NORMAL DISTRIBUTIO
HISTORIO-PHILOSOPHIC C

BY

SOLOMON GEBRECHRISTOS TEDLA (A

This dissertation is submitted in partial fulfillment of degree of

Statistics in the Department of Mathematics

UNIVERSITY OF NAIROBI

JUNE, 1998

DECLARATION

This is my original work and has not been presented for a degree in any other university.

Signature: ______ 28/7/98 ______,

SOLOMON GEBRECHRISTOS TEDLA (Abba)

This dissertation has been submitted for examination with my approval as university supervisor .

Signature: ______ 28/7/98 ______,

Prof. J. W. ODHIAMBO

DEPARTMENT OF MATHEMATICS

UNIVERSITY OF NAIROBI

P. O. BOX 30197,

NAIROBI - KENYA

DEDICATION

This work is dedicated to my beloved parents Captain **GEBRECHRISTOS TEDLA** and **LETTEKIDAN HAILU** who inculcated in me a true Christian spirit of service and dedication.

TABLE OF CONTENTS

Chapter Three

Normal Distribution:Properties and its Characterisations

Chapter Four

Theory of Normal Distribution in Hilbert Space

SUMMARY OF CONTENTS

In general the set-up of the research work has a logical intention. That is the chapters and their respective sections are logically arranged so that one can read and arrive at the results inductively or deductively.

Chapter one gives a preliminary background on the important research questions, namely, the schematic discernment of statistical and probabilistic research works and method of approach at University of Nairobi, and the characterization of normal distribution in Hilbert Space.

Chapter two focuses on the historio-philosophic development of probability and statistical thoughts and theories. As a concluding remark on this chapter, a tentative conclusion on the historio-philosophic approach of the Statistical Section of Mathematics at University of Nairobi on the mathematical probability and statistics is given.

In *chapter three* first we come across the historical context of normal distribution and its philosophic applications. Furthermore, different approaches in defining and analyzing the unique properties of normal distribution are given: modern and classical approaches. In addition, different models of investigations, like De Moivre-Laplace method, Adrians method, Hegen's hypothesis and so on, are studied in deriving the normal distribution. Also, The central importance of normal distribution in probabilistic and statistical studies is illustrated by the relationship between normal distribution with discrete and continuous distribution as well as properties of pure and applied mathematics; normal distribution is analyzed using the number theory and Maxwell's distribution of velocities and law of error. The highly developed mathematical methodologies, theorems and techniques of the characterization of normal distribution are elaborated. Using these techniques the theory of normal distribution in Hilbert space is studied.

Chapter four develops the central issue of the statement of problem logically. That is, first after a brief historical analysis, the fundamental definitions and properties of Hilbert space are described. Then the geometry of Hilbert space with operator theory are presented. Finally, the necessary properties and theorems of normal distribution in Hilbert space are clarified.

ACKNOWLEDGMENTS

I present my deep gratitude to my religious Order, namely, Franciscan Capuchin Friar Minors (OFM Cap.) and my home Province of Eritrea, who gave me this opportunity of challenge and happily covering without any delay all the necessary expenses for my studies. Also I cannot forget the warm hospitality of my Kenyan confriars and other jurisdictions, and the Vice-Province of Kenya.

Words can't reflect accurately one's innermost gratitude and appreciation for your beneficiaries materially and spiritually. Nonetheless, I acknowledge wholeheartedly the scholarly dedication and fraternal approaches of all my lecturers in the last five academic years, particularly my supervisor Prof. J. W. Odhiambo, and Dr. J. O. Owino, Dr. M. M. Manene, Dr. F. Njui, Dr. K. Gacii, my former CUEA lecturers, Mr. C. Achola and others. I say humbly and sincerely may God reword you accordingly.

I do appreciate all dear friends, brothers and sisters who encouraged me tirelessly, in particular, Fr. Andemaraim Tesfamariam, Daughters of St. Anne, Comboni Sisters, Capuchin Sisters, Christian Brothers and other religious institutes. Unforgettable companion John Mwangi Nderitu has been true witness to the spirit of cooperation and understanding in all our studies, research works and fraternal relationship.

May the true creator of natural order and human wisdom be always glorified through our life and works.

Chapter One

INTRODUCTION

1.1 General Introduction

The present study has two important research questions, namely, (i) the schematic discernment of statistical and probabilistic research works and method of approach at University of Nairobi, and (ii) the characterization of normal distribution in Hilbert Space.

The first question comes to light as a logical consequence of historio-philosophic development of thoughts, as well as discussion and analysis of probability theory, statistics, and comparative description and exploration of different approaches to probability and corresponding schools. Generally speaking, following their respective approaches to mathematical statistics and probability theory, there are four approaches to probability: (a) *the classical approach*, which adheres to the notion of equally probable cases by reason of symmetry, or probability is defined as the ratio of the favourable to the possible cases, (b) *the empirical approach*, in which by virtue of the so called empirical law of chance, based on the notion of repeatable events whose frequency on a large number of trials gives the probability almost certainly and exactly, (c) *the asymptotic approach*(frequency theory), by considering an infinite sequence of trials defines probability as the limiting value of the frequency, and (d) *the subjectivistic*(degree of belief) approach, considers probability as a measure of the degree of belief of a given subject in the occurrence of an event.

In chapter two, using some specific criteria, we will pose to answer the question that says "is there any affinity of University of Nairobi to a specific school of probability and statistics?." Hence, as historical fact and dialectic progress of schools we will try to identify the Statistical Section of Mathematics Department at University of Nairobi with a school of statistics it belongs to.

The heart of the present synthetic research is the comparative analysis of statistical entities with pure mathematical properties. In other words, we will study the characterization of normal distribution in Hilbert Space, as well as the Normal Law in Number Theory.

Normal distribution possesses very rich properties, especially for the Applied Statistician. Many scholars, starting from De Moivre-Laplace up to Lindeberg-Feller, attributed great importance to the properties of normal law. Contemporary scholars too adhere to this affirmation, and due to the rapid progress and application of probability theory and statistical principles in different fields of science, this dominant factor is vivid. In line with this we will try to identify the characterization of normal law in the well known and rapidly expanding branch of pure mathematics, namely Hilbert Space.

Hilbert Space, named after a German mathematician - David Hilbert(1862-1943), plays an important role in the functional analysis. Hilbert space theory is a useful language for applied mathematics, engineers and scientists who apply mathematics. In general Hilbert space theory deals with a wide range of topics. We will concentrate on the properties that expose our characterization of normal law.

1.2 Statement of the problem

The research problem is stated in question format.

Major questions:

- How can we analyze the normal distribution in historio-philosophic context?
- How can we present "a characterization of normal distribution in Hilbert space?"

Guiding questions:

- What are the important probabilistic and stastistical historical periods, and philosophical importance of probability theory?
- Can we classify the different approaches to probability theory and corresponding schools?

- With respect to the historical development of schools of statistics and probability, and comprehensive framework of the applications of normal law, using certain criteria, can we say that the University of Nairobi belongs to one specific school or many schools?
- Can we compare probabilistic and statistical theories of normal distribution with properties of number theory and Hilbert space?
- What do we mean by the characterization of normal distribution?
- Is there any advantage and contribution in statistical mathematics when we study the characterization of normal distribution in Hilbert Space and number theory?
- What are the prominent properties of Hilbert space that fit Normal distribution?

1.3 Objectives of the study

The objectives of the present study are:

- To explore the characterization of Normal distribution in Hilbert space.
- To investigate the relationship between statistical principles and pure mathematical properties.
- To assess if the Statistical Section of Mathematics at University of Nairobi has affinity with specific schools of probability and statistics in historio-philosophic context, hence, to give a tentative conclusion.

1.4 Brief literature review

The present literature review examines (i) the efforts that have been made over time to explore the theory of characterization of normal distribution, and (ii) the characterization of normal distribution in Hilbert space, as well as (iii) the statistical research works done at University of Nairobi

The characterization of a distribution is the investigation of those unique properties enjoyed by that distribution. Mathai and Pederzoli (1977) have compiled and put together their studies with recent research papers and published then in a form of a

- With respect to the historical development of schools of statistics and probability, and comprehensive framework of the applications of normal law, using certain criteria, can we say that the University of Nairobi belongs to one specific school or many schools?
- Can we compare probabilistic and statistical theories of normal distribution with properties of number theory and Hilbert space?
- What do we mean by the characterization of normal distribution?
- Is there any advantage and contribution in statistical mathematics when we study the characterization of normal distribution in Hilbert Space and number theory?
- What are the prominent properties of Hilbert space that fit Normal distribution?

1.3 Objectives of the study

The objectives of the present study are:

- To explore the characterization of Normal distribution in Hilbert space.
- To investigate the relationship between statistical principles and pure mathematical properties.
- To assess if the Statistical Section of Mathematics at University of Nairobi has affinity with specific schools of probability and statistics in historio-philosophic context, hence, to give a tentative conclusion.

1.4 Brief literature review

The present literature review examines (i) the efforts that have been made over time to explore the theory of characterization of normal distribution, and (ii) the characterization of normal distribution in Hilbert space, as well as (iii) the statistical research works done at University of Nairobi.

The characterization of a distribution is the investigation of those unique properties enjoyed by that distribution. Mathai and Pederzoli (1977) have compiled and put together their studies with recent research papers and published then in a form of a monograph. In their monograph entitled "*Characterization of the Normal Probability Law*" (1977), they deal thoroughly with the highly mathematical topic of characterization and try to motivate students to undertake research work in this area. Thus the material is developed from the very elementary level to the research level.

There are properties that will uniquely determine a normal distribution, that is to say, the normal law is the only distribution to enjoy such properties. Investigation of such properties and the determination of the resulting distributions are known as *characterizations of distributions*.

There are two distinct methods developed one following the other: (i) the *functional equation method*, and (ii) the *axiomatic approach*.

The *functional equation method* is developed to its present format by Kagan, Linnik and Rao (1973). In their method they developed techniques in characterization problems as follows: (a) Use the properties and derive a functional equation. Then solve the functional equation for a unique solution by imposing additional conditions if necessary. (b) Use the properties to derive a difference, differential or difference-differential equation, and then seek a unique solution for it. (c) Use the properties and analyze some general structures to classify or separate certain distributions. For more details see Mathai and Pederzoli(1977).

The *axiomatic approach* is advanced by Mathai and Rathie (1975). The axiomatic approach to characterization of normal law proceeds as following: An axiomatic definition is provided for a basic concept itself such as variance, correlation, entropy, affinity, information and the like. In problems of this nature a few postulates are put forward and the resulting concepts are uniquely determined, thus providing axiomatic definitions for these measures. The main techniques used in the characterization of basic concepts are also the same as the techniques used in the characterizations of probability laws. Mathai and Pederzoli(1977) remark that historically, the basic concepts are introduced mainly on the basis of intuitive or heuristic considerations.

As a result of the above methods, normal distribution is characterized through *linear independence* [Darmois,1951; Basu,1951; Linnik,1952; Skitovich,1954], *linear and quadratic independence*[Laha,1956,1957;Chanda,1955; Linnik,1956; Gordon,1968; Gordon and Mathai, 1972 Mathai,1977], *regression properties*[Laha and Lukacs,1960; Zinger and Linnik,1964; Mathai, 1967; Gordon,1968; Gordon and Mathai, 1972], *by solutions of certain functional equations* [Rao, 1967; Linnik, 1960; Zinger and Linnik, 1955], *from the Student's law* [Mauldon, 1956; Kotlarski, 1966], *structural set-up* [Mathai, 1967; Patil and Seshadri, 1963,1964], maximization principle[Mathai, 1977] and other *miscellaneous techniques* [Mathai and Gordon, 1972].

The Characterization of Normal Distribution in Hilbert Space was initiated by Prohorov and Fisz (1957). In their article they came up with a theorem of random elements in Hilbert space, and the theorem is stated as follows:

Consider random element with values in a real separable Hilbert space H [that is measurable mapping $\xi(u)$ of a fundamental probability field into the space H]. Let the probability distribution and characteristic function be denoted respectively as.

P^{ξ} and $\phi(f,\xi) = \int_H e^{i(f,\xi)} dP^{\xi}$, $f \in H$.

Let $P^{\xi} = P^{\eta}$ be denoted as $\xi \sim \eta$ and let $\|\xi\|$ be the linear functional (f,ξ), $f \in H$, stochastic variables. The mathematical expectation of the random element ξ is such an element $M\xi \in H$ such that for every $f \in H$, $M(f,\xi) = (f, M\xi)$. Consider the conditions, (α): $M\|\xi\|^2 < \infty$; (β): $M(f,\xi)^2 > 0$ for any $f \in H$, $f \neq \theta$; (γ): $M\xi = \theta$ where θ is the null element in H, and let $\xi^{(1)} \sim \xi^{(2)} \sim \xi^{(3)}$, be random element in H, subject to (α), (β) and (γ) and let (δ) $\xi^{(1)}$ and $\xi^{(2)}$ be independent and (ε) $\xi^{(1)} + \xi^{(2)} = \Lambda \xi^{(3)}$, where Λ is a (Λ1) linear, (Λ2) bounded, (Λ3) self conjugate, (Λ4) positive operator in H. Then the distribution P of each of the random elements $\xi^{(i)}$ is normal and $A = \sqrt{2}E$ where E is a unitary operator. Additional explanation of Characterization of Normal Distribution will be given in chapter four.

Then, Eaton and Pathak (1969) picked up the topic and studied it more comprehensively. Furthermore, Pathak(1970) made another study on this topic and gave additional results. Parallel to these research studies, we will continue to ponder upon the characterization of normal law in Hilbert space, and present some propositions which can enable us to give an analogy of inner product.

Normal Law in Number Theory is advanced mainly by Kac in 1949 in his article "Probability methods in some problems of analysis and number theory," and in 1959 in his monograph entitled *Statistical Independence in Probability, Analysis and Number Theory*. Other prominent scholars who laboured on this research area are like Borel(1909), Hardy and Ramanujan(1917), Champernowne(1933), Kac and Erdös(1939), Rényi(1955), Kubilus(1956), and Rényi and Turán(1958).

The history of probability theory and mathematical statistics has been studied, though not in a thorough manner, by prominent scholars, namely, mathematical historians. Among the authoritative works on the history of probability and statistics the following are prominent, in English: Isaac Todhunter, *A History of Mathematical Theory of Probability from the Time of Pascal to that of Laplace*, 1865,1965, covers 1654-1812; C. C. Heyde and E. Seneta, *Studies in the History of Probability and statistics*, Vol. 2,, 1975; D. B. Owen (ed.), *On the History of Statistics and Probability*, 1976, prepared by selected famous American scholars; L. E. Maistrov, *Probability Theory: A Historical Sketch*, 1974 [Translated by Samuel Kotz], well documented but based on an economic theory; Ian Hacking, *The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference*, 1975, philosophical analysis of the development of probabilistic areas in mathematics; J. Koren (ed.), *The History of Statistics*, 1918, 1970, contains the development and progress of Statistics in France; E. S. Pearson and M. G. Kendall (eds.), *Studies in the History of Statistics and Probability*, 1970.

In line with this observation, we can give a brief review about historical note and discernment on the original statistical work done at University of Nairobi. The pioneers or the initiators of the Statistical Section of Mathematics Department at University of Nairobi are Professor M. S. Patel and his colleagues, in late 1960's.

Basing our observation on the articles published as concrete documents and dissertations as further confirmative works, we note that the *model techniques or design theory* has gained ground at University of Nairobi. Furthermore, categorizing broadly, it can be said that four research groups are emerging: *group screening, educational and manpower planning, biological population modelling,* and *AIDS modelling* (Epidemiological modelling).

The group screening technique was developed during World War II by Dorfman (1943) and studies by Sterret(1957). It was improved by introducing more than two stages by Sobel and Groll(1959), Watson (1961), Li(1962), Patel(1962), and Finucan(1964).

The first and second published works, as outcome of research works at University of Nairobi, are "Two stage group screening designs with unequal a-priori probabilities," and "Optimum two-stage group screening designs," by M. S. Patel and J. A. M. Ottieno, both publications occuring in 1984. The third one is "Three-stage group screening with error in observations," by J. W. Odhiambo and M. S. Patel in 1985.

The subsequent published articles followed the same pattern, except a few, namely, stressing on "group screening": Odhiambo and Owino(1985), "A stochastic model for estimating academic survival in an education system"; Patel and M. Ottieno(1985), "Optimum Two-stage group-screening designs with unequal a-priori probabilities and with error in decisions," paper presented at the joint statistical meetings of A.S.A., E.N.A.R., WNAR, IMS at Las Vegas, USA 5th - 8th August, 1985; Odhiambo and Patel(1986), "On multiple group screening designs"; Odhiambo(1986), "The performance of multistage group screening designs"; Odhiambo and Khogali(1986), "A transition model for estimating academic survival through cohort analysis"; Patel and

Ottieno(1987), "Optimum Two-stage group-screening with unequal group sizes and errors in decisions"; Odhiambo and Patel(1987), "Three-stage group screening with unequal group sizes and with errors in observations"; Patel and Manene(1987), "Step-wise group screening with equal prior probabilities and no errors in observations"; Odhiambo, and Manene(1987), "Step-wise group screening designs with errors in observations"; Owino and Philips(1988), "A comparison of retention properties of the Kenya primary education system before and after 1970"; Adhikary and Chaudhuri(1989), "A note on handling linear randomized response,"; Adhikary and Chaudhuri(1989), "On Two properties of an unequal probability sampling scheme"; Adhikary and Chaudhuri(1990), "A note on interpreting subsamples of unequal sizes drawn with and without replacement"; Adhikary(1991), "On the performance of the nearest proportional to size sampling design".

Recently in the conference of the Kenya Mathematical Society further review of the works done at University of Nairobi are exposed again: Odhiambo(1993) presented a paper on "A review of the factor screening method"; Ottieno(1993)presented a statistical analysis on "Mortality levels and determinants in Kenya"; Owino(1993) presented a paper on "A mathematical model for comparison of educational characteristics of males and females"; Weke(1993) presented a paper on "IBNR claims reserving and GLIM".

Other more recent publication, also are of great help to discern more closely into the research progress at the University: Owino and Odhiambo(1994), "A statistical method for planning an educational system"; Getao and Odhiambo(1996), "The potential of information technology in the management of an African crisis: Computers and AIDS". In particular the *Fifth Scientific Conference of the East Central & Southern Africa Network of the International Biometric Society*, 22nd -25th September 1997 held in Mombasa-Kenya shows a great achievement in the research work done at University of Nairobi: Gachii and Odhiambo(1997), "Deletion designs in estimation of low order interactions"; Luboobi and Simwa(1997), "HIV/AIDS epidemic curves for Kenya and

Uganda: A parametric statistical approach"; Luboobi and Simwa(1997), "HIV/AIDS epidemic curves for Kenya and Uganda: A nonparametric statistical approach"; M. M. Manene(1997), "On two-type stepwise group screening designs"; Mwambi, Odhiambo and Duchateau(1997), " A multiple matrix model to study the population dynamics of R. appendiculatus in Zimbabwe"; Odhiambo and Getao(1997), "The potential of group screening method in the management of AIDS crisis in Africa"; Owino and Omolo(1997), "Optimal harvesting in poultry farming".

Thus, focusing our analysis on these published articles, research projects and dissertations, we will asses the original statistical work done at University of Nairobi, and come up with a tentative conclusions.

1.5 Significance of the study

Besides the written reference, as indicated in the literature review, formal and informal discussions with the prominent scholars in the Departments of Mathematics at the University of Nairobi have been held. The actual protagonists on the research area of probability and statistics are the main source of discernment on the historical and philosophical contextualization of statistical and probabilistic research fields at the University of Nairobi.

The result of this study will give rise to an interest in three research areas, namely the philosophical assessment on the foundation of mathematical statistics and probability, discernment on normal law on number theory and the characterization of the theory of normal distribution in Hilbert space.

The historio-philosophical outlook on mathematical probability and statistics can open a real interest on philosophy of randomness, inductive reasoning, inductive behaviour and other related topics. In other words, one can venture into the deep knowledge of philosophy of logic and language of mathematical probability and statistics in diversified perspective. For instance a statement of research problem can be stated as what is *Nairobi Studies on Philosophy of Mathematical Probability and Statistics* ?

The relationship between the properties of normal distribution and number theory and the characterization of normal distribution in Hilbert space is amazing. Since in probability theory we study mathematics, statistical data or measurements, theory of nature and theory of knowledge itself, we can say that normal law, a phrase preferred by pure mathematicians, is either the law of nature or a mathematical theorem. Thus, many properties and fascinating relationships between normal law and number theory, and the characterization of normal distribution in Hilbert space can be re-discovered

Chapter Two

HISTORICAL NOTE ON PROBABILITY THEORY AND STATISTICS

2.1 INTRODUCTION

The prehistory of probability theory and the origins of probability theory as a science is not thoroughly investigated. Nonetheless, the basic stimuli in the rise of probability theory can be stated as: processing numerical data and results of observations in various sciences, the practical requirements of insurance companies and abstract problems connected with games of chance, calculating the number of various possible outcomes in throwing several dice. The origins of probability theory was generally attributed to investigations by the renowned French mathematician Fermat(1601-1665) of problems posed by a gambling contemporary to Pascal(1623-1662). Now this conviction have been pushed back a century earlier to the Italian mathematicians Cardono(1501-76) and Tartaglia(1499-1557) about 1570. Cardano (about1526) is considered to be the author of the first book on probability theory. He is also the author of remarkable text of algebra *Ars Magna*(1545), whereby causing disputes as to the real authorship of certain phases of mathematical development. *Ars Magna* contains the works of Scipione del Ferro(about 1500), Tartaglia and Lodovico Ferrari(1522-1565).

According to F. N. David(1955), a probability theory historian, with regard to the beginner of modern probability theory, we have the following observation:

> I do not think that the fact that Cardano did not quite see the mathematical abstraction clearly can detract from the fact that he did, on paper at any rate, as far as we know, calculate the first probability by theoretical argument, and in so doing he is the real begettor[sic] of modern probability theory. ... It is true that Galileo wrote on one problem only and fairly briefly at that, but it is difficult to see why Pascal and Fermat should be preferred as the originators of probability theory before Galileo or Cardano.[David, 11-13]

2.2 Important Probabilistic and Statistical Historical Periods

Using the criteria of (i) time- chronological aspect, (ii) the method of approach to statistics and probability, and (iii) great influence or impact on the study of statistics and probability, we can say that the historical development of Probability and Statistics is classified in the following periods: (1) The completion of the emergence of Probability with Bernoulli (1713), (2) The initiation and development of Statistical and Probabilistic Theory with Laplace (1812), Gauss (1816), Poisson (1837) and Chebyshev (1846) in the nineteenth century, (3) The Advent of prominent twentieth century probabilistic philosophers and mathematical statisticians, and Modern Schools of Probability and Statistics.

2.2.1 The completion of the emergence of Probability with Bernoulli (1713),

2.2.1.1 Etymological meaning of Probability

Of the precise meaning of probability there are conflicting views among experts, philosophers, mathematicians, statisticians. As Copi and Cohen(1990), and von Wright(1977) observed the reason for this may be partly grasped from a survey of the various channels through which a scientific concept of probability has emerged.

Etymologically the Latin word *Probabilis* (probable), is a term applied generally to any belief that is reasonable without being certain. The vagueness of this formulation permits a wide variety of uses, especially in modern times. The first broad distinction that must be made is that between *the degree of credibility* attaching to a proposition and *mathematical probability* which belongs to prepositional functions or related propositions. The first of these is the common sense view - the maxim, "one who sleeps intentionally, can not be awakened by an elephant," is accepted by many people. The second is a well defined part of general mathematical theory. Durbin (1967) says that the basic distinction with respect to mathematical probability is between the abstract theory and its applications and interpretations. The latter have the same general subdivision: 1) *Classical approch* (equally probable cases), 2)*Empirical or statistical* (repeatable events- empirical law of chance), 3)*The Assymptotic approach* (limiting value of the frequency), 4)*Logical* (degree of evidence theory) or *Subjectivist or Personal* (degree of belief theory). This classification and the notion of method of approach will be dealt with in detail in connection to the school of statistics.

Though these broad categories can represent irreducible differences in individual interpretations, it seems difficult to do without this basic division. Furthermore, there is the notion of probability that was common to the Greeks and medieval scholastics; for them probability meant any argument that gave rise merely to *opinion* and not to *demonstrative certitude*. Finally, there is the notion of inductive probability - the possible utilisation of mathematical probability for induction in science, either as its justification or as an aid in developing a logic of discovery.

In depicting the historical note on probability, statistics and normal distribution, it is hard to come up with a complete picture in few pages. Nonetheless, the intention of the present work is not just for the sake of history, but primarily, for the verification and identification of *schools of statistics* and to illuminate the role of *normal distribution* in statistical mathematics, and to classify the University of Nairobi, if it is possible, as a school with respect to the historical development of schools of statistics and probability.

2.2.1.2 Renaissance

In the period of Renaissance we note the presence of mathematics of randomness. About 1660 probability, as we know now it, started to emerge with dual factors: as stated earlier, on the degree of belief and with devices tending to produce stable long-run frequencies.

Commercial insurance against risks was developed in the Italian cities of the early Renaissance. The theoretical foundations of life insurance were laid in the 17th century. The English statistician John Graunt in 1662 drew attention to the stability of statistical series obtained from registers of deaths. Soon after, the English astronomer Edmund Halley showed how to calculate annuities from mortality tables.

chances but able to apprehend the fact that evidence and causation are in different categories, could perfectly well start measuring *epistemic probability*. He developed combinatorial methods primarily for the purpose of logical deductions, which is closely connected with investigations aimed at the construction of the "universal Characteristic." Todhunter(1865), in his monumental work on *History of the Mathematical theory of Probability*, points out that the mathematical treatment of the subject of combinations as given by Leibniz(1880) is far inferior to that given by Pascal 1623-1662 (1963). In the first printed textbook of probability, Christiaan Huygens (1629-1695), *Calculating in Games of Chance*(De ratiociniis in ludo aleae - 1657), we come across an important term expectatio (expectation); his perception of his work is also interesting: "I would like to believe that in considering these matters closely, the reader will observe that we are dealing not only with games but rather with the foundations of a new theory, both deep and interesting" [1695]; this is a foreword, a letter from Huygens to van Schooten, dated 27 April 1657. Huygens' work is published as an appendix in Latin to a volume entitled *Exercitationes Mathematicae* (Mathematical Studies) by Francis van Schooten, which appeared in 1657.

2.2.1.3 Political Arithmetic

Statistics began as the systematic study of quantitative facts about the state. John Graunt -1662 tells us in the preface to his *Natural and Political Observations* upon selfsame bills. He and William Petty - whose various essays on "political Arithmetic" make him the founder of economics - seem to have been the first people to make good use of these *population statistics*. Petty was a man who wanted to put statistics to the service of the state and saw real importance of collecting statistics for testing a wide range of hypotheses even the ones about medical efficacy.

2.2.1.4 Annuities

Annuity as opposed to loans, was a secured income for an assigned period, where a standard way to raise public money, partly because it was possible for a government to chances but able to apprehend the fact that evidence and causation are in different categories, could perfectly well start measuring *epistemic probability*. He developed combinatorial methods primarily for the purpose of logical deductions, which is closely connected with investigations aimed at the construction of the "universal Characteristic." Todhunter(1865), in his monumental work on *History of the Mathematical theory of Probability*, points out that the mathematical treatment of the subject of combinations as given by Leibniz(1880) is far inferior to that given by Pascal 1623-1662 (1963). In the first printed textbook of probability, Christiaan Huygens (1629-1695), *Calculating in Games of Chance*(De ratiociniis in ludo aleae - 1657), we come across an important term expectatio (expectation); his perception of his work is also interesting: "I would like to believe that in considering these matters closely, the reader will observe that we are dealing not only with games but rather with the foundations of a new theory, both deep and interesting" [1695]; this is a foreword, a letter from Huygens to van Schooten, dated 27 April 1657. Huygens' work is published as an appendix in Latin to a volume entitled *Exercitationes Mathematicae* (Mathematical Studies) by Francis van Schooten, which appeared in 1657.

2.2.1.3 Political Arithmetic

Statistics began as the systematic study of quantitative facts about the state. John Graunt -1662 tells us in the preface to his *Natural and Political Observations* upon selfsame bills. He and William Petty - whose various essays on "political Arithmetic" make him the founder of economics - seem to have been the first people to make good use of these *population statistics*. Petty was a man who wanted to put statistics to the service of the state and saw real importance of collecting statistics for testing a wide range of hypotheses even the ones about medical efficacy.

2.2.1.4 Annuities

Annuity as opposed to loans, was a secured income for an assigned period, where a standard way to raise public money, partly because it was possible for a government to mathematician of the Renaissance. The subject was developed into a "geometry of the die" (*Aleae geometria*) by Blaise Pascal, Pierre de Fermat, and Christiaan Huygens in the 17th century. Fermat treated the problems within a general theory of combinations, which was further developed by the Swiss mathematician Jakob Bernoulli. The latter can be regarded as *the founder of probability theory* as a branch of mathematics: his posthumously published Ars Conjectandi of 1713 can be said to aim at a fusion of the a priori methods of combinatoric probability and the a posteriori methods of early statistical theory.

2.2.1.6 The art of Conjecturing - Ars conjectandi

Jacques Bernoulli's *Ars conjectandi* presents the most decisive conceptual innovations in the early history of probability. The author died in 1705. He had been writing the book off and on for twenty years. Although the chief theorem was proved in 1692, he was never satisfied and he never published. **Ian Hacking(1975)** in his book entitled *The Emergence of Probability* sums up the historical fact, probabilistic and philosophical importance as follows:

> The work was finally given to the printer by his nephew Nicholas, and appeared in Basle in 1713. In,that year probability came before the public with a brilliant portent of all the things we know about it now: its mathematical profundity, its invitation for philosophizing. Probability had fully emerged. [Hacking, 143].

The *Ars conjectandi* comes in **four parts**; the **first** is an improved version of Huygen's book on games of chance. The **second** is a general essay on the theory of combinations. The **third** application of the theory of combination to a sequence of further exercises on games of chance. The **fourth** part of the book revolutionazed the probability theory: for the first time a subjective conception of probability is explicitly avowed and the first limit theorem is proved. In this part it is intended to show the application of probability mathematics to matters of economics, morality and politics. It is this part that justifies the very tittle, *Ars conjectandi*(The Art of Conjecturing).

Furthermore, on this part Bernoulli announces that 'Probability is degree of certainty and differs from absolute certainty as the part differs from the whole.' Further detailed analysis is given by Hacking(1971) in "Jacques Bernoulli's Art of Conjecturing."

The first mathematical contribution of the work is the formalization of the *first limit theorem* in probability. Secondly, he has been regarded as father of the first subjective conception of probability or Bernoulli became father of different schools: *frequentist* (Richard von Mises - 1951), *inductivist* (Rudolf Carnap - P. M. Boudot 1967), *inference via confidence intervals* (Jerzy Neyman 1957, Dempster 1966).

Bernoulli introduced the term subjective into probability and through the subsequent centuries, and in particular in the twentieth century it gave rise to different meaning and schools: (i) the most extreme subjectivism- personalism, probabilities are unknown - Bruno de Finetti, L. J. Savage; (ii) the theories of logical or inductive probability, which can not be detached from evidence, one may fail to know only through failure to do probabilistic logic - J. M. Keynes and others; (iii) the concept of subjectivity prominent among current philosophers of quantum physics.

2.2.1.7 The first limit theorem

The first limit theorem known as "the weak law of large numbers" is a theorem of pure probability theory, and holds under any interpretation of calculus.

Bayes'(1763) paper published half a century after the appearance of the *Ars conjectandi* is the first systematic attempt to compute values for conditional probability (p is in $s_n \pm \varepsilon/s_n$)

The brief description of the historical note on the emergence of probability comes to an end with the publication of *Ars Conjectandi*. Meanwhile, Abraham de Moivre [1667-1754] published *De mensura sortis* (1711), which soon was to culminate in *The Doctrine of Chances, or a Method of Calculating the Probability of Events in Play*,(1718,1738,1756)where the mathematics of probability was recognized as an independent discipline in its own right, and although not in clear words, it is claimed that a statistical law is the course of nature. Furthermore, a curious pre-Bernoullian paper of 1710, by John Arbuthnot(1667-1735), is said to be the first published test of significance of a statistical hypothesis and contributed to the work of a group of men who endeavoured to relate Newtonian science to natural religion, like John Wilkins. In brief the remarks of Karl Pearson(1857-1936) can fittingly express the historical fact of this period:

> Newton's idea of an Omnipresent deity, who maintains mean statistical values, formed the foundation of statistical development through Derham, Susmilch, Niewentyt, price to Quetlet and Florence Nightingale ... De Moivre expanded the Newtonian theology and directed statistics into the new channel down which it flowed for nearly a century. The causes which led De Moivre to his 'Approximatio' or Bayes to his theorem were more theological and sociological than purely mathematical, and until one recognizes that the post-Newtonian English mathematicians were more influenced by Newton's theology than by his mathematics, the history of science in the 18th century - in particular that of the scientists who were members of the Royal Society - must remain obscure.[1926, 551-2]

David Hume(1711-1776), in his work *A Treatise of Human Nature*, being an attempt to introduce the experimental method of reasoning into moral subjects, in 1739 poses for sceptical problem about the future, the problem of induction. He doubts that any known facts about past objects or events give any reason for beliefs about future objects or events. A similar problem arises also for inference about unremembered past events, and unobserved present ones. This basic sceptical problem is expressed and negated as follows: "An expectation that the future will be like the past must be either knowledge or opinion. But all reasoning concerning the future must be based on causes and effects. Reasoning concerning causes and effect is not knowledge. Therefore it must be opinion, or probability. But all probable reasoning is founded on the supposition that the future will resemble the past, so opinion cannot be justified without circularity. Knowledge and probability are exhaustive alternatives. Hence expectation about the future is unjustified." Probability emerged from the Renaissance transformation in opinio. Although the emergence of probability is a transformation in opinion, the emergence of 'probability-and-induction' is a more complete event depending on parallel transformations in high science and low science.

It is good to note that by a widening of the aim of decisions, obtained in 1738 by Daniel Bernoulli(1700-1782), in 1738 through the introduction, under the name of *moral expectation*(or *utility*) and a widening of the domain of probabilities, initiated by James Bernoulli(1713) and more profoundly by Bayes(1763) by relating them to statistical observation "inductive reasoning" is established.

2.2.2 The initiation and development of Statistical and Probablistic Theory with Laplace (1812), Gauss(1816), Poisson (1837) and Chebyshev (1846) in the nineteenth century

Some of the basic concepts of statistical theory were initiated during the first quarter of the nineteenth century by Laplace in his fundamental *Théorie Analytique des Probabilités* (Paris,1812), and by Gauss in his papers on the method of least squares, and his monumental work *Theoria Motus corporum coelestium in sectionibus conicis solem ambientium*, (Hamburg,1809).

2.2.2.1 Pierre-Simon de Laplace

Pierre-Simon de Laplace (1749-1827) introduces the purely subjective criterion of equal possibility of events, considering that two events are equally possible if there is no reason to believe one of them will occur rather than the other. Assuming that our knowledge is incomplete concerning many objects and events, Laplace proposes applying probability theory to all problems of the natural sciences and society, such as moral sciences. Laplace believed that phenomena and the actual nature of things do not coincide and the purpose of science is to correct the illusions and deceptions of our senses, by perceiving true objects in their deceptive appearance and manifestations. Nature should be approached by comparing various factors; the phenomena should be examined from various points of view in their development; a collection of facts is not sufficient; one should compare and experiment.

His first papers in probability theory are dated at the middle seventies of the eighteenth century: the first one being *Mémoire sur la Probabilité des causes par les évènements* in 1774. In 1810 Laplace obtained his most important result in probability theory, presently known as *Laplace's theorem*. The essence of this theorem is that the binomial probability distribution under suitable normalization and unlimited increase in the number of trials approaches the *normal probability distribution*. After this Laplace published his classical treatise "Théorie analytique des Probabilités" in 1812. In this volume Laplace presented all his basic results in probability theory; laid the foundations for study of various statistical regularities, successfully applied probability theory to estimations of errors in observations and so on.

Lehmann(1958,1959) suggest that Laplace was the first to produce a general solution to what sometimes is called *Bernoulli's problem*. Pierre-Simon de Laplace(1812),*Théorie Analytique des Probabilités*, appears to compute estimators F_δ with property that

$$\text{Probability } \{ p \text{ is in } F_\delta(s_n) \} = 1 - \delta, \quad \text{for all } p \text{ in } [0,1]; \tag{2.1}$$

This is the universal quantification of an expression like probability (p is in $s_n \pm \varepsilon/p$). Furthermore, it is expressing that regardless of the true value of p, the probability of making a right estimate is 1-δ. This is the exact security level. But there is a suggestion that (2.1) is not obtained by Laplace. The reason is that at one moment he has to substitute the observed s_n for the unknown p, and hence the solution is only asymptotically correct. Perhaps W. S. Gosset's famous statistic 't' (1908) was the first device to overcome this kind of inexactness. However, if we ignore this kind of inexactness we can regard the theory of 'probable errors' produced by Gauss in 1816 as using interval estimators with a security level of 0.5.

Laplace in his classical work solved several problems. One of them is to "find the best combination of observations for the determination of an unknown quantity under the condition that positive and negative errors are equally probable and the number of observations is indefinitely large." Without any assumptions on the distribution law of the errors. Laplace obtains that the method of least squares yields the best possible combination of observations. He also gives a new proof for James Bernoulli's theorem. He obtains the asymptotic formula for the probability of the sum of independent random variables each one of which admits only all the integer values between -a and +a. In his derivation he actually employs the basic ideas of the *theory of characteristic functions*.

Furthermore. Laplace in *Oeuvres completes*, VII(1878-1912), investigated the problem of credibility of estimates made after observations, and presented a straight forward Bayesian analysis. Although Laplace found an estimator which at least asymptotically has an exact security level. it is not unique. Hence other desires are required to choose among the estimators of given security level. The best known solution is due to Jerzy Neyman(1957). His theory explains that in many interesting situations there exists a unique estimator of given security level that for every false value of *p* minimizes the chance of including p in the interval estimate. An interval estimate got from such an estimator is a *confidence interval*. Neyman advocates *inductive behaviour* rather than *inductive inference*. We can behave in a way that is usually right, but we cannot measure the credibility of our doing the right thing on any individual occasion. According to Hacking(1975) t*his is one of the chief bones of contention in contemporary philosophy of statistics*. The Bayesian school, for example, has quite the opposite opinion. The logic of the *confidence interval* approach was made clear in the 1930s by J. Neyman and E. S. Pearson.

One of the applications of probability theory of special interest to Laplace is in the field of *demography*. He discusses methods of indirect population counts, and estimates of precision in such counts: he also develops the theory of sampling census and other problems. This result is supported by his study conducted in 1802 about a sampling population census of France. Thus he played a significant role in the development of statistics, in particular he contributed greatly to the application of probability theory to demography.

The classical definition of probability is given in his book "Théorie analytique des Probabilités, 1812": *the probability P(A) of event A is equal to the ratio of the number of possible outcomes of a trial which are favourable to event A to the number of all possible outcomes of the trial.* Here equi-probability is assumed.

If we consider the expansion $(x + x^2 + x^3 + x^4 + x^5 + x^6)^n$, then the value of the coefficient of x^s is equal to the number of outcomes with *n* dice, giving the sum of points equal to *s*. Laplace generalizes this method of calculation to the method of generating functions widely used at present. A function $f(t) = \Sigma^{\infty}_{n=0} f_n t^n = f_0 + f_1 t + f_2 t^2 + ... + f_n t^n + ...$ is called the *generating function* of the sequence $f_0, f_1, f_2, ..., f_n, ...$. Generating functions are used not only in probability theory, but also in algebra and other branches of mathematics.

2.2.2.2 Carl Friedrich Gauss

However, if we ignore Laplace for his inexactness, we can regard the theory of 'probable errors' produced by Gauss in 1816 as using interval estimators with a security level of 0.5. Carl Friedrich Gauss (1777-1855), a German scientist and mathematician, first work related to probability theory was the famous *Theoria Motus corporum coelestium in sectionibus conicis solem ambientium*, (Hamburg,1809). In the last part of this work Gauss for the first time presented his theory of errors in observations. Two other papers are related to this topic: "Disquisitio de elementis ellipticis Palladis," *Comment. (Göttingen)* **I** (1808 -1811) and "Bestimmung der Genauigkeit der Beobachtungen" *(Z. für Astron.* (1816), pp. 185-197). These works were generalized and supplemented in his treatise "Theoria combinationis observationum erroribus minimis obnoxiae." *Comment. (Göttingen)* **V** (1819 -1822), which appeared in 1823. In 1828 a supplement of this work: "Supplementum theoriae combinationis observationum erroribus minimis obnoxiae." *Comment. (Göttingen)* **VI** (1823 -1827) was published. In 1845-1851 Gauss wrote "Application of probability theory for determination of the balances of widows' pension funds" and also computed "Tables for determination of the time periods for various types of obligatory incomes for survivors." Gauss's notes and letters are also of great value and interest from the point of view of probability theory.

Adrien-Marie Legendre(1806) in his treatise "New methods for the determination of the orbits of comets" developed the method of least squares in the appendix entitled "On the method of least squares". He writes "Among all the principles, which may be suggested for this purpose, there is none simpler than the one we utilized in the previous discussion: the method is to minimize the sum of squares of the errors." He formulates this principle in a clear manner and observes that it should be very useful in various problems of physics and astronomy, where the derivation of the most precise results possible from observations is required.

Gauss presented his method for the first time in 1809. However, he observes that "Our principle, which we have made use of since the year 1795, has lately been published by Legendre in the work *Nouvelles méthodes pour la détermination des orbites des comètes*, Paris, 1806". He also gives this date in his letter to Laplace of 30 January 1812. In a letter to H. M. W. Olberns in 1802 he mentions, however, that "starting from 1794, I have been utilizing the method ... which has also been applied in Legendre's work ... ." Gauss mentions two dates, 1794 and 1795. Contemporary authorities are inclined to accept 1794 as the correct one. For more details on this controversy between Legendre and Gauss we can see the work of Bell (1937), "Men of Mathematics," *Simon and Schuster*.

The normal distribution was considered a *universal law* for a long period of time. This state of affairs resulted in a delay in the development of quantitative methods for discarding some observations, since the normal law admits the possibility of errors of any magnitude. Hence it was assumed that all the observations should be retained. Only in the middle of the nineteenth century did the first probabilistic criteria for rejecting observations begin to appear. Gauss in 1816 investigated the estimation of h based on results of observations [where the normal law is given by

$$\varphi(\Lambda) = h\,(\pi)^{-½}\,exp[-h^2\Lambda^2]. \qquad (2.2)$$

He introduces the error function

$$\theta(t) = 2(\pi)^{-1/2} \int_0 exp\,(-t^2)dt \tag{2.3}$$

and presents a small table of the values of this function. The value of the argument r such that $\theta(\rho) = 0.5$ is specially singled out. This value is $\rho = 0.4769363$. The quantity $\rho/h = r$ is called *the probable error for the function $\theta(ht)$*. The first table of a normal curve appeared in D. Bernoulli's treatise of 1771.

The most complete exposition of the *theory of errors* is contained in Gauss's paper "theoria combinationis observationum erroribus minimis obnoxiae." He assumes that there are two types of errors, namely, *random errors* and *prediction errors*. His results in the theory of errors are presented almost without modification even nowadays in many textbooks on statistics. It is worthy to note that Gauss contributed to the development of probability theory as well as to different branches of mathematics and science.

2.2.2.3 Simeon Denis Poisson

During this period we come across another prominent mathematician Simeon Denis Poisson [1781-1840]. His works are many with respect to the probability theory: *On the probability of mean results of observations, 1827; Continuation of the memoir on mean results of observations, 1832.; Sur l'avantage du Banquier au jeu de Pharaon* [1832? 1837?]; *On the probability of births of boys and girls*, and several others. All these papers were included in various forms in Poisson's main work on probability theory. " Recherches sur la probabilité des jugements en matière criminelle et en matière civile" published in 1837. His celebrated theorem is also contained in this volume. In his book he first presents a brief survey of previous results in probability theory, and in particular Laplace's and Condorcet's contributions on moral probability. He himself also believes that the analytic theory of probabilities is applicable to the evaluation of the correctness of court decisions. For this purpose he deduced "the law of large numbers", which is different from Bernoulli's theorem. In mathematical notation it is formulated as

$$\lim_{n \to \infty} \left(\left| \frac{m}{n} - \bar{p} \right| < \varepsilon \right) = 1, \tag{2.4}$$

where *n* is number of independent trials, *p* arithmetic mean, *m/n* relative frequency of the occurrence of event A. If the probability of the occurrences of events remains constant from trial to trial, then $\bar{p} = p$, and the Poisson theorem in this case reduces to Bernoulli's theorem.

For Poisson, all events of a moral as well as of a physical nature are subject to this universal law. He viewed this theorem not only as mathematical fact, but also as a philosophical truism. It served as ground for his investigations concerning the correctness of court decisions and phenomena of a moral nature. He believes that by means of this principle the probability of any human decision may be determined regardless of the reasons for these decisions.

In his book he also derived the so-called "law of small numbers". As the deviation of the value of *p* from the value ½ increases, the asymptotic representation of $P_{m,n}$ in the form $(2\pi)^{-1/2} \exp(-x^2/2)$ becomes less and less accurate. In order for Laplace's theorem to give a reasonably accurate approximation to $P_{m,n}$, the number of observations must be substantially increased, which is not always convenient or even possible. The problem arises of obtaining an asymptotic formula which will be particularly suitable for small *p*. This problem was solved by Poisson. He obtains that, as $p_n \to 0$ with $n \to \infty$, the probability that an event will occur *m* times approaches

$$P_{m,n} = (\lambda^m e^{-\lambda}) / m!, \quad \text{where } \lambda = np_n. \tag{2.5}$$

This formula of Poisson can be utilized as an approximating expression for $P_{m,n}$ for a fixed but small *p* and large *n*. David(1955) and others credit this discovery of "Poisson's binomial exponential limit" to de Moivre(1718).

The Polish statistician L. Bortkiewicz (1868-1931) renamed the Poisson distribution *the law of small numbers*. He also applied this distribution to rare events, such as deaths by horse-kick in the Prussian Army, births of triplets, and so on.

Commenting on the state of probability theory at that period of time, Gnedenko(1948) writes that:

> In spite of the fact that Laplace and Poisson concluded an important and fruitful initial period in the development of probability theory, a period of philosophical cementation of the basis of this science, this period resulted in an indifferent attitude toward probability theory in the West and in a definite rejection of the possibilities of utilizing its methods in studying natural phenomena. This led to the beginning of a long period of stagnation in probability in the West. [Gnedenko, 394; Maistrov, 160]

In this connection the scientific works of a famous Belgian statistician Adolphe Quetelet (1794-1874) are indicative. Quetelet(1842), who attended Laplace's lectures[in 1823-4], proposed that the rules of probability theory are those that govern and direct the activities of human society. The degrees of inclination to crime, marriage, etc., are according to Quetelet, nothing but mathematical probabilities. He describes the average man as everlasting and invariable, the absolute perfect type, while separate individuals are a distorted representation of this type.

2.2.2.4 Probability theory in Russia and the St. Petersburg's school

Next it is worthwhile to pose for a moment, and ponder through the "Probability theory in Russia and the St. Petersburg's school".

The teaching of probability in Russia started in 1829-30, by Revkovskiĭ, at Vilnus University, and somehow he follows the footsteps of Bernoulli and Laplace, and favoured and recommended in 1830 by M. V. Ostrogradskiĭ (1801-1862). In Moscow University the first courses in probability theory were given starting from 1850 by A. Yu. Davidov (1823-1885). Davidov published several papers on this subject in the years 1854-7; "An application of probability theory to statistics(1855)." The first course in probability theory at the University of St. Petersburg was offered in 1837 by V. A. Ankudovich.

Among the first important works on probability theory in Russian were the works of Nikolai Jvanovich Lobachevskii, in which through his experiment arrived at the problem of determining the distribution law of a sum of a given number of mutually independent identically distributed random variables. The problem is solved in the "new elements of geometry", "probability of the average results obtained from repeated observations". This paper is published in Crelle's Journal, 1842. According to Maistrov(1974), he rigorously derived accurate and convenient practical formulas, like obtaining the distribution of the arithmetic mean, the distribution of the sum of mutually independent varibles, and defined probability in accordance with the definition given by Laplace.

The first works on probability theory carried out in Moscow University were N. D. Brashman's paper "Solutions of problems in the calculus of probabilities" (1835) and N. E. Zernov's(1843) long Memoir (*Probability Theory*, Moscow). Zernov, following Bernoulli and Laplace, in his book is portrayed as representative of the deterministic approach which was prevalent at that tjme; and applies probability theory to demographic statistics, insurance, the theory of errors and legal procedures. Zernov asserts that hardly any other science can be found, except for probability theory, which bears a "direct relation" to so many and such diversified disciplines. Maistrov(1974) observes that another Russian mathematician V. Ya. Bunyakovskiĭ (1804-1889), using Laplace's works and by translating to and compiling in Russian, disseminated the knowledge of probability theory in Russia. It is believed that Bunyakovskiĭ's paper on self-calculators(1867) prompted Chebyshev to construct his arithmometer. Also another prominent representative of the Russian school of probability theory was M. V. Ostrogradskiĭ (1801-1862). His contributions to probability theory were prompted mainly by practical considerations, and were influenced by Laplace.

2.2.2.5 Pafnutiĭ Lvovich Chebyshev

The above mathematicians gave way for the establishment of the *St. Petersburg School*, and to the advent of Pafnutiĭ Lvovich Chebyshev (1821-1894), who is the creator and ideological leader of the pre-revolutionary mathematical school in Russia. Chebyshev, as well as many other mathematicians, was influenced by the works of Ostrogradskiĭ and Bunyakovskiĭ. Chebyshev contributed a lot to the development of mathematics; his investigations span the theory of approximating functions of polynomials, theory of numbers, theory of mechanisms, probability theory and other areas. The mathematical school guided by him, in 1860-1883, played an important role in the advancement of mathematics in Russia. The most prominent representatives of this school were A. N. Korkin(1837-1908), E. I. Zolotarev(1847-1878), A. A. Markov(1856-1922), G. F. Voronoi (1868-1908), A. M. Lyapunov (1857-1918), D. A. Grave (1863-1939), V. A. Steklov (1864-1926).

Maistrov(1974) exaplains that the school was united in common interest and problems, method of discussion of problems and formulation of inquires and materialistic approach to science and mathematics. Chebyshev believed that the harder the problem, the more productive the methods for its solution and the wider the scope of its possible applications. The close relationship between theory and practice was the determining factor in his mathematical activities. According to him an approximate solution is accurate if it is possible to determine bounds for the errors. The pedagogical activities of Chebyshev were expressed in his students, like Lyapunov and Markov; i.e. he was a remarkable lecturer and instructor, thus he was able to lay dawn the establishment of the Russian mathematical school: the majority of P. L. Chebyshev's works and that of his followers tended toward a detailed investigation of problems important from the point of view of applications and which at the same time present special theoretical difficulties and require the construction of new methods, the results being extended into a general theory.

Chebyshev investigated at a great length in probability theory the limit theorems. He wrote only four papers in probability theory, but their influence on the future development of this science was immense. In his first work, which constituted his Master's dissertation, "An essay on elementary analysis of probability theory,"(at Moscow University in 1846) he introduced and utilized his basic premise of deriving accurate bounds on approximating expressions. In his thesis he proved Bernoulli's theorem and also presented corresponding bounds on obtaining approximations; he proves Poisson's theorem for a finite number of different probabilities and also a general elementary proof of this theorem with corresponding bounds on the errors. Next, he worked on the theorems that he calls the basic theorem of probability theory: the addition rule, multiplication rule, and Theorem about conditional probability.

On 17 December 1866, Chebyshev presented at a session of the Academy of Science, his paper "On mean values"[Des Valeurs moyennes]. This is published in 1867 in the *journal Matematicheskii Sbornik*, **II**, pp. 1-9, as well as in Liouville's *Journal de Mathématiques Pures et Appliquées*, pp. 177-184. Here Chebyshev proved an important inequality known nowadays as the Chebyshev inequality. Using this inequality, he obtains a theorem known as Chebyshev's theorem or Chebyshev's form of the law of large numbers from which Poisson's and Bernoulli's theorems follow as particular cases. The theorem can be written mathematically as follows:

$$\lim_{n\to\infty} P\left(\left|\frac{x_1+x_2+\ldots+x_n}{n}-\frac{\bar{x}_1+\bar{x}_2+\ldots+\bar{x}_n}{n}\right|\leq\varepsilon\right)=1, \qquad (2.6)$$

that is, approaching the limit with n→ ∞ we obtain the law of large numbers.

According to Hayde and Seneta(1972) here we should note that the basic proof of this inequality was contained in I. J. Bienaymé's(1796-1878) paper the method of least squares. Hence, it is called the Bienaymé - Chebyshev inequality. Markov(1924) states that we associate this remarkable and simple inequality with the two names, Bienaymé and Chebyshev, because Chebyshev was the first to clearly express and prove it, while the basic idea of the proof was pointed out much earlier by Bienaymé in a memoir containing the inequality itself, albeit in a not particularly obvious form.

The second basic problem that occupied Chebyshev's attention was the central limit theorem. He devoted his attention to this problem in his paper in 1887 in the *Proceedings of the Academy of Sciences*. The paper is entitled "On two theorems concerning probabilities": i) the law of large numbers and ii) the limit theorem for the sum of independent random variables, and the construction of the method of moments in probability theory. And in his paper "Sur les valeurs limitées des integrales",1874, [On integral residual which yield approximate values of the integrals] constructed the moments of order *k*. These moments can be rewitten in the form

$$C_k = \int_a^b x^k \, f(x) \, dx \tag{2.7}$$

where $f(x)$ is *integrable in the Riemann sense.*

As a concluding remark we can quote Kolmogorov's(1947) observation:

> P. L. Chebyshev impelled Russian probability theory into first place in the world. From the methodological point of view the basic change, due to Chebyshev, is not the fact that he was the first who strongly insisted on complete rigour in proving theorems, but mainly that he always strove to obtain exact estimates on deviations from the limiting laws in the form of inequalities applicable for any number of trials." [Kolmogorov, 56]

2.2.2.6 A. A. Markov and A. M. Lyapunov

The two prominent figures and advocates of this school are Andrei Andreevich Markov (1856-1922), who replaced Chebyshev in 1883 till 1905 to guide the school of St. Petersburg, and A. M. Lyapunov (1857-1918). The principal works of Markovin probability theory are related to the limit theorem for the sum of independent variables, in particular those connected in a chain. He was also the originator of a very important branch of probability- *the study of dependent random variables*. He was interested in the following two problems: the applicability of the law of large numbers and of the central limit theorem to sums of dependent variables. Markov's investigation on a sequence of random variables which form a "chain" has far reaching applications. These chains of dependent random variables are now referred to as Markov chains. The study of Markov processes and Markov chains has become a large branch of probability theory with an enormous literature, and is of great significance in the application of probability theory to various branches of the natural sciences and engineering. The model of an atom proposed by Bohr is an example of such a system.

Lyapunov in his two papers proves the limit theorem with weaker restrictions, that is, by using the method of *characteristic function* which overcomes the problem of moments; that is the mathematical expectation of the powers of the random variables, may not exist in all cases, while the characteristic functions exist for any random variable. This method originated in the works of Laplace and Lagrange, and it was used in 1892 by I. Sleshinkii of the University of Odessa. This fact is mentioned in Lyapunov's 1900 paper "Sur une proposition de la théorie des probabilités," (On a theorem in the calculas of probabilities). His second paper is "Nouvelle forme du théorème sur la limite de probabilité," (A new form of a theorem on the limit of proabliities). This method is adhered by many prominent scholars of probability theory, like Gnedenko and Kolmogorov(1968). This method became the basic method for solution of problems on sums of random variables, mainly due to the following property: the characteristic function of sums of independent random variables equals the product of their characteristic functions, i.e.,

$$\varphi_{x+y}(t) = \varphi_x(t)\cdot\varphi_y(t) \tag{2.8}$$

Thus the method of characteristic function is more general than the method of moments. Characteristic functions exist for any random variable and determine completely the moments of the distribution, provided the latter exist. The characteristic function determines uniquely the distribution function, independently of whether the moments exist or not. He also obtained an upper bound on the error committed in replacing the exact distribution of the sum by its limiting distribution.His theorem is called the central limit theorem of probability theory. It explains why so many random variables obey the normal law. It follows from Lyapunov's theorem that, if the random variable X is a sum of a large number of independent random variables, each one of which has only an insignificant contribution to the sum, then the distribution of X will be close to normal. His result is improved after 20 years by Y. W. Linderberg in 1922 by obtaining a new sufficient conditon, and in 1935 by W. Feller by showing the necessary condition.

This necessary and sufficient condition of Linderberg and Feller is stated as follows:

If $F_n(x)$ is the distribution function of X_n, and *h* is a fixed positive number, then, as $n \to \infty$

$$\frac{1}{n} \int_{|x|>h} x^2 d\sum_{i=1}^{n} F_i(x) \to 0. \qquad (2.9)$$

This condition is necessary and sufficient for the convergence of the distribution of $\frac{1}{n}\sum_{i=1}^{n} X_i$ to the normal distribution. Thus, the central limit theorems, in the case of independent random variables, starting with the De Noivre-Laplace version culminated with that of the Linderberg-Feller.

In the 19th century there were many scientists who applied probability theory and statistical methods to advance their researches, especially in physics, namely, Robert Brown(177-1858), James Clerk Maxwell(1831-1879), Rudolf Clausius(1822-1888), Ludwig Boltzmann(1844-1906), Josiah Willard Gibbs (1839-1903). There is a recent advancement on the application of probability theory on "Brownian motion" by Loève(1978). In particular Boltzmann and Gibbs worked hard so that probabilistic theories and statistical methods may be applied in physical sciences. Boltzmann(1964) is primarily connected with the initiation and development of statistical physics. His main contribution was the molecular-kinetic interpretation of the second law of thermodynamics and the derivation of the statistical interpretation of entropy. Gibbs in his book "Basic Principles of Statistical Mechanics" (1902) achieved a logical conclusion of the classical statistical physics. The following remark by Frankfurt and Frank(1964) can express the historical scene in the 19th century: "Gibbs lives, because profound scholar, matchless analyst that he was, he did for statistical mechanics and for thermodynamics what Laplace did for celestial mechanics and Maxwell did for electrodynamics, namely, made his field a well-nigh finished theoretical structure."

Francis Galton(1889), a biologist, revealed the usefulness of statistical methods in biological research and explored what we call regression analysis by introducing the concepts of regression line and correlation coefficient. His research on regression analysis originated from the study of the correlation between characteristics of parents and children, but he failed to realize the difference between population characteristics and sample characteristics. Following Galton, K. Pearson(1857-1936) developed the theory of regression and correlation, with which he succeeded in establishing the basis of biometrics. He arrived at the concept of population in statistics: a statistical population is a collective consisting of observable individuals, while a sample is a set of individuals drawn out of the population and containing something telling us about characteristics of the population. Here de Finetti(1975) states that the idea that all natural characteristics have to be normally distributed is one that can no longer be sustained: it is a question that must be settled empirically.

2.2.3 The Advent of prominent twentieth century Probabilistic Philosophers and Mathematical Statisticians, and Modern Schools of Probability and Statistics.

2.2.3.1 The axiomatic foundations of probability theory

Towards the beginning of the twentieth century, probability theory developed enormously as a result of the contributions of the Russian school, application to physics, and the advent of prominent probabilistic philosophers and mathematical statisticians. The necessity of re-evaluating the logical foundations of probability theory in order to secure its position as a genuine mathematical discipline, and to construct rigorously and to develop probability theory became more and more evident.

A direct predecessor of the founders of axiomatization in probability theory was Henri Poincaré (1854-1912), mathematician, philosopher, and physicist. He wrote the book "Culcul des Probabilités" (1912), which is one of the most rigorous and interesting books on probability theory written at the beginning of the twentieth century. As Maistrov(1974) noted he made remarkable contributions to the fields of differential equations, integral equations, algebra, theory of number, geometry, theory of electricity, thermostatics, theory of Hertz's waves, the kinetic theory of gases, and wrote a number of books and articles of philosophical nature in which he sometimes discusses philosophical and methodological problems of probability theory as well. Poincaré(1912), following Laplace's deterministic approach, defines random events in a deterministic way, and he says:

> If we had an exact knowledge of the laws of nature and the position of the universe at the initial moment, we could predict exactly the position of the same universe in a succeeding moment. ... It may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the latter. Prediction becomes impossible, and we have a fortuitous phenomenon. ... We do not[know] to what are due accidental errors, and precisely because we do not know, we are aware they obey the law of Gauss. Such is the paradox. ... Chance is only the measure of our ignorance. Fortuitous phenomena are, by definition, those laws we are ignorant of. [Poincaré, 1-5,51]

As a conclusion of the classical definition of probability, he makes his remarks saying "How can we determine that all the cases are equally probable? Mathematical determination is not possible in this case; in each application we must put conditions and stipulate that we shall consider these particular cases as equiprobable. These assumptions are not completely arbitrary, but they may escape the mathematician, if he does not analyze them after they have been made. From his remarks we can observe the necessity of more rigorous approach to the concepts of the foundations of probability theory.

There are also other probabilistic philosophers and pure mathematicians who paved the way on the axiomatization of probability theory. For the establishment of a logical order and consistency for any kind of inference, rules of inference, and to show the absence of contradiction in all the results obtained by *an axiomatic method*, the totality of objects studied by a given mathematical theory is necessary. In other words, propositions are set as the basis of the theory and all the subsequent proportions are deduced from these axioms, and the rules of deduction are distinctly formulated.

2.2.3.1.1 Formal axiomatic method

Copleston (1985a) on his philosophical note on G. W. Leibniz(1646-1716), mathematician and philosopher, observes that a deductive system of logic or of mathematics is an illustration or example of the general truth that the universe is a system. Furthermore, he distinguishes two types of truth: truths of reason and truths of facts. Truths of reason are analytic propositions and embrace the sphere of the possible, while of truths of facts are synthetic propositions and embrace the spare of the existential. Existential propositions are truths of act not of reason. Among truths of reason are those primitive truths which Leibniz calls ‘identicals,’ which are known by intuition, their truth being self-evident. A true proposition asserts existence of a subject, it is a truth of fact, a contingent proposition, and not truth of reason. Copleston (1985a) affirms that there is a philosophical and epistemological approach to mathematical proposition: mathematical propositions do not give us factual information about the world; they state, as Hume[1711-1776] put it, relations between ideas. for factual information about the world indeed about reality in general, we have to turn to experience, to sense perception and to introspection. If we wish for factual information about the world, we must content ourselves with probabilities, which is all that inductively-based generalization can give us. With this concept in mind, now we can give analysis of the formal axiomatic approach to mathematical truth.

Towards the end of nineteenth and the beginning of the twentieth century, the axiomatic method penetrated various branches of mathematics. This fact followed after the discoveries, independently, by János Bolyai (Hungarian, 1825) and Nikolay I.

Lobachevsky (Russian, 1826), that there is a possibility of constructing geometry on axioms different from Euclid's, that is, by assuming that for some plane, some line ℓ in the plane, and some point p in the plane and not on ℓ, there exist at least two distinct lines in the plane passing through point p and not intersecting line ℓ. In pure mathematics the systems of axioms for geometry was carried out by Mortiz Pasch(German, 1882), David Hilbert *(Grundlagen der Geometrie, 1899; The Foundations of Geometry, 1902).*, Guiseppe Peano and V. F. Kagan; axiomatization for arithmetic were initiated by Peano(1889) and Hilbert(1897).

The postulational method when the undefined terms are treated as meaningless is called the formal postulational method, or formal axiomatic method. Thus, axiom and postulate are often synonymous. The deduction of such a theory independent of any interpretation makes it a mathematical tool prepared in advance for diverse applications. Furthermore, transcending the mathematical branches, it is a method for discovering new facts in general.

In the beginning of the twentieth century the inadequacy of the classical foundation of probability stemming from Laplace was noted; especially the highly restrictive nature of its applicability to problems of physics, statistics, biology and the technical sciences. Hence, new logical foundation for probability theory, in line with the other branches of mathematics, on the axiomatic method were needed.

2.2.3.1.2 Promoters of the axiomatic foundations of probability theory

The first works in this approach are due to S. N. Bernstein (1880-1968). Bernstein's book "Probability Theory" (1946), served as a text book for mathematicians, physicists and other disciplines, and presents a detailed axiomatization of probability theory.

Bernstein(1917) introduced three axioms, namely,

> i) *The axiom of comparability of probabilities*: If a is a particular case of A in the strict sense, then $P(a) < P(A)$; conversely, if for events a_1 and A the inequality $P(a_1) < P(A)$ holds, then $P(a_1) = P(a)$, where a is a certain particular case of A in the strict sense.

ii) *The axiom of incompatible(disjoint) events*: If it is known that events A and A_1 are incompatible, and, moreover, that events B and B_1 are also incompatible, while $P(A)=P(B)$ and $P(A_1) = P(B_1)$, then the probability of event C, which consists in the occurrence of event A or event A_1, is equal to the probability of event C_1 consisting in the occurrence of B or B_1, i.e. $P(A \text{ or } A_1) = P(B \text{ or } B_1)$.

iii) *The axiom of combination of events*: If α is a particular case of event A, then the probability of α under given conditions depends only on the probability of events A under the same conditions and on the probability acquired by α in the case when event A occurs.

Two corollaries are deduced from axiom(i): a) the probability of a certain event is larger than the probability of a possible event, and b) the probability of a possible event is larger than that of an impossible one. Furthrmore, The axiom of combination of events can be formulated also as follows: The probability of combination of A and B (under given conditions) depends only on the probability of A (under the same conditions) and on the probability acquired by B after the occurrence of A.

On the basis of these axioms Bernstein constructed the whole structure of probability theory. Kolmogorov(1947) says that the first systematically developed axiomatization of probability theory, based on the notion of qualitative comparison of (random) events according to their (larger or smaller) probability is due to S. N. Bernstein The numerical value of the probability appears in this conception as a derived rather as a primary notion. Glivenko(1939) showed the equivalence of Bernstein's axiomatization with Kolmogorov's set-theoretical axioms and Bernstein's idea was further developed by Koopman(1940). His ideas of axiomatization and the application of probability theory to problems in the natural sciences served as the basis of his "Probability Theory," which is one of the classical works on probability theory.

For the sake of completion of the study of axiomatization of probability theory, before stating the classical work of Kolmogorov, we need to mention other scholars who worked on this field and corresponding publication year of their works: Mises(1919,1931, 1936), Keynes (1921), Lévy(1928), Cantelli(1932,1939), Kamke(1932), Reichenbach

(1935), Jeffreys(1935,1936,1937,1938), de Finetti (1937), Castelnuovo(1937), Wald (1938), Borel(1939), and Fréchet(1939-43). Their respective stands with respect to the foundation of probability will be explained in detail on the sub-section entitled "on different approaches of probability."

Richard von Mises(1883-1953), German-American - the founder of the school of "the frequency approach in probability theory," who was advocating that probability theory is a science investigating phenomena of real world rather than a mathematical discipline, points out the shortcomings of the classical definition of probability and tried to amend it by defining probability as the limiting value of the relative frequency. Von Mises'(1964) conditions or axioms are two: (i)There must exist limits of the relative frequencies of events with particular attributes within the collective. (ii) These limits are invariant with respect to the choice of any subsequence of the collective which is arbitrary(except that it must not be based on distinguishing the elements of the collective in their relation to the attribute under consideration). His approach is a conceptual approach of axiomatization for the limiting-frequency theory.

Another attempt to axiomatize a conceptual approach for the subjectivistic theory is due to Keynes(1921) and de Finetti(1937). Later Jeffreys (1939) developed the notion of probability as the degree of likelihood.

De Finetti(1972) in his analysis on "the axiomatic foundations of probability theory" discusses both from the formal point of view and with reference to the different conceptions about the meaning and role of probability. Himself is an ardent adherent of a *subjectivist or personalist school of probability*. Following Frank Plumpton Ramsey, in 1937, he made a systematic attempt to base the mathematical theory of probability on the notion of partial belief.

2.2.3.1.3 The axioms of Kolmogorov(1933)

For its clarity and conciseness we will follow de Finetti's(1972) description of the axioms of Kolmogorov. But a good descriptive analysis and its influence on statistics is given by J. L. Doob(1976) on 'Axiomatic approach of Kolmogorov' in his article entitled "Foundations of Probability theory and its influence on the theory of statistics."

Kolmogorov, a prominent figure of the Moscow school of probability theory, was able to construct an axiomatization of probability theory which is a decisive stage in its further development. He and his colleagues, like Khinchin(1956,1961) and Gnedenko(1948), were greatly influenced by ***the concepts of set theory*** and ***the metric theory of functions***. In 1920s he was engaged in the logical formulation of the ideas of the metric theory of functions in probability theory. His research resulted in the publication of "Grundbegriffe der Wahrscheinlichkeitsrechnung" in 1933. In his book the analogies between the notions of the measure of a set and the probability of an event, between the integral and the mathematical expectation, orthogonality of functions and the independence of random variables, and others were established.

Thus, probability theory attained an equitable position among other mathematical disciplines.

In Kolmogorov's approach, probability theory is the study of probability domains ($\mathcal{E}$,P) (Wahrscheinlichkeitsfelder) which are defined as follows:

a) A class C of primitive elements, called *elementary cases* is given;

b) sets of elementary cases, that is, *subclasses* of **C**, are called *events*;

c) finally, one considers a class of events $\mathcal{E}$ and a function ***P*** satisfying the following

Axioms:

I. $\mathcal{E}$ is a *field* of sets of C (i.e. the union, intersection and difference of sets in E belong to $\mathcal{E}$);

II. $\mathcal{E}$ contains C (i.e. the "sure events", the set of all elementary cases, must be included among the events in $\mathcal{E}$);

III. A real-valued, non-negative function is defined on $\mathcal{E}$. This function assigns to any set E in $\mathcal{E}$ a number P(E), the *probability* of E (but we could also use such terms as *measure* or *mass* in order to avoid, even in the terminology, any reference to controversial notions);

IV. P(C) = 1 (this is a convention concerning the value of the probability of the sure event);

V. If E_1 and E_2 are *disjoint* events (have no elements in common), then,

$$(E_1 \cup E_2) = P(E_1) + P(E_2);$$

VI. If $E_1, E_2, \ldots, E_n, \ldots$ is a sequence of events in $\mathcal{E}$, such that each event is contained in the preceding one and that their logical product is null (i.e. there is no elementary case belonging to all the E_n), then, $P(E_n) \to 0$ as $n \to \infty$.

Finally, the *conditional probability* of E given H, written P(E/H), where E and H are events in $\mathcal{E}$, is defined by Kolmogorov, according to the compound probability theorem, as follows:

$$P(E/H) = P(EH) / P(H), \quad \text{assuming that } P(H) \neq 0.$$

The *compatibility* of these axioms is proved by Kolmogorov by considering an $\mathcal{E}$ consisting of only two events, the sure and impossible events, with corresponding probabilities equal to one and zero.

We should note that Axiom VI, a "continuity postulate," is equivalent to "complete additivity".

Although the axioms of Kolmogorov help to define chance they are no enough. As Koopman(1940) and Hacking(1974) have shown they do not determine, for instance, an hypothesis about chances, a statistical hypothesis, which is well supported by statistical data. Their central point is the logic of comparative support, i.e., it is concerned with the assertions that one proposition is better or worse supported by one piece of evidence, than another proposition is by other or the same evidence.

2.2.3.2 On Different Approaches to Probability and Corresponding Schools

The issue on the approaches of scholars to probability theory is open-ended. Specifically, after the Heisenberg's(1927) *uncertainty principle*, the whole of science was recognised as ultimately based philosophically, on the concepts of experimental probability. Thus from late 1920's on wards, since probability theory covers a combination of mathematics, measurements or statistical data, theory of nature and theory of knowledge itself, activity in philosophical probability has been intensive. The *principle of uncertainty*, in quantum mechanics, states that the position and momentum(velocity) of a given particle can not be exactly measured simultaneously with complete accuracy. The amount of uncertainty is specified by the formulas

$$\Delta x \cdot \Delta p_x \geq h.$$
$$\Delta y \cdot \Delta p_y \geq h. \qquad (2.10)$$
$$\Delta z \cdot \Delta p_z \geq h.$$

where Δx is the uncertainty in the value of x, etc., and h is Plank's constant, about $6x10^{-27}$ erg-sec. Beside the position-momentum uncertainty relation there is the energy-time uncertainty relation: $\Delta E \cdot \Delta t > h$, but the verbal interpretation is quite different in nonrelativistic quantum mechanics. Now we can pose a question of enquiry: "do probability philosophers admit the presence of subjective and ontological knowledge of nature or randomness?" To get a satisfactory answer an intrinsic and extrinsic discernment on authoritative research works on probability theory must be done.

But in this section we shall confine our selves, in view of variations among scholars, on four approaches to probability, i.e. we can list schools by their respective approaches to mathematical statistics and probability.

a) *The classical approach*

The classical approach, based on the notion of *equally probable cases* by reasons of symmetry, and on the consequent definition of probability as the ratio of the *favourable* to the *possible* cases. As the name indicates, this approach is utilized by the first scholars on probability theory like Bernoulli, Laplace and their followers.

b) *The empirical approach*

The empirical approach, based on the notion of *repeatable events* whose frequency on a large number of trials, by virtue of the so-called *empirical law of chance*, gives the probability almost certainly and exactly. Among the modern scholars, the positions of Castelnuovo, Cantelli, Fréchet, and Lévy can be classified under the empirical approach. Borel too, could be included in this group, although his position is not far from the subjectivists'.

c) *The asymptotic approach*(frequency theory)

The asymptotic approach which, with some idealization, makes the preceding definition more precise by considering an *infinite sequence of trials*, von Mises' *collective*, and defines probability as the *limiting value of the relative frequency*. Both in its basic formulation and in its applications, seems geared mostly to statistical inference. A very similar approach is used in work on the statistical design of experiments, especially by Fisher. The scholars which adhere the asymptotic approach are like von Mises, Kamke and Reichenbach; also Kolmogorov declares himself in favour of the asymptotic approach.

d) *The subjectivistic(Degree of Belief) approach*

The subjectivistic approach considers probability a measure of the degree of belief of a given subject in the occurrence of an event (proposition). There is the so called *the logical theory(degree of evidence)*, fostered by Keynes, Carnap and Jeffreys, is largely deductivists in approach, and seems closes to the pure theoretical formulation of mathematical probability. While de Finetti, Ramsey and Savage agree with the logical theory in making probability a logical relation between a statement and a body of evidence; it disagrees in allowing the evidence to vary in terms of the knowledge available at a given time.

Generally, we note that the positions even of scholars classified in the same group can differ somewhat in terminology, omissions or additions, in strict form or weaker form etc. For further discernment on different approaches to probability refer to the original works of the authors cited on the bibliographical list.

2.2.3.2.1 Further clarifications

To discern more on the unifying and diversifying factors of the different approaches to probability more clarification and elaboration are required. Let us see the positive and negative comments of the respective approaches.

The **classical approach** is reviewed by the subsequent schools. As De Finetti(1972) points out "the disputed pointed alluded to originate from certain tendencies to limit the domain of the theory of probability to narrow zones where notions of symmetry(as for a dice) or of statistical regularities(as for the sex of a future birth) facilities evaluations or probability and agreement about them on the part of diverse individuals.".In plain words discrimination between the total field of uncertain facts and those subfields to which a privilege role has been assigned is not justified. The term equipossibility is more elaborated, by von Mises(1951) and more recently by Boudot(1967), as the *epistemic concept* of probability corresponds to an epistemic concept of possibility, while the *aleatory concept* of probability corresponds to a concept of physical possibility or it possesses the *de re* and *de dicto* modality; That is de re if it pertains to things, and de dicto if it applies to what is said or can be stated. But this kind of elaboration and equipossibility definitions of probability are not fully convincing for modern logicians. The concept of equipossibility is explained using the case of exchangeability or symmetry with respect to order. This approach applies directly onto to special cases, so it is not adequate. According to de Finetti(1972), from the critical viewpoint, this approach can give at best, only an incentive to reduce the axioms of quantitative probability - as a direct numerical assignment of probability - to axioms of a purely quantitative nature - in the form of, inequalities among probabilities. Some scholars use the principle of indifference and the range theory of probability to describe the classical approach.

In other words, those in favour of **empirical approach**, like De Groot(1970) based on what is experienced or seen rather than on theory, maintain that *probability is a logical concept which can be applied to parameters in a much wider class of problems; and in each such problem there is a uniquely defined distribution which is appropriate for a particular parameter and most necessarily be assigned to that parameter*. The critique of the this approach say that for this school an event can be assigned a probability under hypotheses of 'stability of the frequency' which it is difficult to make precise. Thus it is impossible to base a rigorous analysis of such a foundation.

The influence on von Mises' approach on probability, directly or indirectly, stems from the theory of empiricism. The real father of the classical empiricism is John Locke(1632-1704). This philosophy, by vigorously attacking on the theory of innate ideas, propagates that all our ideas come from the elementary data of experience, from sense-perception and from introspection.

As stated above, the basic notion in von Mises' frequency theory of probability theory is the concept of a collective. A collective is an infinite sequence K of similar observations, each of which determines a certain point belonging to a given finite-dimensional space R, i.e. probability is **a limiting value of the frequency**, or *a distribution is appropriate only when values of the parameter clearly have relative frequencies*. According to von Mises the events do not possess probabilities prior to the experiment: the probability is not an objective property of the phenomenon. Phenomenology is the study of objects and events as they appear in experience or immediate object of awareness in experience. Thus, phenomenon is a thing as it appears in mind or thing-in-itself. Probabilities of events arise only as a result of an experiment. In von Mises' view we do not determine the existing objective properties by means of an experiment, but rather attribute them to the phenomena. Thus probability is deprived of its meaning as an objective numerical characteristic of real-world phenomena. Even Kolmogorov(1956), who is in favour of the approach, shows reservation for the extreme stand and says:

> The assumption concerning the probable nature of trials, i.e., concerning the tendency of frequencies to group around a fixed value may be valid on its own only if certain conditions are presented which cannot be retained for an indefinitely long time and with indefinite precision. Therefore, the limiting transition $m/n \rightarrow p$ cannot have real meaning. Moreover, the formulation of the stability of frequencies principle using this limiting process requires the availability of admissible methods for determining infinite sequences of trials which can be a mere mathematical fiction. [Kolmogorov, 274-275]

In the **asymptotic approach**, the imprecise condition of stability of the frequency is replace by the condition of existence of a limiting frequency. In de Finetti(1972) words the defect of this school is also pinpointed as follows:

> probability cannot be assumed to be defined for every event, and those events for which a probability exists do not form a filed ... assuming that events A and B have been assigned some probabilities, it may nonetheless [be] the[sic] impossible to attach a probability to their logical product AB.[de Finetti, 74-7]

As von Wright(1977) noted it is a great merit of von Mises to have stressed the importance of the idea of random distribution to a frequency theory of probability. The demand of randomness is relevant to the question of the adequacy of the frequency view as a proposed analysis of the meaning of probability. But randomness is not relevant to the question of the mathematical correctness of interpreting abstract probability in terms of frequencies. Nevertheless, some form of frequency theory is thought by many writers to offer the best account, for a large category of cases, of the relation between abstract probability and empirical reality.

The primary incentive for this approach seems to be that conviction of David Hume(1738-40): We ought to start with a çlose observation of man's psychological processes and of his moral behaviour and endeavour to ascertain their principles and causes. Our method must be inductive rather than deductive. And where experiments of this kind are judiciously collected and compared, we may hope to establish on them a science which will not be inferior in certainty, and will be much superior in utility, to any other of human comprehension. This approach is also influenced by Kantian(1724-1804), Hegelian(1770-1831) and Leibnizean(1646-1716) philosophical thoughts and Bayes' theorem, the acceptance of a-priori, existing in the mind prior to and independent of experience, from cause to effect, and a-posteriori, based upon actual observation or upon experimental data, from effect to cause.

In **subjective approach** a distinction between events to which a probability can be assigned and those to which it cannot, does not seem acceptable. In other words, the probability of the logical product of events A and B cannot be deduced from the probabilities of the single events. In simple terms in this theory not only the notion of probability but other basic notions such as dependence, independence, equipossibility and others are defined subjectively; even the relationship between probability and frequency is subjective. In simple terms, probability distributions are subjective and that whenever anyone carries out a statistical investigation involving a parameter, he can represent his uncertainty pertaining to the values of that parameter by suitable probability distribution.

As von Wright(1977) states, it is true that the combination of probabilistic ideas with the value-theory notions of preference and utility has had fruitful applications to the mathematical study of economic and related forms of human behaviour. Support of this fact does not exclude taking a somewhat critical view of the epistemological and logical basis of the belief theory of probability.

This approach bases its assumptions on the inductive reasoning, or bases the mathematical theory of probability on the notion of partial belief, taking the point of departure on the measuring a person's belief by proposing a bet and observing the lowest odds that he will accept. That is the laws, of probability may be called rules for consistent(coherent) sets of degrees of belief. Lévy(1953) says that one who wants to get to a certain point must first see it with his eyes(intuition) before he reaches it with his feet(logic), which expresses the range of the field of application of inductive reasoning. A distribution of partial belief contrary to the laws of probability, Ramsey(1931) says that would be inconsistent in the sense that it violated the laws of preference between options, such as that preferability is a transitive asymmetrical relation. In brief this school stresses the role of mathematics in inductive reasoning, i.e., in the theoretical and exact formulation of inductive reasoning.

To see the distinction of interpretations we can take as an example "Kolmogorov's fifth axiom on finite additivity." For the classical, empirical and asymptotic approaches the axiom is a purely arithmetical consequence of the definition, since the number of "favourable cases" as well as the "frequency" and "limiting frequency" are naturally additive; for the subjectivistic approach, the axiom is a necessary condition for the mutual consistency of the assessments of different probabilities.

2.2.3.2.2 Special Remarks

At this preliminary exploration of the different schools of statistics to give a critical examination of controversial aspects of the different approaches may be immature. But as we are dealing with identification of school, it is good to assess their respective deficiencies and novelties.

Already a critical, philosophical and historical survey about this issue has been done by de Finetti(1972), Hacking(1975), Maistrov(1974), von Wright(1977) and others. The most appealing question is "*is probability subjective or objective?*" Subjectivism, in philosophy, stresses the doctrine that all knowledge is limited to experiences by the self, and that transcendent knowledge is impossible. This stresses the role of consciousness or thought. While, objectivism is used to stress the apartness of thing known[the object] from the person who knows it, or the things external to the mind or external elements of cognition. A typical objectivist would conceive his role as that of discovering an order[that pre-existed his mind] in reality. The subjectivist conception of probability as a degree of belief is often contrasted with the objectivist conception of the notion as either a relative frequency or a ratio of measures of ranges. It is questionable, however, whether a sharp contrast can be maintained between objectivism and subjectivism in the philosophy, particularly the epistemology, or probability. Among the various objectivistic schools, the root seems to be identifiable as the antithesis between **inductive reasoning** and **inductive behaviour** (up-held by Neyman-1957). New developments in the theory of inductive behaviour arise only in so far as the decision, unlike the opinions, are made by groups rather than individuals. There are two distinct aspect of all approaches, namely conceptual questions and mathematical questions.

Let us, following De Finetti(1972), present in table form the five points about inductive reasoning and five points about the passage to inductive behaviour or decision theory.

Inductive Reasoning	The passage to inductive behaviour or Decision Theory
1. Of an event, that is, of any verifiable proposition in the domain of the logic of certainty, we can only say whether it is certain, impossible, or possible - that is, whether the answer is either demonstrably "yes" or "no", or one cannot prove either "yes" or "no" - and, in the domain of probabilistic logic, we can only evaluate the probability according to our judgement.	6. Optimal behaviour in the face of uncertainty for a given individual consist in choosing a decision that maximizes the expected utility. If information can be obtained free of cost and the choice can be made afterward, one simply has a widening of the field of possible decisions. An optimal decision is obtained by choosing appropriately the partition about which to request information and then choosing the decision optimal for each of its elements. If the information does entail cost, this cost must be included. If utility and cost can be expressed in monetary terms, one need only subtract the cost of investigation from the expected gain.
2. Nothing can be derived from the ostensible concept of "not knowing anything." Better said, this is an expression that means nothing precise though it corresponds ineffectively and ambiguously to certain ideas.	**7.** A collective decision by several individuals, who agree on their evaluations of utility but not on those of the probabilities, must be optimal for a hypothetical individual whose opinion are convexly comprised among those of the real individuals concerned.
3. Any assertion concerning probabilities of events is merely the expression somebody's opinion and not itself an event. There is no meaning, therefore, in asking whether such an assertion is true or false or more or less probable.	**8.** A factor that can be important here is the tendency of opinions, as information increases, to agree with each other.
4. If we speak of a conditional probability P(E/H), we must repeat for E as well as for H what has been said of E with respect to P(E). Namely, the expression has meaning if and only if E and H are events. Also H must express the assumed or acquired information in its entirety.	**9.** A different, and independent, mechanism for agreement on action may apply when a collective decision is envisaged that may prove to be a bad one for a certain member but only in cases that the member concerned initially considers unlikely.
5. Inductive reasoning is nothing other than reckoning P(H/E), the probability of H after the observation of E, in accordance with Bayes' theorem - or, equivalently, according to the theorem or compound probability, of which Bayes' theorem is corollary.	**10.** Greater complications are encountered with more widely differing attitudes and interests of the individuals. But no new criterion is called for: One has but to apply the criterion of the maximal expected utility in different circumstances.

De Finetti, in defensive mood, tries to justify the subjective approach by presenting the above important points and also gives the opposing views.

Supporters of the frequency view found that an adequate analysis of probability requires them to combine their definition of the concept with the idea of a random distribution of events on a series of occasions. Supporters of the range theory or classical approach have had recourse to some form of a principle of indifference for the determination of equipossibility in certain unit alternatives. The objection raised is that whether randomness and equipossibility can be satisfactorily accounted for without reference to states of knowledge or ignorance.

Now it is good to use von Wright(1977) terminologies in our description of subjective approach to probability as a degree of belief. The belief theory does not necessarily entail an identification of probability with belief as a psychological phenomenon. The attitudes in option between goods may be said to reveal subjective estimations of probability. But the derivation of the laws of probability within the belief theory does not confer on them the status of psychological laws of believing. It rather makes them standards of rationality (consistency) in the distribution of beliefs or in preferences. So we cannot regard belief theory as an account of probability in purely subjectivist, i.e., psychological terms. In other words, according to Durbin(1967), the personal probabilities of the subject-matter experts in the domain in question[can furnish] us with an estimate of probability in the sense of degree-of rational belief.

As a concluding remark we can agree with the following Durbin's(1967) observations:

> Probability theory is a pure mathematical model and as such abstracts from its applications in the real world. Nonetheless, probabilistic models are highly useful in explaining the real world, especially where statistical laws prevail. Finally, there are valid methods for arriving at these probabilistic models from observed data. [Durbin, 816]

2.2.3.3 The development of statistical mathematics in the twentieth century

The transformation of the status of probability theory in contemporary study of science is due to the advent of prominent twentieth century probabilistic philosophers and mathematical statisticians. Synthetically, we can give historical analysis of the development of statistical mathematics using two approaches, namely, the *parametric thesis* and *chronological approach*.

2.2.3.3.1 The Parametric Thesis

The parametric thesis proposed by van Dantzig, Danish, with his colleague Hemelrijk, in 1954, divides the development of mathematical statistics into *four stages* characterized by the use of *one, two, many and no parameters*, respectively.

According to van Dantzig, the **first stage** of development began with the discovery of regularities of certain statistical ratios. In the area of demography where sample surveys were first used, and the English merchant John Graunt's famous work(1662), William Petty, Edmund Halley, Per Wargentin of Sweden, and Johann Peter Susmilch of Germany are good examples of this stage of development. In modern terminology, this stage of knowledge about demographic phenomena is characterized by *one parameter*, i.e., the mean of the population under investigation.

The **second stage** is marked by the growing awareness of variability. Various laws of error were suggested by the eighteenth century astronomers, culminating in the works of Laplace and Gauss on the normal law of error. Mathematically speaking, the population was characterised at this sate by *two parameters*, i.e., the mean and the precision constant(standard deviation), and in the more general case of multivariate distributions, by the *first two moments*. Consequently, all statistical theories based on the *normal law* or *error* belong to this stage, including the least-squares method, theories of correlation and regression, and the analysis of variance and covariance.

The works of Laplace and Gauss generated excessive reliance on the *normal law*, especially by such Quetelet, Airy and Galton. However, toward the end of the ninetieth century, empirical investigations gradually demonstrated normality to be the exception rather than the rule. This led to the development of Karl Pearson's system of skew curves and the development of the Gram-Charlier series, the theory of curve-fitting by the method of moments, and Pearson's χ^2 goodness-of-fit test. This **third stage** is therefore characterized mathematically by the use of *many parameters* in frequency distributions.

The large sample theory developed during the second and third stages gave way to a more realistic small sample theory in the twentieth century. The search for new foundations of statistics in the 1920s and 1930s led to the replacement of the *inverse approach* by those of Fisher (the principle of randomazition) in agricultural surveys, and the Neyman-Pearson (confidence interval) in social surveys. This desire of logical rigour, according to van Dantzig, was responsible for the increased interest in the *non-parametric* or *distribution-free approach*, which characterizes the **fourth stage** of development.

Furthermore, Wei-Ching Chang(1976), in elaborating van Dantzig's thesis, emphasises the importance of the works of three prominent scholars in the history of theory of probability and mathematical statistics, namely, Laplace, Fisher and Neyman. And in line to the history of sampling survey, he suggests that to complement van Dantzig's thesis, adherence to the inferential procedures which is used by different schools of thought should be emphasized.

2.2.3.3.2 Development of Statistical Mathematics - Chronologically

As Lehmann(1959) observed a period of intensive development of statistical method and a systematic use of hypothesis testing began towards the end of the century with the work of Karl Pearson, his χ^2 paper of 1900- Chi-square test for goodness of fit, and this attitude towards scientific investigation can be seen from his article:

> "No scientific investigation is final; it merely represents the most probable conclusion which can be drawn from the data at the disposal of the writer. A wide range of facts, or more refined analysis, experiment, and observation will lead to new formulas and new theories. This is the essence of scientific progress."[1899,169-244]

In connection with his system of curves, Karl Pearson developed a method of point estimation known as the method of moments

Neyman(1976) as well as other scholars stress that in the late 19th and early 20th centuries mathematical statistics emphasized the Kollecktivmasslehre, a mathematical discipline concerned with collective characteristics of populations. In other words, statisticians at this period thought of a population as a collective having infinitely many individuals, which led to the idea that the larger the size of a sample the more precisely could the sample give information about the population. The term descriptive statistics was introduced to mean the use of a variety of methods for describing such characteristics of populations. An important method of descriptive statistics is to consider a family of flexible curves or surfaces that can be used to approximate the empirical frequency distribution. A number of such families were developed, all representing interpolation formulas. The most successful system seems to be that of due to Karl Pearson.

In the 1920s the Hungarian-American mathematician George Polyá constructed a system of chance mechanisms that can generate almost all the distributions of Karl Pearson's system. Thus mathematical statistics shifted from Kollektivmasslehre to the construction of chance mechanisms or the so called stochastic models of phenomena. This idea was explicitly stated by Emile Borel of France : *The basic problem of mathematical statistics is to invent a system of simple chance mechanisms, such as throws of a coin, so that the probabilities determined by this system agree with the observed relative frequencies of the various details of the phenomena studied.* Borel's definition was good but it does not allow mathematical statistics to stand as a field of its own. Depending on the attitudes of given research workers, stochastic models belong to the relevant substantive fields or the theory of probability.

Following the footsteps of K. Pearson, F. Y. Edgeworth of England, in 1908, found that consistent use of the method of moments must yield an excessive frequency of large errors of estimation and proposed a new method of estimation, conjectured to be much better. In 1922 R. Fisher discussed the same ground more rigorously and intensively. He introduced the term method of maximum likelihood. The English mathematician *Student(1908)*, pen name of William Sealy Gosset- by his discovery of the exact distribution of *t* opened the new epoch of *exact sampling theory or distribution*, and Fisher developed a number of tests of particular hypotheses: he introduced the concepts of null hypothesis and significance test, and added the concepts of consistency, efficiency and sufficiency to the list of possible properties of estimators. Lehmann(1959) noted that problems of testing hypotheses and of estimation got recognition as an independent field for systematic study when Fisher in 1922 gave a new definition of statistics, and J. Neyman's(1928,1933,1935,1938) and E. S. Pearson's(1928) principal ideas came to light. Stating that the object of statistical methods is the reduction of bulky data, Fisher distinguished three basic problems: those of specification of the kind of population from which the data come - *Kollektivmass*, of *estimation* and of *distribution* - probabilistic problems connected with point estimation. In these papers, that is, Fisher(1920, 1922,1925) and Nayman(1935), the concept of sufficiency is developed, principally in connection with the theory of point estimation. The factorization theorem is given in a form which is formally weaker but essentially equivalent to

$$p^{\chi}_{\theta}(x) = g_{\theta}[T(x)]h(x). \qquad (2.11)$$

The above definition can be clarified using the descriptive definition of Herman Chernoff(1976):

> The word *statistic* applies to a number that summarizes aspects of the data and is typically assumed to behave *randomly* according to some law of probability or *probability distribution* determined by θ. The relationship between θ and the inferences with regard to θ. Thus, the statistician must cope with *random variation*, and the mathematical statistician is concerned with the probability distribution of possibility complicated functions of the data. These are called *sampling distribution*. [Chernoff, 208-9]

Hansen and Madow(1976) confirm that In India, during the World War II, P. C. Mahalanobis (died 1971) has contributed a lot in the development of statistical mathematics, independent of other schools, namely, by creating and directing the work of the Indian Statistical Institute, developing a strong program of sample-surveys. Among the numerous contributions made by him and his colleagues is the extensive use of interpenetrating samples from initial sample selection through the successive stages of data collection, data processing and analysis. The Indian Statistical Institute, under the direction of C. R. Rao, after 1971, has continued to be a major source of contributions to statistical theory and practice. And another who contributed for the statistical application in agricultural area is P. V. Sukhatme, who lead the Indian Council for Agricultural Research.

According to Lehmann,(1959) a formal unification of the theories of estimation and hypothesis testing, which also contains the possibility of many other specialization, was achieved by Wald(1939,1950,1958), who gave a single comprehensive formulation in his general theory of decision procedures.

Kiyosi Itô(1987) states that after the publication of Savage's book in 1954, there was a revival of the Bayesian approach, that is, one based on the concept of subjective probability, and now the group of those statisticians who accept the Bayesian approach are called Bayesians or neo-Bayesians.

2.2.3.4 Different Branches of Statistics

Broadly speaking, the branches of Statistics can be classified or listed as: i) Statistical Inference - Theory of Estimation and testing of hypothesis, ii) Probability and Distribution Theory, and iii) Design of Experiments

i) Statistical Inference - Theory of Estimation and testing of hypothesis

The most authoritative books on this branch are mainly Lehmann's(1959) and Zacks'(1971). Lehmann in presenting competently the "Testing Statistical Hypothesis" gives also the historical development of theory of estimation and test of hypothesis. The main contributors to this development are mainly Neyman(1928-38) and Pearson(1928,1933), Fisher(1922), Savage(1962) and A. Wald(1950,1971). Zacks, using the measure-probabilistic approaches, gives a special impetus to the theory of estimation;

others who worked on this area are De Groot(1970), Lehmann. Bayesian method can be seen in the light of statistical inference - there is a standard text by Raiffa(1968).

ii) Probability and Distribution Theory

This branch has a long history. It is already dealt with at length. We note that the standard works on this field are done by Kolmogorov(1933), Gnedenko and Kolmogorov (1968), Saks(1937), Halmos(1950), Doob(1953), Loéve(1962), Chow and Teicher(1978), Feller(1965, 1968), de Finetti (1972,1975) etc. Stochastic processes can be affiliated to applied probability theory - the standard texts are those of Doob and Feller.

iii) Design of Experiments

The design of experiments consists of two parts: a) the analysis and b) the constructions and combinatorial problems. The analysis of design of experiments has its origin principally in the work of R. A. Fisher, much of it contained in his books (1925,1935). A comprehensive treatment is given by Kempthorne(1952), Cochran and Cox (1957), Scheffe(1959), Rao(1973). While the constructions and combinatorial problems is treated elegantly by Bose(1938,1939,1947, 1963), John(1971), Raghavarao (1971), Raktoe et al. (1981) etc.

Currently *statistical mathematics* is progressing rapidly and has a vast branches, in which as part of applied mathematics, applicable to different fields of science. The "structure in data" embraces embraces many fields in statistical branches: multivariate analysis, time series analysis, sample survey, quality control, information theory, sequential analysis, nan-parametric statistics, mathematical population. One of interesting new application of statistics is to genetics- in which Kempthorne's(1957) book *An Introduction to the Genetic Statistics* is a good example. Many scholars have produced standard texts in their respective fields, like Anderson(1958), Rao(1952) and Kendal(1957), and Mardia(1979) in multivariate analysis; Anderson(1971) Grenander (1957), and Wold(1954) in time series analysis; Cochran (1992) in sampling survey; Kullback(1968) in information theory and coding; Wald in sequential analysis, Hollander and Wolfe(1973) in non-parametric statistics; Keyfitz(1968) in mathematics of population or demography.

2.3. Schematic Discernment on Statistical and Probabilistic Research Works and Method of Approach at Nairobi University

2.3.1 Historical Background

The analysis of our historical background consists of identification of the initiators of the Statistical Section of Mathematics at University of Nairobi and the background of methodological approach to probability theory with respect to modality of lectures and research works. The following are guiding questions: "who initiated Statistical Section of Mathematics at University of Nairobi, when and how was it started, which one of the different approaches to probability theory is adhered to as a background of methodological approach?" The reliable answers to the above interesting questions can be deducted from the first protagonists of the period of initiation of Statistical Mathematics at University of Nairobi, as well as from the original research done by probability mathematicians.

Professor M.S. Patel with some of his first Kenyan students
From Left: C. Achola(Mathematics), Prof. J.W. Odhiambo, Prof. M. S. Patel, Dr. Ebi Kimanani, Prof. J.A.M. Ottieno

The Statistical Section of Mathematics at University of Nairobi started with the "Unit-Course-Work" system in 1975. The first group to follow this system are Dr. M. M. Manene, Dr. F. Njui, Mr. Kinya and other student colleagues. The main players in this sparking insight and future focus as a successful statistical research centre in the whole country are Professor M. S. Patel and his colleague lecturers. The initial foundation of research approach, basing our observation on the dissertations, was firmly established on the emphasis of making new research works on group screening. In 1984-1985 we see a slight shift to the question of "Rotatable Designs." A new direction of study is also opened by 1991 with a doctoral dissertation entitled "A stochastic Model for Stocks and Flows of Students in an Education System." In 1993 and 1995 we see again a new research work on Designs. Finally, the 1997 dissertations affirm an interest in a new area of research, i.e., biological population and epidemiological modelling. Meanwhile, it must be noted that many research works have been done at M. Sc. level since the initiation of the Section of Statistical Mathematics at University of Nairobi in all these areas. As a special remark, it can be said that the period 1984-1987 is an era of publishing a bulk of articles on the four emerging research groups, namely group screening, educational and manpower planning, biological population modelling and AIDS modelling (Epidemiological modelling).

R. C. Bose was a man of reputation and a main factor of Patel's appreciation of Statistical Mathematics. His works, contributions and influence in design of experiments can be easily seen from Raghavarao's book. Raghavarao(1971) through out his work "Constructions and Combinatorial Problems in Design of Experiments" implicitly and explicitly confirms that he is presenting in more orderly way Bose's method and scholarly studies. Particularly, in chapter 9, "Graph Theory and Partial Geometries," the works of Bose, Bruck and Hoffman are discussed. With respect to partial geometry (r,k,t), Bose(1963), following Bruck(1951,1963), after giving its definition, like Kolmogorov who axiomatized probability theory, axiomatizes partial geometry for the purpose of designs of experiment. Raghavarao (1966) extended the scope of partial geometry to include three-associate geometric designs. The contribution of Hoffman(1963, 1965) is the characterization of projective planes, affine planes, and symmetrical BIB designs, in which the characteristic roots and their multiplicities of the adjacency matrices of their line graphs determine the corresponding configurations up to isomorphisms, excepting the case of the symmetrical BIB design with the parameters v=b=4, r=k=3, l=2.

We should also note that Patel, while he was in Kenya, 1968-1993, contributed a lot in the research of group-screening with more than two stages and in collaboration with his students put to light many academic research works on different areas of statistics.

2.3.2 Doctoral Dissertations

Names	Title	Year	Comment on the method of Approach
J. A. M. Otticno	Two Stage Group Screening Designs	1981	Analysis of optimality of group testing designs
J. W. Odhiambo	Three Stage Group Screening Designs	1982	Analysis of optimality of factorial designs
J. K. Arap Koske	Fourth Order Rotatable Designs	1984	Matrix algebraic approach
F. Njui	Fifth Order Rotatable Designs	1985	Non-linear and linear regression, matrices
M. M. Manene	Further Investigations of Group-Screening Designs: Step-wise Designs	1985	Analysis of optimality of group testing designs
J. Owino	A Stochastical Model for Stocks and Flows of Students in an Education System	1991	Matrix algebraic approach
K. N. Gacii	On the Construction of Deletion Designs	1993	Geometric approach
F. Onyango	On Theory of Random Search	1995	The methods used are PG(2,s) & EG(2,s)
H. G. Mwambi	Generalized Matrix and Compartimental Population Models	1997	Analytic - geometric approach
R. O. Simwa	Mathematical and Statistical Analysis of HIV/AIDS Epidemic with reference to Kenya and Uganda	1997	A parametric and non parametric statistical approach

2.3.3 Research Papers and Current Research Interests

a) Research Papers

Name of Researchers	Published Articles: title, name of periodical, date
Adhikary, A. K. & Chaudhuri, A.	"A note on handling linear randomized response," *Journal of statistical planning and Inference*, Vol. 22(1989), p. 263. "On Two properties of an unequal probability sampling scheme," *Metika*, 36(1989), p.161 "Variance estimation with randomized response." *Comm. Statist. -Theory Meth.*, 19(3), (1980), p.1119. "A note on interpreting subsamples of unequal sizes drawn with and without replacement," *Comm. Statist. -Theory Meth.*, 19(4), (1990), p. 1475.
Adhikary, A. K.	"On the performance of the nearest proportional to size sampling design," *Comm. Statist. -Theory Meth.*, 20(21), (1991), pp. 3933 -3941.
Gachii, K. M. & Odhiambo, J. W.	"Deletion designs in estimation of low order interactions," *The Fifth Scientific Conference of the East Central & Southern Africa Network of the International Biometric Society*, 22nd -25th September 1997(Kenya), pp. 20-22.
Getao, J. L. and Odhiambo, J. W.	"The potential of information technology in the management of an African crisis: Computers and AIDS," *Global Information Technology and Socio-Economic Development*, (Ivy League Publishing, 1996), pp. 53-59.
Koske, J. K. A.	"Response Surface Designs with missing observations," *Biometry for Development Proceedings of the First Scientific Meeting of the Biometric Society, Kenya Group and East/Central African Network* , April 2-6, 1990(Nairobi: ICIPE Science Press), pp. 51-54. "The variance function of the difference between two estimated fourth order response surface," *J. S. P. I.* (1989), pp. 263-266.
Luboobi, L. S. & Simwa, R. O.	"HIV/AIDS epidemic curves for Kenya and Uganda: A parametric statistical approach," *The Fifth Scientific Conference of the East Central & Southern Africa Network of the International Biometric Society*, 22nd -25th September 1997(Kenya), pp. 39-43. "HIV/AIDS epidemic curves for Kenya and Uganda: A nonparametric statistical approach," *The Fifth Scientific Conference of the East Central & Southern Africa Network of the International Biometric Society*, 22nd -25th September 1997(Kenya), pp. 44-50.
Manene , M. M.	"On two-type stepwise group screening designs," *The Fifth Scientific Conference of the East Central & Southern Africa Network of the International Biometric Society*, 22nd -25th September 1997 (Kenya), pp. 57-62.
Mwambi , H. G., Odhiambo, J. W. & Duchateau, L.	"A multiple matrix model to study the population dynamics of R. appendiculatus in Zimbabwe," *The Fifth Scientific Conference of the East Central & Southern Africa Network of the International Biometric Society*, 22nd - 25th September 1997(Kenya), pp. 65-71.
Munyinyi, D. & Nokoe, S.	"Estimation probability distribution function for female ticks on unimproved Zebu Cattle," *Biometry for Development Proceedings of the First Scientific Meeting of the Biometric Society, Kenya Group and East/Central African Network*, April 2-6, 1990 (Nairobi: ICIPE Science Press), pp. 93-98.
Patel, M. S.	"Group-screening with more than two stages," *Technometrics*, 4(2), (1962), pp. 209-217. "A critical look at two stage group-screening method," *Kenya Journal of Science and Technology*, Vol. 4, No. 2(1983). "Group screening for isolating defective factors of a population," *Biometry for Development Proceedings of the First Scientific Meeting of the Biometric Society, Kenya Group and East/Central African Network*, April 2-6, 1990 (Nairobi: ICIPE Science Press), pp. 84-92.

Patel, M. S. & Arap Koske, J. K	"Conditions for fourth order ratatability in *k* dimensions," *Comm. Statist. - Theory Meth.* (14)(6), (1985), pp. 1343-1351.
Patel, M. S. & Ottieno, J. A. M.	"Two stage group-screening designs with unequal a-priori probabilities," *Comm. Statist. -Theory Meth.*, 13(6), (1984), pp. 761- 779. "Optimum Two-stage group-screening designs," *Comm. Statist. -Theory Meth.*, 13(21), (1984), pp. 2649-2663. "Optimum Two-stage group-screening with unequal group sizes and errors in decisions," *Comm Statist. -Theory Meth.*, 14, (1987), pp. 799-820. "Optimum Two-stage group-screening designs with unequal a-priori probabilities and with error in decisions," paper presented at the joint statistical meetings of A.S.A., E.N.A.R., WNAR, IMS at Las Vegas, USA 5th - 8th August, 1985
Patel, M. S. & Manene, M. M.	"Step-wise group screening with equal prior probabilities and no errors in observations," *Comm. Stat. -Simulation Computa.* (1987).
Odhiambo, J. W.	"The performance of multistage group screening designs," *Comm Statist. - Theory Meth.*, 15 (1986), pp.2467-2481. "A review of the factor screening method," *Proceedings of the 1st Conference of the Kenya Mathematical Society*, 19-21 August, 1992, (1993), pp. 54 -57.
Odhiambo, J. W. & Kimanani, E.	"A logistic model for the sterile male technique," *Biometry for Development Proceedings of the First Scientific Meeting of the Biometric Society, Kenya Group and East Central African Network*, April 2-6, 1990(Nairobi: ICIPE Science Press), pp. 176-186.
Odhiambo, J. W & Patel, M. S.	"Three-stage group screening with error in observations," *Comm Statist -Theory Meth.*, 14, (1985), pp. 647- 666. "On multiple group screening designs," *Comm. Statist. -Theory Meth.*, 15(5), (1986), pp.1627-1645. "Three-stage group screening with unequal group sizes and with errors in observations," *Comm. Statist. -Theory Meth.*, 16(10), (1987), pp. 2957-2979.
Odhiambo, J. W. & Getao, K. W.	"The potential of group screening method in the management of AIDS crisis in Africa," *The Fifth Scientific Conference of the East Central & Southern Africa Network of the International Biometric Society*, 22nd - 25th September 1997(Kenya), pp. 81-84.
Odhiambo, J. W. & Manene, M. M.	"Step-wise group screening designs with errors in observations," *Comm. Statist. - Theory Meth.*, 16(10), (1987), p. 3095ff.
Odhiambo, J. W. & Khogali, A. K.	" A transition model for estimating academic survival through cohort analysis," *Int. J. Sci. Technol.*, 17(1986), pp. 339-346.
Odhiambo, J. W. & Owino, J. O	"A stochastic model for estimating academic survival in an education system," *Kenya J. Sci. Technol.*, A6(1), (1985), pp. 59-67.
Ottieno, J. A. M.	"Mortality levels and determinants in Kenya," *Proceedings of the 1st Conference of the Kenya Mathematical Society*, 19-21 August, 1992, (1993), p. 73 [abstract].
Owino, J. O. & Philips, C. M.	" A comparison of retention properties of the Kenya primary education system before and after 1970," *Kenya J. Sci. Technol.*, A9, (1988), pp. 5-10.
Owino, J. O.	"A mathematical model for comparison of educational characteristics of males and females," *Proceedings of the 1st Conference of the Kenya Mathematical Society*, 19-21 August, 1992, (1993), pp. 61-63.
Owino, J. O. & Odhiambo, J. W.	"A statistical method for planning an educational system," *Discovery and Innovation*, 6(2), (1994), pp. 140-144.
Owino, J. O. & Omolo, B. O.	"Optimal harvesting in poultry farming," *The Fifth Scientific Conference of the East Central & Southern Africa Network of the International Biometric Society*, 22nd -25th September 1997(Kenya), pp. 103-107.
Weke, P. G. O.	"IBNR claims reserving and GLIM," *Proceedings of the 1st Conference of the Kenya Mathematical Society*, 19-21 August, 1992, (1993), pp. 67-69.

b) Current Research Interests

Gacii, K. N.	Designs of experiments, Deletion designs
Manene, M. .M.	Group-screening designs mathematical modelling
Mwambi, A. H.	Matrix modelling for population dynamics, Matrix modelling
Njui, F.	Designs of experiments with emphasis on rotable designs, regression analysis, response surface designs
Odhiambo, J. W.	Mathematical Modelling induction and man power planning; theoretical ecology; group screening designs, factor screening
Ottieno, J. A. M.	Statistical demography and quality control theory
Owino, J.	Stochastic model of educational planning, mathematical modelling of social processes
Simwa, R. O.	AIDS modelling or epidemiological modelling with a parametric and non parametric statistical approach

2.3.4 Reference Books / Text Books

Courses/Units	Reference/Text Books
Measure Theory and Probability	Chow, Y. S. and Teicher, H. *Probability Theory: Independence, Interchangeability, Martingales*, 1978. Loeve, M. *Probability Theory*, 1963, 1978. Kingman, J. F. and Taylor, S. J. *Introduction to Measure and Probability Theory*, 1966 Halmos, P. R. *Measure Theory*, 1950.
Tests of Hypothesis	Ferguson, T. N. *Mathematical Statistics: A Decision Theoretic Approach*, 1967 Lehmann, E. L. *Testing Statistical Hypothesis*, 1959 Wald, A. *Statistical Decision Functions*, 1950, 1971 Savage, L. J. *The Foundations of Statistical Inference*, 1962. Vorober, N. N. *Game Theory*, 1977.
Theory of Estimation	Zacks, A. *The Theory of Statistical Inference*, 1971. DeGroot, M. H. *Optimal Statistical Decisions*, 1970. Cox, D. R. and Hinkley, D. V. *Theoretical Statistics*, 1974
Multivariate Analysis	Anderson, T. W. *An Introduction to Multivariate Statistical Analysis*, 1958 Mardia, K. V., Kent, J. T., and Bibby, J. M. *Multivariate Analysis*, 1979 Khirsagar, A. M. *Multivariate Analysis*, 1972. Morrison, D. F. *Multivariate Statistical Methods*, 1976. Kendall, M. G. *Multivariate Analysis*, 1975.
Analysis of Variance	Scheffe, H. *The Analysis of Variance*, 1959. Kempthorne, O. *The Design and Analysis of Experiments*, 1952. Cochran , W. G. & Cox, G. M. *Experimental Designs*, 1957.
Design of Experiment	John, P. W. M. *Statistical Design and Analysis of Experiments*, 1971. Raghavarao, D. *Constructions and Combinatorial Problems in Design of Experiments*, 1971. Raktoe, B. L. et al., *Factorial Designs*, 1981
Probability and Stochastic Processes	Feller, W. *Introduction to Probability Theory and Its Applications*, Vol. I and II , 1993 Medhi, J. *Stochastic Processes*, 1982 Doob, J. L. *Stochastic Processes*, 1953. Taylor, H. M. and Karlin, S. *A First Course in Stachoastic Processes*, 1975. Bhat, U. N. *Elements of Applied Stochastic Processes*, 1984. Cox, D. R. The Theory of Stochastic Process, 1965
Time Series Analysis	Anderson, T. W. *The Statistical Analysis of Time Series*, 1971 Wold, H. O. A. *A Study in the Analysis of Stationary Time Series*, 1954. Chatfield, C. *The Analysis of Time Series*, 1987. Kendall, M. G. *Time-Series*, 1976. Wold, H. O. A. *A Study in the Analysis of Stationary Time Series*, 1954.
Sample Surveys	Cochran, W. G. *Sampling Techniques*, 1992. Muthy, M. N. *Sampling Theory and Methods*, 1967.

	Hansen, M. H., Hurwitz, W. N. and Madow, W. G. *Sample Survey Methods and Theory*, Vols I and II, 1953
Non-Parametric Methods	Fraser, D. A. *Nonparametric Methods in Statistics*, 1957 Gibbons, *Non-parametric Statistical Inference*, 1976 Gottfried E. Noether, *Elements of Nonparametric Statistics*, 19 Hollander, M. and Wolfe, D. A. *Nonparametric Statistical Methods*, 1973. Lehmann, E. L. *Non Parametrics:Statistical Methods based on ranks*, 1975 Maritz, J. S. *Distribution-Free Statistical Methods*, 1981

2.3.5 Tentative conclusion

The pioneers or the initiators of the statistical section of mathematics at Nairobi University are Professor M. S. Patel(1968) and his colleagues. Basing our observation on the articles published as concrete documents and dissertations as further confirmative works, we note the *model techniques or design theory* has gained ground at University of Nairobi. Furthermore, categorizing broadly, it can be said that four research groups are emerging: (i) *group screening*, (ii) *educational and manpower planning*, (iii) *biological population modelling*, and (iv) *AIDS modelling* (Epidemiological modelling).

Since the present work is the first survey of its kind, it is better to be frank with respect to the decisive achievement. The primary impulse is that, it calls to reflection and have a critical philosophical assessment of one's approach and methodology in statistics and probability; and if possible, to invite for the study of logic and philosophical history of mathematics, probability and statistics.

We can note two trends in making of tentative conclusions about possibility of identification of Statistical Section of University of Nairobi with a specific school or many schools of statistics and probability: impediments and plausibilities. The two aspects of observations, for the sake of conciseness and clarity, can be presented in a summary form.

The following observations are deduced through the formal discussions and as an outcome of informal inquiry:

(i) Impediments

- The absence of a strong conviction or attachment to a specific method of approach or school of philosophy of probability.

- The trend to accept randomness as it is and absence of any venture on the epistemological meaning of the term in probability and statistics
- The lack of a course of mathematical philosophy or systematic philosophical and historical analysis of probability theory and statistical mathematics.
- The absence of clear philosophical training policy of mathematics, in particular of probability theory and statistical mathematics, at Public Universities; focusing mainly on job orientation courses or stressing on the income generating policy. Generally, there is a pragmatic approach to training policy.

(ii) Plausibilities

- The presence of philosophical and personal reflection on research works - individual conviction.
- The background of founders and their method of approach to probability theory and statistical mathematics.
- The courses offered in relation to the foundation of probability and related topics, especially Measure Theory and Probability, Statistical Inference.

From the criteria used above, namely, trying to identify the Statistical Section of Mathematics at University of Nairobi, if it is possible, as a school with respect to the historical development and dialectic progress of schools of statistics and probability using the criteria: (i) The background of the founders, (ii) Text books, reference books and hand-outs used in the units or courses offered, (iii) Research papers and dissertations, we can give a tentative conclusion.

The affinity is much more expressed through the methodology of research works and frequently books used. This relation is apparent with respect to the subjectivistic or degree of belief approach. A concrete example can be quoted for clarity: DeGroot(1970), in his book "Optimal Statistical Decisions", states that "Subjective, or Bayesian, statistical decision theory is applicable to those problems in which the information and uncertainty about the parameters can, at any time, be summarized by a probability distribution of their possible values. Therefore, this book will deal only with those statistical decision problems which meet the following two requirements: (1) The conditions can be formulated in terms of a manageable number parameters. (2) Although the values of these parameters are not known exactly, any uncertainty about the values can be represented by a suitable probability distribution. ... It will be assumed in all problems in this book that each parameter can be assigned a particular probability distribution."[DeGroot, 4]

The other supportive motivation for this deduction, with the conviction among the current scholars present at the university, is a tendency that the degree of belief theory answers a lot of vague ideas and can be justified using conceptual and mathematical questions. This influence can be seen in the studies of probability theory related to the Bayesian statistics decision theory. The definition of statistical theory itself gives a clue: in modern formulation of statistical theory it is generally held that Statistic is a science which deals with decision making in the case of uncertainty.

But, also the preference of asymptotic approach is cited from research work as well as from applications of some probabilistic and statistical principles. Chow and Teicher(1978) in their book "Probability Theory: Independence, Interchangeability, Martingales", after describing the two important approaches of probability theory, give their concern and method of study of probability theory: "The concern of this book is with the measure-theoretic foundations of probability theory and the body of laws and theorems that emerge thereform." They believe that "the frequency approach appears to have lost out to the measure-theoretic." This possibility is accepted by the prominent scholars and there is high esteem for Kolmogorov's work too.

The results of the external criticism, deducted using the above criteria, may not lead to the actual situation at the ground. We should note that complete certainty have never been the trade mark of a scientific fact, although it is the primary duty of scientific endeavour to minimise the uncertainty as much as possible. As Prof. J. W. Odhiambo said *now lecturers are more preoccupied with the issue of survival rather than on how to think.*

Chapter Three

NORMAL DISTRIBUTION: PROPERTIES AND ITS CHARACTERISATIONS

3.1 Historical Perspective of Normal Distribution

3.1.1 Beginners

3.1.1.1 De Moivre - Laplace

The investigations by Eggenberger(1894), Pearson(1924,1925,1926,1929), Archibald (1926), and Sheynin (1966,1968,1970a,1970b) reveal that Abraham de Moivre(1667-1754) was the first to derive the normal law. De Moivre's theorem(1730) or limit of Bionomial and his contribution to the development of normal distribution is dealt with in detail in chapter 3, section 3.2.2.1. Laplace (1749-1827) is mainly remembered in probability theory by his proof of one of the most important limit theorems. This theorem deals with the distribution of deviations of the frequency of occurrence of an event in a sequence of independent trials from its probability. This theorem is called de Moivre-Laplace theorem, since the particular case of $p = ½$ was obtained by de Moivre.

Laplace(1812) states the theorem as follows:

> Let the probability of the occurrence of a given event E in n independent trials be equal to p ($0 < p < 1$) and let m be the number of trials in which event E actually occurred; then the probability of the inequality $z_1 < (m-np) / (npq)^{½} < z_2$ ($q = 1-p$), differ by an arbitrarily small amount from $(2\pi)^{-½}\int_{z_1}^{z_2} \exp(-z^2/_2)\, dz$, provided n is sufficiently large [the integral theorem of Laplace, or global Laplace theorem]. The probability of exactly m occurrences of event E in n trials is approximately equal to $(2\pi npq)^{-½} \exp(-z^2/_2)$, where $z = (m-np) / (npq)^{½}$ (Local Laplace Theorem).

Laplace attributed great importance to his theorem. He believed that his law explained completely the behaviour of random mass schemes to which, according to Laplace, the majority of real-world phenomena belong, that is the model based on his law is almost universal. Maistrov(1974) believes that only after this work of Laplace did the wide spread applications of probability theory become feasible as a scientifically justified method. It is Laplace's(1952) view that all the regularities of any field of mass phenomena are reducible to the unique normal law, as the celestial phenomena are reduced to the unique law of universal gravitation. Based on this point of view he attempts to apply probability theory to court procedures, decisions at gatherings, and so on. This is misinterpretation of the far-reaching conclusions of his contributions to probability theory. His logical deduction is that he considers the history of human society as a field governed by pure chance and therefore assumes that probability theory is the science capable of rendering a complete analysis and explanation of this history, so that the analysis of social phenomena falls within the realm of probability theory. This law was termed *normal law* by Henri Poincaré(1854-1912) a mathematician and physicist, accepted as a direct predecessor of the founders of axiomatization in probability theory.

3.1.2 *Advancement*

3.1.2.1 Adrian - Gauss

We observe that random errors occur in the course of observations of any kind. The problem of how to avoid them or at least to cope with them has attracted the attention of scientists since early times. However, this can be solved satisfactorily only by means of probabilistic methods. This problem was considered in detail in the early nineteenth century. Two mathematicians, Robert Adrian(1775-1843), American and Carl Friedrich Gauss (1777-1855), German, independently and almost simultaneously obtained the basic result, the derivation of the normal "law for the distribution of random errors," the so-called "law of errors". They reached their result using different methods.

Adrian deduced independently and published in 1808, before Gauss, the famous law of errors," which was published in 1808 in *The Analyst, or Mathematical Museum*, and serves as the foundation of present-day classical probability and statistics. He was solving a particular problem and, when generalizing it, obtained the distribution law of random errors. The interesting parts of this paper are two derivations of the *normal law* for the distribution of random errors in observations:

(i) Using the a certain distance measure:

The resulting function is

$$U = \exp\,(a_1 + mx^2/2a),$$

which is called by Adrian "the general equation of the curve of probability," for $m < 0$.

(ii) Considering the measurements of a segment AB with equally probable errors in the length and in the azimuth:

The result that of the probabilities of errors are $\exp(c + \frac{1}{2}nx^2)$ and $\exp(c + \frac{1}{2}ny^2)$. For detailed description refer to section 3.2.2, of chapter 3.

Gauss on the other hand, was investigating the general theory of errors in observations and the normal distribution of random errors became a necessary and most important part of this theory. The derivation of the distribution of random errors as given by Gauss served as a basis for further development of the theory of errors.

Gauss published his derivation of the normal law of distribution of random errors in observation in 1809 in his famous work *Theoria motus corporum coelentium in sectionibus conicis solem ambientium*. The derivation of the normal law followed the following assumptions: given equidistant observations of a certain quantity, let the random errors posses the differential density of probability distribution $\varphi(\Delta)$. It is required to determine $\varphi(\Delta)$ under the assumption that the most probable value of the quantity under consideration is equal to the arithmetic mean of the observed values. Gauss obtains that the function $\varphi(\Delta)$ is given by

$$\varphi(\Delta) = h\,(\pi)^{-\frac{1}{2}}\,exp\,(-h^2\,\Delta^2\,).$$

He denotes the value h as "the measure of precision of the observations." According to this law, errors of any magnitude are possible.

The main stimuli for his result is his occupation with problems of astronomy and geodesy and was able to develop methods of processing results of observations. The results of these observations are not immune of errors. Thus the problem of determining the most probable value of the observed quantity arises. These problems led Gauss to develop the theory of errors, which is directly connected with the ideas and notions of probability theory.

3.1.2.2 Chebyshev and his followers

Chebyshev in his paper "On integral residua which yield approximate values of the integrals," 1874, obtained the Laplace-de Moivre distribution using the moment of the function approach. Kolmogorov(1948) remarks that the results of Chebyshev's investigations on the problem of moments are applied here to the determination of the form of the probability distribution law of a sum of a large number of independent random variables and it is established that, under certain very general conditions, this distribution law, with the increase in the number of summands, approaches in the limit the normal distribution law of de Moivre-Laplace [the so-called basic limit theorem of probability theory]; moreover the possibility of a further refinement of its result is pointed out in this paper, although without a rigorous proof. In his last paper on probability theory "On two theorems concerning probabilities" (1887), Chebyshev actually summarizes all his reasearch in this field. First he states the first of his theorems - the law of large numbers. In the second theorem, one of his most important results, he established that under certain very general conditions the distribution law of probabilities of the sum of a large number of independent random variables approaches the normal distribution in the limit as the number of summands increases. Markov and his companion, like Lyapunov, advanced the understanding of normal law by using *the* *method of moments* and especially by *the method of characteristic functions*. Further clarification is given in section 3.4.2. chapter 3.

3.1.3 Solidification and Applications

Soon after the derivation of the normal distribution, we observe that, it some how became the focal point of the study of probability theory and for application of statistical methods in several branches of physical science, and even for social sciences.

Its central importance in statistics stems from three facts: (i) many actual populations approximate closely to normal forms. (ii) It forms the limiting distribution of many widely used statistics. (iii) Under general conditions, the role of normal distribution as a limit, the asymptotic behaviour, of distribution functions of normalized sums of random variables is widely accepted (Central Limit Theorems).

3.1.3.1 Lindeberg-Feller

Lyapunov's(1901) central limit theorem of probability theory is improved by Y. W. Lindeberg in 1922 in which he obtained a new sufficient condition, and in 1935 by W. Feller who showed the necessity of this condition. In his celebrated investigation of normal convergence, Lyapunov(1948) examined not only conditions for, but also the speed of this convergence. His results were greatly improved by Berry(1941), and independently by Esseen(1945). Bernstein(1939), commenting on Lyapunov's achievemnt, says a classical result which constitutes a culmination point of Lyapunov's investigation in probability theory.

This necessary and sufficient condition of Lindeberg(1922) and Feller(1935) is stated as follows: If $F_n(x)$ is the distribution function of X_n, and h is a fixed positive number, then, as $n \to \infty$

$$\frac{1}{n}\int_{|x|>h} x^2 \, d \sum_1^n F_i(x) \to 0.$$

This condition is necessary and sufficient for the convergence of the distribution of $(1/n)\sum_1^n X_i$ to the *normal distribution*.

As a final remark we can use Loève's(1963) words:

> The real liberation which gave birth to the Central Limit Problem came with a new approach due to P. Lévy. He stated and solved the following problem: Find the family of *all possible limit laws* of normed sums of independent and identically distributed random variables. ... Find conditions for convergence to *any specified law of this family* ... The solution of the problem is due to the introduction , by de Finetti, of the 'infinitely decomposable' family of laws and to the discovery of their explicit representation by Kolmogorov in the case of finite second moments and by P. Lévy in the general case. ... The final form is essentially due to Gnedenko. [Loève, 289-90]

3.1.3.2 Several Applications

Given that the normal distribution is widely used, and somewhat abused, in statistics, it is natural that the most familiar problems of inference are those which involve this distribution. But our present interest is to ponder through the applications of normal law or distribution in different fields.

The observation that in complicated situations where some kind of disorder prevails that something having the appearance of order often emerges gave rise to the applications or assumption of normal distribution in several fields. This phenomena *order out of chaos* draws a lot of attention and is expressed when considering the distribution of velocities in the kinetic theory of gases, the same variance corresponds to kinetic energy being constant, which suggests a connection with Maxwell's conclusions - see chapter three section 3.2.2.5 Francis Galton's(1889) observations about *regression analysis* also give tendency to the presence of the normal distribution in natural inheritance. At one moment, around 1900, Poincaré made an observation that everyone believes in normal law, and said that experimentalists believe that the normal distribution is a mathematical theorem, while mathematicians believe that it is an empirical fact. Even though it is an exaggerated view, it expresses the wide range of application of normal distribution in scientific investigations. The different methods utilized to derive the normal distribution themselves indicate the importance of the law.

The role of normal distribution in the various branches of statistics is widely recognized and appreciated. As Anderson(1958) points out in his book *An Introduction to Multivariate Statistical Analysis*, normal distribution is used as a model for analysing sampling theory, factor analysis. In these cases, as well as a host of others in agricultural experiments, in engineering problems, in certain economic problems, and other fields, the multivariate normal distributions have been found to be sufficiently close approximations to the populations so that statistical analyses based on these models are justified. Furthermore, as the central limit theorem leads to the univariate normal distribution for single variables, so does the general central limit theorem for several variables lead to the multivariate normal distribution.

Another basic reason that the application of normal distribution is accepted mathematically is that *normal theory* is amenable to exact mathematical treatment. The multivariate methods, which deal with the variety of problems, based on the normal distribution are extensively developed and can be studied in a rather organized and systematic way. The suitable methods of analysis are mainly based on standard operations of matrix algebra; the distributions of many statistics involved can be obtained exactly or at least characterized by moments; and in some cases optimum properties of procedures can be deduced.

It is worthy to use Maistrov's(1974) observation for the explanation of the role of normal distribution in science: the feasibility of replacing the exact distribution by its limit follows from the so-called *central limit theorem*. The essence of this theorem is establishing the conditions under which the distribution function of the sum of independent random variables approaches the *normal distribution* as the number of summands increases. There are in nature a vast number of phenomena subject to the action of a large number of causes where each individual cause acts independently and exerts only a very small influence on the course of the phenomenon. That is why this theorem is of such importance for the science.

3.2 Definitions and Derivations of Normal Distribution

3.2.1 Definitions

There are two approaches that are used to define a normal distribution: i) classical and ii) modern. In the classical perspective Normal Distribution is depicted through the density function. While in Modern approach the distribution is characterized in such a way that the concept involved can be extended into more complex random variables with countable and uncountable dimensions. Furthermore, Rao(1973) in his monumental book *Linear Statistical Inference and Its Applications* gives further explanations and defintions on normal distribution.

3.2.1.1 Classical Definition of Normal Distribution

The Univariate normal distribution function can be written as

$$k\,e^{-\frac{1}{2}\alpha(x-\beta)^2} = k\,e^{-\frac{1}{2}(x-\beta)\alpha(x-\beta)}$$

where α is positive and k is chosen so that the integral of $F(x) = \Pr\{X \leq x\}$, over the entire x-axis is unity. The *cumulative distributioh function* defined for every pair of real number x, where F(x) is absolutely continuous, implying that $dF(x)/dx = f(x)$, exists almost everywhere. It is assumed that F(x) has the following properties: (i) F(x) is nondecreasing, (ii)$F(-\infty)=0$, $F(\infty)=1$, and (iii) F(x) is continuous at least from the left defines a random variable of which F is the distribution function.

The density function of a multivariate normal distribution of $x_1, \ldots, x_p$, following Anderson(1958) we can say that has an analogous form. The scalar variable x is replaced by a vector $\underline{x} = (x_1, \ldots, x_p)'$; the scalar constant β is replaced by vector $\underline{\beta} = (\beta_1, \ldots, \beta_p)'$ and the positive constant α is replaced by a positive definite (symmetric) matrix

$$(\underline{x} - \beta)' A (\underline{x} - \beta) = \sum_{i,j=1}^{p} a_{ij} (x_i - \beta_j)(x_j - \beta_i).$$

Thus the density function of a *p*-variate normal distribution is

$$f(x_1, ..., x_p) = K e -1/2 (x- b)' A (x - b),$$

where $K > 0$ is chosen so that the integral over the entire *p*-dimensional Euclidean space of $x_1,...,x_p$ is unity.

3.2.1.2 Modern Definition of Normal Distribution

Next we shall deal with the modern approach in defining the multivariate normal distribution, and present two different definitions. It must be noted that the distribution is not defined by probability density function. It is characterized by the property that every linear function of the p-variables has a univariate normal distribution. Such characterization is exploited in deriving the distributions of sample statistics. We can note that corresponding to any known result in the univariate theory, the generalization to the multivariate theory can be written down with a little or no further analysis.

For example knowing the joint distribution of the sample mean and sample variance in the univariate theory we can write down the joint distribution of the sample means of multiple measurements and sample variances and covariances. The entire theory of multivariate tests of significance by analysis of dispersion is obtained as a generalization of the univariate analysis of variance. Through this method we can encounter a number of characterizations of the multivariate normal distribution which will be useful to study the theory of normal distributions in Hilbert and other more general spaces, like Banach space. It is worthy to note that multivariate normal distribution plays an important role in statistical inference involving multiple measurements.

Next, in our definition, we will follow Rao's method(1973) of presentation and explanation of the properties of normal distribution.

Definition - 1: A p-dimensional random variable $\underline{u}$, that is, a random vector $\underline{u}$ taking values in E_p (Euclidean space of p-dimensions) is said to have a p-variate normal distribution N_p if and only if every linear function of $\underline{u}$ has a univariate normal distribution.

This definition of multivariate normal distribution is inspired by the result due to Cramer(1937) and Wold(1938), which states that "the distribution of a p-dimensional random variable is completely determined by the one-dimensional distributions of linear functions $\underline{t}'\underline{u}$, for every fixed real vector $\underline{t}$."

This result is indicating that if a random vector $\underline{u}$ exists satisfying definition-1, then its distribution is uniquely determined. According to Rao(1973) this definition of N_p can be extended to the definition of a normal probability measure on more general spaces such as Hilbert or Banach spaces by demanding that the induced distribution of every linear functional is univariate normal.

Following the **definition-1**, the following properties of normal distribution can be listed:

a) Expected value, $E(\underline{u})$ and dispersion matrix, $D(\underline{u})$ exist which we denote by μ and Σ respectively. Further for a fixed vector $\underline{t}$, $\underline{t}'\underline{u} \sim N_1(\underline{t}'\mu, \underline{t}'\Sigma\underline{t})$, that is univariate normal with mean $\underline{t}'\mu$ and variance $\underline{t}'\Sigma\underline{t}$.

b) The characteristic function of $\underline{u}$ is $\exp(i\underline{t}'\underline{u} - \frac{1}{2}\underline{t}'\Sigma\underline{t})$.

c) The p-variate normal distribution is completely specified by the mean vector μ and the dispersion matrix Σ of the random variable, since the characteristic function involves only μ and Σ. We may therefore, denote a p-variate normal distribution by $N_p(\mu, \Sigma)$, involving μ and Σ as parameters.

- If there exits a vector and matrix such that for every $\underline{t}$, $\underline{t}'\underline{u} \sim N_1(\underline{t}'\mu, \underline{t}'\Sigma\underline{t})$, then

  $\underline{u} \sim N_p(\mu, \Sigma)$.

- If $\Sigma = \Lambda$ (a diagonal matrix), the components $u_1, \ldots, u_p$ are independent and each is univariate normal.

- Let u_1 and u_2 be two subsets of variable $\underline{u}$. We can write

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} D(U_1) & Cov(U_1, U_2) \\ Cov(U_1, U_2) & D(U_2) \end{pmatrix} \quad (3.1)$$

where Σ_{12} and Σ_{12} are the dispersion matrices of u_1, u_2, and Σ_{12} is covariance matrix of u_1, u_2. The random variables u_1, u_2 are independently distributed if and only if $\Sigma_{12} = 0$.

- If $u_1, u_2, \ldots, u_k$ of u are independent pairwise, they are mutually independent.
- The function $\exp(it'u - ½ t'\Sigma t)$ is indeed a characteristic function so that N_p of **definition-1** exists.
- $u \sim N_p(\mu, \Sigma)$ with rank k if and only if,

$$u = \mu + b\,g, \quad b\,b' = \Sigma$$

where b is $(p \times k)$ matrix of rank k and $g \sim N_k(0, I)$, that is, the components $g_1, g_2, \ldots, g_k$ are independent and each is distributed as $N_1(0,1)$.

d) If $u \sim N_p$, the marginal distribution of any subset of q components of u is N_q.

e) The joint distribution of q linear functions of u is N_q. If $y = c\,u$, where c is $(q \times p)$, represents the q linear functions, then

$$y \sim N_q(c\mu,\ c\,\Sigma c').$$

Definition-2: A p-dimensional random variable u is said to have a normal distribution N_k if it can be expressed in the form $u = \mu + b\,g$, where b is $p \times m$ matrix of rank m and g is a $m \times 1$ vector of independent N_1 (univariate normal) variables, each with mean zero and unit variance.

Observations:

a) The relationship $u = \mu + b\,g$ shows that the random vector $u \in M(\mu{:}\Sigma)$, the linear manifold generated by the columns of μ and Σ, with probability 1.

b) Let $u_i \sim N_{p_i}(\mu_i, \Sigma_i)$, $i=1, 2, \ldots, k$, be independent and T be a function of $u_1, \ldots, u_k$. Using the representation

$$u_i = \mu_i + b_i g_i,$$

we have

$$T(\underline{u}_1, \ldots, \underline{u}_k) = T(\mu_1 + \underline{b}_1 g_1, \ldots, \mu_k + \underline{b}_k g_k)$$

$$= t(g_1, \ldots, g_k).$$

Then the study of T, a statistic based on $\underline{u}_1, \ldots, \underline{u}_k$, reduces to the study of function of independent univariate normal variables $g_1, \ldots, g_k$. Such a reduction helps, as the known results in the univariate theory can be immediately applied or deduce the results in the multivariate theory.

c) It is possible to note that $\underline{u} = \mu + \underline{b}\, g$, with $\underline{u}_i = \mu_i + \underline{b}_i g$ as in **definition-2** and $\underline{u} = \mu + \underline{c}\, \underline{f}$ where $\underline{c}$ is $p \times q$ matrix and $\underline{f}$ is a q-vector of independent N(0,1) variables, have the same distribution if $\underline{b}\, \underline{b}' = \underline{c}\, \underline{c}'$, so that no restriction on q or $R(\underline{c})$ need be imposed. But a representation with restriction on $\underline{b}$ as in **definition-2** is useful in practical applications.

Following the **definition-2**, the following properties of normal distribution can be listed:

a) Let $\underline{u}'_1 = (\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_r)$, $\underline{u}'_2 = (\underline{u}_{r+1}, \underline{u}_{r+2}, \ldots, \underline{u}_p)$ be two subsets of $\underline{u}$ and Σ_{11}, Σ_{12}, Σ_{22} be the partitions of Σ as defined in (3.1). Then the conditional distribution of $\underline{u}_2$ given $\underline{u}_1$ is

$$N_{p-r}(\mu_2 + \Sigma_{21}\Sigma^{-}_{11}(\underline{u}_1 - \mu), \; \Sigma_{22} - \Sigma_{21}\Sigma^{-}_{11}\Sigma_{12})$$

where $E(\underline{u}_i) = \mu_i$, i= 1, 2, and Σ^{-}_{11} is a generalized inverse of Σ_{11}.

b) The reproductive property of N_p. Let $\underline{u}_i \sim N_p(\mu_i, \Sigma_i)$, i= 1, ..., n be all independent. Then for fixed constants $a_1, \ldots, a_n$,

$$\underline{y} = a_1 \underline{u} + \ldots + a_n \underline{u}_n \sim N_p(\sum a_i \mu_i, \sum a_i^2 \Sigma_i).$$

c) Let u_i, i=1,2, ..., n, be independent and identically distributed as. Then,

$$(1/n)\sum u_i = \overline{U} \sim N_p(\mu, 1/n\, \Sigma).$$

d) Let $\underline{u} \sim N_p(\mu, \Sigma)$. Then a necessary and sufficient condition that

$$Q = (\underline{u} - \mu)'\, \underline{a}(\underline{u} - \mu) \sim \chi^2(k)$$

is $\Sigma(a\Sigma\underline{a} - \underline{a})\Sigma = 0$ in which case $k = \text{trace}(\underline{a}\,\Sigma)$

3.2.2 Derivations of Normal Distribution

In the present section we will investigate the different methods or models used to derive the normal distribution function. One striking factor in the derivation of normal distribution is the fact that many scholars independently using different approaches arrived at the same conclusion As Poincaré said, there must be something mysterious about the normal law, since mathematicians think it is a law of nature, whereas physicists are convinced that it is a mathematical theorem.

Some general pattern for the "derivation of Normal Law" are dealt with by prominent scholars, like Kac(1959), Parratt(1961), Rao(1973), Maistrov(1974) and Mathai(1977).

3.2.2.1. De Moivre's Theorem - Limit of Binomial

Abraham De Moivre (1667 - 1754), a French-English mathematician, is best known for his investigation of the concepts of normal distribution and probable error, for his generalization of Cotes' theorem [$\exp i\theta = \cos\theta + i\sin\theta$], and anticipation of Stirling's approximation [$n! \approx (2\pi n)^{1/2} e^{-n} n^n$]. He was a friend of Edmund Halley(1656-1742) and Isaac Newton(1642-1727), and corresponded with Jean Bernoulli. His major works are *Philosophical Transactions*(from 1695 to 1715), *Doctrine of Chances*(1718), *Annuities Upon Lives*(1725), and *Miscellanea Analytica*(1730 - a compilation of his researches in trigonometry and calculus). His pamphlets on Stirling's approximation and the normal curve appeared in 1730 and 1733, respectively.

Let k be a number of successes in a sequence of Bernoulli trials with probability θ for success and define the random variable

$$X = \frac{k - n\theta}{[n\theta(1-\theta)]^{1/2}},$$

then the limit distribution function is *normal*.

The characteristic function of k is

$$E(e^{itk}) = (1 - \theta + \theta e^{it})^n,$$

and the characteristic function of $f_x(t)$ of x is

$$E(e^{itX}) = E(e^{it(k-n\theta)/[n\theta(1-\theta)]^{1/2}})$$

$$= e^{-itn\theta/[n\theta(1-\theta)]^{1/2}}\, E(e^{itk/[n\theta(1-\theta)]^{1/2}})^n$$

$$= e^{-itn\theta/[n\theta(1-\theta)]^{1/2}}\, E(1-\theta+\theta e^{it/[n\theta(1-\theta)]^{1/2}})^n$$

$$= [(1-\theta)e^{-itn\theta/[n\theta(1-\theta)]^{1/2}} + \theta e^{it/[n\theta(1-\theta)]^{1/2}}]^n \qquad (3.2)$$

We note that

$$e^{iz} = 1 + (iz) + \ldots + (iz)^r / r! + 0(z)^r,$$

and expanding the exponential inside (3.2), we find

$$f_x(t) = [\, 1 - t^2/2n + 0(t^2/n)\,]^n \rightarrow \exp[-t^2/2\,], \quad \text{as } n \rightarrow \infty$$

which is the characteristic function of N(0,1). Hence by the continuity theorem, the limit of distribution function of X is normal. In effect the result means that for large *n*, the distribution function of the binomial variable *k* can be approximated by the distribution function of normal variable with mean $n\theta$ and variance $n\theta(1-\theta)$.

3.2.2.2 Adrian's Methods

Robert Adrian (1775-1843), an American, published his results in 1808 in *The Analyst*, or *Mathematical Museum*. As Maistrov(1974) noted the most interesting parts of this paper are two derivations of the normal law for the distribution of random errors in observations.

Let AB be the true value of any quantity, for example of a certain distance. The measure of this quantity is A*b*, the error being *b*B (fig. 1).

Let AB, BC, ... be several successive distances of which the value by measure are A*b*, *bc*, ... , the whole error being C*c*; now suppose the measures A*b*, *bc*, are given and also the whole error C*c* (fig.1). Adrian Assumes 'as an evident principle' that the errors in measurements of AB, BC are proportional to their lengths. Introducing the notation A*b* = *a*, *bc* = *b*, C*c* = *c*; and denoting the errors of the measures Ab, bc by x, y, respectively, we obtain for the 'greatest probability' the equation $x/a = y/b$. Let X and Y be the probability of the mutual occurrence of these errors in equal to XY. It is required to find X and Y under the condition that the probability Xy will be maximal

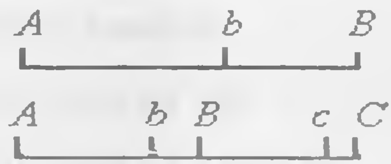


figure 1:Measure of distance

Introducing the notation

$$f(x) = \ln X,\ \varphi(y) = \ln Y$$

Then the maximum of XY corresponds to the relation

$$f(x) + \varphi(y) = \max.$$

Differentiating the last relation, we obtain

$$f'(x)\,x' + \varphi'(y)\,y' = 0 \Leftrightarrow f'(x)\,x' = -\varphi'(y)\,y'.$$

As we can observe, and is also noted by Maistrov(1974), Adrian does not indicate with respect to which argument tthe derivative is taken. He does not even mention the arguments of the function. All this respects, of course, a certain defect in his work. But for the maximal probability

$$x + y = \text{const}$$

and

$$x' + y' = 0 \Leftrightarrow x' = -y'$$

Dividing the equations we obtain

$$f'(x) = \varphi'(y).$$

Now this equation ought to be equivalent to

$$x/a = y/b.$$

This is satisfied in the simplest form if

$$f'(x) = mx/a$$

and

$$-\varphi'(y)\,y' = my/a.$$

Consider the first relation

$$f'(x) = mx/a \text{ or } df(x) = (mx/a)\,dx;$$

$$\int df(x) = \int (mx/a)\,dx$$

$$f(x) \quad = a_1 + mx^2 / 2a$$

$$f(x) \quad = \ln X = a_1 + mx^2 / 2a$$

$$X \qquad = \exp(a_1 + mx^2 / 2a).$$

The function

$$U = \exp(a_1 + mx^2 / 2a)$$

is called by Adrian "the general equation of the curve of probability." Next he proves that $m < 0$.

In the same paper Adrian presents a second derivation of the distribution law for random errors in observations. In this derivation he considers the measurements of a segment AB with equally probable errors in the length and in the azimuth. Adrian assumes that the locus of the equal probability of the location of point B, determined by the measurements of the length of AB, is the simplest curve, i.e. a circle with the centre at point B. Under these conditions he obtains that the probabilities of errors are $\exp(c + \frac{1}{2}nx^2)$ and $\exp(c + \frac{1}{2}ny^2)$ correspondingly, where x and y are the errors, $Bm = x$, $mn = y$, and $c = \text{const}$(fig. 2)

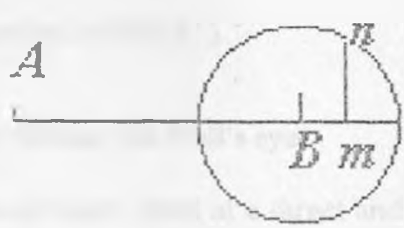


figure 2: Measurement of length

In the same article by Adrain the derivation of the least-squares principle is given as well as the derivation of the principle of the arithmetic mean, and a method "to correct the dead reckoning at sea by an observation of the latitude" are presented.

3.2.2.3 Theory of Error - Hagen's Hypothesis

Hagen based his proof of the normal law of error under the following assumptions:

(i) An error is the sum of a large number of infinitesimal errors, all of equal magnitude, due to different causes.

(ii) The different components of errors are independent.

(iii) Each component of error has an equal chance of being positive or negative.

By assumption (iii), each component of error takes the value $\pm\varepsilon$ with probability ½ for each, so that the mean is zero and the variance is ε^2. If

$$x = \varepsilon_1 + \dots + \varepsilon_n$$

is the total error due to n independent components, then

$$E(x) = E(\varepsilon_1) + \dots + E(\varepsilon_n) = 0$$

$$V(x) = \sum V(\varepsilon_i) = n\,\varepsilon^2 = \sigma^2.$$

Let us find the limiting distribution of x as $n \to \infty$ and $\varepsilon \to 0$ in such away that σ^2 is finite and fixed. The characteristic function of ε_j is

$$\frac{1}{2}\left(e^{it\varepsilon} + e^{-it\varepsilon} \right),$$

and that of $x = \varepsilon_1 + \dots + \varepsilon_n$ is

$$\left(1 - \frac{t^2\varepsilon^2}{2!} + \frac{t^4\varepsilon^4}{4!} \dots \right)^n = \left[1 - t^2\sigma^2/2n + 0(t^2/n) \right]^n \to e^{-t^2/2}, \text{ as } n \to \infty,$$

which is the characteristic function of $N(0,\sigma^2)$

3.2.2.4 Herschel's Hypothesis -Hitting the Bull's eye

Consider a distribution of shots fired at a target and let (x,y) be the co-ordinates (random variables) representing the deviation(errors) of a shot with respect to two orthogonal axes through the target point. Let the following hypotheses be true:

(i) The marginal density functions p(x), q(y) of the errors X and Y are continuous.

(ii) The probability density at (x,y) depends only on the distance $r = \sqrt{x^2 + y^2}$ from the origin (radial symmetry).

(iii) The errors in x and y directions are independent. Then the probability density function of the deviation Z in any direction is the normal density $\exp\{-z^2/2\sigma^2\}/\sigma\sqrt{20}$.

Using (ii) and (iii) the density at (x,y) is

$$p(x)q(y) = s(r), \quad r^2 = x^2 + y^2. \tag{3.3}$$

Putting x = 0, we find that the functions s and q are proportional to each other, while putting y = 0, we find that s and p are proportional to each other. Therefore, the functional equation (3.3) reduces to, writing

$$f(x)=\log\left[\frac{p(x)}{p(0)}\right],$$

$$f(x)+ f(y) = f(r), \quad r^2 = x^2 + y^2. \tag{3.4}$$

Further,

$$f(x) = f(-x) = f(|x|),$$

obtained by putting y=0, x = -x in (3.4). Thus, if

$$x^2 = x_1^2 + x_2^2,$$

$$f(r) = f(y) + f(x_1) + f(x_2), \qquad r^2 = y\, x^2 = y^2 + x_1^2 + x_2^2,$$

and so in general

$$f(r) = f(x_1) + ... + f(x_k), \qquad \sum x_i^2 = r^2.$$

Choosing $k = n^2$ and putting $x = x_1 = \ldots = x_2$, we see that

$$f(nx) = n^2 f(x)$$

or

$$f(n) = n^2 f(1) \qquad \text{for } x = 1.$$

For x = m/n where m is an integer,

$$n^2 f(m/n) = f(nm/n) = f(m) = m^2 f(1)$$

or

$$f(m/n) = c(m/n)^2,$$

where $c = f(1)$, so that $f(x) = cx^2$ for all rational x, and because of continuity the relation is true for all x. Hence

$$p(x) = p(0)\exp[cx^2]. \qquad (3.5)$$

For (3.5) to be a probability density, c must be negative and may be written as $-1/2\sigma^2$. Integrating from $-\infty$ to ∞ and equating the result to unity we find

$$p(0) = 1/(\sqrt{2\pi\sigma^2}),$$

so that

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}x^2} \qquad (3.6)$$

which is the well celebrated *normal distribution*, $N(0,\sigma^2)$ with $E(x) = 0$ and $v(x) = \sigma^2$.

The joint probability density of the errors X, Y is

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

The error in any direction$(\cos\theta, \sin\theta)$ is

$$Z = X\cos\theta + Y\cos\theta.$$

To find the probability density of Z, we consider the transformation

$$z = x\cos q + y\cos q,$$

$$z = x\cos q - y\cos q.$$

The Jacobian of the transformation $\partial(z,u) / \partial(x,y) = 1$. The density transforms to

$$\exp[(-(z^2 + u^2) / 2\sigma^2)] / (2\pi\sigma^2),$$

which shows that U and Z are independent and p.d. of Z is

$$\exp[-z^2 / 2\sigma^2],$$

which is the required result.

3.2.2.5 Maxwell's Hypothesis

Maxwell arrived at the normal distribution in deriving the distribution of velocities of molecules under the following assumptions:

(i)The components of velocity u, v, w in three orthogonal directions are independently distributed.

(ii) The marginal distributions of u, v, w are the same

(iii) The phase space is isotropic, that is, the density of molecules with given velocity components is a function of total velocity and not the direction.

If $f(.)$ denotes the probability density of any component of velocity, the assumptions (i) to (iii) lead to the functional equation

$$f(u)\,f(v)\,f(w) = g(V)\,, \qquad V^2 = u^2 + v^2 + w^2.$$

Thus, $f(u)$ is of the form (3.6) which is normal distribution for any single component of velocity and

$$g(V) = \text{const.}\ \exp[-\alpha\,(u^2 + v^2 + w^2)\,].$$

3.2.2.6 Markoff's [Markov's] Method

Attempts to generalize the result of De Moivre provided one of the strongest motivations for developing analytical tools of probability theory. A powerful method was proposed by Markoff(1912), but he was unable to make it rigorous. Some twenty years layter, the method was justified by Paul Lévy.

The attempted conclusion is, let $g(x) = 1$, for $\omega_1 < x < \omega_2$, and $g(x) = 0$ otherwise, then

$$\lim_{n\to\infty} \mu\left\{\omega_1 < \frac{r_1(t) + \ldots + r_n(t)}{\sqrt{n}} < \omega_2\right\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{i\omega_2 v} - e^{i\omega_1 v}}{iv}\,dv = \frac{1}{\sqrt{2\pi}}\int_{\omega_1}^{\omega_2} e^{-y^2/2}\,dy$$

The main problem is that the function $\dfrac{e^{i\omega_2 v} - e^{i\omega_1 v}}{iv}$ is not absolutely integrable taking the limit $n \to \infty$, since the limits of integration are $-\infty$ and $+\infty$.

Markov unable to overcome this difficulty abandoned the method. The justification of Markov's method was given by introduction of

$$g(x) = g^{+}(x) - g^{-}(x),$$

where

$$g^{+}(x) = \max\,(g(x), 0), \quad g^{-}(x) = \max\,(-g(x), 0),$$

and either

$$E[g^+(x)] < \infty \text{ or } E[g^-(x)] < \infty,$$

then,

$$E[g(x)] = E[g^+(x)] - E[g^-(x)].$$

This implies g(x) is absolutely integrable function of v in (-∞,∞). Thus the argument can be proved rigorously.

A close inspection of the method of the derivation indicates that the following affirmation is true. Let $f_n(t)$, $0 < t < 1$, be a sequence of measurable functions such that for every v

$$\lim_{n \to \infty} \int_0^1 e^{-iv f_n(t)}\, dt = e^{-v^2/2}$$

Then,

$$\lim_{n \to \infty} \mu\{\omega_1 < f_n(t) < \omega_2\} = \frac{1}{\sqrt{2\pi}} \int_{\omega_1}^{\omega_2} e^{-y^2/2} dy.$$

3.3 Principles of Convergence and The Relation of Normal Distribution With Other Theoretical Distributions

3.3.1 Principles of Convergence

3.3.1.1 Central Limit Theorem

A thorough investigation is done by Loève(1963), Rao(1973), Chow and Teicher (1978) and particularly by Gnedenko and Kolmogorov(1968) in their book entitled *Limit Distributions for Sums of Independent Random Variables*. Primarily the Central Limit Problem of probability theory is the problem of convergence of laws of sequences of sums of random variables. The general Central Limit Problem was solved using the characteristic function tools, and the truncation and symmetrization methods. The Central Limit Problem can be stated as follows:

Let

$$S_{nk_n} = \sum_{k=1}^{k_n} X_{nk}$$

be sums of uniformly asymptotically negligible independent summands X_{nk}, that is

$X_{nk} \xrightarrow{P} 0$ uniformly in k, with $k_n \to \infty$.

(i) Find the family of all possible limit laws of these sums.

(ii) Find conditions for convergence to any specified law of this family.

Following historical development of the problem, there are three limit theorems and corresponding limit laws at the heart of the classical problem. In turn the three limit laws give rise to the three limit types.

The *first theorem* (Bernoulli's) of probability theory, published in 1713, says that $S_n/n \xrightarrow{P} p$, where, S_n is the number of occurrences of an event of probability p in n independent and identical trials, $E(X_k)=p$, $V(X_k)=np(1-p)$, $k=1, 2 \ldots$ $\xrightarrow{P}$ convergence in probability. The result is achieved by direct analysis of the asymptotic behaviour of the binomial probabilities.

Developing the analysis, A. De Moivre(1732), as indicated before, obtained the *second limit theorem* which, in the integral form due to Laplace(1801), says that

$$P\left[(S_n - np)/(np(1-p))^{1/2} < x\right] \to (1/\sqrt{2\pi}) \int_{\infty}^{x} \exp(-\tfrac{1}{2}y^2)\, dy, \quad -\infty \le x \le \infty.$$

The *third theorem* was obtained by Poisson(1832), who modified the Bernoulli case by assuming that the probability $p = p_n$ depends upon the total number n of trials in such a manner that $np_n \to \lambda > 0$. Therefore, writing now X_{nk} and S_{nn} instead of X_k and S_n, the Poisson case corresponds to sequences of sums $S_{nn} = \sum_{k=1}^{n} X_{nk}$, n=1,2, ..., where , for every fixed n, the summands X_{nk} are independent and identically distributed indicators with

$$P[X_{nk} = 1] = \lambda/n + o(1/n).$$

By direct analysis of the asymptotic behaviour of the binomial probabilities, Poisson proved that

$$P[X_{nn} = k] \to \lambda^k e^{-\lambda}/k!, \qquad k = 0, 1, 2, \ldots .$$

Hence, the corresponding three basic laws of probability are deduced.

- The *degenerate law* $\mathcal{L}(0)$ of random variables degenerate at 0 with distribution function having one point of increase at x=0 and characteristic function reduced to 1.
- The *normal law* N(0,1) of normal random variables with density function defined by

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy.$$

and characteristic function given by $\phi(t) = \exp(-t^2/2)$.

- The *Poisson law* $\mathcal{P}(\lambda)$ of Poisson random variables with density function defined by

$$F(x) = e^{-\lambda} \sum_{k=0}^{[x]} \frac{\lambda^k}{k!},$$

and the characteristic function is given by $e^{\lambda[\exp(it)-1]}$.

In brief the three theorems can be summarized as: in the Bernoulli case

$$\mathcal{L}[(S_n - E(S_n))/n] \to \mathcal{L}(0), \text{ and } \mathcal{L}[(S_n - E(S_n))/\sigma S_n] \to N(0,1),$$

while in the Poisson case

$$\mathcal{L}(S_{nn}) \to \mathcal{P}(\lambda).$$

As a conclusion we note that there are three limit types following the three limit laws (i) The *degenerate type* of the degenerate laws $\mathcal{L}(a)$ with $f(t) = e^{ita}$,

(ii) The *normal type* of normal laws $N(a,b^2)$ with $f(t) = \exp\{ita - b^2t^2/2\}$

(iii) The *Poisson type* of Poisson laws $\mathcal{P}(\lambda; a,b)$ with $f(t) = \exp\{ita + \lambda(e^{itb} - 1)\}$

As a remark it can be said that the prominent players in handling and reformulating and finally solving the problem are Kolmogorov(1937) and Lévy(1937). And there are also other scholars like Linderberg(1922), Feller(1937), Doeblin(1939), and Gnedenko(1950).

3.3.1.2 Laws of Large Numbers

A thorough investigation is done by Loève(1963), Rao(1973), Chow and Teicher (1978), Gnedenko and Kolmogorov(1968), and particularly by Sheynin(1968) in his article entitled "On the early history of the law of large numbers." For Kingman and

Taylor(1966) laws of large numbers are precise formulations of the rough interpretation of the expectation of a random variable as the average of its values in a large number of independent trials Beside the two well known *(Borel-1909) strong law of large numbers and the (Bernoulli 1713) weak laws of large numbers*, there are other laws which depict the behaviour of large numbers, like Kolmogorov inequalities, symmetrization inequalities, Lévy inequalities.

There are a class of probability distributions called *symmetric stable laws*. These are distributions whose characteristic functions are of the form exp $(-c\,|\,t\,|^{\alpha})$, where $c > 0$. Such distributions exist if and only if $0 < \alpha \leq 2$. The *Cauchy distribution* has $\alpha = 1$; when $\alpha = 2$ the distributions are *normal distributions*. If X_n have characteristic function exp $(-c\,|\,t\,|^{\alpha})$, it is easy to verify that, for any *n*, $S_n / n^{1/\alpha}$ has the same distribution as X_n.

3.3.2 The Relation with the other Theoretical Distributions

The central importance of normal distribution in statistics stems from three facts: (i) many actual populations approximate closely to normal forms; (ii) it forms the limititing distribution of many widely used statistics and (iii) under general conditions, the means of many distributions tend to be normally distributed in large samples.

3.3.2.1 Discrete Distributions

Following Feller(1993), we can say that a sample space is called discrete if it contains only finitely many points or infinitely many points which can be arranged into a simple sequence E_1, E_2, ... Given a discrete sample space G with sample points E_1, E_2, ... it is assumed that with each point E_i there is associated a number, called the probability of E_i and denoted by $P\{E_i\}$. It is non negative and $P\{E_1\} + P\{E_2\} + ... = 1$. The probability $P\{A\}$to any event A is the sum of the probabilities of all sample points in it.

In more elegant way a discrete distribution function can be defined as follows: a distribution function G on R is called *discrete* if

$$G(x) = \sum_{j:\, x_j < x} p_j \qquad x \in R,$$

where $p_j > 0$ for all j, $\sum_{all\,j} p_j = 1$, and $S = \{x_j; 1 \leq j \leq n \leq \infty\}$ is a subset of $(-\infty, \infty)$. The associated function

$$f(x) = p_j \text{ for } x = xj,\ f(x) = 0 \text{ for } x = xj, \qquad 1 \leq j \leq n \leq \infty$$

is termed a probability density function(p.d.f.) on $S = \{x_j : 1 \leq j \leq n \leq \infty\}$. A p.d.f. is totally determined by S and $\{p_j, 1 \leq j \leq n \leq \infty\}$; and S is the set of positive or nonnegative integers or some finite subset thereof. The probability space (Ω, F, P) and random variable X on it whose F(x) with discrete d.f. G, can be constructed if we choose Ω=S, F=class of all subsets of Ω,

$$P\{A\} = \sum_{j:xj \in A} p_j$$

and

$$X(\omega) = \omega;\ P(\omega : X(\omega) = x_j\} = p_j\ ,\ 1 \leq j \leq n \leq \infty,$$

where $\sum_1^n p_j = 1$

Also Rao(1973), presenting the discrete distribution function as *step function*, depicts the distribution in a measure-probabilistic approach and derives and explains distribution function as a Borel field B generated by the interval [0,1], Ω = [0,1], and P the Lebesgue measure so that $P([0,\omega)) = \omega,\ 0 \leq \omega \leq 1$.

i) From Bernoulli Trials to Binomial Distribution

Repeated independent trials are called Bernoulli trials if there are only two possible outcomes for each trial and their probabilities remain the same throughout the trials. Let b(k;n,θ) be the probability that n Bernoulli trials with probabilities θ for success and 1- θ for failure result in k successes and n-k failures. Then we have

$$b(k;n,\theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

Again let

$$b(k;n,\theta) = P\{S_n = k\},$$

where S_n is the number of successes in n trials as a random variable. Then b(k;n,θ) is our *binomial distribution*. The term binomial refers to the fact that b(k,n,θ) represents the k^{th} term of the *binomial expansion* of $(1 - \theta + \theta)^n$

When n is large and θ is small and $\lambda = n\theta$ of moderate magnitude it is preferable to use the approximation to b(k;n,θ) which is due to Poisson. Siméon D. Poisson (1781 - 1840), a French mathematician deduced this relation in 1832 and his main work is published as "*Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précedées des règles générales du calcul des probabilités*" in 1837. The following derivation will help us to have insight in the relation between the two types of density function.

For k=0

$$b(0;n,\theta) = (1-\theta)^n = (1-\theta/n)^n .$$

Using Taylor expansion rule and logarithms we get

$$\log b(0;n,\theta) = n \log(1-\lambda/n) = -\lambda - \lambda^2/2n - \ldots$$

so that for large *n*

$$b(0,n,\theta) \approx e^{-\lambda}.$$

For fixed k and sufficiently large *n*

$$\frac{b(k;n,\theta)}{b(k-1;n,\theta)} = \frac{\lambda-(k-1)\theta}{k(1-\theta)} \approx \frac{\lambda}{k} .$$

Next, using recurrence method we get

$$b(1;n,\theta) \approx \lambda \,.\, b(0;n,\theta) \approx \lambda e^{-\lambda},$$

$$b(2;n,\theta) \approx \lambda/2 \,.\, b(1;n,\theta) \approx \lambda^2 e^{-\lambda}/2$$

and hence by induction

$$b(k;n,\theta=\lambda/n) \approx \frac{e^{-\lambda}\lambda^k}{k!} = \mathcal{P}(k;\lambda).$$

This is the classical *Poisson approximation to the binomial distribution.* An alternate method is used to deduce this result by Chow and Teicher(1978).

As we can observe from research works and statistical applications the normal approximation to the binomial distribution is of considerable theoretical and practical values Feller's(1993) approach can be used for alternative method to the approximation of binomial distribution by normal distributiion. Furthermore, as it noted the De Moivre-Laplace limit theorem above, it played an important role in the development of probability theory because it lead to the first limit theorem.

In generalized form the approximation formula can be written as:

For fixed z_1 and z_2 as $n \to \infty$

$$P\{n\theta + z_1(n\theta(1-\theta))^{\frac{1}{2}} \leq S_n \leq n\theta + z_2(n\theta(1-\theta))^{\frac{1}{2}}\} \to N(z_1) - N(z_2).$$

That is for

$$P\{\alpha \leq S_n \leq \beta\} = \sum_{k=\alpha}^{\beta} b(k;n,\theta) \approx N(\alpha - n\theta) - N(\beta - n\theta),$$

where $[\alpha,\beta]$ is a fixed real interval.

ii) Poisson Distribution

For large values of λ it is possible to approximate the Poisson distribution by the normal distribution. The deduction is simple. If n is large and θ small, then $b(k;n,\theta)$ will be found to be near the Poisson probabilities $P(k; \lambda)$ with $\lambda = n\theta$. For small λ only the Poisson approximation can be used, but for large λ we can use either the normal or the Poisson approximation.

In general using the Stirling's formula normal approximation to the Poisson distribution can be presented as follows:

If $\lambda \to \infty$, then for fixed $\alpha < \beta$,

$$\sum_{\lambda+\alpha\sqrt{\lambda} < k < \lambda+\beta\sqrt{\lambda}} p(k;\lambda) \to N(\alpha) - N(\beta).$$

iii) Hypergeometric Distribution

Hypergeometric distribution, mainly used for combinatorial problems, is defined us

$$p_k = \frac{\binom{r}{k}\binom{n-r}{n_1-k}}{\binom{n}{n_1}},$$

where $r \le n$ (both integers), $n_1 \le n$ (integer), $k = \max(0, n_1 - n + r), \ldots, \min(r, n_1)$. The probabilities p_k are defined only for k not exceeding r or n_1, and if either $k > r$ or $k > n_1$ then $p_k = 0$.

The name is explained by the fact that the generating function of $\{p_k\}$ can be expressed in terms of Hypergeometric functions. It can be approximated by binomial and Poisson distributions. If n is large and $n_1/n = 0$, then the probability p_k is close to

$$\binom{r}{k}(\theta^k(1-\theta)^{r-k}).$$

More precisely

$$\binom{r}{k}\left(\theta - \frac{k}{n}\right)^k\left((1-\theta) - \frac{r-k}{n}\right)^{r-k} < p_k < \binom{r}{k}\theta^k(1-\theta)^{r-k}\left(1-\frac{r}{n}\right)^{r}.$$

And if $n \to \infty$ and $r \to \infty$ so that the average number $\lambda = r/n$ then $p_r \to \mathcal{P}(\lambda)$.

The normal approximation to the Hypergeometric distribution goes like this. Let n,m,k be positive integers and suppose that they tend to infinity in such a way that

$$\frac{r}{n+m} \to t, \quad \frac{n}{n+m} \to \theta, \quad \frac{m}{n+m} \to (1-\theta), \quad h\{k - r\theta\} \to x$$

where $h = \dfrac{1}{\sqrt{(n+m)\theta(1-\theta)(1-t)}}$.

Furthermore, using the normal approximation to the binomial distribution we get the following result:

$$\binom{n}{k}\binom{m}{r-k}\bigg/\binom{n+m}{r} \approx hN(x)$$

3.3.2.2 Absolutely Continuous Distributions:

A distribution function is called *absolutely continuous* if there exists a Borel function g on $\mathbf{R} = [-\infty, \infty]$ such that

$$G(x) = \int_{-\infty}^{x} g(t)dt, \; x \in \mathbf{R}$$

in which $g(t)$ is said to be the probability density and necessarily satisfies $g \geq 0$ almost everywhere (a.e.),

$$\int_{-\infty}^{\infty} g(t)dt, = 1.$$

If $g(t)$ is continuous at t, $\dot{G}(t) = g(t)$.

For different definitions of absolute continuous distribution we can see the works of Chow and Teicher(1978), Rao(1973), and Feller(1991).

The following distributions fall in this category of absolutely continuous, and we are going to assess the relationship between them and normal distribution.

i)Student's t -Distribution.

The history of t-distribution is interesting. The English mathematician *Student* - pen name of William Sealy Gosset discovered the exact distribution of t in 1908 , and opened the new epoch of *exact sampling theory or distribution*. This work of *Student* made it possible to perform statistical inference by means of small samples and consequently changed statistical research from the study of collectives to that of uncertain phenomena. The concept of population was once again related to a probability space with a probability distribution containing unknown parameters. Thus it began to be emphasised that a sample has to be drawn at random from the population if we are to make an inference about a parameter based on the sample.

In order to define t-distribution let us follow *Students*'(1908) definition and Rao's method:

a) Let $y \sim N(0,1)$ and $x \sim \chi^2(k)$ be independent variables. Then,

$$t = y \,/\, (x/k)^{1/2}$$

which is the ratio of a normal variable to the square root of an independent χ^2 variable divided by the degrees of freedom.

The joint distribution of y and x is

$$c.\exp[-y^2/2]\exp[-x^2/2]\, x^{(k/2)-1}\, dy\, dx. \qquad (3.7)$$

By making the transformation to polar co-ordinates ($0 < r < \infty$, $-\pi/2 < \theta < \pi/2$),

$$y = r \sin \theta, \quad x = r^2 \cos^2 \theta,$$

$$dx\, dy = 2r^2 \cos \theta\, dr\, d\theta$$

transforms to

$$c \exp[-r^2/2]\, r^k (\cos \theta)^{k-1} dr\, d\theta. \tag{3.8}$$

The distribution of θ alone, which is seen to be independent of r, is

$$c\,.\,(\cos \theta)^{k-1}\, d\theta = [\beta(½, k/2)]^{-1} (\cos \theta)^{k-1}\, d\theta, \tag{3.9}$$

thus, supplying the constant to make the total integral unity. The statistic whose distribution is to be found is

$$t = \sqrt{k}\, y / \sqrt{x} = k \tan \theta, \ (-\infty < t < \infty),$$

$$dt = k \sec^2 \theta\, d\theta = k(1 + t^2/k)\, d\theta.$$

The expression (3.9) transforms to

$$S(t/k)\, dt = [\sqrt{k}\, \beta(½, k/2)]^{-1} (1 + t^2/k)^{-(k+1)/2}\, dt, \tag{3.10}$$

which is called *Student's t* distribution on *k* degrees of freedom and is represented by S(*k*).

b) Let $y \sim N(\mu, \sigma^2)$ and $(x/\sigma^2) \sim \chi^2(k)$ be independent.

Then, since $(y - m)/s \sim N(0, 1)$ and applying (3.10) to the ratio of

$$[(y - \mu)/\sigma] \text{ to } (x/k\sigma^2)^{½}$$

we get

$$(y - \mu)/(x/k)^{½} \sim S(k). \tag{3.11}$$

c) Let $y \sim N(\mu, \sigma^2)$ and $(x/\sigma^2) \sim \chi^2(k)$ be independent.

The probability density of $t = y/(x/k)^{½}$ is

$$S(t/k, \delta) = \frac{k^{k/2}}{\Gamma(k/2)} \frac{e^{-\delta^2/2}}{(k + t^2)^{(k+1)/2}} \sum_{s=0}^{\infty} \Gamma\left(\frac{k+s+1}{2}\right)\left(\frac{\delta^s}{s!}\right)\left(\frac{2t^2}{k+t^2}\right)^{s/2}$$

where $\delta = \mu/\sigma$, which is called the non-central t distribution.

ii) Hotelling T^2 - statistic.

The multivariate analogue of the square of Student's t -distribution is

$$T^2 = N(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu).$$

where $\bar{x}$ is the mean vector of a sample of N and S is the sample covariance matrix. Hotelling (1931) *in his paper "The generalization of Student's ratio,"* proposed the T^2-statistic for two samples and derived the distribution under the null hypothesis. But the representation of T^2 as the ratio of independent χ^2's leading to an elegant derivation of its distribution is due to Wijsman(1957). For detailed and scholarly presentation we can see Hotelling(1931) and Anderson(1958).

iii)Gamma Distribution

The general gamma distribution has the p.d

$$G(x/\alpha,p) = [\alpha^p / G(p)] e^{-\alpha x} x^{p-1}, \qquad \alpha > 0,\ p > 0,\ 0 < x < \infty.$$

The r^{th} raw moment is seen to be

$$\Gamma(p+r)/\alpha^r \Gamma(p)$$

so that

$$E(x) = p/\alpha \text{ and } V(x) = p/\alpha^2.$$

Let $x_i \sim G(\alpha,p_i)$, i=1, ..., k be all independent. Then,

$$x_1 + \ldots + x_k \sim G(\alpha, p = \Sigma p_i),$$

that is, the gamma distribution has the reproductive property like the normal distribution but not for variations in both the parameters.

Let $x \sim G(\alpha,p_1)$ and $y \sim G(\alpha,p_2)$. Then, p.d. of $g = x/(x+y)$ is

$$[\Gamma(p_1+p_2)/\Gamma(p_1)\Gamma(p_2)]\, g^{p_1-1}(1-g)^{p_2-1}, \quad 0 < g < 1,$$

which is called the *beta distribution*, $B(p_1,p_2)$. The beta distribution involving two parameters γ, δ has the probability density

$$B(x/\gamma,\delta) = [\,b(g,d)]\ \text{-}1\ x\ g\text{-}1(1\text{-}x)d\text{-}1, \quad 0 < x < 1,$$

where b(γ,δ) is the *Beta function*

$$\Gamma(\gamma)\Gamma(\delta) / \Gamma(\gamma+\delta).$$

iv) χ^2 - Distribution

The special case of the *gamma distribution* with $\alpha=1/2$ and $p=k/2$, where k is an integer is called the χ^2 *distribution* on *k* degrees of freedom. The density function is

$$\chi^2(x/k) = 1/(2^{k/2}\,\Gamma(k/2)\,e^{-x/2}\,x^{(k/2)-1}.$$

Following the property of gamma distribution, we observe if $x_i \sim \chi^2(k_i)$, i=1, ..., m, are independent, then $\Sigma x_i \sim \chi^2(\Sigma k_i)$.

v) Wishart Distribution.

John Wishart(1928) in his article entitled "The Generalized product moment distribution in samples from a normal multivariate population" came up with **Wishart Distribution**. The sample covariance matrix,

$$S = \frac{1}{N-1}\,\Sigma_\alpha(x_\alpha - \bar{x})(x_\alpha - \bar{x})'$$

is an estimate of the population covariance matrix Σ. When Σ=I, this distribution is in a sense a generalisation of the χ^2-distribution. The distribution of S, often called the Wishart distribution, is fundamental to multivariate statistical analysis.

vi) Cauchy Distribution

The Cauchy density centred at the origin is defined by

$$C(x/\mu,\tau) = (1/\pi)\,(\tau/\,\tau^2 + x^2), \qquad -\infty < x < \infty,\ \mu = 0,$$

where $\tau > 0$ is a scale parameter. the Corresponding distribution function is π^{-1} arc tan (x/t). The graph of $C(x/\mu,\tau)$resembles that of the normal density but approaches the axis so slowly that an expectation does not exist. The importance of the Cauchy densities is due to the convolution formula

$$C_s * C_t = C_{s+t}. \tag{3.12}$$

It states that the family of Cauchy densities is closed under convolutions.

The convolution formula has the amazing consequence that for independent variables X_1, ..., X_n with the common density, the average $(X_1 + ... + X_n)$ / n has the same density as the X_i. It has the curious property that if X has density C_t then 2X has density

$C_{2t} = C_t * C_t$. Thus 2X = X + X is the sum of two dependent variables, but its density is given by the convolution formula. Moreover, if U and V are two independent variables with common density C_t and $X = aU + bV$, $Y = cU + dV$, then X+Y has density $C_{(a+b+c+d)t}$ which is the convolution of the densities $C_{(a+b)t}$ of X and $C_{(c+d)t}$ of Y; nevertheless, X and Y are not independent.

The Cauchy density corresponds to the special case n=1 of the family of Student's t densities. In other words, If X and Y are independent random variables with the normal density N, then $X / |Y|$ has the Cauchy density with t=1. The convolution property of the gamma densities looks exactly like (3.12) but there is an important difference in that the parameter α of the gamma densities is essential whereas (3.12) contains only a scale parameter. With the Cauchy density the type is *stable*. This stability under convolutions is shared by the normal and Cauchy densities; the difference is that the scale parameters compose according to the rules $\sigma^2 = \sigma_1^2 + \sigma_2^2$ and $\alpha = \alpha_1 + \alpha_1$, respectively. There exist other stable densities with similar properties, and with a systematic terminology, the normal and Cauchy densities are called "symmetric, stable of index 2 and 1." We note that the law of large numbers describing the behaviour of the mean as the number of observations increases does not hold in the case of the *Cauchy distribution*.

3.4 The Relation with Pure and Applied Mathematics

3.4.1 Maxwell's Distribution of Velocities

James Clerk Maxwell(1831-1879), an immediate predecessor of L. Boltzmann (1844-1906) - who is known as the initiator and developer of statistical physics as well as one of the founders of modern physics and theoretical physicist - thought of molecules as elastic solids. Starting from this premise, Maxwell constructed a theory of gases that was related to the works of Clausius In his paper, 1875, by stating special contribution of Clausius, in the development of methods for investigating systems consisting of infinitely many molecules in motion, "opened up a new field of mathematical physics."

Maxwell states that "By following this method, which is the only one available either experimentally or mathematically, we pass from the methods of strict dynamics to those of statistics and probability. When an encounter takes place between two molecules, they are transferred from one pair of groups to another, but by the time that a great many encounters have taken place, the number which enter each group is, on an average, neither more nor less than the number which leave it during the same time. When the system has reached this state, the numbers in each group must be distributed according to some definite law."

This distribution law of the velocities of molecules was derived by Maxwell. For this purpose, he proceeds form the following consideration:

Let $\varphi(x)dx$ be the probability that the projection of the velocity of a molecule on the x axis is contained between x and $x + dx$, and let the corresponding definitions be given for $\varphi(y)dy$ and $\varphi(z)dz$. The probability that the vector from the origin representing the velocity will be contained between x, y, z and $x + dx, y + dy, z + dz$, is equal to

$$P = \varphi(x)\varphi(y)\varphi(z)dxdydz.$$

This probability, on the other hand, should be a function of the distance from the origin, i.e.,

$$\varphi(x)\varphi(y)\varphi(z) = f(x^2 + y^2 + z^2).$$

Taking logarithms on both sides, we obtain:

$$\ln \varphi(x) + \ln \varphi(y) + \ln \varphi(z) = \ln f(x^2 + y^2 + z^2).$$

Differentiating with respect to x yields:

$$\frac{\varphi'(x)}{\varphi(x)} = \frac{2x f'(x^2 + y^2 + z^2)}{f(x^2 + y^2 + z^2)} \quad \text{or} \quad \frac{\varphi'(x)}{x\,\varphi(x)} = \frac{2 f'(x^2 + y^2 + z^2)}{f(x^2 + y^2 + z^2)}.$$

Analogously,

$$\frac{\varphi'(x)}{x\,\varphi(x)} = \frac{\varphi'(y)}{y\,\varphi(y)} = \frac{\varphi'(z)}{z\,\varphi(z)} = \frac{2 f'(x^2 + y^2 + z^2)}{f(x^2 + y^2 + z^2)}.$$

Taking into account certain additional physical considerations, we easily determine a function that satisfies this relation:

$$\varphi(x) = \exp(-k x^2); \quad \varphi(x)\varphi(y)\varphi(z) = f(x^2 + y^2 + z^2) = \exp[-k^2(x^2 + y^2 + z^2)].$$

This formula represents the *Maxwell law of velocities*.

3.4.2 Gaussian Distribution and the Law of Errors

Gauss published his derivation of the normal law of distribution of random errors in observations in 1809 in his famous work "Theoria motus corporum coelestium." Along with the unusually wide scope of Gauss's activities, a characteristic feature of his investigations in a deep interrelation between theoretical and applied problems. He often discovered general mathematical ideas as a result of solving specific problems. This is particularly relevant to his work in the field of probability theory.

After deriving the "normal law of the distribution of random errors"

$$[\varphi(\Delta) = h\pi^{-1/2} \exp(-h^2 \Delta^2)],$$

Gauss points out a certain defect in this law. According to this law, errors of any magnitude are possible. It should be emphasised that in deriving the normal distribution Gauss made extensive use of the principle of the arithmetic mean.

The most complete exposition of the theory of errors is contained in Gauss's(1828) paper "Theoria combinationis observationum erroribus minimis obnoxiae." Gauss writes in this memoir that no matter how carefully the observations are carried out, errors are unavoidable. Some errors may be random, others may be predicted and evaluated since these are either constant or vary in a regular manner. The latter type of error is referred to as a systematic error. Gauss, however, points out that such a division of errors into two kinds is relative, and in many cases depends upon the problem at hand.

This paper of Gauss's is devoted to the study of laws governing the distribution of random errors. Gauss proceeds with the most general assumptions concerning the probability density of the errors $\varphi(x)$. Since positive and negative errors appear equally often, $\varphi(-x)=\varphi(x)$. Next small errors occur more often than the large ones, hence the value of $\varphi(x)$ will be maximal at $x=0$ and diminishes constantly with the increase of x. Clearly, the value of the integral $\int\varphi(x)dx$ in the limits form $x = -\infty$ up to $x = +\infty$ always equals 1.

The basic problem considered by Gauss is actually as follows: Let the variables y, x_1, x_2, ..., x_n be linearly related, i.e. $y = \sum^{n}_{s=1} a_s x_s$, while the a_s are unknown. To determine these unknowns, the values of $y_r = \sum^{n}_{s=1} a_s x_{sr}$, $r = 1,2, ..., N$, are obtained from the experimental data.

But the experimental determination of y_r is subject to error. Thus, we actually obtain, instead the value y_r, the value $\eta_r = y_r + \Delta_r$. Given x_{sr} and the obtained values of η_r, it is required to determine the best possible approximate values α_s of the quantities a_s. According to Gauss, these ought to be determined from the condition

$$\sum^{N}_{r=1} (\eta_r - \sum^{n}_{s=1} \alpha_s x_{sr})^2 = \min. \qquad (3.13)$$

α_s are then uniquely determined form the system of equations derived form condition (3.13). The equation are called normal equations and are of the form:

$$\sum^{n}_{s=1} \alpha_s \sum^{N}_{r=1} x_{sr} x_{ir} = \sum^{N}_{r=1} \eta_r x_{ir}, \quad i = 1,2, ..., n.$$

Condition (3.13) is minimised if the expression in each of the square brackets vanishes, i.e.

$$\eta_r = \sum^{n}_{s=1} \alpha_s x_{sr},$$

and the values of η_r ($r = 1,2, ..., N$) satisfy the system of the normal equations. The obtained approximations α_s of the values of a_s are free from systematic bias, that is, the mathematical expectation of a α_s is equal to a_s.

3.4.3 The Normal Law in Number Theory

The normal law in number theory, in comparison to pure mathematics, is studied by many scholars, namely, Erdös and Kac(1939), Kac(1949,1959), Rényi(1955), Rényi and Turan(1958), Kubilus(1956), and others.

In order to analyze the properties of normal law in number theory we need preliminary definitions on basic concepts.

3.4.3.1 From Vieta to the Notion of Statistical Independence

Every real number t, $0 \le t < 1$, can be written uniquely in the form

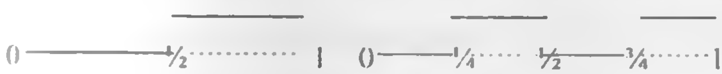
$$t = \frac{\varepsilon_1}{2} + \frac{\varepsilon_2}{2^2} + \frac{\varepsilon_3}{2^3} + \ldots \qquad (3.14)$$

where each ε is either 0 or 1. This is the familiar *binary expansion* of t.

Since the digits ε_i are functions of t, and thus (3.14) can be written as

$$t = \frac{\varepsilon_1(t)}{2} + \frac{\varepsilon_2(t)}{2^2} + \frac{\varepsilon_3(t)}{2^3} + \ldots \tag{3.15}$$

With the convention about terminating expansions, the graphs of $\varepsilon_1(t)$, $\varepsilon_2(t)$, $\varepsilon_3(t)$, . . . are as follows:



Let the functions $r_i(t)$, introduced and studied by Rademacher(1922) or known as Rademacher functions, be defined by the equations

$$r_i(t) = 1 - 2\varepsilon_i(t), \qquad k = 1, 2, 3, \ldots \tag{3.16}$$

whose graphs look as follows:

Now we can write (3.15) in the form of

$$1 - 2t = \sum_{k=1}^{\infty} \frac{r_k(t)}{2^k} \tag{3.17}$$

Next, we note that

$$\int_0^1 e^{ix(1-2t)}\,dt = \frac{\sin x}{x}$$

and

$$\int_0^1 \exp\left(ix\frac{r_k(t)}{2^k}\right)dt = \cos\frac{x}{2^k}$$

Now

$$\frac{\sin x}{x} = \prod_{k=1}^{\infty} \cos\frac{x}{2^k},$$

whose special case when $x = \pi/2$ is the *classical formula of Vieta*, that is,

$$\frac{2}{\pi} = \prod_{n=1}^{\infty} \cos\frac{\pi}{2^{n+1}},$$

assumes the form

$$\frac{\sin x}{x} = \int_0^1 e^{ix(1-2t)}\,dt = \int_0^1 \exp\left(ix \sum_{k=1}^{\infty} \frac{r_k(t)}{2^k} \right) dt = \prod_{k=1}^{\infty} \cos\frac{x}{2^k} = \prod_{k=1}^{\infty} \int_0^1 \exp\left(ix \frac{r_k(t)}{2^k} \right) dt.$$

This implies that

$$\int_0^1 \prod_{k=1}^{\infty} \exp\left(ix \frac{r_k(t)}{2^k} \right) dt = \prod_{k=1}^{\infty} \int_0^1 \exp\left(ix \frac{r_k(t)}{2^k} \right) dt$$

An integral of a product is a product of integrals!

Using the above formulae the Vieta's formula is connected to binary digits.

Let us consider the set of t's for which

$$r_1(t) = +1, \quad r_2(t) = -1, \quad r_3(t) = -1.$$

The graphs of r_1, r_2 and r_3 indicate that the set, except possible for end points, is simply the interval $\left(\frac{3}{8}, \frac{4}{8}\right)$.

The length or measure of this interval is clearly $\frac{1}{8}$, and

$$\frac{1}{8} = \tfrac{1}{2} \cdot \tfrac{1}{2} \cdot \tfrac{1}{2}.$$

This observation can be written in the form

$$\mu\{r_1(t) = +1, r_2(t) = -1, r_3(t) = -1\} = \mu\{r_1(t) = +1\}\, \mu\{r_2(t) = -1\}\, \mu\{r_3(t) = -1\}$$

where μ stands for measure(length) of the set defined inside the braces.

In general if $\delta_1, \delta_2, \ldots, \delta_n$ is a sequence of +1's and -1's then

$$\mu\{r_1(t) = \delta_1, \ldots, r_n(t) = \delta_n\} = \mu\{r_1(t) = \delta_1\}\, \mu\{r_2(t) = \delta_2\} \ldots \mu\{r_n(t) = \delta_n\}$$

This may seem to be merely a complicated way of writing

$$(\tfrac{1}{2})^n = \tfrac{1}{2} \times \tfrac{1}{2} \times \ldots \times \tfrac{1}{2} \text{ (n times)},$$

but it expresses a deep property of the functions $r_k(t)$, and hence binary digits.

Now using this property we can prove that "an integral of a product is a product of integrals."

$$\int_0^1 \exp\left(i \sum_{k=1}^{n} c_k r_k(t) \right) dt = \sum_{\delta_1,\ldots,\delta_n} \exp\left(i \sum_{k=1}^{n} c_k \delta_k(t) \right) \mu\{c_1(t) = \delta_1, \ldots, c_n(t) = \delta_n\}$$

$$= \sum_{\delta_1,\ldots,\delta_n} \prod_{k=1}^{n} e^{ic_k\delta_k} \prod_{k=1}^{n} \mu\{r_k(t) = \delta_k\} = \sum_{\delta_1,\ldots,\delta_n} \prod_{k=1}^{n} e^{ic_k\delta_k} \mu\{r_k(t) = \delta_k\}$$

$$= \prod_{k=1}^{n} \sum_{\delta_1,\ldots,\delta_n} e^{ic_k\delta_k} \mu\{r_k(t) = \delta_k\} = \prod_{k=1}^{n} \int_0^1 e^{ic_k\delta_k}\,dt.$$

After this initial connection between *Vieta's formula* and *binary digits* next we turn to the theory of coin tossing. The elementary theory of coin tossing starts with two assumptions: (i) the coin is "fair" and (ii) the successive tosses are independent.

The first assumption helps us to have equiprabability[$p = \frac{1}{2}$], while the second is used to justify the rule of multiplication of probabilities. That is if events $A_1, \ldots, A_n$ are independent, then the probability of their joint occurrence is the product of the probabilities of their individual occurrences. Thus, the functions $r_k(t)$ can be used as *model* for coin tossing.

Next let us have a glossary for our model.

Probability Theoretic	**Number-Theoretic**
symbol H (Head)	+
symbol T (Tail)	−
kth toss (k = 1, 2, ...)	$r_k(t)$ (k = 1,2, ...)
event	set of t's
probability of an event	measure of the corresponding set of t's
probability theoretic r. v.'s	number theoretic functions - f(n)
expectation of r.v.'s	mean value of a function $M\{f(t)\} = \lim_{n \to \infty} \frac{1}{2T} \int_{-T}^{T} f(t)dt$

The analogue of probability and number terminologies can be clarified through the following example

- Find the probability that in n independent tosses of a fair coin, exactly ℓ will be heads.
- Find the measure of the set of t's such that exactly ℓ of the n numbers $r_1(t), r_2(t), \ldots, r_n(t)$ are equal to +1.

The condition that exactly ℓ among $r_1(t), r_2(t), \ldots, r_n(t)$ are equal to 1 is equivalent to the condition that

$$r_1(t) + r_2(t) + \ldots + r_n(t) = 2\ell - n \tag{3.18}$$

Next we note that, for m an integer we have

$$\frac{1}{2\pi}\int_0^{2\pi} e^{imx}dx = \begin{cases} 1, & m = 0 \\ 0, & m \neq 0, \end{cases}$$

After this initial connection between *Vieta's formula* and *binary digits* next we turn to the theory of coin tossing. The elementary theory of coin tossing starts with two assumptions: (i) the coin is "fair" and (ii) the successive tosses are independent.

The first assumption helps us to have equiprabability[$p = ½$], while the second is used to justify the rule of multiplication of probabilities. That is if events $A_1,...,A_n$ are independent, then the probability of their joint occurrence is the product of the probabilities of their individual occurrences. Thus, the functions $r_k(t)$ can be used as *model* for coin tossing.

Next let us have a glossary for our model.

Probability Theoretic	**Number-Theoretic**
symbol H (Head)	+
symbol T (Tail)	–
kth toss (k = 1, 2, ...)	$r_k(t)$ (k = 1,2, ...)
event	set of t's
probability of an event	measure of the corresponding set of t's
probability theoretic r. v.'s	number theoretic functions - f(n)
expectation of r.v.'s	mean value of a function $M\{f(t)\} = \lim_{n\to\infty} \frac{1}{2T}\int_{-T}^{T} f(t)dt$

The analogue of probability and number terminologies can be clarified through the following example:

- Find the probability that in n independent tosses of a fair coin, exactly ℓ will be heads.
- Find the measure of the set of t's such that exactly ℓ of the n numbers $r_1(t)$, $r_2(t)$, ..., $r_n(t)$ are equal to +1.

The condition that exactly ℓ among $r_1(t)$, $r_2(t)$, ..., $r_n(t)$ are equal to 1 is equivalent to the condition that

$$r_1(t) + r_2(t) + \ldots + r_n(t) = 2\ell - n. \qquad (3.18)$$

Next we note that, for m an integer we have

$$\frac{1}{2\pi}\int_0^{2\pi} e^{imx}dx = \begin{cases} 1, & m = 0 \\ 0, & m \neq 0, \end{cases}$$

$$\lim_{n\to\infty} \frac{r_1(t) + \dots + r_n(t)}{n} = 0,$$

and is equivalent to saying that, for almost every t,

$$\lim_{n\to\infty} \frac{\varepsilon_1(t) + \dots + \varepsilon_n(t)}{n} = \frac{1}{2}.$$

In other words, almost every number t has, asymptotically, the same number of zeros and ones in its binary expansion. This is the arithmetical content of Borel's theorem.

If g is an integer greater than 1, we can write

$$t = \frac{\omega_1(t)}{g} + \frac{\omega_n(t)}{g^2} + \dots, \qquad 0 \le t \le 1,$$

where each digit $\omega(t)$ can now assume the values 0, 1, ..., g - 1.

For almost every t ($0 \le t \le 1$)

$$\lim_{n\to\infty} \frac{F_n^{(k)}(t)}{n} = \frac{1}{g},$$

where $F_n^{(k)}(t)$ denotes the number of times the digit k, $0 \le k \le g - 1$, occurs among the first n ω's.

From the fact that a denumerable union of sets of measure 0 is of measure 0, it follows that almost every number t, $0 \le t \le 1$, is such that in every system of notation, that is, for every $g > 1$, each allowable digit appears with proper and just frequency. In other words, almost every number is *normal*. A simple example by Champernowne(1933) is the number written in decimal notation

0 123456789101112131415161718192021222324 2 . . .,

where after the decimal point we write out all positive integers in succession.

3.4.3.3 A law of nature or a mathematical theorem?

In the study of the normal law in number there is a striking question of inquiry. Is Normal Law a law of nature or a mathematical theorem? In order to have satisfactory answer let us define three important phrases, namely, the relative measure, the mean value of a function and linear independence of real numbers.

- *The relative measure* - Let A be a set of real numbers, and consider the subset of A which lies in (-T,T), i.e., $A \cap (-T,T)$. The relative measure $\mu_R\{A\}$ of A is defined as the limit

$$\mu_R\{\Lambda\} = \lim_{T\to\infty} \frac{1}{2T}\ \mu\{\Lambda \cap (-T,T)\},$$

if the limit exists. The relative measure is not completely additive, for if $\Lambda_i = (i, i+1)$, $i = \pm 0, \pm 1, \pm 2, \ldots$, then

$$\mu_R\left\{\bigcup_{i=-\infty}^{\infty} \Lambda_i\right\} = 1,$$

while

$$\sum_{i=-\infty}^{\infty} \mu_R\{\Lambda_i\} = 0.$$

- *The mean value of a function* - The mean value $M\{f(t)\}$ of the function $f(t)$, $-\infty < t < \infty$, is defined as the limit

$$M\{f(t)\} = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f(t)\,dt,$$

if the limit exists.

- *Linear independence of real numbers* - Real numbers λ_1, λ_2, ... are called linearly independent or independent over the field of rationals if the only solution $(k_1, k_2, k_3, \ldots)$ in integers of the equation

$$k_1\lambda_1 + k_2\lambda_2 + \ldots = 0 \quad \text{is} \quad k_1 = k_2 = k_3 = \ldots = 0.$$

Next let λ_1, λ_2, ... be linearly independent, and consider the function

$$\sqrt{2}\,\frac{\cos\lambda_1 t + \ldots + \cos\lambda_n t}{\sqrt{n}} \tag{3.19}$$

Let $\Lambda_n(\omega_1,\omega_2)$ be the set on which

$$\omega_1 < \sqrt{2}\,\frac{\cos\lambda_1 t + \ldots + \cos\lambda_n t}{\sqrt{n}} < \omega_2\,. \tag{3.20}$$

Theorem: $\mu_R\{\Lambda_n(\omega_1,\omega_2)\}$ is defined and moreover that

$$\lim_{n\to\infty} \mu_R\{\Lambda_n(\omega_1,\omega_2)\} = \frac{1}{\sqrt{2\pi}}\int_{\omega_1}^{\omega_2} e^{-y^2/2}\,dy. \tag{3.21}$$

Proof: Using similar notations used in Markov's method we have

$$\begin{aligned} &\frac{1}{2T}\int_{-T}^{T} g_\varepsilon\left(\sqrt{2}\,\frac{\cos\lambda_1 t + \ldots + \cos\lambda_n t}{\sqrt{n}}\right)dt \\ &\le \frac{1}{2T}\int_{-T}^{T} g\left(\sqrt{2}\,\frac{\cos\lambda_1 t + \ldots + \cos\lambda_n t}{\sqrt{n}}\right)dt \\ &\le \frac{1}{2T}\int_{-T}^{T} g_\varepsilon^1\left(\sqrt{2}\,\frac{\cos\lambda_1 t + \ldots + \cos\lambda_n t}{\sqrt{n}}\right)dt \end{aligned} \tag{3.22}$$

where $g_\varepsilon^-(x) \le g(x) \le g_\varepsilon^+(x)$.

Furthermore,

$$\frac{1}{2T}\int_{-T}^{T} g_\varepsilon^+\left(\sqrt{2}\,\frac{\cos\lambda_1 t + \ldots + \cos\lambda_n t}{\sqrt{n}}\right)dt$$

$$\le \frac{1}{2T}\int_{-\infty}^{\infty} G_\varepsilon^+(\xi)\left[\frac{1}{2T}\int_{-T}^{T}\exp\left(i\xi\sqrt{2}\,\frac{\cos\lambda_1 t + \ldots + \cos\lambda_n t}{\sqrt{n}}\right)dt\right]d\xi \tag{3.23}$$

where both $G_\varepsilon^+(\xi)$ and $G_\varepsilon^+(\xi)$ are absolutely integrable in $(-\infty, \infty)$.

Now we prove that

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\exp\left(i\xi\sqrt{2}\,\frac{\cos\lambda_1 t + \ldots + \cos\lambda_n t}{\sqrt{n}}\right)dt = J_0^{\,n}\left(\sqrt{2}\,\frac{\xi}{\sqrt{n}}\right) \tag{3.24}$$

where J_0 is the familiar Bessel Function.

Let us prove the theorem for n = 2 since the proof for arbitrary n is exactly the same.

Letting

$$\eta = \left(\sqrt{2}\,\frac{\xi}{\sqrt{n}}\right),$$

and recalling that

$$G_\varepsilon^\pm(\xi) = \int_{-\infty}^{\infty} g_\varepsilon^\pm(x)\, e^{i\xi x}\, dx.$$

we have

$$\frac{1}{2T}\int_{-T}^{T} e^{i\eta(\cos\lambda_1 t + \cos\lambda_2 t)}\,dt = \sum_{k,\ell=0}^{\infty}\frac{(i\eta)^k (i\eta)^\ell}{k!\,\ell!}\,\frac{1}{2T}\int_{-T}^{T}\cos^k\lambda_1 t\,\cos^\ell\lambda_2 t\,dt. \tag{3.25}$$

Now, we must find that

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\cos^k\lambda_1 t\,\cos^\ell\lambda_2 t\,dt = M\{\cos^k\lambda_1 t\,\cos^\ell\lambda_2 t\}.$$

It is known that

$$\cos^k\lambda_1 t\,\cos^\ell\lambda_2 t = \frac{1}{2^k}\,\frac{1}{2^\ell}\left(e^{i\lambda_1 t} + e^{-i\lambda_1 t}\right)^k\left(e^{i\lambda_2 t} + e^{-i\lambda_2 t}\right)^\ell$$

$$= \frac{1}{2^k}\,\frac{1}{2^\ell}\sum_{r=0}^{k}\sum_{s=0}^{\ell}\binom{k}{r}\binom{\ell}{s} e^{i[(2r-k)\lambda_1 + (2s-\ell)\lambda_2]t}.$$

and

$$M\{e^{i\alpha t}\} = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{i\alpha t}\,dt = \begin{cases}1, & \alpha = 0\\ 0, & \alpha \ne 0.\end{cases}$$

Because of linear independence,

$$(2r-k)\lambda_1 + (2s-\ell)\lambda_2$$

can be zero only if $2r = k$ and $2s = \ell$, and thus it follows almost immediately that

$$M\{\cos^k \lambda_1 t \ \cos^\ell \lambda_2 t\} = \frac{1}{2^k}\binom{k}{\frac{k}{2}}\frac{1}{2^\ell}\binom{\ell}{\frac{\ell}{2}} \tag{3.26}$$

if both k and ℓ are even and 0 in all other cases. We can write (3.26) in the form

$$M\{\cos^k \lambda_1 t \ \cos^\ell \lambda_2 t\} = M\{\cos^k \lambda_1 t\}\, M\{\cos^\ell \lambda_2 t\}, \tag{3.27}$$

and combining this with (3.25) we obtain

$$M\{e^{i\eta(\cos\lambda_1 t + \cos\lambda_2 t)}\} = M\{e^{i\eta\cos\lambda_1 t}\}\, M\{e^{i\eta\cos\lambda_2 t}\} \tag{3.28}$$

It is clear that

$$M\{e^{i\eta\cos\lambda t}\} = \frac{1}{2\pi}\int_0^{2\pi} e^{i\eta\cos\theta}\,d\theta = J_0(\eta) \tag{3.29}$$

and hence from (3.28) we get

$$M\{e^{i\eta(\cos\lambda_1 t + \cos\lambda_2 t)}\} = J_0^{\,2}(\eta).$$

Thus we can consider (3.24) as having been proved. Letting $T \to \infty$ in (3.22) and using (3.23) and (3.24) we obtain

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\infty}^{\infty} G_\varepsilon^-(\xi) J_0^{\,n}\left(\sqrt{2}\,\frac{\xi}{\sqrt{n}}\right) d\xi \\
&\le \liminf_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} g\left(\sqrt{2}\,\frac{\cos\lambda_1 t + \ldots + \cos\lambda_n t}{\sqrt{n}}\right) dt \\
&\le \limsup_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} g\left(\sqrt{2}\,\frac{\cos\lambda_1 t + \ldots + \cos\lambda_n t}{\sqrt{n}}\right) dt \\
&\le \frac{1}{2\pi}\int_{-\infty}^{\infty} G_\varepsilon^+(\xi) J_0^{\,n}\left(\sqrt{2}\,\frac{\xi}{\sqrt{n}}\right) d\xi.
\end{aligned} \tag{3.30}$$

It is well known that as $\eta \to +\infty$

$$J_0(\eta) = O\left(\frac{1}{\sqrt{|\eta|}}\right),$$

and consequently, for $n \ge 3$,

$$J_0^{\,n}\left(\sqrt{2}\,\frac{\xi}{\sqrt{n}}\right)$$

is absolutely integrable in ξ. This implies that for $n \geq 3$

$$\lim_{\varepsilon \to 0} \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{\varepsilon}^{-}(\xi) J_0^{\,n}\left(\sqrt{2}\,\frac{\xi}{\sqrt{n}}\right) d\xi = \lim_{\varepsilon \to 0} \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{\varepsilon}^{+}(\xi) J_0^{\,n}\left(\sqrt{2}\,\frac{\xi}{\sqrt{n}}\right) d\xi$$

and hence that

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} g\left(\sqrt{2}\,\frac{\cos\lambda_1 t + \ldots + \cos\lambda_n t}{\sqrt{n}}\right) dt = \mu_r\{\Lambda_n(\omega_1, \omega_2)\}$$

exists. Next (3.30) can be written in the form

$$\frac{1}{2\pi} \int_{-T}^{T} G_{\varepsilon}^{-}(\xi) J_0^{\,n}\left(\sqrt{2}\,\frac{\xi}{\sqrt{n}}\right) d\xi \leq \mu_r\{\Lambda_n(\omega_1, \omega_2)\}$$

$$\leq \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{\varepsilon}^{+}(\xi) J_0^{\,n}\left(\sqrt{2}\,\frac{\xi}{\sqrt{n}}\right) d\xi,$$

and one verifies easily that

$$\lim_{n \to \infty} (\xi) J_0^{\,n}\left(\sqrt{2}\,\frac{\xi}{\sqrt{n}}\right) = e^{-\xi^2/2}.$$

The proof of (3.21) can now be completed exactly as in Markov's method.

If we look upon

$$q_n(t) = \sqrt{2}\,\frac{\cos\lambda_1 t + \ldots + \cos\lambda_n t}{\sqrt{n}}$$

as a result of superposition of vibrations with incommensurable frequencies, the theorem embodied in (3.21) gives precise information about the relative time $q_n(t)$ spends between ω_1 and ω_2. That we are led here to the normal law

$$\frac{1}{\sqrt{2\pi}} \int_{\omega_1}^{\omega_2} e^{-y^2/2}\, dy$$

usually associated with random phenomena is perhaps an indication that the deterministic and probabilistic point of view are not as irreconcilable as they may appear at first sight.

3.4.3.4 Theorems of Normal Law in Number Theory

It is good to note that a number theoretic function $f(n)$ is a function defined on the positive integers 1, 2, 3, ...; and the mean $M\{f(n)\}$ of f is defined as the limit (if it exists)

$$M\{f(n)\} = \lim_{n \to \infty} \frac{1}{N} \sum_{n=1}^{N} f(n).$$

If A is a set of positive integers, we denote by A(N) the number of its elements among the first N integers.

If

$$\lim_{n\to\infty}\frac{A(N)}{N}=D\{A\}$$

exists, it is called the *density* of A. The density is analogous to the *relative measure*, and like relative measure it is not completely additive. Consider the integers divisible by a prime *p*. The density of the set of these integers is clearly $1/p$. Take now the set of integers divisible by both *p* and *q* (*q* another prime). To be divisible by *p* and *q* is equivalent to being divisible by *pq*, and consequently the density of the new set is $1/pq$. Now $1/pq = 1/p \cdot 1/q$, and we can interpret this by saying that the "events" of being divisible by *p* and *q* are independent. This holds, of course, for any number of primes, and we can say, using a picturesque but not a precise language, that the primes play a game of chance! This simple, nearly trivial, observation is the beginning of a new development which links in a significant way number theory on the one hand and probability theory on the other.

The fact that $\nu(m)$, the number of prime divisors of *m*, is the sum

$$\sum_p \rho_p(m) \tag{3.31}$$

of independent functions suggests that, in some sense, the distribution of values of $\nu(m)$ may be given by the normal law. This is indeed the case, and in 1939 Erdös and Kac proved the following theorem:

Let $K_n(\omega_1,\omega_2)$ be the number of integers *m*, $1 \le m \le n$, for which

$$\log\log n+\omega_1\sqrt{\log\log n}<\nu(m)<\log\log n+\omega_2\sqrt{\log\log n} \tag{3.32}$$

then,

$$\lim_{n\to\infty}\frac{K_n(\omega_1,\omega_2)}{n}=\frac{1}{\sqrt{2\pi}}\int_{\omega_1}^{\omega_2}e^{-y^2/2}dy. \tag{3.33}$$

Because of the slowness with which log log n changes the result (3.33) is equivalent to the statement:

$$D\{\log\log n + \omega_1\sqrt{\log\log n} < \nu(n) < \log\log n + \omega_2\sqrt{\log\log n}\} = \frac{1}{\sqrt{2\pi}}\int_{\omega_1}^{\omega_2} e^{-y^2/2}dy. \quad (3.34)$$

There are several different proofs of this result, but all of them are long and not elementary in their approach. Next we will follow classical result of Landau.

If $\pi_k(n)$ denotes the number of integers not exceeding n having exactly k prime divisors, then

$$\pi_k(n) \sim \frac{1}{(k-1)!}\frac{n}{\log n}(\log\log n)^{k-1}. \quad \text{[known as Landau's theorem]} \qquad (3.35)$$

For k = 1, this is the familiar prime number theorem; for k >1, (3.35) can be derived from the prime number-theorem by entirely elementary considerations.

Now

$$K_n(\omega_1,\omega_2) = \sum_{\log\log n + \omega_1\sqrt{\log\log n} < k < \log\log n + \omega_2\sqrt{\log\log n}} \pi_k(n), \qquad (3.36)$$

and hence one might expect that

$$\frac{K_n(\omega_1,\omega_2)}{n} \sim \frac{1}{\log n}\sum_{\log\log n + \omega_1\sqrt{\log\log n} < k < \log\log n + \omega_2\sqrt{\log\log n}} \frac{(\log\log n)^{k-1}}{(k-1)!} \qquad (3.37)$$

Using Markov's method, i.e. by proving Laplace's formula, namely,

$$\lim_{x\to\infty} e^{-x}\sum_{x+\omega_1\sqrt{x} < k < x+w_2\sqrt{x}} \frac{x^k}{k!} = \frac{1}{\sqrt{2\pi}}\int_{\omega_1}^{\omega_2} e^{-y^2/2}dy$$

and setting

$$x = \log\log n \ (e^{-x} = 1/\log n), \qquad (3.38)$$

we can obtain

$$\frac{K_n(\omega_1,\omega_2)}{n} \sim \frac{1}{\sqrt{2\pi}}\int_{\omega_1}^{\omega_2} e^{-y^2/2}dy \quad \text{or (3.33).}$$

Furthermore, let's prove (3.16) by letting $K_n(\omega_1,\omega_2)$ be the number of integers m, $1 \le m \le n$, for which

$$\nu(m) < \log\log n + \omega_2\sqrt{\log\log n},$$

and setting

$$\sigma_n(\omega) = K_n(\omega)/n. \qquad (3.39)$$

It is clear the is a distribution function, and

$$\frac{1}{n \log \log n}\sum_{m=1}^{n}\left(\nu(m) - \log \log n\right)^2 = \int_{-\infty}^{\infty}\omega^2 d\sigma_n(\omega). \tag{3.40}$$

If we use the precise estimate

$$\sum_{p<n,p} \frac{1}{p} = \log \log n + C + \varepsilon_n, \varepsilon_n \to 0, \tag{3.41}$$

then the argument that says almost every integer m has approximately log log m prime divisors gives

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}\omega^2 d\sigma_n(\omega) = 1 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^2 e^{-y^2/2}dy. \tag{3.42}$$

We have also

$$\lim_{n\to\infty}\frac{1}{\sqrt{\log \log n}}\sum_{m=1}^{n}\left(\nu(m) - \log \log n\right) = 0,$$

and hence

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}\omega\, d\sigma_n(\omega) = 0 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^2 e^{-y^2/2}dy. \tag{3.43}$$

If we could prove that for every integer k >2

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}\omega^k d\sigma_n(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^k e^{-y^2/2}dy. \tag{3.44}$$

it would follow that

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} e^{i\xi\omega} d\sigma_n(\omega) = e^{-\xi^2/2}$$

for every ξ and hence that

$$\lim_{n\to\infty}\sigma_n(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\omega} e^{-y^2/2}dy. \tag{3.45}$$

This, in view of (3.39), is nothing but the theorem. Proving (3.44) is, of course, equivalent to proving that

$$\lim_{n\to\infty}\frac{1}{n(\log \log n)^{k/2}}\sum_{m=1}^{n}\left(\nu(m) - \log \log n\right)^k = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^k e^{-y^2/2}dy. \tag{3.46}$$

and this in turn depends on asymptotic evaluations of sums

$$\sum_{p_{l_1}\cdots p_{l_k} < n} \frac{1}{p_{l_1}\cdots p_{l_k}}.$$

This, remarkably enough, is not at all easy, but Halberstam succeeded in carrying out the proof along these lines. This approach, without doubt, is the most straightforward and closest in spirit to the traditional lines of probability theory. The ultimate triumph of the probabilistic method in number theory came with the proof by Renyi and Turan that the error term

$$\frac{K_n(\omega)}{n} - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\omega} e^{-v^2/2}dv$$

is of the order of

$$\frac{1}{\sqrt{\log \log n}}.$$

That error is of order $(\log \log n)^{-1/2}$ was conjectured by Le Veque by analogy with similar estimates in probability theory - the primes, indeed, play a game of chance! For references to the work of Davenport, Erdős, Erdős and Kac, Halberstam, and Schoenberg and Turán we can confer to articles of Kac(1949), Kubilus(1956), and Rényi and Turán(1958).

The above theorem can be re-written in the following format:

Using (3.34) and from the fact that $M\{\omega(n)-v(n)\} < \infty$, deducing first that the density of the set of integers for which $\omega(n) - v(n) > g_n$, $g_n \to \infty$, is 0, we can get the result that

$$D\{ 2^{\log \log n + w_1\sqrt{\log \log n}} < d(n) < 2^{\log \log n + w_2\sqrt{\log \log n}} \} = \frac{1}{\sqrt{2\pi}}\int_{w_1}^{w_2} e^{-v^2/2}dv$$

where $d(n)$ denotes the number of divisors of n.

3.5 Some Characterizations of the Normal Probability Law

The characterization of a distribution is the investigation of those unique properties enjoyed by that distribution. Mathai and Pederzoli (1977) have compiled and put together their studies with recent research papers and published in a form of a monograph. In their monograph entitled "*Characterization of the Normal Probability Law*," they dealt thoroughly with the highly mathematical topic of characterization and try to motivate students to undertake research work in this area. Thus the material is developed from the very elementary level to the research level. Indeed, they work is interesting; through presenting different levels of exercises invites us to undertake serious research works.

There are properties that will uniquely determine a normal distribution, that is to say, the normal law is the only distribution to enjoy such properties. Investigation of such properties and the determination of the resulting distributions are known as *characterizations of distributions*.

There are two distinct methods developed one following the other: (i) the *functional equation method*, and (ii) the *axiomatic approach*.

The *functional equation method* is developed to its present format by Kagan, Linnik and Rao (1973). In their method they developed techniques in characterization problems as follows: (a) Use the properties and derive a functional equation. Then solve the functional equation for a unique solution by imposing additional conditions if necessary. (b) Use the properties to derive a difference or difference-differential equation, and then seek a unique solution for it. (c) Use the properties and analyze some general structures to classify or separate certain distributions.

The *axiomatic approach* is advanced by Mathai and Rathie (1975,1976). The axiomatic approach to characterization of normal law proceeds as follows: An axiomatic definition is provided for a *basic concept* itself such as variance, correlation, entropy, affinity, information and the like. In problems of this nature a few postulates are put forward and the resulting concepts are uniquely determined, thus providing axiomatic definitions for these measures. The main techniques used in the characterization of basic concepts are also the same as the techniques used in the characterizations of probability laws. Generally, moment generating function- $M_x(t)$- and characteristic function- $\phi_x(t)$- play a great role, i.e under certain conditions these determine the corresponding distributions uniquely through the uniqueness of properties of Laplace and Fourier transforms respectively. Mathai and Pederzoli(1977) remark that historically, the basic concepts are introduced mainly on the basis of intuitive or heuristic considerations.

As a result of the above methods, normal distribution is characterized through *linear independence* (Darmois,1951; Basu,1951; Linnik,1952; Skitovich,1954), *linear and quadratic independence* (Laha,1956,1957; Chanda,1955; Linnik,1956; Gordon,1968; Gordon and Mathai, 1972; Mathai,1977), *regression properties* (Laha and Lukacs,1960; Zinger and Linnik,1964; Mathai, 1967; Gordon,1968; Gordon and Mathai, 1972), *by solutions of certain functional equations* (Rao, 1967; Linnik,1960; Zinger and Linnik, 1955), *from the Student's law* (Mauldon, 1956; Kotlarski, 1966), *structural set-up* (Mathai, 1967; Patil and V. Seshadri 1963,1964), *maximization principle and other miscellaneous techniques* (Mathai, 1977; Mathai and Gordon, 1972).

3.5.1 The Functional Equation Method

3.5.1.1 Characterisation through structural set-up

Characterisation through structural set-up depends mainly on the conditional densities, that is, if (X,Y) is a stochastic vector and if f(x,y) is the joint density then f(x,y) = g(x/y) h(y) where g(x/y) is the conditional density of X given Y and h(y) is the marginal density of Y. Even though in general, by knowing the conditional density of X the marginal densities of X and Y can not be determined, but in certain cases if we know the conditional density has a certain structural set-up then the marginal densities are uniquely determined by this structural property.

Before giving the general characterisation theorem for the linear exponential family of distributions let us see the basic definition:

If a stochastic variable X has the probability function

$$f(x) = \begin{cases} \dfrac{a(x)e^{\theta x}}{g(\theta)} & \text{for } x \in S,\ g(\theta) > 0,\ \theta \in \Omega \\ 0 & \text{, elsewhere} \end{cases}$$

where S is a subset of the set of real numbers and g(θ) is the *normalising factor*, that is,

$$g(\theta) = \int_S a(x)\, e^{\theta x}$$

where S denotes the integral or summation depending upon X is continuous or discrete and Ω is some parameter space, then $\{f(x),\ \theta\in\Omega\}$ is said to be a *linear exponential family of distributions*. Binomial, Poisson, logarithmic, normal with location parameter, negative exponential, gamma with one parameter belong to the linear exponential family of distributions.

Next follows a general characterisation theorem for the linear exponential family of distributions:

Let $X, X_1, \ldots, X_n$ be independent non-degenerate continuous real stochastic variables whose probability functions do not vanish at the origin. Let the conditional distribution of X given

$X_1, \ldots, X_{n-1}$, $X + X_1 + \ldots + X_n$ have the structural form C(x,z) where $Z = X + X_1 + \ldots + X_n$. Let the conditional distributions of X_i have the structural forms $C_i(x_i, z)$ for all i and for every subset. If C(x,z) is such that,

$$\frac{C(x,z)\,C(x_1,z)\ldots C(x_n,z)\,C(0,z)}{C(0,z)\,C(0,z)\ldots C(z,z)\,C(z,z)} = \frac{h(x)\,h(x_1)\ldots h(x_n)}{h(z)}$$

for some non-negative function h(x) then $X, X_1, \ldots, X_n$ all belong to the linear exponential family and further $X_1, \ldots, X_n$ are identically distributed. As a corollary to this result a characterisation for the normal distribution can be derived. Let $X, X_1, \ldots, X_n$ be as defined in the above theorem. Let $C(x,z) = \text{const.}\ \exp[-(x-z/2)^2/2\sigma^2]$. Then $X, X_1, \ldots, X_n$ are identically normally distributed.

3.5.1.2 Characterisation through independence of linear forms

Characterisation of normal distribution through independence of linear forms is advanced mainly by Darmois(1951), Basu (1951), Linnik (1952), Skitovich(1954) and others. Next we will state the two important theorems of characterisation of normal distribution through independence of linear forms.

- Let $X_1, \ldots, X_k$ be a set of independent, but need not be identically distributed, stochastic variables and let $U = a_1X_1 + \ldots + a_kX_k$ and $V = b_1X_1 + \ldots + b_kX_k$ where the a's and b's are constant. If U and V are independently distributed then X_i for which $a_ib_i \neq 0$ is normally distributed.

- Let the linear forms $\sum_{i=1}^{\infty} a_ix_i$ and $\sum_{j=1}^{\infty} b_jx_i$ converge with probability one to the stochastic variables U and V respectively. Let U and V be independent. Let the sequences $\{a_i/b_i, a_ib_i \neq 0\}$ and $\{b_i/a_i, a_jb_i \neq 0\}$ be both bounded. Then for every j for which $a_ib_i \neq 0$, X_i is normally distributed.

3.5.1.3 Characterisation through independence of linear and quadratic forms

It is clear that if we have a simple random sample of size n from a normal population, then the sample mean and the sample variance are independently distributed. With respect to the characteristic property of normal distribution the following theorems depict the behaviour of sample mean and sample variance.

- Let $(X_1, \ldots, X_n)$ be a simple random sample from a population with distribution function F(x) and characteristic function $\phi(t)$. Then the sample mean $\bar{X}$ and the sample variance S^2 are independently distributed if and only if the population is normal. This theorem was first proved by Geary(1936) and later by Lukacs(1942)

- A necessary and sufficient condition for the independence of a linear statistic L and a quadratic statistic Q, where $L = a_1X_1 + \ldots + a_nX_n$ and $Q = \sum_{i=1}^{n} X_i^2 - L^2$ with $\sum_{i=1}^{n} a_i^2 = 1$ and with $(X_1, \ldots, X_n)$ a simple random sample of size n from some population with distribution function F(x), is that the population is normal.

- Let $(X_1, \ldots, X_n)$ be a simple random sample from a population where the second moment exists. Let $L = X_1 + \ldots + X_n$ and $Q = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}X_iX_j$. Let $B_1 = \sum_{i=1}^{n} a_{ii} \neq 0$ and $B_2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} = 0$.

  Then, L and Q are independently distributed if and only if the population is normal.

3.5.1.4 Characterisation through regression

If X and Y have a joint bivariate normal distribution then the regression of X on Y, that is, the conditional expectation of X given Y, is linear in Y. Next, we can note that linear regression with some properties of conditional variance implies normality. This result is verified in the following theorems.

- Let $(X_1, \ldots, X_n)$ be a simple random sample from a population with finite variance σ^2.

  $L = X_1 + \ldots + X_n$ and $Q = \sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik} X_i X_k + \sum_{j=1}^{n} b_j x_j$.

  and let B_1 $\sum_{i=1}^{n} a_{ii} \neq 0$, B_2 $\sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik} = 0$ and B_3 $\sum_{j=1}^{n} b_j = 0$.

  Then the population is normal if and only if Q has a constant regression on L.

- Let $X_1, \ldots, X_n$ be independent stochastic variables with finite variances. Consider the linear forms $L_1 = a_1X_1 + \ldots + a_nX_n$ $L_2 = b_1X_1 + \ldots + b_nX_n$ with $a_jb_j \neq 0$ for $j = 1, \ldots, n$. Then $E(L_1/L_2) = \alpha + \beta L_2$ and $var(L_1/L_2) = \sigma^2_0$ (costant) if and only if

  (a) the x_j for which $b_j \neq \beta a_j$ are normal and

  (b) $\beta = (\Sigma' a_j b_j \sigma^2_j) / (\Sigma' a^2_j \sigma^2_j)$, $\sigma^2_0 = \Sigma' (b_j - ba_j)^2 \sigma^2_j$

  where $\sigma^2_j = var(X_j)$ and Σ' indicates that the summation is taken over all j for which $b_j \neq \beta a_j$.

3.5.1.5 Characterisation by solutions of certain functional equations

As introductory clue and guiding results there are, among the many theorems, two important theorems, which use this approach for the characterisation of normal law, namely, Darmois-Skitovich(1951,1954) and Rao(1967). The former imposes a stronger condition of independence of linear forms, and states that if two linear forms, $a_1X_1 + \ldots + a_nX_n$ and $b_1X_1 + \ldots + b_nX_n$ where $a_ib_i \neq 0$, $i=1, \ldots, n$ of independent stochastic variables $X_1, \ldots, X_n$ are independently distributed then $X_1, \ldots, X_n$ are normally distributed. While the later applies a weaker condition that the regression of one linear form on the other is zero, i.e., $E(a_1X_1 + \ldots + a_nX_n / b_1X_1 + \ldots + b_nX_n) = 0$.

Furthermore the following theorems vividly represent the characterisation of normal law by solutions of certain functional equations deduced by Rao(1967):

- Let X_1, X_2 be two independently and identically distributed stochastic variables such that $E(X_1) = 0$. Let there exist linear functions $a_1X_1 + a_2X_2$ and $b_1X_1 + b_2X_2$ where $a_ib_i \neq 0$, $i=1,2$, such that

  $$E(a_1X_1 + a_2X_2 / b_1X_1 + b_2X_2) = 0$$

  and $|b_2/b_1| \leq 1$. Then,

  (i) If $|a_2/a_1| < 1$ or if $|a_2/a_1| = 1$ and $|b_2/b_1| < 1$, then X_1, X_2 have degenerate distributions.

  (ii) If $E(x_1^2) < \infty$, $a_1b_1 + a_2b_2 = 0$, and $|b_2/b_1| < 1$, then X_1, X_2 are normally distributed.

- Let $X_1, \ldots, X_n$ be independently and identically distributed stochastic variables such that $E(X_1) = 0$.

  Let.

$$U = a_1X_1 + \ldots + a_nX_n$$

and

$$V = b_1X_1 + \ldots + b_nX_n$$

such that $E(U/V) = 0$, $|b_n| > \max(|b_1|, \ldots, |b_{n-1}|)$ and $a_n \neq 0$. Let $E(x_1^2) < \infty$, $\Sigma a_ib_i = 0$ (or $E(x_1^2) \neq 0$) and $(a_ib_i/ a_nb_n) < 0$ for $i = 1, \ldots, n-1$, then the X_i are normally distributed.

- Consider the following conditions

  (i) $a_i \neq 0$, $i=1,2, \ldots, n$,

  (ii) b_i and c_i are not simultaneously zero for each i,

  (iii) $b_i \neq b_j$ for i and j such that b_i, b_j, c_i, c_j are all different from zero,

  (iv) all a_i defined are of the same sign and all δ_i defined are of the same sign.

  Then $E(U_1/U_1,U_3) = 0$ implies that $X_1, \ldots, X_n$ are all normally distributed.

3.5.1.6 Characterisation from the student's law

This procedure is mainly advanced by Kotlarski (1966).

If $(X_0,X_1, \ldots, X_n)$ $(n \geq 1)$ is a random sample of size $n+1$ from a normal population $N(0,\sigma^2)$ then it is well-known that

$$Y_1 = \frac{X_1\sqrt{1}}{|X_0|}, Y_2 = \frac{X_2\sqrt{2}}{\sqrt{X_0^2 + X_1^2}}, Y_3 = \frac{X_3\sqrt{3}}{\sqrt{X_0^2 + X_1^2 + X_3^2}}, \ldots, Y_n = \frac{X_n\sqrt{n}}{\sqrt{X_0^2 + X_1^2 + \ldots + X_{n-1}^2}} \quad (3.30)$$

are independently distributed as student-t variates with 1,2, ..., n degree of freedom respectively.

Kotlarski(1966) showed that, when $n \geq 2$, the above independent student variates characterise the normal variates $X_0, X_1, \ldots, X_n$ under some conditions on the distributions of $X_0, X_1, \ldots, X_n$. It is good to note that if a stochastic variable X is symmetric about the origin, the distribution of X is uniquely determined by the distribution of $U=X^2$. That is, there is a unique correspondence between the distributions of U and X when X is symmetric about the origin. Furthermore, if $X_0, X_1, \ldots, X_k$ are normal variates with zero mean and with the common variance σ^2 then evidently $U_i = X_i^2$, $i=0,1, ..., k$ have the gamma density function,

$$f(u) = \begin{cases} \dfrac{u^{1/2}}{\sigma\sqrt{2\pi}} e^{-(u/2\sigma^2)} & \text{for } u > 0 \\ 0 & \text{elesewhere, for all } i = 0,1, ..., k. \end{cases}$$

*H*is theorem is as follows:

- Let $X_0, X_1, ..., X_n$ be n+1 real independent stochastic variables ($n \geq 2$) such that $P(X_i=0) = 0$, $k=0,1,..., n$ and having distribution symmetric about zero, then the necessary and sufficient condition for $X_0, X_1, ..., X_n$ to be identically normally distributed as $N(0,\sigma^2)$ is that $Y_1,...,Y_n$ of (3.30) are independently distributed as student-t with 1,2, ..., n degree of freedom respectively. This theorem can be stated in terms of $U_1, ..., U_n$ as follows:
- Let $U_0, U_1, ..., U_n$ be n+1 real independent positive stochastic variables ($n \geq 2$). Let $V_1=U_1/U_0$, $V_2= U_2 / U_0+U_1$, $V_3= U_3 / U_0+U_1+U_2$, ..., $V_n= U_n / U_0+U_1+... +U_{n-1}$. Then a necessary and sufficient condition for U_k, $k = 0, 1, ..., n$ to be identically distributed according to the density

$$f(u) = \begin{cases} \dfrac{u^{-1/2}}{\sigma\sqrt{2\pi})} e^{-(u/2\sigma^2)} & \text{for } u > 0 \\ 0 & \text{, elesewhere, for all } i = 0,1, ..., k. \end{cases}$$

is that $V_1, ...V_n$ are independently distributed according to

$$g_k(v) = \begin{cases} \dfrac{\Gamma\left(\frac{k+1}{1}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\kappa}{2}\right)} v^{-1/2} (1+v)^{-(k+1)/2} & \text{for } v > 0 \\ 0 & \text{, elesewhere,} \end{cases}$$

correspondingly.

It must be noted that there are other methods for the characterszation of the normal distribution: Maximum likelihood characterization, characterization through admissibility of estimators, characterization through sample variance and so on. We note that since the variables Y_k, $k=1,2,..., n$ in (3.30) are symmetric about the origin the distributions of Y_k are uniquely determined by the distributions of $V_k=Y_k^2/k$

(k=1,2,...,n). When Y_k is student-t with k degrees of freedom then Y_k^2 has an F-distribution with 1 and k degrees of freedom and further, $V_k = Y_k^2 / k$ has the density

$$g_k(v) = \begin{cases} \dfrac{\Gamma\left(\frac{k+1}{1}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{K}{2}\right)} v^{-1/2}(1+v)^{-(k+1)/2} & \text{for } v > 0 \\ 0 & \text{, elesewhere,} \end{cases}$$

3.5.1.7 Characterisations of the multivariate normal law

As Mathai and Pederzoli(1977) and Rao (1973) verify most of the above results (section 3.5.1.1 up to 3.5.1.6) can be extended to cover the multivariate normal distribution. There is also one basic result which allows us to transform a characterisation problem on the multivariate normal to one on a univariate normal. That is, with the help of this result, which is due to Cramér(1937) and Wold(1938), many results of univariate normal can be generalised to the case of the multivariate normal and vice-versa. The result states that the distribution of a p-dimensional random(stochastic) variable $\underline{x}$ is completely determined by the *one*-dimensional distributions of linear functions $\underline{t}'\,\underline{x}$, for every fixed real vector $\underline{t}$.

This result in the case of normal distribution can be stated as follows. For an arbitrary real $px1$ vector of constants $\underline{t}$ if $\underline{t}'\,\underline{x}$ has a univariate normal distribution, then $\underline{x}$ has a p-variate normal distribution.

The following theorem is a generalisation of the theorem in section 3.5.1.1, characterisation through structural set-up, on the *independence of linear forms(statistics)* in scalar variables.

Let $X_1, \ldots, X_k$ be mutually independent vector stochastic variables each of order p. That is if we have X_i then its transpose form $X_i' = (X_{i1}, X_{i2}, \ldots, X_{ip})$ where X_{ij}, j=1, ..., p are scalar stochastic variables. Let $A_1, \ldots, A_k$ and $B_1, \ldots, B_k$ be pxp non-singular matrices. Let us consider $L_1 = \sum_{j=1}^{k} A_j X_j$ and $L_2 = \sum_{j=1}^{k} B_j X_j$.

- If the linear statistics L_1 and L_2 are independently distributed, then each vector X_i(i=1,..., k) has a p-variate normal distribution.

The following theorem corresponds to the characterisation of the univariate normal *through regression*, and is helpful in extending a number of results on the univariate normal to the case of multivariate normal.

- Let X, X_1, ..., X_n be n+1 symmetric stochastic matrices of the same finite order. Let

$$\Phi(T, T_1, ..., T_n) = e^{i\,tr(TX + \sum_{j=1}^{n} T_j X_j)}$$

where T, T_1, ..., T_n are square matrices of the real constants. Let $P(X_1, ..., X_n)$ and $Q(X_1, ..., X_n)$ be polynomials in X_1, ..., X_n of degrees p and q respectively.

Further let

$$P(iX_1, ..., iX_n) = i^{r-1} P(X_1, ..., X_n)$$

and

$$Q(iX_1, ..., iX_n) = i^{r} Q(X_1, ..., X_n) \text{ for some } r \geq 1.$$

Let

$$E(X, X_1^{r_1}, ..., X_1^{r_n}) \text{ for } r_1 + ... + r_n \leq p$$

and

$$E(X_1^{s_1}, ..., X_n^{s_n}) \text{ for } s_1 + ... + s_n \leq q \text{ exist.}$$

Then the necessary and sufficient condition for the regression of X on $X_1, ..., X_n$ to be of the form

$$E(X/ X_1, ..., X_n)\, P(X_1, ..., X_n) = Q(X_1, ..., X_n)$$

for all given X_1, ..., X_n is that

$$\left[\frac{\partial}{\partial T} P\left(\frac{\partial}{\partial T_1}, ..., \frac{\partial}{\partial T_n}\right) \Phi(T, T_1, ..., T_n)\right]_{T=0} = Q\left(\frac{\partial}{\partial T_1}, ..., \frac{\partial}{\partial T_n}\right) \Phi(0, T_1, ..., T_n)$$

where $\frac{\partial}{\partial T}\Phi = \frac{\partial \Phi}{\partial T_m}$ and $\left(\frac{\partial}{\partial T_j}\right)^r = \frac{\partial^r}{\partial T^r_j}$ for j = 1, ..., n.

The following two characterisation theorems can help us to obtain a number of similar characterisation theorems.

- Let $X_1, \ldots, X_n$ be a set of *n* independently and identically distributed stochastic $1xp$ vectors, each with finite $E(X)$, $E(X'X)$, and $E(XX'X)$. Suppose that,

  $n\beta_1-\Lambda_1=0$, $n(n+1)\beta_3-2\Lambda_1\Lambda_3=0$, $n(n^2-n+1)\beta_3-\Lambda_1-\Lambda_2-\Lambda_3-\Lambda_4=0$, $C-n\beta_1=\sigma(n^2\beta_3-\Lambda_1-\Lambda_2)\neq 0$ where $\sigma=\sum_{i=1}^{p}\sigma_{ii}$. Then, in the class of populations whose characteristic functions are pseudo-analytic of type I of the vector variable T, a necessary and sufficient condition for *S* to have cubic regression on *L* is that the pupulation be multivariate normal with characteristic function,

  $$\Phi(T)=\exp[\,i\mu T' - T(1/p)(\sigma I)T'\,]$$

  where $S=\Sigma_{j,k,m}a_{jkm}X_jX_k'X_m+\sum_{j=1}^{n}c_jX_j$ and $L=\sum_{j=1}^{n}c_jX_j$.

- Let $X_1, \ldots, X_n$ be a set of *n* independently and identically distributed stochastic $1xp$ vectors, each with finite $E(X)$, $E(X'X)$, and $E(XX'X)$. Suppose that,

$n\beta_1-\Lambda_1=0$, $n^2\beta_3-\Lambda_1-\Lambda_2=0$, $n(n^2-n+1)\beta_3-\Lambda_1-\Lambda_2-\Lambda_3-\Lambda_4=0$, $(C-n\beta_1)I=n(n+1)\beta_3-2\Lambda_1-\Lambda_3)\Sigma\neq 0$. Then, in the class of populations whose characteristic functions are pseudo-analytic of type II of the vector variable T, a necessary and sufficient conditon for S to have cubic regression on L is that the pupulation be multivariate normal with the characteristic function.

$$\Phi(T)=\exp[\,i\mu T' - (1/2)T\Sigma T'\,]$$

where Σ is a constant multiple of the identity.

3.5.2 Other Types of Characterizations

The above methods follow the first type of characterization, that is, the main technique used in this kind of characterisations is to employ the property so as to arrive at some differential equations, functional equations or some structural forms then obtain unique solutions of them.

The second type of characterization problems in statistics are characterisations of basic concepts such as varaince, covariance, correlation, entropy, affinity etc. These characterisations lead to axiomatic or mathematical definitons of these basic concepts. In order to use this method first it is required to define the various information and statistical measures, and then to list their general and particular properties, respectively. Then we can discuss the solutions of some functional equations which are useful for the characterisation of the various information and statistical measures, namely, functional equations in one variable, and two or more varaibles.

Characterization Theorems

There are many characterization theorems for information and statistical concepts. Let us define the concept of entropy: The *entropy* of the distribution P, in information measure, is defined as $H_n(p_1,...,p_n) = -\sum_{i=1}^{n} p_i \log p_i$. An important special case of this definition for n=2 is defined as *entropy function*, $H_2(p, 1-p) = -p \log p - (1-p) \log (1-p)$ for $p \in [0,1]$. The following are some of them deduced by Mathai and Rathie (1975,1976):

- Shannon's noiseless coding theorem states that the minimum of $\Sigma p_i n_i$ is the entropy $H_n(p_1,...,p_n)$ with equality iff $n_i = -\log p_i$ for all i. Here $p_1, ..., p_n$ are the probabilities of *n* input symbols $x_1, ..., x_n$ where x_i is represented by a sequence of n_i characters from the binary alphabet. Also it is assumed that n_i's satisfy the inequality $\Sigma 2^{-n_i} \leq 1$.
- If the function F_n satisfying the postulates

  (i) $F_n(p_1, ..., p_n)$ is a continuous function of its variables,

  (ii) $F_n(1/n, ..., 1/n)$ is a monotonic increasing function of n,

  (iii) $F_2(½, ½) = 1$, normalization principle,

  (iv) $F_n(p_1, ..., p_n) = F_n(p_{a1},...,p_{an})$, for every arbitrary permutation $\{a_1,...,a_n\}$ for $\{1,...,n\}$, and

  (v) $H_n(p_1, ..., p_{m-1}, p_m q_1, p_m q_2, ..., p_m q_{n-m+1}) = H_m(p_1, ..., p_m) + p_m H_{n-m+1}(q_1, q_2, ..., q_{n-m+1})$ where $\sum_{i=1}^{m} p_i = 1$ and $\sum_{i=1}^{n-m+1} q_i = 1$, then it is uniquely determined by the definition of entropy, in information measure, that is,

$$H_n(p_1,...,p_n) = -\sum_{i=1}^{n} p_i \log p_i .$$

- The postulates (i)$F_n(p_1, ..., p_n) = F_n(p_{a1}, ..., p_{an})$, (ii)$F_n(p_1, ..., p_n) = F_{n-1}(p_1+p_2, p_3,...,p_n) + A(p_1, p_2) F_2[p_1/(p_1, p_2), p_2/(p_1, p_2)]$ with $A(p_1,p_2) = p_1+p_2$, (iii)$F_2(p, 1-p)$ - Lebesgue integrability in [0,1], and (iv) $F_2(\frac{1}{2},\frac{1}{2})=1$, normalization principle, imply $H_n(p_1,\ldots,p_n) = -\sum_{i=1}^{n} p_i \log p_i$ and conversely.

- If the functions $F_n: S_n \to R$, $(n = 2,3,\ldots)$ satisfy the set of independent properties

  (i) $F_n(p_1,\ldots,p_n) = F_n(p_{\alpha 1},\ldots,p_{\alpha n})$,

  (ii) $F_n(p_1,\ldots,p_n) = F_{n+1}(p_1,\ldots,p_n,0)$, expansibility or zero-indifferent,

  (iii) for P, R, $U \in S_n$ and Q, S, $V \in S_m$, $F_{mn}(p_1q_1,\ldots,p_mq_n) = F_n(p_1,\ldots,p_n) + F_m(q_1,\ldots,q_m)$, the principle of additivity,

  (iv) for $p_{ij} \geq \sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} = 1$, $n, m \geq 2$, $F_{mn}(p_{11},\ldots,p_{1n},\ldots,p_{m1},\ldots,p_{mn}) \leq F_m(\sum_{j=1}^{n} p_{1j},\ldots,\sum_{j=1}^{n} p_{mj}) + F_n(\sum_{i=1}^{m} p_{i1},\ldots,\sum_{i=1}^{m} p_{in})$, rule of sub-additivity, and

  (v) $F_2(\frac{1}{2},\frac{1}{2})=1$, normalization principle, and $\lim_{p\to 0^+} F_2(p, 1-p) = F_2(0,1)$

  then, F_n is uniquely given by

$$H_n(p_1,\ldots,p_n) = -\sum_{i=1}^{n} p_i \log p_i.$$

- Let $n_1,\ldots,n_N$ satisfy the inequality $\sum_{i=1}^{N} 2^{-n_i} \leq 1$, then

$$t^{-1} \log(\sum_{i=1}^{N} p_i 2^{t n_i}) \geq H_{n,\alpha}(p_1,\ldots,p_N), \text{ where } \alpha = (1+t)^{-1}, 0 < t < \infty.$$

Chapter Four
THEORY OF NORMAL DISTRIBUTION IN HILBERT SPACE

4.1 DAVID HILBERT: Brief Life History, Works and Contributions

4.1.1 Brief Life History

David Hilbert was born January 23, 1862 in Königsberg, East Prussia- Germany. He attended Königsberg from 1882 to 1885, when he received his doctoral degree with a thesis on the theory of invariants. It was there that he established a life-long friendship with H. Minkowski. He was Lecturer(Assi. Prof.) 1886-1892, associate professor 1892/3; and in 1892 he became a professor at the University. He married to Käthe Jerosch in 1892 and got a child named Franz, and in 1895 he was appointed to a professorship at the University of Göttingen, a position he held until his death, February 14, 1943. Incidentally, we note that Gauss(1777-1855), Dirichlet(1805-1859) and Riemann(1826-1866) are all associated with the University of Göttingen. The authoritative biography of D Hilbert is written by Constance Reid, student and life long colleague; he wrote two biographical sketches: the first appeared in the 1922 *Naturwissenschaften* and the second at the end of the collected works(1970).

4.1.2 His works and Contributions

He obtained his basic theorem on invariant between 1890 and 1893 - that all invariant can be expressed in terms of a finite number, and hence he modified the mathematics of invariants; and next began research on the foundations of geometry and the theory of algebraic number fields. Concerning the former, he published *Grunlagen der Geometrie* (first edition 1899), in which he gave the complete axioms of Euclidean geometry and a logical examination of them. Concerning the latter, he systematised all the important known results of algebraic number theory in his monumental *Zahlbericht*

(1897). In Number theory, he enunciated his significant conjecture on class field theory. A substantial part of Hilbert's fame rests on a list of *23 research problems* he enunciated at the international congress of mathematicians held in Paris in 1900. In his address, "The Problems of Mathematics," he surveyed nearly all the mathematics of his day and endeavoured to set forth the problems he thought would be significant for mathematicians in the 20th century.

Between 1904 and 1906 he conducted research on the Dirichlet principle of potential theory and on the direct method in the calculus of variation. Around 1909 he established the foundations of *the theory of Hilbert spaces - infinite-dimensional space;* a concept that is useful in mathematical analysis, quantum mechanics and relativity theory. His works in the integral equations about 1909 led directly to 20th century research in functional analysis- the branch of mathematics in which functions are studied collectively. Also he proved in 1909 the conjecture in number theory that for any *n*, all positive integers are sums of a certain fixed number of n^{th} powers – e. g. $5 = 2^2 + 1^2$, in which n=2.

After 1910 he was chiefly involved in research on the foundations of mathematics, and he advocated the standpoint of formalism. Making use of his results on integral equations, Hilbert contributed to the development of mathematical physics by his important memoirs on kinetic gas theory and the theory of radiations. He addressed the citizens of the city of Kōnigsberg in 1930, after receiving his honorary citizenship, entitled by "The understanding of Nature and Logic"; and his last six words "we must know, we shall know" sum up his enthusiasm for mathematics and the devoted life he spent raising it to a new level. Indeed, by founding the formalist school of mathematical philosophy, contributing to many branches of mathematics, and presenting many illuminating mathematical papers. He is one of the greatest mathematicians on the first half of the 20th century.

The *Gesammelte Abhandlungen*, 3 Vols., 1932-5 (reprinted 1965, second edition 1971), contains almost all of Hilbert's papers, including *Zahlbericht*; there are also assessments of his works by other mathematicians. His Biography is written by his student and life long colleague Constance Reid(1996), first appearing in 1922 *Naturwissenschaften* and the second the collected works. Herman Weyl, Hilbert's leading student, in his article "Obituary Notice," *Bull Am. Math. Soc.* **50**(1944), pp.612-654, gives a definite assessment of Hilbert. Also *Gedenkband*(1971), edited by Kurt Reidemeister, contains some previously unpublished papers of Hilbert and the recording of his 1930 speech.

4.2. Fundamental Definitions, Properties and Axioms of Abstract Hilbert Space

4.2.1 General Remarks

The theory of Hilbert space arose from problems in the theory of integral equations. Hilbert noticed that a linear integral equation can be transformed into an infinite system of linear equations for the Fourier coefficients of the unknown function. He considered the linear space $\mathcal{L}_2$ consisting of all sequences of numbers $\{x_n\}$ for which $\Sigma_{n=1}^{\infty} |x_n|^2$ is finite, and defined for each pair of elements $x=\{x_n\}$, $y=\{y_n\} \in \mathcal{L}_2$ their inner product as $(x,y) = \Sigma_{n=1}^{\infty} x_n \bar{y}_n$. The space $\mathcal{L}_2$ can be regarded as an infinite-dimensional extension of the notion of a Euclidean space. In fact, Hilbert space is a direct generalisation of Euclidean space; hence, its "geometry" comes closer to Euclidean geometry than in the case of any other Banach-space. It possesses a great many of the properties of Euclidean space not possessed by Banach-spaces of general type.

A *Hilbert Space* is a Banach space whose norm has the *parallelogram property*. Any normed linear space over the reals which is complete in the topology determined by the norm is called (real) *Banach space*. From the definition of Hilbert space, it follows that any Banach space will be a Hilbert space provided that there is an inner product defined satisfying $\|f\|^2 = \langle f, f\rangle$

The question immediately arises as to whether or not all Banach spaces are Hilbert spaces; or is it always possible to define an inner product in a Banach space? We can settle this as follows. If there is to be an inner product, then

$$\|f+g\|^2 = \langle f+g, f+g\rangle = \langle f, f\rangle + 2\langle f, g\rangle + \langle g, g\rangle$$

$$\|f-g\|^2 = \langle f-g, f-g\rangle = \langle f, f\rangle - 2\langle f, g\rangle + \langle g, g\rangle$$

so that on adding

$$\|f+g\|^2 + \|f-g\|^2 = 2\|f\|^2 + 2\|g\|^2. \tag{4.1}$$

Thus the relation (4.1) for all f, g in the space is a necessary condition for the Banach space to have an inner product. It can be assumed of the condition (4.1) as a generalisation of the Euclidean theorem that in any parallelogram the sum of the squares on the diagonals is twice the sum of the squares on two adjacent sides. If this is not valid in the Banach space **K**, then it is not possible to define an inner product on **K**. This allows us to show that $\mathcal{L}_p$ is not a Hilbert space for $p\neq2$. The Banach space of normed linear functionals on a Banach space is said to be its adjoint; but a Hilbert space is adjoint to itself.

Riesz(1955) considered the space of functions now termed $\mathcal{L}_2$-space and succeeded in giving a satisfactory answer to the Fourier expansion problem. Abstract Hilbert spaces were introduced by von Neumann(1929). In his book, von Neumann(1932) established a rigorous foundation of quantum mechanics employing Hilbert spaces and the spectral expansion of self-adjoint operators. Weyl(1944) later justified the Dirichlet principle of Riemann by the method of orthogonal projection in a Hilbert space, and thus paved the way for the funcitonalytic study of differential equations. This has enabled functional analysis to be developed far more widely and completely in the context of Hilbert space than in the context of general normed spaces, so that Hilbert space theory has grown to be an important independent branch of functional analysis with its own results and methods.

4.2.2 Definition

a. Fundamentals

Before giving the axiomatic definition of Hilbert Space it is preferable and logical to deal with the fundamental elementary properties of functional analysis, namely, linear spaces, scalar product and some topological concepts.

4.2.2.1 Linear Spaces

A set R of elements $f, g, h, \ldots$, [also called points or vectors] forms a *linear space* if (i) there is an operation, called *addition* and denoted by the symbol +, with respect to which R is an abelian group;

(ii) multiplication of elements of R by (real or complex) numbers $\alpha, \beta, \gamma, \ldots$ is defined as that

$$\alpha(f+g) = \alpha f + \alpha g, \quad (\alpha+\beta)f = \alpha f + \beta f, \quad \alpha(\beta f) = (\alpha\beta)f, \quad 1.f = f, \quad 0.f = 0.$$

Elements $f_1, f_2, \ldots, f_n$ in R are *linearly independent* if the relation

$$a_1 f_1 + a_2 f_2 + \ldots + a_n f_n = 0 \tag{4.2}$$

holds only in the trivial case with $a_1 = a_2 = \ldots = a_n = 0$; otherwise $f_1, f_2, \ldots, f_n$ are linearly *dependent*. The left member of equation (4.2) is called *linear combination* of the elements $f_1, f_2, \ldots, f_n$.

A linear space **R** is *finite dimensional* and, moreover, *n-dimensional* if **R** contain n linearly independent elements and if any $n+1$ elements of **R** are linearly dependent. If a linear space has arbitrarily many linearly independent elements, then it is *infinite dimensional*.

4.2.2.2 Some Topological Concepts

Let us have a brief introduction to the study of point sets in an arbitrary metric space. By denoting a metric space by E and distance D[x,y] between two elements of E, we can recall that . If x_o is a fixed element of E, and ρ is a positive number, then the set of all points x for which $D[x, x_0] < \rho$ is called the *sphere* in E with *centre* x_0 and *radius* ρ; such a sphere us a neighbourhood, more precisely a ρ-neighbourhood of the point x_0.

A sequence of points $x_n \in E$ (n=1,2,3....) has the limit point $x \in E$, and is written as $x_n \to x$ or $\lim_{n \to \infty} x_n = x$ when $\lim_{n \to \infty} D[x_n, x] = 0$. This implies that $\lim_{m,n \to \infty} D[x_m, x_n] = 0$, where *m* and *n* tend to infinity independently. If this is true, then the sequence is called *fundamental*. Thus, by the triangle inequality, a fundamental sequence may or may not converge to an element of the space. A metric space E is called *complete* if every fundamental sequence in E converges to some element of the space.

If each neighbourhood of $x \in E$ contains infinitely many points of a set M in E, then x is called a limit point of M. If a set contains all its limit points, then it is said to be *closed*. The set consisting of M and its limit points is called the *closure* of M and is denoted by M. If the metric space E is the closure of some countable subset of E, then E us said to be *separable*. Thus, in a separable space there exists a countable set N such that, for each point $x \in E$ and each $\varepsilon > 0$, there exists a point $y \in N$ such that $D[x, y] < \varepsilon$.

b. Axiomatic Definition of Hilbert Space

The following axiomatic definition of Hilbert spaces is due to von Neumann(1932).

Let K be the field of complex or real numbers, the elements of which we denote by $\alpha, \beta, \ldots$ Let H be a linear space over K, and to any pair of elements $x, y \in H$ let there correspond a number $\langle x, y \rangle \in K$ satisfying the following five conditions:

(i) $\langle x_1 + x_2, y \rangle = \langle x_1, y \rangle + \langle x_2, y \rangle$,

(ii) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$,

(iii) $\langle x, y \rangle = \overline{\langle y, x \rangle}$,

(iv) $\langle x, x \rangle \geq 0$; and

(v) $\langle x, x \rangle = 0 \Leftrightarrow x = 0$.

Then we call H a *pre-Hilbert* space and $\langle x, y \rangle$ the *inner product* of x and y.

With the norm

$$\|x\| = \sqrt{\langle x, x \rangle},$$

H is a normed linear space. If H is complete with respect of the distance

$$\|x - y\|, \text{ i.e. } \|x_n - y_m\| \to 0 \ (m, n \to \infty)$$

implies the existence of $\lim x_n = x$, then we call H a **Hilbert space**. According to as K is complex or real we call H a *complex* or *real Hilbert space*, in which case axiom (iii) becomes $\langle x, y \rangle = \langle x, y \rangle$ for all $x, y \in H$.

Hilbert Space can be also defined as follows: "A **Hilbert Space** H is an infinite dimensional inner product space which is a complete metric space with respect to the metric generated by the inner product." This definition has an axiomatic character. Various concrete linear spaces satisfy the conditions in the definition. Therefore, H is often called an **abstract Hilbert** space, and the concrete spaces mentioned are called **examples** of this abstract space.

4.2.2.3 Example of Abstract Hilbert Space

One of the important examples of H is the space $\mathcal{L}_2$. The construction of the general theory was begun for this particular space by Hilbert in connection with his theory of linear integral equations.

The elements of the space $\mathcal{L}_2$ are sequehces of real or complex numbers

$$f = \{x_n\}_{n=1}^{\infty}, \quad g = \{y_n\}_{n=1}^{\infty}, \ldots,$$

such that

$$\Sigma_{n=1}^{\infty} |x_n|^2 < \infty, \quad \Sigma_{n=1}^{\infty} |y_n|^2 < \infty, \ldots .$$

The number $x_1, x_2, x_3, \ldots$, are called components of the vector f or co-ordinates of the point f. The zero vector is the vector with all components zero. the addition of vectors is defined by the formula

$$f + g = \{x_n + y_n\}_{n=1}^{\infty}.$$

The relation

$$\Sigma_{n=1}^{\infty} | x_n + y_n |^2 < \infty$$

follows from the inequality

$$|x_n + y_n|^2 \leq 2|x_n|^2 + 2|y_n|^2.$$

The multiplication of a vector f by a number λ is defined by

$$\lambda f = \{\lambda x_n\}_{n=1}^{\infty}.$$

The scalar product in the space $\mathcal{L}_2$ has the form

$$(f,g) = \Sigma_{n=1}^{\infty} x_n \bar{y}_n.$$

The series on the right converges absolutely because

$$|xy| \leq ½|x|^2 + ½|y|^2.$$

The inequality

$$|(f,g)| \leq \|f\| \cdot \|g\|$$

now has the form

$$|\Sigma_{n=1}^{\infty} x_n \bar{y}_n| \leq (\Sigma_{n=1}^{\infty}|x_n|^2)^{½} \cdot (\Sigma_{n=1}^{\infty}|y_n|^2)^{½}.$$

and is due to Cauchy.

The space $\mathcal{L}_2$ is *separable*. A particular countable dense subset of $\mathcal{L}_2$ consists of all vectors with only finitely many nonzero components and with these components rational, i.e., the components are of the form $\xi + i\eta$ where ξ and η are rational numbers. In addition to this, the space $\mathcal{L}_2$ is complete. In fact, if the sequence of vectors

$$f^{(k)} = \{x_n^{(k)}\}_{n=1}^{\infty} \ (k = 1, 2, 3, ...)$$

is fundamental, then each of the sequences of numbers

$$\{x_n^{(k)}\}_{n=1}^{\infty} \ (n = 1, 2, 3, ...)$$

is fundamental and, hence, converges to some limit x_n ($n = 1, 2, 3, ...$).

Proof.

Now, for each $\varepsilon > 0$ there exists an integer N such that for $r > N$, $s > N$

$$\sqrt{\sum_{n=1}^{\infty}\left|x_n^{(r)} - x_n^{(s)}\right|^2} < \varepsilon.$$

Consequently, for every m,

$$\sqrt{\sum_{n=1}^{m}\left|x_n^{(r)} - x_n^{(s)}\right|^2} < \varepsilon.$$

Let s tend to infinity to obtain

$$\sqrt{\sum_{n=1}^{m}\left|x_n^{(r)} - x_n\right|^2} \leq \varepsilon.$$

But, because this is true for every m,

$$\sqrt{\sum_{n=1}^{m}\left|x_n^{(r)} - x_n\right|^2} \leq \varepsilon.$$

Hence, it follows that

$$f = \{x_n\}_{n=1}^{\infty} \in \mathcal{L}_2,$$

and, since $\varepsilon > 0$ is arbitrary, $f^{(k)} \to f$. Thus the completeness of the space $\mathcal{L}_2$ is established. In the definition of an abstract Hilbert space the requirement of separability is not included but completeness is included, since it is essential for almost all of our considerations.

The space $\mathcal{L}_2$ is infinite dimensional because the *unit vectors*

$$e_1 = \{1,0,0, \ldots\}, \quad e_2 = \{0,1,0, \ldots\}, \quad e_3 = \{0,0,1, \ldots\}, \ldots\ldots,$$

are linearly independent. The space $\mathcal{L}_2$ is the infinite dimensional analogue of E_m, the (complex) m-dimensional *Euclidean space*, the elements of which are finite sequences

$$f = \{x_n\}_1^m,$$

and most of the theory which we present consists of generalisations to H of well-known facts concerning E_m.

4.3 The Geometry of Hilbert Space

In this section we will have a quick introduction to the important results concerning the geometry of Hilbert space. That is to say the subject of discussion is the geometry of linear spaces, in which a scalar product is defined in a certain axiomatic way and the norm is derived from this scalar product as in the geometric vector space. We will observe that the normed spaces thus obtained have richer structure and are more similar to the geometric vector space that those not having this property.

4.3.1 Scalar Product

The scalar product of vectors a and b in the gometric vector space is defined by

$$(a/b) := \|a\| \, \|b\| \cos \gamma$$

where $\|\cdot\|$ is the absolute value of the vector and γ is the angle between a and b.

A linear space R is metrizable if for each pair of elements $x, y \in R$ there is a (real or complex) number $\langle x, y \rangle$ which satisfies the conditons:

(i) $\langle x, y \rangle = \overline{\langle y, x \rangle}$,

(iia) $\langle \alpha_1 x_1 + \alpha_2 x_2, y \rangle = \alpha_1 \langle x_1, y \rangle + \alpha_2 \langle x_2, y \rangle$,

(iii) $\langle x, x \rangle \geq 0$, with equality only for $x=0$.

The number $\langle x, y \rangle$ is called the *scalar product* or *inner product* of x and y. Property (ii) expresses the linearity of the scalar product with respect to its first argument. The analogous property with respect to the second argument is

(iib) $\langle x, \beta_1 y_1 + \beta_2 y_2 \rangle = \overline{\beta}_1 \langle x, y_1 \rangle + \overline{\beta}_2 \langle x, y_2 \rangle$.

If a scalar product is defined in a linear space X then X is called a *scalar product space* or *pre-Hilbert space.*

This property is derived from (i) above. The positive square root $\sqrt{\langle x, x \rangle}$ is called the *norm (the absolute value)* of the element(vector) x and is denoted by the symbol $\|x\|$. The norm is analogous to the length of a line segment. As with line segments, the norm of a vector is zero if and anly if it is the zero vector. In addition, it follows that

$$\|\alpha x\| = |\alpha| \cdot \|x\|.$$

This can be verified using (iia) and (iib) conditions.

The norm in a Hilbert space was defined by means of the inner product. In turns out that the inner product can be recovered from the norm.

Another very important theorem, associated with the above conditons and properties, is the *Cauchy-Schwarz-Bunyakovski Inequality.*

For any two vectors x and y,

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|,$$

with equality if and only if x and y are linearly dependent. Furthermore, it is good to mension another property of the norm, namely, *Triangle inequality*

$$\|x + y\| \leq \|x\| + \|y\|.$$

Within the general framework of scalar product the following theorems and properties can be described:

(a) $\|x\| = \langle x, y \rangle^{1/2}$ is a norm;

(b) If $x_n \to x$ and $y_n \to y$, then $\langle x_n, x_n \rangle \to \langle x, y \rangle$;

(c) The completion of a scalar product space is a scalar product space with the scalar product - a complete scalar product space is called a Hilbert space.

4.3.2 Orthogonality

Inner product spaces allow us to introduce the important notion of orthogonality. An inner product space R becomes a metric space, if the distance between two points x, $y \in R$ is defined as $D[x,y] = \|x - y\|$. It follows from the properties of the norm that the distance function satisfies the usual conditons.

These conditions are:

(i) $D[x,y] = D[y,x] > 0$ for $x \neq y$,

(ii) $D[x,x] = 0$,

(iii) $D[x,y] \leq D[x,h] + [h,y]$ (triangular inequality).

Two vectors $x, y \in H$ are *orthogonal*, $x \perp y$, if $\langle x, y \rangle = 0$. Given a set *M* we write $x \perp M$ if $x \perp m$ for all $m \in M$. A set of vectors $\{x_\alpha\}$ is called an *orthogonal* set if $\langle x_\alpha, x_\beta \rangle = 0$ whenever $\alpha \neq \beta$. A vector x is normalized if $\|x\| = 1$. An orthonormal set is defined as an orthogonal set of normalized vectors. Thus $\{e_\alpha\}$ is an *orthonormal set* if $\langle e_\alpha, e_\beta \rangle = \delta_{\alpha\beta}$.

4.3.2.1 Orthoganal Systems

In the geometric vector space, every vector is the linear combination of fixed sets of three orthogonal vectors and any finite-dimensional linear space possesses a finite basis, i.e. n fixed vectors such that every vector of the linear space is a linear combination of these fixd vectors. In certain infinite-dimensional spaces a fixed infinite sequence of elelments can be found with similar properties.

A sequence $\{e_k\}$ in a scalar product space H is called *orthogonal* if $\langle x_\alpha , x_\beta \rangle = 0$ if $\alpha \neq \beta$. If $\| e_k \| = 1$ for i = 1, 2, ... is also satisfied, then $\{e_k;\ k=1, 2, ...\}$ is called *orthonormal*. An orthogonal or orthonormed sequence is also called an *orthogonal* or *orthonormal system*.

Next, our main object is the construction of an infinite basis in a separable Hilbert space or in a separable scalar product space.

Let $\{e_k\}$ be an orthonormal system, n a fixed integer, and x an element of H; then we can determine the scalars γ_k ; k = 1,2, ..., n in such a way that the distance

$$\left\| x - \sum_{k=1}^{n} \gamma_k e_k \right\|$$

is *minimal*.

The solution of a real scalar product space is as following:

$$\left\| x - \sum_{k=1}^{n} \gamma_k e_k \right\|^2 = \left\langle x - \sum_{k=1}^{n} \gamma_k e_k , x - \sum_{k=1}^{n} \gamma_k e_k \right\rangle$$

$$= \langle x, x \rangle - 2\sum_{k=1}^{n} \gamma_k \langle x, e_k \rangle + \sum_{k=1}^{n} \gamma_k^{\,2}$$

For the minimum of this quadratic form

$$\frac{\partial}{\partial \gamma_k} \left\| x - \sum_{k=1}^{n} \gamma_k e_k \right\|^2 = 2\gamma_k - 2\langle x, e_k \rangle = 0 \quad k = 1,2, ..., n$$

it follows that the desired minimum is obtained if and only if

$$\gamma_k = \langle x, e_k \rangle \quad k = 1, 2, ...$$

In the case of complex space H the solution is the same but a more lengthy calculation is required, since in this case we have to seek the minimum of a quadratic form of *2n real variables*.

It can be conclude that the minimum of the distance $\left\| x - \sum_{k=1}^{n} \gamma_k e_k \right\|$ is obtained if and only if

$$\gamma_k = \langle x, e_k \rangle, \qquad k = 1, 2, \ldots$$

and then

$$\left\| x - \sum_{k=1}^{n} \langle x, e_k \rangle e_k \right\|^2 = \|x\|^2 - \sum_{k=1}^{n} \left| \langle x, e_k \rangle \right|^2.$$

As a remark it can be said that the consideration connected with the Cauchy-Schwarz inequality is the special case of this problem when n=1. As a consequance of this fact we can deduce the following results:

- If e_k is the infinite sequence whose k^{th} element is 1 and all other elements are 0, then $e^k \in \mathcal{L}_2$ and $\{e_k;\ k=1,2,\ldots\}$ is an orthonormal system in $\mathcal{L}_2$.
- The sequence $\left\{ \frac{1}{(2\pi)^{1/2}} e^{ikt};\ k = 0, \pm1, \pm2, \ldots \right\}$ is an orthonormal system in $\mathcal{L}_2^0\ [0,2\pi]$

  and

$$e_k = e^{ikt},$$

$$\langle x, e_k \rangle = \frac{1}{(2\pi)^{1/2}} \int_0^{2\pi} x(t) e^{-ikt} dt,$$

and

$$\sum_{k=-n}^{n} \langle x, e_k \rangle e_k$$

is the n^{th} partial sum of the (complex) Fourier series of $x \in \mathcal{L}_2\ [0,2\pi]$. On the basis of this result, the coefficients, $k = 1,2, \ldots$ can be considered as the generalisation of the Fourier coefficients. These coefficients are therefore called the Fourier coefficients of x with respect to the orthogonal system $\{e_k;\ k = 1,2, \ldots\}$.

- If $x \in H$ space can be given in the form of an orthogonal seriess, then

$$x = \sum_{k=1}^{\infty} \langle x, e_k \rangle e_k \tag{4.3}$$

- An element x of the pre-Hilbert space H can be given in the form (4.3) if and only if

$$\|x\|^2 = \sum_{k=1}^{\infty} \left| \langle x, e_k \rangle \right|^2.$$

4.3.2.2 Complete Orthonormal Systems

A sequence $\{a_k; k=1,2,\ldots\}$ is called *complete* if $\langle x, a_k \rangle = 0$ for $k=1,2,\ldots$ implies $x=0$. Furthermore, an orthonormal system M is complete in H if M is not contained in a larger orthonormal system in H, i.e., if there is no nonzero vector in H which is orthogonal to every vector of the system M. In this section we shall see theorems and examples related to the complete orthonormal systems, as well as show that any separable Hilbert space contains a ***complete orthonormal system***. In Hilbert space a complete orthonrmal system contains an infinite number of elements, and there arises the problem of the cardinality of such systems. This problem is solved easily for separable spaces.

- For every $x \in H$,

$$x = \sum_{k=1}^{\infty} \langle x, e_k \rangle e_k$$

if and only if the orthonormal system $\{e_k\}$ is complete.

A standard method for the construction of orthonormal systems, called the *Gram-Schmidt process*, proceeds as follows. In the *Gram-Schmidt process*, n linearly independent vectors a_k; $k = 1, 2, \ldots, n$ are converted into the n elements of an orthonormal system $\{e_k; k=1,2,\ldots\}$. In this processs e_1 is the scalar multiple of a_1, e_2 is a linear combination of a_2 and e_1, e_3 is a linear combination of a_3, e_1 and e_2 and so on. The computation is organized in terms of the minimal number of vectors and operations. Let $\{a_k; k=1,2,\ldots\}$ be linearly independent. The first member of the orthonormal system is

$$e_1 = \frac{a_1}{\|a_1\|}$$

and for the second member e_2,

$$z_2 = a_2 - \lambda_{21} e_1$$

where the scalar λ_{21} is determined by the condition

$$\langle z_2, e_1 \rangle = \langle a_2, e_1 \rangle - \lambda_{21} = 0$$

and hence

$$\langle a_2, e_1 \rangle = \lambda_{21}.$$

So if $e_2 = z_2 / ||z_2||$ then $\{e_1, e_2\}$ is an orthonormal system with two elements.

For the third member, e_3,

$$z_3 = a_3 - \lambda_{31}e_1 - \lambda_{32}e_2$$

where the scalars λ_{31} and λ_{32} are determined by the conditions

$$\langle z_3, e_1 \rangle = \langle a_3, e_1 \rangle - \lambda_{31} = 0$$

$$\langle z_3, e_2 \rangle = \langle a_3, e_2 \rangle - \lambda_{32} = 0.$$

Hence $\lambda_{31} = \langle a_3, e_1 \rangle$ and $\lambda_{32} = \langle a_3, e_2 \rangle$. So if $e_3 = z_3 / ||z_3||$ then $\{e_1, e_2, e_3\}$ is an orthogonal system with three members.

Now if $e_1, e_2, \ldots, e_n$ have already been obtained, then for e_{n+1},

$$z_{n+1} = a_{n+1} - \sum_{k=1}^{\infty} \lambda_{n+1,k} e_k$$

where the sclars $\lambda_{n+1,k}$; $k = 1, 2, \ldots, n$ are determined by the condition

$$\langle z_{n+1}, e_k \rangle = \left\langle a_{n+1} - \sum_{k=1}^{n} \lambda_{n+1,k}\, e_k, e_k \right\rangle = \langle a_{n+1}, e_k \rangle - \lambda_{n+1,k} = 0.$$

and hence

$$\lambda_{n+1,k} = \langle a_{n+1}, e_k \rangle \quad k = 1, 2, \ldots, n.$$

So if

$$e_{n+1} = z_{n+1} / ||z_{n+1}||,$$

then $\{ e_k;\ k = 1, 2, \ldots, n+1 \}$ is an orthonormal system obtained from the lineear space generated by the n+1 vectors a_k; k+1, k+2, ..., n+1.

Sometimes the above method is called *orthogonalization*.

Using the Gram-Schmidt process for $1, t, t^2, \ldots, t^n, \ldots$ in $\mathcal{L}_2\,[-1,+1]$, a sequence of orthogonal polynomials known as Legendre polynomials is obtained, the n^{th} element of which is of exactly (n-1) th degree; the first four members of this sequence are

$$1,\ \ t,\ \ \tfrac{1}{2}(3t^2 - 1),\ \ \tfrac{1}{2}\,5(t^3 - 3t).$$

A complete orthonormal system is also called an *orthonormal (orthogonal) basis* since it is a basis for the scalar product space *H*. Important theoretical conclusions of the above are as follows:

- Every separabble scalar product space contains a (finite or infinite) basis.
- If the space *H* is separable, then every orthonormal system of vectors in *H* consists of a finite or countable number of elements.
- An infinite orthonormal sequence e_1, e_2, e_3, ... is complete in *H* if and only if the sequence is closed in *H*.
- The space *H* contains a complete orthonormal sequence if and only if it is separable.
- Any two complete orthonormal systems in a Hilbert space have the same cardinal number.

There are interesting results and inequalities about orthogonal sets and orthogonality: Pythagorean theorem, polarization identiy, parallelogram identity, Bessel's inequality etc.

- Let $\{x_i\}_{i=1}^{n}$ be an orthogonal set in the inner product space *H*, then

$$\left\|\sum_{i=1}^{n} x_i\right\|^2 = \sum_{i=1}^{n} \|x_i\|^2 \qquad \text{[Pythagorean theorem]}$$

- For all x, y in *H* we have

[Schwarz inequality]

- For all x, y $\in$ *H* we have

$$\|x+y\| \le \|x\| + \|y\| \qquad \text{[Triangle inequality]}$$

- For all vectors x, y $\in$ *H*

$$(x,y) = \frac{1}{4}\left\{\|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2\right\} \qquad \text{[Polarization identiry]}$$

- For all vectors x, y $\in$ *H* we have

$$\|x+y\|^2 + \|x-y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right) \qquad \text{[Parallelogram idenity]}$$

- Let $\{e_i\}$ be any orthonormal set then for each vector x $\in$ *H* we have

$$\|x\|^2 \ge \sum |(x, e_i)|^2 \qquad \text{[Bessel's inequality]}$$

- The inner product is a continuous function in each of its variables.

4.3.3 The projection principle

The projection principle for finite and infinite-dimensional subspace can be explained as follows:

Let M be a linear subspace of the pre-Hilbert space M; then $x_m \in M$ is called the nearest *vector* or the *best approximation* of $x \in M$ if

$$\|x - x_m\| \leq \|x - m\| \quad m \in M.$$

Let M be a linear subspace of the pre-Hilbert space H, then $x_p \in M$ is called the (orthogonal) *projection* of $x \in M$ if, for every $m \in M$,

$$\langle x - x_p, m \rangle = 0.$$

Through these abstract formulations we have a connection between projection and best approximation similar to that in the geometric vector space.

The following results of finite and infinite dimensional projection principles can be listed below:

- $x_p \in M$ is the best approximation of $x \in M$ if and only if it is the orthogonal projection of x in M.
- If M is the linear subspace generated by $\{a_1, a_2, \ldots, a_n\}$ then the orthogonal projection of $x \in M$ onto M is

$$x_p = \sum_{k=1}^{n} \gamma_k a_k$$

where $\{\gamma_k;\ k=1,2, \ldots, n\}$ is the solution of the following system of linear equations:

$$\sum_{k=1}^{n} \langle a_k, a_i \rangle \gamma_k = \langle x, a_i \rangle \qquad i=1, 2, \ldots, n.$$

- $\sum_{k=1}^{n} \gamma_k a_k = 0$ if and only if $\{\gamma_k;\ k=1,2, \ldots, n\}$ is the solution of the system of linear equations

$$\sum_{k=1}^{n} \langle a_k, a_i \rangle \gamma_k = 0 \quad i=1, 2, \ldots, n.$$

- If M is a complete subspace of a scalar product space M, then there exists a projection $x_p \in M$

Projection Theorem

Next we see the Projection Theorem in the light of probability theory. We should note that Loeve(1963) have already dealt with the property of orthogonality and projection theorem in the light of probability theory.

Before stating the projection theorem let us define important terms with respect to the probability theory. X and Y are *orthogonal*, and we write $X \perp \bar{Y}$, - the *bar* means complex-conjugate - if $E(X\bar{Y}) = 0$. In particular, $X \perp Y$ if and only if $E|X|^2 = 0$, that is X = 0 almost surely; in fact, X=0 almost surely is orthogonal to every Y. A *linear subspace* $\mathcal{L}$ is a family of random variables closed under formation of all almost surely linear combinations of its elements. If, also, $\mathcal{L}$ is closed under passages to the limit in quadratic mean, then it is a *closed linear subspace*. A random variable X is *orthogonal* to $\mathcal{L}$ and we write $X \perp \mathcal{L}$, if $X \perp Y$ whatever be $Y \in \mathcal{L}$.

Projection Theorem - Let X be a closed linear subspace. For every X there exists an almost surely unique orthogonal decomposition

$$X = X' + X'', \quad X' \perp \mathcal{L}, \quad X' \in \mathcal{L}.$$

It is good to note that within a strongly normal family, *orthogonality* is equivalent to *independence* and projection is equivalent to conditioning.

If we consider sequences, and more generally random functions, formed by random variables whose second moments and hence mixed second moments are finite, then random variables in question can be interpreted as points in a Hilbert space, and such spaces are a natural generalization of Euclidean spaces for which all the classical tools were developed.

the operator T. This extension is uniquely defined for each operator T. The *extension by continuity of a functional* is defined analogously.

4.4.2 Operators

Before dealing with operators, as a bridge, let us see the *definition* and *properties of linear functionals.*

A functional Φ is said to be *linear* if:

(i) its domain D is a linear manifold and $\Phi(af + bg) = a\,\Phi(f) + b\,\Phi(g)$

for $f, g \in D$ and any complex numbers a and b;

(ii) the inequality $\sup_{f \in D, \|f\| \le 1} |\Phi(f)| < \infty$ is satisfied.

The left member of this inequality is called the *norm* of the functional Φ and is denoted by the symbol $\|\Phi\|_D$ or, if D=H, simply by $\|\Phi\|$.

If $f \in D$ and $f \neq 0$ then, by the definition of the norm of a functional,

$$\left|\Phi\left(\frac{f}{\|f\|}\right)\right| \le \|\Phi\|_D.$$

Hence for $f \in D$,

$$|\Phi(f)| \le \|\Phi\|_D \cdot \|f\|. \qquad (4.4)$$

Relation (4.4) shows that the linear functional Φ continuous. In fact, by (4.4)

$$|\Phi(f) - \Phi(f_0)| = |\Phi(f - f_0)| \le \|\Phi\|_D \cdot \|f - f_0\| \quad \text{for } f, f_0 \in D.$$

From (4.4) it also follows that, if $f \in D$ and $\|f\| \le 1$, then $|\Phi(f)| \le \|\Phi\|_D$,

with strict inequality if $\|f\| < 1$. Therefore, the norm $\|\Phi\|_D$ can be defined by

$$\sup_{f \in D, \|f\| = 1} |\Phi(f)| = \|\Phi\|_D$$

or equivalently, by

$$\sup_{f \in D} \frac{|\Phi(f)|}{\|f\|} = \|\Phi\|_D.$$

4.4.2.1 Linear and Bounded Linear Operators

An operator T is *linear* if its domain of definition D is a linear manifold and if

$$T(\alpha f + \beta g) = \alpha Tf + \beta Tg$$

for any $f, g \in D$ and any complex numbers α and β.

A linear operator T is *bounded* if

$$\sup_{f\in D,\|f\|=1} \|Tf\| < \infty.$$

The left member of this inequality is called the norm of the operator T in D and is denoted by the symbol $\|T\|$ or, sometimes, by $\|T\|_D$.

The properties of linear functionals are also valid for bounded linear operators, namely:

a) The norm of a bounded linear operator T can be defined equivalently by

$$\|T\| = \sup_{f\in D,\|f\|\le 1} |Tf| = \sup_{f\in D} \frac{\|Tf\|}{\|f\|}$$

b) A bounded linear operator is continuous.

c) If a linear operator is continuous at one point, then it is bounded.

d) The extension by continuity of a bounded linear operator T leads to a unique linear operator with the same norm as the original operator.

e) If S and T are linear Operators, then $\alpha S + \beta T$, where α and β are complex numbers, is a linear operator with the intersection $D_S \cap D_T$ of the domains D_S and D_T as the domain of definition. Each of the products ST and TS is also a linear operator. If S and T are bounded linear operators defined everywhere in H, then the operators ST and TS are also bounded linear operators defined everywhere in H, and

$$\|ST\| \le \|S\|\cdot\|T\|, \quad \|ST\| \le \|T\|\cdot\|S\|.$$

4.2.2 Bilinear Functionals

Ω is a bilinear function defined in H, if to each pair of elements $f,g\in H$ there corresponds a definite complex number $\Omega\langle f, g\rangle$, and

i) $\Omega\langle a_1f_1 + a_2f_2, g\rangle = a_1\Omega\langle f_1, g\rangle + a_2\Omega\langle f_2, g\rangle$

ii) $\Omega\langle f, b_1g_1 + b_2g_2\rangle = \bar{b}_1\Omega\langle f, g_1\rangle + \bar{b}_2\Omega\langle f, g_2\rangle$

iii) $\sup_{\|f\|=1,\|g\|=1} |\Omega\langle f, g\rangle| < \infty$

An example of a bilinear functional is the scalar product $\langle f, g \rangle$. The number $\sup_{\|f\|=\|g\|=1} |\Omega\langle f, g\rangle|$ is called the *norm* of the bilinear functional Ω, and is denoted by $\|\Omega\|$.

We note that

$$\|\Omega\| = \sup_{\|f\|=\|g\|=1} |\Omega\langle f, g\rangle| = \sup \frac{|\Omega\langle f, g\rangle|}{\|f\| \cdot \|g\|}.$$

Therefore, for any $f, g \in H$,

$$\Omega|\langle f, g\rangle| \leq \|\Omega\| \cdot \|f\| \cdot \|g\|.$$

Each bilinear functional $\Omega\langle f, g \rangle$ has a representation of the form

$$\Omega\langle f, g \rangle = \langle Af, g \rangle.$$

In this equation A is a bounded linear operator with domain H which is uniquely determined by Ω. Furthermore, $\|A\| = \|\Omega\|$.

4.4.2.3 Adjoint operators

Let A be an arbitrary bounded linear operator defined on H. The expression $\langle f, Ag \rangle$ defines a bilinear functional on H with norm $\|A\|$. From the above result we note that there exists a unique bounded linear operator A^* defined on H with norm $\|A^*\| = \|A\|$ such that

$$\langle f, Ag \rangle = A^*\langle f, g \rangle \text{ for } f, g \in H.$$

This operator A^* is called the *adjoint* of A. We note that the operator $(A^*)^* = A^{**}$ is equivalent to the original operator A. If A is bounded and $A^* = A$, then A is said to be *self-adjoint*. A bounded linear operator A, defined on H, is said to be *normal* if it commutes with its adjoint, i.e., if $A^*A = AA^*$.

Let A and B be two bounded linear operators defined on H. Then,

$$\langle ABf, g \rangle = \langle Bf, A^*g \rangle = \langle f, B^*A^*g \rangle,$$

which implies that $(AB)^* = B^*A^*$. Therefore, the product of two *self-adjoint operators* is self adjoint if and only if the operators commute. In other words, if T is a bounded linear operator of a Hilbert space(i.e., the range is also contained in *H*) then it may happen that T = T*. In this case T is called *self adjoint* or *Hermitian*.

The linear operator T is called *positive* if $\langle Tx, x \rangle \geq 0$ for every $x \in H$. It is called strictly postive if $\langle Tx, x \rangle = 0$ only if x=0. For self-adjoint operators A and B we write A≤B if B-A is a *positive operator*.

4.4.2.4 Linear operators in a separable space

We recall that a Hilbert space H is separable if there exists a countable dense subset in H. Or, a Hilbert space H is separable if and only if it has a countable orthonormal basis.

Bounded operators admit matrix representations completely analogous to the well known matrix representations of operators on finite dimensional spaces. That is, if T is a linear operator of a finite-dimensional Hilbert space, then T can be represented by matrix multiplication by means of an orthonormal basis $\{e_k\}$ of *H*.

Definition- If the operator T is defined everywhere in *H* and if its value for any vector

$$x = \sum_{k=1}^{n} \langle x_k, e_k \rangle e_k = \sum_{k=1}^{n} x_k e_k$$

is given by the formulas

$$Tx = \sum_{k=1}^{n} x_k Te_k$$

and

$$y_i = \langle Tx, e_i \rangle = \sum_{k=1}^{n} t_{ik} x_k .$$

then we say that the operator A admits a matrix representation relative to the orthogonal basis $\{e_k\}_{k=1}^{n}$.

The following are examples which can help us to understand how many problems of finite-dimensional Hilbert space are connected with linear mappings may lead to matrix problems.

- Let **T** be the matrix with entries $t_{ik}=\langle Te_k, e_i \rangle$ and let **x**, **y** be column matrices with entries $x_k=\langle x,e_k \rangle$, $y_k=\langle Tx,e_k \rangle$, where $x\in H$; then **y**=**Tx** (4.5)

- If an Operator T, defined everywhere in a separable space *H*, admits a matrix representation with respect to some orthogonal basis, then it is bounded.

- In order that the matrix (t_{ik}) represent a bounded linear operator defined everywhere in H, it is necessary and sufficient that, for some constant M, the inequality

$$\left|\sum_{i=1}^{p}\sum_{k=1}^{q} t_{ki}x_i\bar{y}_k\right| \le M\sqrt{\sum_{i=1}^{p}|x_i|^2}\sqrt{\sum_{k=1}^{q}|y_k|^2}$$

hold for any numbers $x_1, x_2, \ldots, x_p$ and $y_1, y_2, \ldots, y_q$.

- The mapping T→ **T** from the linear operator of the *n*-dimensional Hilbert space H onto the set(algebra) of n×n matrices has the following properties. If $T_1 \to \mathbf{T}_1$ and $T_2 \to \mathbf{T}_2$ then

(i) $T_1 = T_2$ if and only if $\mathbf{T}_1 = \mathbf{T}_2$;

(ii) $aT_1 + bT_2 \to a\mathbf{T}_1 + b\mathbf{T}_2$ where *a* and *b* are scalars;

(iii) $T_1 T_2 \to \mathbf{T}_1\mathbf{T}_2$;

(iv) $T_1^* \to \mathbf{T}_1^*$.

(v) The inverse operator exists if and only if the inverse matrix $\mathbf{T}^{-1}$ exists, and then

$$T^{-1} \to \mathbf{T}^{-1}.$$

4.4.2.5 Normal Operator

The operator T can be represented in an orthonormal basis by diagonal matrix with only real elements if and only if t is self-adjoint. The operator T can be represented in an orthonormal basis by a diagonal matrix if and only if $T^*T = TT^*$. T is then called a *normal operator*.

The following theorem clarifies how eigenvectors are used as basis of H and T is represented by the previous theorem (4.5) as a diagonal matrix with real elements: In a finite-dimensional space T is a self-adjoint operator if and only if every eigenvalue λ of T is real and there is an orthonormal basis of H formed of eigenvectors of T. If A is a self-adjoint operator, then the point l is a regular point of A if $\Lambda_\Lambda(\lambda) = H$ and λ is a *point of the spectrum* if $\Lambda_\Lambda(\lambda) \neq H$. Furthermore, it is said that the point λ belongs to the *point(discrete) spectrum* of the self-adjoint operator A if $\overline{\Lambda_\Lambda(\lambda)} \neq H$ and λ belongs to the continous spectrum if $\Lambda_\Lambda(\lambda) \neq \overline{\Lambda_\Lambda(\lambda)}$. And we note that the spectrum of a self-adjoint operator is a closed set. λ is called a *regular value* of the operator T if the inverse operator $(\lambda I\text{-}T)^{-1}$ exist, and if it is not regular value then it belongs to the *spectrum* $\sigma(T)$ of T, i.e., all the other points of the complex plane comprise the spectrum of the operator T. If λ belongs to the spectrum of T, then there exists a solution of the equation $\lambda x\text{-}Tx=0$, that is different from 0. If there are solution $x \neq 0$, then λ is called an *eigenvalue* of T and the solutions are the corresponding ***eigenvectors***. If λ is a regular value of T, then $x=(\lambda I\text{-}T)^{-1}f$ is the unique solution of the equation $\lambda x\text{-}Tx = f,\ f\in H$.

If T is a normal operator, i.e. $T^*T = TT^*$, then $Tx = \lambda x$ if and only if $T^*x = \bar{\lambda}x$. If the operator T is not normal, then it may happen that there is only a single eigenvalue of T.

4.4.2.6 Unbounded Linear and Closed Operators .

Let T be a linear map whose domain of definition D_T is a linear manifold in a Hilbert space H_1 and whose range is included in a Hilbert space H_2. We define the graph of T as the set $\Gamma(T)$ of all pairs $\{[x, Tx] \mid x\in D_T\}$ in $H_1 \oplus H_2$ the direct sum of H_1 and H_2. The operator T is called *closed* if its graph $\Gamma(T)$ is a *closed* linear manifold, that is, a subspace, of $H_1 \oplus H_2$.

Equivalently stated T is closed if for any sequence x_n in D_1 for which $x_n \to x$ and $Tx_n \to y$ we have necessarily $x\in D_1$ and $y=Tx$. Every bounded linear operator T from H_1 to H_1 is closed.

We define the resolvent set $\rho(T)$ and the resolvent function $R(\lambda,T)$ for *unbounded operator* just as for bounded ones. Thus $\lambda\in\rho(T)$ if and only if $R(\lambda,T)=(\lambda I-T)^{-1}$ exists as a bounded operator, that is, $R(\lambda,T)$ is bounded and

$$(\lambda I-T)R(\lambda,T)x=x \quad \text{for } x\in H$$

and

$$R(\lambda,T)(\lambda I-T)x=x \quad \text{for } x\in D_T$$

In other words, an operator T(not necessarily linear) is closed if the relations

$$x_n\in D_T, \quad \lim x_n=x, \quad \lim Tx_n=y \ \text{ imply that } \ x\in D_T,\ Tx=y.$$

Thus, the difference between closedness and continuity consists of the following: if the operator T is continuous, then the existence of *lim* x_n $(x_n\in D_T)$ implies the existence of *lim* Tx_n; but if the operator T is only closed, then the convergence of the sequence

$$x_1, x_2, x_3, \ldots (x_n\in D_T) \tag{4.6}$$

does not imply the convergence of the sequence

$$Tx_1, Tx_2, Tx_3, \ldots \tag{4.7}$$

If T is closed, then in particular, it has the property that two sequences of the type (4.7) cannot converge to different limits if the corresponding sequences (4.6) converge to the same limit. An operator t having the property mentioned in the preceding sentence may not be closed; but it has *closed extensions*. Among these is the so-called *minimal closed extension*, which is contained in every closed extension of the operator T. The minimal closed extension is uniquely defined for each operator T. It is denoted by $\bar{T}$ and is called the *closure* of T.

As a concluding remark it can be said that, if the operator T is closed, then each operator $T-\lambda E$ is closed, and if the inverse operator T^{-1} exists then it is closed.

4.4.2.7 Compact Operators

A point set is said to be *compact* if every sequence belonging to it contains a convergent subsequence. Corresponding to the two types of convergence (strong and weak) are strong (or ordinary) compactness and weak compactness. We say that the sequence $x_k \in H$, ($k=1, 2, 3, ...$) converges *weakly* to the vector x and we write $x_k \xrightarrow{w} x$ if $\lim (x_k, h) = (x, h)$, for $h \in H$. The concepts of weakly fundamental sequence and of weak completeness are defined analogously. If the sequence $\{x_k\}_{k=1}^{\infty}$ converges to x, i.e., if $\lim \| x_k - x \| = 0$, then the sequence converges *strongly* to x. Strong convergence implies weak convergence, but not conversely. The usual definition of *compact operators* is as follows. The linear operator T is called compact if the range $\{Tx; x \in B\}$ of any bounded set B is pre-compact. The following results follow the above definition.

- Every bounded point set in H is weakly compact.
- For the weak convergence of the sequence of vectors $\{x_k\}_{k=1}^{\infty}$ it is necessary and sufficient that:

  (i) the numerical sequence (x_k, y) ($k=1,2,3, ...$) converge for each y of some set M which is dense in H; and

  (ii) the sequence $\{x_k\}_{k=1}^{\infty}$ be bounded, i.e., the inequality $\| x_k \| \le C < \infty$ ($k=1,2,3,...$)
- The compact operators of a Hilbert space H form a closed subalgebra of $B(H)$.

  (i) If T_1, T_2 are compact operators then T_1T_2 and $aT_1 + bT_2$ ($a, b \in \Phi$) are also compact operators.

  (ii) If $\{T_n\}$ is a sequence of compact linear operators and $\| T_k - T \| \to 0$ then T is also compact.

4.4.2.8 Conjugation Operators

A *conjugation operator* is an operator I defined on H such that

(i) $\langle If, Ig \rangle = \overline{\langle f, g \rangle}$,

(ii) $I^2 f = f$ for $f, g \in H$.

From (ii) it follows that the range of the operator I is the whole space H. In fact, each vector $h \in H$ can be represented in the form $h = Ig$ merely by taking $g = Ih$. Instead of the usual linearity, the operator I has the following property, which is sometimes called *conjugate linearity*:

$$I(af + bg) = \bar{a} If + \bar{b} Ig.$$

Indeed, letting $g = Ih$ in (i) we get $\langle If, h\rangle = \overline{\langle f, Ih\rangle}$

An example of a conjugation operator in $\mathcal{L}_2$ is the operation of transition to the complex conjugate function:

$$I\varphi(t) = \overline{\varphi(t)}$$

For each conjugation operator in a separable space it is possible to select an orthonormal basis $\{e_k\}_1^\infty$, such that if

$$f = \sum_{k=1}^{\infty} x_k e_k \quad \text{then} \quad If = \sum_{k=1}^{\infty} \overline{x_k} e_k .$$

Definition: A symmetric operator A is said to be real with respect to a given conjugation operator I, if the operators A and I commute, i.e., if $f \in D_A$ implies that $If \in D_A$ and $IAf = AIf$.

4.4.3. Isomorphic Hilbert space and Isomorphic operators

In three-dimensional Euclidean space the simplest operation after that of projection is rotation of the space, which changes neither the lengths of vectors nor the angles between pairs of them. We now consider an analogous operation in Hilbert space.

Definition: The operator U with domain H ($D_U = H$) and range H ($\Delta_U = H$) is *unitary* i

$$\langle Uf, Ug\rangle = \langle f, g\rangle \quad \text{for } f, g \in H.$$

The following are some properties of unitary operator:

- Unitary operator has an *inverse operator*, which is also unitary, i.e., the operator U^{-1} exists, and since $D_{U^{-1}} = \Delta_U$ and $\Delta_{U^{-1}} = D_U$ the operator U^{-1} is defined in the whole space and maps it onto the whole space.
- A unitary operator is necessarily *linear*, i.e., if $f = a_1 f_1 + a_2 f_2$, then $Uf = a_1 Uf_1 + a_2 Uf_2$.
- If a linear operator T satisfies the condition $\langle Uf, Uf\rangle = \langle f, f\rangle$ and if $D_T = \Delta_T = H$, then T is unitary.

If there is a unitary operator from H_1 onto H_2, then the Hilbert spaces H_1 and H_2 are considered to be identical in a certain sense. Due to the existence of an inner product the notion of isomorphism can be specialized. Generally two spaces H_1 and H_2 are isomorphic if there exists an invertible transformation T from H_1 onto H_2.

Definition: The operator V with domain H_1 ($D_V = H_1$) and range H_1 ($\Delta_V = H_2$) is *isometric* if :

$$\langle Vf, Vg \rangle_2 = \langle f, g \rangle_1 \quad \text{for } f, g \in H_1.$$

A unitary operator in H is a special case of an isometric operator for which $H_1 = H_2 = H$. Many properties of unitary operators carry over to arbitrary isometric operators, and some list of these properties follows.

- Each isometric operator has an inverse operator which is also isometric.
- If the operator V is linear, and maps all the space H_1 onto the space H_2 and if $\langle Vf, Vf \rangle_2 = \langle f, f \rangle_1$ for $f \in H$, then V is an isometric operator.
- Every isometric operator is linear.

Definition: The Hilbert spaces H_1 and H_2 are called *isomorphic* or *congruent* if there exists a unitary operator U mapping H_1 onto H_2. In Other words, let T_1 and T_2 be linear operators defined, respectively, in spaces H_1 and H_2, so that $D_{T_1} \subset H_1$, $\Delta_{T_1} \subset H_1$, $D_{T_2} \subset H_2$, $\Delta_{T_2} \subset H_2$ (In particular, the spaces H_1, H_2 may coincide). The operators T_1 and T_2 are called *isomorphic* or *unitarily equivalent* if there exists an isometric operator V, which maps H_1 onto H_2 and D_{T_1} onto D_{T_2}, such that $VT_1f = T_2Vf$ for each $f \in D_{T_1}$. That is, T_1 and T_2 are *unitarily equivalent* if

$$D_{T_2} = VD_{T_1} \quad \text{and} \quad T_1 = V^{-1}T_2V.$$

An operator $U: H_1 \to H_2$ is called an *isometry* if it satisfies $U^*U = I$, and *coisometry* if $UU^* = I$ is satisfied.

We note that the following conditions for an operator T mapping a Hilbert space H_1 onto another Hilbert space H_2 are equivalent:

- T is isomeric, i.e. $\|Tf\| = \|f\|$ for every $f \in H_1$
- $T^*T = E_1$ (identity operator in H_1)
- $\langle Tf, Tg \rangle = \langle f, g \rangle$ for $f, g \in H_1$

4.4.4 Important Theorems

The following theorems are prominent in the study of Hilbert space. We will state and their respective proves can be conferred on Akhiezer and Glazman(1961), Kingman and Taylor(1966), Fuhrmann(1981) and Máté(1989).

4.4.4.1 Riesz Representation Theorem for Hilbert Spaces

The following theorem of Riesz provides a representation for each linear functional in H.

Theorem - Each linear functional Φ in the Hilbert space H can be expressed in the form Φ(h)=(h,x), where x is an element of H which is uniquely determined by the functional Φ; furthermore, $\|\Phi\|=\|x\|$.

4.4.4.2 Hahn-Banach Theorem for Hilbert Spaces (Extension theorem)

Theorem - If f is a normed linear functional on a linear subspace A of a normed linear space, then f can be extended to a normed linear functional on the whole space without changing its norm.

The Hahn-Banach extension theorem can be stated as follows: Suppose K is a linear subspace of a linear space H. Then any bounded linear functional on K can be extended to a bounded linear functional on H with the same norm. Or, suppose K is a linear subspace of a linear space H, p is a subadditive functional on H such that a $p(ax)=ap(x)$ for $a \geq 0$, $x \in H$; and f is a linear functional on K such that $f(x) \leq p(x)$ for all $x \in K$. Then there is a linear functional $\bar{f}: H - R$ such that $\bar{f}(x)=f(x)$ for $x \in K$, $\bar{f}(x) \leq p(x)$ for $x \in H$.

4.4.4.3 Riesz-Fisher Theorem

This theorem is formulated and proved in Hilbert space. Since $\mathcal{L}_2$-spaces are realisations of Hilbert pace, we will deduce the classical theorem about the Fourier expansion as a trigonometric series of a function in $\mathcal{L}_2$ as a special case.

Given an orthonormal family (e_j), $j \in J$ on a Hilbert space H, and any point $x \in H$, the real numbers $c_j=(x,e_j)$ $(j \in J)$, are called the *Fourier coefficients* of x on the orthonormal family and the series $\sum_{j \in J} c_j e_j$ is called the *Fourier series* of x.

Theorem - Let $\{e_j\}$ ($j \in J$) be any orthonormal system (not necessarily complete) in a Hilbert space H, and let $\{\beta_j\}$ $j \in J$ be any set of real numbers such that $\sum_{j \in J} \beta_j^2$ converges. Then there is a point $x \in H$ with Fourier coefficients $\beta_j = (x, e_j)$ such that the finite partial sums $s_I = \sum_{i \in I} \beta_i e_i$ converge to x in norm.

4.4.5. The 23 Problems of Hilbert

A substantial part of Hilbert's fame rests on a list of *23 research problems* he enunciated at the international congress of mathematicians held in Paris in 1900. In his address, "The Problems of Mathematics," he surveyed nearly all the mathematics of his day and endeavoured to set forth the problems he thought would be significant for mathematicians in the 20th century.

The following table is taken from Encyclopedic Dictionary of Mathematics, Vol. II edited by Itô (1987,736-7).

1	To rove the continuum hypothesis.
2	To investigate the consistency of the axioms of arithmetic.
3	To show that it is impossible to prove the following fact utilizing only congruence axioms: Two tetrahedral having the same altitude and base area have the same volume. Solved by M. Dehn(1900).
4	To investigate geometries in which the line segment path between any pair of points gives the shortest path between the pair.
5	To obtain the conditions under which a topological group has the structure of Lie group. Solved by A. M. Gleason and D. Montgomery and L. Zippin (1952), and H. Yamabe (1953).
6	To axiomatize those physical sciences in which mathematics plays an important role.
7	To establish the transcendence of certain numbers. The transcendence of $2^{\sqrt{2}}$, which was one of numbers put for by Hilbert, was shown by A. Fel'fond (1934) and T. Schneider (1935).
8	To investigate problems concerning the distribution of prime numbers; in particular, to show the correctness of Riemann hypothesis. [Unsolved]
9	To establish a general law of reciprocity. Solved by T. Takagi (1921) and E. Artin (1927).

10. To establish effective methods to determine the solvability of Diophantine equations. Solved affirmatively for equations of two unknowns by A. Baker, *Philos. Trans. Roy. Soc. London*, (A) 263 (1968); solved negatively for the general case by Yu. V. Matiyasevich (1970).
11. To investigate the theory of quadratic forms over an arbitrary algebraic number field of finite degree.
12. To construct class fields of algebraic number fields.
13. To show the impossibility of the solution of the general algebraic equation of the seventh degree by compositions of continuous functions of two variables. Solved negatively. In general, V. I. Arnold proved that every real, continuous function $f(x_1,x_2,x_3)$ on [0,1] can be represented in the form $\sum_{i=1}^{9} h_i(g_i(x_1,x_2),x_3)$, where h_i and g_i are real, continuous functions, and A. N. Kolmogorov proved that $f(x_1,x_2,x_3)$ can be represented in the form $\sum_{i=1}^{7} h_i(g_{i1}(x_1), g_{i2}(x_2), g_{i3}(x_3))$, where h_i and g_{ij} are real, continuous functions and g_{ij} can be chosen once for all independently of f [*Dokl. Nauk SSSR*, 114(1957), *Amer. Math. Soc. Transl.*, 28(1963)].
14. Let k be a field, $x_1, \ldots, x_n$ be variables, and $f_i(x_1, \ldots, x_n)$ given polynomials in $k[x_1,\ldots,x_n]$ ($i=1, \ldots, m$). Furthermore, let **R** be the ring formed by rational functions $F(X_1, \ldots, X_m)$ in $k(X_1, \ldots, X_m)$ such that $F(f_1, \ldots, f_m) \in k[x_1, \ldots, x_n]$. The problem is to determine whether the ring **R** has a finite set of generators. Solved negatively by M. Nagata, *Amer. J. Math.*, 81(1959).
15. To establish the foundations of algebraic geometry. Solved by B. L. van der Waeden (1938-1940), A. Weil (1950), and others.
16. To conduct topological studies of algebraic curves and surfaces.
17. Let $f(x_1, \ldots, x_n)$ be a rational function with real coefficients that takes a positive value for any real n-tuple $(x_1, \ldots, x_n)$. The problem is to determine whether the function f can be written as the sum of squares of rational functions. Solved in the affirmative by E. Artin (1927).
18. To express Euclidean n-space as a disjoint union $\bigcup_{\lambda} P_{\lambda}$, where each P_{λ} is congruent to one of a set of given polyhedra.
19. To determine whether the solutions of regular problems in the calculus of variations are necessarily analytic. Solved by S. N. Bernshtein, I. G. Petrovskii and others.
20. To investigate the general boundary value problem.
21. To show that there always exists a linear differential equations of the Fuchsian class with given singular points and monodromic group. Solved by H. Rohrl and others (1957).
22. To uniformize complex analytic functions by means of automorphic functions. Solved for the case of one variable by P. Koebe (1907).
23. To develop the methodology of the calculus of variations.

4.5. A Characteristic Property of Normal Distribution in Hilbert Space

As a preliminary remark it can be said that this topic is advanced by statisticians who combine pure mathematical theories and probabilistic and statistical principles. It combines theory of operator theory and properties of characteristic functions; and comes up with theorems of a characterisation of the normal law in Hilbert space.

The Characterisation of Normal Distribution in Hilbert Space was initiated by Prohorov and Fisz (1957). In their article they came up with a theorem of random elements in Hilbert space, and the theorem.

Then, Eaton and Pathak (1969) picked up the topic and studied it more comprehensively and came up with the theorem of probability measure in Hilbert Space Furthermore, Pathak(1970) made another study on this topic and gave additional results.

4.5.1 Prohorov & Fisz Theorem (1957) of Random Element in Hilbert Space

Theorem 4.1

Consider random element with values in a real separable Hilbert space H [that is measurable mapping $\xi(u)$ of a fundamental probability field into the space H]. Let the probability distribution and characteristic function be denoted respectively as,

$$P^{\xi} \text{ and } \quad \phi(f,\xi) = \int_H e^{i(f,\xi)} dP^{\xi}, \quad f \in H.$$

Let $P^{\xi} = P^{\eta}$ be denoted as $\xi \sim \eta$ and let $||\xi||$ be the linear functional (f,ξ), $f \in H$, stochastic variables. The mathematical expectation of the random element ξ is such an element $M\xi \in H$ that for every $f \in H$, $M(f,\xi) = (f, M\xi)$. Consider the conditions, (α): M "ξ"^{2} $< \infty$; (β): $M(f,\xi)^2 > 0$ for any $f \in H$, $f \neq 0$; (γ): $M\xi = 0$ where 0 is the null element in H, and let $\xi^{(1)} \sim \xi^{(2)} \sim \xi^{(3)}$, be random elements in H, subject to (α), (β) and (γ) and let (δ): $\xi^{(1)}$ and $\xi^{(2)}$ be independent and (ε): $\xi^{(1)} + \xi^{(2)} \sim \Lambda\, \xi^{(3)}$, where Λ is a (A1) linear, (A2) bounded, (A3) self conjugate, (A4) positive operator in H. Then the distribution P of each of the random element $\xi^{(i)}$ is normal and $A = \sqrt{2}E$ where E is a unitary operator.

4.5.2 Eaton-Pathak Theorems(1969) for Probability Measures on Hilbert Space

After Prohorov & Fisz(1957), we observe that Eaton and Pathak (1969) picked up the topic and studied it more comprehensively.

Their article is divided into four sections, namely, (i)introduction, (ii)preliminaries, (iii)a characterisation of the normal law, and (iv)semi-stable laws in *H*.

In the *introductory part* the importance of the Rao-Ramachandran(1968) or the Pathak-Pillai(1968) theorem on the characterisation of a distribution. The main *problem* is presented as follows: Let $X_0, X_1, \ldots, X_k$ be independently and identically distributed real-valued random variables, and let $Y_1 = X_0 - \sum_{i=1}^{k} c_i X_i$ and $Y_2 = X_0 - \sum_{i=1}^{k} b_i X_i$. If we assume further that $E(Y_1/Y_2) = 0$, then, can we characterise the distribution X_0? This is verified by Rao(1967) with the conditions that

(i) X_0 has finite variance,

(ii) $|b_i| < 1$, $i = 1, \ldots, k$,

(iii) $c_i/b_i > 0$ and

(iv) $\sum_{i=1}^{k}(c_i / b_i)b_i^2 = 1$, then the distribution of X_0 is normal.

Thus, it is shown that $E(Y_1/Y_2) = 0$, is equivalent to

$$\varphi(t) = \prod_{i=1}^{k}\left[\varphi(b_i t)\right]^{a_i} \text{ (functional equation)} \tag{4.8}$$

where $a_i = c_i/b_i$ and $\varphi(t)$ is the characteristic function of X_0.

Furthermore, the equation (4.8) is considered in a real separable Hilbert space $(H(.,.))$. If A and B are two Hermitian or self-adjoint linear operators on *H* to *H*, (for more clarification see section 4.4.2.3), it is written as A≥B to mean A-B is positive semi-definite. The extension of equation (4.8) considered is

$$\hat{\mu}(y) = \prod_{i=1}^{k}\left[\hat{\mu}(B_i y)\right]^{a_i} \tag{4.9}$$

where $\hat{\mu}(|\hat{\mu}| \neq 0)$ is the characteristic function of a probability measure μ on *H*, $a_i > 0$, $y \in H$, B_i is a bounded operator on *H* with a bounded inverse, and suppose there exists a constant λ_0, $0 < \lambda_0 < 1$, such that $\| B_i \| < \lambda_0$, i=1, ..., k. With aid of these assumptions they deduced the results in section (iii), the characterisation of the normal law, of their article.

The *preliminaries* deal with some results concerning probability measures on a real separable Hilbert space. A detailed discussion on some of these can be found in Parthasarathy's (1967) book, *Probability measures on Metric Spaces.*

The *third section* is the heart of the research and after considering the function equation, (4.9), and other assumptions the following results are shown:

(a) $\hat{\mu}$ is infinitely divisible,

(b) If $\sum_{i=1}^{k} a_i B_i B_i^{*} \geq I$ then $\hat{\mu}$ corresponds to the normal distribution (possible degenerate) on

H, and (c) if $\sum_{i=1}^{k} a_i \leq 1$, then $\hat{\mu}$ corresponds to the distribution degenerate at $0 \in H$.

Theorem 4.2

If $\mu \in M(H)$ and $\hat{\mu}$ satisfies

$$\hat{\mu}(y) = \left[\hat{\mu}(B_i y)\right]^{n}$$

then μ is infinitely divisible, where $M(H)$ is the space of all probability measures on *H* and $M_0(H)$ be the space of all infinitely divisible(i.d.) measures in $M(H)$, for each μ in $M(H)$, $\hat{\mu}$ denote the characteristic function of μ and $(\hat{x})$ the characteristic function of the distribution degenerate at $x \in H$.

Proof:

Iteration of (4.9) yields

$$\mu(y) = \prod_{i=1}^{k^n} \left[\hat{\mu}(D_{n,i} y)\right]^{c_{n,i}} \tag{4.10}$$

where $c_{n,i}$ is of the form $a_{i_1} a_{i_2} \ldots a_{i_n}$ and $D_{n,i}$ is of the form $B_{i_1} B_{i_2} \ldots B_{i_n}$.

Let

$$\hat{\mu}_{n,i}(y) = \hat{\mu}(D_{n,i} y).$$

Since $\| D_{n,i} \| < \lambda_0^n$, it follows that $\{\mu_{n,j}\}$ j=1, ..., k^n, n=1,2, ... is a sequence of infinitesimal probability measures. The infinite divisibility of μ now follows from Corollary, that is, let

$\hat{\mu}$ be a non-vanishing characteristic function on H, then there exists a unique continuous function λ on H with $\lambda(0) = 0$ such that $\hat{\mu}(y) = e^{\lambda(y)}$. This completes the proof.

Theorem 4.3

If $\hat{\mu}$ satisfies (4.9) and if

$$\sum_{i=1}^{k} a_i B_i B_i' > I,$$

then $\hat{\mu}$ corresponds to a normal law (possible degenerate).

Proof:

Since $\hat{\mu}$ is i. d. and satisfies (4.9) we have

$$\log \hat{\mu}(y) = i(x_0, y) - \tfrac{1}{2}(Sy, y) + \int K(x,y)\, d M(x)$$

$$= i\sum_{j=1}^{k} a_j (x_0, B_j y) - \tfrac{1}{2} \sum_{j=1}^{k} a_j (B_j S B_j y, y) + \sum_{j=1}^{k} a_j \int K(x, B_j y) dM(x)$$

$$= i\left(\sum a_j B_j' x_0, y\right) - \tfrac{1}{2}\left(\sum a_j B_j' S B_j y, y\right) + \sum a_j \int K(B_j' x, y) dM(x)$$

$$+ \sum a_j \int \left[K(x, B_j y) - K(B_j' x, y)\right] dM(x) \qquad (4.11)$$

However,

$$\int \left[K(x, B_i y) - K(B_i' x, y)\right] dM(x) = i\int \left\{ \frac{(y, B_j' x)}{1 + \|x\|^2} - \frac{(y, B_i' x)}{1 + \|B_i' x\|^2} \right\} dM(x) \qquad (4.12)$$

which exists from the conditions on M.

Let $\gamma_i \in H$ be such that

$$\int \left[K(x, B_i y) - K(B_i x, y)\right] dM(x) = i(\gamma_i, y) \quad \text{for all } y \in H \qquad (4.13)$$

Let

$$u_0 = \sum a_j B_i' x_0 + \sum a_j \gamma_j.$$

Then (4.11) can be written as

$$i(x_0, y) - \tfrac{1}{2}(Sy, y) + \int K(x, y) dM(x) = i(x_0, y) - \tfrac{1}{2}\left(\sum a_j B_j' S B_j y, y\right) + \sum a_i \int K(x, y) dM((B_i')^{-1} x) \quad (4.14)$$

Since $\sum a_i B_i' S B_i$ is Hermitian or self-adjoint with finite trace, it follows from the uniqueness of the representation for

$$\log \hat{\mu}(y) \text{ that } dM(x) = \sum a_i\, dM((B_i')^{-1} x) \qquad (4.15)$$

To complete the proof, (4.15) together with the assumption that $\sum a_i B_i' B_i \geq I$ should imply that M=0.

Let

$$S(r)=\{x \mid \|x\| \leq r\}, \qquad \text{for } r>0.$$

$$\int_{S(r)} \|x\|^2 dM(x) = \sum a_i \int_{S(r)} \|x\|^2 dM((B_i')^{-1}x)$$

$$= \sum a_j \int_{(B_j')^{-1}S(r)} \|B_j' x\|^2 dM(x) = a_1 \int_{(B_1')^{-1}S(r)} \|B_1' x\|^2 dM(x) + \sum_{i=2}^{k} a_i \int_{(B_i')^{-1}S(r)} \|B_i' x\|^2 dM(x). \qquad (4.16)$$

Now, since

$$\|B_i\| \leq \lambda_0 < 1,$$

we have

$$(B_i')^{-1}S(r) \supset S(r).$$

Setting

$$C(r)=(B_1')^{-1}S(r) \cap S(r)^c,$$

we then have

$$\int_{S(r)} \|x\|^2 dM(x) \geq a_1 \int_{C(r)} \|B_1' x\|^2 dM(x) + \sum a_i \int_{S(r)} \|B_i' x\|^2 dM(x)$$

$$\geq a_1 \int_{S(r)} \|B_1' x\|^2 dM(x) + \int_{S(r)} \|x\|^2 dM(x) \qquad (4.17)$$

The last inequality follows by noting that

$$\sum a_i \|B_i' x\|^2 = \sum a_i (B_i B_i' x, x) \geq (x, x)$$

since

$$\sum a_i B_i' B_i \geq I.$$

Since $a_1 > 0$, we conclude that

$$\int_{C(r)} \|B_1' x\|^2 dM(x) = 0. \qquad (4.18)$$

Now,

$$C(r) = \{x \mid \|B_1' x\| \leq r\} \cap \{x \mid \|x\| > r\}.$$

Since

$$\|B_1' x\| \leq \lambda_0 \|x\| < \|x\|,$$

it is easy to see that if $u \in H$, $u \neq 0$, then there exists rational number r_0 such that $u \in C(r_0)$. Thus, setting $D = \bigcup C(r)$ where the union is over positive rational r, we see that $D = H - \{0\}$. However (4.18) implies that

$$\int_D \|B_r^* x\|^2 \, dM(x) = 0.$$

Since $\| B_r^* x \|^2 > 0$ for all $x \in D$, $M(D) = 0$ so that $M(H) = 0$ since $M\{0\}) = 0$. This completes the proof.

Semi-stable laws in Hilbert space

In semi-stable laws in H the following functional equation:

$$\hat{\mu}(y) = [\hat{\mu}(By)]^a \tag{4.19}$$

is considered, where a is a positive real number, and B is a bounded operator such that B^{-1} exists, is bounded, and $\| B \| \leq \lambda_0$ for some positive $\lambda_0 < 1$. Characteristic functions which satisfy (4.19) afford one possible generalisation of Levy's(1937) semi-stable laws on the real line. These laws are dealt with at length by P. Lévy (1937), in his book entitled *Théore de l'Addition des Variables Aléatores*. Hence one can obtain a representation for characteristic functions on H which satisfy (4.19).

Theorem 4.4

Let $a > 1$, then we have the following theorem:

let $\mu \in M(H)$. Then the following are equivalent:

(a) $\hat{\mu}$ satisfies (4.19)

(b) $\log \hat{\mu}(y) = i(\gamma, y) - \frac{1}{2}(Sy, y) + \sum_{r=0}^{\infty} a^r \Psi(B^r y) + \sum_{r=-1}^{-\infty} a^r [\Psi(B^r y) + i(x_0, B^r y)]$

where,

(i) $\Psi(y) = \int_{A_0} [e^{i(x,y)} - 1 - i(x,y)] \, dM_0(x)$

$A_0 = \{x | \|x\| \leq 1\} \cap \{x | \|(B^*)^{-1} x\| > 1\}$

and M_0 is a finite measure defined on Borel sets of A_0 such that

$$\lim_{n \to \infty} \int_{A_0} \left(\sum_{r=1}^{n} a^r (bBB^*)^r x, x \right) dM_0(x) < +\infty$$

ii) $x_0 \in H$ is defined by $(x_0, y) = \int_{\Lambda_0} (x,y)\,dM_0(x)$ for all $y \in H$,

iii) the vector $\gamma \in H$ is such that $(I - aB')\gamma = x_0$,

iv) S is a non-negative Hermitian operator with finite trace such that $S = aB'SB$.

Proof:

Suppose that $\hat{\mu}$ satisfies (4.19). Then $\hat{\mu}$ is i. d. and we have

$$\log \hat{\mu}(y) = i(\gamma_1, y) - \tfrac{1}{2}(Sy, y) + \int K(x,y)\,dM(x) \tag{4.20}$$

From the uniqueness of the above representation and equation (4.19), it follows, as in theorem 4.3, that

$$S = aB'SB \tag{4.21}$$

or

$$dM(x) = a\,dM((B')^{-1}x).$$

Now, let

$$\Lambda_r = \{x \mid \|(B')^r x\| \le 1\} \cap \{x \mid \|(B')^{r-1}x\| > 1\} \text{ for } r = 0, \pm 1, \pm 2, \dots \tag{4.22}$$

It is clear that the sets Λ_r are disjoint and

$$\bigcup_{-\infty}^{0} \Lambda_r = \{x \mid \|x\| \le 1\} - \{0\}$$

and

$$\bigcup_{1}^{\infty} \Lambda_r = \{x \mid \|x\| > 1\}.$$

Consequently,

$$\int K(x,y)\,dM(x) = \int_{\{\|x\| > 1\}} K(x,y)\,dM(x) + \int_{\{\|x\| < 1\}} K(x,y)\,dM(x)$$

$$\begin{aligned} \int K(x,y)\,dM(x) = {} & \int_{\{\|x\| < 1\}} [e^{i(x,y)} - 1 - i(x,y)]\,dM(x) + i \int_{\{\|x\| \le 1\}} \frac{(x,y)\|x\|^2\,dM(x)}{1 + \|x\|^2} \\ & + \int_{\{\|x\| > 1\}} [e^{i(x,y)} - 1]\,dM(x) + i \int_{\{\|x\| > 1\}} \frac{(x,y)}{1 + \|x\|^2}\,dM(x) \end{aligned} \tag{4.23}$$

The existence of the second and fourth integrals is easy to establish. Define $\gamma_2 \in H$ by

$$(\gamma_2, y) = \int_{\{\|x\| \le 1\}} \frac{(x,y)\|x\|^2}{1 + \|x\|^2}\,dM(x) + \int_{\{\|x\| > 1\}} \frac{(x,y)}{1 + \|x\|^2}\,dM(x) \tag{4.24}$$

Now, we write

$$\int_{\{\|x\|<1\}}[e^{i(x,y)}-1-i(x,y)]dM(x) = \sum_{r=0}^{\infty}\int_{\Lambda_r}[e^{i(x,y)}-1-i(x,y)]dM(x)$$

$$= \sum_{r=0}^{\infty}\int_{\Lambda_0}[e^{i((B')^{-r}x,y)}-1-i((B')^{-r}x,y)]dM((B')^{-r}x) \tag{4.25}$$

$$= \sum_{r=0}^{\infty}a^{-r}\int_{\Lambda_0}[e^{i(x,B^{-r}y)}-1-i(x,B^{-r}y)]dM(x)$$

$$= \sum_{r=0}^{\infty}a^{r}\Psi(B^{r}y)$$

where Ψ is defined in (i) with M_0, the restriction of M to Λ_0. A similar analysis yields

$$\int_{\{\|x\|>1\}}[e^{i(x,y)}-1]dM(x) = \sum_{r=-1}^{-\infty}\int_{\Lambda_r}[e^{i(x,y)}-1-)]dM(x)$$

$$= \sum_{r=-1}^{-\infty}\int_{\Lambda_0}[e^{i((B')^{-r}x,y)}-1-i((B')^{-r}x,y)+i((B')^{-r}x,y)]dM((B')^{-r}x) \tag{4.26}$$

$$= \sum_{r=-1}^{-\infty}a^{r}[\Psi(B^{r}y)+i(x_0,B^{r}y)]$$

where x_0 is defined in (ii). Defining γ to be $\gamma_1+\gamma_2$, we obtain the representation for $\log\hat{\mu}(y)$ as given in (b). The assertions in (iii) and (iv) follow immediately from the uniqueness of the representation, i.e.,

$$\hat{\mu}(y) = \exp[\,i(x_0,y) - ½(Sy,y) + \int K(x,y)\, dM(x)]. \tag{4.13}$$

Further, it can be verified that

$$\lim_{n\to\infty}\int_{\Lambda_n}\left(\sum_{i=1}^{n}a^{i}(BB')^{i}x,\ x\right)dM_n(x) = \int_{\{\|x\|<1\}}\|x\|^2\ dM(x) < +\infty \tag{4.27}$$

by the representation theorem for i.d. laws.

To prove the converse, we first obtain a σ-finite measure M on H from M_0 as follows. For any Borel set $A \subset H$ such $0\notin A$, define M(A) by

$$M(A) = \sum_{r=-\infty}^{\infty}a^{r}M_0((B')^{r}(A\cap\Lambda_r)) \tag{4.28}$$

where $(B')^{r}(C) = \{x\ (B')^{r}x \in C\}$ for any set C. It is not difficult to verify that M is a σ-finite measure on H-{0} and M satisfy (4.11). Further, since $a > 1$,

$$M\{x \mid \|x\| > 1\} < +\infty$$

and

$$\int_{\{\|x\|<1\}}\|x\|^2\ dM(x) = \lim_{n\to\infty}\int_{\Lambda_n}\left(\sum_{j=1}^{n}a^{i}(BB')^{i}v,\ v\right)dM(x) < +\infty \tag{4.29}$$

by (i).

By proceeding backwards through the argument for (a) implies (b), we conclude that

$$\log \hat{\mu}(y) = i(\gamma_1, y) - \tfrac{1}{2}(Sy, y) + \int K(x,y)dM(x) \tag{4.30}$$

where $\gamma_1 = \gamma_1 - \gamma_2$, γ and S are given in (iii) and (iv) respectively, and γ_2 is defined in (4.24). Hence $\hat{\mu}$ corresponds to an i. d. law. That $\hat{\mu}$ satisfies (4.19) is clear form its representation in (b). This completes the proof.

Remarks:

When $aB' = I$, then x_0 defined in (ii) is necessarily 0 and γ can be arbitrary. If $\|B^{-1}\| < a$, then the series

$$\sum_{r=1}^{\infty} a^r (x_0, B^r y) \tag{4.31}$$

converges absolutely and in this case $\hat{\mu}$ admits the representation

$$\log \hat{\mu}(y) = -\tfrac{1}{2}(Sy, y) + \sum_{r=-\infty}^{\infty} a^r \Psi(B^r y) \tag{4.32}$$

when $\hat{\mu}$ satisfies (4.19).

In the particular case when H is the real line, it follows from theorem 4.4 that if $ab^2 > 1$, then $\hat{\mu}$ corresponds to the normal distribution (degenerate when $ab^2 > 1$). However, when $ab^2 < 1$, $\hat{\mu}$ has the representation

$$\log \hat{\mu}(y) = i\gamma y + \sum_{r=0}^{\infty} a^r \Psi(B^r y) + \sum_{r=-1}^{-\infty} a^r [\Psi(b^r y) + ib^r x_0 y] \tag{4.33}$$

since the normal component must vanish. Also, in this case, the condition

$$\lim_{n \to \infty} \int_{A_n} \left(\sum_{j=1}^{n} (ab^2)^j x^2 \, dM_n(x) \right) < +\infty \tag{4.34}$$

is automatically satisfied since $ab^2 < 1$.

Moreover, if $a|b| > 1$, then the corresponding semi-stable law has the form (4.32) with S=0, and the distribution possesses a finite first moment.

4.5.3 Pathak-Theorems(1970) for Probability Measures on Hilbert Space

Furthermore, Pathak(1970), after the establishment of the theorem for probability measures on Hilbert space, made another study on this topic and gave additional results.

In his note P. K. Pathak attempts to generalise the theorem of Rao-Ramachandran(1968) to the case when the a_i's in 4.9,

$$\hat{\mu}(y) = \prod_{i=1}^{k}\left[\hat{\mu}(B_i y)\right]^{a_i},$$

are not all positive. Under certain restriction on the B_i's in (4.9), he showed that the problem of characterisation normality on Hilbert space through (4.9) reduces to an essentially univariate problem of characterising normality on the real line.

To establish the theorem preliminary expositions are required.

Let $(H,(.,.))$ be a real separable Hilbert space and let x, y etc. stand for generic elements of H. Let $M(H)$ be the space of probability measures on H. For each $\mu \in M(H)$, let $\hat{\mu}$ denote the characteristic function of μ. A probability measure $\mu \in M(H)$ is called normal if its characteristic function is given as follows:

$$\hat{\mu}(y) = \exp\{i(x_o,y) - \tfrac{1}{2}(Sy,y)\} \tag{4.35}$$

where $x_o \in H$ and S is a non-negative Hermitian operator with finite trace.

Subsequently, the following theorem concerning the spectral representation of Hermitian operators with discrete spectrum is needed:

Let $B_1, \ldots, B_k$ be bounded Hermitian operators with discrete spectrum and suppose that

$$B_i B_j = B_j B_i \quad \text{for all } i, j = 1, \ldots, k.$$

Then there exists a complete orthonormal sequence $\{e_n\}$ of vectors such that, for each i,

$$B_i = \sum_{n=1}^{\infty} \lambda_{in} P_n \tag{4.36}$$

where P_n denotes the projection operator on the subspace spanned by the vector e_n.

Next let us establish necessary and sufficient conditions under which $\hat{\mu}$, given under (4.9), corresponds to a normal law.

Let us select a system of co-ordinates in H, that is, a complete orthonormal sequence $\{e_n\}$ of vectors in H such that, for each i,

$$B_i = \sum_{n=1}^{k} \lambda_{in} P_n$$

as is given by (4.36). Let y_n be the n-th co-ordinate of y, i.e., $y_n = (y,e_n)$.

e H is, then, isomorphic with the sequence ℓ_2 of all real sequences $\ldots y_n^2 < \infty$. Every $\mu \in M(H)$ now induces a measure μ in ℓ_2 through ism corresponding to the basis $\{e_n\}$.

$$\phi_n(y_1, \ldots, y_n) = \hat{\mu}(y_1 e_1 + \ldots + y_n e_n).$$

ow that $\phi_n(y_1, \ldots, y_n)$ is a characteristic function of the finite on of μ induced by the projection:

$$x = (x_1, x_2, \ldots) \to (x_1, \ldots, x_n).$$

that $\phi_n(y_1, \ldots, y_n)$ satisfies the following equation.

$$\phi_n(y_1, \ldots, y_n) \quad \prod_{i=1}^{k} \left[\phi(\lambda_{i1} y_1, \ldots, \lambda_{in} y_n)\right]^{\alpha_i} \tag{4.37}$$

tain necessary and sufficient conditions under which $\phi_n(y_1, \ldots, y_n)$, esponds to a normal law. Normality of μ can then be established eorem.

e orthonormal sequence in H. For $\mu \in M(H)$ and $n = 1, 2, \ldots$, let

$$\phi_n(y_1, \ldots, y_n) = \hat{\mu}(y_1 e_1 + \ldots + y_n e_n)$$

ich n, $\phi_n(y_1, \ldots, y_n)$ corresponds to a finite dimensional normal is normal.

subspace spanned by the vectors $e_1, \ldots, e_n$. Definite a probability follows:

$$\mu_n(A) = \mu_{(n)}(A \cap H_n). \tag{4.38}$$

e projection of μ on H_n

$$\hat{\mu}_n(y) \quad \int \exp[i(x,y)]d\mu_{(n)} \quad \int_{H_n} \exp[i(x_1 y_1 + \ldots + x_n y_n)]d\mu_{(n)} \tag{5.40}$$

$$\hat{\mu}(x_1 y_1 + \ldots + x_n y_n) = \phi_n(y_1 + \ldots + y_n).$$

Further since $\phi_n(y_1, \ldots, y_n)$ corresponds to a normal law, we have

$$\hat{\mu}_n(y) = \exp [i(\gamma_{n1}y_1 + \ldots + \gamma_{nn}y_n - 1/2\,(y_1 + \ldots + y_n)\, S_n(y_1, \ldots, y_n)']$$

$$= \exp [i(\gamma_n, y) - 1/2\,(S_n y, y)] \quad \text{(say)} \tag{4.41}$$

where S_n is a positive definite operator with finite trace. Consequently μ_n is normal for each n. It is easy to see from (4.40) that

$$\lim_{n\to\infty} \hat{\mu}_n(y) = \hat{\mu}(y) \tag{4.41}$$

An argument similar to that theorem of Parthasarthy(1967) shows that the sequence of measures$\{\mu_n\}$ is weakly conditionally compact. Consequently μ_n converges weakly to μ. Since μ_n is normal for each n, μ must be normal. This completes the proof

Theorem 4.6:

Let be μ_n a probability measure on an n-dimensional Euclidean space. Suppose that μ_n has moments of all order and $\hat{\mu}_n$, the characteristic function of μ_n, satisfies the following functional equation

$$\hat{\mu}_n(y_1, \ldots, y_n) \quad \prod_{i=1}^{k} \left[\hat{\mu}_n(\lambda_{i1}y_1, \ldots, \lambda_{in}y_n)\right]^{a_i} \tag{4.42}$$

where $|\lambda_{ij}| < 1$. Then μ_n is normal (possibly degenerate).

Proof:

It can be seen that $\hat{\mu}_n$ has no zeros so that (4.42) can be written as

$$\log \hat{\mu}_n(y_1, \ldots, y_n) = \Sigma\, a_i \log \hat{\mu}_n(\lambda_{i1}y_1, \ldots, \lambda_{in}y_n). \tag{4.42}$$

Since μ_n has moments of all order, $\hat{\mu}_n$ possesses derivatives of all order. Consequently

$$\left(1 - \sum_{i=1}^{k} a_i \lambda_{ij}^{r_1} \cdots \lambda_{ij}^{r_1}\right) \frac{\partial^m}{\partial y_1^{r_1} \cdots \partial y_n^{r_n}} \log \hat{\mu}_n(y_1, \ldots, y_n) \Bigg|_{\substack{y_1 = 0 \\ y_n = 0}} 0 \tag{4.44}$$

where $r_1, \ldots, r_n$ are non-negative integers such that $r_1 + \ldots + r_n = m$. Since $|\lambda_{ij}| < 1$,

$$\sum_{i=1}^{k} a_i \lambda_{ij}^{r_1} \ldots \lambda_{ij}^{r_n} \neq 1$$

for sufficiently large m. It now follows from (4.44) that, for m large,

$$\frac{\partial^m}{\partial y_1^{r_1}\dots\partial y_n^{r_n}} \log \hat{\mu}_n(y_1, \dots, y_n)\Bigg|_{\substack{y_1=0\\ y_n=0}} 0 \qquad (4.45)$$

Therefore $\log \hat{\mu}_n(y_1, \dots, y_n)$ is a polynomial of some degree in $y_1, \dots, y_n$. Hence μ_n must be normal. This completes the proof.

The following theorem is the main note in the Pathak's article(1970).

Theorem 4.7

Let $\mu \in M(H)$ and suppose that $\hat{\mu}$ satisfies (4.9). Let $\{e_n\}$ be a complete orthonormal system in H such that, for each i,

$$B_i = \sum_{n=1}^{\infty} \lambda_{in} P_n ,$$

where P_n denotes the projection operator on the subspace spanned by the vector e_n. Then the following two assertions are equivalent.

(a) The probability measure μ is normal.

(b) For each n, the functional equation

$$\psi_n(y_n) = \prod_{i=1}^{k} \left[\psi_n(\lambda_{in} y_n)\right]^{a_i}, \qquad (4.46)$$

where $\psi_n(y_n) = \hat{\mu}(y_n e_n)$, implies that $\psi_n(y_n)$ corresponds to a normal law on the real line.

Proof:

It suffices to establish that (a)$\Rightarrow$ (b). Suppose that (b) holds. Let μ_n denote the probability measure on n-dimensional Euclidean space that corresponds to the characteristic function

$$\phi_n(y_1, \dots, y_n) = \hat{\mu}_n(y_1 e_n + \dots + y_n e_n).$$

It is easy to show that $\phi_n(y_1, \dots, y_n)$ satisfies the functional equation:

$$\phi_n(y_1, \dots, y_n) = \prod_{i=1}^{k} \left[\phi_n(\lambda_{i1} y_1, \dots, \lambda_{in} y_n)\right]^{a_i} \qquad (4.47)$$

A little consideration will now show that (4.46) implies that the probability measure corresponding to ϕ_n has moments of all order. Consequently by theorem 4.6, it follows that, for each n, μ_n is normal. Hence by virtue of theorem 4.5, μ is normal. This completes the proof.

CONCLUDING REMARKS

The intrinsic beauty and day-to-day application of normal distribution or normal law is amazing, and has generated tremendous philosophical curiosity. Many scholars, since its derivation in 1730, which is accredited to Abraham de Moivre (1667-1754), a French mathematician, are struck by its versatility and fascinating depicting character of natural phenomena. It is a focus of study from different angles, namely, pure, applied and statistical mathematics. Thus, in this perspective, we can see the advantage of studying probabilistic theories in the light of normal distribution in historio-philosophic contextualization.

Its historical aspect with respect to the study of probability theory gives us a *truth-in-life* interpretation of mathematical probability. That is, the history of normal distribution, as dealt with in chapter three, clarifies its central importance in the progress of probabilistic and statistical thoughts. It is widely used, and somehow abused, in inferential problems. Its amenability to exact mathematical treatment has generated highly sophisticated mathematical methods, and enabled scholars to study a variety of problems in a rather organized and systematic way.

The *philosophical interpretation* with respect to the general contextualization of mathematical theorems or laws of nature is still a baffling question to mathematicians. Kac(1959) poses the question several times and tries to give his scientific observation. Nonetheless, we need further investigations to verify and accept normal distribution as law of nature. Parallel to this we note that Poisson(1832) believed that all events of a moral as well as of physical nature are subject to his formula - Poisson distribution. This tendency invites us to go deep into the relation between the mathematical discoveries and natural order. A serious scholar can venture into this realm and come up with satisfactory explanations. As a primary clue it can be said that there is a *possibility* of presenting natural phenomena or order with mathematical probability with admissible minimum error.

These possibilities are the primary stimuli for the diversifying factors of the different approaches to mathematical probability theory and rise of conflicting schools of probability. Mathematics, since the inspiring studies of János Bolyai (Hungarian, 1825) and Lobachevskiĭ (Russian, 1826), that there is a possibility of constructing geometry on axioms different from Euclid's - non-Euclidean geometry, is considered to be a highly abstract science having both pure or speculative, and practical or applied aspects. The promotion of the *axiomatic foundation* of probability theory can be seen in this viewpoint. The most successful scholar in this research field is Kolmogorov(1933), even though his axioms are not sufficient for all aspects of probabilistic studies - they do not give room to the theory of chances. In other words, they do not determine which of a pair of hypotheses about a distribution is better supported by a given data. Thus, further improvements of his measure-probabilistic axioms are required.

Next, after these preparatory comments, we can deepen our knowledge on the *Nairobi studies in philosophy of mathematical probability and statistics*. This issue is addressed in section 2.3 in detail. Nonetheless, further clarification can purify the air for the curious on lookers. The method of analysis is mainly based on the external criticism, that is, the background of the founders, the method of teaching and books of references, the research papers and dissertations. But this method, without a thorough internal criticism, is not enough to reach a solid conclusion. In the process of internal criticism we are able to investigate the deep meaning of the mathematical probability discoveries. In this way we can go beyond the symbolic reading of formulae or what the author intended and what they typify, and arrive at convincing philosophical interpretations of the new contributions and stand of outlook. All in all the measure-theoretic approach has a high esteem among the scholars and as Prof. Odhiambo says hitherto the priority of scholars is mainly determined by survival factor rather than with philosophical thought of one's conviction. So at this stage it is difficult to give a specific standpoint in a scholarly manner. But with extensive internal and external criticism of the research works and related topics it is possible to give a clear picture of the philosophical works on the mathematical probability and statistics at the University of Nairobi. In other words, Nairobi studies on philosophy of mathematical probability and statistics can be explored effectively using the *external criticism* of the over all background, methodology and books of reference, and *internal criticism* on the research works in general, and philosophical analysis of ten mathematical probabilistic and statistical dissertations (1981-1997) in particular.

A fare criticism on the presentations and analyses of historical facts can guide us to a comprehensive study on historio-philosophic study on the development of mathematical probabilistic and statistical thoughts. The authoritative works on these fields sometimes start with their own point of view and tend to conclude accordingly. To mention but a few, like Maistrov(1974), with an economic theory, Hacking(1975), with philosophical point of view, and Owen(1976), with geographical limitation. These show that historical presentations and analyses of mathematical probability theories are dependent on the availability of raw materials, mainly on the work done by the prominent scholars in that specified field, the subjective interpretation or personal impressions and preferences. So in the process of acceptance of the *historio-philosophic analysis* a critical eye and substantial study on the essence of the subject itself are unavoidable factors.

The general subject of probability at present covers mathematics, measurements or statistical data, theory of nature and theory of knowledge it self, hence, *probabilistic philosophers* venture on the methodological and epistemological issues. Good means of comparison of discernment is that of Heisenberg's(1927) uncertainty principle - to formulate other similar or different observations; since philosophy as such is a reflected knowledge or experienced insight on being, nature, or essence and accidents of things.

Neyman(1957) advocates *inductive behaviour* as the sole means of foundation of science, while de Finetti(1972) and his colleagues ardently claim that *inductive reasoning* is the true ladder of knowledge. These two points of view are the chief bones of contention in contemporary philosophy of probability and statistics. There are two distinct aspect of all approaches, namely, conceptual and mathematical questions. Nonetheless, we can't work out philosophy of science in general and mathematical probability and statistics in particular without inductive reasoning and inductive behaviour. They complement each other.

The concept of *randomness*, generally in science, propagated by Fisher(1922)and his colleagues, is intrinsically an open-ended concept. We should note that it is misleading to give a rigorous operational definition of randomness, but its subjective meaning can be grasped easily. It refers to the chance happening of a future event, always about the unknown. This property of randomness, that is past versus future has fascinating philosophical connotation. Prediction of nature or a generalized description of the behaviour of nature, is an important goal of science in general, and of probability theory in particular, but the degree of reliability of the prediction is not absolutely perfect. This degree of uncertainty can be determined by using the methods of probability and statistical theories. Analysis of scientific facts, including their intrinsic uncertainties, enrich the philosophical intuition and scientific facts can no longer be seen as probable. They, if the probable error or random error is known and determined, are the most reliable knowledges acquired by human means, because they include a realistic self-appraisal.

The diversity of the definitions and derivations of normal distribution indicate that it attracted wide range of spectrum. Consequently, we observe that the principles of convergence, namely the *central limit theorem* and laws of *large numbers* are associated with normal distribution. The relationship with the other theoretical distributions is equally amazing. The discrete and continuous distributions can be approximated easily by normal distribution. Its relation with pure and applied mathematics is appealing. Especially ***normal law*** in ***number theory*** can be further investigated and enrich mathematical probability. Here we have an attractive central question of a probabilistic philosopher, that is, "Is normal law a mathematical theorem or law of nature?"

The following remarks reflect some already explained properties of the *characterisation of normal distribution in Hilbert space* and future problems for research.

- The proof of the above theorem 4.7 mainly depend on the assumption that the B's in (4.9) are commuting Hermitian operators with discrete spectrum. In finite dimensional Euclidean spaces this assumption is satisfied if the B's are commuting symmetric matrices. In finite dimensional separable Hilbert space the assumption is satisfied when the B's are commuting compact Hermitian operators.
- An interesting question that can be raised now is whether an analogous theorem is valid when the B's are not commuting and/or do not have a discrete spectrum. This problem is not yet solved and a satisfactory answer in this case would be extremely interesting.
- In the above proof, the assumption that B's are bounded operators with $\|B_i\| < 1$ for all *i* is used. This assumption can be relaxed slightly. Consider a functional equation of the form

$$\prod_{i=1}^{k} [\hat{\mu}(B_n y)]^{a_i} = 1$$

and suppose that there exists a complete orthonormal sequence $\{e_n\}$ of vectors such that, for each *i*,

$$Bi = \Sigma \lambda_{in} P_n \text{ and } |\lambda_{in}| < |\lambda_{1n}| \text{ for all n.}$$

Then μ is normal if and only if the equation

$$\prod_{i=1}^{k} [\psi_n(\lambda_{in} y_n)]^{a_i} = 1$$

where $\psi_n(y_n) = \hat{\mu}(y_n e_n)$, implies that $\psi_n(y_n)$ corresponds to a normal law on the real line.

- An immediate consequence of theorem 4.7 is that to assert the normality of μ, whose characteristic function satisfies (4.9), it suffices to establish that, for each *n*, the characteristic function $\psi_n(y_n)$ corresponds to a normal law. The verification of this last assertion is much simpler in practice. The theorem of Rao and Ramachandran (1968) can now be directly applied to (4.46) to establish that $\psi_n(y_n)$ corresponds to a normal law.

- Theorem 4.7 strengthens the Eaton-Pathak theorems(1969). In Eaton-Pathak theorems it is proved only for invertible operators, but in Theorem 4.7 there is no such assumption for the operators B's. Indeed, Eaton(1970) affirms that the Eaton-Pathak theorem can be proved for non invertible operators.

Parallel to these research studies, we can ponder upon the characterisation of normal law in Hilbert space, and present some propositions which can enable us to give an *analogy of inner product* with the covariance of two random variables with zero expectations is manifested.

With the help of the definition and properties of Hilbert space, we note that the inner product(u,v) of two functions defined by

$$(u, v)^2 = \int_{-\infty}^{+\infty} u(x)\, \overline{v(x)}\, dx,$$

the inner product $\mathcal{L}_2$ becomes a Hilbert space.

Then, after Feller's(1991) assessment, we can arrive at the required proposition.

LIST OF REFERENCES

ADRIAN, R. (1808): "Research concerning the probabilities of the errors which happen in making observations," *The Analyst*, or *Mathematical Museum* 1, No. 4, (Philadelphia), 83-87, 93-109.

AKHIEZER, N. I. and GLAZMAN, I. M.(1961): *Theory of Linear Operators in Hilbert Space, Vol. II*. New York: Frederick Ungar Publishing Co.

AKHIEZER, N.I. and M. GLAZMAN, I. M. (1963): *Theory of Linear Operators in Hilbert* Space, *Vol. II*. New York: Frederick Ungar Publishing Co.

ALEXANDROV, P. S. *et al*. (1969): "Bernstein's obituary,"*Uspehi Mat. Nauk* 24,211-9.

ANDERSON, T. W. (1971): The Statistical Analysis of Time Series. New York: Wiley.

ANDERSON, T. W.(1958): *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley & Sons, Inc.

ARBUTHNOT, J. (1710):"An argument for divine providence taken from the constant regularity observed in the births of both sexes," *Philosophical Transactions of the Royal Society of London*, 27, 186-90.

ARCHIBALD, R. C. (1926): "A rare pamphlet of De Moivre and some of his discoveries," *Isis* 8, no. 4(28), 671-684.

BASU, D. (1951): "On the independence of linear functions of independent chance variables," *Bull. Inst. Internat. Statist.*, 33, 83-96.

BAYES, T. (1950): "An essay toward solving a problem in the doctrine of chance," *Bulletin Dept. Agric.*, Washington (originally published in 1763, *Phil. Trans. Roy. Soc.*).

BELL, E. T.(1937): "Men of Mathematics," *Simon and Schuster*, 259.

BENNETT, J. H. ed.(1990):*Statistical Inference and Analysis: Selected Correspondence of R. A. FISHER*. Oxford: Clarendon Press.

BERNOULLI, D.(1954): "Specimen theoriae novae de mensura sortis," *Econometrica*, 22, 23-26 (originally published in 1738 *Comm. Acad. Sci. Imp. Petropolitanae*,5,175-192).

BERNOULLI, J.(1713): *Ars conjectandi*. Basel.

BERNSTEIN, S. N.(1917): "An essay on the axiomatic foundations of probability theory," *Proceedings of Kharkov Mathematical Association*, Vol 15.

BERNSTEIN, S. N.(1939): "Several Comments concerning the Limit theorem of Liapounoy," *Dokl. Akad. Nauk. SSSR* 24, 3-7

BERNSTEIN, S. N.(1946): *Probability Theory*.Moscow-Leningrad:1927-1st edition, 1946-4th edition.

BERRY, A.(1941): "The accuracy of Gaussian approximation to the sum of independent variates," *Trans. Am. Math. Soc.* 49 , , 122-136.

BIENAYMÉ, I. J. (1853): "Considérations à l'appui de la découverte de Laplace sur la loi des probabilités dans la méthode des moindres carrés," *Comptes Rwndus Acad. Sci. Paris* 37.

BOLTZMANN, L. (1964): *Lectures on Gas Theory*. Berkeley: University of California Press.

BOREL, E. (1909): "Sur les probabilités dénombrables et leurs applications arithmétiques," *Rend. Circ. Mat. Palermo*, 27, 247-271.

BOREL, E. (1939): "Valeur pratique et philosophie des probabilités," *Traité*, Vol 3, Gauthier-Villars, Paris.

BOREL, E. (1953): "Game theory," *Econometria*, 21, 97-174(originally published in 1921, *Comptes Rend., Paris*)

BORTKIEWICZ, L. (1898): *Das Gesetz der kleinen Zahlen* - in which the Poisson distribution was revived after a half century of nearly perfect vacuum.

BOSE, R. C. and BUSH, K. A.(1952): "Orthogonal arrays of strength two and three," *Ann. Math. Stat.*, 23, 508-524.

BOSE, R. C. and CARETER, R. L.(1959): "Complex representation in the construction of rotatable designs," Ann. Math. Stat., 30, 771-780.

BOSE, R. C. and CONNOR, W. S.(1952): " Combinatorial properties of group-divisible incomplete block designs," *Ann. Math. Stat.*, 23, 367-383.

BOSE, R. C. and CONNOR, W. S.(1960): "Analysis of fractionally replicated $2^m \times 3^n$ designs," *Bull. Intern. Stat. Inst.*, 37, 142-160.

BOSE, R. C. and DRAPER, N. R.(1959): "Second-order rotatable designs in three dimensions," *Ann. Math. Stat.*, 30, 1097-1107.

BOSE, R. C. and KISHEN, K.(1940): "On the problem of confounding in general symmetrical factorial design," *Sankhyā*, 5, 21-36.

BOSE, R. C. and SHRIKHANDE, S. S.(1963): "On the composition of balanced incomplete block designs," *Can. J. Math.*, 12, 177-188.

BOSE, R. C. and SRIVASTAVA, J. N.(1964): "Analysis of irregular factorial fractions," *Sankhyā*, A26, 117-44.

BOSE, R. C. and SRIVASTAVA, J. N.(1964): "Mathematical theory of factorial designs. I. Analysis. II. Construction" *Bull. Intern. Stat. Inst.*, 40, 780-794.

BOSE, R. C.(1938): "On the application of properties of Galois fields to the problem of construction of hypergraecolatin squares," *Sankhyā*, 3, 328-338.

BOSE, R. C.(1939): "On the construction of balanced incomplete block designs," *Ann. Eugenics*, 9, 353-399.

BOSE, R. C.(1942a): "On some new series of balanced incomplete block designs," *Bull. Calcutta Math. Soc.*, 34, 17-31.

BOSE, R. C.(1942b): "A note on the resolvability of incomplete block designs," *Sankhyā*, 6, 105-110.

BOSE, R. C.(1942c): "A note on two series of incomplete block designs," *Bull. Calcutta Math. Soc.*, 34, 129-130.

BOSE, R. C.(1947): "On a resolvable series of balanced incomplete block designs," *Sankhyā*, 8, 249-56.

BOSE, R. C.(1959): "On the application of finite projective geometry for deriving a certain series of balanced Kirkman arrangements," *Bull. Calcutta Math. Soc.*, Golden Jubilee number, 341-54.

BOSE, R. C.(1961): "On some connections between the design of experiments and information theory," *Bull. Intern. Stat. Inst.*, 38, 257-271.

BOSE, R. C.(1963a): "Strongly regular graphs, partial geometries and partially balanced designs," *Pacific J. Math.* 13, 389 - 419.

BOSE, R. C.(1963b): "Combinatorial properties of partially balanced designs and association schemes," *Contributions to Statistics*. Presented to Prof. P. C. Mahalanobis in his 70th Birthday. Pergamon Press, New York, 21-48.

BOSE, R. C., SHRIKHANDE, S. S. and BHATTACHARYA, K. N.(1953) "On the construction of group divisible incomplete block designs,"*Ann. Math. Stat.*,24,167-95.

BRUCK, R. H.(1951): "Finite nets. I. Numerical invariants," *Can. J. Math.*, 3, 94-107.

BRUCK, R. H.(1963): "Finite nets. II. Uniqueness and imbedding," *Pacific J. Math.*,13, 421-457.

BUNYAKOVSKII, V. Ya. (1846): *Foundations of the Mathematical Theory of Probabilities*[in Russian], St Petersburg, II-III.

CANTELLI, F. P.(1932): "Una teoria astratta del calcolo delle probabilità," *Giorn. Ist. It. Attuari*, 3, 257-265.

CANTELLI, F. P.(1935): "Sulla esstensione del principio delle probabilità totali ad una successione illimitata di eventi incompativili," *Giorn. Ist. It. Attuari*, 6(4), 415-427.

CANTELLI, F. P.(1939): "Su una teoria astratta del calcolo delle probabilità e sulla sua applicazione al teorema detto delle probabilità zero e uno," *Giorn. Ist. It. Attuari*, 10, 1-9

CANTELLI, F. P.(1958): "Alcune memorie matematiche," *Giuffrè*, Milano[papers 23(1916) and 24(1917) are printed in de Finetti(1972), 175-188 and 189-213].

CARDANO, G.(1501 -1576): *Liber de Ludo Aleae* [The Book on Games of Chance], 1663 - deals with combinatorial problems, casting of dice.

CARNAP, R.(1950): *Logical Foundations of Probability*. University of Chicago Press.

CARNAP, R.(1952): *The Continuum of Inductive Methods*. University of Chicago Press.

CARNAP, R.(1962): "The Aim of Inductive Logic," *Logic, Methodology and Philosophy of Science*, 303-318.

CASTELNUOVO, G.(1937a): *Calcolo delle Probabilità*. Bologna: Zanichelli

CASTELNUOVO, G.(1937b):"La probabilité dans les différentes branches de la science," *Act. Sci. Ind.* 463

CHAMPERNOWNE, D. G.(1933): "The construction of decimals normal in the scales of ten," *Jour. London Math.*, 8, 254-260.

CHANDA, K. C. (1955): "On some moment properties when two polynomials have independent distributions," *Calcutta Statist. Assoc. Bull.*, 6, 40-44.

CHANG, Wei-Ching (1976): "Statistical Theories and Sampling Practice," *On the History of Statistics and Probability*, 299.

CHEBYSHEV, P. L. (1946): *Selected Mathematical Works*. Moscow-Leningrad.

CHEBYSHEV, P. L. (1947): *Polnoe sobranie sochineni (Complete Collected Works)*, Vol. 2; Vol. 3 (1948), Vol 5 (1951) (Moscow-Leningrad).

CHEBYSHEV, P. L. (1948): *Complete Collected Works*, Vol. 3.

CHERNOFF, H.(1976): "The interaction Between Large sample Theory and Optimal Design of Experiments," *On the History of Statistics and Probability*. New York: Marcel Dekker, 208-23.

CHOW, Y. S. and TEICHER, H. (1978): *Probability Theory: Independence, Interchangeability, Martingales*. New York: Springer-Verlag.

COCHRAN, W. G. (1976): "Early development of techniques in comparative experimentation," *On the History of Statistics and Probability*. New York Marcel Dekker, Inc. 1-25.

COCHRAN, W. G. (1992): *Sampling Techniques*. New Delhi: Wiley Eastern Limited

COHEN, J (1956): *Risk and Gambling: The Study of Subjective Probability*. Longmans, London.

COPI, I. M. and COHEN, C.(1990): *Introduction to Logic*. New York Macmillan Publishing Company.

COPLESTON, F. (1985a): *A History of Philosophy*- Vol. IV. New York: Image Books.

COPLESTON, F. (1985b): *A History of Philosophy* - Vol. V. New York: Image Books.

CASTELNUOVO, G.(1937b):"La probabilité dans les différentes branches de la science," *Act. Sci. Ind.* 463

CHAMPERNOWNE, D. G.(1933): "The construction of decimals normal in the scales of ten," *Jour. London Math.*, 8, 254-260.

CHANDA, K. C. (1955): "On some moment properties when two polynomials have independent distributions," *Calcutta Statist. Assoc. Bull.*, 6, 40-44.

CHANG, Wei-Ching (1976): "Statistical Theories and Sampling Practice," *On the History of Statistics and Probability*, 299.

CHEBYSHEV, P. L. (1946): *Selected Mathematical Works*. Moscow-Leningrad.

CHEBYSHEV, P. L. (1947): *Polnoe sobranie sochineniy (Complete Collected Works)*. Vol. 2; Vol. 3 (1948), Vol 5 (1951) (Moscow-Leningrad).

CHEBYSHEV, P. L. (1948): *Complete Collected Works*, Vol. 3.

CHERNOFF, H.(1976): "The interaction Between Large sample Theory and Optimal Design of Experiments," *On the History of Statistics and Probability*. New York: Marcel Dekker, 208-23.

CHOW, Y. S. and TEICHER, H. (1978): *Probability Theory: Independence, Interchangeability, Martingales*. New York: Springer-Verlag.

COCHRAN, W. G. (1976): "Early development of techniques in comparative experimentation," *On the History of Statistics and Probability*, New York: Marcel Dekker, Inc. 1-25.

COCHRAN, W. G. (1992): *Sampling Techniques*. New Delhi: Wiley Eastern Limited.

COHEN, J.(1956): *Risk and Gambling: The Study of Subjective Probability*. Longmans, London.

COPI, I. M. and COHEN, C.(1990): *Introduction to Logic*. New York: Macmillan Publishing Company.

COPLESTON, F. (1985a): *A History of Philosophy*- Vol. IV. New York: Image Books.

COPLESTON, F. (1985b): *A History of Philosophy* - Vol. V. New York: Image Books.

CRAMÉR, H.(1937): *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics and Math. Phy. 36, Cambridge University Press.

CRAMÉR, H.(1946): *Mathematical Methods of Statistics*. Princeton University Press.

DARMOIS, G, (1951): "Sur diverse propriétés caractéristiques de la loi de probabilité de Laplace-Gauss," *Bull. Int. Statist. Inst.*, 33, 79-82.

DAVID, F. N. (1955): "Studies in the history of probability and statistics I. Dicing and Gaming," *Biometrika* 42, 1-15.

de FINETTI, B. (1931c): "Sul significato soggettivo della probabilità," Fundamenta Mathematicae, 17, 298-329.

de FINETTI, B. (1936-41): Statistica e probabilità nella concezione di R. von Mises. Probabilisti di Cambridge, Punti di vista: Hans Reichenbach e D. von Dantzig. *Statistica* and *Supplemento Statistico*, Padova.

de FINETTI, B. (1937): "La prévison: ses loi logiques, ses sources subjectives," *Ann. Inst. H. Poicaré(Paris)*, 7, 1-68 [English translation - Foresight: its logical laws, its subjective sources," *Studies in Subjective Probability*, New York: Wiley, 1961].

de FINETTI, B. (1951): "Recent Suggestions for the reconciliation of theories of probability," *Proc. Second Berkeley Symposium*, 217-225.

de FINETTI, B. (1959): "La probabilità e la statistica nei rapporti con l'innduzione, secondo i diversi punti di vista," *Corso C.I.M.E. su Induzione e Statistica*, Cremonese, Rome.

de FINETTI, B. (1972): *Probability, Induction and Statistics: The art of guessing*. London: John Wiley & Sons.

de FINETTI, B. (1975): *Theory of Probability: A Critical Introductory Treatment*, Vol.2. London: John Wiley and Sons. Here de Finetti states that "the idea that all natural characteristics have to be normally distributed is one that can no longer be sustained: it is a question that must be settled empirically."

de FINETTI, B. and SAVAGE, L. J.(1962): "Sul modo di scegliere le probabilità iniziali," *Bibl. d. Metron*, 1, 81-154.

de FINETTI, B. (1972): *Probability, Induction and Statistics*. London: John Wiley & Sons. [Himself is an ardent adherent of a *subjectivist or personalist school of probability*. Following F. P. Ramsey, in 1937, he made a systematic attempt to base the mathematical theory of probability on the notion of partial belief- contains many past papers.]

De MOIVRE, A.(1711): "De mensura sortis, seu, de probabilitate eventuum in ludis a casu fortuito pendentibus," *Philosophical Transactions of the Royal Society*, 27, 213 - 64. [Translated in Philosophical Transactions Abridged, 4, 190 - 210.]

De MOIVRE, A.(1756): *The Doctrine of Chances, or a Method of Calculating the Probability of Events in Play*, London: 1718; second edition 1738, third edition 1756.

DeGROOT, M. H. (1970): *Optimal Statistical Decisons*. New York: McGraw-Hill Book Company.

DOOB, J. L. (1976): "Foundations of Probability theory and its influence on the theory of statistics," *On the History of Statistics and Probability*, 195-204.

DORFMANN, R.(1943): "The detection of defective members of large populations," *Annals of Mathematical Statistics*, 14, 436 - 440.

DURBIN, P. R. (1967): "Probability," *New Catholic Encyclopedia*, 11, 815

EATON, M. L. and PATHAK, P. K.(1969): "A characterization of the normal law on Hilbert space," *Sankhyā, Ser. A.*, 31, 259-268.

EGGENBERGER, J. (1894): "Beitrage zur Darstellung des Bernoullischen Theorems, etc." *Mitt. Naturforsch. Gesellsch.Bern* for 1893, no.1305-1334, 110-182. Also published as a separate edition(1906), Berlin: Fischer.

ESSEEN, G.(1945): "Fourier Analysis of distribution functions. A Mathematical Study of the Laplace-Gaussian law," *Acta Math.*, 1-125.

FELLER, W.(1991): *An Introduction to Probability Theory and Its Applications*, Vol. II. New Delhi: Wiley Eastern Limited.

FELLER, W. (1993): *An Introduction to Probability Theory and Its Applications*. Vol. I. New Delhi: Wiley Eastern Limited.

FELLER, W.(1935): "Über den Zentralen Grenzwertsatz der Wahrscheinlinchkeitsrechnung,"*Math. Z.* 40, 521-559.

FINUCAN, H. M.(1964): "The blood testing problem," *Applied Statistics*, Vol. 13, 43-50.

FISHER, R. A. (1920): "A Mathematical examination of the methods of determining the accuracy of an observation by the mean error and by the mean square error," *Monthly Notices Roy. Astron. Soc.*, Vol. 80, 758-770.

FISHER, R. A. (1922): "On the Mathematical foundations of theoretical statistics," *Phil. Trans. Roy. Soc., London, Ser. A*, Vol. 222, 309-368-Development of a theory of point estimation on the basis of the maximum likelihood principle.

FISHER, R. A.(1925): "Theory of statistical estimation," *Proc. Cambridge Phil. Soc.*, Vol. 22, 700-725.

FISHER, R. A.(1925): *Statistical Methods for Research Workers*. Edinburgh Oliver & Boyd,11th ed.,1950.

FISHER, R. A.(1950): *Contributions to Mathematical Statistics*. New York: John Wiley.

FISHER, R. A.(1955): "Statistical methods and scientific induction," *J. Roy. Stat. Soc.*B 17, 69-78.

FRANKFURT, U. I. and FRANK, A. M. (1964): *Dzozaja Villard Gibbs* (J. W. Gibbs), Moscow.

FRÉCHET, R. M. (1939-43): *Les Probabilités Associées à un Systemè d'Evénements Compatibles et Dépendants*. Paris: Hermann.

FRÉCHET, R. M.(1951): "Generalisation de la loi probabilite de Laplace,"*Ann. de l'Institute Henri Poincare* 12, 1-29.

FUHRMANN, P. A.(1981): *Linear Systems and Operators in Hilbert Space*. New York: McGraw-Hill.

GALILEI, G. (1564-1642): *Considerazione sopra il Giuco dei Dadi* [On Outcomes in the Game of Dice], 1718.

GALTON, F.(1889). *Natural Inheritance*.

GAUSS, C. F.(1957): "Izbrannyc geodezicheskie sochineniya,"[Selected Geodesic Works],Vol. I Moscow.

GNEDENKO, B. V. and KOLMOGOROV, A. N.(1968): *Limit Distributions for sums of Independent Random Variables*. London: Addison-Wesley Publishing Company,.

GNEDENKO, B. V.(1948): *Trudy Inst. Istor. Estestvoznam(Proc. Inst. History Natural Sciences)*, 2, 394.

GORDON, F. S. (1973): "Characterizations of populations using regression properties," *Ann. of Statist.*, 1, 114-126.

GORDON, F. S. and MATHAI, A. M. (1972): "Characterizations of the multivariate normal distribution using regression properties," *Ann. Math. Statist.*, 43, 205-229.

GRAUNT, J.(1662): *Natural and Political Observations upon the Bills of Mortality*.

HACKING, I. (1971): "Jacques Bernoulli's Art of Conjecturing," *Brit. J. Phil. Sci.* 22(3), 209-249.

HACKING, I. (1974): *Logic of Statistical Inference*. London: Cambridge University Press.

HACKING, I. (1975): *The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference*. London: Cambridge University Press. [It is among the authoritative works on the history of probability and statistics in English].

HALMOS, P. R.(1967): *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*. New York: Chelsea.

HALMOS, P. R.(1967): *A Hilbert Space Problem Book*. Princeton: D. Van Nostrand Company, Inc.

HANSEN, M. H. and MADOW, W. G.(1976): "Some Important Events in the Historical Development of Sample Surveys," *On the History of Statistics and Probability*, 84-86.

HAYDE, C . C. and SENETA, E. (1972): "Studies in the History of probability and statistics, XXXI. The simple branching process, a turning point test and a fundamental inequality: A historical note on I. J. Bienaymé," *Biometrika*, 59, 680 - 683.

HEISENBERG, W.(1927): "The Physical Principles of the Quantum Theory, tr. C. Eckart and F. C. Hoyt (Chicogo 1930, pa repr. New York n. d.), "Uber den anschaulichen Inhalt der quanten theoretischen Kinematik und Mechanik,"*Zeitschrift für Physik* 43,172-98.

HEYDE, C. C. and SENETA, E.(1975): *Studies in the History of Probability and statistics*, Vol. 2.

HOFFMAN, A. J. and RAY-CHAUDHURI, D. K. (1965a): "On the line graph of a finite affine plane," *Can. J. Math.*, 17, 687-94.

HOFFMAN, A. J. and RAY-CHAUDHURI, D. K. (1965b): "On the line graph of a symmetric balanced incomplete block design," *Trans. Am. Math., Soc.*, 116, 238 - 252.

HOFFMAN, A. J. and SINGLETON, R. R.(1960): "On Moore graphs with diameters 2 and 3," *I.B.M J. Res. Develop.*, 4, 497-504.

HOFFMAN, A. J.(1963): "On polynomial of a graph," *Am. Math. Monthly*, 70.

HOFFMAN, A. J.(1965): "On the line graph of a projective plane,"*Proc. Am. Math. Soc.*, 16, 297-302.

HOTELLING, H. (1931): " The Generalization of Student's ratio," *Ann. Math. Statist* 2, 360-378.

HUME, D. (1739): *A Treatise of Human Nature, being an attempt to introduce the experimental method of reasoning into moral subjects*. London.

HUME, D. (1748): *An enquiry Concerning Human Understanding*. London

ITÔ, K. ed., (1987): "Hilbert, David", ," *Encyclopedic Dictionary of Mathematics*, Vol. II Cambridge, The MIT Press, 736-737.

ITÔ, K. ed., (1987): "Statistical Inference," *Encyclopedic Dictionary of Mathematics*, Vol. II Cambridge, The MIT Press, 1508-1511.

JEFFREYS, H. (1931): *Scientific Inference* [1957 - 2nd ed. Cambridge University Press].

JEFFREYS, H. (1939): *Theory of Probability*. London: Oxford Uni. Press, [third ed., 1969].

JEFFREYS, H. (1955): "The present position of the theory of probability," *Brit. J. for the Philosophy of Science*, 5, 275-289.

KAC, M. (1959): *Statistical Independence in probability, Analysis and Number Theory*, New Jersey: The Mathematical Association of America, John Wiley and Sons.

KAC, M. and Steinhaus, H.(1938): "Sur les functions independantes IV," *Studia Math*.,7,1-15

KAC, M.(1943): "Convergence of certain gap series," *Ann. Math*., 44, 411- 415

KAC, M.(1949): "Probability methods in some problems of analysis and number theory," *Bull. Amer. Math. Soc*., 55, 641-665

KAGAN, A. M., LINNIK, Yu. V. and RAO, C. R. (1973): *Characterization Problems in Mathematical Statistics*. U.S.S.R. Academy of Science (English translation, Wiley, New York).

KAMKE, E. (1932): *Einführung in die Wahrscheinlichkeitstheorie*. Leipzig.

KANTOROVICH, L. V. and AKILOV, G. P.(1964): *Functional Analysis in Normed Spaces*. London: Pergamon Press.

KAPLANSKY, I.(1974): "Hilbert, David", *Encyclopaedia Britannica*, 8, 872

KEMPTHORNE, O. (1952): *The Design and Analysis of Experiments*. New York: Wiley.

KEMPTHORNE, O. (1957): *An Introduction to the Genetic Statistics*. New York: Wiley.

KENDALL, M. G. (1947 and 1948): *The Advanced Theory of Statistics, Vols. I and II*. London Griffin.

KEYNES, J. M.(1929): *A Treatise on Probability*, 2nd ed. London: Macmillan

KINGMAN, J.F. and TAYLOR, S. J. (1966): *Introduction to Measure and Probability Theory*. Cambridge: The University Press .

KLEENE, S. C. (1977): "Mathematics, Foundation of," *The New Encyclopaedia Britannica*, Vol 11, 630 - 639.

KOLMOGOROV, A. N. (1933): *Grundbegriffe der Wahrscheinlichkeitsrechnung. Ergebnisse der Mathematik*, Berlin. Springer [English translation by Chelsea, New York,1950].

KOLMOGOROV, A. N. (1947): "The role of Russian science in the development of probability theory," *Ucen. Zap. Moskov. Gos. Inst.* 91, 56- 60

KOOPMAN, B. O. (1940a): "The axioms and algebra of intuitive probability," *Ann. Math*., 41, 269-292.

KOOPMAN, B. O. (1940b): "The bases of probability," *Bull. Amer. Math. Soc.*, 46, 763-774.

KOOPMAN, B. O. (1941): "Intuitive probabilities and sequences," *Ann. Math.*, 42, 169-187.

KOOPMAN, B. O.(1957): *Quantum Theory and the Foundations of Probability*. New York: McGraw-Hill.

KOREN, J. (ed.) (1918): *The History of Statistics*, 1918,1970[contains the development and progress of Statistics in France].

KOTLARSKI, I. I.(1966): "On characterizing the normal distribution by student's law," *Biometrica*, 53, 603-606.

KUBILUS, I. P.(1956): "Probability methods in number theory,"[in Russian],*Usp. Mat. Nauk*,68, 31-66.

KULLBACK, S. and LEIBER, E. A. (1951): "On information and sufficiency,"*Ann. Math. Stat.*, 22, 79-86.

KULLBACK, S.(1968): *Information Theory and Statistics*. New York: Dover.

KUZNECOV, B. G. (1955): "Maxwell's electrodynamics, its origins, development and historical value," *Trudy Inst. Istor. Estestvoznan. Tehn.* 5, 148-149.

KYBURG, H. E. and SMOKLER, H. E. (1964): *Studies in Subjective Probability*. New York: Wiley [It contains the original works of prominent up-helders of subjective statistics, like de Finetti, Savage].

LAHA, R. G. (1956): "On a characterization of multivariate normal distribution," *Sankhyā*,14, 367-8.

LAHA, R. G. (1957): "On a characterization of the normal distribution from properties of suitable linear statistics," *Ann. Math. Statist.*, 28, 126-139.

LAHA, R. G. (1957): "On some characterization problems connected with linear structural relations," *Ann. Math. Statist.*, 28, 405-414.

LAHA, R. G. and LUKACS, E. (1960): "On a problem connected with quadratic regression," Biometrika, 47, 335-343.

LAPLACE, P. S. M. de (1812): *Théorie Analytique des Probabilités*, 1812.

LAPLACE, P. S. M. de (1878 - 1912): *Oeuvres completes*, VII, Paris.

LAPLACE, P. S. M. de (1952): *A Philosophical Essay on Probabilities*. New York: Dover Publications [Translated by Truscott, F. W. and Emory, F. L. - *Essai Philosophique sur les Probabilités* (1820), an introduction to Laplace's great work *Théorie Analytique des Probabilités*].

LEGENDRE, Andrien-Marie(1806): *Nouvelles méthodes pour la détermination des orbites des comètes*, Paris.

LEHMANN, E. L.(1958): *Some early instances of Confidences of Confidence Statements*. Statistical Laboratory, University of California, Berkeley. ONR 5 Technical Report to the Office of Naval Research.

LEHMANN, E. L.(1959): *Testing Statistical Hypotheses*. New York: John Wiley & Sons.

LEIBNIZ, G. W. (1666): *De arte combinatoria*.

LEIBNIZ, G. W.(1880): *Die Philosophiche Schriften*, Vol. 4 (Berlin), 27-102

LEVY, P. (1925): *Calcul des probabilités*.

LEVY, P. (1937): *Théore de l'Addition des Variables Aléatores*. Paris: Gauther-Villars(1954 other edition).

LEVY, P. (1949): "Les paradoxes de l'infini et le calcul des probabilités."*Bull. Sci. Math*.,73, 185-95.

LEVY, P.(1953): *Random Functions: General Theory with Special Reference to Laplacian Random Functions*. University of California Press.

LI, C. H.(1962): "A sequential method of screening experimental variables." *Journal of American Statistical Association*, 57, 455-477.

LINDERBERG, J. (1922): "Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlinchkeitsrechnung." *Math. Z.*, 15, 211-225.

LINDLEY, D. V. (1965): *Introduction to Probability and Statistics from a Bayesian Viewpoint*, in two volumes: I, *Probability*; II, *Inference*. Cambridge University Press.

LINNIK, Yu. V. (1952): "Linear statistics and the normal law."*Dokl. Akad. Nauk. SSSR*., 83, 353-5.

NNIK, Yu. V. (1956): "On the question of finding a general distribution from the distribution of a statistic," *Theor. Prob. Appl.*, 1, 422-434.

NNIK, Yu. V. (1960): *Decomposition of Probability Distributions.* [English translation, 1964, Edinburg: Oliver and Boyd Ltd.]

OBACHEVSKII, N. I.(1951): *Polnoe sobranie sochinenii* (Complete Colledted Works), Vol. 5. Moscow - Leningrad.

LOEVE, M. (1963): *Probability Theory.* Toronto: D. Von Nostrand Company.

LOEVE, M. (1978): *Probability Theory II*, 4th edition, Graduate Texts in Mathematics, Volume 46.

LUKACS, E. (1960): *Characteristic Functions.* London: Graffin.

LUKACS, E. and LAHA, R. G. (1964): *Applications of Characteristic Functions.* New York: Hafner Publishing Co.

LYAPUNOV, A. M.(1900): "Sur une proposition de la théorie des probabilités." (On a theorem in the calculas of probabilities) *Izv. Akad. Nauk. Ser. 5* (13), 359-386.

LYAPUNOV, A. M.(1901): "Nouvelle forme du théorème sur la limite de probabilité." (A new form of a theorem on the limit of probabilities) *Notes Acad. Sci. Phys. - Math. Sect. Ser. 8* 12, 1-24; This method is adhered by many prominent scholars of probability theory.

LYAPUNOV, A. M.(1948): *Selected Works.* Acad. Sci. USSR, Moscow.

MAISTROV, L. E. (1974): *Probability Theory: A Historical Sketch.* New York: Academic Press. [Translated by Samuel Kotz].

MARKOV, A. A.(1898): "Sur les racines de l'équation $\frac{e^{x^2}\partial^m e^{-x^2}}{\partial x^m} = 0$," *Izv. Akad. Nauk. Ser. 5* (9), 435-446.

MARKOV, A. A.(1898): "The law of large numbers and the method of least squares." *Izd. Fiz.-Mat. Ob-va Pri Kazan. Ser.* 2 [*Notices(Izvestiya) of the Physical-Mathematcal Society at Kazan University Ser.* 2] 8, 110-118.

ARKOV, A. A.(1906):"Extension of the law of large numbers to dependent variables," *Izd. Fiz.-Mat. Ob-va Pri Kazan. Ser. 2* (15), 155-6.

ARKOV, A. A.(1913): *Ischislenie veroyatnoste* (The Culculus of Probabilities). Moscow.

MARKOV, A. A.(1924): *The Calculus of Probabilities*, (in Russian). Moscow.

MARKOV, A. A.(1951): *Izbrannye trudy* (Selected Works, Acad. Sci. USSR). Moscow.

MATE, L. (1989): *Hilbert Space Methods in Science and Engineering*. Bristol: Adam Hilger.

MATHAI, A. M. and PEDERZOLI, G. (1977): *Characterization of the Normal Probability Law*. New Delhi: Wiley Eastern Limited.

MATHAI, A. M. (1967): "On structural properties of the conditional distributions," *Canadian Math. Bull.*, 10, 239-245.

MATHAI, A. M. and RATHIE, P. N. (1975): *Basic Concepts in Information Theory and Statistics: Axiomatic Foundations and Applications*. New York: Wiley Holsted,New Delhi: Wiley Eastern.

MATHAI, A. M. and RATHIE, P. N. (1976): *Recent Contributions to Axiomatic Definitions of Information and Statistical Measures*. Ogawa Birthday Volume.

NAYMAN, J. (1937): "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans. Roy. Soc.* A, 236, 333-380.

NAYMAN, J. (1957): "Inductive behaviour as a basic concept of philosophy of science," *Revue I. S. I.*, 25, 7-22.

NAYMAN, J and PEARSON, E. S. (1928): " On the use and interpretation of certain test criteria for purposes of statistical inference," *Biometrika*, Vol 20A, 175-240, 263-294- Proposes the likelihood ratio principle for obtaining reasonable tests, and applies it to a number of specific problems.

NAYMAN, J and PEARSON, E. S. (1932-3): "On the testing of statistical hypotheses in relation to probability a priori," *Proc. Cambridge Phil. Soc.*, Vol 29, 492-510 - In connection with the problem of hypothesis testing, suggests assigning weights for the various possible wrong decisions and the use of the minimax principle.

NAYMAN, J. (1935): "Sur un teorema concernente le cosidette statistiche sufficienti," *Giorn. Ist. Ital. Att.*, Vol 6, 320-334.

NAYMAN, J. (1938): "L'estimation statistique traitée comme un probléme classique de probabilité," *Actualités sci. et ind.*, No. 739, 25-57 - Puts forth the point of view that statistics is primarily concerned with how to behave under uncertainty rather than with determining the values of unknown parameters, with inductive behaviour rather than with inductive inference.

NAYMAN, J and PEARSON, E. S. (1933): "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trams. Roy. Soc.* A, 231, 289-337.

NAYMAN, J. and PEARSON, E. S. (1936): "Sufficient statistics and uniformly most powerful tests of statistical hypotheses," *Stat. Research Mem.* 1, 113.

NAYMAN, J. and PEARSON, E. S. (1936-8): Contributions to the theory of testing statistical hypotheses: I. Unbiased critical regions of Type A and Type B. *Stat. Res. Mem.*, 1, 1. II. Certain theorems on unbiased critical regions of Type A. III. Unbiased tests of simple statistical hypotheses specifying the values of more than one unknown parameter. *Stat. Res. Mem.*, 2, 25.

NAYMAN, J.(1938): *Lectures and Conferences on Mathematical Statistics and Probability*, Washington, 1938, 1952 - where many of his principal ideas are summarized.

NAYMAN, J.(1976): "The Emergence of Mathematical Statistics: A Historical Sketch with Particular Reference to the United States," *On the History of Statistics and Probability*, 154-8.

ORE, O.(1953): *The Gambling Scholar*. Princeton University Press (N.Y. 1961).

ORE, O.(1960): "Pascal and the invention of probability theory,"Amer. Math.Monthly 67, 409-19.

OSTROGRADSKII, M. V.(1961): *Complete Collection of Works*, Vol. 3. Academy of Sciences of URRSSR, Kiev.

OWEN, D. B. (ed.) (1976): *On the History of Statistics and Probability* [prepared by selected famous American scholars].

PARRATT, L. G.(1961): *Probability and Experimental Errors in Science*. New York: John Wiley & Sons.

PARTHASARATHY, K. R. (1967): *Probability measures on Metric Spaces*. New York: Academic Press[See in particular theorem 2.4, theorem 4.10, theorem 5.4 and theorem 6.2].

PARTHASARATHY, K. R (1977): *Introduction to Probability and Measure*. Delhi: The Macmillan Company of India Limited.

PATEL, M. S.(1962): "Group screening with more than two stages," *Technometrics*, 4(2), 209-217.

PATHAK, P. K. (1970): "A note on a characterization of the normal law on Hilbert space," *Sankhyā, Ser. A.*,32, 31-36.

PATHAK, P. K. and PILLAI, R. N.(1968): "On a characterization of the normal law," *Sankhyā, Series A*, 30, 141-144.

PATIL, G. P. and SESHADRI, V. (1963): "A Characterization of a bivariate distribution by the marginal and the conditional distributions of the same component,"*Ann. Inst. Statist. Math.*, 15, 215-221.

PATIL, G. P. and SESHADRI, V. (1964); "Characterization theorems for some univariate probability distributions," *J. Roy. Statit. Soc. Ser.* B, 26, 286-292.

PEARSON, E. S. and KENDALL, M. G. (eds.) (1970): *Studies in the History if Statistics and Probability*. London: Griffin.

PEARSON, E. S.(1965): "Studies in the history of probability and statistics," *Biometrika*, 53, 3-18.

PEARSON, K. (1899): "Mathematical contribution to the theory of evolution. On the reconstruction of the stature of prehistoric races," *Philos. Trans. Roy. Soc.* A192, 169-244.

PEARSON, K. (1900): "On the criterion that a given system of deviation from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *Phil. Mag., Ser. 5*, Vol. 50, 157-172.

PEARSON, K. (1924): "Historical note on the origin of the normal curve of errors,"*Biometrika*,16, 402-4.

PEARSON, K. (1925): "James Bernoulli's theorem," *Biometrika*, 17, 201-210.

PEARSON, K. (1926): "Abraham de Moivre," *Nature*, 117, 551-2.

PEARSON, K. (1929): "Laplace," *Biometrika*, 21, 202 -16.

POINCARÉ, H. (1906): *La Science et l'Hypothèse*. Flammarion, Paris[English edition "Science and Hypothesis" 1952, New York: Dover Publications].

POINCARÉ, H. (1908): *Science et Méthode*. Flammarion, Paris. [English edition "Science and Method" 1910, London: Thomas Nelson and Sons, with Preface by B. Russel].

POINCARÉ, H. (1912): *Calcul des Probabilités* (Calculus of Probabilities). Paris.

POINCARÉ, H. (1946): *The Foundations of Science*. Lancaster, Pennsylvania: The Science Press.

POISSON, S. D.(1837): *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précedées des règles générales du calcul des probabilités*.

POLYA, G.(1954): *Mathematics and Plausible Reasoning* Vol I. *Induction and Analogy in Mathematics*. Vol. II. *Patterns of Plausible Reasoning*. Princeton University Press., Princeton, New Jersey [revised edition with appendix, 1968]

PRECHET, M.(1951): "Generalisation de la loi probabilite de Laplace,"*Ann. de l'Institute Henri Poincare* 12.

PROHOROV, Yu. V. and FISZ, M.(1957): "A characterization of normal distributions in Hilbert space," *Teoria Veroiatnostei iee Primeneniia*, 2, 475-477.

QUETELET, A. J.(1843a): *A Treatise on Man*. Edinburgh; New York, 1968.

QUETELET, A. J.(1863b):*Social Systems and the Laws Governing Them*. St. Petersburg.

RADEMACHER, H.(1922): "Einige Sätze über Reihen von allgemeinen Orthogonal-funktinen," *Math. Ann.*, 87, 112-138.

RAGHAVARAO, D.(1966). "An extended partial geometry," *J. Indian Soc. Agric. Stat.*, 18, 99-107.

RAGHAVARAO, D.(1971): *Construction and Cobinatorial Problems in Design of Experiments.* New York: John Wiley.

RAMSEY, F. D. (1931): The Foundations of Mathematics and Other Logical Essays. Kegan, London [It contains "Truth and Probability," 1926 and "Further considerations," 1928].

RAO, C. R. (1952): "Some theorems on minimum varaince estimation," *Sankhyā* 12, 27-42.

RAO, C. R. (1967): "On some characterizations of the normal law," *Sankhyā, Series A*, 29, 1-14.

RAO, C. R. (1973): *Linear Statistical Inference and Its Applications.* New York: John Wiley & Sons.

RAO, C. R. and RAMACHANDRAN, B.(1968): "Some results on characteristic functions and characterizations of the normal and generalized stable laws," *Sankhyā, Series A*, 30, 125-40.

REICHENBACH, H. (1935): *Wahrscheinlichkeitslehre.* Leiden: Sijthoff.

REID, C. (1970): *With an appreciation of Hilbert's mathematical work by H. Weyl.* Springer.

REID, C. (1996): *Hilbert(David).* San Francisco: CA - a recent edition. The authoritative biography of D. Hilbert is written by Constance Reid, student and life long colleague; he wrote two biographical sketches: the first appeared in the 1922 *Naturwissenschaften* and the second at the end of the collected works.

REID, C. (1996): *Hilbert(David).* San Francisco: CA.

RÉNYI, A. (1955a): "On a new axiomatic theory of probability," *Acta Math. Acad. Scient. Hungaricae*, 6, 285-535.

RÉNYI, A. and TURÁN, P. (1958): "On a theorem of Erdős - Kac," *Acta Arith.*, 4, 71-84.

RÉNYI, A.(1955): "On the density of certain sequences of integers," *Publ. Inst. Math. Belgade*, 8, 157-162.

RIESZ, F. and NAGY, B. (1955): *Leçons d'analyse fonctionelle*. Akademiai Kiadó, third edition [English translation, *Functional analysis*, Ungar, 1955].

RUBEN, H. (1974): "A further characterization of the normality through the sample variance," *Sankhyā*, *Ser.*, *A* 37, 72-81.

RUBEN, H. (1975): "A new characterization of the normal distribution through the sample variance," *Sankyā*, *Ser.*, *A* 36, 379-388.

SAVAGE, L. J. et al. (1962): *The Foundations of Statistical Inference*. New York : Wiley.

SAVAGE, L. J.(1954): *The Foundations of Statistics*. Wiley, New York.

SAVAGE, L. J.(1961):"The foundations of statistics reconsidered,"*Proc. IV Berkeley Symp.*,1, 575- 586.

SAVAGE, L.J.(1961): "Elicitation of Personal Probabilities and Expectations," *J. Amer. Stat. Assn.*, 66, 783-801.

SCHEFFE, H. (1959): *The Analysis of Variance*. New York: Wiley.

SHEYNIN, O. B. (1965): "On the works of R. Adrian in the theory of errors and its applications," *Isotr.-Mat. Issled*, 16, 325-336.

SHEYNIN, O. B. (1966): "Origin of the theory of errors, " *Nature*, 211, 1003-4;

SHEYNIN, O. B. (1968): "On the early history of the law of large numbers," *Biometrika*, 55, 459-467.

SHEYNIN, O. B. (1970a): "On the history of the limiting theorems of de Moivre and Laplace," *Istor. i Metodologiia Estestvennykh Nauk*, 9, 199 - 211

SHEYNIN, O. B. (1970b): "Daniel Bernoulli, on the normal law," *Biometrika*, 57, 199-202.

SHEYNIN, O. B. (1970c): "On the early history of the law of large numbers," *Biometrika*, 55, 459 - 67.

SHEYNIN, O. B.(1971): "On the history of some statistical laws of distribution," *Biometrika*, 58, 234-6.

SHIMONY, A. (1955): "Coherence and the Axioms of Confirmation," J. *Symbolic Logic*, 20, 1-29.

SKITOVICH, V. P. (1954): "Linear forms in independent random variables and the normal distribution law," *Izv. Akad. Nauk. SSSR, Ser. Mat.*, 18, 185-200.

SOBEL, M. and GROLL, P. A. (1959): "Group testing to eliminate efficiently all defectives in a biological sample," *The Bell System Journal*, Vol. 38, 1179-1252.

STERRET, A (1957): "On the detection of defective members of large populations," *Annals of Mathematical Statistics*, 28, 1033-1036.

STUDENT (W. S. Gosset), (1908): "The probable error of a mean," *Biometrika*, 6, 1-25.

TARTAGLIA, N. (1556): *Trattato generale di numeri e misure* [General Treatise on Number and Measure] - various poroblems of combinatorics and probability theory are closely interrelated.

TODHUNTER, I. (1865): *History of the Mathematical theory of Probability from the Time of Pascal to that of Laplace.* New York: 1965 [It covers 1654-1812 and is among the authoritative works on the history of probability and statistics in English].

TUKEY, J. (1957): "Some examples with fiducial relevance," *Ann. Math. Stat.* 28, 687-695.

van DANTZIG, D. (1957): "Statistical Priesthood: Savage on Personal probabilites," *Statistica Neerlandica*, 11, 1-16.

van DANTZIG, D. and HEMELRIJK, J. (1954): "Statistical methods based on a few assumptions," *Bull. Int. Statist. Inst.* 342, 239-267.

von MISES, R. E.(1919): "Grundlagen der Wahrscheinlichkeitsrechnung,"*Math. Zeitschrift*, 5, 52-99.

von MISES, R. E.(1930): *Probability and Statistics.* Moscow-Leningrad [English ed., New York: Academic Press, 1964].

von MISES, R. E.(1931): *Wahrscheinlichkeitsrechnung und ihre Anwendung.* Leipzig: Deuticke.

von MISES, R. E.(1936): *Wahrscheinlichkeit, Statistik und Wahrheit.* Vienna: Springer[English translation by Macmillan, New York, 1957].

von NEUMANN, J.(1929): "Allgemeine Eigenwerttheorie Hermitescher Functional-operataren," *Math. Ann.* 102, 49-139.

von NEUMANN, J.(1932): Mathematische Grundlagen der Quantenmechanik, Springer.

von WRIGHT, G. H. (1977): "Mathematics, History of," *The New Encyclopaedia Britannica*, Vol. 11, 666-71.

WALD, A.(1938): "Die Widerspruchsfreiheit des Kollektivbegriffes," *Act. Sci. Ind.*, No. 735, 75-99.

WALD, A.(1947): *Sequential Analysis*. New York: John Wiley.

WALD, A.(1950): *Statistical Decision Functions*. New York: John Wiley.

WALD, A.(1955): *Selected Papers in Statistics and Probability*. New York: McGraw-Hill.

WALD, A.(1958): *Selected Papers in Statistics and Probability by Abraham Wald*. Stanford: University Press.

WATSON, G. S.(1961): "A study of the group screening method," *Technometrics*, Vol. 3, 371-388.

WEYL, H.(1944): "David Hilbert and his mathematical work," *Bull. Amer. Math. Soc.*, 50, 612-654.

WHITESIDE, D. T. (1987): "De Moivre," *Encyclopedia Americana*, 8, 699.

WIJSMAN, R. A. (1957): "Random orthogonal transformations and their use in some classical distribution problems in multivariate analysis," *Ann.Math. Statist.* 28,415-23.

WISHART, J. (1928): "The Generalized product moment distribution in samples from a normal multivariate population," *Biometrika* 20 A, 32-52.

WOLD, H. (1938): *A Study in the Analysis of Stationary Time Series*. University of Uppsala [1954 ed. Stockholm: Almqvist and Wicksell].

ZERNOV, N. E.(1843): *Probability Theory*. Moscow.

ZINGER, A. A. and LINNIK, Yu. V. (1955): "An Analytic generalization of Cramer's theorem and its applications," *Vestnik Leningrad. Univ.*, 10, 51-56.

ZINGER, A. A. and LINNIK, Yu. V. (1964): A characteristic property of the normal distribution," *Theor. Prob. Appl.*, 9, 624-626.