the function

$$\varphi(x) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{x} e^{-u^2}du.$$

In this case it is easy to give an explicit expression of the polynomials $Q_n(x)$. Let

$$H_n(x) = e^{x^2}\frac{d^n e^{-x^2}}{dx^n}.$$

Integrating by parts, one can prove that for $l \leqq n - 1$

$$\int_{-\infty}^{\infty} e^{-x^2}x^l H_n(x)dx = 0.$$

Hence, one may conclude that $Q_n(x)$ differs from $H_n(x)$ by a constant factor. Let

$$Q_n(x) = c_n H_n(x).$$

To determine c_n, we may use the relation

$$H_n(x) = -2xH_{n-1}(x) - 2(n - 1)H_{n-2}(x)$$

which can readily be established. Introducing polynomials Q_n, this relation becomes

$$Q_n(x) = -2x\frac{c_n}{c_{n-1}}Q_{n-1}(x) - 2(n - 1)\frac{c_n}{c_{n-2}}Q_{n-2}(x).$$

Hence,

$$\frac{c_n}{c_{n-2}} = \frac{1}{2n - 2}, \qquad \alpha_n = -2\frac{c_n}{c_{n-1}}, \qquad \beta_n = 0.$$

Since $H_0(x) = Q_0(x) = 1$, we have $c_0 = 1$; also

$$\alpha_1 = \frac{1}{m_0} = 1 = -2\frac{c_1}{c_0}$$

whence $c_1 = -\frac{1}{2}$. The knowledge of c_0 and c_1 together with the relation

$$c_n = \frac{c_{n-2}}{2n-2}$$

allows determination of all members of the sequence $c_2, c_3, c_4, \ldots$. The final expressions are as follows;

$$c_{2m} = \frac{1}{2^m \cdot 1 \cdot 3 \cdot 5 \cdots (2m-1)}$$

$$c_{2m+1} = \frac{-1}{2^{m+1} \cdot 2 \cdot 4 \cdot 6 \cdots 2m}.$$

From the above relation between $H_n(x)$, $H_{n-1}(x)$, $H_{n-2}(x)$ and owing to the fact that $H_n(x)$ is an even or odd polynomial, according as n is even or odd, one finds

$$H_{2m}(0) = (-2)^m \cdot 1 \cdot 3 \cdot 5 \cdots (2m-1),$$

while another relation

$$H_n'(x) = -2nH_{n-1}(x),$$

following from the definition of $H_n(x)$, gives

$$H'_{2m-1}(0) = (-2)^m \cdot 1 \cdot 3 \cdot 5 \cdots (2m-1).$$

These preliminaries being established, we shall prove now that

$$\chi_n(v) = \frac{1}{c_n c_{n+1}(H'_{n+1}(v)H_n(v) - H'_n(v)H_{n+1}(v))}$$

attains its maximum for $v = 0$. Let

$$\Omega(v) = H'_{n+1}(v)H_n(v) - H'_n(v)H_{n+1}(v).$$

Then, taking into account the differential equation for polynomials $H_n(v)$:

$$H_n''(v) = 2vH_n'(v) - 2nH_n(v)$$

we find that

$$\frac{d\Omega}{dv} = 2v\Omega - 2H_n(v)H_{n+1}(v).$$

On the other hand,

$$\Omega = -H_{n+1}(v)^2 \frac{d}{dv}\frac{H_n(v)}{H_{n+1}(v)},$$

and denoting roots of the polynomial $H_{n+1}(v)$ in general by ξ,

$$\frac{d}{dv}\frac{H_n(v)}{H_{n+1}(v)} = -\sum \frac{H_n(\xi)}{H'_{n+1}(\xi)} \frac{1}{(v-\xi)^2}.$$

Consequently

$$\Omega = H_{n+1}(v)^2 \sum \frac{H_n(\xi)}{H'_{n+1}(\xi)} \frac{1}{(v-\xi)^2}.$$

Again

$$H_n(v)H_{n+1}(v) = H_{n+1}(v)^2 \sum \frac{H_n(\xi)}{H'_{n+1}(\xi)} \frac{v-\xi}{(v-\xi)^2},$$

and so

$$\frac{d\Omega}{dv} = 2H_{n+1}(v)^2 \sum \frac{H_n(\xi)}{H'_{n+1}(\xi)} \frac{\xi}{(v-\xi)^2} = \frac{-H_{n+1}(v)^2}{n+1} \sum \frac{\xi}{(v-\xi)^2}.$$

Roots of the polynomial $H_{n+1}(x)$ being symmetrically located with respect to 0, we have:

$$\sum \frac{\xi}{(v-\xi)^2} = -\sum \frac{\xi}{(v+\xi)^2} = 2v \sum \frac{\xi^2}{(v^2-\xi^2)^2},$$

and finally

$$\frac{d\Omega}{dv} = -2v\frac{H_{n+1}(v)^2}{n+1} \sum \frac{\xi^2}{(v^2-\xi^2)^2}.$$

Hence

$$\frac{d\Omega}{dv} > 0 \quad \text{if} \quad v < 0; \qquad \frac{d\Omega}{dv} < 0 \quad \text{if} \quad v > 0$$

that is, $\Omega(v)$ attains its maximum for $v = 0$ and $\chi_n(v)$ attains its maximum for $v = 0$. Referring to the above expressions of c_{2m}, c_{2m+1}; $H_{2m}(0)$, $H'_{2m+1}(0)$, we find that

$$\chi_{2m}(0) = \frac{2 \cdot 4 \cdot 6 \cdots 2m}{3 \cdot 5 \cdot 7 \cdots (2m+1)}$$

$$\chi_{2m+1}(0) = \frac{2 \cdot 4 \cdot 6 \cdots 2m}{3 \cdot 5 \cdot 7 \cdots (2m+1)}.$$

In Appendix I, page 354, we find the inequality

$$\frac{2 \cdot 4 \cdot 6 \cdots 2m}{1 \cdot 3 \cdot 5 \cdots (2m-1)} \frac{1}{\sqrt{4m+2}} < \frac{\sqrt{\pi}}{2}$$

whence

$$\frac{2 \cdot 4 \cdot 6 \cdots 2m}{3 \cdot 5 \cdot 7 \cdots (2m+1)} < \sqrt{\frac{\pi}{4m+2}}.$$

Thus, in all cases

$$\chi_n(v) \leqq \chi_n(0) < \sqrt{\frac{\pi}{2n}},$$

whence, by virtue of inequality (19),

$$|\varphi_1(v) - \varphi(v)| < \sqrt{\frac{\pi}{2n}}.$$

Thus any distribution function $\varphi_1(v)$ with the moments

$$m_0 = 1, \qquad m_{2k-1} = 0, \qquad m_{2k} = \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{2^k} \qquad (k \leqq n)$$

corresponding to

$$\varphi(v) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{v} e^{-u^2} du$$

differs from $\varphi(v)$ by less than

$$\sqrt{\frac{\pi}{2n}}.$$

Since this quantity tends to 0 when n increases indefinitely, we have the following theorem proved for the first time by Tshebysheff:

The system of infinitely many equations

$$\int_{-\infty}^{\infty} d\varphi(x) = 1; \qquad \int_{-\infty}^{\infty} x^{2k-1} d\varphi(x) = 0; \qquad \int_{-\infty}^{\infty} x^{2k} d\varphi(x) = \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{2^k}$$

$$k = 1, 2, 3, \ldots$$

uniquely determines a never decreasing function $\varphi(x)$ *such that* $\varphi(-\infty) = 0$; *namely,*

$$\varphi(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-u^2} du.$$

8. Tshebysheff-Markoff's Fundamental Theorem. When a mass $= 1$ is distributed according to the law characterized by a function $F(x, \lambda)$ depending upon a parameter λ, we say that the distribution is variable. Notwithstanding the variability of distribution, it may happen that its moments remain constant. If they are equal to moments of normal distribution with density

$$\frac{e^{-x^2}}{\sqrt{\pi}}$$

then by the preceding theorem we have rigorously

$$F(x, \lambda) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-u^2} du$$

no matter what λ is.

Generally moments of a variable distribution are themselves variable. Suppose that each one of them, when λ tends to a certain limit (for instance ∞), tends to the corresponding moment of normal distribution. One can foresee that under such circumstances $F(x, \lambda)$ will tend to

$$\varphi(x) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{x} e^{-u^2}du.$$

In fact, the following fundamental theorem holds:

Fundamental Theorem. *If, for a variable distribution characterized by the function* $F(x, \lambda)$,

$$\lim \int_{-\infty}^{\infty} x^k dF(x, \lambda) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-x^2}x^k dx; \qquad \lambda \to \infty$$

for any fixed $k = 0, 1, 2, 3, \ldots$, *then*

$$\lim F(v, \lambda) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{v} e^{-x^2}dx; \qquad \lambda \to \infty$$

uniformly in v.

Proof. Let

$$m_0, m_1, m_2, \ldots m_{2n}$$

be $2n + 1$ moments corresponding to a normal distribution. They allow formation of the polynomials

$$Q_0(x), Q_1(x), \ldots Q_n(x) \text{ and } Q(x)$$

and the function designated in Sec. 6 by $\psi(x)$. Similar entities corresponding to the variable distribution will be specified by an asterisk. Since

$$m_k^* \to m_k \qquad \text{as} \qquad \lambda \to \infty$$

and since $\Delta_n > 0$, we shall have

$$\Delta_n^* > 0$$

for sufficiently large λ. Then $F(x, \lambda)$ will have not less than $n + 1$ points of increase and the whole theory can be applied to variable distribution. In particular, we shall have

$$0 \leq \varphi(v) - \psi(v - 0) \leq \chi_n(v)$$

(20)

$$0 \leq F(v, \lambda) - \psi^*(v - 0) \leq \chi_n^*(v).$$

Now $Q_s^*(x)(s = 0, 1, 2, \ldots n)$ and $Q^*(x)$ depend rationally upon

$m_k^*(k = 0, 1, 2, \ldots 2n)$; hence, without any difficulty one can see that

$$Q_s^*(x) \to Q_s(x); \qquad s = 0, 1, 2, \ldots n$$
$$Q^*(x) \to Q(x)$$

as $\lambda \to \infty$; whence,

$$\chi_n^*(v) \to \chi_n(v).$$

Again

$$\psi^*(v - 0) \to \psi(v - 0)$$

as $\lambda \to \infty$. A few explanations are necessary to prove this. At first let $Q_n(v) \neq 0$. Then the polynomial $Q(x)$ will have $n + 1$ roots

$$\xi_1 < \xi_2 < \xi_3 < \cdots < \xi_{n+1}.$$

Since the roots of an algebraic equation vary continuously with its coefficients, it is evident that for sufficiently large λ the equation

$$Q^*(x) = 0$$

will have $n + 1$ roots:

$$\xi_1^* < \xi_2^* < \xi_3^* < \cdots < \xi_{n+1}^*$$

and ξ_k^* will tend to ξ_k as $\lambda \to \infty$. In this case, it is evident that $\psi^*(v - 0)$ will tend to $\psi(v - 0)$. If $Q_n(v) = 0$, it may happen that ξ_1^* or ξ_{n+1}^* tends respectively to $-\infty$ or $+\infty$ as $\lambda \to \infty$, while the other roots tend to the roots

$$x_1, x_2, \ldots x_n$$

of the equation

$$Q_n(x) = 0.$$

But the terms in $\psi^*(v - 0)$ corresponding to infinitely increasing roots tend to 0, and again

$$\psi^*(v - 0) \to \psi(v - 0).$$

Now

$$\chi_n(v) < \sqrt{\frac{\pi}{2n}}.$$

Consequently, given an arbitrary positive number ϵ, we can select n so large as to have

$$\chi_n(v) < \sqrt{\frac{\pi}{2n}} < \epsilon.$$

Having selected n in this manner, we shall keep it fixed. Then by the preceding remarks a number L can be found so that

$$\chi^*(v) < \sqrt{\frac{\pi}{2n}} < \epsilon$$
$$|\psi(v-0) - \psi^*(v-0)| < \epsilon$$

for $\lambda > L$. Combining this with inequalities (20), we find

$$|F(v, \lambda) - \varphi(v)| < 3\epsilon$$

for $\lambda > L$. And this proves the convergence of $F(v, \lambda)$ to $\varphi(v)$ for a fixed arbitrary v. To show that the equation

$$\lim F(v, \lambda) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{v} e^{-x^2}dx$$

holds uniformly for a variable v we can follow a very simple reasoning due to Pólya. Since $\varphi(-\infty) = 0$, $\varphi(+\infty) = 1$ and $\varphi(x)$ is an increasing function, one can determine two numbers a_0 and a_n so that

$$\varphi(x) \leqq \varphi(a_0) < \frac{\epsilon}{2} \quad \text{for} \quad x \leqq a_0$$
$$1 - \varphi(x) \leqq 1 - \varphi(a_n) < \frac{\epsilon}{2} \quad \text{for} \quad x \geqq a_n.$$

Next, because $\varphi(x)$ is a continuous function, the interval (a_0, a_n) can be subdivided into partial intervals by inserting between a_0 and a_n points $a_1 < a_2 < \cdots < a_{n-1}$ so that

$$0 < \varphi(a_{k+1}) - \varphi(a_k) < \frac{\epsilon}{2}$$

for $k = 0, 1, 2, \ldots n - 1$. By the preceding result, for all sufficiently large λ

$$F(a_0, \lambda) < \frac{\epsilon}{2}; \qquad 1 - F(a_n, \lambda) < \frac{\epsilon}{2}$$

and

$$|F(a_k, \lambda) - \varphi(a_k)| < \frac{\epsilon}{2}; \qquad k = 1, 2, \ldots n - 1.$$

Now consider the interval $(-\infty, a_0)$. Here for $v \leqq a_0$

$$0 \leqq F(v, \lambda) < \frac{\epsilon}{2}; \qquad 0 < \varphi(v) < \frac{\epsilon}{2}$$

and

$$|F(v, \lambda) - \varphi(v)| < \epsilon.$$

For v belonging to the interval $(a_n, +\infty)$

$$0 \leqq 1 - F(v, \lambda) < \frac{\epsilon}{2}, \qquad 0 < 1 - \varphi(v) < \frac{\epsilon}{2},$$

whence again

$$|F(v, \lambda) - \varphi(v)| < \epsilon.$$

Finally, let

$$a_k \leqq v < a_{k+1} \qquad (k = 0, 1, 2, \ldots n - 1).$$

Then

$$\begin{aligned} F(v, \lambda) - \varphi(v) &\geqq F(a_k, \lambda) - \varphi(a_{k+1}) = \\ &= [F(a_k, \lambda) - \varphi(a_k)] + [\varphi(a_k) - \varphi(a_{k+1})] \\ F(v, \lambda) - \varphi(v) &\leqq F(a_{k+1}, \lambda) - \varphi(a_k) = \\ &= [F(a_{k+1}, \lambda) - \varphi(a_{k+1})] + [\varphi(a_{k+1}) - \varphi(a_k)]. \end{aligned}$$

But

$$F(a_k, \lambda) - \varphi(a_k) > -\frac{\epsilon}{2}; \qquad \varphi(a_k) - \varphi(a_{k+1}) > -\frac{\epsilon}{2}$$

$$F(a_{k+1}, \lambda) - \varphi(a_{k+1}) < \frac{\epsilon}{2}; \qquad \varphi(a_{k+1}) - \varphi(a_k) < \frac{\epsilon}{2},$$

whence

$$-\epsilon < F(v, \lambda) - \varphi(v) < \epsilon.$$

Thus, given ϵ, there exists a number $L(\epsilon)$ depending upon ϵ alone and such that

$$|F(v, \lambda) - \varphi(v)| < \epsilon$$

for $\lambda > L(\epsilon)$ no matter what value is attributed to v.

The fundamental theorem with reference to probability can be stated as follows:

Let s_n be a stochastic variable depending upon a variable positive integer n. If the mathematical expectation $E(s_n^k)$ for any fixed $k = 1, 2, 3, \ldots$ tends, as n increases indefinitely, to the corresponding expectation

$$E(x^k) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} x^k e^{-x^2}dx$$

of a normally distributed variable, then the probability of the inequality

$$s_n < v$$

tends to the limit

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{v} e^{-x^2}dx$$

and that uniformly in v.

In very many cases it is much easier to make sure that the conditions of this theorem are fulfilled and then, in one stroke, to pass to the limit theorem for probability, than to attack the problem directly.

Application to Sums of Independent Variables

9. Let $z_1, z_2, z_3, \ldots$ be independent variables whose number can be increased indefinitely. Without losing anything in generality, we may suppose from the beginning

$$E(z_k) = 0; \qquad k = 1, 2, 3, \ldots.$$

We assume the existence of

$$E(z_k^2) = b_k$$

for all $k = 1, 2, 3, \ldots$. Also, we assume for *some* positive δ the existence of absolute moments

$$E|z_k|^{2+\delta} = \mu_k^{(2+\delta)}; \qquad k = 1, 2, 3, \ldots.$$

Liapounoff's theorem, with which we dealt at length in Chap. XIV, states that the probability of the inequality

$$\frac{z_1 + z_2 + \cdots + z_n}{\sqrt{2B_n}} < t,$$

where

$$B_n = b_1 + b_2 + \cdots + b_n$$

tends uniformly to the limit

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{t} e^{-x^2}dx$$

as $n \to \infty$, provided

$$\frac{\mu_1^{(2+\delta)} + \mu_2^{(2+\delta)} + \cdots + \mu_n^{(2+\delta)}}{B_n^{1+\frac{\delta}{2}}} \to 0.$$

Liapounoff's result in regard to generality of conditions surpassed by far what had been established before by Tshebysheff and Markoff, whose proofs were based on the fundamental result derived in the preceding section. Since Liapounoff's conditions do not require the existence of moments in an infinite number, it seemed that the method of moments was not powerful enough to establish the limit theorem in such a general form. Nevertheless, by resorting to an ingenious artifice, of which we made use in Chap. X, Sec. 8, Markoff finally succeeded in proving the limit theorem by the method of moments to the same degree of generality as did Liapounoff.

Markoff's artifice consists in associating with the variable z_k two new variables x_k and y_k defined as follows:

Let N be a positive number which in the course of proof will be selected so as to tend to infinity together with n. Then

$$\begin{aligned} x_k &= z_k, & y_k &= 0 & &\text{if} & |z_k| &\leqq N \\ x_k &= 0, & y_k &= z_k & &\text{if} & |z_k| &> N. \end{aligned}$$

Evidently z_k, x_k, y_k are connected by the relation

$$z_k = x_k + y_k$$

whence

$$E(x_k) + E(y_k) = 0. \tag{21}$$

Moreover

$$E(x_k^2) + E(y_k^2) = E(z_k^2) = b_k$$

(22)

$$E|x_k|^{2+\delta} + E|y_k|^{2+\delta} = E|z_k|^{2+\delta} = \mu_k^{(2+\delta)},$$

as one can see immediately from the definition of x_k and y_k.

Since x_k is bounded, mathematical expectations

$$E(x_k^l)$$

exist for all integer exponents $l = 1, 2, 3, \ldots$ and for $k = 1, 2, 3, \ldots$. In the following we shall use the notations

$$\begin{gathered} |E(x_k^l)| = c_k^{(l)}; \qquad l = 1, 2, 3, \ldots \\ c_1^{(2)} + c_2^{(2)} + \cdots + c_n^{(2)} = B_n' \\ \mu_1^{(2+\delta)} + \mu_2^{(2+\delta)} + \cdots + \mu_n^{(2+\delta)} = C_n. \end{gathered}$$

Not to obscure the essential steps of the reasoning we shall first establish a few preliminary results.

Lemma 1. *Let q_k represent the probability that $y_k \neq 0$; then*

$$q_1 + q_2 + \cdots + q_n \leqq \frac{C_n}{N^{2+\delta}}.$$

Proof. Let $\varphi_k(x)$ be the distribution function of z_k. Since $y_k \neq 0$ only if $|z_k| > N$, the probability q_k is not greater than

$$\int_{-\infty}^{-N} d\varphi_k(x) + \int_N^{\infty} d\varphi_k(x).$$

On the other hand,

$$\int_{-\infty}^{-N} |x|^{2+\delta} d\varphi_k(x) + \int_N^{\infty} |x|^{2+\delta} d\varphi_k(x) \leqq \mu_k^{(2+\delta)}.$$

But

$$\int_{-\infty}^{-N}|x|^{2+\delta}d\varphi_k(x) + \int_N^{\infty}|x|^{2+\delta}d\varphi_k(x) \geqq N^{2+\delta}\left\{\int_{-\infty}^{-N}d\varphi_k(x) + \int_N^{\infty}d\varphi_k(x)\right\},$$

whence

$$q_k \leqq \int_{-\infty}^{-N}d\varphi_k(x) + \int_N^{\infty}d\varphi_k(x) \leqq \frac{\mu_k^{(2+\delta)}}{N^{2+\delta}}.$$

The inequality to be proved follows immediately.

Lemma 2. *The following inequality holds:*

$$1 \geqq \frac{B'_n}{B_n} \geqq 1 - \frac{C_n}{B_nN^\delta}.$$

Proof. From

$$E|y_k|^{2+\delta} \leqq \mu_k^{(2+\delta)}$$

which is a consequence of the second equation (22) it follows that

$$E(y_k^2) \leqq \frac{\mu_k^{(2+\delta)}}{N^\delta}.$$

The first equation (22)

$$c_k^{(2)} + E(y_k^2) = b_k$$

gives

$$b_k \geqq c_k^{(2)} \geqq b_k - \frac{\mu_k^{(2+\delta)}}{N^\delta}.$$

Taking the sum for $k = 1, 2, 3, \ldots n$, we get

$$B_n \geqq B'_n \geqq B_n - \frac{C_n}{N^\delta},$$

whence

$$1 \geqq \frac{B'_n}{B_n} \geqq 1 - \frac{C_n}{B_nN^\delta}.$$

Lemma 3. *For* $e \geqq 3$,

$$\frac{c_1^{(e)} + c_2^{(e)} + \cdots + c_n^{(e)}}{B_n^{\frac{e}{2}}} \leqq \left(\frac{N^2}{B_n}\right)^{\frac{e-2}{2}}.$$

Proof. This inequality follows immediately from the evident inequalities

$$c_k^{(e)} \leqq E|x_k|^e \leqq N^{e-2}E(x_k^2) \leqq N^{e-2}b_k.$$

Lemma 4. *The following inequality holds*

$$\frac{c_1^{(1)} + c_2^{(1)} + \cdots + c_n^{(1)}}{B_n^{\frac{1}{2}}} \leqq \left(\frac{C_n}{N^{2+\delta}}\right)^{\frac{1}{2}}.$$

Proof. Since

$$E(x_k) + E(y_k) = 0,$$

we have

$$c_k^{(1)} = |E(x_k)| = |E(y_k)| \leqq E|y_k|.$$

On the other hand, by virtue of Schwarz's inequality

$$[E|y_1| + E|y_2| + \cdots + E|y_n|]^2 \leqq$$

$$\leqq (q_1 + q_2 + \cdots + q_n)\sum_{k=1}^{n} E(y_k^2) \leqq B_n\frac{C_n}{N^{2+\delta}},$$

whence the statement follows immediately.

If the variable integer N should be subject to the requirements that both the ratios

$$\frac{C_n}{N^{2+\delta}} \quad \text{and} \quad \frac{N^2}{B_n}$$

should tend to 0 when n increases indefinitely, then the preceding lemmas would give three important corollaries. But before stating these corollaries we must ascertain the possibility of selecting N as required. It suffices to take

$$N = (B_nC_n)^{\frac{1}{4+\delta}}.$$

Then

$$\frac{N^2}{B_n} = \frac{C_n}{N^{2+\delta}} = \left(\frac{C_n}{B_n^{1+\frac{\delta}{2}}}\right)^{\frac{2}{4+\delta}} \to 0$$

by virtue of Liapounoff's condition.

Also

$$\frac{C_n}{B_nN^\delta} = \left(\frac{C_n}{N^{2+\delta}}\right)^{\frac{\delta}{2+\delta}} \cdot \left(\frac{C_n}{B_n^{1+\frac{\delta}{2}}}\right)^{\frac{2}{2+\delta}}$$

will tend to 0. By selecting N in this manner we can state the following corollaries:

Corollary 1. *The sum*

$$q_1 + q_2 + \cdots + q_n$$

tends to 0 *as* $n \to \infty$.

Corollary 2. *The ratio*

$$\frac{B_n'}{B_n}$$

tends to 1.

Corollary 3. *The ratio*

$$\frac{c_1^{(e)} + c_2^{(e)} + \cdots + c_n^{(e)}}{B_n^{\frac{e}{2}}}$$

tends to 0 *for all positive integer exponents e except e* = 2.

10. Let $F_n(t)$ and $\phi_n(t)$ represent, respectively, the probabilities of the inequalities

$$\frac{z_1 + z_2 + \cdots + z_n}{\sqrt{2B_n}} < t$$
$$\frac{x_1 + x_2 + \cdots + x_n}{\sqrt{2B_n}} < t.$$

By repeating the reasoning developed in Chap. X, Sec. 8, we find that

$$|F_n(t) - \phi_n(t)| \leqq q_1 + q_2 + \cdots + q_n.$$

Hence,

$$\lim (F_n(t) - \phi_n(t)) = 0 \qquad \text{as} \qquad n \to \infty$$

by Corollary 1. It suffices therefore to show

$$\phi_n(t) \to \frac{1}{\sqrt{\pi}}\int_{-\infty}^{t} e^{-x^2}dx \qquad \text{as} \qquad n \to \infty,$$

and that can be done by the method of moments. By the polynomial theorem

$$\left(\frac{x_1 + x_2 + \cdots + x_n}{\sqrt{2B_n}}\right)^m = \sum \frac{m!}{\alpha!\beta! \cdots \lambda!} \frac{S_{\alpha,\beta,\ldots\lambda}}{2^{\frac{m}{2}}B_n^{\frac{m}{2}}}$$

where the summation extends over all systems of positive integers $\alpha \geqq \beta \geqq \cdots \geqq \lambda$ satisfying the condition

$$\alpha + \beta + \cdots + \lambda = m$$

and $S_{\alpha,\beta,\ldots\lambda}$ denotes a symmetrical function of letters $x_1, x_2, \ldots x_n$ determined by one of its terms

$$x_1^\alpha x_2^\beta \ldots x_l^\lambda$$

if l represents the number of integers $\alpha, \beta, \ldots \lambda$. Since variables $x_1, x_2, \ldots x_n$ are independent, we have

$$E\left(\frac{x_1 + x_2 + \cdots + x_n}{\sqrt{2B_n}}\right)^m = \sum \frac{m!}{\alpha!\beta! \cdots \lambda!} \frac{G_{\alpha,\beta,\ldots\lambda}}{2^{\frac{m}{2}} B_n^{\frac{m}{2}}}$$

where $G_{\alpha,\beta,\ldots\lambda}$ is obtained by replacing powers of variables by mathematical expectations of these powers. It is almost evident that

$$\frac{|G_{\alpha,\beta,\ldots\lambda}|}{B_n^{\frac{m}{2}}} \leqq \frac{c_1^{(\alpha)} + c_2^{(\alpha)} + \cdots c_n^{(\alpha)}}{B_n^{\frac{\alpha}{2}}} \cdot \frac{c_1^{(\beta)} + c_2^{(\beta)} + \cdots + c_n^{(\beta)}}{B_n^{\frac{\beta}{2}}} \cdots \frac{c_1^{(\lambda)} + c_2^{(\lambda)} + \cdots + c_n^{(\lambda)}}{B_n^{\frac{\lambda}{2}}}.$$

Now if not all the exponents $\alpha, \beta, \ldots \lambda$ are $= 2$ (which is possible only when m is even), by virtue of Corollary 3 the right member as well as

$$\frac{G_{\alpha,\beta,\ldots\lambda}}{B_n^{\frac{m}{2}}}$$

tends to 0. Hence

$$E\left(\frac{x_1 + x_2 + \cdots + x_n}{\sqrt{2B_n}}\right)^m \to 0$$

if m is *odd*.

But for *even* m we have

$$(23) \qquad E\left(\frac{x_1 + x_2 + \cdots + x_n}{\sqrt{2B_n}}\right)^m - \frac{m!}{2^m} \frac{G_{2,2,\ldots 2}}{B_n^{\frac{m}{2}}} \to 0.$$

Let us consider now (m being even)

$$\left(\frac{B_n'}{B_n}\right)^{\frac{m}{2}} = \left(\frac{c_1^{(2)} + c_2^{(2)} + \cdots + c_n^{(2)}}{B_n}\right)^{\frac{m}{2}} = \sum \frac{\frac{m}{2}!}{\lambda!\mu! \cdots \omega!} \frac{H_{\lambda,\mu,\ldots\omega}}{B_n^{\frac{m}{2}}}$$

where summation extends over all systems of positive integers

$$\lambda \geqq \mu \geqq \cdots \geqq \omega$$

satisfying the condition

$$\lambda + \mu + \cdots + \omega = \frac{m}{2}$$

and $H_{\lambda,\mu,\ldots\omega}$ is a symmetric function of $c_1^{(2)}, c_2^{(2)}, \ldots c_n^{(2)}$ determined by its term

$$(c_1^{(2)})^\lambda (c_2^{(2)})^\mu \ldots (c_l^{(2)})^\omega,$$

l being the number of subscripts $\lambda, \mu, \ldots \omega$. Apparently

$$\frac{H_{\lambda,\mu,\ldots\omega}}{B_n^{\frac{m}{2}}} \leqq \frac{(c_1^{(2)})^\lambda + (c_2^{(2)})^\lambda + \cdots + (c_n^{(2)})^\lambda}{B_n^\lambda} \cdots \frac{(c_1^{(2)})^\omega + (c_2^{(2)})^\omega + \cdots + (c_n^{(2)})^\omega}{B_n^\omega}.$$

Besides

$$c_k^{(2)} \leqq N^2, \qquad (c_k^{(2)})^e \leqq N^{2e-2}c_k^{(2)} \leqq N^{2e-2}b_k$$

and

$$\frac{(c_1^{(2)})^e + (c_2^{(2)})^e + \cdots + (c_n^{(2)})^e}{B_n^e} \leqq \left(\frac{N^2}{B_n}\right)^{e-1} \to 0$$

if $e > 1$. Thus

$$\frac{H_{\lambda,\mu,\ldots\omega}}{B_n^{\frac{m}{2}}} \to 0$$

if not all subscripts $\lambda, \mu, \ldots \omega$ are equal to 1. It follows that

$$\left(\frac{B_n'}{B_n}\right)^{\frac{m}{2}} - \left(\frac{m}{2}\right)! \frac{H_{1,1,\ldots 1}}{B_n^{\frac{m}{2}}} \to 0.$$

But by Corollary 2

$$\frac{B_n'}{B_n} \to 1$$

and evidently $H_{1,1,\ldots 1} = G_{2,2,\ldots 2}$. Hence

$$\left(\frac{m}{2}\right)! \frac{G_{2,2,\ldots 2}}{B_n^{\frac{m}{2}}} \to 1$$

and this in connection with (23) shows that for an even m

$$E\left(\frac{x_1 + x_2 + \cdots + x_n}{\sqrt{2B_n}}\right)^m \to \frac{m!}{2^m\left(\frac{m}{2}\right)!}.$$

Finally, no matter whether the exponent m is odd or even, we have

$$\lim E\left(\frac{x_1 + x_2 + \cdots + x_n}{\sqrt{2B_n}}\right)^m = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} x^m e^{-x^2}dx.$$

Tshebysheff-Markoff's fundamental theorem can be applied directly and leads to the result:

$$\lim \phi_n(t) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{t} e^{-x^2} dx$$

uniformly in t. On the other hand, as has been established before,

$$\lim [F_n(t) - \phi_n(t)] = 0$$

uniformly in t. Hence, finally

$$\lim F_n(t) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{t} e^{-x^2} dx$$

uniformly in t.

And this is the fundamental limit theorem with Liapounoff's conditions now proved by the method of moments. This proof, due to Markoff, is simple enough and of high elegance. However, preliminary considerations which underlie the proof of the fundamental theorem, though simple and elegant also, are rather long. Nevertheless, we must bear in mind that they are not only useful in connection with the theory of probability, but they have great importance in other fields of analysis.

APPENDIX III

ON A GAUSSIAN PROBLEM

1. In a letter to Laplace dated January 30, 1812,[1] Gauss mentions a difficult problem in probability for which he could not find a perfectly satisfactory solution. We quote from his letter:

Je me rappelle pourtant d'un problème curieux duquel je me suis occupé il y a 12 ans, mais lequel je n'ai pas réussi alors à résoudre à ma satisfaction. Peut-être daignerez-vous en occuper quelques moments: dans ce cas je suis sur que vous trouverez une solution plus complète. La voici: Soit M une quantité inconnue entre les limites 0 et 1 pour laquelle toutes les valeurs sont ou également probables ou plus ou moins selon une loi donnée: qu'on la suppose convertie en une fraction continue

$$M = \cfrac{1}{a' + \cfrac{1}{a'' + \cdots}}$$

Quelle est la probabilité qu'en s'arrêtant dans le développement à un terme fini $a^{(n)}$ la fraction suivante

$$\cfrac{1}{a^{(n+1)} + \cfrac{1}{a^{(n+2)} + \cdots}}$$

soit entre les limites 0 et x? Je la designe par $P(n, x)$ et j'ai en supposant toutes les valeurs également probables

$$P(0, x) = x.$$

$P(1, x)$ est une fonction transcendante dépendant de la fonction

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{x}$$

que Euler nomme inéxplicable et sur laquelle je viens de donner plusieurs recherches dans un mémoire présenté à notre Société des Sciences qui sera bientôt imprimé. Mais pour le cas ou n est plus grand, la valeur exacte de $P(n, x)$ semble intraitable. Cependant j'ai trouvé par des raisonnements très simples que pour n infinie

$$P(n, x) = \frac{\log (1 + x)}{\log 2}$$

[1] Gauss' Werke, X, 1, p. 371.

Mais les efforts que j'ai fait lors de mes recherches pour assigner

$$P(n, x) - \frac{\log (1 + x)}{\log 2}$$

pour une valeur très grande de n, mais pas infinie, ont été infructueux.

The problem itself and the main difficulty in its solution are clearly indicated in this passage. The problem is difficult indeed, and no satisfactory solution was offered before 1928, when Professor R. O. Kuzmin succeeded in solving it in a very remarkable and elegant way.

2. Analytical Expression for $P_n(x)$. We shall use the notation $P_n(x)$ for the probability which Gauss designated by $P(n, x)$. The first question that presents itself is how to express $P_n(x)$ in a proper analytical form. Let $\delta(v_1, v_2, \ldots v_n, x)$ be an interval whose end points are represented by two continued fractions:

$$\frac{1}{v_1} + \frac{1}{v_2} + \cdots + \frac{1}{v_n + x} \quad \text{and} \quad \frac{1}{v_1} + \frac{1}{v_2} + \cdots + \frac{1}{v_n}$$

with positive integer incomplete quotients $v_1, v_2, \ldots v_n$, while x is a positive number $\leqq 1$. Two such intervals corresponding to two *different* systems of integers $v_1, v_2, \ldots v_n$ and $v_1', v_2', \ldots v_n'$ do not overlap; that is, do not have common inner points. For, if they had a common inner point represented by an irrational number N (which we can always suppose), we should have for some positive $x' < 1$ and $x'' < 1$

$$N = \frac{1}{v_1} + \frac{1}{v_2} + \cdots + \frac{1}{v_n + x'} = \frac{1}{v_1'} + \frac{1}{v_2'} + \cdots + \frac{1}{v_n' + x''}.$$

But that is impossible unless $v_1' = v_1, v_2' = v_2, \ldots v_n' = v_n$.

A number M being selected at random between 0 and 1 and converted into a continued fraction

$$M = \frac{1}{v_1} + \frac{1}{v_2} + \cdots + \frac{1}{v_n + \xi}$$

if the quantity ξ turns out to be contained between 0 and $x < 1$, M must belong to one (and only one) of the intervals $\delta(v_1, v_2, \ldots v_n, x)$ corresponding to one of all the possible systems of n positive integers $v_1, v_2, \ldots v_n$. Since M has a uniform distribution of probability and since the length of the interval $\delta(v_1, v_2, \ldots v_n, x)$ is

$$(-1)^n\left[\cfrac{1}{v_1 + \cfrac{1}{v_2 + \cdots + \cfrac{1}{v_n + x}}} - \cfrac{1}{v_1 + \cfrac{1}{v_2 + \cdots + \cfrac{1}{v_n}}}\right]$$

the required probability $P_n(x)$ will be expressed by the sum

$$P_n(x) = \sum_{v_1, v_2, \ldots v_n} (-1)^n\left[\cfrac{1}{v_1 + \cfrac{1}{v_2 + \cdots + \cfrac{1}{v_n + x}}} - \cfrac{1}{v_1 + \cfrac{1}{v_2 + \cdots + \cfrac{1}{v_n}}}\right]$$

extended over all systems of positive integers $v_1, v_2, \ldots v_n$. In general let

$$\frac{P_i}{Q_i} = \cfrac{1}{v_1 + \cfrac{1}{v_2 + \cdots + \cfrac{1}{v_i}}} \qquad (i = 1, 2, \ldots n)$$

be a convergent to the continued fraction

$$\cfrac{1}{v_1 + \cfrac{1}{v_2 + \cdots + \cfrac{1}{v_n}}}.$$

Then the above expression for $P_n(x)$ can be exhibited in a more convenient form:

$$(1) \qquad P_n(x) = \sum_{v_1, v_2, \ldots v_n} (-1)^n\left[\frac{P_n + xP_{n-1}}{Q_n + xQ_{n-1}} - \frac{P_n}{Q_n}\right].$$

By the very definition of $P_n(x)$ we must have $P_n(1) = 1$; hence the important relation

$$(2) \qquad \sum \frac{1}{Q_n(Q_n + Q_{n-1})} = 1.$$

This result can also be established directly by resorting to the original expression of $P_n(1)$ and performing summation first with respect to v_1, then with respect to v_2, etc.

Relation (2) can be interpreted as follows: Let δ in general be the length of an interval $\delta(v_1, v_2, \ldots v_n, 1)$. Then

$$\Sigma\delta = 1$$

summation being extended over the (enumerable) set of intervals δ.

3. The Derivative of $P_n(x)$. In attempting to show that $P_n(x)$ tends uniformly to a limit function as $n \to \infty$ it is easier to begin with its derivative $p_n(x)$. Series

$$\sum \frac{1}{(Q_n + xQ_{n-1})^2}$$

obtained by formal derivation of (1) is uniformly convergent in the interval (0, 1). For

$$Q_n > \frac{Q_n + Q_{n-1}}{2}$$

whence

$$\frac{1}{(Q_n + xQ_{n-1})^2} < \frac{2}{Q_n(Q_n + Q_{n-1})}$$

and the series

$$\sum \frac{2}{Q_n(Q_n + Q_{n-1})} = 2$$

is convergent. Hence

$$\frac{dP_n(x)}{dx} = p_n(x) = \sum \frac{1}{(Q_n + xQ_{n-1})^2}.$$

Since

$$Q_n = v_nQ_{n-1} + Q_{n-2}$$

we have

$$p_n(x) = \sum_{v_1, v_2, \ldots v_n} \frac{1}{\left(Q_{n-1} + \dfrac{1}{v_n + x}Q_{n-2}\right)^2} \cdot \frac{1}{(v_n + x)^2}$$

and, performing summation with respect to $v_1, v_2, \ldots v_{n-1}$ for constant v_n

$$\sum_{v_1, v_2, \ldots v_{n-1}} \frac{1}{\left(Q_{n-1} + \dfrac{1}{v_n + x}Q_{n-2}\right)^2} = p_{n-1}\left(\frac{1}{v_n + x}\right),$$

whence

$$p_n(x) = \sum_{v_n=1}^{\infty} p_{n-1}\left(\frac{1}{v_n + x}\right)\frac{1}{(v_n + x)^2}$$

or else

$$p_n(x) = \sum_{v=1}^{\infty} p_{n-1}\left(\frac{1}{v+x}\right)\frac{1}{(v+x)^2} \tag{3}$$

—an important recurrence relation which permits determining completely the sequence of functions

$$p_1(x),\ p_2(x),\ \ldots$$

starting with $p_0(x) = 1$.

4. Discussion of a More General Recurrence Relation. In discussing relation (3) the fact that $p_0(x) = 1$ is of no consequence. We may start with any function $f_0(x)$ subject to some natural limitations, and form a sequence

$$f_1(x),\ f_2(x),\ f_3(x),\ \ldots$$

by means of the recurrence relation

$$f_n(x) = \sum_{v=1}^{\infty} f_{n-1}\left(\frac{1}{v+x}\right)\frac{1}{(v+x)^2}. \tag{4}$$

The following properties of $f_n(x)$ follow easily from this relation:

a. If

$$f_0(x) = \frac{a}{1+x}$$

then

$$f_n(x) = \frac{a}{1+x}; \qquad n = 1, 2, 3, \ldots$$

For

$$f_1(x) = a\sum_{v=1}^{\infty}\left(\frac{1}{v+x} - \frac{1}{v+x+1}\right) = \frac{a}{1+x}$$

whence the general statement follows immediately.

b. If

$$\frac{m}{1+x} \leqq f_0(x) \leqq \frac{M}{1+x}$$

then

$$\frac{m}{1+x} \leqq f_n(x) \leqq \frac{M}{1+x}.$$

Follows from (*a*) and equation (4) itself.

As a corollary we have: Let M_n and m_n be the precise upper and lower bounds of

$$(1 + x)f_n(x) \qquad (n = 0, 1, 2, \ldots)$$

in the interval $0 \leqq x \leqq 1$. Then

$$M_0 \geqq M_1 \geqq M_2 \geqq \cdots$$
$$m_0 \leqq m_1 \leqq m_2 \leqq \cdots$$

c. We have

$$\int_0^1 f_n(x)dx = \sum_{v=1}^{\infty} \int_0^1 f_{n-1}\left(\frac{1}{v + x}\right)\frac{dx}{(v + x)^2} =$$
$$= \int_1^{\infty} f_{n-1}\left(\frac{1}{u}\right)\frac{du}{u^2} = \int_0^1 f_{n-1}(x)dx = \int_0^1 f_0(x)dx.$$

d. The following relations can easily be established by mathematical induction:

$$f_n(x) = \sum f_0\left(\frac{P_n + xP_{n-1}}{Q_n + xQ_{n-1}}\right)\frac{1}{(Q_n + xQ_{n-1})^2}$$
$$f_{2n}(x) = \sum f_n\left(\frac{P_n + xP_{n-1}}{Q_n + xQ_{n-1}}\right)\frac{1}{(Q_n + xQ_{n-1})^2}$$
$$f_{3n}(x) = \sum f_{2n}\left(\frac{P_n + xP_{n-1}}{Q_n + xQ_{n-1}}\right)\frac{1}{(Q_n + xQ_{n-1})^2}.$$
$$\cdots\cdots$$

Let us suppose now that the function $f_0(x)$ defined in the interval

$$0 \leqq x \leqq 1$$

possesses a derivative everywhere in this interval and let μ_0 be an upper bound of $|f'_0(x)|$ while M is an upper bound of $|(1 + x)f_0(x)|$. Then by property (*b*)

$$|f_n(x)| \leqq M; \qquad |f_{2n}(x)| \leqq M; \qquad |f_{3n}(x)| \leqq M, \ldots.$$

The function $f_n(x)$ represented by the series

$$f_n(x) = \sum f_0(u)\frac{1}{(Q_n + xQ_{n-1})^2}$$

where u stands for

$$\frac{P_n + xP_{n-1}}{Q_n + xQ_{n-1}},$$

has a derivative; for the series obtained by a formal differentiation

$$f'_n(x) = \sum f'_0(u)\frac{(-1)^n}{(Q_n + xQ_{n-1})^4} - 2\sum f_0(u)\frac{Q_{n-1}}{(Q_n + xQ_{n-1})^3}$$

is uniformly convergent and represents $f'_n(x)$. Now

$$\frac{Q_{n-1}}{(Q_n + xQ_{n-1})^3} < \frac{1}{Q_n^2}$$

and

$$Q_n^2 > \frac{Q_n(Q_n + Q_{n-1})}{2}.$$

Hence

$$\left|2\sum f_0(u)\frac{Q_{n-1}}{(Q_n + xQ_{n-1})^3}\right| < 4M\sum \frac{1}{Q_n(Q_n + Q_{n-1})} = 4M$$

by virtue of (2). On the other hand, the inequality

$$Q_n(Q_n + Q_{n-1}) = (v_nQ_{n-1} + Q_{n-2})[(v_n + 1)Q_{n-1} + Q_{n-2}] >$$
$$> 2Q_{n-1}(Q_{n-1} + Q_{n-2})$$

holding for $n \geqq 2$ together with an evident inequality

$$Q_1(Q_1 + Q_0) \geqq 2$$

shows that

$$Q_n(Q_n + Q_{n-1}) > 2^n \qquad (n \geqq 2).$$

Thus

$$(Q_n + xQ_{n-1})^4 > Q_n^2 \cdot Q_n^2 > \frac{Q_n(Q_n + Q_{n-1})}{2} \cdot \frac{Q_n(Q_n + Q_{n-1})}{2} >$$
$$> 2^{n-2}Q_n(Q_n + Q_{n-1})$$

and consequently

$$\left|\sum f'_0(u)\frac{(-1)^n}{(Q_n + xQ_{n-1})^4}\right| < \frac{\mu_0}{2^{n-2}}.$$

Hence, we may conclude that

$$\mu_1 = \frac{\mu_0}{2^{n-2}} + 4M$$

is an upper bound of $|f'_n(x)|$. Similarly, starting with the second equation in (d), we find that

$$\mu_2 = \frac{\mu_1}{2^{n-2}} + 4M$$

is an upper bound of $|f'_{2n}(x)|$, and so forth. In general, the recurrence relation

$$\mu_k = \frac{\mu_{k-1}}{2^{n-2}} + 4M \qquad (k = 1, 2, 3, \cdots)$$

determines upper bounds of

$$|f'_n(x)|, \; |f'_{2n}(x)|, \; |f'_{3n}(x)|, \; \ldots .$$

It is easy to see that in general

$$\mu_k < \frac{\mu_0}{2^{k(n-2)}} + \frac{4M}{1 - 2^{-(n-2)}}$$

so that for sufficiently large n

$$\mu_k < 5M.$$

5. Main Inequalities. Let

$$\varphi_0(x) = f_0(x) - \frac{m_0}{1 + x}.$$

Then

$$f_n(x) - \frac{m_0}{1+x} = \varphi_n(x) =$$

$$= \sum \varphi_0(u) \frac{1}{(Q_n + xQ_{n-1})^2} > \frac{1}{2} \sum \varphi_0(u) \frac{1}{Q_n(Q_n + Q_{n-1})}.$$

Since the intervals δ defined at the end of Sec. 2 do not overlap and cover completely the whole interval (0, 1), we may write:

$$l = \frac{1}{2}\int_0^1 \varphi_0(x)dx = \frac{1}{2}\sum \int_{(\delta)} \varphi_0(x)dx = \frac{1}{2}\sum \varphi_0(u_1)\frac{1}{Q_n(Q_n + Q_{n-1})},$$

the latter part following from the mean value theorem and u_1 being a number contained within the interval δ. By subtraction we find

$$f_n(x) - \frac{m_0}{1+x} - l > \frac{1}{2}\sum [\varphi_0(u) - \varphi_0(u_1)]\frac{1}{Q_n(Q_n + Q_{n-1})}$$

and, since both u and u_1 belong to the same interval δ,

$$\varphi_0(u) - \varphi(u_1) > -\frac{\mu_0 + m_0}{Q_n(Q_n + Q_{n-1})} > -\frac{\mu_0 + m_0}{2^n}.$$

Consequently,

$$f_n(x) - \frac{m_0}{1+x} - l > -\frac{\mu_0 + m_0}{2^{n+1}},$$

and a fortiori

$$f_n(x) > \frac{m_0 + l - 2^{-n}(\mu_0 + m_0)}{1 + x}.$$

It follows that

(5) $$m_1 \geqq m_0 + l - 2^{-n}(\mu_0 + m_0).^*$$

In a similar way, considering the function

$$\psi_0(x) = \frac{M_0}{1 + x} - f_0(x)$$

and setting

$$l_1 = \tfrac{1}{2}\int_0^1 \psi_0(x)dx,$$

we shall have

$$f_n(x) < \frac{M_0 - l_1 + 2^{-n}(\mu_0 + M_0)}{1 + x},$$

whence

(6) $$M_1 \leqq M_0 - l_1 + 2^{-n}(\mu_0 + M_0).$$

Further, from (5) and (6)

$$M_1 - m_1 \leqq M_0 - m_0 + 2^{-n+1}(\mu_0 + M_0) - l - l_1.$$

But

$$l + l_1 = \tfrac{1}{2}\log 2 \cdot (M_0 - m_0) = (1 - k)(M_0 - m_0); \qquad k < 0.66,$$

so that finally

$$M_1 - m_1 < k(M_0 - m_0) + 2^{-n+1}(\mu_0 + M_0).$$

Starting with $f_n(x), f_{2n}(x), \ldots$ instead of $f_0(x)$, in a similar way we find

$$\begin{aligned} M_2 - m_2 &< k(M_1 - m_1) + 2^{-n+1}(\mu_1 + M_1) \\ M_3 - m_3 &< k(M_2 - m_2) + 2^{-n+1}(\mu_2 + M_2) \\ &\cdots\cdots \\ M_n - m_n &< k(M_{n-1} - m_{n-1}) + 2^{-n+1}(\mu_{n-1} + M_{n-1}). \end{aligned}$$

From these inequalities it follows that

$$M_n - m_n < (M_0 - m_0)k^n + 2^{-n+1}[\mu_0 k^{n-1} + \mu_1 k^{n-2} + \cdots + \mu_{n-1} + \\ + M_0 k^{n-1} + M_1 k^{n-2} + \cdots + M_{n-1}].$$

Without losing anything in generality, we may suppose that $f_0(x)$ is a positive function. Then

* M_i, m_i are used here with the same meaning as M_{ni}, m_{ni} in Sec. 4.

$$M_k \leqq M_0, \qquad \mu_k < 5M_0 \qquad (k = 1, 2, 3, \ldots)$$

at least for sufficiently large n. Owing to these inequalities we shall have

$$M_n - m_n < (M_0 - m_0)k^n + \mu_0\left(\frac{k}{2}\right)^{n-1} + \frac{6M_0}{(1-k)2^{n-1}}. \tag{7}$$

This inequality shows that sequences

$$M_0 \geqq M_1 \geqq M_2 \geqq \cdots$$
$$m_0 \leqq m_1 \leqq m_2 \leqq \cdots$$

approach a common limit a. The following method can be used to find the value of this limit. Let N be an arbitrary sufficiently large integer and n the integer defined by

$$n^2 \leqq N < (n+1)^2.$$

Then

$$\frac{m_n}{1+x} \leqq f_{nn}(x) \leqq \frac{M_n}{1+x},$$

and therefore

$$\frac{m_n}{1+x} \leqq f_N(x) \leqq \frac{M_n}{1+x}.$$

The last inequality permits presenting $f_N(x)$ thus:

$$f_N(x) = \frac{a}{1+x} + \theta(M_n - m_n); \qquad |\theta| < 1, \tag{8}$$

whence

$$\int_0^1 f_N(x)dx = \int_0^1 f_0(x)dx = a \log 2 + \theta'(M_n - m_n), \qquad |\theta'| < 1,$$

and, because $M_n - m_n$ ultimately becomes as small as we please in absolute value,

$$a \log 2 = \int_0^1 f_0(x)dx.$$

Equation (8) shows clearly that the sequence of functions

$$f_0(x), f_1(x), f_2(x), \ldots$$

defined by the recurrence relation (4) approaches uniformly the limit function

$$\frac{a}{1+x}$$

where

$$a = \frac{1}{\log 2}\int_0^1 f_0(x)dx.$$

6. Solution of the Gaussian Problem. It suffices to apply the preceding considerations to the case $f_0(x) = p_0(x) = 1$. In this case $M_0 = 2$, $m_0 = 1$, $\mu_0 = 0$ and

$$a = \frac{1}{\log 2}.$$

Consequently,

$$p_N(x) = \frac{1}{(1 + x)\log 2} + \theta\left(k^n + \frac{3}{(1 - k)\cdot 2^{n-3}}\right); \qquad |\theta| < 1$$

where $n = [\sqrt{N}]$. It suffices to integrate this expression between limits 0 and $t < 1$ to find

$$P_N(t) = \frac{\log(1 + t)}{\log 2} + \lambda\left(k^n + \frac{3}{(1 - k)2^{n-3}}\right); \qquad |\lambda| < t.$$

As $N \to \infty$

$$P_N(t) \to \frac{\log(1 + t)}{\log 2}$$

as stated by Gauss. Moreover,

$$\left|P_N(t) - \frac{\log(1 + t)}{\log 2}\right| < t\left(k^n + \frac{3}{(1 - k)2^{n-3}}\right)$$

for sufficiently large, but finite N.

Table of the Probability Integral

$$\phi(z) = \frac{1}{\sqrt{2\pi}}\int_0^z e^{-\frac{1}{2}t^2}dt$$

z	$\phi(z)$	z	$\phi(z)$	z	$\phi(z)$	z	$\phi(z)$
0.00	0.0000	0.65	0.2422	1.30	0.4032	1.95	0.4744
0.01	0.0040	0.66	0.2454	1.31	0.4049	1.96	0.4750
0.02	0.0080	0.67	0.2486	1.32	0.4066	1.97	0.4756
0.03	0.0120	0.68	0.2517	1.33	0.4082	1.98	0.4761
0.04	0.0160	0.69	0.2549	1.34	0.4099	1.99	0.4767
0.05	0.0199	0.70	0.2580	1.35	0.4115	2.00	0.4772
0.06	0.0239	0.71	0.2611	1.36	0.4131	2.02	0.4783
0.07	0.0279	0.72	0.2642	1.37	0.4147	2.04	0.4793
0.08	0.0319	0.73	0.2673	1.38	0.4162	2.06	0.4803
0.09	0.0359	0.74	0.2703	1.39	0.4177	2.08	0.4812
0.10	0.0398	0.75	0.2734	1.40	0.4192	2.10	0.4821
0.11	0.0438	0.76	0.2764	1.41	0.4207	2.12	0.4830
0.12	0.0478	0.77	0.2794	1.42	0.4222	2.14	0.4838
0.13	0.0517	0.78	0.2823	1.43	0.4236	2.16	0.4846
0.14	0.0557	0.79	0.2852	1.44	0.4251	2.18	0.4854
0.15	0.0596	0.80	0.2881	1.45	0.4265	2.20	0.4861
0.16	0.0636	0.81	0.2910	1.46	0.4279	2.22	0.4868
0.17	0.0675	0.82	0.2939	1.47	0.4292	2.24	0.4875
0.18	0.0714	0.83	0.2967	1.48	0.4306	2.26	0.4881
0.19	0.0753	0.84	0.2995	1.49	0.4319	2.28	0.4887
0.20	0.0793	0.85	0.3023	1.50	0.4332	2.30	0.4893
0.21	0.0832	0.86	0.3051	1.51	0.4345	2.32	0.4898
0.22	0.0871	0.87	0.3078	1.52	0.4357	2.34	0.4904
0.23	0.0910	0.88	0.3106	1.53	0.4370	2.36	0.4909
0.24	0.0948	0.89	0.3133	1.54	0.4382	2.38	0.4913
0.25	0.0987	0.90	0.3159	1.55	0.4394	2.40	0.4918
0.26	0.1026	0.91	0.3186	1.56	0.4406	2.42	0.4922
0.27	0.1064	0.92	0.3212	1.57	0.4418	2.44	0.4927
0.28	0.1103	0.93	0.3238	1.58	0.4429	2.46	0.4931
0.29	0.1141	0.94	0.3264	1.59	0.4441	2.48	0.4934
0.30	0.1179	0.95	0.3289	1.60	0.4452	2.50	0.4938
0.31	0.1217	0.96	0.3315	1.61	0.4463	2.52	0.4941
0.32	0.1255	0.97	0.3340	1.62	0.4474	2.54	0.4945
0.33	0.1293	0.98	0.3365	1.63	0.4484	2.56	0.4948
0.34	0.1331	0.99	0.3389	1.64	0.4495	2.58	0.4951
0.35	0.1368	1.00	0.3413	1.65	0.4505	2.60	0.4953
0.36	0.1406	1.01	0.3438	1.66	0.4515	2.62	0.4956
0.37	0.1443	1.02	0.3461	1.67	0.4525	2.64	0.4959
0.38	0.1480	1.03	0.3485	1.68	0.4535	2.66	0.4961
0.39	0.1517	1.04	0.3508	1.69	0.4545	2.68	0.4963
0.40	0.1554	1.05	0.3531	1.70	0.4554	2.70	0.4965
0.41	0.1591	1.06	0.3554	1.71	0.4564	2.72	0.4967
0.42	0.1628	1.07	0.3577	1.72	0.4573	2.74	0.4969
0.43	0.1664	1.08	0.3599	1.73	0.4582	2.76	0.4971
0.44	0.1700	1.09	0.3621	1.74	0.4591	2.78	0.4973
0.45	0.1736	1.10	0.3643	1.75	0.4599	2.80	0.4974
0.46	0.1772	1.11	0.3665	1.76	0.4608	2.82	0.4976
0.47	0.1808	1.12	0.3686	1.77	0.4616	2.84	0.4977
0.48	0.1844	1.13	0.3708	1.78	0.4625	2.86	0.4979
0.49	0.1879	1.14	0.3729	1.79	0.4633	2.88	0.4980
0.50	0.1915	1.15	0.3749	1.80	0.4641	2.90	0.4981
0.51	0.1950	1.16	0.3770	1.81	0.4649	2.92	0.4982
0.52	0.1985	1.17	0.3790	1.82	0.4656	2.94	0.4984
0.53	0.2019	1.18	0.3810	1.83	0.4664	2.96	0.4985
0.54	0.2054	1.19	0.3830	1.84	0.4671	2.98	0.4986
0.55	0.2088	1.20	0.3849	1.85	0.4678	3.00	0.49865
0.56	0.2123	1.21	0.3869	1.86	0.4686	3.20	0.49931
0.57	0.2157	1.22	0.3888	1.87	0.4693	3.40	0.49966
0.58	0.2190	1.23	0.3907	1.88	0.4699	3.60	0.499841
0.59	0.2224	1.24	0.3925	1.89	0.4706	3.80	0.499928
0.60	0.2257	1.25	0.3944	1.90	0.4713	4.00	0.499968
0.61	0.2291	1.26	0.3962	1.91	0.4719	4.50	0.499997
0.62	0.2324	1.27	0.3980	1.92	0.4726	5.00	0.499997
0.63	0.2357	1.28	0.3997	1.93	0.4732		
0.64	0.2389	1.29	0.4015	1.94	0.4738		

INDEX

A

Arrangements, 18

B

Bayes' formula (theorem), 61
Bernoulli criterion, 5
Bernoulli theorem, 96
Bernoulli trials, 45
Bernstein, S., inequality, 205
Bertrand's paradox, 251
Buffon's needle problem, 113, 251
  Barbier's solution of, 253

C

Cantelli's theorem, 101
Cauchy's distribution, 243, 275
Characteristic function, composition of, 275
  of distribution, 240, 264
Coefficient, correlation, 339
  divergence, 212, 214, 216
Combinations, 18
Compound probability, theorem of, 31
Continued fractions, 358, 361, 396
  Markoff's method of, 52
Continuous variables, 235
Correlation, normal (*see* Normal correlation)
Correlation coefficient, distribution of, 339

D

Difference equations, ordinary, 75, 78
  partial, 84
Dispersion, definition, 172
  of sums, 173
Distribution, Cauchy's, 243, 275
  characteristic function of, 264
  of correlation coefficient, 339
Distribution, determination of, 271
  equivalent point, 369
  general concept of, 263
  normal (Gaussian), 243
  Poisson's, 279
  "Student's," 339
Distribution function of probability, 239, 263
Divergence coefficient, empirical, 212
  Lexis' case, 214
  Poisson's case, 214
  theoretical, 212
  Tschuprow's theorem, 216

E

Elementary errors, hypothesis of, 296
Ellipses of equal probability, 311, 328
Estimation of error term, 295
Euler's summation formula, 177, 201, 303, 347
Events, compound, 29
  contingent, 3
  dependent, 33
  equally likely, 4, 5, 7
  exhaustive, 6
  future, 65
  incompatible, 37
  independent, 32, 33
  mutually exclusive, 6, 27
  opposite, 29
Expectation, mathematical, 161
  of a product, 171
  of a sum, 165

F

Factorials, 349
Fourier theorem, 241
French lottery, 19, 108
Frequency, 96
Fundamental lemma (*see* Limit theorem)
Fundamental thçorem (*see* Tshebysheff-Markoff theorem)

G

Gaussian distribution, 243
Gaussian problem, 396
Generating function of probabilities, 47, 78, 85, 89, 93, 94

H

Hermite polynomials, 72
Hypothesis of elementary errors, 296

I

Independence, definition of, 32, 33

K

Khintchine (*see* Law of large numbers)
Kolmogoroff (*see* Law of large numbers; Strong law of large numbers)

L

Lagrange's series, 84, 150
Laplace-Liapounoff (*see* Limit theorem)
Laplace's problem, 255
Laurent's series, 87, 148
Law of large numbers, generalization by Markoff, 191
  for identical variables (Khintchine), 195
  Kolmogoroff's lemma, 201
  theorem, 185
  Tshebysheff's lemma, 182
Law of repeated logarithm, 204
Law of succession, 69
Lexis' case, 214
Liapounoff condition (*see* Limit theorem)
Liapounoff inequality, 265
Limit theorem, Bernoullian case, 131
  for sums of independent vectors, 318, 323, 325, 326
  fundamental lemma, 284
  Laplace-Liapounoff, 284
Line of regression, 314
Lottery, French (*see* French lottery)

M

Marbe's problem, 231
Markoff's theorem, infinite dispersion, 191
Markoff's theorem, for simple chains, 301
Markoff-Tshebysheff theorem (*see* Tshebysheff-Markoff theorem)
Mathematical expectation, definition of, 161
  of a product, 171
  of a sum, 165
Mathematical probability, definition of, 6
Moments, absolute, 240, 264
  inequalities for, 264
  method of (Markoff's), 356*ff*.

N

Normal correlation, 313
  origin of, 327
Normal distribution, Gaussian, 243
  two-dimensional, 308

P

Pearson's "χ^2-test," 327
Permutations, 18
Point, of continuity, 261, 356
  of increase, 262, 356
Poisson series, 182, 293
Poisson's case, 214
Poisson's distribution, 279
Poisson's formula, 137
Poisson's theorem, 208, 294
Polynomials, Hermite (*see* Hermite)
Probability, approximate evaluation of, by Markoff's method, 52
  compound, 29, 31
  conditional, 33
  definition (classical) of, 6
  total, 27, 28
Probability integral, 128
  table of, 407

R

Relative frequency, 96
Runs, problem of, 77

S

Simple chains, 74, 223, 297
  Markoff's theorem for, 301
Standard deviation, 173

Stieltjes' integrals, 261
Stirling's formula, 349
Stochastic variables, 161
Strong law of large numbers (Kolmogoroff), 202
"Student's" distribution, 339

T

Table of probability integral, 407
Tests of significance, 331
Total probability, theorem of, 27, 28
Trials, dependent, independent, repeated, 44, 45
Tschuprow (*see* Divergence coefficient)
Tshebysheff-Markoff theorem, fundamental, 304, 384
  application, 388
Tshebysheff's inequalities, 373
Tshebysheff's inequality, 204
Tshebysheff's lemma, 182
Tshebysheff's problem, 199

V

Variables, continuous, 235
  independent, 171
  stochastic, 161
Vectors (*see* Limit theorem)