LONDON
Cambridge University Press
FETTER LANE

NEW YORK · TORONTO
BOMBAY · CALCUTTA · MADRAS
Macmillan

TOKYO
Maruzen Company Ltd

THEORY OF PROBABILITY

BY THE LATE

WILLIAM BURNSIDE
Sc.D., LL.D., F.R.S.

HONORARY FELLOW OF
PEMBROKE COLLEGE, CAMBRIDGE

CAMBRIDGE
AT THE UNIVERSITY PRESS
1936

First edition 1928
Reprinted 1936

PRINTED IN GREAT BRITAIN

PREFACE

THE present small volume on the theory of probability represents a manuscript which Professor Burnside had practically completed some time before his death.

The theory had begun to occupy his thought during the war; and the earliest (1918) of his papers, relating to any of its topics, deals with what is manifestly a military question, reduced (for purposes of calculation) to a purely mathematical form. As was his wont in any subject, his interest in its developments grew: a number of isolated papers by him appeared from time to time, in a widening range of treatment. Ultimately, he set himself to make a systematic account of the theory as it presented itself to him.

So far as can be remembered by Mrs Burnside, the draft was written at intervals before the middle of 1925. At the time when he had finished his account, it contained all the issues which he proposed to discuss: but marginal references in the manuscript shew that he intended to add a number of Notes elucidating or establishing statements in the text. Of these Notes, only one* was actually written; and no memoranda have been found which might have indicated the intended range of the remainder. His work was interrupted by a serious illness late in 1925. After a recovery which was only partial, he occasionally longed to return to the draft, so as to make additions and amplifications: but the necessary strength was lacking. The manuscript remained unaltered.

It has seemed desirable to publish his draft exactly as he left it. The Syndics of the Cambridge University Press have been willing to undertake the publication; and Mrs Burnside desires me to express her thanks to the Syndics for their action.

* It occupies pp. 101, 102 of the volume.

At the request of Mrs Burnside, and by the willing acquiescence of the Syndics, my notice of Professor Burnside which was written for the Royal Society is prefixed to the volume. And I have appended (p. 104) a list of the papers which, in his later years, he published on questions cognate with the range of the volume.

During the progress of the printing, I have owed much to the Secretary of the Syndics, Mr S. C. Roberts, and to the Staff of the University Press, for their unstinted and ready help; and I return to them my sincere thanks for their considerate cooperation.

A. R. FORSYTH

March 1928

CONTENTS

CHAPTER IV

METHODS OF APPROXIMATION

CHAPTER V

PROBABILITY OF CAUSES

CHAPTER VI

PROBABILITIES CONNECTED WITH GEOMETRICAL QUESTIONS

CHAPTER VII

THEORY OF ERRORS

CHAPTER VIII

GAUSS'S LAW OF ERRORS

[From the *Proceedings of the Royal Society*, Ser. A, vol. 117 (1928), pp. xi–xxv.]

WILLIAM BURNSIDE

WILLIAM BURNSIDE was born on July 2, 1852, the son of William Burnside, a merchant, of 7, Howley Place, Paddington, London. His father was of Scottish ancestry: his grandfather, who had gone to London, was a partner in the bookselling firm of Seeley and Burnside.

Left an orphan at the age of six, Burnside was educated at Christ's Hospital, where he was a Grecian: there, besides his distinction in the grammar school, he attained the highest place in the mathematical school. Having been elected to an entrance scholarship at St John's College, Cambridge, he went into residence in October, 1871, and was regarded as the best man of his year in the college. In accordance with the general custom of capable students of mathematics in Cambridge, he "coached" for the Tripos, his private tutor being W. H. Besant, one of the few rivals of the famous Routh. For some reason, Burnside migrated to Pembroke College in the same university, the change being made late in his second year (May, 1873). He graduated in the Mathematical Tripos of 1875 as second wrangler, being bracketed with George Chrystal, who afterwards was professor at Edinburgh; the fourth wrangler was R. F. Scott, now* Master of St John's College. In the subsequent Smith's Prize Examination, Burnside was first and Chrystal was second.

A fellowship at Pembroke was the worthy sequel of such a degree; he continued a fellow from 1875 until 1886. He was at once appointed to lecture in his college: and he lectured also at Emmanuel in 1876 and at King's in 1877. At that time, college teaching for the best students was sometimes shared by a few colleges, in isolated groups, and included subjects selected from the average normal course for Honours; and Burnside, in addition, gave lectures on hydrodynamics, an advanced course open to all the University. That particular subject was coming

* The writer is indebted to Sir Robert Scott, for several of the personal records in this notice.

into vogue again at Cambridge; attention, regularly paid to the established work of Stokes, was stimulated by the then new work of Greenhill and especially of Lamb. Burnside also examined for the Mathematical Tripos from time to time. Occasionally, he did some private coaching. But later it appeared that, instead of restricting himself mainly to Tripos subjects in furtherance of his lectures and his inevitable share in examinations, he had launched himself upon a broad sea of study, then far removed from the Tripos domain.

As an undergraduate, he had proved an expert oarsman. While at St John's College, even as a freshman, he had rowed in the Lady Margaret First Boat which, with the famous Goldie as stroke, went head of the river in 1872. Rather light in weight as an undergraduate, too light (according to the canons of the day) to be considered for the University Boat, he was always rather spare of build and he retained a wonderful power of endurance; and he kept his rowing form for many years. He rowed in the Pembroke Boat after graduation, as long as he continued in residence; he was a splendid "7," and had a full share in its steady rise on the river. For some years after he left Cambridge, his reputation as an oar survived as a tradition in college circles.

After going out of residence, similar opportunities for rowing were not accessible. But in the course of holidays frequently spent in Scotland, Burnside had acquired a zest for fishing; and for many a summer onwards he continued to go there, pursuing what grew to be his favourite sport. As in rowing, so in fishing, he developed skill and became an expert fisherman; indeed, with all he undertook, nothing short of his best was sufficient.

In 1885, at the instance of Mr (afterwards Sir) William Niven, the Director of Naval Instruction—himself a Cambridge man, devoted to natural philosophy, as it was styled by good Newtonians—Burnside was appointed professor of mathematics in the Royal Naval College at Greenwich. The rest of his teaching life was spent in that post. There was a current belief, a belief now known to be justified by fact, that his old college had invited him to return to important office; but he remained at

Greenwich. His work was to his liking. It was a course, well-defined in extent and in demands on time, within a variety of congenial subjects, though only touching in part upon the regions of his constructive thought. The actual teaching, with its incident duties, left him adequate opportunity to keep abreast of progress, even to advance progress, in the subjects of professional duty. It also left him leisure, which was carefully and diligently used, to pursue his own researches, whatever their direction. Best of all to him, he was free from the interruptions and the incessant small demands, business and social, that are inseparable from official administration. For at all times, and in all ways, multifarious detail—whether incidental to the non-scientific side of official duty, or the current presidency of a scientific society such as the London Mathematical, even the purely algebraical garniture and the side-issues in mathematical investigations—such detail was inexpressibly irksome to his spirit.

At Greenwich, Burnside's work was devoted to the training of naval officers. It consisted of three ranges. There was a junior section for gunnery and torpedo officers; the chief subject of study was the principles of ballistics. There was a senior section for engineer officers: the chief subjects of study were strength of materials, dynamics, and heat engines. The advanced section—perhaps that in which he exercised the greatest influence on his students—was reserved for the class of naval constructors; in that range, Burnside's special mastery of kinematics, kinetics, and hydrodynamics, proved invaluable. Records and remembrance declare that he was a fine and stimulating teacher, patient with students in their difficulties and their questions—though elsewhere, as in discussion with equals, his manner could have a directness that, to some, might appear abrupt. He certainly earned the gratitude of his students, as appeared from their spontaneous token of tribute to him when he left in 1919; the address, which they then presented, was treasured by him and his family.

Burnside had married Alexandrina Urquhart in 1886, soon after he was appointed professor at Greenwich. She survives him, with their family of two sons and three daughters.

After his work at the Naval College had ended, the whole family retired to West Wickham in Kent. Burnside, happy as he had been in that work and regretting its actual termination, enjoyed his leisure, spending it among his books, in fishing holidays in Scotland and, not least, in his researches, some continued in regions recognised as specially his own, some of them in the systematic development of ideas in still another branch of mathematics upon which his intellectual interests had settled. The last year of his life was marked by failing health: and the proximate cause of his death was a recurrence of cerebral hæmorrhage. He died on August 21, 1927; and he is buried in West Wickham churchyard.

In recognition of his eminence as a mathematician, not a few academic honours came to Burnside during his life. He was never avid of honours; indeed, he was eager to avoid those forms of academic recognition constituted by official positions of dignity, when they demanded the performance of any public duty set in formal pomp or circumstance. He received honorary degrees, Sc.D. from Dublin, LL.D. from Edinburgh. He was elected a Fellow of the Royal Society in 1893, on the first occasion of candidature: he served on the Council of that body from 1901 to 1903; and he was awarded one of the two Royal medals for the year 1904. He was a member of the Council of the London Mathematical Society for the long continuous period from 1899 to 1917: there, he was a tower of strength, in advice during the Council's meetings, and by his many reports as a referee upon a multitude of varied original papers submitted by a small army of authors. He was awarded the De Morgan medal of the Society in 1899. From 1906 to 1908 he served as President: while willingly allowing his name to be submitted for membership of the Council year after year, he accepted their highest office only with grave and characteristic reluctance. The honour, in which he appeared to shew most interest, was conferred on him in 1900. In that year he was elected an Honorary Fellow of his old college, Pembroke; and at the time of his death he had become the senior on the small roll of Honorary Fellows. Yet, even in the few and far from fluent remarks of thanks which he made at the college dinner welcoming, by courteous custom, the newly elected honorary members of the foundation, he urged that the happy and successful pursuit of research was its own reward; and the sincerity of his plea was appreciated not least by those who had done their part in recognition of his labours.

Burnside was frequently called upon to examine for the Mathematical Tripos and for the open Civil Service examinations of the highest grade. Occasionally, he acted as external examiner for one or other of the English Universities, as well as for the Naval College after his retirement. He was not an easy examiner—before his early days of such duty, the phrase "easy problems" at Cambridge had come to bear a perverse significance. His questions could be of the type which, gathered in one of his papers, might justify the epithet beautiful: they were certainly too beautiful for the candidates in the 1881 Tripos, the first university occasion when he examined. Yet, though they often were difficult and always on a high level, they were set with the design of evoking an examinee's thought, rather than of providing an opportunity for the facile display of trained manipulative skill along familiar lines.

Through many years, Burnside was in constant requisition as a referee, for the Royal Society and for the London Mathematical Society. He could not be called lenient: for, however sympathetic with writers, and especially young writers, he held a high standard of the attainment that was deserving of publication. He was often fruitful in suggestion. He could even be severe on occasion: yet he would mitigate a judgment when grounds for its reconsideration were submitted. Similarly, as a critic of a friend's proof-sheets, he could be severe, yet always objectively so: he obviously assumed, without the possibility of question, that the friend's standard and his own were alike in practice. Thus, at the end of a discussion, the friend would find that added light had been cast upon the whole matter—surely the best criterion of sympathetic criticism. And if severe with others, he was stern with himself—a mental discipline that exercised its influence towards the directness and the precision both of form and of substance in his writings.

Valuable as were his teaching, his activity as an examiner, and his influence as a referee, it is by the contributions which he has made to his science that Burnside's name will be held in remembrance.

His range was wide; for it branched out, through applied mathematics from the days of his early training, into great tracts of pure mathematics in the years of his matured powers. Yet, even in the later time, when specialisation has tended to become acute, he could specialise with the best. Though of course not comparable with an Euler, a Cauchy, or a Cayley, in the variety or the amount of work he has left, he has delved in many fields and has left his trace in many directions. He published over one hundred and fifty papers, as well as one treatise, the *Theory of Groups*, of which a second (and greatly amplified) edition was issued also under his own care. He has also left a manuscript, fairly complete as far as it was carried, on the theory of probability. He himself did not regard this work as finished; on various issues, he was in correspondence from time to time with the present President of the Royal Society, the Astronomer Royal, and others; and he certainly did not consider that he had resolved all his own questions. Had life in health lasted appreciably longer, there is no doubt that he could have attained, as he intended to pursue, further development in a subject which occupied much of the thought of his later years.

In that considerable tale of papers, most are short. Very many of them occupy only a few pages. His longest individual paper—he never used the more ambitious title "memoir"—deals with automorphic functions: it really consists of two parts connected, though not consecutive, in matter; and the whole occupies no more than fifty-three octavo pages. Brief however as his papers are, it can fairly be asserted that each one of them contains some definite and recognisable result or results. He never discussed side-issues; he would not even dwell on the minute details of a main issue. Indeed, he could be intellectually bored by processes, that halted in their march to settle subsidiary questions as they arose; with him, auxiliary necessary material was set out before the main advance. When once an issue was attained, he was content to let it stand by its own significance:

to others he would leave attempts "to gild refined gold, to paint the lily."

He happily was saved mathematical controversy, which he detested. On one occasion he was surprised, even disturbed, by the receipt of an unseemly letter the very tone of which amazed him (not unjustifiably): it concerned a question of priority which, in so far as it could affect a man punctilious in his acknowledgment of the work of others, to Burnside was as thin as air, though manifestly not so to the writer of the letter. The quiet firmness of Burnside's answer to his ungracious correspondent ended the matter. On occasion, his work has been known to provide ammunition for others. Thus in 1887 and 1888 he wrote papers on the kinetic theory of gases, a subject which at that date led to much disagreement in opinion; stating his assumptions, he dealt with the average exchange of energy during the impact of elastic spheres and with the partition of energy between motions of translation and of rotation. These papers can only have been the outcome of some appeal emanating from Tait. The result was used (but Burnside took no direct part) in an onslaught upon Boltzmann's work made by Tait, a "bonnie fechter," never reluctant in the use of the controversial tomahawk.

In his writings, Burnside had a style which precisely, and habitually (as if it were an instinct), contributed to efficiency of presentation. Even while an undergraduate, he had been noted for the style of his mathematical work; he was reputed to be the most "elegant," though not the most widely read (Chrystal was thus reputed), among the young mathematicians of his own standing. In pure literature, critics, whether analytic or constructive, do not always agree upon the necessary essentials of general style, though they can select individual characteristics. In scientific productions, the task is assuredly no easier than in the humanities. Burnside had two of the essential secrets of an effective style: he exercised a power of clear and precise thinking that was maintained until the achievement of a definite issue; and he possessed a faculty of lucid (if condensed) expression of the whole course of a constructive argument. He was intolerant of approach to vague meandering: "Words, words" would be his caustic comment on an unconstructive passage. The elusive charm of the sudden thought, that in itself is a revelation, is rare in mathematics, though it can be found in a Fourier or a Salmon. But such was not Burnside's aim, perhaps never his dream; he did not seek for aught else than clearness, directness, terseness most of all. He would practise no art in trying to secure the attention of an inexpert beginner. In exposition, conciseness was his rule. Once, the attempt of a friend, to obtain from him a more expanded treatment of some early stages in his Theory of Groups, was met by a declaration of regret that he had been unable to effect further condensation. The consequence is that all Burnside's published work is close and firm in texture; yet, to an attentive reader, it is never lacking in clearness and movement.

Throughout Burnside's residence at Cambridge, the University had been in the finest flower of her activity in applied mathematics. Stokes, Cayley, Adams, were long-established professors; Maxwell's appointment had been more recent. The staple subjects for the most capable mathematical students were physical astronomy, dynamics, light, sound, and heat. The range of electricity and magnetism, except for a slight infusion of some of the work of Sir William Thomson (afterwards Lord Kelvin), was academic and unconnected with laboratory knowledge; and Maxwell's presentation, based on the researches of Faraday, had still to make its place in the Cambridge course, men scarcely even dreaming of the revolution it was to accomplish later. Pure mathematics, save for the rare appearance of a Clifford, a Pendlebury, or a Glaisher, was left to Cayley's domain, unfrequented by aspirants for high place in the Tripos. Much of the original thought of her mathematicians in those years found its expression in problems, a veritable mine of isolated results propounded as conundrums in the Senate House and in College examinations. Even so, the worship of the mathematical spirit at the shrine of natural philosophy was maintained in a well-defined conservative range.

At the beginning of his work, Burnside could hardly fail to conform to this Cambridge use; indeed as regards the subjects (though not as regards all methods for the subjects) in applied mathematics, he largely remained in the older round to the end. Yet even while he continued in Cambridge, he was gradually emerging into his own domain. Bred an applied mathematician in the Cambridge school of natural philosophy, which tended to regard all mathematics as a useful tool—no more than a tool—in so-called practical applications, he came to find that there was a world of pure mathematics different from that which filled the receptive stage of his student days. In the creative stage of thinking for himself beyond the range of learning and of teaching for the Tripos, he gradually made his way into that new world. He took rank with the constructive pure mathematicians, without losing hold of his earlier studies. Indeed to him, as to others with a similar experience, the new knowledge shed fresh light upon the older interests; but any effective combination of the old and the new could only be made by an intellect of the type such as Burnside happily possessed.

Thus, as already stated, Burnside's earliest advanced lectures were devoted to hydrodynamics. Elsewhere, the old-fashioned methods for conjugate functions, stream-lines, and velocity-potential, were being analytically transformed through the introduction of functions of a complex variable. For many a day, Cambridge had preserved an almost invincible repulsion to the then objectionable $\sqrt{-1}$, cumbrous devices being adopted to avoid its use or its occurrence wherever possible. But some teachers could shew that, in two-dimensional fluid motion, simplicity and new results alike were easily attainable by its means; and its formal debut within the Cambridge enclosure was made in Lamb's treatise. To Burnside's intellect the new calculus appealed; and as a matter of record, his first published paper (1883) is concerned with elliptic functions, not with hydrodynamics.

Three examples will suffice to indicate the development in Burnside's thought, thus indicated.

In 1888 he investigates three main questions connected with deep-water waves resulting from a limited initial disturbance, a research probably suggested by certain phenomena noted in the Krakatoa eruption. In that paper, he proceeds by analysis which belongs to what would now be called the classical methods of

Fresnel, Poisson, and Stokes; it requires much elaborate work in definite integrals with real variables, without any reference to the (happily satisfied) convergence of those integrals; and Burnside arrives at direct results of observable significance, which relate to the greatest amplitude of displacement, the range of propagation, and the governance of the wave-length. It is not without interest, in connection with his increasing grasp of newer methods, to note that in this paper he "justifies" the use of a complex value for a constant—while, two years later in a paper which deals with streaming motion, he uses complex variables without a word of prelude to superfluous justification.

The problem of the two-dimensional potential, as envisaged by the applied mathematicians in the middle third of the last century, such as Green, Stokes, Thomson and Tait, has been completely changed by the ideas of the theory of functions. Old assumptions have had their significance and their limitations revealed, the earlier physicists not always in sympathy with exacting refinements which to them smack of pedantry, the later mathematicians not always respectful to the intuitions content with a semblance of proof. Burnside knew both attitudes of mind—the earlier from his training, the later from his continued study; and so he could bring old results to new issues. Thus in a paper (1891) on the theory of the two-dimensional potential, satisfying the equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0,$$

and determined by prescribed conditions within an area and assigned values along a boundary, he returns to the old property —the possession of every undergraduate—that the potential can have no maximum or minimum within the boundary. Pointing out that maxima and minima must therefore lie on the boundary and that conditions of continuity require their aggregate to be an even integer, he obtains a relation between that integer, the integer denoting the number of distinct portions of the boundary, and the integer representing the number of double points on the equipotential contour lines as they pass from a boundary arc over the area back to another boundary arc. Moreover, he obtains the relation for the most general case when the conditions are extended so as to admit discontinuities (in the form of logarithmic or algebraic infinities) within the boundary; and he indicates the bearing of the relation on the graphs of these contour lines.

In 1894 he published a paper discussing Green's Function for a system of non-intersecting spheres. There, beginning with the known result for two spheres, he transformed it by a property he had deduced from a geometrical interpretation of homographic substitutions. He extended the transformed result to any number of spheres. By inversions which are represented by point transformations, and by sets of inversions which accumulate into a group of transformations, he obtains a pseudo-automorphic function, in the form of a series where the coefficients of the successive terms are powers of the magnification at the successive inversions. Lord Kelvin would not have recognised his theory of images in that final form: yet the development into that form is only a continued amplification of the theory. Burnside, moreover, carried it further, by connecting the application with any solution of Laplace's equation, instead of the inverse distance alone as in the theory of images. Here, as in all his investigations, it was only too evident that he had wandered far from the ancient Cambridge fold.

Various well-marked stages in the progress of Burnside's knowledge almost indicate themselves, from the evidence of his original papers.

Apparently, the first large new subject, of which he made a profound study, was elliptic functions: its rudiments had hardly been admitted to his Cambridge course. At every turn he devised something novel. Is it the transformation of the simplest elliptic differential element? Noting the general characteristic of the four critical points in the Riemann interpretation, he deals with the successive possibilities of the transformation: (*a*) into itself, by interchanging these four points in pairs, with the obvious inference that there are three modes, which are explicitly obtained; (*b*) into the Weierstrass normal form, with one of the critical points sent to infinity, and the remaining three practically arbitrary; (*c*) into the Legendre normal form, with the four points symmetrically arranged round the origin along an axis; and (*d*) into the Riemann normal form, with $0, 1, \infty$, as three canonical points for all, and the fourth defined by the parametric invariant of the element. Is it so simple an issue as the division of the periods by 3 or by 9? Even for the simplest form of that issue, he treats it by a general method and not by any special artifice: a short paper in 1883 achieves the trisection for the Jacobian elliptic functions; a later paper in 1887 achieves the same problem for the Weierstrass elliptic functions; a still later paper uses the same method, supplemented by the introduction of resolvents, to obtain the results for division by 9. Is it the extension of Jacobi's expression of the apparently hyperelliptic integral

$$\int \{x(1-x)(x-\lambda)(x-\kappa)(x-\kappa\lambda)\}^{-\frac{1}{2}} dx,$$

under the (quadratic) transformation

$$z = x + \frac{\kappa\lambda}{x},$$

as the sum of two elliptic integrals? Burnside deals with the cubic and the quintic transformations in odd degree, with the quartic transformation in even degree, and obtains the respective types of degenerate hyperelliptic integrals; characteristically leaving other instances as "exercises" (though, not "easy" exercises) in the method expounded. And, almost as an incident, he notes a case when an apparently elliptic integral

$$\int \frac{x-p}{x-q} \{(x-\alpha)(x-\beta)(x-\gamma)(x-\delta)\}^{\frac{1}{2}} dx,$$

where the relation

$$\frac{y-p}{y-q} = -\frac{x-p}{x-q}$$

transforms the elementary elliptic differential into itself, is only simply periodic. Or, to take only a last example in this range, he completes the known proposition that the co-ordinates of a point on the intersection of two quadrics are expressible in terms of elliptic functions, by constructing the actual arguments; and he shews that the two invariants in the Weierstrass form are the quadrinvariant and the cubinvariant of the customary quartic equation occurring in the reference of the quadrics to their common self-conjugate tetrahedron.

Another subject that absorbed his attention was differential geometry, which also, save for some rarely read sections in Salmon's *Geometry of Three Dimensions*, hardly entered into the Cambridge course. Burnside gathers together fundamental propositions, then accessible only by search among widely scattered authorities; and he applies them with effect. Before 1890, the parameters of nul lines on a surface had not appeared (or perhaps, only with Cayley) in English memoirs. In one paper, Burnside uses them, with severe ingenuity, to obtain the different classes of surfaces that possess plane lines of curvature. In another paper, he uses them to construct the differential equation of all confocal sphero-conics, proving that the co-ordinates of points are expressible in terms of elliptic functions of a parametric argument which is obtained explicitly. There, as always in his papers, Burnside's work marches forward to a definite issue and constitutes a contribution to knowledge.

Comparative simple known properties are given a widened significance. Thus he takes the known property that two finite screws compound into a single screw; and (1890) he devises a simple geometrical construction for the axis of the resultant screw. He notes that, as the proof does not require the use of parallels, the result is valid for elliptic space and for hyperbolic space. Five years later, he returns to the matter in a paper on the kinematics of non-Euclidean space; and now he notes that displacements correspond to point-transformations, sets of displacements to groups of transformations. The theory of groups is beginning to affect his work.

He can derive new results from elementary results in ordinary geometry, as well as from the range of abstract geometry. His interpretation of a homographic substitution

$$w = \frac{az + b}{cz + d}$$

as inversion at two fixed circles—this 1891 paper seems the first occasion when the specific mention of a group is made in his published work—is used to assign the criteria, necessary and sufficient, to determine whether a group, formed of assigned fundamental transformations, will or will not contain a loxodromic substitution. Or he will deal with the ancient problem of drawing a straight line between two points, for which the ruler suffices in the Euclidean postulate when the points lie at an implicitly supposed finite distance apart; and he gives a construction for the cases, when one of the points is at infinity, when both of them are at infinity, when one of them is the ideal point required in projective geometry; his construction applies to any space, Euclidean, elliptic, hyperbolic. Or he will take a proposition (analytically established) concerning the four rotations by which a triply orthogonal frame of lines can be displaced into coincidence with a similar frame; by the use of a known (Hamilton) proposition in rotations, he gives a geometrical construction for the displacement, a construction which seems almost obvious—after it has been obtained. Or he will proceed to abstract space: he discusses a configuration of 27 hyperplanes and 72 points in space of four dimensions, such that six of the planes pass through each point and sixteen of the points lie in each of the planes. To him it is a natural extension of the customary configuration of the 27 lines on an ordinary cubic surface in three dimensions.

Burnside's investigations in elliptic functions compelled him to range in the wider field of the theory of functions in general; so thither he had proceeded and, in his progress, he became an investigator.

His contributions are, as ever, varied in range. Fifty years ago, it was a surprise—to-day, it is almost a commonplace—to learn that functions of real variables exist, which are always finite, are always continuous, and never possess a determinate differential coefficient: the now classical example, due to Weierstrass, is that of the series

$$\sum_{n=0} b^n \cos a^n\theta,$$

where a is any uneven positive integer, and b is a real positive quantity such that $ab > 1 + \frac{3}{2}\pi$. Burnside made a step in advance (1894). He shewed that there are functions of real variables, everywhere finite, everywhere uniformly convergent, everywhere possessing the unrestricted complement of successive differential coefficients, yet never expansible in power-series; and, as an illustration, he constructs the function

$$\sum_{n=0}^{\infty} \frac{1}{n!} \frac{1}{1 + a^{2n}(x - \tan n\alpha)^2},$$

where a is real and >1, and where α/π is not a rational fraction. His proof is concise and demands no acquaintance with elaborate theory; as usual, it leads direct to a definite result that completes the investigation.

On another occasion, he deals with the Schwarz solution of the problem of representing a closed convex polygon in one plane conformally upon the half of another plane—a result that has rendered signal service in mathematical investigations in matters so diverse as heat, hydrodynamics, and electricity. In these last applications, only the simplest examples are used: in the general Schwarz solution, an Abelian integral occurs the use of which is gravely handicapped by its multiplicity of periods; so that additional conditions become necessary to render the analysis specific in application. Burnside, already skilled in polyhedral functions and general automorphic functions, investigates the aggregate of instances where, at the utmost, doubly-periodic functions will suffice. But he goes on to deal with multiply-connected spaces having polygonal boundaries: in particular, he gives the solution for the conformal representation of the doubly-connected area which lies between two concentric similarly placed squares, the side of one square being double that of the other.

He seizes upon the existence-theorem which establishes the possibility of expressing the co-ordinates of a point on an algebraic curve by means of uniform functions that are automorphic under sets of transformation. The lack of determination of the group, appropriate to a postulated equation, has left the solution as one merely of existence without specific determination. Burnside, combining his knowledge of groups, of elliptic functions, and of Klein's icosahedral functions, gives a complete specific resolution of the problem for the (apparently) hyperelliptic equation

$$y^2 = x(x^4 - 1).$$

It is unnecessary to accumulate more instances. Burnside's matured development flashed out in his double paper on automorphic functions, published in 1892. The subject belonged to a new section of mathematical knowledge, mainly inaugurated by Henri Poincaré and systematically expounded in a series of memoirs, now classical, in the initial volumes of *Acta Mathematica*. The underlying idea is simple. Trigonometrical functions are singly-periodic: that is, each such function is unchanged when its argument suffers an increment or a decrement which is any integer multiple of a single quantity. Elliptic functions are doubly-periodic: that is, each such function is unchanged when its argument similarly suffers an increment or a decrement which is a linear combination of any independent integer multiples of two quantities (the ratio of these quantities must not be real). Jacobi had proved long ago that uniform functions of triple periodicity (and, *à fortiori*, of periodicity higher than triple) in a single variable do not exist. But in every such instance the modification of the argument consists solely of an additive increment or decrement. The question arises: What is the most general type of periodicity for a function of one argument? And it naturally entails the further question: What are the functions possessing that type of periodicity? Isolated results were known, such as Jacobi's elliptic modular functions and Klein's polyhedral functions: their significance as examples of a wider theory had not appeared. It was Poincaré who presented the first general treatment of these questions.

Into this work of Poincaré, Burnside plunged. In it he revelled. His new results are embodied in the paper on automorphic functions which has just been cited. In particular, Poincaré had overstated an exclusive central result. Burnside detected the overstatement and the fundamental cause; and he devised a new class of automorphic functions, simpler than any of the classes devised by Poincaré. The full theory, even now, remains to be established: it awaits the construction (or the equivalent of the construction) of a central function or functions which, while palpably automorphic, shall be amenable to ordinary analytical manipulation as are the corresponding central theta-functions of purely incremental periodicity. When the history of that theory comes to be written, Burnside's name will hold an honourable place in the record.

The consideration of the very foundation of these automorphic functions led Burnside further afield, along a way already opening out before him in his progress, into a region which he explored with ample discovery. It was to provide the most continuous and most conspicuous of his contributions to his science. The characteristic property of every automorphic function of a single variable is that, without change in the value of the function, its argument is subject to a number of discrete operations, which are independent of one another, are capable of unlimited repetition and reversion, and admit all possible combinations, repetitions, and reversions, in unrestricted sequence. The aggregate of all the operations, which thus emerge, is termed a group, so that a function can be automorphic under a group of transformations (or substitutions). But just as the properties of the integers, which occur in the arithmetic of any calculation, merge into the general theory of number which ignores all specific application, so the properties of transformations in a group merge into a more comprehensive calculus. That calculus deals with the composition, the construction, the resolution, and the essential properties, of a group regarded as an abstract entity whose component elements are subjected to mathematical laws of combination. It is no part of that calculus to take account of possible regions of application: instances present themselves in algebraic equations, in analytic functions, in differential equations, in divisions of space of different orders of dimension, in the displacements of a solid body, in invariants and covariants of all kinds—a selection of subjects manifestly not complete.

The earliest expression of the notion and its initial development are due to Galois: he indicated the kind of relation that could exist between the properties of an algebraic equation and some corresponding group of finite order. The early growth of the theory was due to French mathematicians, Cauchy in particular, then Serret. Somewhat later came the fine exposition by Jordan who, it may be mentioned, had Klein and Lie as pupils at the outbreak of the Franco-Prussian war in 1870. Down to that date, the subject revolved round algebraic equations as a centre.

The interest in the theory began to spread. The next real extension was due to Sylow, in a memoir on groups of substitutions. Then followed a partial construction of its mathematics as a pure calculus, without regard to applications: the contributions of Cayley and of Weber may be noted. The theory soon divided itself into two co-ordinate sections, sometimes advancing as pure calculus, sometimes extending to new regions of application. A theory of continuous groups branched off into complete independence; it became a great body of mathematical doctrine, under the inspired researches of Sophus Lie and his disciples. The theory of discontinuous groups attracted an equally ardent band of investigators: the names of Klein, Burnside, Frobenius, Hölder, and Dyck, recall diverse developments in theory and in use.

It was to the theory of discontinuous groups of finite order that Burnside mainly devoted his attention. Scattered references to such groups occur in some of his papers already cited. At first, their occurrence seems merely incidental; then they almost prove that his thought was gradually accumulating the evidences of a connected theory. From the early nineties onward through much of the remainder of his life, Burnside's constructive thought concentrated on the subject. Paper after paper appeared from him, on a vast variety of associated topics, in ordered development, each providing some fresh contribution, all of them marked by imaginative insight and compelling power. They found their first culmination in his book on the *Theory of Groups*, published in 1897. That volume was a systematic and continuous exposition of the pure calculus of the theory as it then stood; and it embodied the researches of other workers in Europe and America (always with ample references) as well as his own. His papers on the theory of groups continued, unhastingly, unrestingly. A second edition of the book, considerably more extended than the first, appeared in 1909. Even so, his activity in the subject still continued, though with a gradually decreasing production. He published over fifty separate papers on this range of knowledge alone; each of them, even the briefest, contained some definite result or results of significance. All this work, original from himself, is a splendid contribution emanating from one mind and, of itself, is sufficient to secure the remembrance of his name.

With the coming of the war in 1914 and during its course, there was a comparative cessation in Burnside's productivity. His frame was almost as lithe as ever and apparently as full of easy spring, as though to belie the passage of years. Some of his constructive activity passed silently into the service of his country in certain naval matters. In those years he undoubtedly continued to produce papers; but the main body of his work could be regarded as verging towards its termination.

One new subject, however, secured some regular attention from him, even amid his unbroken interest in groups. It may have originated from the mathematics of some war problems, and its interest may have been fostered as he pondered over the combinations of diverging results of observations. In the year 1918 he produced a short paper dealing with a question in probability, purely mathematical as propounded; and it was followed, from time to time, by other papers, some suggested by practical problems. Probability, as a mathematical theory, has not yet lent itself to a single process of organised development based on any unique set of ideas, which are generally accepted as fundamental. Even the method of almost universal use in astronomical observations depends upon the Gauss assumption of the arithmetic mean of a number of discordant observations, as the best measure of the unknown quantity. But that assumption stands as only one out of many inferences from the less arbitrary assumption that the probability of an error, in any observation, is some function solely of the deviation from the unknown accurate measure; with that less arbitrary assumption, a more general inference is that the difference between the unknown measure and the arithmetic mean is some symmetric function of the differences between the observed magnitudes. (Of course, the occurrence of the symmetric function modifies the law of facility of error: or the adoption of an admissible law, not inconsistent with the assumption and differing from the exponential law, determines the form of the symmetric function.) Burnside deals only with the arithmetic mean: thus tacitly, with other writers, making the symmetric function to be zero. As indicated earlier, he did not consider that he had resolved all his difficulties. Ever a severe critic, he remained critical of himself; he was not afraid to modify an opinion; he did not hesitate to abandon an opinion, if ever he regarded it as not fully tenable, as indeed happened in fact. The manuscript, which he has left and which will be published by the Cambridge University Press*, is the expression of his views so far as they had been framed into a system.

There is one activity in human nature which exercises a perennial lure for living minds. When a worker of recognised distinction in any field has completed his contribution to thought, some survivors delight in assigning him his place in an ordered hierarchy of memorable names. The task demands an easy omniscience which shall gauge all knowledge and all intellect, if the estimate of precedence in relative merit is to be promulgated with authority and received with belief. Yet, somehow, such estimates lack the quality of permanence. Nearly two thousand years ago Lucretius, the brilliant expositor of natural philosophy in an age of culture, described Epicurus as a man

Qui genus humanum ingenio superavit,

a tribute paid two full centuries after the death of the Greek philosopher of the atom: the world to-day, if it ever hears of the name thus lauded, greets the judgment with a smile. Less confident men may, in their own day, render a more modest yet equally sincere homage to a passing spirit, from their reverence for the genius that has striven and in their remembrance of the worldly task that has been done. Burnside, during a life of steadfast devotion to his science, has contributed to many an issue. In one of the most abstract domains of thought, he has systematised and amplified its range so that, there, his work stands as a landmark in the widening expanse of knowledge. Whatever be the estimate of Burnside made by posterity, contemporaries salute him as a Master among the mathematicians of his own generation.

A. R. F.

November 11, 1927.

* It is embodied in the present volume.

THEORY OF PROBABILITY

CHAPTER I

INTRODUCTION

1. The words "probable" and "likely" continually occur in conversation, as also does the substantive "probability" though not so frequently.

"The glass fell a lot last night, it will probably rain today" or "The barometer has fallen half-an-inch since yesterday, there is a probability of rain before night" are statements such as we all have heard at the breakfast table. The hearer, if he treats them as anything more than attempts at starting conversation, will regard them as more or less vague judgments founded on the speaker's previous experience. He will certainly not recognize any numerical precision in them.

A more speculative acquaintance having examined the barometer might say, "I'll bet you 2 to 1 in half-crowns it will rain before night"; to which the answer might be, "No, but I'll take 3 to 1." Here both speakers do apparently make some rough kind of numerical estimate of the probability of its raining before night. Their estimates however apparently do not agree, nor would an audience infer that either speaker attached numerical precision to his estimate.

Let us take another set of statements involving the words "likely, probable, probability." The captains of two cricket teams habitually determine the choice of innings by spinning a coin. They would certainly repudiate the suggestion that a coin is more likely to fall head than to fall tail. They would assent to the statements:—

"When a coin is spun, it is equally likely to fall head or tail";

"When a coin is spun, the probability of its falling head is the same as the probability of its falling tail."

Though they might find a difficulty in explaining, without using the words "likely, probable, probability," the meaning of these statements they assent to, they undoubtedly act upon them.

In the above sentences between inverted commas, the words—probable—likely—probability—are used in a more or less vague conversational sense. Now when the calculation of probabilities is spoken of, it is implied that the probabilities in question are capable of being measured or specified by numbers. About such probabilities there can be nothing vague. It may very well be the case that some of the probabilities of ordinary conversation cannot in any way be brought under the head of calculable probabilities; while others, by making suitable assumptions, can. For instance, probabilities connected with the question of whether it will rain before night may be found to belong to the first class, while those connected with the fall of a coin may belong to the second.

A probability can only be said to be measured or represented by a number, when a rule exists by means of which the number can be calculated from a sufficiently extensive set of data. Before stating such a rule, it will be convenient to begin with explanations of both the phraseology and the notation that will be used.

When making a trial or choice is spoken of, it is implied that the result of this trial or choice is uncertain. The degree of uncertainty will depend upon the nature of the trial or choice. A somewhat typical case is the choice of an integer. When no condition is imposed on the result of the choice, the number of results is clearly unlimited. When an integer is expressed in the scale 10, the sum of its digits, s, and the number of its zero digits, z, are definite numbers. All integers may be divided into two classes: those for which s does not exceed 20, and those for which s does exceed 20. The condition that "in the integer chosen, s does not exceed 20" is a limitation on the trial or choice, for it cuts out a number of what were possible results. There are still however an unlimited number of results. In the same way the condition that "in the integer chosen, z does not exceed 10" is a limitation on the choice; and with this condition by itself satisfied the number of results is still unlimited. If, however, both conditions are satisfied, the number of results is no longer unlimited. It is clear that there must be an unlimited variety of sets of conditions, such that when those of a single set are all satisfied, the number of results of the choice is finite.

If the two above conditions are called conditions A and B, then when they are satisfied there is a finite number $n_{AB}\ (=n_{BA})$ of results of the choice.

Now consider some further condition C, for instance, that "the leading digit in the integer is 3." The number of results of the trial when, in addition to conditions A and B, a condition C is also satisfied will, in general, be less than n_{AB}. Denote it by n_{ABC}. This set of n_{ABC} results belong to the n_{AB} results; and in those of the n_{AB} results, which do not belong to the n_{ABC}, the condition C is not satisfied. If then

$$n_{AB} = n_{ABC} + n_{ABC'},$$

$n_{ABC'}$ is the number of the n_{AB} results in which the condition C is not satisfied. When n_{ABC} is greater than unity, it will clearly be possible to divide the n_{ABC} results in which conditions A, B and C are all satisfied, into two sets by means of a fourth condition D which is satisfied by some and is not satisfied by the rest. In general the $n_{ABC'}$ results will also be divided into two sets by the condition D, so that

$$n_{AB} = n_{ABCD} + n_{ABCD'} + n_{ABC'D} + n_{ABC'D'},$$

where, for instance, $n_{ABC'D}$ is the number of integers which satisfy conditions A, B, D, and do not satisfy condition C. The order in which the letters in a suffix are written in this notation is immaterial.

If any one of the numbers n_{ABCD} is greater than unity, this process may be continued by introducing new conditions.

Finally the n_{AB} results of the choice, which satisfy conditions A and B, may be distinguished from each other by a finite number of other conditions $C, D, E, F, \ldots$. To each one of them will correspond a symbol $CD \ldots E'F' \ldots$, implying that for it conditions $C, D, \ldots$ are satisfied and conditions $E, F, \ldots$ are not satisfied.

The suffix notation, that has been introduced and explained in the case of the choice of an integer, is quite independent of the particular case to which it has been applied. In the preceding illustration, with regard to conditions A, B and C which have been stated explicitly, it is clear that, as regards each, an integer must either satisfy it or not satisfy it. There are no ambiguous cases. It is assumed, once for all, that a condition introduced in connection with other trials or choices is such that a result either does satisfy it or does not. With this assumption, the suffix notation may clearly be extended to distinguish between the results of any trial or choice.

Even when the results of a trial are subjected to no conditions, their number may be limited owing to the nature of the trial itself. In such cases, the introduction of conditions to be satisfied by the results will in general involve a further limitation of the number of results.

2. With these explanations, the rule for calculating calculable probabilities may now be stated.

Rule. The results of a trial or choice, or the trial itself, or both the trial and the results, are subject to such conditions that, wherever whenever and by whomever the trial is made, there are just n possible results, of which one must occur and only one can occur. If in n_A of these results the condition A is satisfied, while in the remaining $n - n_A$ it is not satisfied, the probability that the condition A is satisfied, when a trial is made, is n_A/n; provided that each two of the n results are assumed to be equally likely.

The rule on which the calculation of probabilities depends has been stated in a variety of forms. For instance, Poincaré puts it in the following form*:—

> La probabilité d'un événement est le rapport du nombre des cas favorables à cet événement au nombre total des cas possibles; à condition que tous les cases soient également vraisemblables.

In a Note (p. 101), some reasons will be given for the form chosen here, and especially for the way in which the assumption of equal likelihood has been made. The number n_A is some integer from 0 to n, both inclusive. If n_A is neither 0 nor n, the results of the trial are divided into two sets by condition A, namely those in which condition A is satisfied and those in which it is not satisfied; but, if n_A is either zero or n, this is not so. Condition A will be said to be relevant to the trial in the first case and not relevant in the second.

Suppose now that condition A is relevant to the trial, and consider the new trial which is subject to the further condition

* *Calcul des Probabilités*, 1re éd., pp. 1, 3; 2me éd., pp. 24, 26.

that condition B is satisfied. This new trial has just n_A possible results. In n_{AB} of these, the condition B is satisfied; and in the remaining $n_A - n_{AB}$, it is not satisfied. It follows from the rule that the probability that in the new trial condition B is satisfied is n_{AB}/n_A; since each two of the n_A possible results have already been assumed to be equally likely. In the same way, in the new trial which is subject to the condition that condition A is *not* satisfied, the probability that condition B is satisfied is $n_{A'B}/n_{A'}$. Hence when the trial is made and the condition A is satisfied, the probability that condition B is satisfied is not in general the same, as when the trial is made and condition A is not satisfied.

If however condition A is not relevant to the trial, the probability, n_B/n, that the condition B is satisfied does not depend on whether n_A is zero or n; i.e. on whether A is satisfied or not.

The suffix notation, which has been introduced for distinguishing between the results of a trial, will also be used in representing probabilities. Thus p_A will denote the probability that, when the trial is made, condition A is satisfied: $p_{AB'}$ $(= p_{B'A})$ will denote the probability that condition A is satisfied and condition B is not satisfied, and so on. A convenient extension of this notation is to use $p_{(B)A}$ for the probability that, when condition B is satisfied, condition A may be satisfied.

Suppose that $A_1, A_2, \ldots, A_{s-1}$ is a set of conditions no two of which are both satisfied in any result of the trial, so that the n_{A_i} results in which A_i is satisfied are all distinct from the n_{A_j} results in which A_j is satisfied. Then

$$n = n_{A_1} + n_{A_2} + \ldots + n_{A_{s-1}} + n_{A'_1 A'_2 \ldots A'_{s-1}};$$

or, if A_s is the condition that no one of the $s-1$ conditions $A_1, A_2, \ldots, A_{s-1}$ is satisfied,

$$n = n_{A_1} + n_{A_2} + \ldots + n_{A_s}.$$

Hence, since $p_A = n_A/n$,

$$1 = p_{A_1} + p_{A_2} + \ldots + p_{A_s} \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(i)},$$

where $A_1, A_2, \ldots, A_s$ is a set of conditions of which one must be satisfied and only one can be satisfied when a trial is made. In particular,

$$1 = p_A + p_{A'}.$$

Again, since the n_{A_iB} results in which conditions A_i and B are both satisfied are distinct from the n_{A_jB} results in which conditions A_j and B are both satisfied,

$$n_B = n_{A_1B} + n_{A_2B} + \dots + n_{A_sB},$$

so that

$$p_B = p_{A_1B} + p_{A_2B} + \dots + p_{A_sB} \quad \text{.............(ii).}$$

The last equation but one may be written

$$n_B = \frac{n_{A_1B}}{n_{A_1}} n_{A_1} + \frac{n_{A_2B}}{n_{A_2}} n_{A_2} + \dots + \frac{n_{A_sB}}{n_{A_s}} n_{A_s};$$

and it has been seen that $n_{A_1B}/n_{A_1} = p_{(A_1)B}$. Hence

$$p_B = p_{(A_1)B}\, p_{A_1} + p_{(A_2)B}\, p_{A_2} + \dots + p_{(A_s)B}\, p_{A_s} \quad \text{......(iii).}$$

3. Among the conditions that are not relevant to the result of a trial, there are some which call for special notice. Consider for instance the condition—that the trial has been made at some other time or in some other place than those in question. If this were relevant, it would be satisfied by some and not satisfied by other results of the trial at the particular time and place considered; and the number of possible results at the time and place considered would be less than n, contrary to the supposition in the rule. In precisely the same way it follows that a condition—that another trial, whether of the same kind or not, shall have a particular result—is also not relevant.

If a trial is repeated, and it is proposed to consider probabilities connected with the repeated trial, it is necessary to make an assumption of equal likelihood. Suppose there are N possible results each two of which are equally likely for the repeated trial, and that in N_{ij} of them the ith result occurs at the first trial and the jth at the second. For the repeated trial, subject to the condition that the ith result occurs in the first, there are just

$$N_{i1} + N_{i2} + \dots + N_{in}$$

results; and each two of them are equally likely. Now it has just been seen that the result of the first trial is not relevant to the second, so that the probability that the jth result occurs at the second trial is

$$\frac{N_{ij}}{N_{i1} + N_{i2} + \dots + N_{in}},$$

for all values of i. From this it follows at once that

$$\frac{N_{ij}}{N} = \frac{1}{n^2},$$

or in words, each two of the n^2 results of the repeated trial, arising by combining any result at the first trial with any result at the second, are equally likely. This reasoning may clearly be used to shew that when the trial is repeated m times, each two of the n^m results are equally likely.

If attention is directed, in the repeated trial, to whether condition A is satisfied or not, the N results may be divided into four sets N_{AA}, $N_{AA'}$, $N_{A'A}$, $N_{A'A'}$ in number, the notation being that already used. Then $N_{AA}/(N_{AA}+N_{AA'})$ is the probability that, if the condition A is satisfied at the first trial, it is satisfied at the second. It has been seen that the proviso is not relevant to the result of the second trial. Hence

$$\frac{N_{AA}}{N_{AA}+N_{AA'}}=p_A.$$

Similarly
$$\frac{N_{A'A}}{N_{A'A}+N_{A'A'}}=p_A,$$

and
$$\frac{N_{AA}+N_{AA'}}{N}=p_A,$$

the last equation expressing directly that the probability, that condition A is satisfied at the first trial, is p_A. These relations give

$$\frac{N_{AA}}{N}=p_A^2,\quad \frac{N_{AA'}}{N}=\frac{N_{A'A}}{N}=p_A(1-p_A),\quad \frac{N_{A'A'}}{N}=(1-p_A)^2.$$

In precisely the same way it is shewn that if the trial is repeated $r+s$ times, the probability, that at r specified trials in the set the condition A is satisfied and that at the remaining s trials it is not satisfied, is

$$p_A^{\,r}(1-p_A)^s.$$

If a second trial has just n' possible results, wherever and whenever it is made, and if each two of these results are assumed to be equally likely, it may be proposed to deal with the probabilities regarding the results of the two trials when performed together. As in the previous cases, an assumption of equal likelihood must be made. Let there be N results for the double trial in N_{ij}, of which the ith result of the first and the jth result of the second occur. Of the N results, there are

$$N_{1j}+N_{2j}+\dots+N_{nj}$$

results satisfying the condition that the second component of the double trial has its jth result. This condition is not relevant to the result of the first component of the double trial, so that

$$N_{1j} = N_{2j} = \ldots = N_{nj},$$

for each value of j. It is similarly shewn that

$$N_{i1} = N_{i2} = \ldots = N_{in},$$

for each value of i. Hence

$$\frac{N_{pq}}{N} = \frac{1}{nn'}.$$

In words, each two of the nn' results of the double trials, formed by taking any result of the first component with any result of the second, are equally likely.

This result also may clearly be extended to a multiple trial with any number of different components.

Still using the same suffix notation, $(n_{AB'} + n_{A'B})/n$ is the probability that either condition A or condition B, but not both of them, may be satisfied. This may be rather more conveniently expressed by saying that:—

Probability that just one of the conditions A and B is satisfied

$$= p_{AB'} + p_{A'B}.$$

Similarly, the probability that at least one of the two conditions A and B is satisfied

$$= p_{AB} + p_{AB'} + p_{A'B}.$$

Now $\qquad p_A = p_{AB} + p_{AB'}, \quad p_B = p_{AB} + p_{A'B}.$

Hence the expressions

$$(\alpha) \quad p_A + p_B - 2p_{AB},$$

$$(\beta) \quad p_A + p_B - p_{AB},$$

give the probabilities that of the two conditions A and B, (α) just one, (β) at least one, is satisfied.

The corresponding formulæ, in relation to a number of conditions greater than two, will now be established.

Let $A_1, A_2, \ldots, A_n$ be n distinct conditions each of which is satisfied or is not satisfied, when a trial is made.

There are just 2^n symbols such as $p_{A_1 A_2 \ldots A_i A'_{i+1} \ldots A'_n}$ in which each suffix occurs either accented or unaccented. It is a result of (ii) above that

$$p_{A_1} = \Sigma\, p_{A_1 A_2 \ldots A_i A'_{i+1} \ldots A'_n},$$

where the summation extends to all the symbols with n suffixes in which A_1 is unaccented. Similarly,

$$p_{A_1A_2} = \Sigma p_{A_1A_2A_3\ldots A_iA'_{i+1}\ldots A'_n},$$

where the summation extends to all the symbols in which both A_1 and A_2 are unaccented; and so on.

4. A typical symbol in which r suffixes are unaccented, and $n-r$ are accented, will be denoted by

$$p_{r,\,n-r}.$$

Since r suffixes can be chosen from the n in $\binom{n}{r}$ or $\dfrac{n!}{r!\,(n-r)!}$ ways, $p_{r,\,n-r}$ stands for any one of a set of $\binom{n}{r}$ symbols, a particular one being specified by the particular set of r suffixes which are accented.

Similarly particular ones of the symbols $p_{A_1}, p_{A_1A_2}, p_{A_1A_2A_3}, \ldots$ will be denoted by $p_1, p_2, p_3, \ldots$; so that p_s, for instance, is typical of any one of a set of $\binom{n}{s}$ symbols.

In what follows, $\Sigma p_{r,\,n-r}$ denotes the sum of the $\binom{n}{r}$ symbols of which $p_{r,\,n-r}$ is the type; and Σp_s denotes the sum of the $\binom{n}{s}$ symbols of which p_s is the type.

If now Σp_r is expressed in terms of the 2^n symbols with n suffixes, any particular symbol $p_{s,\,n-s}$ will occur in the sum once for each distinct set of r unaccented symbols which it contains. Hence if $s < r$, $p_{s,\,n-s}$ will not occur at all, while if $s \geqslant r$, $p_{s,\,n-r}$ will occur $\binom{s}{r}$ times.

It follows that

$$\Sigma p_r = \sum_{s=r}^{s=n} \binom{s}{r} \Sigma p_{s,\,n-s},$$

this relation holding for all values of r from 1 to n. Hence

$$\sum_{r=t}^{r=n} (-1)^{r-t} \binom{r}{t} \Sigma p_r = \sum_{r=t}^{r=n} \sum_{s=r}^{s=n} (-1)^{r-t} \binom{r}{t}\binom{s}{r} \Sigma p_{s,\,n-s}$$

$$= \sum_{s=t}^{s=n} \sum_{r=t}^{r=s} (-1)^{r-t} \binom{r}{t}\binom{s}{r} \Sigma p_{s,\,n-s}.$$

Now $$\sum_{r=t}^{r=s}(-1)^{r-t}\binom{r}{t}\binom{s}{r}=0, \quad s \neq t,$$
$$=1, \quad s=t.$$

Hence $$\Sigma p_{t,\,n-t}=\sum_{r=t}^{r=n}(-1)^{r-t}\binom{r}{t}\Sigma p_r \ldots\ldots\ldots\ldots\text{(iv)}.$$

Further, $$\sum_{s=t}^{s=n}\Sigma p_{s,\,n-s}=\sum_{s=t}^{s=n}\sum_{r=s}^{r=n}(-1)^{r-s}\binom{r}{s}\Sigma p_r$$
$$=\sum_{r=t}^{r=n}\sum_{s=t}^{s=r}(-1)^{r-s}\binom{r}{s}\Sigma p_r;$$

and since $$\sum_{s=t}^{s=r}(-1)^{r-s}\binom{r}{s}=(-1)^{r-t}\binom{r-1}{t-1},$$

it follows that
$$\sum_{s=t}^{s=n}\Sigma p_{s,\,n-s}=\sum_{r=t}^{r=n}(-1)^{r-t}\binom{r-1}{t-1}\Sigma p_r \ldots\ldots\ldots\text{(v)}.$$

Now $\Sigma p_{t,\,n-t}$ is the sum of the $\binom{n}{t}$ symbols $p_{A_1A_2\ldots A_tA'_{t+1}\ldots A'_n}$ in which just t of the n suffixes are accented. It is therefore the probability that just t of the n conditions are satisfied; and similarly $\sum_{s=t}^{s=n}\Sigma p_{s,\,n-s}$ is the probability that either t or more of the n conditions, in other words that at least t of the n conditions, are satisfied. The formulæ (iv) and (v) therefore give the extension to the case of n conditions of the previous formulæ (α) and (β) for the case of two. In particular, the probability that at least one of the n conditions shall be satisfied is
$$\Sigma p_1-\Sigma p_2+\Sigma p_3-\ldots+(-1)^{n+1}p_n;$$
and therefore the probability that no one of them is satisfied is
$$1-\Sigma p_1+\Sigma p_2-\ldots+(-1)^n p_n.$$

5. Returning to the formulæ
$$p_A=p_{AB}+p_{AB'},\quad p_B=p_{AB}+p_{A'B},$$
it follows that
$$p_Ap_B=p_{AB}(p_{AB}+p_{AB'}+p_{A'B})+p_{AB'}p_{A'B}$$
$$=p_{AB}(1-p_{A'B'})+p_{AB'}p_{A'B}.$$
Hence if the condition
$$p_{AB}p_{A'B'}-p_{AB'}p_{A'B}=0$$
is satisfied, then
$$p_{AB}=p_Ap_B,\qquad p_{AB'}=p_A(1-p_B),$$
$$p_{A'B}=(1-p_A)p_B,\quad p_{A'B'}=(1-p_A)(1-p_B).$$

In this case, the conditions A and B are generally said to be independent.

It might be expected that when, in this sense, each pair of the three conditions A, B, C are independent, all probabilities connected with them could be expressed in terms of p_A, p_B and p_C; but this is not necessarily the case.

Suppose that

$$p_{AB} = p_A p_B, \quad p_{AC} = p_A p_C, \quad p_{BC} = p_B p_C,$$

and that at the same time

$$p_{ABC} = p_A p_B p_C + k.$$

Then

$$\begin{aligned} p_{A'BC} = p_{BC} - p_{ABC} &= (1 - p_A) p_B p_C - k, \\ p_{AB'C} &= (1 - p_B) p_A p_C - k, \\ p_{ABC'} &= (1 - p_C) p_A p_B - k. \end{aligned}$$

Also

$$\begin{aligned} p_{AB'C'} = p_{B'C'} - p_{A'B'C'} &= (1 - p_B)(1 - p_C) - p_{A'B'C'}, \\ p_{A'BC'} &= (1 - p_A)(1 - p_C) - p_{A'B'C'}, \\ p_{A'B'C} &= (1 - p_A)(1 - p_B) - p_{A'B'C'}. \end{aligned}$$

Entering these values in

$$p_{ABC} + p_{A'BC} + p_{AB'C} + p_{ABC'} + p_{AB'C'} + p_{A'BC'} + p_{A'B'C} + p_{A'B'C'} = 1,$$

it follows that

$$p_{A'B'C'} = (1 - p_A)(1 - p_B)(1 - p_C) - k.$$

Hence when A and B, B and C, C and A, are respectively independent, all the probabilities connected with them can be expressed in terms of p_A, p_B, p_C and another number k, which may have any value between $-k_1$ and k_2, where k_1 is the greatest of

$$p_A p_B p_C, \qquad p_A (1 - p_B)(1 - p_C),$$
$$p_B (1 - p_A)(1 - p_C), \quad p_C (1 - p_A)(1 - p_B),$$

and k_2 is the least of

$$(1 - p_A)(1 - p_B)(1 - p_C), \quad (1 - p_A) p_B p_C,$$
$$(1 - p_B) p_A p_C, \qquad (1 - p_C) p_A p_B.$$

Also

$$p_{A'BC} = p_B p_C - p_{ABC},$$
$$p_{AB'C'} = (1 - p_B)(1 - p_C) - p_{A'B'C'},$$

$$\begin{aligned} p_{ABC} + p_B p_C + p_C p_A + p_A p_B - 3p_{ABC} \\ + 3 - 2(p_A + p_B + p_C) + p_B p_C + \ldots - 3p_{A'B'C'} + p_{A'B'C'} = 1, \\ 2 - 2(p_A + p_B + p_C) + 2(p_B p_C + \ldots) - 2p_{ABC} - 2p_{A'B'C'} = 0, \\ (1 - p_A)(1 - p_B)(1 - p_C) + p_A p_B p_C - p_{ABC} - p_{A'B'C'} = 0, \end{aligned}$$

$$p_{AB'C'} = (1 - p_B)(1 - p_C) - (1 - p_A)(1 - p_B)(1 - p_C) - p_A p_B p_C + p_{ABC}$$

$$= p_A (1 - p_B)(1 - p_C) - p_A p_B p_C + p_{ABC}$$

$$= p_{ABC} + p_A (1 - p_B - p_C).$$

The various results and formulæ that have been now deduced from the rule, especially formula (iii), will be found to simplify very materially the calculation of probabilities in complicated cases. In particular, they enable the calculator to dispense with a continual reference to the rule by utilizing probabilities that have already been determined. But, in general, the real difficulties of calculation are connected with those cases, in which the number denoted by n is great. The determination of p_A involves the distinguishing and picking out of those of the n results, in which the condition A is satisfied. That it may be impossible to do this by a direct enumeration a simple example will shew.

Let us assume that when a coin is spun it is equally likely to fall head or tail. Then we have seen that when the coin is spun n times in succession each two of the 2^n results, as regards head and tail, are equally likely. When the coin is spun three times, what is the probability that there will be a sequence of at least two heads? The enumeration is quite simple. In the cases symbolized by

HHH, HHT, THH,

and in no others, the required condition is satisfied. Hence the probability is $\frac{3}{8}$.

When the coin is spun 100 times, what is the probability that there will be a sequence of at least 10 heads? The problem is of just the same nature as the preceding one, except that larger numbers are involved. The number giving the possible results contains 31 digits. Assuming that an inspection lasting one second would enable one to say whether a particular result satisfied the condition or not, it would take over 3×10^{22} years to complete the enumeration. It is not therefore going too far to say that in this case a direct enumeration for n_A is impossible. Some indirect method must be used. A large part of the following chapters will be devoted to these indirect methods and the approximate calculations which are necessarily connected with them.

Equal Likelihood.

6. Before going on to this, it is well to consider shortly the assumption or assumptions of equal likelihood that must be made, if probabilities are to be calculated by means of the rule. It is to be noticed that two calculators making the same formal assumptions of equal likelihood necessarily obtain the same numerical value for a probability, assuming them not to make mistakes of calculation. The resulting value in no way depends on the meaning that either calculator may attach to the assumptions, nor on whether or no the assumptions appear to them reasonable assumptions. So far, in fact, as the calculations go, the assumptions are purely formal. It is only when the calculated probabilities are applied to questions of interest outside the calculations themselves that the assumptions cease to be merely formal. These applications are continually being made. As a particular instance, a good deal of modern molecular physics is bound up with certain calculated probabilities. When such applications are made, the assumptions of equal likelihood, on which the calculations are made, can no longer be regarded as purely formal. They become in fact, directly or indirectly, assumptions about physical phenomena; and the question of whether the assumptions are reasonable becomes at once of fundamental importance.

It is quite obvious that two different assumptions of equal likelihood will, in general, lead to different values of the calculated probabilities. It has been seen that the probability of a sequence of at least two heads, when a coin is spun three times, is $\frac{3}{8}$; the assumption having been made that at a single spin head and tail are equally likely. No surprise would be felt at getting a number other than $\frac{3}{8}$ had a different assumption been made as regards the result of a single spin. The apparently paradoxical result of getting two different values for the same probability is always to be explained in this way. One of the most noted of these is due to M. Bertrand. He solves* the question:—

> What is the probability that a chord of a given circle drawn at random is greater than the side of the inscribed equilateral triangle?

* *Calcul des Probabilités*, p. 4; the question is asked for 'smaller than' and is solved for 'greater than.'

He carries out the calculation in three different ways and arrives at the results $\frac{1}{3}$, $\frac{1}{2}$, $\frac{1}{4}$, respectively. An analysis of the three calculations shews that the assumptions of equal likelihood are not the same in the three ways. This fact however is masked by the comparative complication of the assumptions; hence the apparent paradox.

Another possibility that may be mentioned here is that the data, from which it is proposed to calculate a probability, are insufficient whatever assumption of equal likelihood is made. A well-known example of this is given by the following question due* to Professor Boole:—

> The *à priori* probabilities of two causes A_1 and A_2 are c_1 and c_2 respectively. The probability that, if cause A_1 occurs, an event E will accompany it (whether as consequence of A_1 or not) is p_1; and the probability that E will accompany A_2, when A_2 occurs, is p_2. The event E cannot happen in the absence of both causes A_1 and A_2. What is the probability of the event E?

With the notation used in this chapter, there are eight possibilities to be taken into account, denoted by

$$A_1A_2E,\quad A_1A_2'E,\quad A_1'A_2E,\quad A_1'A_2'E,$$
$$A_1A_2E',\quad A_1A_2'E',\quad A_1'A_2E',\quad A_1'A_2'E'.$$

In any case,

$$p_E = p_{A_1A_2E} + p_{A_1A_2'E} + p_{A_1'A_2E} + p_{A_1'A_2'E}$$
$$= p_{A_1E} + p_{A_2E} - p_{A_1A_2E} + p_{A_1'A_2'E}.$$

The data give

$$p_{A_1E} = c_1p_1,\quad p_{A_2E} = c_2p_2,\quad p_{A_1'A_2'E} = 0,$$

so that

$$p_E = c_1p_1 + c_2p_2 - p_{A_1A_2E}.$$

Since the data give no information at all about the simultaneous occurrence of the causes A_1 and A_2, nothing is known about $p_{A_1A_2E}$, other than the necessary relations that it is equal to or less than both p_{A_1E} and p_{A_2E}. Hence the data are insufficient to determine the probability of E. It is remarkable that Prof. Boole himself and other distinguished mathematicians arrived at definite values of p_E, which did not agree with each other.

* *Laws of Thought*, p. 321.

CHAPTER II

DIRECT CALCULATION OF PROBABILITIES

7. The methods and formulæ of the previous chapter will now be illustrated by considering a number of particular examples. In each set of cases the assumption of equal likelihood will be stated explicitly. The words "chance" and "probability" will be used indifferently as being equivalent to each other.

The first set of illustrations will be drawn from the game of bridge. There are 52!/39! 13! ways in which a set of 13 cards can be taken from a pack of 52. It will be assumed that each two of these are equally likely.

I. What is the chance of holding at least one long suit (i.e. a suit of five or more) at bridge?

In a hand which does not contain a long suit, the distribution of the 13 cards in four suits must be according to one of the schemes

(i) 4, 3, 3, 3; (ii) 4, 4, 3, 2; (iii) 4, 4, 4, 1.

The number of sets of 13 cards, of which 4 are hearts, 3 diamonds, 3 clubs and 3 spades, is

$$\frac{13!}{9!\,4!}\left(\frac{13!}{10!\,3!}\right)^3.$$

The suit of 4 cards may be any one of the four suits. Hence the number of hands corresponding to the first scheme is

$$4\cdot\frac{13!}{9!\,4!}\left(\frac{13!}{10!\,3!}\right)^3.$$

The number of ways, in which 4 hearts, 4 diamonds, 3 clubs and 2 spades may be chosen, is

$$\left(\frac{13!}{9!\,4!}\right)^2\frac{13!}{10!\,3!}\cdot\frac{13!}{11!\,2!}.$$

There are six ways of choosing the two 4-suits, and then the other two may be taken in two ways. Hence the number of hands corresponding to the second scheme is

$$12\left(\frac{13!}{9!\,4!}\right)^2\frac{13!}{10!\,3!}\cdot\frac{13!}{11!\,2!}.$$

Similarly it may be shewn that the number of hands corresponding to the third scheme is

$$4\left(\frac{13!}{9!\,4!}\right)^3 . 13.$$

Hence the chance of *not* holding a long suit is

$$\frac{4 \cdot \frac{13!}{9!\,4!}\left(\frac{13!}{10!\,3!}\right)^3 + 12\left(\frac{13!}{9!\,4!}\right)^2 \frac{13!}{10!\,3!} \cdot \frac{13!}{11!\,2!} + 4\left(\frac{13!}{9!\,4!}\right)^3 . 13}{\frac{52!}{39!\,13!}}$$

After reduction this is found to be ·351, to three places. Hence the chance of holding a long suit is ·649.

II. What is the chance that one's hand at bridge shall contain just one card of some suit?

A single heart may be chosen in 13 ways; and the remaining 12 cards may be chosen from diamonds, clubs and spades in $39!/27!\,12!$ ways. The same applies to a single card of any other suit. Hence the required chance is

$$\frac{4 . 13 . \frac{39!}{27!\,12!}}{\frac{52!}{39!\,13!}}$$

On reduction this is found to be ·320, to three places.

III. The number of ways, in which a hand at bridge can be chosen to hold n (< 13) assigned cards, is the number of ways in which $13 - n$ cards can be selected from $52 - n$. Hence the probability that a hand at bridge contains n assigned cards is

$$\frac{13!}{(13-n)!} \cdot \frac{(52-n)!}{52!}.$$

For $n = 1, 2, 3, 4$, this gives $\frac{1}{4}$, $\frac{1}{17}$, $\frac{11}{850}$, $\frac{11}{4165}$.

IV. What is the chance in a hand at bridge of holding (i) at least one ace, (ii) just one ace, (iii) at least two aces?

From the previous case the chances of holding one, two, or three assigned aces, or all four, are

$$\tfrac{1}{4}, \quad \tfrac{1}{17}, \quad \tfrac{11}{850}, \quad \tfrac{11}{4165}.$$

There are six sets of two assigned aces, and four sets of three. Hence, with the notation of (iv) and (v), p. 10,

$$\Sigma p_1 = 1, \quad \Sigma p_2 = \tfrac{6}{17}, \quad \Sigma p_3 = \tfrac{44}{850}, \quad \Sigma p_4 = \tfrac{11}{4165}.$$

It follows that the probability of holding at least one ace

$$=1-\tfrac{6}{17}+\tfrac{44}{850}-\tfrac{11}{4165}=\cdot 694, \text{ to three places;}$$

the probability of holding just one ace

$$=1-2\,.\,\tfrac{6}{17}+3\,.\,\tfrac{44}{850}-4\,.\,\tfrac{11}{4165}=\cdot 434, \text{ to three places;}$$

the probability of holding at least two aces

$$=\tfrac{6}{17}-2\,.\,\tfrac{44}{850}+3\,.\,\tfrac{11}{4165}=\cdot 255, \text{ to three places.}$$

As already stated, the assumption underlying the above calculations is that, when 13 cards are chosen from a pack of 52, each two sets are equally likely to be chosen. The question clearly arises as to whether this assumption is a necessary consequence of the assumption that, when one card is chosen from a pack of 52, each two cards are equally likely to be chosen.

Let us make the assumption that, when one object is chosen from a set of n, each two objects are equally likely to be chosen; and let us subject the choice to the condition that one particular object, say the ith, is not chosen. Then in the restricted choice, subject to this condition, there are just $n-1$ possible results; and by the assumption already made, each two of these are equally likely. Hence the probability of any one of them is $\dfrac{1}{n-1}$.

Now if it is proposed to deal with the probabilities connected with drawing two objects simultaneously from the set of n, an assumption of equal likelihood must be made. Suppose there are just N results each two of which are equally likely, and that in N_{ij} of these the ith and jth objects are chosen ($N_{ij}=N_{ji}$, $N_{ii}=0$). Imposing the further condition that one of the chosen objects is the ith, there are $\sum_j N_{ij}$ possible results of the restricted trial; and each two of these are equally likely by the assumption already made. Hence the probability that the other object chosen is the jth is $N_{ij}/\sum_j N_{ij}$, and

$$\frac{N_{ij}}{\sum_j N_{ij}}=\frac{1}{n-1}.$$

This being true for all values of i and j, it follows that

$$\frac{N_{ij}}{N}=\frac{2}{n(n-1)}.$$

In a similar way it may be shewn that, when a set of r objects are chosen from n, the equal likelihood of each two sets of r follows as a consequence of the assumption that, when *one* object is chosen, each two are equally likely to be chosen.

V. A box contains a white and b black balls, and $N\,(=p+q)$ balls are simultaneously drawn from it. If $p \leqslant a$, $q \leqslant b$, what is the chance that p white and q black are drawn? When N is given, for what value of p is this chance as great as possible?

From $a+b$ balls, N may be drawn in $(a+b)!/N!\,(a+b-N)!$ ways; and by assumption each two of these are equally likely. From a balls, p may be drawn in $a!/p!\,(a-p)!$ ways; and each of these may be combined with the $b!/q!\,(b-q)!$ ways in which q may be drawn from b. Hence the required chance is

$$\frac{a!\,b!\,N!\,(a+b-N)!}{(a+b)!\,p!\,q!\,(a-p)!\,(b-q)!}.$$

This will be greater than the chance for $p+1$ white and $q-1$ black or for $p-1$ white and $q+1$ black, if

$$q\,(a-p) < (p+1)\,(b-q+1),$$

and

$$p\,(b-q) < (q+1)\,(a-p+1).$$

Moreover, if the first of these inequalities holds, so also does the one derived from it by increasing p and diminishing q by the same number, and a similar statement is true of the second. The chance then is as great as possible when

$$\frac{p+1}{q} > \frac{a+1}{b+1} > \frac{p}{q+1},$$

or

$$p+1 > (N+1)\,\frac{a+1}{a+b+2} > p.$$

The chance is therefore greatest when p is the greatest integer in $(N+1)(a+1)/(a+b+2)$.

VI. There are n boxes of which the ith contains a_i white objects and b_i black objects. One of the boxes is chosen and from it an object is drawn. What is the chance that it is white?

If we denote by A_i the condition that the ith box is chosen, and by B the condition that a white object is drawn, then by formula (iii), p. 6,

$$p_B = \sum_{i=1}^{n} p_{A_i} p_{(A_i)B}.$$

On the assumptions, that each box is equally likely to be chosen and that from any box each object is equally likely to be drawn,

$$p_{A_i} = \frac{1}{n}, \quad p_{(A_i)B} = \frac{a_i}{a_i + b_i},$$

and
$$p_B = \frac{1}{n} \sum_i^n \frac{a_i}{a_i + b_i}.$$

VII. With the conditions of the previous case, what is the probability that N consecutive drawings give white objects, each object drawn being replaced before the next drawing: (i) when a box is chosen before each drawing, (ii) when the box chosen for the first drawing is used throughout?

Putting
$$\frac{a_i}{a_i + b_i} = p_i,$$

in the first case the probability of a white object at each drawing is

$$\frac{\Sigma p_i}{n},$$

so that (Chap. I) the probability of N consecutive white objects is

$$\left(\frac{\Sigma p_i}{n}\right)^N.$$

If the drawings are all made from the ith box, the probability of N consecutive white objects is p_i^N; and therefore, by formula (iii), p. 6, the required probability in the second case is

$$\frac{\Sigma p_i^N}{n}.$$

It follows, from a well-known inequality, that the probability in the second case (if $N > 1$) is always greater than that in the first.

VIII. What is the chance that an integer chosen from the first N integers is divisible by at least two of the four primes 2, 3, 5, 7?

Of the integers from 1 to N, the number that are divisible by a prime p is the integral part of N/p, which is represented by $[N/p]$. Hence if, when an integer is chosen from the first N, each two are equally likely to be chosen, the chance that it is divisible by p is

$$\frac{[N/p]}{N}.$$

Unless N is divisible by p, this is less than $1/p$, but its difference from $1/p$ is less than $1/N$. If N is sufficiently large the chance is very nearly $1/p$. If q is another prime, the chance that the number chosen is divisible by both p and q is

$$\frac{[N/pq]}{N};$$

and when N is large enough, this is sensibly $1/pq$. This reasoning may be repeated for more than two primes.

With the notation (p. 10) of (iv) and (v), the relations

$$\Sigma p_2 = \frac{1}{2.3} + \frac{1}{2.5} + \frac{1}{2.7} + \frac{1}{3.5} + \frac{1}{3.7} + \frac{1}{5.7},$$

$$\Sigma p_3 = \frac{1}{2.3.5} + \frac{1}{2.3.7} + \frac{1}{2.5.7} + \frac{1}{3.5.7},$$

$$\Sigma p_4 = \frac{1}{2.3.5.7}$$

are approximately true when N is large enough, the errors approaching zero as N increases. Now with these values

$$\Sigma p_2 - 2\Sigma p_3 + 3\Sigma p_4 = \tfrac{1}{3}.$$

Hence, when N is large enough, the chance that an integer, chosen from the first N integers, is divisible by at least two of the first four primes is sensibly $\frac{1}{3}$.

In the same way it may be shewn that the probability of divisibility by at least two of the first 10 primes is ·48 to two places; or by at least two of the first 20 primes is ·55. It should be noted that as the number of primes considered increases, so also must the number N to ensure reasonable accuracy in the inference.

The general statement is that, when N is large enough, the chance that an integer chosen from the first N is divisible by at least two of the set of distinct primes $p_1, p_2, \ldots, p_n$ is

$$1 - \left(1 + \frac{1}{p_1 - 1} + \frac{1}{p_2 - 1} + \ldots + \frac{1}{p_n - 1}\right)\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\ldots\left(1 - \frac{1}{p_n}\right).$$

IX. There are n letters and n corresponding envelopes, and one letter is put into each envelope. This can be done in $n!$ ways. It is assumed that each two of these distributions are equally likely. What is the probability (i) that just r letters go into their corresponding envelopes, (ii) that no letter goes into its corresponding envelope?

There are $\frac{n!}{(n-r)!}$ ways in which a given set of r letters can be put into r out of n envelopes: and in just one of these ways will each of the r letters go into its corresponding envelope, so that with the notation of (iv) and (v)

$$p_r = \frac{(n-r)!}{n!}.$$

Also r letters can be chosen out of n in $\frac{n!}{r!(n-r)!}$ ways, so that

$$\Sigma p_r = \frac{1}{r!}.$$

The probability, that just r letters go into their corresponding envelopes, is

$$\Sigma p_r - (r+1)\,\Sigma p_{r+1} + \frac{(r+1)(r+2)}{1.2}\Sigma p_{r+2} - \dots$$

$$+ (-1)^{n-r}\frac{(r+1)(r+2)-n}{(n-r)!}p_n$$

$$= \frac{1}{r!}\left\{1 - 1 + \frac{1}{1.2} - \frac{1}{1.2.3} + \dots + (-1)^{n-r}\frac{1}{(n-r)!}\right\}.$$

Unless $n-r$ is quite small, this is very nearly the same as $\frac{e^{-1}}{r!}$.

The probability, that at least one letter goes into its corresponding envelope, is

$$\Sigma p_1 - \Sigma p_2 + \dots + (-1)^{n-1}p_n$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n-1}\frac{1}{n!}.$$

It follows that, unless n is quite small, the probability that no letter goes right is sensibly e^{-1}. The numerical value is ·368, to three places.

X. A line of unit length is divided into M equal parts; and it is assumed that when a point is marked on the line it is as likely to be in any one part as in any other. N points are marked on the line. What is the chance that they all lie on a segment of the line made up of r consecutive parts?

There are $M-r+1$ segments of the line made up of r consecutive parts. That segment, which starts from the tth division reckoning from the left-hand end, will be called the tth segment. Denoting the chance that all N points lie on the part of the line common to the ith, jth, ... segments by $p_{ij\ldots}$, the required chance q is

$$\Sigma p_i - \Sigma p_{ij} + \Sigma p_{ijk} - \ldots,$$

the sums applying to all sets of 1, 2, 3, ... segments respectively; for this is the chance that there is at least one segment which contains all N points.

Now if i and j are not consecutive numbers, the part of the line common to the ith and the jth segments, is also common to the ith, jth, kth, lth, ... segments, where $k, l, \ldots$ are any numbers lying between i and j. If $j > i$, just $r+i-j$ parts are common to the ith and jth segments, so long as this number is not negative. Hence, if $r+i > j > i$,

$$p_{ij} = \left(\frac{r+i-j}{M}\right)^N = p_{ikj} = p_{iklj} = \ldots,$$

where $k, l, \ldots$ are any numbers lying between i and j; and if $j > r+i$,

$$0 = p_{ij} = p_{ikj} = p_{iklj} = \ldots.$$

If $j = i+s+1$, p_{ij} occurs once in Σp_{ij}, s times in Σp_{ijk}, $\frac{1}{2}s(s-1)$ times in Σp_{ijkl}, and so on, when the above equalities are taken account of. Hence if $s > 0$, p_{ij} will occur in the expression for q with zero as coefficient, so that

$$q = \Sigma p_i - \Sigma p_{i,\,i+1}$$

$$= (M-r+1)\left(\frac{r}{M}\right)^N - (M-r)\left(\frac{r-1}{M}\right)^N.$$

The probability, that the N points are in some segment of r parts and in no smaller segment, is

$$\Sigma p_i - 2\Sigma p_{ij} + 3\Sigma p_{ijk} - \ldots.$$

If $j = i + s + 1$, the coefficient of p_{ij} is

$$2 - 3s + 4\frac{s(s-1)}{2} - \ldots.$$

This is 1, if $s = 1$; and 0, if $s > 1$.

Hence the probability q', that there is just one segment of r parts in which all N points lie, is

$$q' = (M - r + 1)\left(\frac{r}{M}\right)^N - 2(M - r)\left(\frac{r-1}{M}\right)^N + (M - r - 1)\left(\frac{r-2}{M}\right)^N.$$

If, in the formulæ for q and q', we write

$$\frac{r}{M} = x,$$

then

$$q = Nx^{N-1} - (N-1)x^N - \frac{N \cdot N-1}{1 \cdot 2} \cdot \frac{x^{N-2}(1-x)}{M} + \ldots,$$

$$q' = \frac{N\{N - 1 - (N-3)x\}x^{N-2}}{M} - \ldots,$$

where the unwritten terms contain M^{-2} and higher negative powers. When M is very large, the unwritten terms will be very small. Hence the probability, that there is some continuous segment of the line, of length x, which contains all N points, is very nearly

$$Nx^{N-1} - (N-1)x^N;$$

and the probability, that the distance between the two extreme points lies between x and $x + \frac{1}{M}$, is very nearly

$$\frac{N\{N - 1 - (N-3)x\}x^{N-2}}{M}.$$

XI. The conditions being as in the previous example, what is the probability that, N being greater than M, at least one of the divisions shall contain no point?

If q_i is the chance that the ith division contains no point, q_{ij} the chance that neither the ith nor the jth division contains a point, and so on,

$$q_i = \left(1 - \frac{1}{M}\right)^N, \quad q_{ij} = \left(1 - \frac{2}{M}\right)^N, \quad q_{ijk} = \left(1 - \frac{3}{M}\right)^N, \ldots;$$

and the required chance is

$$q = \Sigma q_i - \Sigma q_{ij} + \Sigma q_{ijk} - \ldots,$$

so that

$$1 - q = \sum_{r=0}^{M-1} (-1)^r \cdot \frac{M!}{r!\,(M-r)!}\left(1 - \frac{r}{M}\right)^N.$$

For quite small values of M and N, the numerical value of q may be calculated from this formula; but clearly some method of approximation must be used when M and N are large.

Suppose, for instance, that

$$N = M(\log M + k),$$

where M is large and $k/\log M$ is small and positive. Then

$$\log\left(1 - \frac{r}{M}\right)^N = (\log M + k)\left(-r - \frac{r^2}{2M} - \ldots\right),$$

so that

$$\frac{M!}{r!\,(M-r)!}\left(1 - \frac{r}{M}\right)^N = \frac{e^{-kr}}{r!} \cdot \frac{M!}{M^r\,(M-r)!}\, e^{-\frac{r^2}{2} \cdot \frac{\log M + k}{M} - \ldots}$$

For any given value of r, the factor

$$\frac{M!}{M^r\,(M-r)!}\, e^{-\frac{r^2}{2} \cdot \frac{\log M + k}{M} - \ldots}$$

rapidly approaches unity as M increases, while

$$\frac{e^{-kr}}{r!}$$

rapidly approaches zero as r increases. Hence for the value of N assumed,

$$1 - q = \sum_0^\infty \frac{e^{-kr}}{r!} \text{ very nearly}$$

$$= e^{-e^{-k}} \text{ very nearly}$$

$$= e^{-Me^{-\frac{N}{M}}} \text{ nearly.}$$

As a particular case, if N is nearly equal to

$$M(\log M - \log\log 2),$$

the probability that every division contains a point is very nearly $\frac{1}{2}$.

XII. A coin is spun n times. The probability of its shewing head at the first spin is p'; while at any subsequent spin the probability, that the coin shews the same face as at the previous spin, is p. What is the probability that the coin shews head at the nth spin?

If the probability is p_n, then, by applying the formula (iii) on p. 6 to the nth spin.

$$p_n = pp_{n-1} + (1-p)(1-p_{n-1}),$$

or

$$p_n = (2p-1)\,p_{n-1} + 1 - p.$$

Similarly

$$p_{n-1} = (2p-1)\,p_{n-2} + 1 - p,$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots;$$

while, for the second spin,

$$p_2 = (2p-1)\,p' + 1 - p.$$

Hence

$$\begin{aligned} p_n &= (2p-1)^{n-1}\,p' \\ &\quad + [1 + (2p-1) + (2p-1)^2 + \dots + (2p-1)^{n-2}](1-p) \\ &= \tfrac{1}{2} + (2p-1)^{n-1}\,(p' - \tfrac{1}{2}). \end{aligned}$$

Unless p is either very small or very nearly 1, the required probability is very nearly $\frac{1}{2}$ after quite a moderate number of spins.

CHAPTER III

INDIRECT METHODS OF CALCULATING PROBABILITIES

8. If p is the probability of an event on a single trial, then p^n is the probability that it will happen at each of n consecutive trials; and therefore $1-p^n$ is the probability that it will fail to happen at least once in n consecutive trials. Hence, in N sets of n consecutive trials, the probability that the event will fail to happen at least once in each set is $(1-p^n)^N$. This probability clearly approaches zero as N increases. Therefore $1-(1-p^n)^N$, which is the probability that in one at least of the N sets of n consecutive trials the event happens every time, approaches unity as N increases.

The result may be expressed in a slightly different way, since the N sets of n consecutive trials constitute a set of Nn consecutive trials. Thus, in Nn consecutive trials, the probability that the event happens n times consecutively, once or more, the sequences consisting either of the first n, or the second n, ..., or the Nth n, is $1-(1-p^n)^N$. The probability that, at some stage of the Nn trials, the event will happen n or more times consecutively, is clearly greater than $1-(1-p^n)^N$. To take a particular case, the probability that in 7200 trials a spun coin will fall head 10 or more times consecutively is greater than $1-\left(1-\frac{1}{2^{10}}\right)^{720}$, i.e. is greater than $\frac{1}{2}$, assuming head and tail equally likely.

It should be noted that, no matter how small p and how great n may be, $1-(1-p^n)^N$ will differ very little from unity when N is large enough.

Though the above process gives a lower limit to the probability required, the limit is clearly much too small, and the true value must be arrived at in some other way. In this and a number of similar questions, which are generally classed under the head of "duration of play," the required probability can often be determined by obtaining a finite difference equation which it must satisfy.

Let p be the probability of an event at a single trial. Denote by u_M the probability that, by or before the Mth trial, the event has happened n or more times consecutively; and by v_M the probability that, at the Mth trial, the event has happened n times consecutively without having done so before the Mth trial. Then

$$u_M = u_{M-1} + v_M.$$

In order that the Mth trial may complete the first set of n consecutive happenings, the following conditions must be satisfied:

(i) It must not have happened n or more times consecutively, up to and including the $(M-n-1)$th trial;

(ii) It must not happen at the $(M-n)$th trial;

(iii) It must happen at each trial from the $(M-n+1)$th to the Mth.

The probabilities of these three independent events are

$$1-u_{M-n-1}, \quad 1-p, \quad p^n.$$

Hence $$v_M = (1-u_{M-n-1})(1-p)\,p^n,$$

or $$u_M - u_{M-1} = (1-u_{M-n-1})(1-p)\,p^n.$$

Putting $$1-u_M = w_M,$$

so that w_M is the probability that the event has not happened n times up to and including the Mth trial,

$$w_M - w_{M-1} + (p^n - p^{n+1})\, w_{M-n-1} = 0.$$

When $M = 1, 2, \ldots, n-1$, the value of w_M is unity, and $w_n = 1-p^n$. The probability that, in the first $n+1$ trials the event shall happen at least n times consecutively, is clearly $p^n + (1-p)\,p^n$; so that $w_{n+1} = 1 - 2p^n + p^{n+1}$.

To determine w_M, put

$$f(y) = \sum_1^\infty w_M y^{M-1}.$$

Then

$$\begin{aligned}\{1 - y + (p^n - p^{n+1})\, y^{n+1}\} f(y) &= w_1 + (w_2 - w_1)\, y + \ldots + (w_{n+1} - w_n)\, y^n \\ &\quad + \sum_{M=n+1}^{\infty} y^M \{w_M - w_{M-1} + (p^n - p^{n+1})\, w_{M-n-1}\} \\ &= 1 - p^n y^{n-1} - (p^n - p^{n+1})\, y^n.\end{aligned}$$

Hence w_M is the coefficient of y^{M-1} in the expansion of

$$\frac{1 - p^n y^{n-1} - (p^n - p^{n+1}) y^n}{1 - y + (p^n - p^{n+1}) y^{n+1}}$$

in ascending powers of y. Since

$$1 + yf(y) = \frac{1 - p^n y^n}{1 - y + (p^n - p^{n+1}) y^{n+1}},$$

the value of w_M is expressed rather more simply as the coefficient of y^M in this latter expression. It remains to determine this numerically.

If x_i $(i = 1, 2, \ldots, n+1)$ are the roots of the equation

$$x^{n+1} - x^n + p^n - p^{n+1} = 0,$$

which are easily shewn to be all distinct,

$$1 - y + (p^n - p^{n+1}) y^{n+1} = \prod_1^{n+1} (1 - x_i y).$$

When $1 + yf(y)$ is represented as the sum of partial fractions in the usual way, it takes the form

$$\sum_1^{n+1} \frac{x_i (x_i^n - p^n)}{x_i^n - (n+1)(p^n - p^{n+1})} \cdot \frac{1}{1 - x_i y},$$

so that

$$w_M = \sum_1^{n+1} \frac{x_i (x_i^n - p^n)}{x_i^n - (n+1)(p^n - p^{n+1})} \cdot x_i^M.$$

If p is either very small or very nearly unity, the solution of the equation for x requires special treatment. It will be assumed that this is not the case, and that n is not too small a number. It may then be shewn that the equation for x has one root very nearly equal to unity, given approximately by

$$x_1 = 1 - p^n + p^{n+1}.$$

It may also be shewn that the moduli of the remaining roots do not differ much from $p(1-p)^{\frac{1}{n}}$, and that they cannot therefore be near unity.

Hence, except for comparatively small values of M, the quantities $x_2^M, x_3^M, \ldots$ are all extremely small compared to x_1^M; and a good approximation is given by

$$w_M = \frac{(1 - p^n + p^{n+1})^n - p^n}{(1 - p^n + p^{n+1})^n - (n+1)(p^n - p^{n+1})} (1 - p^n + p^{n+1})^{M+1}.$$

In particular, if $p=\frac{1}{2}$ and n is not too small,

$$w_M = \frac{2^{n+1}-n-2}{2^{n+1}-2n-1}\left(1-\frac{1}{2^{n+1}}\right)^{M+1} \text{ very nearly,}$$

except for comparatively small values of M. For instance, if a spun coin is equally likely to fall head or tail, the probability, that in M spins it will at some stage fall head at least n times running, is nearly

$$1-\left(1-\frac{1}{2^{n+1}}\right)^{M+1}.$$

The numerical significance of such a formula as this is rather difficult to grasp. As an illustration, if the coin is spun steadily at the rate of 12 spins a minute, the probability of each of the following events is almost exactly $\frac{1}{2}$, viz.

(i) The occurrence of a sequence of 10 or more heads, in 1 hour 58 minutes;

(ii) The occurrence of a sequence of 20 or more heads, in 85 days;

(iii) The occurrence of a sequence of 40 or more heads, in 241,724 years.

In particular, the probability of the case referred to at the end of Chap. I, when $n=10$, $M=100$, is $\cdot 03$ very nearly.

Examples.

9. I. Three persons play as follows. Two play a single game; and the loser sits out, while the third person comes in for the second game. At the end of each game, the loser sits out; and the one who was not playing comes in. In each single game the two engaged have equal chances of winning. The play goes on till one player has won n consecutive games. What is the chance that the play will be over by or before the end of the Nth game?

If play is not over at the end of the Nth game, denote by $u_{N,r}$ the chance that at the end of the Nth game its winner has won r $(r=1, 2, \ldots, n-1)$ consecutive games. He has an even chance of winning the $(N+1)$th game; and therefore

$$u_{N+1,r+1}=\tfrac{1}{2}u_{N,r}, \quad (r=1, 2, \ldots, n-2).$$

Now $u_{N+1,1}$ is the chance that the winner of the Nth game loses the $(N+1)$th, whatever number of games he had won consecutively. Hence

$$u_{N+1,1}=\tfrac{1}{2}(u_{N,1}+u_{N,2}+\ldots+u_{N,n-1}).$$

Using the previous relation, this is

$$u_{N+1,1}-\frac{1}{2}u_{N,1}-\frac{1}{2^2}u_{N-1,1}-\ldots-\frac{1}{2^{n-1}}u_{N-n+2,1}=0;$$

and since $$u_{N+1,\,r+1}=\tfrac{1}{2}u_{N,\,r},$$

each of the quantities $u_{N,r}$ $(r=1, 2, \ldots, n-1)$ satisfies this linear difference equation. Now v_N, the chance that play is not over at the end of the Nth game, is given by

$$v_N=u_{N,1}+u_{N,2}+\ldots+u_{N,\,n-1}.$$

Hence $$v_{N+1}-\frac{1}{2}v_N-\frac{1}{2^2}v_{N-1}-\ldots-\frac{1}{2^{n-1}}v_{N-n+2}=0.$$

Moreover $v_N=1$, where $N=1, 2, \ldots, n-1$. Hence, if x_i $(i=1, 2, \ldots, n-1)$ are the roots of the equation

$$x^{n-1}-\frac{1}{2}x^{n-2}-\frac{1}{2^2}x^{n-3}-\ldots-\frac{1}{2^{n-1}}=0,$$

$$v_N=\sum_{i=1}^{n-1}A_i x_i^N,$$

where $$1=\sum_{i=1}^{n-1}A_i x_i^m, \quad (m=1, 2, \ldots, n-1).$$

It follows that

$$\begin{vmatrix} v_N, & x_1^N, & x_2^N, & \ldots, & x_{n-1}^N \\ 1, & x_1, & x_2, & \ldots, & x_{n-1} \\ 1, & x_1^2, & x_2^2, & \ldots, & x_{n-1}^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1, & x_1^{n-1}, & x_2^{n-1}, & \ldots, & x_{n-1}^{n-1} \end{vmatrix}=0.$$

The developed form of this determinant is easily found to be

$$v_N=\sum_{i=1}^{n-1}x_i^{N-1}\frac{f(1)}{(1-x_i)f'(x_i)},$$

where

$$f(x)=x^{n-1}-\frac{1}{2}x^{n-2}-\frac{1}{2^2}x^{n-3}-\ldots-\frac{1}{2^{n-1}}=\frac{x^n-x^{n-1}+\dfrac{1}{2^n}}{x-\frac{1}{2}}.$$

The equation $$x^n-x^{n-1}+\frac{1}{2^n}=0$$

is practically the same as that which has been discussed in the previous example. It has one root, x_1, very nearly equal to $1-\dfrac{1}{2^n}$; and the moduli of the others differ only slightly from $\frac{1}{2}$. Hence when N is large, the first term in the above formula is very much larger than any of the others, i.e.

$$v_N = \frac{x_1^{N-1} f(1)}{(1-x_1) f'(x_1)} \text{ very nearly.}$$

Now $$f'(x_1) = \frac{(n-1)x_1^n - (\frac{3}{2}n - 2)x_1^{n-1} + \frac{1}{2}(n-1)x_1^{n-2} - \frac{1}{2^n}}{(x_1 - \frac{1}{2})^2};$$

if terms containing $\frac{1}{2^{2n}}$ are neglected,

$$f'(x_1) = 2\left(1 - \frac{1}{2^n}\right)^{2n-4}.$$

Hence, when n is not too small and N is large compared to n,

$$v_N = \left(1 - \frac{1}{2^n}\right)^{N-2n+3}.$$

II. If the probability of a coin falling head is p, what is the probability that, at some stage in N consecutive spins, the number of heads exceeds the number of tails by r?

(This is the same problem as that of a person playing against another with unlimited resources, which may be expressed as follows:—If a person receives a counter each time he wins a game and pays one each time he loses, and if he starts with r counters, what is the chance that before N games are over he will have lost all his counters, assuming that the chance of his winning any games is p?)

Suppose $u_{N,r}$ is the required probability. It is clear, from considering the two possibilities with respect to the first spin, that

$$u_{N,r} = p u_{N-1,r-1} + (1-p) u_{N-1,r+1}.$$

Moreover, from the meaning of the symbols, $u_{N,0}$ is unity; and $u_{N,N}$ is p^N, for all values of N.

Consider now a sequence of $r + 2s$ spins, in which there are $r + s$ heads and s tails. At the end of the sequence, the heads exceed the tails by r. Denote by $f(r, s)$ the number of such distinct sequences, for which the heads do not exceed the tails by r until the end of the sequence. The probability of any one of these sequences is $p^{r+s}(1-p)^s$. Hence the probability, that in $r + 2s$ spins the heads exceed the tails by r at the end and by less than r at each previous stage, is $f(r, s)\, p^{r+s}(1-p)^s$.

It follows at once that

$$u_{r+2s,r} = p^r \sum_{t=0}^{t=s} f(r,t)\, p^t (1-p)^t.$$

Entering this expression for $u_{r+2s,r}$ in the above difference equation, and noting that the result is true for all values of p, it is found that

$$f(r,t) = f(r-1,t) + f(r+1,t-1).$$

Also, from the meaning of $f(r,s)$, it follows that

$$f(a,0) = 1, \quad f(0,b) = 0,$$

for all values of a and b. Direct calculation gives

$$f(r,1) = r, \quad f(r,2) = \tfrac{1}{2} r(r+3), \quad f(r,3) = \tfrac{1}{6} r(r+4)(r+5).$$

This suggests that

$$f(r,s) = \frac{1}{s!} r(r+s+1)(r+s+2)\dots(r+2s-1),$$

which is verified immediately on entering this value in the functional equation. Hence

$$u_{r+2s,r} = p^r \sum_0^s \frac{r(r+t+1)\dots(r+2t-1)}{t!} p^t (1-p)^t.$$

This is clearly also the value of $u_{r+2s+1,r}$.

The numerical determination of $u_{r+2s,r}$ for given values of r and s from this formula would be laborious. It may however be shewn that, for all values of x from 0 to $\frac{1}{4}$ inclusive, the series

$$1 + rx + \frac{r(r+3)}{1.2} x^2 + \dots + \frac{r(r+t+1)\dots(r+2t-1)}{t!} x^t + \dots$$

is convergent and has the sum

$$\left(\frac{1-\sqrt{1-4x}}{2x}\right)^r,$$

where $\sqrt{1-4x}$ denotes the positive square root. When $p=\frac{1}{2}$, $p(1-p)=\frac{1}{4}$; and for any other value of p, $p(1-p)$ is less than $\frac{1}{4}$. Also when $x = p(1-p)$, $\sqrt{1-4x}$ is $1-2p$ or $2p-1$, according as p is less or greater than $\frac{1}{2}$. The series in $u_{r+2s,r}$, when taken to infinity, is therefore $(1-p)^{-r}$ or p^{-r}, according as p is less than or greater than $\frac{1}{2}$. It follows that, when N is great enough,

$$u_{N,r} = \left(\frac{p}{1-p}\right)^r, \quad p < \tfrac{1}{2}$$
$$= 1, \qquad p > \tfrac{1}{2}.$$

It has been seen that, however small p may be, the probability of a run of r consecutive heads approaches unity, as N is taken greater for all values of r.

The present section shews that, if $p < \frac{1}{2}$, the probability of the heads exceeding the tails by r, however large N may be, diminishes towards zero as r increases.

III. What is the probability that, in a sequence of M spins, the tails shall never be in excess of the heads, assuming head and tail equally likely?

This question has several interesting applications. Consider first the case, in which a sequence of $2N$ spins results in N heads and N tails. Denote by $\psi(N)$ the number of these sequences, in which the heads are in excess of the tails *at every stage except the last*; and by $\phi(N)$ the number of the sequences, in which the heads are either equal to or in excess of the tails *at each stage.* In each of the $\psi(N)$ sequences, the first two spins must be head and the last must be tail. Removing the first and the last, there remains a sequence of $N-1$ heads and $N-1$ tails, in which the heads are equal to or in excess of the tails at each stage. Conversely, by prefixing a head and annexing a tail to each sequence of $N-1$ heads and $N-1$ tails in which the heads are equal to or in excess of the tails at each stage, a sequence of N heads and N tails is formed such that the heads are in excess of the tails at each stage except the last. Hence

$$\psi(N) = \phi(N-1).$$

Now each of the $\phi(N)$ sequences must either be a $\psi(N)$ sequence, or there must be a maximum value of n, such that it begins with a $\psi(n)$ sequence; and in the latter case, what follows the first $2n$ spins must form a $\phi(N-n)$ sequence. Hence

$$\phi(N) = \sum_{n=1}^{n=N} \psi(n)\,\phi(N-n),$$

with the convention $\phi(0) = 1$. It is also clear that $\psi(1) = 1$.

Combining these two equations, it follows that

$$\psi(N+1) = \sum_{n=1}^{n=N} \psi(n)\,\psi(N-n+1).$$

Now put $$f(x) = \sum_{1}^{\infty} \psi(a)\, x^a.$$

Then $$[f(x)]^2 = \sum_{a,b=1}^{\infty} \psi(a)\,\psi(b)\,x^{a+b}$$
$$= \sum_{n=2}^{n=\infty} \sum_{a=1}^{a=n-1} \psi(a)\,\psi(n-a)\,x^n$$
$$= \sum_{n=2}^{n=\infty} \psi(n)\,x^n = f(x) - x.$$

Since $\psi(a)$ is essentially positive, this gives
$$2f(x) = 1 - \sqrt{1-4x},$$
where the positive square root must be taken, and the infinite series used is convergent if $4x$ is less than unity. Then, comparing coefficients,
$$\psi(N) = \frac{(2N-2)!}{N!\,(N-1)!}, \quad \phi(N) = \frac{2N!}{(N+1)!\,N!}.$$

Suppose now that there are just $F(M)$ sequences of M spins, in which the tails are at no stage in excess of the heads. If M is odd, say $2N-1$, we can pass to a sequence of $2N$ spins, in which the tails are at no stage in excess of the heads, by annexing either a head or tail to the sequence of $2N-1$; and in this way all sequences of $2N$, satisfying the condition, may be formed.

Hence $$F(2N) = 2F(2N-1).$$

If M is even, say $2N$, a tail cannot be annexed to any one of the $\phi(N)$ sequences, in which the heads are equal to or in excess of the tails at every stage except the last. Hence
$$F(2N+1) = 2F(2N) - \phi(N).$$
Combining these two equations, we have
$$F(2N) - 4F(2N-2) = -2\phi(N-1).$$

Since $F(2)$ is 2, this gives
$$F(2N) = 2^{2N-1} - \sum_{r=1}^{r=N-1} 2^{2N-2r-1}\,\phi(r).$$
It follows that in a sequence of $2N$ spins the probability, that the tails are never in excess of the heads, is
$$\frac{1}{2}\left[1 - \sum_{1}^{N-1} \frac{2r!}{2^{2r}\,(r+1)!\,r!}\right].$$
This may also be put in the form
$$\frac{1}{2}\sum_{N}^{\infty} \frac{2r!}{2^{2r}\,(r+1)!\,r!};$$

and it will be shewn in the next chapter (p. 42) that, if N is not too small, it is very nearly equal to $\frac{1}{\sqrt{\pi N}}$.

A more general question of the type of those just considered is the determination of the probability that, in a sequence of spins, the heads shall never exceed the tails by more than r, and the tails shall never exceed the heads by more than s. The question may also be put as follows.

IV. A and B play a game at which A's chance of losing is p. To begin with, A has $2r$ counters and B has $2s$, where $r+s=n$. Each time A wins a game, B pays him a counter; and each time he loses a game, he pays B a counter. What is the probability that A will have lost all his counters before or at the end of the $2N$th game?

Let $u_{r,N}$ denote the probability; so that, for all values of N, $u_{0,N}=1$, $u_{r,N}=0$. The chance that A loses two games consecutively is p^2, the chance that he gains one and loses the other is $2p(1-p)$, and the chance that he gains both is $(1-p)^2$. Hence

$$u_{1,N}=p^2+2p(1-p)\,u_{1,N-1}+(1-p)^2\,u_{2,N-1},$$

$$u_{2,N}=p^2\,u_{1,N-1}+2p(1-p)\,u_{2,N-1}+(1-p)^2\,u_{3,N-1},$$

....

$$u_{r,N}=p^2\,u_{r-1,N-1}+2p(1-p)\,u_{r,N-1}+(1-p)^2\,u_{r+1,N-1},$$

....

$$u_{n-1,N}=p^2\,u_{n-2,N-1}+2p(1-p)\,u_{n-1,N-1}.$$

This is a system of linear difference equations for

$$u_{1,N},\ u_{2,N},\ \ldots,\ u_{n-1,N}.$$

If the particular solution, independent of N, is

$$u_{r,N}=C_r,$$

then $$p^2C_{r-1}+[2p(1-p)-1]\,C_r+(1-p)^2C_{r+1}=0,$$

with the conditions

$$C_0=1,\quad C_n=0.$$

These conditions give

$$C_r=p^{2r}\,\frac{p^{2n-2r}-(1-p)^{2n-2r}}{p^{2n}-(1-p)^{2n}}.$$

If now
$$u_{r,N} = C_r + A_r \alpha^N,$$
then
$$\begin{aligned} A_1\alpha &= \qquad\quad 2p(1-p)A_1 + (1-p)^2 A_2, \\ A_2\alpha &= p^2A_1 + 2p(1-p)A_2 + (1-p)^2A_3, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ A_r\alpha &= p^2A_{r-1} + 2p(1-p)A_r + (1-p)^2A_{r+1}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ A_{n-1}\alpha &= \qquad\quad p^2A_{n-2} + 2p(1-p)A_{n-1}; \end{aligned}$$
and α is given by
$$\begin{vmatrix} 2p(1-p)-\alpha, & (1-p)^2, & 0, & 0, & 0, & 0 \\ p^2, & 2p(1-p)-\alpha, & (1-p)^2, & 0, & 0, & 0 \\ 0, & p^2, & 2p(1-p)-\alpha, & (1-p)^2, & 0, & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0, & 0, & 0, & 0, & p^2, & 2p(1-p)-\alpha \end{vmatrix} = 0,$$
where there are $n-1$ rows and columns.

Let x_1 and x_2 denote the roots of the equation
$$p^2x^2 + \{2p(1-p)-\alpha\}x + (1-p)^2 = 0.$$
The determinant equation is
$$\frac{x_1^n - x_2^n}{x_1 - x_2} = 0;$$
on putting
$$\alpha = 2p(1-p)(1+\cos\theta),$$
the determinant equation becomes
$$\frac{\sin n\theta}{\sin\theta} = 0,$$
so that the $n-1$ values of α are
$$4p(1-p)\cos^2\frac{t\pi}{2n}, \qquad (t = 1, 2, \ldots, n-1).$$

The difference equation for the coefficients A_r is
$$(1-p)^2A_{r+1} + \{2p(1-p)-\alpha\}A_r + p^2A_{r-1} = 0,$$
giving
$$A_r = \left(\frac{p}{1-p}\right)^r \left[B_t \sin\frac{rt\pi}{n} + B_t' \cos\frac{rt\pi}{n}\right].$$

Since $u_{0,N} = 1$, $u_{n,N} = 0$, it follows, taking account of the particular solution, that $A_0 = A_n = 0$; and therefore $B_t' = 0$. Hence
$$u_{r,N} = p^{2r}\frac{p^{2n-2r} - (1-p)^{2n-2r}}{p^{2n} - (1-p)^{2n}} + \sum_{t=1}^{n-1}\left(\frac{p}{1-p}\right)^r B_t \sin\frac{rt\pi}{n}\left\{4p(1-p)\cos^2\frac{t\pi}{2n}\right\}^N.$$

The $n-1$ constants B_t will be determined by the conditions

$$u_{1,1}=p^2, \quad u_{2,1}=u_{3,1}=\ldots=u_{n-1,1}=0.$$

The probability, that B will lose all his $2s$ counters by or before the end of the $2N$th game, say $v_{s,N}$, will be found, by putting $1-p$ for p and $n-r$ for r in $u_{r,N}$. The probability, that neither player has lost the whole of his counters at the end of the $2N$th game, is $1-u_{r,N}-v_{s,N}$. This is the probability that in $2N$ spins with a coin, for which the chance of falling head is p, the excess of heads above tails at each stage is between $2r-2$ and $-2s+2$.

If $p=\frac{1}{2}$,

$$u_{r,N}+v_{s,N}=1+\sum_{t=1}^{t=n-1} B_t\left\{\sin\frac{rt\pi}{n}+\sin\frac{(n-r)\,t\pi}{n}\right\}\cos^{2N}\left(\frac{t\pi}{2n}\right)$$

$$=1+2\Sigma' B_t \sin\frac{rt\pi}{n}\cos^{2N}\left(\frac{t\pi}{2n}\right),$$

where Σ' is the sum for *odd* values of t from 1 to $n-1$.

To fix the ideas, suppose n even and equal to $2m$. Then if $w_{r,s,N}$ is the probability that, at each stage in $2N$ spins of a coin for which head and tail are equally likely, the excess of heads above tails lies between $2r-2$ and $-2s+2$ inclusive,

$$w_{r,s,N}+2\sum_{i=1}^{i=m} B_{2i-1}\sin\frac{r\,(2i-1)\,\pi}{2m}\cos^{2N}\frac{(2i-1)\,\pi}{4m}=0;$$

and
$$w_{1,2m-1,1}=\tfrac{3}{4}, \quad w_{r,2m-1,1}=1 \quad (r>1).$$

Hence

$$0=\begin{vmatrix} \frac{1}{2}w_{r,s,N}, & \sin\frac{r\pi}{2m}\cos^{2N}\frac{\pi}{4m}, & \sin\frac{3r\pi}{2m}\cos^{2N}\frac{3\pi}{4m}, & \ldots, & \sin\frac{(2m-1)\,r\pi}{2m}\cos^{2N}\frac{(2m-1)\,\pi}{4m} \\ \frac{3}{8}, & \sin\frac{\pi}{2m}\cos^{2}\frac{\pi}{4m}, & \sin\frac{3\pi}{2m}\cos^{2}\frac{3\pi}{4m}, & \ldots, & \sin\frac{(2m-1)\,\pi}{2m}\cos^{2}\frac{(2m-1)\,\pi}{4m} \\ \frac{1}{2}, & \sin\frac{2\pi}{2m}\cos^{2}\frac{\pi}{4m}, & \sin\frac{6\pi}{2m}\cos^{2}\frac{3\pi}{4m}, & \ldots, & \sin\frac{2\,(2m-1)\,\pi}{2m}\cos^{2}\frac{(2m-1)\,\pi}{4m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{2}, & \sin\frac{m\pi}{2m}\cos^{2}\frac{\pi}{4m}, & \sin\frac{3m\pi}{2m}\cos^{2}\frac{3\pi}{4m}, & \ldots, & \sin\frac{m\,(2m-1)\,\pi}{2m}\cos^{2}\frac{(2m-1)\,\pi}{4m} \end{vmatrix}$$

determines $w_{r,s,N}$. When N is sufficiently great,

$$\cos^{2N}\frac{(2i-1)\,\pi}{4m}, \qquad (i=2, 3, \ldots, m)$$

is very small compared to

$$\cos^{2N}\frac{\pi}{4m};$$

so that, for large values of N,

$$w_{r,s,N} = A\cos^{2N}\frac{\pi}{4m}\sin\frac{r\pi}{2m},$$

where A is very nearly constant.

V. A box contains n objects; when one is drawn, each is equally likely to be drawn. An operation consists of drawing an object from the box and replacing it by a white object. What is the probability that, after r operations, there are just x white objects in the box, assuming that there were a white objects initially?

Denote the required probability by $p_{x,r}$. If x is less than a, $p_{x,r} = 0$ for all values of r.

If there are just x white objects after $r+1$ operations, there must be either x or $x-1$ after r operations.

If there are x after r operations, a white one must be drawn in order that there may be x after $r+1$ operations; the probability of this is $\frac{x}{n}$. If there are $x-1$ after r operations, no one of them must be drawn in order that there may be x white after $r+1$ operations; the probability of this is $\frac{n-x+1}{n}$. Hence

$$p_{x,r+1} = \frac{x}{n}p_{x,r} + \frac{n-x+1}{n}p_{x-1,r},$$

for all values of x from a to n.

Taking a for x, this gives

$$p_{a,r} = \left(\frac{a}{n}\right)^r.$$

Taking $a+1$ for x, it gives

$$p_{a+1,r+1} - \frac{a+1}{n}p_{a+1,r} = \frac{n-a}{n}p_{a,r} = \frac{(n-a)a^r}{n^{r+1}}.$$

The general solution of this is

$$n^r p_{a+1,r} = C(a+1)^r - (n-a)a^r;$$

and since

$$p_{a+1,0} = 0,$$

we find

$$p_{a+1,r} = (n-a)\left[\left(\frac{a+1}{n}\right)^r - \left(\frac{a}{n}\right)^r\right].$$

Taking $a+2$ for x,

$$p_{a+2,\,r+1} - \frac{a+2}{n}\, p_{a+2,\,r}$$

$$= \frac{n-a-1}{n}\, p_{a+1,\,r} = (n-a-1)(n-a)\{(a+1)^r - a^r\},$$

the solution of which, when $p_{a+2,\,0} = 0$, is

$$p_{a+2,\,r} = \frac{(n-a-1)(n-a)}{2}\left[\left(\frac{a+2}{n}\right)^r - 2\left(\frac{a+1}{n}\right)^r + \left(\frac{a}{n}\right)^r\right].$$

There is no difficulty in verifying, by an induction, the general formula suggested by these particular cases, viz.

$$p_{x,\,r} = \frac{(n-a)!}{(n-x)!\,(x-a)!}\left[\left(\frac{x}{n}\right)^r - (x-a)\left(\frac{x-1}{n}\right)^r + \frac{(x-a)(x-a-1)}{1\,.\,2}\left(\frac{x-2}{n}\right)^r + \ldots + (-1)^{x-a}\left(\frac{a}{n}\right)^r\right].$$

The probability, that after $r\,(>n-a)$ operations all the objects in the box are white, is

$$1 - (n-a)\left(1-\frac{1}{n}\right)^r + \frac{(n-a)(n-a-1)}{1\,.\,2}\left(1-\frac{2}{n}\right)^r - \ldots.$$

When $r = \rho n$, and n is not too small, this is sensibly

$$(1-e^{-\rho})^{n-a}.$$

For instance, if $n = 100$, $a = 0$, the probability, that the objects will all be white after 1200 operations, exceeds $\frac{999}{1000}$.

CHAPTER IV

METHODS OF APPROXIMATION

10. It will have been seen in the previous chapters that, while the formula for a probability connected with a comparatively small number of trials is often a complicated numerical function, the approximate expression for the probability, when the number of trials is great, takes a relatively simple form.

In the present chapter, it is proposed to obtain approximate expressions for the probabilities of various results, on the understanding that the number of trials is very great.

The particular case chosen for investigation is that of a series of N spins of a coin, where N is a large number; but it will be clear that the method of approximation has other applications. The case, in which the coin is equally likely to fall head or tail, is first dealt with.

11. A. In a sequence of N spins there are 2^N possible results, all of which are equally probable, taking account of the order in which heads and tails follow each other. To determine the number of these which give r heads and $N-r$ tails, is the same as finding the number of distinct ways in which r things may be chosen from N. Hence the probability, that the series of spins results in r heads and $N-r$ tails, is

$$\frac{N!}{r!\,(N-r)!}\frac{1}{2^N}.$$

So long as N and r are small numbers, there is no difficulty in evaluating this formula numerically; but it is obvious that direct numerical calculation is out of the question, when numbers running into thousands have to be dealt with. For instance, the labour of determining directly the probability, that in 10,000 spins the number of heads lies between 4900 and 5100, would be prohibitive.

To deal with any such calculations some method of approximation is absolutely necessary; that is to say, some approximate formula for $n!$, when n is a large number. What is known as Stirling's theorem serves this purpose. It states that

$$n! = \sqrt{2\pi n}\, n^n e^{-n}\left(1 + \frac{1+\theta_n}{12n}\right),$$

where θ_n approaches the limit zero as n increases.

The θ_n of the formula is already small, when n is comparatively small. If $n > 10$, $0 < \theta_n < \cdot 01$; so that, for values of n exceeding 1000, the proportional error involved in omitting the factor $1 + \frac{1+\theta_n}{12n}$ is less than ·00009.

This formula will first be used to obtain a convenient approximation to the probability of a given excess of heads in a sequence of spins. To make the calculation as symmetrical as possible, suppose that the number of spins is $2N$ and the number of heads $N + r$, so that $2r$ is the excess of heads over tails. The probability of this is

$$\frac{2N!}{(N+r)!\,(N-r)!}\frac{1}{2^{2N}}.$$

If, in the above formula α_n be written for $\frac{1+\theta_n}{12n}$, then

$$\frac{2N!}{(N+r)!\,(N-r)!}\frac{1}{2^{2N}}$$

$$= \sqrt{\frac{N}{\pi(N^2-r^2)}}\frac{N^{2N}}{(N+r)^{N+r}(N-r)^{N-r}}\frac{1+\alpha_{2N}}{(1+\alpha_{N+r})(1+\alpha_{N-r})}$$

$$= \sqrt{\frac{1}{\pi N}}\frac{1}{\left(1+\frac{r}{N}\right)^{N+r}\left(1-\frac{r}{N}\right)^{N-r}}\left(1-\frac{r^2}{N^2}\right)^{-\frac{1}{2}}\frac{1+\alpha_{2N}}{(1+\alpha_{N+r})(1+\alpha_{N-r})}.$$

Let $$D = \left(1+\frac{r}{N}\right)^{N+r}\left(1-\frac{r}{N}\right)^{N-r};$$

then, since r/N is a proper fraction,

$$\log D = (N+r)\left[\frac{r}{N} - \frac{1}{2}\frac{r^2}{N^2} + \frac{1}{3}\frac{r^3}{N^3} - \dots\right]$$
$$- (N-r)\left[\frac{r}{N} + \frac{1}{2}\frac{r^2}{N^2} + \frac{1}{3}\frac{r^3}{N^3} + \dots\right]$$
$$= \frac{r^2}{N} + \frac{1}{6}\frac{r^4}{N^3} + \frac{1}{15}\frac{r^6}{N^5} + \dots,$$

and $$\frac{1}{D} = e^{-\left(\frac{r^2}{N} + \frac{r^4}{6N^3} + \frac{r^6}{15N^5} + \dots\right)}.$$

Hence

$$\frac{2N!}{(N+r)!\,(N-r)!}\,\frac{1}{2^{2N}}=\frac{1}{\sqrt{\pi N}}\,e^{-\frac{r^2}{N}}f(r,\,N),$$

where

$$f(r,\,N)=\left(1-\frac{r^2}{N^2}\right)^{-\frac{1}{2}}e^{-\left(\frac{r^4}{6N^3}+\frac{r^6}{15N^5}+\dots\right)}\,\frac{1+\alpha_{2N}}{(1+\alpha_{N+r})\,(1+\alpha_{N-r})}.$$

When N is large, it has been seen that the last factor is very nearly unity. In the other factors, write x for r^2/N. They become

$$\left(1-\frac{x}{N}\right)^{-\frac{1}{2}}e^{-\left(\frac{x^2}{6N}+\frac{x^3}{15N^2}+\dots\right)}.$$

When x/N is small, this is very nearly unity; but it diminishes as x/N increases. Hence

$$\frac{1}{\sqrt{\pi N}}\,e^{-\frac{r^2}{N}}$$

is a good approximation to the required probability, so long as r^2/N is not too large; but it gives too great a value as r^2/N increases. It is of course to be noticed that, unless r^2/N is small enough, the numerical value of the probability is inappreciable.

B. A precisely similar result can be obtained when the number of spins is odd; but in dealing with large values of N, there would be no real loss of generality in taking the number of spins always even.

That this approximation lends itself readily to calculation will be clear, by considering the question suggested above. In $2N$ spins the chance, that the number of heads lies between $N+p$ and $N-p$, is approximately

$$\sum_{r=-p}^{r=p}\frac{1}{\sqrt{\pi N}}\,e^{-\frac{r^2}{N}}.$$

Putting $\qquad r=x\sqrt{N},\quad r+1=(x+\delta x)\sqrt{N},$

this is sensibly the same as

$$\frac{1}{\sqrt{\pi}}\int_{-\frac{p}{\sqrt{N}}}^{\frac{p}{\sqrt{N}}}e^{-x^2}\,dx.$$

If, as in the question referred to above, $2N = 10{,}000$, $p = 100$, $p/\sqrt{N}$ is $\sqrt{2}$; and tables give the value of the integral to be ·95.

As a further illustration, since (from tables)

$$\frac{1}{\sqrt{\pi}}\int_{-1\cdot8}^{1\cdot8} e^{-x^2}\,dx = \cdot 99,$$

the probability, that the difference between the number of heads and tails in N spins will exceed $2{\cdot}54\sqrt{N}$, is less than 1/100.

12. C. If the probability of an event happening at a single trial is p, the probability q that it will happen r times in N trials is

$$q = \frac{N!}{r!\,(N-r)!}\,p^r(1-p)^{N-r}.$$

(i). Put $$r = pN + x,$$
so that $$N - r = (1-p)\,N - x;$$
and suppose N so great, that the factor $1+\alpha_N$ in the approximate expression for N may be safely replaced by unity. Then

$$\begin{aligned}\log q = &-\tfrac{1}{2}\log 2\pi + (N+\tfrac{1}{2})\log N - (pN + x + \tfrac{1}{2})\log(pN+x)\\ &-\{(1-p)\,N - x + \tfrac{1}{2}\}\log\{(1-p)\,N - x\}\\ &+(pN+x)\log p + \{(1-p)\,N - x\}\log(1-p)\end{aligned}$$

$$\begin{aligned}= &-\tfrac{1}{2}\log 2\pi p(1-p)\,N\\ &-(pN+x+\tfrac{1}{2})\left(\frac{x}{pN} - \frac{x^2}{2p^2N^2} + \frac{x^3}{3p^3N^3} - \dots\right)\\ &+\{(1-p)\,N - x + \tfrac{1}{2}\}\left[\frac{x}{(1-p)\,N} + \frac{x^2}{2(1-p)^2N^2}\right.\\ &\left.\qquad + \frac{x^3}{3(1-p)^3N^3} + \dots\right]\end{aligned}$$

$$\begin{aligned}= &-\tfrac{1}{2}\log 2\pi p(1-p)\,N\\ &-\frac{1}{2N}\left(\frac{x^2+x}{p} + \frac{x^2-x}{1-p}\right) + \frac{1}{6N^2}\left\{\frac{x^3+\frac{3}{2}x^2}{p^2} - \frac{x^3-\frac{3}{2}x^2}{(1-p)^2}\right\}\\ &+\text{terms in }\frac{1}{N^3};\end{aligned}$$

and therefore

$$q = \frac{1}{\sqrt{2\pi p(1-p)\,N}}\times e^{-\frac{1}{2N}\frac{x^2+(1-2p)\,x}{p(1-p)} + \frac{1}{6N^2}\frac{(1-2p)\,x^3 + \frac{3}{2}(1-2p+2p^2)\,x^2}{p^2(1-p)^2} + \dots}.$$

When N is large enough, the factor

$$e^{\frac{1}{6N^2}\frac{(1-2p)x^3+\frac{3}{2}(1-2p+2p^2)x^2}{p^2(1-p)^2}}$$

cannot differ sensibly from unity, until x is of the order $N^{\frac{2}{3}}$. But when x is of this order, the factor

$$e^{-\frac{1}{2N}\frac{x^2+(1-2p)x}{p(1-p)}}$$

is excessively small. Again, the factor

$$e^{-\frac{1}{2N}\frac{(1-2p)x}{p(1-p)}}$$

will not differ sensibly from unity, until x is of the order N, and then, again,

$$e^{-\frac{1}{2N}\frac{x^2}{p(1-p)}}$$

is excessively small.

Hence
$$q = \frac{1}{\sqrt{2\pi p(1-p)N}}e^{-\frac{x^2}{2Np(1-p)}}$$

is a good approximation to the value of q, so long as this value is appreciable. If, however, it were necessary to determine the numerical value of q for values of x, for which q is excessively small, this approximation might not hold.

(ii). In the case, in which either p or $1-p$ is very small, another approximation to q may be obtained, which is for some calculations more convenient than the preceding. Suppose that p is very small; and put $pN = \nu$. Then

$$q = \frac{\nu^r}{r!}\cdot\frac{N!\left(1-\frac{\nu}{N}\right)^{N-r}}{(N-r)!\,N^r};$$

and, replacing the factorials by their approximate values,

$$q = \frac{\nu^r}{r!}\left(\frac{N}{N-r}\right)^{\frac{1}{2}}\cdot\frac{N^N e^{-N}\left(1-\frac{\nu}{N}\right)^{N-r}}{(N-r)^{N-r}e^{-N+r}N^r}.$$

Now ν is very small compared to N. If r differs very much from ν, the value of q is excessively small. On the other hand, if r does not differ too much from ν,

$$\left(1-\frac{\nu}{N}\right)^N = e^{-\nu}, \quad \left(1-\frac{r}{N}\right)^N = e^{-r} \text{ very nearly,}$$

and
$$q = \frac{e^{-\nu}\nu^r}{r!} \cdot \frac{1}{\left(1-\frac{r}{N}\right)^{\frac{1}{2}}} \frac{\left(1-\frac{\nu}{N}\right)^{-r}}{\left(1-\frac{r}{N}\right)^{-r}},$$

so that
$$q = \frac{e^{-\nu}\nu^r}{r!} \text{ very nearly.}$$

Probable Value: Most Probable Value.

13. It is convenient here to introduce two conceptions which prove to be of great value in many applications of the theory of probabilities; viz. those of the "probable value," and the "most probable value."

If a number can take any one of the distinct values

$$a_i, \quad (i = 1, 2, \ldots, n),$$

and if the probability that the number takes the value a_i is p_i, $(i = 1, 2, \ldots, n)$, so that $\Sigma p_i = 1$, then

$$\sum_i p_i a_i$$

is called the probable value of the number; and if p_m is the greatest of the p's, a_m is called the most probable value.

It is to be noticed that the probable value is not necessarily one of the values that the number actually takes; it is the mean of the values when the weight, given to each in taking the mean, is proportional to its probability.

14 D. Since $\dfrac{N!}{(N-r)!\,r!}$ is as great as possible when $r = \frac{1}{2}N$ or $\frac{1}{2}(N+1)$, according as N is even or odd, the most probable value, of the difference in number of heads and tails in a sequence of N spins, is 0 or 1 according as N is even or odd.

The probable value of the excess of heads over tails (or tails over heads) in N spins is

$$\sum_{r=0}^{N} \frac{N!}{(N-r)!\,r!} \frac{2r-N}{2^N};$$

and this is obviously zero, as would be expected.

In $2N$ spins, the probable value of the square of the excess of heads over tails is

$$\sum_{r=-N}^{N} \frac{2N!}{(N+r)!\,(N-r)!} \frac{4r^2}{2^{2N}}.$$

This may be evaluated as follows. We have

$$\sum_{r=-N}^{N} \frac{2N!}{(N+r)!\,(N-r)!} x^r = \frac{(1+x)^{2N}}{x^N},$$

$$\sum_{r=-N}^{N} \frac{2N!}{(N+r)!\,(N-r)!} r^2 x^r = \left(x\frac{d}{dx}\right)^2 . \frac{(1+x)^{2N}}{x^N};$$

so that, putting $x = 1$,

$$\sum_{r=-N}^{N} \frac{2N!}{(N+r)!\,(N-r)!} \frac{4r^2}{2^{2N}} = 2N.$$

It may be shewn, in a similar way, that the probable value of the fourth power of the excess of heads over tails is $12N^2 - 4N$; while the probable value of any odd power of the excess is zero.

It is interesting to note that, if the approximate value

$$\frac{1}{\sqrt{\pi N}} e^{-\frac{r^2}{N}}$$

of the probability is used instead of the true value, the probable value of the square of the excess is

$$\sum_{r=-N}^{N} \frac{1}{\sqrt{\pi N}} e^{-\frac{r^2}{N}} 4r^2,$$

which is sensibly equal to

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} 4Nx^2 e^{-x^2} dx, \text{ that is, } 2N.$$

A similar calculation of the probable value of the fourth power of the excess gives

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} 16N^2 x^4 e^{-x^2} dx, \text{ that is, } 12N^2.$$

This is about $\left(1 + \frac{1}{3N}\right)$ of the true value. Since the approximate expression for the probability has been seen to be too great for considerable values of r, an over-estimate of the probable value was to be expected, when making use of the approximation.

E. Any sequence of N spins will fall into a series of sequences of heads and tails. The first spin necessarily starts a sequence of heads or tails. Suppose that, from the remaining $N-1$ spins $M-1$ are chosen in any way; and that these $M-1$, and these only, start the sequences of heads and tails other than the first. This implies that the sequence of N spins falls into M sequences of heads and tails. Corresponding to each way in which the $M-1$ are chosen, there will be two distinct sets of sequences of heads and tails; for the first sequence may be either heads or tails. Now the number of ways, in which $M-1$ things may be chosen from $N-1$, is

$$\frac{(N-1)!}{(M-1)!\,(N-M)!}.$$

There are therefore just twice this number of ways, in which the N spins may fall into M sequences of heads and tails. It follows that the probability, that a sequence of N spins falls into M sequences of heads and tails, is

$$\frac{(N-1)!}{(M-1)!\,(N-M)!}\,\frac{1}{2^{N-1}}.$$

The most probable number of sequences of heads and tails is that for which this number is as great as possible; i.e. for which $M-1$ and $N-M$ are equal or differ by unity. Hence, if N is odd, the most probable number of sequences of heads and tails is $\frac{1}{2}(N+1)$; while, if N is even, it is either $\frac{1}{2}N$ or $\frac{1}{2}N+1$.

The probable number of sequences of heads and tails is

$$\sum_{M=1}^{M=N} \frac{(N-1)!}{(M-1)!\,(N-M)!}\,\frac{M}{2^{N-1}};$$

and, by the method already used, this is found to be $\frac{1}{2}(N+1)$. For M sequences, the average number of heads or tails in a sequence is N/M; hence, in a long series of N spins, the probable value and also the most probable value of the average number of heads or tails in a sequence is 2.

Suppose now that, in a set of M sequences, there are m_i sequences of $i\,(i=1, 2, 3, \ldots)$. These numbers are obviously connected by the two equations

$$\Sigma m_i = M, \quad \Sigma i m_i = N.$$

Twice the number of solutions of these equations in positive integers gives the number of ways, in which a set of N spins may fall into M sequences of heads and tails. The number of solutions of these equations is the same as the coefficient of x^N in

$$(x + x^2 + x^3 + \dots)^M;$$

and there is no difficulty in verifying in this way the result already obtained.

Suppose, next, that each sequence of heads or tails is limited to contain not more than r members. Then the number of solutions of the two equations between the m's, with this limitation, is the coefficient of x^N in

$$(x + x^2 + \dots + x^r)^M.$$

Hence twice the coefficient of x^N in

$$\sum_M (x + x^2 + \dots + x^r)^M,$$

that is, in

$$\frac{1 - x}{1 - 2x + x^{r+1}},$$

is the number of ways in which a set of N spins may fall in sequences of heads and tails not exceeding r in a sequence. This gives a new, but in general much less effective, solution of the question of p. 47.

F. The number of ways, in which a set of M sequences of heads and tails, in which there are m_i sequences of i $(i = 1, 2, 3, \dots)$, can occur, is the same as the number of permutations of M things which are alike in sets of m_i, $(i = 1, 2, 3, \dots)$; and this is

$$\frac{(m_1 + m_2 + \dots)!}{m_1!\, m_2! \dots}.$$

Hence the most probable set of M sequences is that for which

$$\frac{(m_1 + m_2 + \dots)!}{m_1!\, m_2! \dots} \frac{1}{2^{N-1}}$$

is as great as possible. This involves determining the least value of $m_1!\, m_2! \dots$, subject to the conditions

$$\Sigma m_i = M, \quad \Sigma i m_i = N.$$

When N is large enough, a roughly approximate solution of this problem may be found as follows. Thus

$$\log(m_1!\, m_2! \dots) = \Sigma \{\tfrac{1}{2}\log 2\pi + (m_i + \tfrac{1}{2}) \log m_i - m_i\},$$

which differs by a constant from

$$\Sigma (m_i + \tfrac{1}{2}) \log m_i.$$

If the m's are treated as continuously varying quantities, the minimum value of the last magnitude, subject to the above two conditions, is given by

$$\log m_i + 1 + \frac{1}{2m_i} + A + Bi = 0, \quad (i = 1, 2, 3, \ldots),$$

where A and B are constants. Omitting the terms $\frac{1}{2m_i}$ in comparison with $\log m_i$, these equations may be written

$$m_i = e^{-1-A-Bi}, \qquad (i = 1, 2, \ldots).$$

Hence $$M = e^{-1-A}\Sigma e^{-Bi} = e^{-1-A}\frac{e^{-B}}{1-e^{-B}},$$

$$N = e^{-1-A}\Sigma ie^{-Bi} = e^{-1-A}\frac{e^{-B}}{(1-e^{-B})^2};$$

so that $$e^{-B} = 1 - \frac{M}{N}, \quad e^{-1-A} = \frac{M^2}{N-M},$$

$$m_i = \frac{M^2}{N-M}\cdot\left(\frac{N-M}{N}\right)^i = \frac{M^2}{N}\cdot\left(\frac{N-M}{N}\right)^{i-1}.$$

For the most probable number of sequences $M = \frac{1}{2}N$: and the most probable set in this case is that for which $m_i = 2^{-i-1}N$.

It is interesting to compare the results, that have been obtained on the supposition that the probability of a coin falling head is definitely known, with those deduced from the complete data regarding a selection of the coin. Suppose there are n coins, and that the chance of the ith coin falling head is p_i $(i = 1, 2, \ldots, n)$. Suppose also that, when one of the coins is chosen, the chance of choosing the jth is q_j $(j = 1, 2, \ldots, n)$.

If a sequence of N spins is made with a chosen coin, the probability of the various results will clearly depend on the number of spins made with a chosen coin before a fresh one is chosen. For instance, if a fresh coin is chosen for each spin, the probability of a head at each spin is $\Sigma q_i p_i$; and the problem is the same as those already considered.

Suppose that the chosen coin is used throughout the N spins: what is the probability of r heads and $N-r$ tails? If the ith coin is chosen, the probability is

$$\frac{N!}{r!\,(N-r)!}p_i^r(1-p_i)^{N-r}.$$

Hence the required probability is

$$\frac{N!}{r!(N-r)!}\sum_{i=1}^{i=n} q_i p_i^r (1-p_i)^{N-r}.$$

Consider the case in which

$$p_i = \frac{i}{n+1}, \quad q_j = \frac{1}{n}, \quad (i, j = 1, 2, \ldots, n),$$

so that the chance of the coin falling head is equally likely to have any one of the values

$$\frac{1}{n+1}, \frac{2}{n+1}, \ldots, \frac{n}{n+1}.$$

The probability of r heads and $N-r$ tails is then

$$\frac{N!}{r!(N-r)!}\sum_{i=1}^{n} \frac{i^r (n+1-i)^{N-r}}{n(n+1)^N}.$$

If n is not too small, the sum in this expression differs very little from

$$\int_0^1 x^r (1-x)^{N-r} dx,$$

the value of which is $$\frac{r!(N-r)!}{(N+1)!}.$$

Hence, if n is large enough, the required probability is very nearly independent of r and equal to $\frac{1}{N+1}$; a marked contrast with the results already obtained.

15. G. So far in dealing with a repeated trial, it is only the probabilities connected with the satisfying or not satisfying of a single condition, that have been considered.

Suppose now that A and B are two different conditions relevant to the results of the trial. When the trial is repeated N times, suppose that, on N_1 specified occasions, A and B are both satisfied; on N_2 occasions, A is satisfied and B is not; on N_3 occasions, A is not satisfied and B is; and on

$$N_4 (= N - N_1 - N_2 - N_3)$$

occasions, neither condition is satisfied. The probability for this combination is

$$p_{AB}^{N_1} p_{AB'}^{N_2} p_{A'B}^{N_3} p_{A'B'}^{N_4}.$$

Now the N_1, N_2, N_3 and N_4 specified occasions can be chosen from the N in

$$\frac{N!}{N_1! \, N_2! \, N_3! \, N_4!}$$

ways. Hence the probability that in the N trials, A and B are both satisfied N_1 times, and so on, is

$$\frac{N!}{N_1!\,N_2!\,N_3!\,N_4!}\,p_{AB}^{N_1}\,p_{AB'}^{N_2}\,p_{A'B}^{N_3}\,p_{A'B'}^{N_4}.$$

If condition A is to be satisfied in just r of the trials, and condition B in just s of the trials, then

$$N_1 + N_2 = r, \quad N_1 + N_3 = s.$$

Hence the probability q, that A is satisfied in just r and B in just s of the N trials, is

$$q = \sum_{N_1} \frac{N!}{N_1!\,(r-N_1)!\,(s-N_1)!\,(N-r-s+N_1)!} \times p_{AB}^{N_1}\,p_{AB'}^{r-N_1}\,p_{A'B}^{s-N_1}\,p_{A'B'}^{N-r-s+N_1},$$

where the sum is taken for those values of N_1 which make no one of the numbers N_1, $r-N_1$, $s-N_1$, $N-r-s+N_1$ negative. If $r \geqslant s$, the greatest value of N_1 is s; and the least value will be 0 or $r+s-N$, according as $r+s$ is less or greater than N.

Putting $\frac{p_{AB}\,p_{A'B'}}{p_{AB'}\,p_{A'B}} = \lambda$, the formula becomes

$$q = p_{AB'}^{r}\,p_{A'B}^{s}\,p_{A'B'}^{N-r-s} \sum_{N_1} f(N_1),$$

where $$f(N_1) = \frac{N!\,\lambda^{N_1}}{N_1!\,(r-N_1)!\,(s-N_1)!\,(N-r-s+N_1)!}.$$

When N is large, an approximate expression for this probability may be obtained, similar to that of p. 44 for a single condition. Thus, if

$$N_1 = p_{AB}N + x_1 = p_1 N + x_1, \quad N_2 = p_{AB'}N + x_2 = p_2 N + x_2,$$
$$N_3 = p_{A'B}N + x_3 = p_3 N + x_3, \quad N_4 = p_{A'B'}N + x_4 = p_4 N + x_4,$$

where $x_1 + x_2 + x_3 + x_4 = 0$, then

$$\log q = \log N! - \log(p_1N + x_1)! - \log(p_2N + x_2)! - \log(p_3N + x_3)! - \log(p_4N + x_4)!$$
$$+ (p_1N + x_1)\log p_1 + (p_2N + x_2)\log p_2 + (p_3N + x_3)\log p_3 + (p_4N + x_4)\log p_4.$$

Writing $\log n! = \frac{1}{2}\log 2\pi + (n + \frac{1}{2})\log n - n$, we have

$$\log q = -\tfrac{3}{2}\log 2\pi - \tfrac{3}{2}\log N - \tfrac{1}{2}\log p_1p_2p_3p_4 - \sum_{i=1}^{i=4}(p_iN + x_i + \tfrac{1}{2})\left(\frac{x_i}{p_iN} - \frac{x_i^2}{2p_i^2N^2} + \dots\right),$$

so that, when N is large enough,

$$q = \frac{1}{\sqrt{p_1 p_2 p_3 p_4}} \frac{1}{(2\pi N)^{\frac{3}{2}}} e^{-\frac{1}{2N}\left(\frac{x_1^2+x_1}{p_1}+\frac{x_2^2+x_2}{p_2}+\frac{x_3^2+x_3}{p_3}+\frac{x_4^2+x_4}{p_4}\right)}$$

very nearly.

The probability that, in the N trials, the condition A is satisfied $(p_1+p_2)N+y$ times and the condition B is satisfied $(p_1+p_3)N+z$ times, is therefore Σq for the values satisfying

$$x_1+x_2+x_3+x_4=0, \quad x_1+x_2=y, \quad x_1+x_3=z.$$

With these values, it will be found that

$$\frac{x_1^2+x_1}{p_1}+\frac{x_2^2+x_2}{p_2}+\frac{x_3^2+x_3}{p_3}+\frac{x_4^2+x_4}{p_4}$$

$$=\left(\frac{1}{p_1}+\frac{1}{p_2}+\frac{1}{p_3}+\frac{1}{p_4}\right) \times \left\{x_4 - \frac{\left(\frac{1}{p_1}+\frac{1}{p_3}\right)y+\left(\frac{1}{p_1}+\frac{1}{p_2}\right)z+\frac{1}{2}\left(\frac{1}{p_1}-\frac{1}{p_2}-\frac{1}{p_3}+\frac{1}{p_4}\right)}{\frac{1}{p_1}+\frac{1}{p_2}+\frac{1}{p_3}+\frac{1}{p_4}}\right\}^2$$

$$+\frac{\left\{\left(\frac{1}{p_1}+\frac{1}{p_3}\right)\left(\frac{1}{p_2}+\frac{1}{p_4}\right)(y+y_0)^2+\left(\frac{1}{p_1}+\frac{1}{p_2}\right)\left(\frac{1}{p_3}+\frac{1}{p_4}\right)(z+z_0)^2 + 2\left(\frac{1}{p_1p_4}-\frac{1}{p_2p_3}\right)(y+y_0)(z+z_0)+C\right\}}{\frac{1}{p_1}+\frac{1}{p_2}+\frac{1}{p_3}+\frac{1}{p_4}},$$

where

$$2y_0 = p_1+p_2-p_3-p_4,$$

$$2z_0 = p_1+p_3-p_2-p_4,$$

$$\begin{aligned}-4C &= \frac{1}{p_1^2}+\frac{1}{p_2^2}+\frac{1}{p_3^2}+\frac{1}{p_4^2} \\ &\quad -2\left(\frac{1}{p_1p_2}+\frac{1}{p_3p_4}\right)\{1-\tfrac{1}{2}(p_1+p_2-p_3-p_4)^2\} \\ &\quad -2\left(\frac{1}{p_1p_3}-\frac{1}{p_2p_4}\right)\{1-\tfrac{1}{2}(p_1+p_3-p_2-p_4)^2\} \\ &\quad +2\frac{1}{p_1p_4}\{1+2(p_1-p_4)^2\} \\ &\quad +2\frac{1}{p_2p_3}\{1+2(p_2-p_3)^2\}.\end{aligned}$$

With the abbreviations

$$\alpha=\left(\frac{1}{p_1}+\frac{1}{p_3}\right)\left(\frac{1}{p_2}+\frac{1}{p_4}\right),\qquad \gamma=\left(\frac{1}{p_1}+\frac{1}{p_2}\right)\left(\frac{1}{p_3}+\frac{1}{p_4}\right),$$

$$\beta=\frac{1}{p_1p_4}-\frac{1}{p_2p_3},\qquad \delta=\frac{1}{p_1}+\frac{1}{p_2}+\frac{1}{p_3}+\frac{1}{p_4},$$

the required probability Σq becomes

$$\Sigma\frac{1}{\sqrt{p_1p_2p_3p_4}}\frac{1}{(2\pi N)^{\frac{3}{2}}}\times e^{-\frac{1}{2N}\left[\delta(x_4-x_0)^2+\frac{\alpha(y+y_0)^2+2\beta(y+y_0)(z+z_0)+\gamma(z+z_0)^2+C}{\delta}\right]}.$$

The sum is taken with respect to values of x_4 increasing by unity at a step. Now the N_1, N_2, N_3, N_4 of the original notation are p_1N+x_4+y+z, p_2N-x_4-z, p_3N-x_4-y, and p_4N+x_4; and no one of these must be negative. Hence x_4 ranges from the larger of the integers $-p_4N$ and $-p_1N-y-z$ to the smaller of the integers p_2N-z and p_3N-y. In other words, the lower and the upper limits of x_4 are of the orders $-N$ and N. The same is true of the lower and the upper limits of x_4-x_0, until either y or z is of the order N, in which case q is excessively small.

Now, if N is not too small, $\Sigma e^{-\frac{\delta}{2N}(x_4-x_0)^2}$ under these conditions, is sensibly

$$\sqrt{\frac{2\pi N}{\delta}}.$$

Further, $e^{-\frac{C}{2\delta N}}$ differs very little from unity under the conditions assumed. Hence, y_0 and z_0 being proper fractions, the required approximation may be written

$$q=\frac{1}{\sqrt{p_1p_2p_3p_4\delta}}\frac{1}{2\pi N}e^{-\frac{\alpha y^2+2\beta yz+\gamma z^2}{2\delta N}}\text{ very nearly.}$$

This quantity varies slowly with y and z. Now y is the excess of the number of times, that condition A is satisfied, over $(p_1+p_2)N$, which is the probable number of times it is satisfied; and a similar statement may be made for z. The probability that the excess for condition A lies between y and $y+\delta y$, and for condition B between z and $z+\delta z$, is $q\,\delta y\,\delta z$.

From this expression the probability, that these excesses should lie between given limits, can be determined approximately by integration.

For instance, the probability, that the excess for condition A lies between y and $y+\delta y$, is

$$\delta y \cdot \frac{1}{\sqrt{p_1 p_2 p_3 p_4 \delta}} \frac{1}{2\pi N} \int_{-\infty}^{\infty} e^{-\frac{\alpha y^2 + 2\beta yz + \gamma z^2}{2\delta N}} dz.$$

Now $$\alpha y^2 + 2\beta yz + \gamma z^2 = \frac{\alpha\gamma - \beta^2}{\gamma} y^2 + \gamma\left(z + \frac{\beta}{\gamma} y\right)^2,$$

and $$\int_{-\infty}^{\infty} e^{-\frac{\gamma}{2\delta N}\left(z+\frac{\beta}{\gamma}y\right)^2} dz = \sqrt{\frac{2\pi\delta N}{\gamma}};$$

hence the required probability is

$$\frac{1}{\sqrt{p_1 p_2 p_3 p_4 \gamma}} \frac{1}{\sqrt{2\pi N}} e^{-\frac{\alpha\gamma - \beta^2}{2\gamma\delta N} y^2} \delta y;$$

or, inserting the values of α, β, γ, it is

$$\frac{1}{\sqrt{2\pi (p_1+p_2)(p_3+p_4) N}} e^{-\frac{y^2}{2(p_1+p_2)(p_3+p_4) N}} \delta y.$$

CHAPTER V

PROBABILITY OF CAUSES

16. When an event has happened which may have been due to any one of a number of different causes, the question arises as to which cause has most probably been in action. Is it possible, from the observed happening of the event, to draw any conclusion as to the relative probability of the various causes that may have led to it?

From the discussion in Chapter I, it has been seen that p_{A_iB}/p_B is the probability, that condition A_i is satisfied when condition B is known to be satisfied.

Suppose that $A_1, A_2, \ldots, A_n$ are n conditions of which one must be satisfied, and only one can be satisfied, when a trial is made. Then

$$p_B = \sum_i p_{A_iB},$$

so that

$$\frac{p_{A_iB}}{p_B} = \frac{p_{A_i}\dfrac{p_{A_iB}}{p_{A_i}}}{\sum_i p_{A_i}\dfrac{p_{A_iB}}{p_{A_i}}} = \frac{p_{A_i}p_{(A_i)B}}{\sum_i p_{A_i}p_{(A_i)B}}.$$

Suppose now that the event E may have any one of n distinct causes, of which in a given trial one and only one can come into play. Let condition B be that the event E shall happen, and condition A_i be that the ith cause has come into play. Then p_{A_i} is the probability before the trial, that the ith cause of E will come into play: $p_{(A_i)B}$ is the probability that E will happen as a result of the ith case; and p_{A_iB}/p_B is the probability, when E has happened, that it has happened as a result of the ith cause. The formula may be conveniently written

$$q_i = \frac{r_i s_i}{\sum_i r_i s_i}:$$

where r_i is the probability of the ith cause, before the result is known (the so-called *à priori* probability of the ith cause); s_i is the probability of the event when the ith cause is in action; and q_i is the probability of the ith cause, when the event is known to have happened (the so-called *à posteriori* probability).

This formula is known as Bayes' formula*; and so long as the r's and the s's are known, there can be no ambiguity in applying it. As a simple illustration, the following case may be considered.

I. There are n boxes, each containing white and black objects. The chance of drawing a white object from the ith box is p_i. In choosing a box from which to draw an object, each box is equally likely to be chosen. An object, observed to be white, is drawn from a box; it is then returned. A second object is drawn from the same box. What is the probability that it is white?

In this case, $r_i = \frac{1}{n}$; so that the probability, that the white object first drawn came from the ith box, is $p_i/\sum_i p_i$. This is the probability that the ith box is used at the second drawing; and therefore the probability, that the second drawing gives a white object, is

$$\frac{\sum_i p_i^2}{\sum_i p_i}.$$

It should be noted that this is greater than $\frac{1}{n}\sum_i p_i$, which is the probability that a first drawing gives a white object.

If from the above statement of conditions the sentence, "In choosing a box from which to draw an object, each box is equally likely to be chosen" is omitted, there are no data from which to calculate r_i; and the question proposed cannot be answered.

Use of the Bayes formula.

17. The hesitation that is undoubtedly felt in making use of Bayes' formula depends upon the fact that, though the s's are generally known, some assumption has to be made with respect to the r's; and the calculated probabilities of cause depend on the particular assumption made. This will be brought out as clearly as possible in some of the following illustrations.

II. A box contains n objects, each of which is either white or black, and each of which is equally likely to be drawn. An object

* It is due to the Rev. Thomas Bayes (elected F.R.S. 1742): for an abstract of his two memoirs, see Todhunter, *History of the Theory of Probability*, ch. XIV.

is drawn and is found to be white. It is returned; and an object is drawn again. What is the probability that it will be white?

Denote by p_r the *à priori* probability that r of the objects are white. Then

$$\frac{rp_r}{\Sigma rp_r}$$

is the *à posteriori* probability that r are white; and the probability of drawing a white object at the second trial is

$$\frac{\Sigma p_r r^2}{n\Sigma p_r r}.$$

If it is assumed that p_r is independent of r and therefore equal to $\frac{1}{n}$, this last probability is

$$\frac{2}{3}+\frac{1}{3n}.$$

If however each object in the box is assumed to be equally likely black or white, then $p_r = \frac{n!}{r!(n-r)!}\cdot\frac{1}{2^n}$; and the required probability is

$$\frac{1}{2}+\frac{1}{2n}.$$

With regard to such a question, it may be suggested that the data are very meagre; therefore it is not surprising that different assumptions about the *à priori* probability lead to very different results.

III. A box contains a number N of objects not greater than M; and it is known that n of them are marked. It is assumed that, when a set of m objects is drawn from the box, all sets of m are equally likely. A set of m is drawn; and it is found that m_1 of them are marked. What is the most probable value of N?

It follows, from the data, that N is equal to or greater than $n+m-m_1$. The probability of the observed event, when the box contains N objects, is

$$\frac{\dfrac{n!}{m_1!(n-m_1)!}\,\dfrac{(N-n)!}{(m-m_1)!(N-n-m+m_1)!}}{\dfrac{N!}{m!(N-m)!}},$$

that is, $$\frac{m!\,n!}{m_1!(n-m_1)!(m-m_1)!}\cdot\frac{(N-n)!(N-m)!}{N!(N-n-m+m_1)!}.$$

Hence, if p_N is the *à priori* probability that the box contains N objects, then, after the event, the probability is

$$q_N = \frac{f(N)\,p_N}{\sum\limits_{n+m-m_1}^{M} f(N)\,p_N},$$

where
$$f(N) = \frac{(N-m)!\,(N-n)!}{N!\,(N-n-m+m_1)!}.$$

The most probable value of N is that which makes $f(N)\,p_N$ as great as possible.

Suppose first that all possible values of N are *à priori* equally probable, so that p_N is independent of N. Then the most probable value of N satisfies the inequalities

$$f(N) > f(N-1), \quad f(N) > f(N+1),$$

which gives, for N, the greatest integer in nm/m_1.

Suppose next that $p_N \propto N$, so that large values of N are *à priori* more likely than small ones. The most probable value of N is then given by

$$Nf(N) > (N-1)f(N-1), \quad Nf(N) > (N+1)f(N+1),$$

which gives, for N, the greatest integer in $(n-1)(m-1)/(m_1-1)$.

If lastly $p_N \propto \dfrac{1}{N+1}$, so that small values of N are *à priori* more likely than large ones, the inequalities are

$$\frac{f(N)}{N+1} > \frac{f(N-1)}{N}, \quad \frac{f(N)}{N+1} > \frac{f(N+1)}{N+2},$$

giving, for N, the greatest integer in

$$\{(m+1)(n+1)-m_1-2\}/(m_1+1).$$

It will be noticed that, so long as m, m_1, n, are not quite small numbers, the three different suppositions with respect to p_N lead to results in close agreement with each other.

IV. A and B play a game, at which A's *à priori* chance of winning is equally likely to be $\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-2}{n}$, or $\frac{n-1}{n}$. Out of a set of $a+b$ games, A wins a and loses b. What is the probable value of A's chance of winning the next game?

If A's chance of winning is $\frac{r}{n}$, the probability of the observed result of the $a+b$ games is

$$\frac{(a+b)!}{a!\,b!}\left(\frac{r}{n}\right)^a\left(1-\frac{r}{n}\right)^b.$$

Hence the *à posteriori* probability, that A's chance of winning the next game shall be $\frac{r}{n}$, is

$$\frac{\left(\frac{r}{n}\right)^a\left(1-\frac{r}{n}\right)^b}{\sum_1^{n-1}\left(\frac{r}{n}\right)^a\left(1-\frac{r}{n}\right)^b};$$

and the probable value of A's chance of winning the next game is

$$\frac{\sum_1^{n-1}\left(\frac{r}{n}\right)^{a+1}\left(1-\frac{r}{n}\right)^b}{\sum_1^{n-1}\left(\frac{r}{n}\right)^a\left(1-\frac{r}{n}\right)^b}.$$

Now, if n is not too small, the quantity $\frac{1}{n}\sum_1^{n-1}\left(\frac{r}{n}\right)^a\left(1-\frac{r}{n}\right)^b$ is very nearly equal to

$$\int_0^1 x^a(1-x)^b\,dx = \frac{a!\,b!}{(a+b+1)!}.$$

Hence, if n is large enough, the required result is very nearly equal to

$$\frac{a+1}{a+b+2}.$$

It has been assumed that the probability, of A's chance of winning being measured by $\frac{r}{n}$, is itself independent of r.

Suppose now that the probability of A's chance of winning being measured by $\frac{r}{n}$, is proportional to $\frac{r}{n}\left(1-\frac{r}{n}\right)$, so that neither A nor B is extremely likely either to win or lose. Then the above expression, for the probable value of A's chance of winning the next game, becomes

$$\frac{\sum_1^{n-1}\left(\frac{r}{n}\right)^{a+2}\left(1-\frac{r}{n}\right)^{b+1}}{\sum_1^{n-1}\left(\frac{r}{n}\right)^{a+1}\left(1-\frac{r}{n}\right)^{b+1}},$$

which is sensibly equal to

$$\frac{a+2}{a+b+4}.$$

It is to be noticed that, if a and b are comparatively large numbers, this result differs very little from the former one. In other words, if the result of a sufficiently large number of games is observed, the hypothesis made as to A's *à priori* chance of winning has but little effect on the result. This is obviously not the fact when the number of games observed is small.

18. V. It is assumed that, when a calculator adds a column of integers, the probability of his making an error of either $+\alpha$ or $-\alpha$ units is $\frac{1}{2^{2+\alpha}}$; while the probability of his getting the correct result is $\frac{1}{2}$. The results of twice adding a given column are s_1 and s_2 ($> s_1$). What is the probable value of the sum?

If s is the true sum, the probability p_s of getting s_1 and s_2 for the sum at two attempts, is

when
$$s < s_1, \quad \frac{1}{2^{4+s_1+s_2-2s}},$$
$$s = s_1, \quad \frac{1}{2^{3+s_2-s_1}},$$
$$s_2 > s > s_1, \quad \frac{1}{2^{4+s_2-s_1}},$$
$$s_2 = s, \quad \frac{1}{2^{3+s_2-s_1}},$$
$$s_2 < s, \quad \frac{1}{2^{4-s_1-s_2+2s}}.$$

Hence, if σ_s is the *à priori* probability that the sum is s, the probable value of the sum is

$$\frac{\Sigma s p_s \sigma_s}{\Sigma p_s \sigma_s},$$

where the above values are used for p_s.

Suppose that s may take any value from A to B, and that *à priori* all these values are equally probable. It will be assumed, to avoid dealing with particular cases, that A is less than s_1 and B greater than s_2. Then the probable value of the sum is

$$\frac{\sum_{s=A}^{B} s p_s}{\sum_{s=A}^{B} p_s}.$$

When the above values of p_s are used, this is found to be

$$\frac{\frac{1}{2}(s_1+s_2)(s_2-s_1+1\frac{1}{3})-\frac{1}{3}(A-\frac{4}{3})2^{2A-2s_1}-\frac{1}{3}(B+\frac{4}{3})2^{2s_2-2B}}{s_2-s_1+1\frac{1}{3}-2^{2A-2s_1}-2^{2s_2-2B}}.$$

If A is very much smaller than s_1 and B much larger than s_2, this probable value is very nearly

$$\tfrac{1}{2}(s_1+s_2).$$

But if $s_1 - A$ and $B - s_2$ are small, the probable value differs sensibly from the arithmetic mean. The supposition that leads to the arithmetic mean as the probable result, viz. that the sum to be found is equally likely to have every value in a long range, does not appear a very reasonable one; indeed, it is out of the question, if A is negative.

It is also to be noticed that, in this case, there is no most probable value. Assuming that the sum is equally likely *à priori* to take any value from A to B, the *à posteriori* probability that the sum is s_1 is the same as that the sum is s_2, and is greater than the probability that the sum has any other value.

It is interesting to compare this result with those, obtained by making other assumptions about the accuracy of the calculator. If the probability of his making an error α were $ke^{-h\alpha^2}$, then in the above calculation

$$p_s = k^2 e^{-h(s-s_1)^2-h(s-s_2)^2}.$$

If it is still assumed that the sum is equally likely to take all values from A to B, the probable value of the sum is

$$\frac{\sum\limits_{s=A}^{B} s e^{-2h\left(s-\frac{s_1+s_2}{2}\right)^2}}{\sum\limits_{s=A}^{B} e^{-2h\left(s-\frac{s_1+s_2}{2}\right)^2}}.$$

A small value of h implies considerable inaccuracy on the part of the calculator. If neither h, $B-\frac{1}{2}(s_1+s_2)$, nor $\frac{1}{2}(s_1+s_2)-A$, be quite small, it is easy to see that this fraction is nearly equal to $\frac{1}{2}(s_1+s_2)$, independently of the actual values of A and B. Moreover, in this case, the most probable value of the sum is clearly the arithmetic mean of s_1 and s_2.

With the second assumption about the calculator's errors, the results, concerning the probable and most probable values of the sum, are more definite and less dependent on the *à priori* probability of a given sum than with the first.

19. VI. An observer watches the spinning of a coin, and notes the sequences of heads and tails. What is the probable number of spins, that have occurred, when he has noted M sequences?

The number, N, of spins must be equal to or greater than M. On the supposition that the number of spins is N, the probability of the observed events is

$$\frac{(N-1)!}{(M-1)!\,(N-M)!}\cdot\frac{1}{2^{N-1}}.$$

Hence, if the *à priori* probability that the number of spins is N be represented by p_N, the probable number of spins is

$$\frac{\Sigma\,\dfrac{N!}{(N-M)!}\,\dfrac{1}{2^{N-1}}\,p_N}{\Sigma\,\dfrac{(N-1)!}{(N-M)!}\,\dfrac{1}{2^{N-1}}\,p_N}.$$

On the assumption that all numbers of spins equal to or exceeding N are *à priori* equally probable, this is

$$\frac{\sum\limits_{N=M}^{\infty}\dfrac{N!}{(N-M)!}\,\dfrac{1}{2^{N-1}}}{\sum\limits_{N=M}^{\infty}\dfrac{(N-1)!}{(N-M)!}\,\dfrac{1}{2^{N-1}}}$$

$$= M\,\frac{1+\dfrac{M+1}{1}\,\dfrac{1}{2}+\dfrac{(M+1)(M+2)}{1\,.\,2}\,\dfrac{1}{2^2}+\dots}{1+\dfrac{M}{1}\,\dfrac{1}{2}+\dfrac{M(M+1)}{1\,.\,2}\,\dfrac{1}{2^2}+\dots}$$

$$= M\,\frac{(1-\frac{1}{2})^{-M-1}}{(1-\frac{1}{2})^{-M}} = 2M.$$

Moreover, on the same assumption, the most probable value of N is $2M$.

Now, in this question, it is not a reasonable assumption that all values of N above M are equally probable. The spinning must take time; and for this reason there must be an upper limit to N. If it is assumed that all values of N from M to M' are equally probable, the probable value of N is

$$M\frac{A}{B},$$

where

$$A = 1 + \frac{M+1}{1}\frac{1}{2} + \frac{(M+1)(M+2)}{2!}\frac{1}{2^2} + \dots + \frac{(M+1)(M+2)\dots M'}{(M'-M)!}\frac{1}{2^{M'-M}},$$

$$B = 1 + \frac{M}{1}\frac{1}{2} + \frac{M(M+1)}{2!}\frac{1}{2^2} + \dots + \frac{M(M+1)\dots(M'-1)}{(M'-M)!}\frac{1}{2^{M'-M}};$$

so that

$$A - B = \frac{1}{2} + \frac{M+1}{1}\frac{1}{2^2} + \dots + \frac{(M+1)(M+2)\dots(M'-1)}{(M'-M-1)!}\frac{1}{2^{M'-M}}$$

$$= \frac{1}{2}\left\{A - \frac{(M+1)(M+2)\dots M'}{(M'-M)!}\frac{1}{2^{M'-M}}\right\}.$$

Hence the probable value of N is

$$2M\left\{1 - \frac{(M+1)(M+2)\dots M'}{(M'-M)!}\frac{1}{2^{M'-M}}\cdot\frac{1}{2B}\right\}.$$

This is always less than $2M$.

It has been seen above that, when N is large, the probable number of sequences in N spins is $\frac{1}{2}N$, the duration of the spins not affecting the question. When however a number of M sequences are observed, and the corresponding probable number of spins is to be determined, the question of duration *does* affect the question, and the probable number of spins is less than $2M$.

VII. There are M counters, marked from 1 to M, in a bag; and one is drawn, each being equally likely to be taken. The counter marked N is drawn, and a coin equally likely to fall head or tail is spun $2N$ times; and the excess $2n_1$ of heads over tails is noted. This is repeated s times, $2N$ spins being made each time; and the excesses of heads over tails are found to be

$$n_1, n_2, \dots, n_s.$$

The whole proceeding with the numbers $M, n_1, n_2, \dots, n_s$, is reported to a calculator, the number N only being withheld from him. What conclusions can he draw about N?

The *à priori* probability, that N has any given value from 1 to M, is $1/M$. If $|n_1|$ is the greatest of the positive numbers $|n_1|, |n_2|, \ldots, |n_s|$, the probability of the observed set of excesses of heads over tails is zero, when $N < |n_1|$, and is

$$\left\{\frac{(2N)!}{2^{2N}}\right\}^s \prod_{i=1}^{i=s} \frac{1}{(N+n_i)!\,(N-n_i)!},$$

when $N \geqslant |n_1|$.

The approximate value of this latter expression is

$$\frac{1}{(\pi N)^{s/2}} e^{-\frac{1}{N}\sum_1^s y_i^2}$$

If then $N > |n_1|$, the calculator infers that the probability, that the counter drawn was marked N, is

$$\frac{\frac{1}{N^{s/2}} e^{-\frac{\sigma}{N}}}{\sum_{n_1}^{M} \frac{1}{N^{s/2}} e^{-\frac{\sigma}{N}}},$$

where $$\sigma = \sum_1^s y_i^2.$$

The most probable value of N is that which makes $e^{-\frac{\sigma}{N}}/N^{\frac{s}{2}}$ as great as possible. The maximum value of this quantity, when N varies continuously, is given by

$$N = \frac{2\sigma}{s};$$

so that the most probable value of N is one of the integers on either side of $2\sigma/s$.

Since $e^{-\frac{\sigma}{N}}/N^{\frac{s}{2}}$, when sensible in value, changes little when N is changed to $N+1$, the probability that N lies between N_1 and N_2 may be written approximately

$$\frac{\int_{N_1}^{N_2} N^{-\frac{s}{2}} e^{-\frac{\sigma}{N}} dN}{\int_0^\infty N^{-\frac{s}{2}} e^{-\frac{\sigma}{N}} dN}.$$

Putting $$\sigma = Nx,$$

this is $$\frac{1}{\Gamma\left(\frac{s}{2}\right)} \int_{\frac{\sigma}{N_2}}^{\frac{\sigma}{N_1}} x^{\frac{s}{2}-1} e^{-x}\,dx.$$

CHAPTER VI

PROBABILITIES CONNECTED WITH GEOMETRICAL QUESTIONS

20. In Chapter II, a problem was discussed (p. 22) in connection with the position of points on a line. The line was divided into n equal parts; and it was assumed that, when a point was marked on it, the point was as likely to be in any one part as in any other. It followed that the chance of the point being on one particular part of the line is $1/n$. Suppose that AB is the line; let P and Q be two particular points on it, of which P lies between the pth and $(p+1)$th points of division, while Q lies between the qth and $(q+1)$th. The segment PQ of the line includes $q-p$ complete parts of the line and portions of two others. The probability, that a marked point lies on the $q-p$ complete parts, is $(q-p)/n$. The probability, that the marked point lies on PQ, is therefore equal to or greater than $(q-p)/n$. It was shewn, in the same way, to be equal to or less than $(q-p+2)/n$. Beyond this it is impossible to go without further data. If n is large, the probability that the marked point lies on PQ is known between narrow limits; and as n is made larger, both $(q-p)/n$ and $(q-p+2)/n$ approach the same value, viz. PQ/AB.

Hence the supposition that, when a line is divided into n equal parts, a marked point is as likely to lie in any one part as in any other, *whatever number n may be*; and the supposition, that the probability of a marked point lying on any particular segment of the line is equal to the length of the segment divided by the length of the line; are equivalent to each other.

Either supposition is often expressed in the form, that all positions of the point are equally probable. It should be noticed that this does not involve all coordinates of the point being equally probable, for the coordinate which defines the position of a point may be chosen in a variety of ways. For instance, the position of P on AB may be defined by x, the length AP, or by y the ratio AP/PB. In terms of x, the probability that the point lies on P_1P_2 is

$$\frac{x_2 - x_1}{AB}.$$

In terms of y, the same probability is

$$\frac{y_2 - y_1}{(1 + y_1)(1 + y_2)}.$$

All values of x between 0 and AB may be described as equally likely; but all values of y cannot be so described.

Assuming that the probability of a marked point on AB lying in the segment AP has a definite meaning, it must depend on the position of P, i.e. it must be a function of x, if $AP = x$. Denote it by $f(x)$. Then $f(x)$ is necessarily a function for which $f(x_2) - f(x_1) \geqslant 0$, if $x_2 - x_1 > 0$; for if P_2 lies between P_1 and B, the probability that a point lies in AP_2 cannot be less than the probability that it lies in AP_1. Suppose that $f(x)$ is discontinuous at $x = x_0$, so that

$$f(x_0 + \alpha) - f(x_0 - \beta) \geqslant k,$$

however small α and β may be. If

$$AC = x_0, \quad AC' = x_0 - \beta, \quad AC'' = x_0 + \alpha,$$

then the probability that the point lies on the segment $C'C''$ is equal to or greater than k, whatever points C', C'' may be to the left and right of C respectively. This clearly implies that there is a finite probability that the marked point has the particular position C. Hence, if there are no particular points on AB of this nature, $f(x)$ must be a continuous function. Assuming further that $f(x)$ has a differential coefficient, the probability that a marked point lies on a segment δx of the line, when δx is small enough, is $f'(x)\,\delta x$; and the fact, that the point must be somewhere between A and B, is given by the condition

$$\int_0^{AB} f'(x)\,dx = 1.$$

Conversely, if it is assumed that the probability of a point lying on a sufficiently small segment δx of a line is proportional to $F(x)\,\delta x$, where x is the distance from one end, then the actual probability is

$$\frac{F(x)\,\delta x}{\int_0^l F(x)\,dx},$$

where l is the length of the line.

21. A precisely similar method may be used with respect to a point marked on a plane area. If x and y are rectangular co-ordinates in the plane of the area, and if it is known that there are no particular points on the area such that the probability of the marked point coinciding with one of them is finite, then when δx and δy are small enough, the probability of the marked point lying in the rectangle bounded by x, y, $x + \delta x$, $y + \delta y$, may be denoted by

$$f(x, y)\,\delta x\,\delta y,$$

subject to the condition

$$\iint f(x, y)\,dx\,dy = 1,$$

where the integral extends over the area within which the point is known to lie.

In particular, if A is the area, and if $f(x, y)$ is a constant, then the probability is

$$\frac{\delta x\,\delta y}{A};$$

and all positions of the marked point are said to be equally likely.

More generally, if $x_1, x_2, \ldots, x_n$ are n independent quantities continuously varying over a certain range, and if the probability of their having values confined to some smaller range has a definite meaning, then when $\delta x_1, \delta x_2, \ldots$ are small enough, the probability of their having a system of values lying between x_1 and $x_1 + \delta x_1$, x_2 and $x_2 + \delta x_2$, ..., will be of the form

$$\phi(x_1, x_2, \ldots, x_n)\,\delta x_1\,\delta x_2 \ldots \delta x_n,$$

subject to the condition that

$$\iint \ldots \int \phi(x_1, x_2, \ldots, x_n)\,dx_1\,dx_2 \ldots dx_n = 1,$$

where the integral is extended over the whole range of the variables.

It may be convenient, for purposes of calculation, to use new variables $y_1, y_2, \ldots, y_n$, functions of the old. The method of the Integral Calculus shews that

$$\iint \ldots \int \phi(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n,$$

extended over a certain range of the x's, which expresses the probability that the x's shall have values within that range, becomes

$$\iint \ldots \int D\psi(y_1, y_2, \ldots, y_n)\, dy_1 dy_2 \ldots dy_n,$$

extended over the corresponding range of the y's; where $\psi(y_1, y_2, \ldots, y_n)$ is $\phi(x_1, x_2, \ldots, x_n)$ expressed in terms of the y's, and

$$D = \begin{vmatrix} \frac{\partial x_1}{\partial y_1}, & \frac{\partial x_2}{\partial y_1}, & \ldots, & \frac{\partial x_n}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2}, & \frac{\partial x_2}{\partial y_2}, & \ldots, & \frac{\partial x_n}{\partial y_2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial x_1}{\partial y_n}, & \frac{\partial x_2}{\partial y_n}, & \ldots, & \frac{\partial x_n}{\partial y_n} \end{vmatrix}.$$

Hence the probability that the y's should have values between y_1 and $y_1 + \delta y_1$, y_2 and $y_2 + \delta y_2$, ..., is

$$D\psi(y_1, y_2, \ldots, y_n)\, \delta y_1 \delta y_2 \ldots \delta y_n.$$

Returning to the case of two independent variables, denote as before by $f(x, y)\,\delta x\,\delta y$ the probability that the variables lie between x and $x + \delta x$, y and $y + \delta y$ respectively.

If the range of the variables is unlimited, the probability that the first lies between x and $x + \delta x$ is

$$\int_{-\infty}^{\infty} f(x, y)\, \delta x\, dy,$$

which is of the form $F(x)\,\delta x$, $F(x)$ being a function of x only. Similarly, if

$$\int_{-\infty}^{\infty} f(x, y)\, dx = G(y),$$

the probability that the second variable lies between y and $y + \delta y$ is $G(y)\,\delta y$.

Suppose now that it is known that the probability p_A, that the first variable lies between x and $x+\delta x$, is $F(x)\,\delta x$, while the similar probability p_B for the second variable is $G(y)\,\delta y$. We have seen that

$$p_{AB}=p_Ap_B+p_{AB}p_{A'B'}-p_{AB'}p_{A'B};$$

and unless $p_{AB}p_{A'B'}-p_{AB'}p_{A'B}$ is zero, it does not follow that the probability, that the two variables lie between x and $x+\delta x$, y and $y+\delta y$ respectively, is $F(x)\,G(y)\,\delta x\delta y$. Now the assumed data give no information as to the value of $p_{AB}\,p_{A'B'}-p_{AB'}\,p_{A'B}$; and therefore, from the assumed data, it is not possible to determine the probability that the two variables simultaneously lie between given limits.

The same is obviously true when the range of the variables is limited. A similar result holds when the number of variables exceeds two.

Illustrations.

22. In the following illustrations it will be assumed, unless the opposite is stated, that all positions of a point marked on a line are equally likely.

I. A point is marked at random on a unit line. What is the probable value of the sum of the squares of the two parts into which it divides the line?

The probability, that the distance of the marked point from one end of the line lies between x and $x+\delta x$, is δx. The corresponding value of the sum of the squares of the two parts s is given by

$$s=1-2x+2x^2.$$

Hence the probable value of s is

$$\int_0^1 (1-2x+2x^2)\,dx=\tfrac{2}{3}.$$

It is instructive to consider this simple example from another point of view. When s is given, there are two values of x, viz.

$$x_1=\tfrac{1}{2}+\sqrt{\tfrac{1}{2}s-\tfrac{1}{4}},\quad x_2=\tfrac{1}{2}-\sqrt{\tfrac{1}{2}s-\tfrac{1}{4}}.$$

Hence $$\delta x_1=\tfrac{1}{4}\frac{\delta s}{\sqrt{\tfrac{1}{2}s-\tfrac{1}{4}}},\quad -\delta x_2=\tfrac{1}{4}\frac{\delta s}{\sqrt{\tfrac{1}{2}s-\tfrac{1}{4}}};$$

and if the sum of the squares of the parts lies between s and $s+\delta s$, the point must lie on one of two segments of the line, each of which is of length $\frac{1}{4}\frac{\delta s}{\sqrt{\frac{1}{2}s-\frac{1}{4}}}$. The probability for this is $\frac{1}{2}\frac{\delta s}{\sqrt{\frac{1}{2}s-\frac{1}{4}}}$.

Hence when the point is marked, the probability that the sum of the squares of the parts lies between s and $s+\delta s$ is $\frac{\delta s}{\sqrt{2s-1}}$. The extreme values of s and $\frac{1}{2}$ are 1. Hence the probable value of s is

$$\int_{\frac{1}{2}}^{1}\frac{s\,ds}{\sqrt{2s-1}}=\tfrac{2}{3}.$$

II. Two points are marked at random on a unit line. What is the probable value of the sum of the squares of the three parts?

The probability, that the distances of the two points from one end of the line lie between x and $x+\delta x$, y and $y+\delta y$ respectively, is $\delta x\,\delta y$. If y is less than x, the sum of the squares is $y^2+(x-y)^2+(1-x)^2$; if y is greater than x, the sum is $x^2+(y-x)^2+(1-y)^2$. Hence the probable value is

$$\int_0^1 dx\left[\int_0^x\{y^2+(x-y)^2+(1-x)^2\}\,dy+\int_x^1\{x^2+(y-x)^2+(1-y)^2\}\,dy\right]=\tfrac{1}{2}.$$

III. A point P_1 is marked at random on a unit line AB, and then a point P_2 is marked at random on P_1B. What is the probable value of the sum of the squares of the three parts?

The probability that AP_1 lies between x and $x+\delta x$ is δx; and the probability that P_1P_2 lies between y and $y+\delta y$ is $\frac{\delta y}{1-x}$. Hence the required probable value is

$$\int_0^1 dx\int_0^{1-x}\frac{dy}{1-x}\{x^2+y^2+(1-x-y)^2\}=\tfrac{5}{9}.$$

IV. When $n-1$ points are marked on a unit line, they divide it into n segments. The lengths $x_1, x_2, \ldots, x_{n-1}$, of $n-1$ of these are arbitrary, subject to the condition that their sum does not exceed unity. The question suggests itself: What is the probability that the $n-1$ segments have lengths lying between x_1 and $x_1+\delta x_1$, x_2 and $x_2+\delta x_2$, ... respectively?

Denote the distances of the $n-1$ points of division from one end of the line by $y_1, y_2, \ldots, y_{n-1}$. When the points are assigned, the segments are unaltered for any permutation of the points among themselves. For one particular sequence of the points, the y's will be in ascending order, so that

$$y_1 = x_1,$$
$$y_2 = x_1 + x_2,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$y_{n-1} = x_1 + x_2 + \ldots + x_{n-1}.$$

Since there are $(n-1)!$ permutations of the points, the probability that the segments will have the values corresponding to this set of points is $(n-1)!\,\delta y_1 \delta y_2 \ldots \delta y_{n-1}$.

Now the Jacobian D of the above set of equations is unity. Hence, by the theorem on p. 68, the probability that the $n-1$ segments have lengths between x_1 and $x_1 + \delta x_1$, x_2 and $x_2 + \delta x_2, \ldots$ is

$$(n-1)!\,\delta x_1 \delta x_2 \ldots \delta x_{n-1}.$$

This agrees with the fact that

$$\int_0^1 dx_1 \int_0^{1-x_1} dx_2 \ldots \int_0^{1-x_1-x_2-\ldots-x_{n-2}} dx_{n-1} = \frac{1}{(n-1)!}.$$

Thus, if a unit line has $n-1$ points marked at random on it, the probable value of the sum of the squares of the n parts into which they divide the line is

$$(n-1)! \int_0^1 dx_1 \int_0^{1-x_1} dx_2 \ldots \int_0^{1-x_1-\ldots-x_{n-2}} S\, dx_{n-1} = \frac{2}{n+1},$$

where S denotes $x_1^2 + x_2^2 + \ldots + x_{n-1}^2 + (1 - x_1 - \ldots - x_{n-1})^2$.

V. It is interesting to notice how these results are modified, when all positions of the marked points are not equally likely. Suppose that the probability of the distance of a point from one end of the unit line lying between x and $x + \delta x$ is proportional to $x(1-x)\,\delta x$, so that the point is more likely to lie in the central part of the line than at the ends. Then, since

$$\int_0^1 x(1-x)\,dx = \tfrac{1}{6},$$

the probable value of the sum of the squares of the two parts, into which a point divides the line, is

$$6\int_0^1 x(1-x)(1-2x+2x^2)\,dx = \tfrac{3}{5}.$$

As would be expected, this is less than when all positions of the point are equally likely.

VI. From the point of view of the present chapter, the result of Ex. X of Chapter II (p. 22) may be expressed as follows:—If n points are marked on a unit line and all positions are equally likely for each point, the probability that the n points all lie on a continuous portion of the line of length x is $nx^{n-1}-(n-1)x^n$.

The problem takes a rather different form when the line, on which the points are marked, is closed. Suppose points are marked on a closed curve of unit length: and assume that the probability, that a marked point lies on a given continuous segment of the curve of length l, is equal to l; i.e. in the sense already used, that all positions of the point are equally likely. Then, when n points are marked on the curve, what is the probability that some continuous segment of the curve of length x is free from points? Assign a positive direction along the curve; and starting from each of the n points, lay off a length $1-x$ in the positive direction. If a portion x of the curve is free from points, the n points must all lie on one of these n segments of length $1-x$. The probability, that the n points all lie on a particular one of these segments, is $(1-x)^{n-1}$.

First, let x be greater than $\frac{1}{2}$. Then, if the points lie on a particular one of the segments, they cannot lie on any other; and the required probability is $n(1-x)^{n-1}$.

Next, suppose that $\frac{1}{2}>x>\frac{1}{3}$. Two segments may now have in common a part of total length $1-2x$, which consists of two arcs starting respectively where the segments start. The case, in which the n points all lie on this portion of length $1-2x$, has been taken into account twice, once with each of the segments. Hence when $\frac{1}{2}>x>\frac{1}{3}$, the required probability is

$$n(1-x)^{n-1}-\frac{n(n-1)}{1\,.\,2}(1-2x)^{n-1}.$$

There is no difficulty in continuing this reasoning. The general result is that, when $\frac{1}{m}>x>\frac{1}{m+1}$, the required probability is

$$\sum_{r=1}^{r=m}(-1)^{r+1}\frac{n!}{r!\,(n-r)!}(1-rx)^{n-1}.$$

It is not difficult to shew that, when n is large and x is $\frac{\log n}{n}$, this result is roughly ·63, and that it diminishes rapidly for larger values of x. Hence, when a large number n of points are marked, the probability of a gap between them materially exceeding $\frac{\log n}{n}$ is very small.

VII. The problem of the comparative regularity of a random distribution of points on a closed curve may be looked at from another point of view. With the notation already used, the probability p, that just m of the n points lie on a continuous portion, length x, of the unit curve, is given by

$$p = \frac{n!}{m!\,(n-m)!}\,x^m (1-x)^{n-m}.$$

Putting $$m = xn + \mu,$$
and using the approximate expressions for the factorials, it is found that, when terms of the order $1/n^2$ are neglected,

$$\log p = -\tfrac{1}{2}\log 2\pi n x(1-x) - \frac{\mu^2 + (1-2x)\mu}{2nx(1-x)};$$

so that, reintroducing m,

$$p = e^{-\frac{(2x-1)^2}{8nx(1-x)}}\,\frac{e^{-\frac{\{m-x(n+1)+\frac{1}{2}\}^2}{2nx(1-x)}}}{\sqrt{2\pi n x(1-x)}}.$$

If x is $n^{-1+\alpha}$, where α is positive, the first factor of p is very nearly unity when n is sufficiently great. Hence p may be written

$$\frac{e^{-\frac{(m-m_0)^2}{h}}}{\sqrt{\pi h}};$$

and the probability, that the number of points on a segment of length $n^{-1+\alpha}$ lies between m_1 and m_2, is

$$\sum_{m=m_1}^{m=m_2} \frac{1}{\sqrt{\pi h}}\,e^{-\frac{(m-m_0)^2}{h}}.$$

Now tables already quoted shew that

$$\int_{m_0-2\sqrt{n}}^{m_0+2\sqrt{n}} \frac{1}{\sqrt{\pi h}}\,e^{-\frac{(m-m_0)^2}{h}}\,dm = \cdot 995 \text{ nearly},$$

and the sum differs very little from the integral. Hence, when n is large enough, the probability that the number of points on an arc of length $n^{-1+\alpha}$ lies between $n^{\alpha} + 2\sqrt{2}n^{\frac{\alpha}{2}}$ and $n^{\alpha} - 2\sqrt{2}n^{\frac{\alpha}{2}}$ is ·995. For such an arc then, however small α may be, the probability is great that the point-density differs very little from its mean value. No such conclusion can be drawn for an arc of length n^{-1}, as is obvious from the nature of the problem.

23. VIII. It will be assumed in what follows that, when a point is marked on a unit sphere, all positions of the point are equally likely, in the sense that, if S is the area of the spherical surface on one side of a closed curve drawn on the surface, then the probability that a marked point lies on that side of the curve is $\frac{S}{4\pi}$.

(i) Two points are marked on a unit sphere. What is the probability that the angular distance between them does not exceed α?

If A is one of the points, and a small circle of radius α is described with A as centre, its area is $2\pi(1 - \cos\alpha)$. Now, if the distance between the two points does not exceed α, the second point must lie either on the small circle or on the same side of it as A. Hence the required probability is $\frac{1}{2}(1 - \cos\alpha)$.

It follows at once that the probability, that the distance between two points marked on the sphere lies between α and $\alpha + \delta\alpha$, is $\frac{1}{2}\sin\alpha\,\delta\alpha$.

(ii) Three points are marked on a unit sphere. What is the probability that there is a small circle of radius $\alpha\,(< \frac{1}{2}\pi)$, on or within which all three points lie?

Let PQR, PQS be two small circles on the sphere of radius α. With P as centre and radius 2α, describe the arc $R_1T_1S_1$ of a small circle, touching the above circles in R_1, S_1; and with Q as centre and radius 2α, describe the arc $R_2T_2S_2$ touching the circles in R_2, S_2. Then an inspection of the figure shews at once that any point U within the closed curve $RR_1T_1S_1SS_2T_2R_2R$ is such that P, Q, U lie within a small circle of radius α; while if U is without this closed curve, P, Q, U, do not lie within any small circle of radius α. Let O, O' be the centres of PQR, PQS.

Then

$$\text{area } RR_1T_1S_1SS_2T_2R_2R = 2 \text{ area } POR_1T_1S_1O'P + 2 \text{ area } OR_1RR_2O - \text{area } POQO'P.$$

If the angles OPO' and POQ are β and γ, then

$$\begin{aligned} \text{area } POR_1T_1S_1O'P &= \beta(1-\cos 2\alpha), \\ \text{area } OR_1RR_2O &= \gamma(1-\cos\alpha), \\ \text{area } POQO'P &= 2\beta + 2\gamma - 2\pi. \end{aligned}$$

Hence the probability, that U lies within $RR_1T_1S_1SS_2T_2R_2R$, is

$$\frac{2\pi - 2\beta\cos 2\alpha - 2\gamma\cos\alpha}{4\pi}.$$

If the distance PP_1 is θ, it has been seen that the probability of the distance of two points on the sphere being between θ and $\theta+\delta\theta$ is $\frac{1}{2}\sin\theta\delta\theta$. Hence the probability, that there is a small circle of radius α containing the three points, is

$$\frac{1}{4\pi}\int_0^{2\alpha} \sin\theta\,(\pi - \beta\cos 2\alpha - \gamma\cos\alpha)\,d\theta,$$

while from the spherical quadrilateral $OPO'P$,

$$\cos\frac{\beta}{2} = \cot\alpha\tan\frac{\theta}{2}, \quad \cos\frac{\gamma}{2} = \sin\frac{\beta}{2}\cos\frac{\theta}{2}.$$

The integral is readily evaluated; and the required probability is found to be

$$(1-\cos\alpha)^2(1+\tfrac{5}{4}\cos\alpha).$$

This is unity when $\alpha = \frac{1}{2}\pi$, as it obviously should be.

From the result it follows that, when three points are marked on a unit sphere, the probability, that the radius of the small circle through them lies between α and $\alpha+\delta\alpha$, is

$$\tfrac{3}{4}\sin\alpha(1-\cos\alpha)(1+5\cos\alpha)\,\delta\alpha.$$

(iii) Three points A, B, C are marked on a unit sphere, all positions of each being equally likely. They are joined by the shorter arcs of the great circles BC, CA, and AB. What is the probable area of the spherical triangle ABC so formed?

The probability, that the arc BC lies between a and $a+da$, is $\frac{1}{2}\sin a\,da$, a lying between 0 and π. Similarly the probability, that the arc CA lies between b and $b+db$, is $\frac{1}{2}\sin b\,db$. The angle ACB lies between 0 and π; and the probability, that it lies between C and $C+dC$, is $\dfrac{1}{\pi}dC$.

Hence, when all positions of each of the three points are equally likely, the probability that the elements BC, CA, and angle ACB lie respectively between a and $a+da$, b and $b+db$, C and $C+dC$, is

$$\frac{1}{4\pi}\sin a\sin b\,da\,db\,dC.$$

Now the area of the spherical triangle is $A+B+C-\pi$. Hence the required probable value of the area is

$$\frac{1}{4\pi}\int_0^\pi\int_0^\pi\int_0^\pi (A+B+C-\pi)\sin a\sin b\,da\,db\,dC.$$

Now $$\int (A+B)\,dC=(A+B)\,C-\int C\,(dA+dB).$$

From the trigonometry of the spherical triangle,

$$\tan\tfrac{1}{2}(A+B)=\frac{\cos\frac{1}{2}(a-b)}{\cos\frac{1}{2}(a+b)}\cot\tfrac{1}{2}C;$$

thus

$$dA+dB=-\frac{\cos a+\cos b}{1+\cos a\cos b+\sin a\sin b\cos C}dC,$$

so that

$$\int_0^\pi (A+B)\,dC=\Big[(A+B)\,C\Big]_{C=0}^{C=\pi}+\int_0^\pi\frac{C\,(\cos a+\cos b)\,dC}{1+\cos a\cos b+\sin a\sin b\cos C}.$$

Now it is clear from a figure that, when $C=\pi$, $A+B$ is 0 or 2π, according as $a+b$ is less or greater than π.

Further, it is easy to verify that

$$\int_0^\pi\int_0^\pi\frac{(\cos a+\cos b)\sin a\sin b\,da\,db}{1+\cos a\cos b+\sin a\sin b\cos C}=0.$$

Hence

$$\int_0^\pi\int_0^\pi\int_0^\pi (A+B)\sin a\sin b\,da\,db\,dC=\int_0^\pi\int_0^\pi k.\sin a\sin b\,da\,db,$$

where k is 0 if $a+b<\pi$, $k=2\pi^2$ if $a+b>\pi$: so the integral $=4\pi^2$.

Also $$\int_0^\pi \int_0^\pi \int_0^\pi (C - \pi) \sin a \sin b \, da \, db \, dC = 2\pi^2.$$

Hence the required probable value of the area is $\frac{3}{2}\pi$.

In a similar way the probability, that the area of the triangle should lie between S and $S + dS$, may be determined.

24. IX. The position of a point on a sphere is determined by two angles θ and ϕ, its co-latitude and longitude measured from a given pole and a given meridian. The position of a *figure*, of given shape, on a sphere may be determined as follows. Let A, B be two marked points of the figure. Denote by θ and ϕ the co-latitude and longitude of A from a pole O, and by ψ the angle OAB. Then the position of the figure is completely determined by the three angles θ, ϕ, ψ; and all positions are given by values of these angles lying respectively between 0 and π, 0 and 2π, 0 and 2π. The probability, that the figure has a position in which the three angles lie between θ and $\theta + \delta\theta$, ϕ and $\phi + \delta\phi$, ψ and $\psi + \delta\psi$, will be of the form

$$F(\theta, \phi, \psi) \, \delta\theta \, \delta\phi \, \delta\psi.$$

Now an element of area of the sphere surrounding A is

$$\sin\theta \, \delta\theta \, \delta\phi.$$

Hence the probability, that A lies in the element of area δS, and OAB lies between ψ and $\psi + \delta\psi$, is

$$\frac{F(\theta, \phi, \psi)}{\sin\theta} \delta S \, \delta\psi.$$

When $F(\theta, \phi, \psi)/\sin\theta$ is a constant, all positions of the figure are said to be equally probable; and the constant is $1/8\pi^2$, since

$$\int_0^\pi d\theta \int_0^{2\pi} d\phi \int_0^{2\pi} d\psi \, . \, F(\theta, \phi, \psi) = 1.$$

Suppose two curves of arbitrary shape, but of finite length, are drawn on the sphere: and that when displaced, without change of shape, all positions of each are equally probable. What is the probable number of their intersections? This question is due to M. Poincaré*; but the solution given here is somewhat different from his.

Whatever positions are given to the two curves, one of them may, by a rotation of the sphere as a whole carrying the

* *Calcul des Probabilités*, p. 122.

curves with it, be brought into a standard position: so that the generality of the question is not affected, by supposing one of the curves fixed and the other equally likely to take any position with respect to it.

First, suppose the two curves to be arcs of great circles subtending angles α and β at the centre, and suppose that either α or β is less than π; so that, if the curves intersect at all, they can only intersect once. The probability of their intersection depends on α, β only; it may be denoted by $f(\alpha, \beta)$, where the function is symmetric in its two arguments. Mark a point C on AB, the α arc, dividing it into two arcs AC, CB, lengths α_1 and α_2. If the β arc intersects AB at all, it must either intersect AC or CB, and it cannot intersect both. Hence

$$f(\alpha_1 + \alpha_2, \beta) = f(\alpha_1, \beta) + f(\alpha_2, \beta).$$

Similarly $$f(\alpha, \beta_1 + \beta_2) = f(\alpha, \beta_1) + f(\alpha, \beta_2).$$

From these, it follows that

$$f(\alpha, \beta) = k\alpha\beta,$$

where k is a constant.

Suppose, next, that the fixed curve consists of two arcs of great circles, subtending α and β at the centre, where α and β are both less than π, and that the other curve is an arc γ $(< \pi)$ of a great circle. In this case, it is possible for the γ arc to intersect both the α arc and the β arc. Denote the probability of this by p. Then the probability that the γ arc intersects the α arc only is $k\alpha\gamma - p$; and the probability that it intersects the β arc only is $k\beta\gamma - p$; so that the probability, that the γ arc intersects the fixed curve at least once, is $k(\alpha + \beta)\gamma - p$. On the other hand, the probable number of intersections of the fixed curve and the moving curve is

$$1(k\alpha\gamma - p) + 1(k\beta\gamma - p) + 2p = k(\alpha + \beta)\gamma.$$

It may be noticed that in the first case, where the curves can only intersect once, the probability of intersecting and the probable number of intersections are the same thing.

There is clearly no difficulty in extending this result to the case, in which both fixed and moving curves consist of any number of small $(< \pi)$ arcs of great circles. The result, if one curve consists of arcs $\alpha_1, \alpha_2, \ldots, \alpha_m$, of great circles, and the second of arcs $\beta_1, \beta_2, \ldots, \beta_n$, is to give

$$k(\alpha_1 + \alpha_2 + \ldots + \alpha_m)(\beta_1 + \beta_2 + \ldots + \beta_n)$$

for the probable number of intersections, i.e.

$$kll',$$

where l is the whole length of one curve and l' that of the other. Now whatever the curves may be, points may be marked on them dividing them into arcs which are ultimately arcs of great circles. The result is therefore general.

The constant k is determined at once, by considering the case of two great circles. Here $l = l' = 2\pi$, and the number of points is 2; so that

$$k(2\pi)^2 = 2.$$

Hence the required result is $\dfrac{ll'}{2\pi^2}$.

CHAPTER VII

THEORY OF ERRORS

25. Practically all magnitude-determinations are liable to error. It is true that, if a basketful of apples is spread out on a table one can determine with certainty the number of apples; but when larger and larger collections of distinct objects are dealt with, a stage must be reached at which it is no longer possible to be certain of the result of counting, if only because of the length of time the process takes.

In general, a magnitude-determination cannot be reduced directly to a process of counting. It nearly always involves the observation of certain coincidences, such as that of a pointer with a division on a scale.

The imperfections, both of our senses and of the instruments used, necessarily imply then an uncertainty as to the result of the determination; and if several determinations of the same magnitude are made, they will be in general different from each other.

The question then arises, if a number of determinations have given

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

as the values of a certain magnitude, what use can be made of this result?

If the a's are given, without any indication of the way in which they were arrived at, a definite result can be obtained only by making some more or less arbitrary assumptions. Under such circumstances, the actual value chosen for the magnitude is, in general, the arithmetic mean of the results, viz.

$$\frac{1}{n}(a_1 + a_2 + \ldots + a_n).$$

If α were the true value of the magnitude, the errors implied in the data, reckoned positive when in excess, would be

$$a_i - \alpha, \qquad (i = 1, 2, \ldots, n).$$

The algebraic sum of the errors is

$$\sum_{i=1}^{n} (a_i - \alpha);$$

and the sum of the squares of the errors is

$$\sum_{i=1}^{n} (a_i - \alpha)^2.$$

If, in the last expression, α is regarded as a variable, its least value is given by

$$\sum_{i=1}^{n} (a_i - \alpha) = 0;$$

i.e., α is the arithmetic mean. Hence a choice, of the arithmetic mean of the a's as the true value, implies that the algebraic sum of the errors is zero, and that the sum of the squares of the errors is as small as possible. If a value in excess of the arithmetic mean is taken, it is implied that negative errors are more numerous, or greater, than positive errors; and if a value less than the arithmetic mean is taken, it is implied that the positive errors exceed the negative.

Conversely, the assumption that the sum of the squares of the errors is as small as possible, leads to the arithmetic mean as the required value.

This however is only one of a great variety of assumptions that might be made with respect to the errors. In general, each assumption will give a different value for the magnitude.

26. I. For instance, it might be assumed that the sum of the absolute values of the errors, that is, their values apart from sign, is as small as possible. If the n given values are taken to be in ascending order of magnitude, and the true value α is assumed to lie between a_r and a_{r+1}, the sum of the absolute values of the errors is

$$(2r - n)\alpha + \sum_{r+1}^{n} a_i - \sum_{1}^{r} a_i.$$

If $2r > n$, this is least when $\alpha = a_r$; and if $2r < n$, it is least when $\alpha = a_{r+1}$. If $2r = n$, it is independent of α.

Hence, if n is odd and equal to $2m + 1$, the sum of the absolute values of the errors is least when $\alpha = a_m$; while if n is even and equal to $2m$, the sum of the absolute values of the errors is constant for all values of α from a_m to a_{m+1}, and this value is less than any other.

II. Again, if it is assumed that the sum of the $2m$th powers of the errors is as small as possible, α will be given by

$$\sum_1^n (a_i - \alpha)^{2m-1} = 0.$$

If a_1 and a_n are the least and the greatest of the a's, and m is sufficiently great, every term in this equation is very small compared to either $(a_1 - \alpha)^{2m-1}$ or $(a_n - \alpha)^{2m-1}$. Hence approximately

$$\alpha = \tfrac{1}{2}(a_1 + a_n).$$

It may be inferred that the assumption, that the sum of the $2m$th powers of the errors is as small as possible, when $m > 1$, makes the determination of the true value depend more on the larger and the smaller observed values than on those in the middle of the series.

In the complete absence of any information as to how the a's were arrived at, it would always then seem most reasonable to take the magnitude equal to the arithmetic mean of the a's, as any other assumption would imply that either an excess of positive errors or an excess of negative errors has occurred.

If the set of n results are divided arbitrarily into two sets of r and $n-r$, and if

$$\frac{1}{r}\sum_1^r a_i = \alpha_1, \quad \frac{1}{n-r}\sum_{r+1}^{i=n} a_i = \alpha_2,$$

(it is no longer supposed that the a's are in order of magnitude), then

$$\frac{1}{n}\sum_1^n a_i = \frac{r\alpha_1 + (n-r)\alpha_2}{n}.$$

The quantity on the right is the *weighted* mean of α_1 and α_2, attaching weights r and $n-r$ to them, where r is the number of observations giving α_1 and $n-r$ is the number giving α_2.

Suppose now that two sets of determinations of a magnitude are given, viz.

$$a_1,\ a_2,\ \ldots,\ a_m,$$
$$b_1,\ b_2,\ \ldots,\ b_n;$$

and that the only thing known about them is, that the first set is obtained by better methods or by more accurate observers than the second. The first set, about which by themselves nothing is known, would give, by the rule of the arithmetic mean,

$$\alpha_1 = \frac{1}{m}\sum_1^m a_i;$$

and similarly the second set would give

$$\alpha_2 = \frac{1}{n}\sum_1^n b_i,$$

for the value required. In deducing the final result from α_1 and α_2, if nothing were known about them but the numbers of observations from which they were derived, weights proportional to these numbers would be used. When however it is known that the a's have been obtained by better methods than the b's, it is reasonable to attach a greater weight to α_1 in consequence of this knowledge, so that the final result would be

$$\frac{km\alpha_1 + n\alpha_2}{km + n},$$

where k is greater than unity. Until k is known, all that can be said from this formula is that the required value lies between

$$\alpha_1 \text{ and } \frac{m\alpha_1 + n\alpha_2}{m + n}.$$

What is always wanted practically is a definite result; and this can only be obtained by giving k a definite numerical value. The mere statement then that the first set of values are obtained by a better method than the second is of little practical use.

What is necessary to get a definite result is a statement of the relative weights to be attached respectively to any one of the first and any one of the second sets of values. This clearly cannot be reduced to a rule, but must be a matter of judgment in each particular case.

Even if it is known that the sets of determinations of a magnitude have been arrived at by the use of the same method and by equally accurate observers, the problem of obtaining a definite result from them remains largely indeterminate. If, however, it is known that the observations made are liable to error following some definite and known law, the problem takes a less indeterminate form.

27. III. Suppose the observations are such that the probability of an error in the determination lying between x and $x+\delta x$ is equal to $f(x)\,\delta x$, with the necessary relation

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

If α is the true value of the magnitude, the probability that n determinations give values lying between a_1 and $a_1+\delta a_1$, ..., a_n and $a_n+\delta a_n$, is

$$f(a_1-\alpha)\,f(a_2-\alpha)\ldots f(a_n-\alpha)\,\delta a_1\,\delta a_2\ldots\delta a_n.$$

Denote by $p(\alpha)\,\delta\alpha$ the *à priori* probability that the true value lies between α and $\alpha+\delta\alpha$. Then after the determinations have been made, the *à posteriori* probability, that the true value lies between α and $\alpha+\delta\alpha$, is by Bayes' formula

$$\frac{f(a_1-\alpha)\,f(a_2-\alpha)\ldots f(a_n-\alpha)\,p(\alpha)\,\delta\alpha}{\int_{-\infty}^{\infty} f(a_1-\alpha)\,f(a_2-\alpha)\ldots f(a_n-\alpha)\,p(\alpha)\,d\alpha};$$

and the probable value of α is

$$\frac{\int_{-\infty}^{\infty} \alpha f(a_1-\alpha)\ldots f(a_n-\alpha)\,p(\alpha)\,d\alpha}{\int_{-\infty}^{\infty} f(a_1-\alpha)\ldots f(a_n-\alpha)\,p(\alpha)\,d\alpha}.$$

If $p(\alpha)$ were known, the latter expression is a formula for the probable value of α, while the most probable value can be deduced from the former one. Which of the two is chosen as giving the true value must be a matter of judgment.

The difficulty is that $p(\alpha)$, from the nature of the case, can never be known: so that some assumption must be made with regard to it. It might be expected that the results would depend to a large extent on the assumptions made with respect to $p(\alpha)$; and this is no doubt the case if such assumptions are made quite arbitrarily. But within the range of the a's, it would be quite unreasonable to assume rapid variation of $p(\alpha)$. On the other hand, if $p(\alpha)$ varies slowly between A and B, where A is less than the smallest and B greater than the greatest of the a's, the values of the two above expressions depend in general but slightly on the way in which $p(\alpha)$ varies. The simplest assumption to make is, that $p(\alpha)$ is constant so long as

$$f(a_1-\alpha)f(a_2-\alpha)\dots f(a_n-\alpha)$$

has a sensible value.

With this assumption, the probable value of α is

$$\frac{\int_{-\infty}^{\infty}\alpha f(a_1-\alpha)f(a_2-\alpha)\dots f(a_n-\alpha)\,d\alpha}{\int_{-\infty}^{\infty}f(a_1-\alpha)f(a_2-\alpha)\dots f(a_n-\alpha)\,d\alpha};$$

and the most probable value of α is given by the equation

$$\frac{f'(a_1-\alpha)}{f(a_1-\alpha)}+\frac{f'(a_2-\alpha)}{f(a_2-\alpha)}+\dots+\frac{f'(a_n-\alpha)}{f(a_n-\alpha)}=0.$$

28. IV. That this last equation will not, in general, give the arithmetic mean for the most probable value is obvious; but it may be asked, for what law of error is the arithmetic mean the most probable value?

Write $$\frac{f'(x)}{f(x)}=F(x).$$

Then if $$\alpha=\frac{1}{n}(a_1+a_2+\dots+a_n),$$

the equation $$F(x_1)+F(x_2)+\dots+F(x_n)=0$$

is true when $$x_1+x_2+\dots+x_n=0,$$

so that

$$F(x_1)+F(x_2)+\dots+F(-x_1-x_2-\dots-x_{n-1})=0$$

is an identity. Hence

$$F'(x_1)=F'(-x_1-x_2-\dots-x_{n-1})=C,$$

$$F(x)=Cx+C',$$

and

$$F(x_1)+F(x_2)+\dots+F(x_n)=C(x_1+x_2+\dots+x_n)+nC',$$

so that $$C'=0.$$

Hence $$\frac{f'(x)}{f(x)}=Cx,$$

$$f(x)=De^{\frac{1}{2}Cx^2}.$$

Now the relation

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

shews that C must be negative, say $-2h$, and gives

$$\frac{1}{D} = \int_{-\infty}^{\infty} e^{-hx^2}\,dx = \sqrt{\frac{\pi}{h}}.$$

Hence the only law of error which makes the arithmetic mean the most probable value is that for which

$$f(x) = \sqrt{\frac{h}{\pi}}\,e^{-hx^2}.$$

This is called Gauss's law of error.

It may be noticed that, with this law, the probable value of α is

$$\frac{\int_{-\infty}^{\infty} \alpha e^{-h\sum_1^n (a_i-\alpha)^2}\,d\alpha}{\int_{-\infty}^{\infty} e^{-h\sum_1^n (a_i-\alpha)^2}\,d\alpha}$$

$$= \frac{\int_{-\infty}^{\infty} \alpha e^{-h\left(\alpha-\frac{1}{n}\sum_1^n a_i\right)^2}\,d\alpha}{\int_{-\infty}^{\infty} e^{-h\left(\alpha-\frac{1}{n}\sum_1^n a_i\right)^2}\,d\alpha}$$

$$= \frac{\int_{-\infty}^{\infty} \left(\alpha + \frac{1}{n}\sum_1^n a_i\right) e^{-h\alpha^2}\,d\alpha}{\int_{-\infty}^{\infty} e^{-h\alpha^2}\,d\alpha} = \frac{1}{n}\sum_1^n a_i;$$

so that, with Gauss's law, both the probable value and the most probable value are the arithmetic mean.

29. Gauss's law of error makes positive and negative errors equally probable. It will therefore certainly not apply to observations which are affected by systematic errors such as the index error of a sextant. Assuming that a set of observations have been corrected in some way for systematic error, it may be asked whether there is any reason to expect that the law of their errors will be Gauss's law.

In general, the actual error of an observation is the sum of a number of minor errors due to different causes and having very various values. To take a specific case, suppose that the actual error is the sum of N minor errors, each of which has one of the values $a_1, a_2, \ldots, a_s$. Denote by p_i the probability, that any minor error has the value a_i, so that

$$\sum_1^s p_i = 1.$$

Then if

$$\sum_1^s n_i = N,$$

the probability, that the actual error has the value

$$\sum_1^s n_i a_i,$$

is

$$\frac{N!}{n_1!\, n_2! \ldots n_s!} p_1^{n_1} p_2^{n_2} \ldots p_s^{n_s},$$

which is the coefficient of $x^{\sum_1^s n_i a_i}$ in

$$\left(\sum_1^s p_i x^{a_i}\right)^N.$$

Hence the probability, that the actual error lies between r_1 and r_2, is the sum of the coefficients of those powers of x in this expression whose indices lie between r_1 and r_2.

In the simplest case, viz. $s = 2$, $p_1 = p_2 = \frac{1}{2}$, $a_1 = -a_2$, it has already been seen in Chapter IV that this leads to Gauss's law; so that, if the actual error is made up of a sufficiently large number of minor errors of the same magnitude, each of which is equally likely to be positive or negative, then it follows Gauss's law.

30. This, no doubt, is a very special assumption; but the result holds good under much more general conditions.

In the above expression for the probability, q, of an error $\sum_1^s n_i a_i$, put

$$n_i = p_i N + x_i, \qquad (i = 1, 2, \ldots, s).$$

Then, when the factorials are replaced by their approximate expressions,

$$q = \frac{1}{(2\pi)^{\frac{s-1}{2}}} \sqrt{\frac{N}{(p_1 N + x_1) \dots (p_s N + x_s)}}$$

$$\times \frac{N^N p_1^{p_1 N + x_1} \dots p_s^{p_s N + x_s}}{(p_1 N + x_1)^{p_1 N + x_1} \dots (p_s N + x_s)^{p_s N + x_s}}$$

$$= \frac{1}{(2\pi)^{\frac{s-1}{2}}} \sqrt{\frac{N}{(p_1 N + x_1) \dots (p_s N + x_s)}} \cdot \frac{1}{D},$$

where $$D = \left(1 + \frac{x_1}{p_1 N}\right)^{p_1 N + x_1} \dots \left(1 + \frac{x_s}{p_s N}\right)^{p_s N + x_s}.$$

This gives

$$\log D = \sum_1^s (p_i N + x_i) \log\left(1 + \frac{x_i}{p_i N}\right)$$

$$= \sum_1^s (p_i N + x_i) \left(\frac{x_i}{p_i N} - \frac{x_i^2}{2p_i^2 N^2} + \frac{x_i^3}{3p_i^3 N^3} - \dots\right)$$

$$= \sum_1^s \left(\frac{x_i^2}{2p_i N} - \frac{x_i^3}{6p_i^2 N^2} + \dots\right),$$

since $$\sum_1^s x_i = 0.$$

Hence, when N is large enough,

$$q = \frac{1}{(2\pi N)^{\frac{s-1}{2}}} \frac{1}{\sqrt{p_1 p_2 \dots p_s}} e^{-\sum_1^s \frac{x_i^2}{2p_i N}}.$$

It follows that the probability of an error lying between r_1 and r_2 is Σq, where the sum is taken for those values of the x's (with integral differences) for which

$$\sum_1^s x_i = 0,$$

$$r_1 \leqslant \sum_1^s (p_i N + x_i) a_i < r_2.$$

31. So far, no supposition has been made about the minor errors. Suppose now that the probable value of a minor error is zero, so that

$$\sum_1^s p_i a_i = 0.$$

The equations of condition are then

$$\sum_1^s x_i = 0,$$

$$r_1 \leqslant \sum_1^s x_i a_i \leqslant r_2.$$

Consider now Σq for all values of the x's (with integral differences) which satisfy the relations

$$0 \leqslant \Sigma x_i \leqslant l,$$

$$r_1 \leqslant \Sigma a_i x_i \leqslant r_2.$$

When N is large enough, this is sensibly

$$\int \dots \int q \, dx_1 dx_2 \dots dx_s,$$

taken over the range given by the above inequalities.

Writing $\qquad x_i = \sqrt{p_i} y_i, \qquad (i = 1, 2, \dots, s),$

the integral becomes

$$\frac{1}{(2\pi N)^{\frac{s-1}{2}}} \int \dots \int e^{-\frac{1}{2N}(y_1^2 + y_2^2 + \dots + y_s^2)} dy_1 dy_2 \dots dy_s,$$

taken over the range

$$0 \leqslant \Sigma \sqrt{p_i} y_i \leqslant l,$$

$$r_1 \leqslant \Sigma \sqrt{p_i} a_i y_i \leqslant r_2.$$

Now make an orthogonal substitution, such that

$$z_1 = \frac{\Sigma \sqrt{p_i} a_i y_i}{\sqrt{\Sigma p_i a_i^2}},$$

$$z_2 = \Sigma \sqrt{p_i} y_i,$$

while $z_3, \dots, z_s$ are chosen consistently with these relations, which can certainly always be done. The integral then becomes

$$\frac{1}{(2\pi N)^{\frac{s-1}{2}}} \int \dots \int e^{-\frac{1}{2N}(z_1^2 + z_2^2 + \dots + z_s^2)} dz_1 dz_2 \dots dz_s,$$

taken over the range

$$r_1 \sqrt{\Sigma p_i a_i^2} \leqslant z_1 \leqslant r_2 \sqrt{\Sigma p_i a_i^2},$$

$$0 \leqslant z_2 \leqslant l,$$

i.e.
$$\frac{1}{\sqrt{2\pi N}}\int_{r_1\sqrt{\Sigma p_i a_i^2}}^{r_2\sqrt{\Sigma p_i a_i^2}} e^{-\frac{z_1^2}{2N}}dz_1\int_0^l e^{-\frac{z_2^2}{2N}}dz_2,$$

or
$$\frac{l}{\sqrt{2\pi N\Sigma p_i a_i^2}}\int_{r_1}^{r_2} e^{-\frac{r^2}{2N\Sigma p_i a_i^2}}dr,$$

if l is small enough. Finally Σq, subject to the conditions

$$0=\Sigma x_i,$$
$$r_1 \leqslant \Sigma a_i x_i \leqslant r_2,$$

is
$$\frac{1}{\sqrt{2\pi N\Sigma p_i a_i^2}}\int_{r_1}^{r_2} e^{-\frac{r^2}{2N\Sigma p_i a_i^2}}dr.$$

This completes the formal proof that the error follows Gauss's law, if it arises as the sum of a sufficiently large number of minor errors, of which the probable value is zero.

It should be noticed that the quantity $\Sigma p_i a_i^2$, occurring in the above expression, is the probable value of the square of a minor error.

32. V. If each component error follows Gauss's law, it may be shewn that the resultant error also follows the law, quite independently of the number of components being great. Suppose that the actual error is the sum of two components, and that the probabilities of the two components lying between x_1 and $x_1+\delta x_1$, x_2 and $x_2+\delta x_2$, respectively are

$$\sqrt{\frac{h_1}{\pi}}e^{-h_1x_1^2}\delta x_1 \text{ and } \sqrt{\frac{h_2}{\pi}}e^{-h_2x_2^2}\delta x_2.$$

The probability that the component errors, assumed to be independent, satisfy these conditions simultaneously is

$$\frac{\sqrt{h_1h_2}}{\pi}e^{-h_1x_1^2-h_2x_2^2}\delta x_1\,\delta x_2.$$

Put
$$x_1+x_2=X, \quad -x_1+x_2=Y,$$
so that
$$\delta x_1\,\delta x_2=\tfrac{1}{2}\delta X\,\delta Y.$$

Then the probability that the sum and the difference of the two component errors lie respectively between X and $X+\delta X$, Y and $Y+\delta Y$, is

$$\frac{\sqrt{h_1h_2}}{\pi}e^{-\frac{1}{4}(h_1+h_2)Y^2-\frac{1}{2}(h_1-h_2)YX-\frac{1}{4}(h_1+h_2)X^2}\,\delta X\,\delta Y.$$

Hence the probability, that the sum of the component errors lies between X and $X+\delta X$, is

$$\frac{\sqrt{h_1 h_2}}{\pi}\,\delta X \int_{-\infty}^{\infty} e^{-\frac{1}{4}(h_1+h_2)\,Y^2-\frac{1}{2}(h_1-h_2)\,YX-\frac{1}{4}(h_1+h_2)\,X^2}\,dY$$

$$=\frac{\sqrt{h_1 h_2}}{\pi}\,\delta X \int_{-\infty}^{\infty} e^{-\frac{1}{4}(h_1+h_2)\left[Y-\frac{h_1-h_2}{h_1+h_2}X\right]^2-\frac{h_1+h_2}{h_1 h_2}X^2}\,dY$$

$$=\sqrt{\frac{h_1 h_2}{(h_1+h_2)\,\pi}}\,e^{-\frac{h_1+h_2}{h_1 h_2}X^2}\,\delta X.$$

This clearly involves the consequence that, if a number of independent component errors all follow Gauss's law and if the resultant error is compounded *linearly* from them in any way, then it also follows Gauss's law.

CHAPTER VIII

GAUSS'S LAW OF ERRORS

33. As has already been stated, when the probability of an error lying between x and $x+\delta x$ is

$$\sqrt{\frac{h}{\pi}}\,e^{-hx^2}\,\delta x,$$

the errors are said to follow Gauss's law.

The constant h in this formula is clearly a kind of measure of the precision of the observations, since the probability of an error of given magnitude diminishes as h increases. Since positive and negative errors are equally likely, the probable error, as defined in Chapter IV, is zero.

If
$$\int_{-d}^{d}\sqrt{\frac{h}{\pi}}\,e^{-hx^2}\,dx=\tfrac{1}{2},$$

the probability of an error lying within the interval from $-d$ to d, is equal to the probability of its lying without this interval. The interval from $-d$ to d is often spoken of as the 50 per cent. zone. Now the preceding equation may be written

$$\frac{1}{\sqrt{\pi}}\int_{0}^{d\sqrt{h}}e^{-x^2}\,dx=\tfrac{1}{4}.$$

In an appendix (p. 103), a table of values of

$$\frac{2}{\sqrt{\pi}}\int_{0}^{x}e^{-x^2}\,dx$$

is given for values of x from 0 to 3 proceeding by differences of ·1, from which by interpolation the value of the integral for intermediate values of x can be determined with considerable accuracy. For instance, the table at once gives

$$d\sqrt{h}=\cdot 477 \text{ to three places:}$$

which may be written

$$h=\frac{\cdot 910}{(2d)^2},$$

or
$$2d=\frac{\cdot 954}{\sqrt{h}};$$

giving the precision-constant in terms of the breadth of the 50 per cent. zone, and *vice versâ*.

If
$$\frac{1}{\sqrt{\pi}}\int_{-d'\sqrt{h}}^{d'\sqrt{h}} e^{-x^2}dx = \cdot 998,$$
the probability of an error lying outside the zone from $-d'$ to d' is $\frac{1}{500}$; or one may say the zone from $-d'$ to d' is practically certain to contain all the errors if the number of observations is not too large. Now the table gives
$$d'\sqrt{h} = 2\cdot 2,$$
so that
$$d' = 4\cdot 61 d,$$
i.e. the breadth of this zone of practical certainty is 4·61 times the breadth of the 50 per cent. zone.

Half the breadth of the 50 per cent. zone, i.e. d, is often called the probable error, though the phrase is *not* used in the sense defined in Chap. IV. There is no risk of confusion if it is remembered that, in this connection, the probable error is a positive quantity d such that the magnitude of an error, apart from sign, is equally likely to be greater or less than d.

Mean Error: Error of Mean Square.

34. There are two other multiples of $1/\sqrt{h}$ which are often used in practice.

The probable value, in the ordinary sense, of $|x|$, that is of the error apart from sign, is
$$= \sqrt{\frac{h}{\pi}}\int_{-\infty}^{\infty} |x|\, e^{-hx^2}dx$$
$$= 2\sqrt{\frac{h}{\pi}}\int_{0}^{\infty} xe^{-hx^2}dx = \frac{1}{\sqrt{h\pi}} = \frac{\cdot 564}{\sqrt{h}} \text{ to three places.}$$
This is called the mean error.

The probable value of the square of the error, is
$$\sqrt{\frac{h}{\pi}}\int_{-\infty}^{\infty} x^2 e^{-hx^2}dx = \frac{1}{2h}\cdot$$
The square root of this, viz. $\frac{\cdot 707}{\sqrt{h}}$ to three places, is called the error of mean square.

The result of differentiating the equation

$$\int_{-\infty}^{\infty} e^{-hx^2}dx = \sqrt{\frac{\pi}{h}}$$

n times with respect to h is

$$\int_{-\infty}^{\infty} x^{2n}e^{-hx^2}dx = \frac{1 \cdot 3 \ldots (2n-1)}{2^n}\sqrt{\frac{\pi}{h^{2n+1}}}.$$

Thus the probable value of the $2n$th power of the error is

$$1 \cdot 3 \ldots (2n-1)\left(\frac{1}{2h}\right)^n.$$

Whatever the value of h, this quantity increases continually with n when n is greater than h; it takes its least value when n is the integer between $h-\frac{1}{2}$ and $h+\frac{1}{2}$. The bearing of the occasional very large errors which Gauss's law implies is perhaps best realized in this way.

35. The probability that two observations give errors lying respectively between x_1 and $x_1+\delta x_1$, x_2 and $x_2+\delta x_2$, is

$$\frac{h}{\pi}e^{-h(x_1^2+x_2^2)}\delta x_1 \delta x_2.$$

Hence the probability that two observations give errors for which $|x_1-x_2|$ is equal to or less than l, is

$$\frac{h}{\pi}\iint e^{-h(x_1^2+x_2^2)}dx_1 dx_2,$$

taken over the range for which

$$-l < x_2 - x_1 < l.$$

As before, put

$$x_2+x_1 = Y, \quad x_2-x_1 = X,$$

so that $$\delta x_1 \delta x_2 = \tfrac{1}{2}\delta X \delta Y.$$

Then the required probability is

$$\frac{h}{2\pi}\iint e^{-\frac{1}{2}h(X^2+Y^2)}dX\,dY,$$

taken over the range $\quad -l < X < l$,

that is, $$\frac{h}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{1}{2}hY^2}dY\int_{-l}^{l} e^{-\frac{1}{2}hX^2}dX,$$

$$= \sqrt{\frac{h}{2\pi}}\int_{-l}^{l} e^{-\frac{1}{2}hX^2}dX.$$

In particular, if $l\sqrt{\frac{h}{2}} = {\cdot}477$ or $l = {\cdot}705 \times 2d$, this probability is $\frac{1}{2}$. The difference of the errors in two observations is therefore equally likely to be greater or less than ·705 times the breadth of the 50 per cent. zone.

This may be expressed by saying that, in two observations, the probable "spread" of the errors is ·705 times the breadth of the 50 per cent. zone. The question of determining, in this sense, the probable spread of the errors in a set of n observations may be treated as follows.

The probability, that the error of the ith of n observations lies between x and $x + \delta x$, and that the errors of the other $n - 1$ lie between x and $x + l$, is

$$\sqrt{\frac{h}{\pi}}\, e^{-hx^2}\delta x \left[\left(\frac{h}{\pi}\right)^{\frac{1}{2}} \int_x^{x+l} e^{-hy^2} dy\right]^{n-1}.$$

Therefore the probability, that the ith of the n observations has the algebraically smallest error and that the spread does not exceed l, is

$$\int_{-\infty}^{\infty} \sqrt{\frac{h}{\pi}}\, e^{-hx^2} dx \left[\left(\frac{h}{\pi}\right)^{\frac{1}{2}} \int_x^{x+l} e^{-hy^2} dy\right]^{n-1}.$$

Now any one of the observations may have the algebraically smallest error; and the cases so obtained are mutually exclusive. It follows that the probability, that the spread of the errors of n observations does not exceed l, is

$$n\left(\frac{h}{\pi}\right)^{\frac{n}{2}} \int_{-\infty}^{\infty} e^{-hx^2} dx \left[\int_x^{x+l} e^{-hy^2} dy\right]^{n-1}.$$

It may be observed that this result can also be established by the method of Example x, Chapter II.

The formula may be written

$$\frac{n}{\pi^{\frac{n}{2}}} \int_{-\infty}^{\infty} e^{-x^2} \left[\int_x^{x+l\sqrt{h}} e^{-y^2} dy\right]^{n-1} dx;$$

and for given numerical values of n and $l\sqrt{h}$, its value may be calculated by means of the table already referred to. When this is done for various values of $l\sqrt{h}$ and the same n, the particular value of $l\sqrt{h}$, which makes the probability $\frac{1}{2}$, may be approximately obtained by interpolation. In this way the following table is arrived at, giving the probable spread in terms of the breadth of the 50 per cent. zone :—

n	4	5	6	8	10	12
$l/2d$	1·47	1·67	1·84	2·07	2·24	2·37

It is easy to see that the probable spread increases without limit as n increases.

Combination of Determinations.

36. Assuming that the errors of observation of a certain quantity do, in fact, follow Gauss's law, the question arises as to what deductions can be drawn from a given set of determinations.

Denote by $$x_1,\ x_2,\ \dots,\ x_n,$$ the numerical values obtained in a set of n determinations: by x_0 the unknown true value of the quantity to be determined: and by h the unknown value of the precision-constant for the observations.

The probability of getting a set of n determinations lying respectively between x_1 and $x_1 + dx_1$, x_2 and $x_2 + dx_2$, ..., x_n and $x_n + dx_n$, is

$$\delta p = \left(\frac{h}{\pi}\right)^{\frac{n}{2}} e^{-h\sum_1^n (x_i - x_0)^2}\, dx_1 dx_2 \dots dx_n;$$

and the probability of getting a set of n determinations, which satisfy definite conditions, is the integral of δp over the range determined by the conditions. In particular, the integral of δp for all real values of the n variables is unity.

It is easy to verify that

$$\sum_1^n (x_i - x_0)^2 = n\left[\left(x_0 - \frac{1}{n}\sum_1^n x_i\right)^2 + \frac{1}{n}\sum_1^n x_i^2 - \left(\frac{1}{n}\sum_1^n x_i\right)^2\right].$$

Introduce new variables defined by

$$y_1 = \frac{1}{n}\sum_1^n (x_i - x_0),$$

$$y_j = \sqrt{\frac{n-j+1}{n(n-j+2)}}\left\{x_{j-1} - \frac{1}{n-j+1}(x_j + x_{j+1} + \dots + x_n)\right\},$$

for $\quad j = 2, 3, \dots, n;$

and denote by $-J$ a numerical constant, the Jacobian of the x's with respect to the y's. With these new variables, it will be found that

$$\frac{1}{n}\sum_1^n x_i^2 - \left(\frac{1}{n}\sum_1^n x_i\right)^2 = \sum_2^n y_i^2,$$

so that $$\delta p = -J\left(\frac{h}{\pi}\right)^{\frac{n}{2}} e^{-nh\sum_1^n y_i^2} dy_1 dy_2 \ldots dy_n.$$

The integral of

$$e^{-nh\sum_2^n y_i^2} dy_2 dy_3 \ldots dy_n,$$

taken over the range defined by

$$(\beta + \delta\beta)^2 \geqslant \sum_2^n y_i^2 \geqslant \beta^2,$$

is the product of $e^{-nh\beta^2}$, and the integral of

$$dy_2 dy_3 \ldots dy_n,$$

taken over the same range. Now the integral of this last differential over the range defined by

$$\beta^2 \geqslant \sum_2^n y_i^2 \geqslant 0$$

is clearly $C'\beta^{n-1}$, where C' is a numerical constant depending only on n. Hence the integral of

$$e^{-nh\sum_2^n y_i^2} dy_2 dy_3 \ldots dy_n,$$

taken over the range defined by

$$(\beta + \delta\beta)^2 \geqslant \sum_2^n y_i^2 \geqslant \beta^2,$$

is $$Ce^{-nh\beta^2}\beta^{n-2} d\beta,$$

where C is a numerical constant; and the integral of δp, taken over the same range, is

$$C - J\left(\frac{h}{\pi}\right)^{\frac{n}{2}} e^{-nh(y_1^2+\beta^2)}\beta^{n-2} dy_1 d\beta.$$

Now again introduce new variables defined by

$$y_1 = \frac{uv}{\sqrt{h}}, \quad \beta = \frac{v}{\sqrt{h}}; \text{ or } u = \frac{y_1}{\beta}, \quad v = \sqrt{h}\beta.$$

For these, $$\frac{\partial(y_1, \beta)}{\partial(u, v)} = \frac{v}{h};$$

so that the immediately preceding differential becomes

$$ke^{-nv^2(1+u^2)}\, v^{n-1}\, du\, dv,$$

where k is a numerical constant.

This is the probability that y_1/β shall lie between u and $u + du$, and $\beta\sqrt{h}$ shall simultaneously lie between v and $v + dv$.

Now $$\int_0^\infty e^{-nv^2(1+u^2)}\, v^{n-1}\, dv = \frac{k_1}{(1+u^2)^{\frac{n}{2}}},$$

where k_1 is a numerical constant, and

$$\int_{-\infty}^{\infty} e^{-nu^2v^2} du = \frac{1}{v}\sqrt{\frac{\pi}{n}}.$$

Hence the probability, that y_1/β lies between u and $u + du$, is $\dfrac{kk_1\, du}{(1+u^2)^{\frac{n}{2}}}$; and the probability, that $\beta\sqrt{h}$ lies between v and $v + dv$, is $k\sqrt{\dfrac{\pi}{n}}\, e^{-nv^2} v^{n-2}\, dv$.

37. The conclusions, which can be drawn from the given set of determinations, are therefore:

(i) the probability, that

$$\frac{x_0 - \frac{1}{n}\sum_1^n x_i}{\sqrt{\frac{1}{n}\sum_1^n x_i^2 - \left(\frac{1}{n}\sum_1^n x_i\right)^2}}$$

lies between u_1 and u_2, is

$$\frac{\int_{u_1}^{u_2} \frac{du}{(1+u^2)^{\frac{n}{2}}}}{\int_{-\infty}^{\infty} \frac{du}{(1+u^2)^{\frac{n}{2}}}};$$

(ii) the probability, that

$$\sqrt{h\left(\frac{1}{n}\sum_1^n x_i^2 - \left(\frac{1}{n}\sum_1^n x_i\right)^2\right)}$$

lies between v_1 and v_2, is

$$\frac{\int_{v_1}^{v_2} e^{-nv^2} v^{n-2}\, dv}{\int_0^\infty e^{-nv^2} v^{n-2}\, dv}.$$

If $$\int_{-\theta_n}^{\theta_n} \frac{du}{(1+u^2)^{\frac{n}{2}}} = \frac{1}{2}\int_{-\infty}^{\infty} \frac{du}{(1+u^2)^{\frac{n}{2}}},$$

then x_0 is equally likely to lie within or without the range from

$$\frac{1}{n}\sum_1^n x_i - \theta_n \sqrt{\frac{1}{n}\sum_1^n x_i^2 - \left(\frac{1}{n}\sum_1^n x_i\right)^2} \text{ to } \frac{1}{n}\sum_1^n x_i + \theta_n \sqrt{\frac{1}{n}\sum_1^n x_i^2 - \left(\frac{1}{n}\sum_1^n x_i\right)^2}.$$

With the phrase that has already been used, the breadth of the 50 per cent. zone as determined by n observations is

$$2\theta_n \sqrt{\frac{1}{n}\sum_1^n x_i^2 - \left(\frac{1}{n}\sum_1^n x_i\right)^2}.$$

For the smaller values of n, the value of θ_n is given by the following table:

n	θ_n
2	·4416
3	·3703
4	·3249
5	·2929
6	·2687
7	·2497
8	·2342
9	·2213
10	·2104.

When n is large, $\theta_n = ·6726/\sqrt{n}$ nearly.

38. It would appear from the result just obtained that, by sufficiently increasing n the number of observations, the probable difference between $\frac{1}{n}\sum_1^n x_i$ and the magnitude to be determined could be made as small as desired. In arriving at this result, it has however been assumed that x, the quantity observed, is susceptible of continuous variation. Now in fact this is never the case. What is taken down as the result of an observation is always a rational number. A series of observations will be represented by a series of numbers, say 11·921, 11·937, 11·918, ... which are written down to a certain decimal place, in this case the third; and each observation is represented by a certain integral multiple of ·001. In taking a reading, say 11·921, what is implied is that the position of the pointer or mark on the scale is nearer to 11·921 than to 11·920 or 11·922, the next two possible adjacent readings. In other words, regarded as a measure of the magnitude, 11·921 and $11·921 \pm \delta$, where δ is a proper fraction not exceeding $\frac{1}{2}$, are not to be distinguished between. It follows that the mean m of the readings and $m \pm \delta$, regarded as measures of the magnitude, are not to be distinguished between; and this means that the part of m, which extends beyond the third decimal place, is of no significance. The determination arrived at for the magnitude must be an integral multiple of ·001.

Hence, in taking each observation as an integer, it is implicitly assumed that the quantity to be determined is itself an integer and that every error is an integer.

If the errors are all integers, the law of error which is most nearly similar to Gauss's law is that, in which the probability of an error n is sensibly

$$\frac{1}{\sqrt{\pi N}} e^{-\frac{n^2}{N}}.$$

The sum of the series, of which this is the general term, for all integral values of n exceeds unity; but unless N is very small, the excess is extremely small. For instance, if $N = 3$ the excess is less than 4×10^{-13}.

Now, assuming this law, if $x_1, x_2, \ldots, x_n$ are the series of observed integral values of a magnitude, and if s is the true integral value, the probability of the set of observations is

$$\frac{1}{(\pi N)^{\frac{n}{2}}} e^{-\frac{1}{N}\sum_1^n (s - x_i)^2}.$$

If $$\frac{1}{n}\sum_1^n x_i = s_1, \quad \sum_1^n x_i^2 - \frac{1}{n}(\Sigma x_i)^2 = s_2,$$

the probability is

$$\frac{1}{(\pi N)^{\frac{n}{2}}} e^{-\frac{n(s - s_1)^2 + s_2}{N}}.$$

Whatever N may be, the greatest value of this for integral values of s is given by $s = s_1'$, where s_1' is the nearest integer to s_1; while for a varying N, the greatest value is given by

$$N = \frac{2s_2}{n} + 2(s_1' - s_1)^2.$$

NOTE

The rule on p. 4 differs from that usually given, mainly as regards the condition of equal likelihood. This proviso is by most writers put in the form:—provided that all the n results are equally likely. It is evident that, so far as the calculation is concerned, it is immaterial whether we say "the n results are equally likely" or "the n results are assumed to be equally likely."

It is not however the same thing to say "assuming all the n results to be equally likely" and "assuming each two of the n results to be equally likely." In the one case, the property of equal likelihood is predicated of the n results as a whole; and in the other, of each pair of them.

Suppose the rule to be modified so that the last clause runs "provided all the N results are assumed to be equally likely." Consider the restricted trial subject to the further limitation that condition A is satisfied. It has just N_A possible results; and in N_{AB} of them, the condition B is satisfied. It is not however possible to say in this case that the probability of condition B being satisfied in the restricted trial is N_{AB}/N_A. In order that this may be true, the property of equal likelihood must apply of the set of N_A results as a whole. Does this necessarily follow as a logical consequence from the assumption that the property of equal likelihood applies to the N results as a whole? It can only do so, if there is some criterion for distinguishing a set of results with the property of equal likelihood from a set which does not possess this property. In the absence of any such criterion, it is not possible to say that the probability of condition B being satisfied at the restricted trial is N_{AB}/N_A; and therefore, from the modified form of the rule, it is impossible to deduce (p. 6) the formula (iii) of Chapter I

$$p_B = p_{(A_1)B}\,p_{A_1} + p_{(A_2)B}\,p_{A_2} + \dots + p_{(A_s)B}\,p_{A_s}$$

without further assumptions.

It must in fact be assumed that, for each condition A, all the N_A results which satisfy condition A are equally likely.

So far as setting up the necessary formulæ, by which calculable probabilities can be determined, is concerned, the last clause of the rule may be stated in either of the forms:—

(i) provided that each two of the N results are assumed to be equally likely; or

(ii) provided that, for each condition A, the N_A results which satisfy condition A are assumed to be equally likely.

Hitherto, no criterion has ever been given for distinguishing between a set of results, which have the property of equal likelihood, and a set which has not. This is the true justification for saying that "each two of the results are assumed to be equally likely" rather than "each two of the results are equally likely."

TABLE OF THE INTEGRAL $I = \frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2}\,dx$

for values of $x = \cdot 1, \cdot 2, \ldots, 2\cdot 9, 3$.

x	I	x	I
·1	·1125	1·6	·9763
·2	·2270	1·7	·9838
·3	·3286	1·8	·9891
·4	·4284	1·9	·9928
·5	·5205	2·0	·9953
·6	·6039	2·1	·9970
·7	·6778	2·2	·9981
·8	·7421	2·3	·9989
·9	·7969	2·4	·9993
1·0	·8427	2·5	·9996
1·1	·8802	2·6	·9998
1·2	·9103	2·7	·99987
1·3	·9340	2·8	·99992
1·4	·9523	2·9	·99996
1·5	·9661	3·0	·99998

LIST OF COGNATE PAPERS

by Professor BURNSIDE.

1. On the probable regularity of a random distribution of points: *Messenger of Math.*, vol. xlviii (1919), pp. 47–48.
2. On errors of observation: *Camb. Phil. Soc. Proc.*, vol. xxi (1923), pp. 482–487.

   [In connection with this paper, consult Mr R. A. Fisher, Note on Dr Burnside's recent paper on errors of observation, *Camb. Phil. Soc. Proc.*, vol. xxi (1923), pp. 655–658.]
3. On errors of observation: *ib.*, vol. xxii (1924), pp. 26–27.
4. The problems of random flight and conduction of heat: *ib.*, vol. xxii (1924), pp. 167–168.
5. On the phrase "equally probable": *ib.*, vol. xxii (1924), pp. 669–671.
6. On the idea of frequency: *ib.*, vol. xxii (1924), pp. 726–727.
7. On an integral connected with the theory of probability: *Messenger of Math.*, vol. liii (1924), pp. 142–144.
8. On the approximate sum of selected terms from the multinomial expansion: *ib.*, vol. liv (1925), pp. 189–192.

INDEX

(*The numbers refer to the pages.*)

REPRINTED BY PHOTOLITHOGRAPHY BY
KIMBLE & BRADFORD
LONDON, W.I