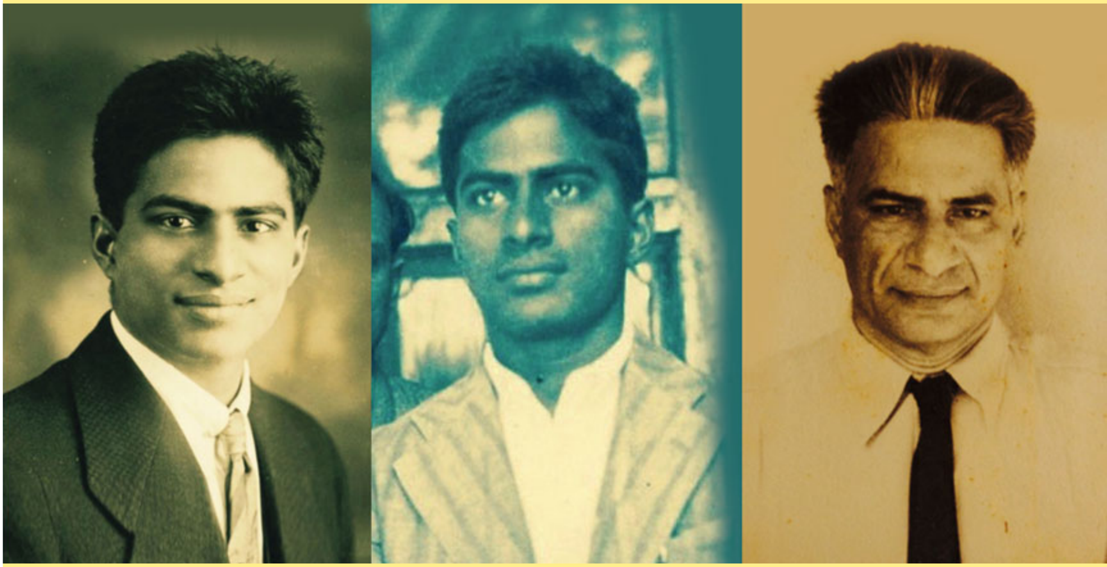
Ramakrishna Ramaswamy *Editor*

D.D. Kosambi



Selected Works in Mathematics and Statistics

Springer

D.D. Kosambi

Ramakrishna Ramaswamy
Editor

D.D. Kosambi

Selected Works in Mathematics and Statistics

Springer

Editor
Ramakrishna Ramaswamy
School of Physical Sciences
Jawaharlal Nehru University
New Delhi
India

ISBN 978-81-322-3674-0 ISBN 978-81-322-3676-4 (eBook)
DOI 10.1007/978-81-322-3676-4

Library of Congress Control Number: 2016953284

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer (India) Pvt. Ltd.
The registered company address is: 7th Floor, Vijaya Building, 17 Barakhamba Road, New Delhi 110 001, India

Preface

Damodar Dharmananda Kosambi[1] was a man of many parts: *phi beta kappa* scholar and Harvard graduate, mathematics professor, historian, archaeologist, epigraphist, polyglot, numismatist, Sanskritist, Indologist, and Marxist: the list of his identities and his personæ is a long and varied one. Over a period of a little over 35 years, Kosambi built a reputation as a major (if somewhat maverick) thinker of modern India, and this reputation has largely remained intact over the years. Widely regarded as one of the founding figures of contemporary Indian historiography, Kosambi quantified numismatics and used statistical inference to inform the study of Indian history [1]. His contributions to Indology and the study of prehistory have been fundamental, and his translations of the poetry of *Bhartrhari* [2] are considered definitive.

As it happens, while the historian, Indologist, and numismatist Kosambi has been written about and his articles and papers in those areas have been published in collections [3] and celebrated, much less has been done with regard to his contributions to mathematics and statistics. This is surprising for at least two reasons. Kosambi was first and last a mathematician in that his first independent paper and his last-known academic contribution were both in mathematics. Indeed, mathematics was the one constant and consistent preoccupation of his professional life: he says as much in the epilogue to his posthumously published autobiographical essay [4]. DDK's first paper[2] [DDK1] was written when he, then 22 years of age, was temporarily at the Banaras Hindu University in 1930, and his final work, a monograph on prime numbers [5], was submitted to publishers very shortly before his death at the age of 59, in 1966. It can be argued that his major contributions in other areas were moulded by his knowledge and style of mathematics—whether the

[1] For convenience, I will henceforth use just the surname Kosambi or the initials DDK. Other abbreviations used frequently are American Mathematical Society (AMS), Mathematical Reviews (MR), International Mathematical Union (IMU), Journal of the Indian Society of Agricultural Statistics (JISAS), Riemann hypothesis (RH), and Tata Institute of Fundamental Research (TIFR).

[2] These papers of DDK are numbered 1 through 67 and are distinguished from the other references by the initials preceding the paper number. See the bibliography on pages xv–xix.

creation of numismatics as a form of historiography through the extensive statistical analysis of large hoards of coins or his deduction of the probable location of the Karasambhale caves [6] through a combination of estimation and logic.

Most scholars who have been influenced by the historical writings of Kosambi are acquainted with a lesser extent with the nature and range of his mathematical contributions [7]. This is mainly a domain issue: as a field, mathematics and history are perceived as separated by a major cultural divide, and there is a general (and reasonable) feeling that the mathematics would be too difficult to understand by any but a trained mathematician. Ironically, Kosambi had in his lifetime experienced the same reaction from the other side—his scientist colleagues at the TIFR had also not appreciated the nature and the extent of his contributions to Indology and the study of Indian history.

Kosambi's intellectual legacy needs to be considered in its totality; the mathematics is integral to his thinking and analysis and cannot be seen as separate from the work in numismatics or, for that matter, history. DDK wrote about 65 papers that were of a mathematical or statistical nature [7]. Some articles were pedagogic expositions rather than original contributions, and some were multidisciplinary in the sense that they integrated linguistics or numismatics along with the mathematics or statistics. Two were the same work in two languages, Chinese [DDK56] and English [DDK59]. In addition, there were original contributions in German [DDK7] and French [DDK5, DDK20, DDK21, DDK42, DDK45], and one of his papers had been translated into Japanese [DDK22]. He wrote at least two mathematical monographs, but regrettably, these never appeared in print, and the manuscripts of both of them are lost. Towards the end of his life, he published two articles [DDK60, DDK64] in the Journal of the Indian Society of Agricultural Statistics that tangentially implied that he had a proof of the Riemann hypothesis. These articles contained an incomplete and flawed approach to this very fundamental mathematical problem; the damage that they caused to his reputation as a serious mathematician was irreparable and irreversible.

Details of Kosambi's professional life are well known and bear only a limited retelling [8]. On completing his BA (*summa cum laude*) at Harvard, Kosambi had, for a complex combination of reasons, to return to India in 1929. He took up a position at the Banaras Hindu University teaching mathematics and gave (optional) German classes on the side [6]. Although he started doing some research in mathematics at BHU, he was soon persuaded to move to Aligarh Muslim University to join a department of mathematics headed by the French mathematician André Weil. It was here that Kosambi first earned a place in the history of mathematics. His paper, *On a generalization of the second theorem of Bourbaki* [DDK2], was written at the provocation of Weil, as "a parodic note passed off as a serious contribution to a provincial journal" [9], the Bulletin of the Academy of Sciences, U. P. [10]. The incident remains somewhat mysterious; according to Weil, Kosambi was having problems with a colleague, and he (Weil) suggested this prank, to name a theorem after a fictitious Russian author. Whether or not this paper deflated the recalcitrant colleague's ego is not clear, but nevertheless, this paper of Kosambi marks the first occurrence of the name of *Bourbaki* in the published literature [11].

Kosambi lasted 2 years in Aligarh before moving back to Pune, to Fergusson College where he stayed until 1945. In this time, he first built up a reputation as a serious mathematician, serious enough that he was elected to the Indian Academy of Sciences by C.V. Raman in 1935 who also probably nominated him for the Ramanujan Medal of the Madras University in 1934. He had started a study of the area he termed "path–geometry" [12] that was to occupy him for several decades subsequently. *A note on the trial of Socrates* appeared in the magazine of Fergusson College in 1939, marking his initial professional foray outside mathematics. In 1940, this was followed by *The emergence of national characteristics among three Indo-European people* [13] in the Annals of the Bhandarkar Oriental Research Institute. By this time, he had also begun his careful analysis of the weights of ancient coins—the first publication on this topic also dates to 1940—and marks the start of his use of quantitative methods in historical analysis.

The years of World War II saw DDK at his creative best. Between 1939 and 1944, he published 35 articles including two papers he wrote in 1943–1944 which brought him considerable renown. One that appeared in the Journal of the Indian Mathematical Society, *Statistics in function space* [DDK36], is a method for decomposing an arbitrary signal into its significant components, a technique termed the principal value decomposition. Today, this is known as the Karhunen–Loève expansion, although both Karhunen and Loève did their work only later, in 1947 and 1948, respectively. It is regrettable that Kosambi's work was not followed up either by him or by others (although it was reviewed in Mathematical Reviews). The second contribution is in his 1944 paper in the Annals of Eugenics [DDK37]. This work in genetics, on what is termed the map distance, quantifies the genetic similarity in terms of the recombination frequency of linked genes. At the time when DDK did the work, his knowledge of genetics was probably minimal, and the structure of DNA was itself largely unknown. Nevertheless, Kosambi provided an interesting and useful method to estimate the map distances from recombination values and this work continues to be used and cited even to this day.

In 1945, DDK left Fergusson College to move to the newly established Tata Institute of Fundamental Research (TIFR) in Bombay following an invitation from the founding director, Homi J. Bhabha, to help establish a School of Mathematics. This remained his address for the next 16 years, although his increasingly meandering intellectual interests, his personal politics, his mathematical obsessions, and his personal angularities all combined to make his tenure at the TIFR a fraught one.

The relationship between Bhabha and Kosambi started off on a cordial note. Bhabha was responsible for having DDK elected president of the Mathematics Section of the Indian Science Congress that was held in Delhi in early 1947 where he gave his presidential address on "Possible applications of the functional calculus" [DDK44], a summary of his ideas on function spaces and the proper orthogonal decomposition [14]. Bhabha also helped arrange a year's visit to the USA for DDK. He gave a course of lectures on tensor analysis at the University of Chicago and also spent time at the Institute for Advanced Studies in Princeton as well as Harvard and MIT in Cambridge.

As his interests in historical analysis increased in the 1950s, DDK's mathematics inevitably slowed down. He travelled to the Soviet Union and China during this period and wrote on a variety of social issues. All these activities were at variance with the TIFR ethos; Bhabha, who was attempting to build a first-class research establishment in nuclear science and mathematics, had little time to indulge DDK in these pursuits. Towards the end of the 1950s, Kosambi started working on the Riemann hypothesis. He published two papers offering a proof of this problem, in the Indian Journal of Agricultural Statistics [DDK60, DDK64]. The motivation for his foray into this work remains unknown since his approach, a probabilistic one, does not evolve out of his earlier work. At any rate, his choice of the journal and the scale of his claim (since the Riemann hypothesis remains unproven today) exposed him to ridicule, both professionally and in person. Mathematicians who knew Kosambi speak of this phase of his life with a distinct air of embarrassment.

The relationship with Bhabha soured, and DDK's contract with the TIFR was not renewed after 1962, making Kosambi one of the very few people to have effectively been fired by the Tata Institute of Fundamental Research. Between 1962 and 1964, DDK was without a formal position although he published papers both in and outside mathematics. Peculiarly, he wrote four of these under the pseudonym S. Ducray [DDK62, DDK63, DDK65, and DDK66]. In 1964, he was appointed a CSIR emeritus professor attached to the Maharashtra Vidnyanvardhini in Pune, a position he held until his death in 1966.

There remain important gaps in writings by or on DDK that need to be filled in the order that an accurate picture of the evolution of his intellectual framework can be drawn. His extensive correspondence with Professor and Mrs. R.J. Conklin between 1930 and 1948, friends of him from his undergraduate years at Harvard, is only partly available. The TIFR correspondence is on record, and the details of the relationship with Bhabha that started out so cordially and ended in so much acrimony that DDK could not bring himself to be generous even after Bhabha died are again well enough known but incompletely analysed. A series of letters exchanged between Divyabhanusinh Chavda and DDK in his final and very bitter years remain essentially unknown. Some of these gaps are being addressed, most recently in *Unsettling the Past*, a collection of essays by and on Kosambi [15].

The present volume brings together the complete bibliography of the mathematics papers of DDK, along with other essays on and by Kosambi. This preface gives a general background, summarizing an earlier essay that was published in the Economic and Political Weekly [8]. Part I of this book contains an introductory essay, *A Scholar in his Time*, which analyses the mathematical development of Kosambi and attempts to situate his contributions in context. This is a reproduction of [16] with small modifications and is followed by selected essays by DDK that help give a perspective on the many strands of thought that he integrated into his work. The autobiographical *Adventures into the Unknown* [4] has appeared in part in several collections as *Steps in Science* [17], but the essay, *On Statistics*, is not widely known. In the war years, when Kosambi was teaching at Fergusson College in Poona in his most intellectually fertile period, he made several interesting mathematical contributions that were in part responsible for his being invited in 1945 to the newly formed Tata Institute of Fundamental Research in Bombay, to help its director, Homi J. Bhabha, to nucleate the School of Mathematics. Also around that time, he received a small grant from the Tata Trust, and the report that he submitted to them and which is reprinted here reveals a side of him that is not evident in his publications. He worked on a diverse set of problems more or less simultaneously, was meticulous in his accounts, and was frugal as well.

Reprinted in Part II are some of the most significant papers written by Kosambi between 1930 and 1964, in particular, those that contributed to his reputation as well as those that were responsible for its loss. The selection of papers and the essays that are reprinted in this book are each accompanied by an introductory paragraph. Part III contains a listing of DDK's papers in languages other than English. Three of these, in German, French, and Chinese, respectively, are reprinted. The articles that are *not* reproduced here are available at the repository of the Indian Academy of Sciences, Bangalore. Along with the personal papers of Kosambi that are now available in the Nehru Memorial Museum and Library, these various resources can only help complete the mosaic of a complex and very gifted scholar.

New Delhi, India
June 2016

Ramakrishna Ramaswamy

Note to the Reader

This volume includes both published papers in mathematics and statistics, as well as essays and commentaries. The footnotes and citations in each of these come in several styles.

- DDK's papers are listed on pages xv–xix. They are cited as [DDK1], [DDK2], etc. throughout the book.
- For biographical information, I have relied to a great extent on Chintamani Deshmukh's *Damodar Dharmanand Kosambi: Jivan ani Karya* (The life and Work of D.D. Kosambi), Mumbai: Granthali, 1993. This was first published in Marathi and subsequently translated into English by Suman Oak, and several versions are freely available online. This is referred to as [DDK-JK] where cited in the commentaries to the papers.
- For each of DDK's published papers that has been reprinted here, the references and footnotes appear within the article. Attempts have been made to remain faithful to the originals.
- References cited in the Preface are listed on the following pages. References in the essays and commentaries in Part I are collectively listed on pages 41–45.

References

1. DDK's books on history are (a) *An Introduction to the Study of Indian History* (Popular Book Depot, Bombay, 1956), (b) *Myth and Reality: Studies in the Formation of Indian Culture* (Popular Prakashail, Bombay, 1962) and (c) *The Culture and Civilisation of Ancient India in Historical Outline* (Routledge & Kegan Paul, London, 1965).
2. DDK edited the following three books on the poetry of Bhartrhari: (a) *The Satakatrayam of Bhartrhari with the Comm. of Ramarsi*, ed. by D.D. Kosambi, K.V. Krishnamoorthi Sharma (Anandasrama Sanskrit Series, No.127, Poona, 1945), (b) *The Southern Archetype of Epigrams Ascribed to Bhartrhari* (Bharatiya Vidya Series 9, Bombay, 1946) and (c) *The Epigrams Attributed to Bhartrhari* (Singhi Jain Series 23, Bombay, 1948).
3. *The Oxford India Kosambi*, ed. by B.D. Chattopadhyaya (Oxford University Press, New Delhi, 2009); *Combined Methods in Indology & Other Writings: Collected Essays*, D.D. Kosambi, Compiled, edited and introduced by Brajadulal Chattopadhyaya (Oxford University Press, New Delhi, 2005); D.D. Kosambi, *Indian Numismatics* (Orient Longman, Hyderabad, 1981); D.D. Kosambi , *Exasperating Essays* (Peoples Publishing House, New Delhi, 1957).
4. D.D. Kosambi, 'Adventure into the Unknown', in *Current Trends in Indian Philosophy*, ed. by K. Satchidananda Murty, K. Ramakrishna Rao (Asia Publishing House, Bombay, 1972).
5. D.D. Kosambi, *Prime Numbers*. The manuscript of this book, that was apparently mailed to his publishers shortly before DDK's death in June 1966, has not been traced.
6. C. Deshmukh, *Damodar Dharmanand Kosambi: Jivan ani Karya* (The life and Work of D.D. Kosambi). (Granthali, Mumbai, 1993). First published in Marathi and subsequently translated into English by Suman Oak; this book is cited as [DDK-JK].
7. The bibliography that now appears on pages xv–xix of this volume is a listing of the complete set of the papers of DDK that are of a mathematical nature. The list has been compiled in part from incomplete sources in the biography by Chintamani Deshmukh [6] as well as Web listings. In addition to the papers listed, many of his essays relate to scientific issues, but these are not included here.
8. R. Ramaswamy, Integrating Mathematics and History: The scholarship of D.D. Kosambi. Econ. Polit. Wkly. **47**, 58–62 (2012). Reproduced in [15].
9. A. Weil, *The apprenticeship of a mathematician* (Birkhäuser, Basel, 1992).
10. In the paper [DDK2], Kosambi thanks Weil for making him aware of the "important work" of this Bourbaki. The French group eventually chose the initial N (Nicolas) for Bourbaki rather than the D given by Kosambi.
11. M. Mashaal, Bourbaki: *A Secret Society of Mathematicians*, (American Mathematical Society, Providence, 2006).
12. Starting with [DDK3], Kosambi developed the idea in a number of papers, including [DDK5, DDK6, DDK8] and [DDK18] and so on. In the 1950s, he was on the editorial board of the Japanese journal, Tensor (New Series) wherein he published [DDK55], possibly his final paper on the topic.
13. The emergence of national characteristics among three Indo-European People. Ann. Bhandarkar Orient. Res. Inst. **20**, 195–206 (1940).
14. In [DDK45], Section 8, Kosambi gives the following examples of where the functional calculus techniques would apply. If average temperature curves are available for any range or period, is it possible to say whether two samples from two different places differ materially? Or do two skulls found by the archaeologist or anthropometrician in two different places differ significantly? The need for a mathematical technique to decide questions of this form is suggestive of how his interests in one area inspired work in the other.
15. M. Kosambi (ed.), *Unsettling the Past: Unknown Aspects and Scholarly Assessments of D.D. Kosambi* (Permanent Black, New Delhi, 2012).

16. R. Ramaswamy, A scholar in his time: Contemporary views of Kosambi the mathematician. Occasional Paper of the Nehru Memorial Museum and Library, Perspectives in Indian Development, New Series **45** (2014).
17. Steps in Science, in *Science and Human Progress: Essays in Honour of Late Prof. D.D. Kosambi, Scientist, Indologist, and Humanist* (Popular Prakashan, Mumbai, 1974).

Acknowledgements

A conversation with Romila Thapar at the Jawaharlal Nehru University a decade or so ago made me aware that the mathematical side of DDK was still largely inaccessible to social scientists. Given the scale of his contributions to historical research and in particular his introduction of quantitative methods into historical analysis, it seemed worthwhile to undertake to collect his contributions in mathematics and statistics in one place, much like what had been done for his historical writing. Putting together a complete bibliography took a little more time than I had anticipated, largely due to the fact that some of the journals that Kosambi published in were not digitized (and some still are not). Nevertheless, with a little help from friends who helped me get the more obscure articles, this was done by 2013.

An invitation from the Nehru Memorial Museum and Library to speak on Kosambi's mathematical contributions gave me the opportunity to put together the material for the essay *A Scholar in his Time* that is the basis of Chap. 1 of this book. DDK's essays that are reprinted as Chaps. 2 and 3 have been published earlier but are still incompletely known and especially in a volume of this sort are worth recalling. I would like to thank the NCRA Library and the TIFR Archives for access to digital versions of DDK's papers. I would also like to thank the Indian Academy of Sciences, the Current Science Association, the Indian Association for the Cultivation of Science, the National Academy of Sciences, India, John Wiley and Sons, the American Mathematical Association, the Indian Mathematical Society, and the Indian Society of Agricultural Statistics for permission to reprint DDK's articles from their journals.

A number of people have helped along the way, and it is my pleasure to thank them all. I have greatly benefited from conversations and/or correspondence with Michael Berry, Divyabhanusinh Chavda, Indira Chowdhury, Shrikrishna G. Dani, Louise Morse, Aban Mukherji, Rajaram Nityananda, Andrew Odlyzko, Oindrila Raychaudhuri, Toshio Yamazaki, and, particularly, Romila Thapar. The mathematical articles were retyped in LaTeX by Mr. Srinivas of the University of Hyderabad in order to make the text more uniform, and Cicilia Edwin of the Indian Academy of Sciences, Bangalore, assisted with the proofreading. Time has not been

kind, not only to DDK's mathematics but also to the various journals in which he published: some of the articles are barely legible and quite difficult to read in the original.

In large part, this book was a joint enterprise with Meera Kosambi who offered considerable help, encouragement, and suggestions in the 5 years during which I grew to know her well. Her death in February 2015 was a great loss, and in her absence, this project feels oddly incomplete. In retrospect though, it seems she knew her time was limited; in her last years, she was anxious to consolidate the intellectual legacies of her grandfather and father in books she wrote and edited.

My family has been greatly supportive over the years, indulging my various preoccupations and obsessions with patience and with grace. My wife Charusita passed away earlier this year, and it is a great personal sadness that she did not live to see this book in its final form. I know she felt that the effort invested in this project was worth the while, and I hope that my children, Krithi and Rohan, will feel the same way.

New Delhi, India Ramakrishna Ramaswamy
June 2016

D.D. Kosambi's Mathematical and Scientific Publications

Given below is a chronology of D.D. Kosambi's articles in different areas of mathematics, statistics, and science. In addition, he wrote two monographs, neither of which were eventually published. The first, a book on *Path Geometry*, was to have been published in the *Annals of Mathematics Studies*, a series edited by Marston Morse at the Institute for Advanced Studies, Princeton. The manuscript sent to Morse and a copy that was given to Homi Bhabha could not be traced. In 2010, when Louise J. (Mrs. Marston) Morse was nearly 100 years old, the Morse Archives were searched one last time. However, it was not possible to locate this manuscript or any reference to it. Shortly before his death, Kosambi mailed the manuscript of another book on *Prime Numbers* to his publishers, Routledge & Kegan Paul. Unfortunately, this was also lost.

The papers that were reviewed in *Mathematical Reviews* are indicated, along with the corresponding MR number and the name of the reviewer where applicable. Some of these "reviews" merely record the publication of the paper and do not offer a serious commentary on the work. The titles of papers that have been reprinted in this collection are in boldface.

1. **Precessions of an elliptical orbit**,
   Indian Journal of Physics **5**, 359–64 (1930)
2. **On a generalization of the second theorem of Bourbaki**,
   Bulletin of the Academy of Sciences, U. P. **1**, 145–47 (1931)
3. *Modern differential geometries*,
   Indian Journal of Physics **7**, 159–64 (1932)
4. *On differential equations with the group property*,
   Journal of the Indian Mathematical Society **19**, 215–19 (1932)
5. *Geometrie differentielle et calcul des variations*,
   Rendiconti della Reale Accademia Nazionale dei Lincei **16**, 410–15 (1932)
6. *On the existence of a metric and the inverse variational problem*,
   Bulletin of the Academy of Sciences, U. P. **2**, 17–28 (1932)
7. **Affin-geometrische Grundlagen der Einheitlichen Feldtheorie**,
   Sitzungsberichten der Preussische Akademie der Wissenschaften,
   Physikalisch-mathematische klasse **28**, 342–45 (1932)

8. **Parallelism and path-spaces**,
Mathematische Zeitschrift **37**, 608–18 (1933)
Review: MR1545422.
The above paper was followed by an extract from the correspondence between É. Cartan and DDK.
Observations sur le memoire precedent: *Extrait d'une lettre à M. D. D. Kosambi.*,
Mathematische Zeitschrift **37**, 619–22 (1933)
Review: MR1545423. The review attributes authorship of the paper to both Cartan and Kosambi
9. *The problem of differential invariants*,
Journal of the Indian Mathematical Society **20**, 185–88 (1933)
10. *The classification of integers*,
Journal of the University of Bombay **2**, 18–20 (1933)
11. *Collineations in path-space*,
Journal of the Indian Mathematical Society **1**, 68–72 (1934)
12. *Continuous groups and two theorems of Euler*,
The Mathematics Student **2**, 94–100 (1934)
13. *The maximum modulus theorem*,
Journal of the University of Bombay **3**, 11–12 (1934)
14. *Homogeneous metrics*,
Proceedings of the Indian Academy of Sciences **1**, 952–54 (1935)
15. *An affine calculus of variations*,
Proceedings of the Indian Academy of Sciences **2**, 333–35 (1935)
16. Systems of differential equations of the second order,
Quarterly Journal of Mathematics (Oxford) **6**, 1–12 (1935)
17. *Differential geometry of the Laplace equation*,
Journal of the Indian Mathematical Society **2**, 141–43 (1936)
18. *Path-spaces of higher order*,
Quarterly Journal of Mathematics (Oxford) **7**, 97–104 (1936)
19. *Path-spaces of higher order*,
Quarterly Journal of Mathematics (Oxford) **7**, 97–104 (1936)
20. *Les metriques homogenes dans les espaces cosmogoniques*,
Comptes Rendus **206**, 1086–88 (1938)
21. **Les espaces des paths generalises qu'on peut associer avec un espace de Finsler**,
Comptes Rendus **206**, 1538–41 (1938)
22. **The tensor analysis of partial differential equations**,
Journal of the Indian Mathematical Society **3**, 249–53 (1939)
Review: MR0001882 (1, 313f) Reviewer: E. W. Titt.
A Japanese translation of this paper appeared in Tensor, **2**, 36–39 (1939)
Review: MR0001075 (1,176c) Reviewer: A. Kawaguchi.
23. **A statistical study of the weights of the old Indian punch-marked coins**,
Current Science **9**, 312–14 (1940)

24. *On the weights of old Indian punch-marked coins*,
Current Science **9**, 410–11 (1940)
25. *Path-equations admitting the Lorentz group*,
Journal of the London Mathematical Society **15**, 86–91 (1940)
Review: MR0002258 (2, 21f) Reviewer: J. L. Vanderslice.
26. *The concept of isotropy in generalized path-spaces*,
Journal of the Indian Mathematical Society **4**, 80–88 (1940)
Review: MR0003125 (2,166g) Reviewer: J. L. Vanderslice.
27. *A note on frequency distribution in series*,
The Mathematics Student **8**, 151–55 (1940)
Review: MR0005390 (3,147h).
28. **A bivariate extension of Fisher's Z–test**,
Current Science **10**, 191–92 (1941)
Review: MR0005589 (3,175h) Reviewer: A. Wald.
29. *Correlation and time series*,
Current Science **10**, 372–74 (1941)
Review: MR0005590 (3,175i) Reviewer: A. Wald.
30. *Path-equations admitting the Lorentz group–II*,
Journal of the Indian Mathematical Society **5**, 62–72 (1941)
Review: MR0005713 (3,192g) Reviewer: J. L. Vanderslice.
31. *On the origin and development of silver coinage in India*,
Current Science **10**, 395–400 (1941)
32. *On the zeros and closure of orthogonal functions*,
Journal of the Indian Mathematical Society **6**, 16–24 (1942)
Review: MR0006770 (4, 39d) Reviewer: E. S. Pondiczery.
33. **The effect of circulation upon the weight of metallic currency**,
Current Science **11**, 227–31 (1942)
34. **A test of significance for multiple observations**,
Current Science **11**, 271–74 (1942)
Review: MR0007235 (4,107b) Reviewer: A. Wald.
35. **On valid tests of linguistic hypotheses**,
New Indian Antiquary **5**, 21–24 (1942)
Review: MR0007247 (4,109a) Reviewer: A. Wald.
36. **Statistics in function space**,
Journal of the Indian Mathematical Society **7**, 76–88 (1943)
Review: MR0009816 (5, 207c) Reviewer: J. L. Doob.
37. **The estimation of map distance from recombination values**,
Annals of Eugenics **12**, 172–75 (1944)
38. *Direct derivation of Balmer spectra*,
Current Science **13**, 71–72 (1944)
39. **The geometric method in mathematical statistics**,
American Mathematical Monthly **51**, 382–89 (1944)
Review: MR0010937 (6, 91c) Reviewer: R. L. Anderson.

40. *Parallelism in the tensor analysis of partial differential equations*,
Bulletin of the American Mathematical Society **51**, 293–96 (1945)
Review: MR0011793 (6, 217e) Reviewer: J. L. Vanderslice.
41. **The law of large numbers**,
The Mathematics Student **14**, 14–19 (1946)
Review: MR0023471 (9, 360i) Reviewer: W. Feller.
42. *Sur la differentiation covariante*,
Comptes Rendus **222**, 211–13 (1946)
Review: MR0015274 (7, 396b) Reviewer: J. L. Vanderslice.
43. *An extension of the least–squares method for statistical estimation*,
Annals of Eugenics **18**, 257–61 (1947)
Review: MR0021290 (9, 49d) Reviewer: J. Wolfowitz.
44. **Possible Applications of the Functional Calculus**,
Proceedings of the 34th Indian Science Congress. Part II: Presidential Addresses, 1–13 (1947)
45. *Les invariants differentiels d'un tenseur covariant a deux indices*,
Comptes Rendus **225**, 790–92 (1947)
Review: MR0022433 (9, 207b) Reviewer: N. Coburn.
46. *Systems of partial differential equations of the second order*,
Quarterly Journal of Mathematics (Oxford) **19**, 204–19 (1948)
Review: MR0028514 (10, 458d) Reviewer: M. Janet.
47. *Characteristic properties of series distributions*,
Proceedings of the National Institute of Science of India **15**, 109–13 (1949)
Review: MR0030731 (11, 42h) Reviewer: J. L. Doob.
48. **Lie rings in path-space**,
Proceedings of the National Academy of Sciences (USA) **35**, 389–94 (1949)
Review: MR0030807 (11, 56a) Reviewer: O. Varga.
49. *The differential invariants of a two-index tensor*,
Bulletin of the American Mathematical Society **55**, 90–94 (1949)
Review: MR0028653 (10, 480b) Reviewer: V. Hlavatý.
50. *Series expansions of continuous groups*,
Quarterly Journal of Mathematics (Oxford, 2) **2**, 244–57 (1951)
Review: MR0045732 (13, 624b) Reviewer: M. S. Knebelman.
51. (with S. Raghavachari) *Seasonal variations in the Indian birth–rate*,
Annals of Eugenics **16**, 165–92 (1951)
Review: MR0046135 (13, 691b) Reviewer: R. P. Boas, Jr.
52. *Path-spaces admitting collineations*,
Quarterly Journal of Mathematics (Oxford, 2) **3**, 1–11 (1952)
Review: MR0047387 (13, 870d) Reviewer: O. Varga.
53. *Path-geometry and continuous groups*,
Quarterly Journal of Mathematics (Oxford, 2) **3**, 307–20 (1952)
Review: MR0051562 (14, 498g) Reviewer: A. Nijenhuis.
54. (with S. Raghavachari) *Seasonal variations in the Indian death–rate*,
Annals of Human Genetics **19**, 100–19 (1954)

55. *The metric in path-space*,
Tensor (New Series) **3**, 67–74 (1954)
Review: MR0061869 (15, 898a) Reviewer: J. A. Schouten.
56. **The method of least–squares**, (in Chinese)
Advancement in Mathematics **3**, 485–491 (1957)
Review: MR0100960 (20 #7385).
57. *Classical Tauberian theorems*,
Journal of the Indian Society of Agricultural Statistics **10**, 141–49 (1958)
Review: MR0118997 (22 #9766) Reviewer: J. Korevaar.
58. (with U. V. R. Rao) *The efficiency of randomization by card–shuffling*,
Journal of the Royal Statistics Society **121**, 223–33 (1958)
59. **The method of least–squares**,
Journal of the Indian Society of Agricultural Statistics **11**, 49–57 (1959)
Review: MR0114265 (22 #5089) Reviewer: R. G. Laha.
60. **An application of stochastic convergence**,
Journal of the Indian Society of Agricultural Statistics **11**, 58–72 (1959)
Review: MR0122792 (23 #A126) Reviewer: W. J. LeVeque.
61. **The sampling distribution of primes**,
Proceedings of the National Academy of Sciences (USA) **49**, 20–23 (1963)
Review: MR0146168 (26 #3690) Reviewer: J. B. Kelly.
62. (as S. Ducray) *A note on prime numbers*,
Journal of the University of Bombay **31**, 1–4 (1962)
63. (as S. Ducray) *Normal Sequences*,
Journal of the University of Bombay **32**, 49–53 (1963)
Review: MR0197433 (33 #5598) Reviewer: B. Volkmann.
64. **Statistical methods in number theory**,
Journal of the Indian Society of Agricultural Statistics **16**, 126–35 (1964)
Review: MR0217024 (36 #119) Reviewer: A. Rényi.
65. (as S. Ducray) **Probability and prime numbers**,
Proceedings of the Indian Academy of Sciences **60**, 159–64 (1964)
Review: MR0179148 (31 #3399) Reviewer: J. Kubilius.
66. (as S. Ducray) *The sequence of primes*,
Proceedings of the Indian Academy of Sciences **62**, 145–49 (1965)
67. *Scientific numismatics*,
Scientific American, February 1966, pages 102–11.

Contents

Part III Select Publications of D.D. Kosambi in Other Languages

About the Editor

Ramakrishna Ramaswamy is a professor in the School of Physical Sciences as well as in the School of Computational and Integrative Sciences at the Jawaharlal Nehru University, New Delhi. He served as the vice chancellor of the University of Hyderabad during 2011–2015. Professor Ramaswamy's main research interests are in nonlinear science and systems and computational biology. He has been interested in Kosambi's life and works over the past decade and in early 2016 edited a book of essays by D.D. Kosambi, *Adventures into the Unknown* (Three Essays Collective, Gurgaon, 2016).

Part I
Essays on and by D.D. Kosambi

Chapter 1
A Scholar in His Time

"Kosambi introduced a new method into historical scholarship, essentially by application of modern mathematics." Bernal [1], who shared some of his interests and much of his politics, summarized the unique talents of DDK in an obituary that appeared in the journal *Nature (London)*, adding, "Indians were not themselves historians; they left few documents and never gave dates. One thing the Indians of all periods did leave behind, however, were hoards of coins. [...] By statistical study of the weights of the coins, Kosambi was able to establish the amount of time that had elapsed while they were in circulation ..."

Today, the significance of Kosambi's mathematical contributions [2] tends to be underplayed, given the impact of his scholarship as historian and Indologist. His work in the latter areas have been collected in several volumes [3] and critical commentaries have appeared over the years [4], but his work in mathematics has not been compiled and reviewed to the same extent [5–8]. Indeed, a complete bibliography is not available so far in the public domain [2]. This asymmetry is unfortunate since, as commented elsewhere [5], an understanding of Kosambi the historian can only be enhanced by an appreciation of Kosambi the mathematician.

There are several contributions that he is known for, some of which like the Kosambi–Cartan–Chern (KCC) theory [9] carry his name, and some like the Karhunen–Loève expansion [10] that do not. The Kosambi mapping function in genetics [DDK37] continues to be used to this day [11], but the path geometry that he studied for much of his life [12] has not found further application. DDK's final years were mired in controversies, both personal and professional. His papers on the Riemann hypothesis (RH) [DDK60, DDK64] brought him a great deal of criticism and not a little ridicule, while his personal politics put him in direct conflict with Homi Bhabha and the Department of Atomic Energy (DAE). These contributed to his eventual and somewhat ignominious ouster from employment at the Tata Institute of Fundamental Research. Although it was a contrary position to hold at the TIFR at that time, his early and passionate advocacy of solar energy was practical and based on sound scientific common sense. In some of his arguments, he seems

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_1

even somewhat Gandhian and indeed, the essential validity of Kosambi's argument remains to this day [13].

DDK was just about 23 years old when he returned to India and took up a temporary position at Banaras Hindu University with a BA (*summa cum laude*) from Harvard. A year later he had moved to the Aligarh Muslim University where he was appointed in the Mathematics Department at the suggestion of Weil [14] who was then already well known as a mathematician and as a prodigy, and who had been invited to the AMU by Syed Ross Masood, the vice chancellor at the time.

Although Weil did not last long in Aligarh, his influence on Kosambi was considerable. In addition to giving him the position and encouraging him on the matter of the Bourbaki prank, Weil helped DDK forge early mathematical links with, among others, Vijayaraghavan [15] and Chowla [16]. He undoubtedly influenced his taste on mathematics, possibly sparking DDK's interest in the Riemann hypothesis. Weil would, in the early 1940s, make important contributions to this field [17], although when DDK turned to it almost thirty years later [DDK60] his efforts were to come a cropper. Weil spent the summer of 1931 in Europe and upon his return to Aligarh, he found that not only had his own position been compromised, but the group of mathematicians that he had put together had also fragmented, with Vijayaraghavan having moved to take up a professorship in Dacca [18]. By early 1932, Weil had returned to Europe, and DDK was to leave Aligarh soon thereafter.

Kosambi started his independent work in Aligarh, choosing the area of *path-geometry*, a term he coined, submitting his papers to leading European journals [DDK3, DDK5, DDK7, DDK8]. The paper that was sent to Mathematische Zeitschrift was also communicated to Élie Cartan who was inspired enough by the result to write a detailed commentary that was published [DDK8] as a note immediately following DDK's paper in 1933. Along with a later paper by the Chinese mathematician, Chern [19], these three works constitute what is now termed the KCC-theory, a topic that has, in recent years, found new applications in physics and biology [9]. (Some years later, in 1946, Kosambi tried to have Chern invited to visit India when he was at the TIFR but nothing came of it given the complexity of the political situation both in India and in China at that time [20].) DDK wrote many papers on path geometry, and in the mid-1940s summarized his work in a manuscript that was submitted to Marston Morse at the Institute for Advanced Study in Princeton. In a letter [21] to Bhabha he says, "The book on path geometry will, according to a letter from Morse, appear in the Annals of Mathematics Studies, Princeton." This book was never published—indeed very few books in this series were published in the post-war years between 1945 and 1948. Efforts to locate a copy of the manuscript in the Morse archives have proved fruitless [22]. DDK makes reference to a second copy of the manuscript that he gave to Bhabha, but that copy has not been located either.

The Nobel laureate, C.V. Raman, had visited Aligarh in 1931 as member of a selection committee, and although there is no specific record of his having met Kosambi, his subsequent actions suggest that he quickly gathered, either directly or indirectly, a very high opinion of DDK. In 1934 when Raman founded the Indian Academy of Sciences in Bangalore, he elected Vijayaraghavan and Chowla. The very next year Kosambi was elected to the IASc at the age of 28, when his mathematical œuvre was slight, and along with others such as P.C. Mahalanobis and V.V. Narlikar. Kosambi was one of the younger of the Founding or Foundation Fellows (namely those elected in 1934 and 1935). Since the initial election to the Academy was almost entirely his decision, the estimation that Raman had of Kosambi's scholarship or of his potential must have been considerable. It is possible that Vijayaraghavan may have played some role in this early recognition [23], and it is also likely that the award of the first Ramanujan Prize of the Madras University in 1934 to S. Chandrashekar, S. Chowla and DDK [24] would have favourably impressed Raman. As it happened, in later years, Kosambi was privately and publicly very critical of Raman's style of functioning [25].

This early recognition, however, stood him in good stead. He published a couple of papers in the Academy journal, Proceedings of the Indian Academy of Sciences in 1935 (and not again until the 1960s when, as S. Ducray, he published two more). Reviews of his papers in other journals began to appear in *Current Science*, the general science journal started by Raman, in addition to original articles that he chose to publish in this journal as well. Indeed, his initial papers on the quantitative approach to numismatics [DDK23, DDK24, DDK31, DDK33] all appeared in *Current Science*.

1.1 Reviews and Commentaries

One of the early references to the work of DDK on numismatics that was brought to the attention of readers of *Current Science* was a review in 1941 [26] by K.A.N. (this was probably the well-known historian K.A. Nilakantha Sastry) of two papers by Kosambi in the *New Indian Antiquary* [27]. By this time, DDK seems to have been well established as an eminent mathematician. While generally admiring of the work, KAN comments on a number of DDK's characteristics: the use of "hard phrases" in his critique of the methods used by others, his exposure "of the hollowness of much pseudo-expertise that has held the field," etc. Nevertheless, the review is not uniformly accepting of DDKs conclusions, and KAN does alert the reader to the potential areas of inaccuracy. In a charming final paragraph, for instance he says "Yet, this conclusion hardly tallies with the impressions of the Mauryan epoch gathered from other sources such as the inscriptions of Asoka, or the polished stone pillars-not to speak of Megasthenes and the *Arthasastra*. There are other statements, *obiter dicta*, which may surprise the reader and even shock him; but there is much, very much in these papers and their method for which he will be grateful."

The journal *Mathematical Reviews* (MR) was started in 1940 by the American Mathematical Society as a way for working mathematicians to keep up with the increasing numbers of papers that appeared each year in diverse journals. The practice was (and still is) to have a brief summary of these papers sometimes with commentary, and sometimes without. Indeed, some papers are merely noted or abstracted, and all reviews are signed.

Of DDK's sixty or so papers in mathematics, about half were reviewed in MR; these are indicated in the bibliography on pages xv–xix. The reviewers include R.L. Anderson, R.P. Boas, Jr., N. Coburn, J.L. Doob, W. Feller, V. Hlavatý, M. Janet, A. Kawaguchi, J.B. Kelly, M.S. Knebelman, J. Korevaar, J. Kubilius, R.G. Laha, W.J. LeVeque, A. Nijenhuis, E.S. Pondiczery (a pseudonym of R.P. Boas Jr), A. Rényi, J.A. Schouten, E.W. Titt, J.L. Vanderslice, O. Varga, B. Volkmann, A. Wald, and J. Wolfowitz.

Several of these reviews are just summaries of the papers, but some are serious commentaries on the work of Kosambi, and, significantly, are by some of the leading contemporaneous mathematicians, probabilists, and statisticians. Indeed, R. P. Boas Jr. who reviewed some of the papers was one of the main editors of *Mathematical Reviews*.

It may be pertinent to note that it is not just DDK's papers that were published in journals outside India that were reviewed in *Mathematical Reviews*; several of the papers published in Indian journals were also commented upon critically. These include the important paper, "Statistics in function space" [DDK36] on which Doob remarks:

> The author discusses statistical problems connected with continuous stochastic processes whose representative functions $x(t)$ are defined by $x(t) = \sum_j x_j \phi_j(t)$, where the ϕ_j determine an orthonormal set and $x_1, x_2, \ldots$ are mutually independent Gaussian chance variables with vanishing means and variances $\sigma_{12}, \sigma_{22}, \ldots$, respectively. It is supposed that $\sum_j \sigma_j^4 < \infty$ and that $K(s,t) = \sum_j \sigma_j^2 \phi_j(s)\phi_j(t)$ defines a continuous integral operator. The process is determined completely by the function K. The samples he considers are functions $x(t)$ rather than merely the values of functions $x(t)$ at a finite number of points. An estimate of the function K is given in terms of a sample of n functions $x(t)$. Various mechanical and electrical methods are suggested for combining functions $x(t)$, given graphically, as necessitated by this type of statistical approach.

This paper was reviewed soon after it was published in 1943. Unfortunately, neither Karhunen nor Loève who essentially rediscovered these results [10] were aware either of the paper or of its review by Doob. In his autobiographical note [28], DDK noted that the lack of computational power had precluded the "effective use" of the results of this paper.

Another of DDK's reviewer's was Abraham Wald (who was later to die in a plane crash in India when he was visiting the country at the invitation of the Indian government) who commented, generally favourably, on four of his papers. What is interesting is that many of the papers were published in journals such as Mathematics Student and the Journal of the Indian Mathematical Society, both of limited circulation, and which to this day remain somewhat difficult to locate.

It should be mentioned that most of DDK's publications in mathematics are independently authored. He did, however, mentor several students, both formally and informally at the TIFR in the 1950s, and among these were S. Raghavachari and U. V. Ramamohana Rao who are his only coauthors [DDK51, DDK54, DDK58].

1.2 The RH Papers

Arguably, the most important as yet unresolved problem in pure mathematics is a hypothesis that was enunciated in 1857 by the celebrated mathematician, Bernhard Riemann. A brief introduction to the nature of the mathematical problem [29] is included here to give some flavour of why it is interesting and a challenge.

The function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s},$$

which is defined and finite for all real $s > 1$ can be extended uniquely, in a natural way, to the whole complex plane, by a process termed analytic continuation; the resulting function is the Riemann zeta function. Recall that when s is a *complex* number, it can be written in the form $s = \alpha + i\beta$, where α is the real part, β the imaginary part, and i is the "square root" of -1. The Riemann zeta function is defined for all complex numbers, and its value is finite except for a collection of isolated points in the plane. The hypothesis concerns the *zeros* of the zeta function, namely those values of s where $\zeta(s) = 0$.

The value of the ζ-function for specific s is a number that can be calculated explicitly in many cases. When s is 1, the sum becomes infinitely large (the $\rightarrow$ in the equation below signifies "tends to"),

$$\zeta(1) = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots \rightarrow \infty;$$

which is the so-called harmonic series. When s equals 2, the sum converges to a finite value,

$$\zeta(2) = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots \rightarrow \frac{\pi^2}{6}.$$

In order to study the questions involved, it is necessary to consider the function $\zeta(s)$ for all s in the complex plane, namely for all values of α and β. The zeta function can also then take values among complex numbers. It turns out that $\zeta(s) = 0$ when s is a negative even integer, namely $\alpha = -2, -4, -6$, and so on, with the imaginary part β being 0. These zeros of the zeta function are termed trivial since the proof is based on a straightforward procedure [30].

The ζ function has in addition an infinite number of "nontrivial" zeros, and Riemann's hypothesis is that for *all* of these, α (namely the real part of s) has the value 1/2. In the complex plane, these zeros therefore all lie on the so-called "critical" line, $\alpha = 1/2$. While being simple enough to state, it remains unproven to this day. Because of connections between the zeta function and prime numbers, a proof of the RH would have significant implications for the distribution of prime numbers, and via this, to essentially all of mathematics.

DDK's mathematical reputation suffered greatly as a result of two papers he published in the Journal of the Indian Society of Agricultural Statistics, on the RH

[DDK60, DDK64]. Notwithstanding its name, the journal does publish serious mathematics, particularly in the area of probability. Although obscure and highly specialized, the journal may not have been as inappropriate for the papers as might appear, since the methods suggested by Kosambi were probabilistic. However, the lack of appropriate refereeing was a real deficiency, and the charge remains that DDK chose to publish the papers in JISAS to be able to pass off a doubtful "proof."

Both papers were reviewed in MR, one by W. J. LeVeque, a number theorist who eventually became executive director of the American Mathematical Society. His critique of "An application of stochastic convergence" [DDK60] goes straight to the point that the claim made by DDK, that

$$\lim_{x\to\infty}\left(\sum_{n\le x} n^{-\sigma} - \sum_{p\le x} p^{-\sigma}\log p\right)$$

exists and is finite for real $\sigma > 1$ is *a result which easily implies the Riemann hypothesis*. However, since the proof is probabilistic in nature, there are major problems that he identifies. "Of the two proofs given for the crucial Lemma 1.2, the reviewer does not understand the first, which seems to involve more "hand-waving" than is customarily accepted even in proofs of theorems less significant than the present one. The second proof appears to be erroneous." The review concludes "The reviewer is unable either to accept this proof or to refute it conclusively. The author must replace verbal descriptions, qualitative comparisons and intuition by precise definitions, equations and inequalities, and rigorous reasoning, if he is to claim to have proved a theorem of the magnitude of the Riemann hypothesis."

The kindest analysis of these works of DDK comes from the Hungarian mathematician, A. Rényi, who says in a posthumous review of the paper "Statistical methods in number theory" [DDK64] that

> The late author tried in the last 10 years of his life to prove the Riemann hypothesis by probabilistic methods. Though he did not succeed in this, he has formulated the following highly interesting conjecture on prime numbers. Put $\mathrm{li}(x,a) = \int_a^x (1/\log t)dt$ for $x \ge a$, $2 \le a < 3$. Let $p_1 < p_2 < \cdots < p_n < \cdots$ denote the sequence of odd primes and consider the numbers $q_n(a) = \mathrm{li}(p_n, a)(n = 1, 2, \ldots)$. Clearly $q_n(a) \sim n$ by the prime number theorem. Let $\pi_n(u,a)(n = 1, 2, \ldots)$ denote the number of points $q_r(a)(r = 1, 2, \ldots)$ lying in the interval $[(n-1)u, nu)$, where $u > 0$. Let $V_k(N,u,a)$ denote the number of those values of $n \le N$ for which $\pi_n(u) = k$. The author's conjecture states that one can choose the values of a and u in such a way that $\lim_{N\to\infty} V_k(N,u,a)/N = u^k e^{-u}/k!(k = 0, 1, \ldots)$. In other words, his conjecture states that the points $q_n(a)$ are distributed as the points in a typical realization of a homogeneous Poisson process with density 1.

Rényi, who had been sent both this and the earlier papers [DDK60, DDK61] prior to publication, goes on to say that "Neither in this paper nor in his previous paper [Proc. Nat. Acad. Sci. U.S.A. 49 (1963), 20–23; MR0146168 (26 #3690)] did the author succeed in proving his hypothesis, nor in deducing from it the Riemann hypothesis." The PNAS paper [DDK61] was reviewed by J. B. Kelley who states, after summarizing the main result, that "The exposition is rather sketchy; in particular, the reviewer could not follow the proof of the crucial Lemma 4."

Either because of the timing of the review or because he may have appreciated the valiant attempts of DDK to prove the Riemann hypothesis by an unusual route, Rényi concludes the review by saying that at that point in time (1968) "one does not have enough knowledge of the fine structure of the distribution of primes to prove or disprove the author's conjecture. The problem seems to be even more difficult than the problem of the validity of the Riemann hypothesis. As a matter of fact, no obvious method exists to prove the author's hypothesis even under the assumption of the Riemann hypothesis. Nevertheless, the conjecture is worthy of study in its own right, and the reviewer proposes to call it *the Kosambi hypothesis* in commemoration of the enthusiastic efforts of the late author."

Rényi's suggestion has not found favour. The probabilistic approach has inherent limitations, but also as these reviews suggest, the rigour emphasized by DDK in his early years had deserted him. What is somewhat surprising is that there are errors in these papers that become evident even with a fairly cursory examination [31] and which could have been detected by an alert referee. The fact that IJSAS published this paper with the errors added to the feeling that DDK deliberately chose the journal to avoid qualified peer review. DDK's mathematical reputation was essentially destroyed by these papers.

Given the continual interest in the RH, only in part increased now by its inclusion as a Millennium Prize problem, there are a number of popular books [32] that summarize the approaches to proving it. Not surprisingly, the work of DDK is not mentioned in any of these. In private correspondence, the mathematical physicist Sir Michael Berry (who has had an abiding interest in the problem) remarks that DDK's "idea for proving RH based on showing that a certain function is nonsingular off the line, is ingenious." Andrew Odlyzko, another mathematician who has worked extensively on the RH and who, even as a graduate student, was aware of DDK's work says [33] that he "was really intrigued by these approaches, but after a while decided that it would take some clever insights far beyond what [he] could think of to accomplish anything rigorous in this area." Among Odlyzko's major contributions to a study of the RH is the computation of a very large number of the zeros to high precision, and for *all* of these, the real part equals 1/2. As an experimental mathematician, he has a good insight into the approach suggested by DDK, adding, "in summary, I think it is a pity that Kosambi did not see the flaws in his arguments and published this paper, but the basic idea is an interesting one, and certainly worth exploring. I would be surprised, but not shocked, if somebody clever managed to do something with it."

1.3 Bhabha and DDK

DDK joined the newly formed Tata Institute of Fundamental Research on 16 June 1945. His appointment, which was for an initial period of five years, was decided at the first meeting of the provisional council of TIFR [34].

The initial correspondence between Bhabha and DDK, although formal, was extremely cordial [35]. In 1946, when Bhabha travelled to England, he appointed

DDK Acting Director, leaving him in charge of the fledgling institute. This was a position of considerable responsibility, and one that DDK clearly enjoyed, and in a long letter [21] written on 8th July he writes, “About building up a School of Mathematics in India, we also think alike; but, as you are fully aware, we have to get people trained in a considerable number of branches for which there are no real specialists in this country.”

The relationship also grew warm, especially since they had to plan the Institute together, concerning themselves with details regarding land acquisition, equipping the laboratories, hiring staff, and planning for the future. That same year DDK was elected Fellow of the Indian National Science Academy, and the next, in 1947, was awarded the Bhabha Prize (named for Bhabha’s father, Jehangir Hormusji Bhabha). He was also chosen the president of the Mathematics section of the 34th Indian Science Congress that was held in Delhi in December 1947 [DDK44] with the active support of Bhabha who also realized that this would bring DDK into contact with Nehru. Kosambi’s mathematical and statistical expertise was also greatly appreciated in the new institute—a number of colleagues, Bhabha among them, acknowledge his advice and help explicitly in their scientific publications. And outside TIFR, the Ministry of Defence sought DDK’s advice on cryptography [35, 36]!

In 1948, when DDK was to go to the USA for a year’s visit, to Chicago and Princeton, Bhabha threw a party for him at his residence in Malabar Hill. This visit was in fact largely arranged by Bhabha, and among other things, DDK was to investigate the possibility of getting a computing machine for the new institute [21] as well as to attract new faculty, K. Chandrasekharan and S. Minakshisundaram in particular. On this trip, he pursued all aspects of his wide-ranging interests, visiting Einstein and von Neumann in Princeton, Norbert Wiener in Boston, as well as the historian, A.L. Basham in London. In Chicago, he was visiting professor at the University, where he gave a course of 36 lectures on tensor analysis. This was a special interest of his: he had been invited to the editorial board of the Hokkaido University journal, Tensor (New Series), and indeed an article of his had been translated into Japanese already in 1939 by the same journal [DDK22].

The position at TIFR gave DDK national prominence as well. In 1950, when the International Mathematical Union was being revived, DDK was effectively asked to head the National Committee on the proposed IMU [37]. Travel money was difficult to come by, so DDK was unable to travel to the USA, and India was represented at the Union Conference in New York [38] by S. Minakshisundaram (of Andhra University, Waltair, but who was already in the USA) and K. Chandrasekharan, who had joined the TIFR by then. Shortly afterwards, K.G. Ramanathan who had obtained his Ph.D. at Princeton also moved to TIFR; Chandrasekharan and he would subsequently play a much more influential role in shaping the TIFR School of Mathematics than Kosambi. Cracks began to surface in the relationship between Bhabha and DDK in the next few years, first in regard to students and then gradually with regard to details such as his attendance in office and other aspects of his working.

The spiral downwards, though, began in 1959 with the publication of the JISAS paper [DDK60] and the subsequent grand obsession with a probabilistic proof of the Riemann hypothesis. His differences became more pronounced with Bhabha who relied more and more on Chandrasekharan's opinion and estimation of DDK's work. The *coup–de–grace* was a letter signed by four of the mathematicians at TIFR stating that Kosambi had become an embarrassment to the Institute with his claim of the proof of the RH and of Fermat's Last Theorem [39] that was being broadcast internationally.

There were other differences with Bhabha which were of a political nature, but these differences were already present in 1945 when Bhabha invited DDK to join TIFR. The largely unknown essay "An Introduction To Lectures On Dialectical Materialism" summarizes a set of 15 lectures given by Kosambi in Poona in 1943 [40]. Later, after he had joined the TIFR, when he gave a lecture in Bombay House, the headquarters of the Tata Group, the notes conclude with an appreciation of Lenin [41]. Indeed, Bhabha facilitated DDK's visits to the Soviet Union and China, and it is not possible that DDK's views were hidden under a bushel until the early 1960s [42].

In July 1960, DDK gave a talk to the Rotary Club of Poona on "Atomic Energy for India." This essay [43] is an unabashed advocacy of solar power over atomic power, mirroring in a sense his ideological conflict with the DAE. Half a century later, many of these issues remain current and the arguments remain valid, as for example the following observation.

> It seems to me that research on the utilisation of solar radiation, where the fuel costs nothing at all, would be of immense benefit to India, whether or not atomic energy is used. But by research is not meant the writing of a few papers, sending favoured delegates to international conferences and pocketing of considerable research grants by those who can persuade complaisant politicians to sanction crores of the taxpayers' money. Our research has to be translated into use.

There is more in these essays on solar energy that merits attention even today such as his observations on energy storage and distribution, and on environmental issues [43]. Eventually matters came to such a pass as to cause the DAE to not renew DDK's contract. As already pointed out, the RH papers had caused a serious blow to Kosambi's mathematical reputation and while this was made out as the proximate cause for his dismissal from TIFR, trouble had been brewing for some time. The letters between Bhabha and DDK grew increasingly formal, bureaucratic, and strained. There was a distinct difference in styles, and the iconoclastic Kosambi was hardly one to fit into the DAE mould.

1.4 Pseudonyms and Aliases

DDK was responsible for the first mention of Bourbaki in the mathematics literature in his publication [DDK2] in 1931, although the obscurity of the Proceedings of the Academy of Sciences, UP, has resulted in the article receiving less attention than it deserved, even from a purely historical point of view [44]. André Weil had suggested a prank that Kosambi ascribe a theorem to a nonexistent (but possibly) Russian mathematician, in order to put down an older colleague in Aligarh who was giving the young Kosambi a difficult time. There is not much more than a paragraph in Weil's autobiography [14] on this episode, so the circumstances surrounding the event are difficult to reconstruct. Nevertheless, this *parodic note passed off as a serious contribution to a provincial journal* is not entirely facetious.

It was not until December 1934 that the Bourbaki idea acquired more momentum [45, 46], when Weil along with Henri Cartan, Claude Chevalley, Jean Delsarte, Jean Dieudonné, and René de Possel, decided "... to define for 25 years the syllabus for the certificate in differential and integral calculus by writing, collectively, a treatise on analysis. Of course, this treatise will be as modern as possible." The book [47] would eventually appear in 1938, authored by the group that now called themselves Nicolas Bourbaki [44]; they then went on to write many more (and extremely influential) volumes. An Indian connection remained: when Boas mentioned (in the Britannica Book of the Year) that Bourbaki was a collective pseudonym, he got an indignant letter of protest, from Bourbaki, writing from his *ashram in the Himalayas* [48]. It should also be noted that Kosambi gives credit to a D. Bourbaki [DDK2] although the forename eventually chosen by the French group was Nicolas [49].

Aliases were used by DDK several times: "Ahriman" in an article published in the magazine of Fergusson College [50], "Indian Scientist" in a piece titled "The Raman Effect" [51], and "Vidyārthi" in a note [52] that used statistics (maybe his nod to William Sealy Gosset, the chemist and statistician who, as "Student" invented the t-test). Apart from these scattered instances, between 1962 and his death in 1966, DDK used the *nom de plume* S. Ducray extensively, both in personal correspondence as well as professionally. This was also by far his most elaborately chosen alias.

It is difficult to discern what led him to use the pseudonym S. Ducray. The alleged etymology is that Bonzo, the Kosambi family dog in the 1960s, was quite plump, and DDK affectionately called him *Dukker*, namely "pig" in Marathi. This evolved into Ducray, a name that sounds vaguely French, with the forename being the Sanskrit for dog, namely *Svana* [53]. The choice of such a name remains enigmatic, and while it may have been prompted initially by his anger with the establishment—to date Kosambi is among the very few persons to have had their appointment terminated by the Department of Atomic Energy—there is enough to suggest that there may be more to the use of this alias than pique. In fact, he signed several letters to his friends as S. Ducray [35].

DDK published four articles as S. Ducray, two in the Journal of the University of Bombay [DDK63, DDK64] and two in the *Proceedings of the Indian Academy of Sciences* [DDK66, DDK67]. The latter two were in fact communicated by C.V. Raman. While this may have been a formal device employed by the journal, it is highly unlikely that Raman knew of the masquerade. Had Raman known, it is also highly unlikely that he would have permitted such subterfuge in a journal of his Academy. These two papers were serious enough as works of mathematics, as were the other two Ducray papers that were submitted to the Journal of the University of Bombay. Indeed, two of these four papers were reviewed in *Mathematical Reviews*. All the four articles show a strong connection to DDK, acknowledging him in one and quoting a private communication from Paul Erdős in another, in addition, of course, to citing his related papers written as D.D. Kosambi.

These papers continued the prime obsession that DDK showed in his last years. Regrettably, the manuscript of his book [54] was lost. If nothing else, it would have provided some clues as to how he hoped to use probability theory in this arena. Although reviewed in MR, the papers had serious shortcomings. J. Kubilius who himself worked in the area of probabilistic number theory says of "Probability and prime numbers" [DDK65] that "The reviewer could not follow the proof of the cardinal Lemma 3." The paper "Normal Sequences" [DDK63] was comprehensively reviewed by B. Volkmann who pointed out a number of inaccuracies and misprints.

One of DDK's earlier papers had been reviewed in *Mathematical Reviews* by E.S. Pondiczery: this was the editor Ralph Boas Jr's pseudonym, a fanciful "slavic" spelling of Pondicherry. The name, which Boas used even when writing serious mathematics, was apparently concocted for its initials, ESP, and was to have been used for writing an article debunking extra-sensory perception. Boas had a well-developed sense of the ludic and was one of the authors of the brilliant article "A Contribution to the Mathematical Theory of Big Game Hunting" that was published in the American Mathematical Monthly under the (collective) pseudonym H.W.O. Pétard [55]. Both Boas and Kosambi were publicly dismissive of extrasensory perception [56]. Perhaps it was these connections that inspired Kosambi when he was to later adopt the Ducray alias.

1.5 Concluding Remarks

History may not have been particularly kind to Kosambi, the mathematician, but in his lifetime DDK was appreciated for his scholarship and intelligence [57] early in his career and by his peers. The manner in which Kosambi was viewed by his contemporaries—many of who were more distinguished than him and had a more significant impact on mathematics—is revealing. From 1930 to 1958 or so, DDK enjoyed the respect and admiration of a large professional circle. As has been noted earlier [5], his contributions in areas such as ancient Indian history, Sanskrit epigraphy, Indology, as well as his writings of a political and pacific nature grew both in volume and in substance in the 1940s and 1950s, overshadowing his mathematics, although the constancy of his work in the area remained. His wide scholarship and his ability to integrate different strands of thought gave him an large and dispersed audience, although his temperament and his politics were also well known and not as widely appreciated.

One important recognition that was accorded him, in part due to his being at the TIFR and the association with Bhabha, but also for his work and his mathematical antecedents [58] was his appointment as a member, in 1950, of the Interim Executive Committee of the International Mathematical Union, to serve along with Harald Bohr, Lars Ahlfors, Karol Borsuk, Maurice Fréchet, William Hodge, A. N. Kolmogorov and Marston Morse. One of the tasks of this rather distinguished group was to choose Fields medalists, and DDK served on this committee for two years.

It is thus noteworthy that in a period that spans three decades, Kosambi was mathematically productive, prolific, original, and was taken seriously by the scientific establishment in the country, as his elections to the Fellowships of the Indian Academy of Sciences and the Indian National Science Academy and the Presidency of the Mathematics section of the 34th Indian Science Congress in 1947, among other distinctions, testify. His papers appeared in leading journals of the world and were communicated by or reviewed by some of the leading mathematicians. And that this happened while his reputation in a diametrically different field was also burgeoning can only be seen as evidence of a complex but nevertheless Promethean intellect.

Chapter 2
Adventure into the Unknown

This essay, published posthumously in the collection Current Trends in Indian Philosophy [28], *resulted from an invitation from scholars at Andhra University to write on his 'personal philosophy as a scientist and research worker'. A somewhat bowdlerised version of this article has been excerpted and published as 'Steps in Science' in the collection* Science and Human Progress: Essays in honour of late Prof. D.D. Kosambi, scientist, indologist and humanist [59]. *The important Epilogue (Sect.* 2.6*) was unfortunately left out of 'Steps in Science'.*

2.1 Why Science?

The question 'Why solve problems?' is psychological. It is as necessary for some as breathing. Why scientific problems, not theology, or literary effort, or some form of artistic expression? Many practising scientists never work out the answer consciously. Those lands where the leading intellectuals speculated exclusively upon religious philosophy and theology remained ignorant and backward, and were progressively enslaved (like India) in spite of a millennial culture. No advance was possible out of this decay without modern techniques of production, towards which the intellectuals' main contribution was through science. There is a deeper relationship: science is the cognition of necessity; freedom is the recognition of necessity. By finding out why a certain thing happens, we turn it to our advantage rather than be ruled helplessly by the event. Science is also the history of science. What is essential is absorbed into the general body of human knowledge, to become technique. No scientist doubts Newton's towering achievement; virtually, no scientist ever reads Newton in the

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_2

original. A good undergraduate commands decidedly more physics and mathematics than was known to Newton, but which could not have developed without Newton's researches. This cumulative effect links science to the technology of mechanised production (where machine saves immense labour by accumulating previous labour) to give science its matchless social power in contrast to art and literature with their direct personal appeal. Archimedes, Newton and Gauss form a chain wherein each link is connected in some way to the preceding; the discoveries of the latter would not have been possible without the earlier. Shakespeare does not imply the pre-existence of Æschylus or of Kalidasa; each of these three has an independent status. For that very reason, drama has advanced far less from the Greeks to the present day than has mathematics or science in general. Even the anonymous statues of Egypt and Greece or the first Chinese bronzes show a command of technique, material and of art forms that make them masterpieces, but the art is not linked to production as such, hence not cumulative. The artist survives to the extent that his name remains attached to some work that people of later ages can appreciate. The scientist, even when his name be forgotten, or his work buried under the wrong tombstone, has only to make some original contribution, however small, to be able to feel with more truth than the poet, 'I shall not wholly die; The greater part of me will escape Libitina'. The most bitter theological questions were argued out with the sword; for science, we have the pragmatic test, experiment, which is more civilised except when some well-paid pseudo-scientist wishes to 'experiment' with thermo-nuclear weapons or bacterial warfare.

2.2 Natural Philosophy

It was obligatory for me to learn several European languages in school and college in the USA. The libraries were the best in the world for accessibility and range of books. Alexander von Humboldt's Cosmos surveyed the whole universe known to the middle of the nineteenth century, from the earth to those mysterious prawn-shaped figures visible through the powerful telescopes, the spiral nebulae. The Einstein theory, arousing passions of theological intensity, had just been regarded as proved, and offered new insight into the structure of space, time and matter. Innumerable outlines made it easy to learn something about every branch of science. Freud had taught men to take an honest look at their own minds. H.G. Wells showed in his *Outline of History* how much the professional annalistic historian had to learn, though Spengler's *Untergang des Abendlandes* made it extremely unlikely that the historian would learn it. The inspiring lives of Pasteur and Claude Bernard proved that man could gain new freedom from disease through the laboratory; the deadliest poison became a tool for the saving of life through investigation of the body's functions. Such were the real *ṛṣis* and *bodhisattvas* of modern times, the sages whose scientific achievement added to man's stature. This contrasted with the supposed inner perfection of mythical Indian sages, expressed in incomprehensible language and fantastically interpreted by commentators. The ability to replace incomprehensible

Sanskrit words by still longer and equally meaningless English terms can make a prosperous career. It cannot produce an Albert Schweitzer, whose magnificent study *Von Reimarus zu Wrede*, analysis of Bach's music and record as medical missionary at Lambarene were impressive even in my irreverent undergraduate years.

Engineering is based upon physics and chemistry, which are qualified as 'exact sciences' precisely because they admit a mathematical basis. Mathematics unlocked the door to the atom and to the movement of celestial bodies equally well. Aptitude granted, mathematical research needed the least financial resources of any science. Mathematical results possess a clarity and give an intellectual satisfaction above all others. They have absolute validity in their own domain, due to the rigorous logical process involved, independent of experimental verification upon which the applications to the exact sciences must depend. This was the very language of nature, *scientiarum clavis et porta* as Roger Bacon put it. Its supreme, transcendental, aesthetic fascination can only be experienced, never explained.

Unfortunately, not every kind of mathematics unlocks every door to nature's secrets. For some twenty years, my main work lay in tensor analysis and path-geometry (my own term). The structure of space–time had been analysed by the measurement of 'distance' in space and time; I showed that it could be done without distance, merely by the racks that explored the 'space', even when the concept of 'space' was generalised beyond physical recognition. In 1949, Einstein pointed out to me during one of several long and highly involved private technical discussions that certain beautifully formulated theories of his would mean that the whole universe consisted of no more than two charged particles. Then, he added with a rueful smile, 'Perhaps I have been working on the wrong lines, and nature does not obey differential equations after all'. If a scientist of his rank could face the possibility that his entire life-work might have to be discarded, why insist that the theorems whose inner beauty brought me so much pleasure after heavy toil must be of profound significance in natural philosophy? Fashions change quickly in physics where theory is so rapidly outstripped by experiment. It seemed and still seems to me that non-associative linear algebras and Markov chains would remove many of the physicists' theoretical difficulties; the experimenters are satisfied with abandoning the principle of parity. The 'redshift' of distant stars will perhaps be explained one day as due to the absorption of energy when light travels at cosmic distances through extremely tenuous matter, rather than evidence for an expanding universe. Such speculations are of no use unless they tally in mathematical detail with observed data.

2.3 Chance and Certainty

Borderline phenomena of classical physics illustrate the inexhaustibility of the properties of matter. Ice, according to the textbooks, melts and water freezes at zero degrees centigrade. But when carefully purified samples of water are slowly cooled and the ice slowly melted again, a considerable gap is found between the melting and freezing points. Fundamental particles that make-up the atom and its nucleus show another type of aberrant behaviour. An electron can cross a potential barrier, as if a stone were of itself to roll uphill against gravity and down the other side. Even the observation of isolated particles becomes difficult, for the very act of observation means some interaction and effect upon the observable. The certainty of classical physics develops only when many fundamental particles are organised into higher units with clear patterns. In the same way, individual molecules of water may move in any direction with almost any speed, but the river as a whole shows directed motion in spite of eddies, so also for aggregates of living matter. In human society, the net behaviour of the group smooths out the vagaries of individual action.

The mathematical analysis best suited for handling such aggregates is the theory of probability. Variation is as important a characteristic of the collective as the mean value. Prediction can only be made within a certain probability, which sounds like the language of the race course. But when the chances of a mistake amount to one in a million, most people take the effect as certain. The level of significance desired may be a personal matter. For example, there is a chance of a letter being lost in the mail; whether or not we register or insure it depends upon our estimate of the risk involved and the expectation of loss. Thus, modern statistical method can be an excellent guide to action. It extends the assurance of exact science to biological and social sciences. Though no man can say when death will come to him, as it certainly must, it is fairly easy to predict within a reasonable margin of error about how many men out of a large group will die after a set number of years. That is why life insurance manages to be a highly paying business, without recourse to astrology. It is further possible to say how occupation and living conditions affect longevity. The man who has to work in a lead mine (without special protection) has his expectation of life reduced by a predictable number of years, more surely than if he were shot at by lead bullets on the battlefield.

Deductions based upon probability differ radically from those of pure mathematics. Conclusions cannot be 'true or false' without qualification, when the variation inherent in the trials is assessed. The standard method is to set up a 'null hypothesis', and take the observed results as due to purely random independent variation. The theory suitably applied (and the application needs profound grasp) then gives one of two conclusions that the numerical observations (if relevant) are compatible with the hypothesis or not. But either conclusion would be true only with a certain calculable probability, which tells us about how often we would go wrong in action. The trick is to set up the experiment in such a way that the desired action may be taken if the null hypothesis is contradicted, for incompatibility implies falsehood whereas compatibility need not imply truth.

This may lead to difficulties when the experimenter's will to believe is stronger than his common sense. Parapsychologists test *ESP*, 'extra-sensory perception' (such as telepathy) by having two people match cards at a distance. The effect is so faint and irregular as to call for delicate statistical tests, which show that the chances are very small, for random matching, wherefore the parapsychologists claim victory. Unfortunately, my own experiments showed that the kind of shuffling practised for *ESP* is inefficient when judged by the same kind of statistics that is applied to card matching. Cards originally next to each other tend too often to stay together.

Claims of *ESP* would be more convincing if one produced supplementary evidence (say matching encephalograms for sender and receiver) for a physical mechanism of transmission. Some regard the effect as beyond normal sensation, transcendental, not accessible to material analysis. In that case, laboratory tests and the statistical 'proof' become mere ritual.[1]

One of my theoretical papers deals with probability and statistics in infinitely many dimensions. There has been no effective use, because we could not get or make the special electronic calculating machine needed to translate this theory into practice. On the other hand, a brief note on genetics was unexpectedly successful. Professional geneticists use it for all kinds of investigations, such as heredity in house mice. It seems to have given a new lease of life to genetical theories which I, personally, should like to see revised. I am accused at times of not appreciating my own formula. It would have been pleasant to see the formula applied to the increase in food production, but the pure scientists of the country which grows the world's greatest food surpluses and suppresses or destroys them to keep grain prices high in a hungry world sneer at 'clever gardening'. There is some difference of opinion here as regards the proper relation of theory to practice.

2.4 Ancient Indian Culture

To teach myself statistics, I decided to take up some practical problems from the very beginning. One such was the study of examination marks of students. It turned out that even the easiest of examinations in India (the first-year college examination) was based on a standard that differed from that of the instruction, if in 25 years no student of the 90 % or more that passed could score more than 82 % overall while the professors who taught and examined had scored much less in their own time. Improvement of the system (whether in examination or instruction) was out of the question in a country where the teaching profession is the waste-basket of all 'white-clothes' occupations and the medium of higher instruction still remains a foreign language.

A more fruitful problem was the statistical study of punch-marked coins. It turned out that the apparently crude bits of 'shroff-marked' silver were coins carefully weighed as modern machine-minted rupees. The effect of circulation on any metal currency is obviously to decrease the average weight in proportion to the time and to increase the variation in weight. This is the mark any society leaves upon its coinage, just by use. The theory of this 'homogeneous random process' is well known, but its application meant the careful weighing, one at a time, of over 7,000 modern coins as control. Numismatics becomes a science rather than a branch of epigraphy and archaeology. The main groups of punch-marked coins in the larger Taxila Hoard

[1] All the well-designed experiments in parapsychology have used random procedures for target selection, and the statistics used in *ESP* research were approved by the American Statistical Institute as early as in the 1930s—K.R. Rao.

could be arranged in definite chronological order, the oldest groups being the lightest in average weight. There seems to have been a fairly regular pre-Mauryan system of checking silver coins.

Arranging coin groups in order of time led naturally to the question: Who struck these coins? The hoard was deposited a few years after Alexander's death: but who left the marks on the coins? The shockingly discordant written sources (*Purāṇas*, Buddhist and Jain records) often give different names for the same king. Study of the records meant knowledge of Sanskrit, of which I had absorbed a little through the pores. Other preoccupations made it impossible to learn the classical idiom like any other beginner. So, the same method was adopted as for the study of statistics: to take up a specific work, of which the simplest was Bhartṛhari's epigrams (*subhāṣitas*). The supposed philosophy of Bhartṛhari, as glorified by commentators, was at variance with his poetry of frustration and escape. By pointing this out in an essay which caused every god-fearing Sanskritist to shudder, I fell into Indology, as it were, through the roof.

There was one defect in the essay, in that the existence and the text of Bhartṛhari were both rather uncertain. This meant text criticism, which ought to have been completed in a few months, as the entire work supposedly contains no more than 300 stanzas. Study of about 400 manuscripts yielded numerous versions with characteristically different stanzas, as well as divergent readings in the common verses. Two and a half years of steady collation work showed that I should never have undertaken such a task, but abandoning it then would mean complete loss of the heavy labour, which could yield nothing to whoever came after me. It took 5 years to edit Bhartṛhari, but even the critics who dislike the editor or his philosophy maintain that the result is a landmark in text criticism. Different methods were needed to edit (with a very able collaborator) the oldest known anthology of classical Sanskrit verse, composed about A.D. 1100 under the Pāla dynasty. The main sources were atrocious photographs of a palm-leaf manuscript in Tibet, and of a most corrupt paper manuscript in Nepal. My judgement of the class character of Sanskrit literature has not become less harsh, but I can at least claim to have rescued over fifty poets from the total oblivion to which lovers of Sanskrit had consigned them.

All this gave a certain grasp of Sanskrit, but hardly of ancient Indian history; the necessary documents simply did not exist. My countrymen eked out doubtful sources with an exuberant imagination and what L. Renou has called 'logique imperturbable'. One reads of the revival of Nationalism and Hinduism under Chandragupta II, of whom nothing is known with certainty. Indian nationalism is a phenomenon of the bourgeois age, not to be imagined before the development of provincial languages (long after the Guptas) under distinct common markets. Our present-day clashes between linguistic groups are an index to the development of local bourgeoisies in the various states. Hinduism came into existence after Mohammedan invasion. Clearly, one of two positions had to be taken. Either India has no history at all, or some better definition of history was needed. The latter I derived from the study of Karl Marx, who himself expressed the former view. History is the development in chronological order of successive changes in the means and relations of production. Thus, slavery in the Graeco-Roman sense was replaced by the caste system in India only because commodity production was at a lower level. Indian history has to be written without the episodes that fill the history books of other countries. But what were the relevant sources? Granted that the plough is more important than a dynasty, when and where was the tool first introduced? What class took the surplus produced thereby? Archaeology provided some data, but I could get a great deal more from the peasants. Field work in philology and social anthropology had to be combined with archaeology in the field as distinguished from the site archaeology of a 'dig'. Our villagers, low caste nomads, and tribal minorities live at a more primitive stage than city people or the brahmins who wrote the *purāṇas*. Their cults, when not masked by brahmin identification with Sanskritised deities, go back to prehistory like the stone axes used in Roman sacrifices. Tracing a local god through village tradition gives a priceless clue to ancient migrations, primitive tracks, early trade routes and the merger of cattle breeding tribesmen with food gatherers which led to firm the agricultural settlement. The technique of observation has to be developed afresh for every province in India. The conclusions published as *An Introduction to the Study of Indian History* had a mixed reception because of the reference to Marx, which automatically classifies them as dangerous political agitation in the eyes of many, while official Marxists look with suspicion upon the work of an outsider.

Field investigation continues to give new and useful results. Experts say glumly that my collection of microliths is unique not only in range of sites but in containing pierced specimens. A totally unsuspected megalithic culture came to light this year. It fell to my lot to discover, read and publish a Brāhmi inscription at Kārle caves, which had passed unnoticed though in plain sight of the 50,000 people who visit the place every year. The suggestion for using the Māḷsheṭ Pass should give Maharashṭra a badly needed key road from Bombay to Ahmadnagar, and save a few million rupees though the funicular railway down Nāneghāṭ would have been more spectacular.

2.5 Social Aspects

The greatest obstacles to research in any backward, underdeveloped country are those needlessly created by the scientist's or scholar's colleagues and fellow citizens. The meretricious ability to please the right people, an attractive pose, glib charlatanism and a clever press agent are indispensable. Mere scientific ability is at a discount. The Byzantine emperor Nikephoros Phokas assured himself of ample notice from superficial observers, at someone else's expense by setting up in his own name at a strategic site in the Roman Forum, a column pilfered from some grandiose temple. Many eminent intellectuals have mastered this technique in India.

The deep question is not what floats to the top of a stagnant class but of fundamental relationship between the great discoverers and their social environment. Conservatives take history as the personal achievement of great men, especially the history of science. The Marxist assertion is that the great man is he who finds some way to fulfil a deep though perhaps unstated social need of his times. Thus, B. Hessen explained Newton's work in terms of the technical and economic necessities of his class, time and place. The thesis was successful enough to be noticed and contested by a distinguished authority on seventeenth-century European history, Sir George Clark. Clark's knowledge of the sources is unquestionably greater than Hessen's, but the refutation manages to overreach the argument. According to Clark, the scientific movement (of the seventeenth century) was set going by 'six interpenetrating but independent impulses' from outside and 'some of its results percolated down into practice and were applied'. The external impulses were 'from economic life, from war, from medicine, from the arts and from religion. What is left then of the independence of science?' The sixth impulse was from the 'disinterested desire to know'. So far as I know, all six impulses applied from the very earliest civilisations of Mesopotamia, Egypt, China and probably the Indus Valley, without producing what we recognise as 'science' from, say, the time of Galileo. What was the missing ingredient, if not the rise of the proto-bourgeoisie in Europe? No Marxist would claim that science can be independent of the social system within which the scientist must function.

Much the same treatment may be given to the literature. Disregarding oversimplification, can one say that Shakespeare's plays manifest the rise of the Elizabethan proto-bourgeoisie, when the said dramas are full of kings, lords and princes? The answer is yes. Compare *Hamlet* or *Richard the Third* with the leading characters in *Beowulf* or the *Chanson de Roland*. The fattest Shakespearean parts such as Shylock and Falstaff are difficult to visualise in any feudal literature. The characters in those plays have a 'modern' psychology, which accounts for their appeal to the succeeding bourgeoisie and hence for the survival value of the dramas. Troilus and Cressida are not feudal characters any more than they are Homeric; Newton's Latin prose and archaic geometrical proofs in the *Principia* make that work unreadable, but do not make it Roman or Greek science.

It would take a whole book to develop this thesis for India's trifling successes and considerable failure in modern science. In what follows, only the most obvious defects in applying science to major Indian problems are considered, without discussion of the extent to which this accounts for the lack of really great scientists in India.

India, the experts tell us, is overpopulated and will remain poor unless birth control and population planning are introduced. But surely, overpopulation can only be with respect to the available food supply. Availability depends upon production, transport and the system of distribution. What is the total amount of food produced? We have theological quarrels between two schools of statisticians, but no reliable estimate of how much is actually grown and what proportion thereof escapes vermin—including middlemen and profiteers—to reach the consumer. If shopkeepers can and do raise prices without effective control, what does a rise in the national income mean? Is the scarcity of grain or of purchasing power? A great deal is said about superstitious common people who must be educated before birth control becomes effective. The superstition which makes the poor long for children has a solid economic foundation. Children are the sole means of support for those among the common people who manage to reach helpless old age. The futility of numerical 'planning' of the population, when nothing is done to ensure that even the able-bodied have a decent level of subsistence, is obvious to anyone but a born expert. Convince the people that even the childless will be fed and looked after when unable to fend for themselves and birth control will become popular.

Let me give examples of scientific effort which could easily have been turned to better account. Considerable funds will be devoted during the Third Plan to research on the uses of bagasse (sugarcane pulp). At present, it is used as fuel and the ashes as fertiliser, whereas paper and many other things could be made from it. But are the other uses (quite well known) the best in the present state of Indian economy? The extra money to be spent on fuel, not to speak of difficulties in getting fuel, would increase the already high cost of sugar manufacture; new factories for by-products mean considerable foreign exchange for the machinery and for the 'experts'. However, if the bagasse is fermented in closed vats, the gas given off can be burned, so that the fuel value is not reduced. The sludge makes excellent fertiliser, which saves money on chemical fertilisers and improves the soil. The scheme (not mine, but due to Hungarian scientists) has apparently been pushed into the background. Again, the proper height of a dam is important in order to reduce the outlay to a minimum, without the risk of running dry more than (say) once in 20 years. The problem is statistical, based upon the rainfall and run-off data where both exist. The principles I suggested were adopted by the Planning Commission, though not as emanating from me. Neither the engineers nor the Planning Commission would consider a more important suggestion, namely, that many cheap small dams should be located by plan and built from local materials with local labour. Monsoon water would be conserved and two or three crops raised annually on good soil that now yields only one. The real obstacle is not ignorance of technique but private ownership of land and lack of cooperation among the owners.

This country needs every form of power available, but is too poor to throw money away on costly fads like atomic energy merely because they look ultra-modern. A really paying development will be of solar energy, neglected by the advanced countries because they have not so much sunlight as the tropics. Our problem lies deeper than power production. The reforestation, indispensable for good agriculture, will not be possible without fuel to replace the firewood and charcoal. Coal mining does not suffice even for industry; fuel oil has to be imported. A good solar cooker would be the answer. Such cookers exist and have been used abroad. The one produced in India was hopelessly inefficient (in spite of the many Indian physicists of international reputation). Neatly timed publicity and a fake demonstration made the gullible public buy just enough useless 'cookers' for a quick profit to the manufacturer.

A flimsy 'Indian Report' on the effects of atomic radiation shows our low moral and scientific calibre by ignoring the extensive data compiled since 1945 in the one country which has had the most painful experience of atomic radiation applied to human beings—Japan. The real danger is not death, which is a release for most Indians, but genetic damage to all humanity. We know what radiation does to heredity in the ephemeral banana-fly *Drosophila melanogaster*. A good deal was found out in the USA about what happens to laboratory mice. What little has been released for the publication is enough to terrify. Man is as much more complicated than a mouse as the mouse than the fruit fly. Humans take a proportionately longer time to breed and to reach maturity, giving fuller scope for genetic derangements to develop. It may take some twenty generations to find out just what these derangements amount to. By then, they will have been bred into many millions of human beings, not as a disease but incurably as a set of hereditary characters. Mankind cannot afford to gamble with its own future in this way, whether that future lies in the hands of communists or not. Atomic war and the testing of nuclear weapons must stop. These views on nuclear war are now fashionable enough to be safely expressed.

2.6 Epilogue

A mathematician must earn that designation by enriching mathematics with original theorems of basic importance. Einstein, for all the stimulus his ideas gave to contemporary differential geometry, was not, and never regarded himself as a mathematician. So, my excursions into statistics, Indology, archaeology and the rest are irrelevant unless some real mathematics emerged at the end. Alternatively, is there something wrong in the philosophy that asserts the unity of theory and practice?

Mathematics is no longer the by-product of a natural philosopher's investigations, as it had been from Pythagoras to Gauss. All sorts of mathematical technique exist today, fully developed long before the physicist feels the need for it. One should contrast G.H. Hardy's *Mathematician's Apology* (Cambridge, 1941) with L. Hogben's *Mathematics for the Million* (London, 1936). The former, though leader and virtually creator of the modern school of British mathematics, was indifferent to the applications and the social context of mathematical discovery. Those were the aspects of mathematics of primary interest to the biologist Hogben, who thereby presented rather elementary mathematics in attractive popularisation. Hardy counted uselessness among the great assets of real mathematics; forgetting Archimedes's military engines, he blamed 'Hogben mathematics' for the senseless destruction of world wars. This was just before the manufacture of nuclear weapons by the 'Science has known Sin' group, in collaboration with outstanding mathematicians like J. von Neumann. If any important mathematics came out of the atomic and hydrogen bombs, the secret has been well kept.

The theory of numbers is the oldest branch of mathematics. Hogben mathematics would not exist without numbers, while Hardy and his associates devoted their best efforts to number theory. Two outstanding problems here are as follows: (1) Fermat's Last Theorem, which can be explained to a schoolboy in spite of its melodramatic title and (2) the Riemann Hypothesis, decidedly more recondite. Both have defeated the efforts of great mathematicians to prove or to disprove them. The Fermat theorem, if true, would lead to no new mathematics; proof of the Riemann conjecture would lay the very foundations of analytic number theory. These unsolved problems gave rise to a distressing possibility in mathematical reasoning: Was there a category of propositions 'neither (demonstrably) true nor false'?

Riemann's conjecture has to do with the distribution of primes, which are those integers (like 257) not divisible by any smaller number except unity. Every whole number can be expressed in just one way as the product of primes, hence their importance. There are infinitely many primes. A given integer is either a prime or not, with no question of probability; yet the occurrence of primes among the integers is highly irregular, without a pattern. Given a specific prime, it is always possible to find the next by hard work, but not by formula. This parallels an experimental situation. Weights of coins of the same denomination fluctuate so much that I could never predict what the next coin would show on delicate balances. However, if there was a next coin, its weight could always be recorded as one more figure of a series. Enough such figures outlined a curve for the distribution of weights. The series of weights formed a *sample* from a population assumed subject to probability laws. Could something of the sort not be proved for the primes? It was necessary to change the scale, because primes occur with less and less frequency (on the whole) as the integers grow larger. The change gave a fixed *average* number of primes per interval of any constant length on the changed scale. Still, the number varied unpredictably from interval to interval. The number of primes per interval was then shown by me to follow a simple though unsuspected probability law, the Poisson distribution. This describes many experimental samples such as the number of cosmic rays per second, of bacteria in thin cultures and of calls in a telephone exchange. The previous failures in prime number theory resulted from the attempt to fit an exact description to an infinite set of infinite random samples.

Every competent judge who saw only this radically new basic result intuitively felt that it was correct as well as of fundamental importance. Unfortunately, the Riemann hypothesis followed as a simple consequence. Could a problem over which the world's greatest mathematicians had come to grief for over a century be thus casually solved in the jungles of India? Psychologically, it seemed much more probable that the interloper was just another 'circle-squarer'. Mathematics may be a cold, impersonal science of pure thought; the mathematician can be thoughtless, heatedly acrid, even rabid, over what he dislikes. Let me admit at once that I made every sort of mistake in the first presentation. There is no excuse for this, though there were strong reasons: I had to fight for my results over three long years between waves of agony from chronic arthritis, against massive daily doses of aspirin, splitting headaches, fever, lack of assistance and steady disparagement. It was much more difficult to discover good mathematicians who were able to see the main point of the proof than it had been to make the original mathematical discovery. How much of this is due to my own disagreeable personality and what part to the spirit of a tight medieval guild that rules mathematical circles in certain countries with an 'affluent society' need not be considered here. There is surely a great deal to be said for the notion that the success of science is fundamentally related to the particular form of society.

Chapter 3
On Statistics

This essay was written after 1945 by which time DDK had moved to TIFR. It is not clear whether this was a lecture given to colleagues there or to a different audience (cf. the reference to Bombay House in the last paragraph). There is not much formal statistics here: indeed, there is not a single equation. The essay is fairly discursive and even somewhat polemical; it reflects the mixture of ideas that Kosambi carried along with the mathematics and statistics that were basic to his analysis [42].

Modern statistics, as contrasted with descriptive statistics of the older type, differs primarily in being a guide to action, which implies more accurate results with an estimate of the error and greater rapidity of working with smaller samples observed. It is not realised that ancient statistics was also in its own way a guide to action, its lack of credit today being due solely to its clumsier apparatus and, with more reason, to the distressing quality of its findings from the point of view of a certain class of people.

The standard types of descriptive statistics are the ancient Roman census which was after all a stock taking for the purposes of the state; its logical continuation in feudal times is the Domesday book of William the Conqueror which (allowing for the changed circumstances) is much the same thing, namely a bit of stocktaking for taxation purposes.

I might illustrate the rise of a new class and a new way of thinking by pointing out to you the change in European literature in the seventeenth and the eighteenth centuries. The older literature dealt with persons of heroic stature who specialise in humanly impossible knightly adventures. The tradition begins with the *Chanson de Roland*, to continue through the entire Arthurian Round Table cycle and the deeds of the Paladins of Charlemagne. On the other hand, when you look down into later literary efforts, you find quite unheroic average figures as, for example, Lesage's *Gil Blas* or Marivaux's *Paysan parvenu* or the most attractive of them all, Voltaire's *Candide* who passes through adventures which are quite romantic in themselves but in which his behaviour is such that the reader can say "this might happen to anyone".

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_3

It is not realised, however, that this change in literature corresponds to an appreciation of the fact that human beings have a certain average and a certain standard or norm, which in itself is well worth studying. The terms “average man“had to wait until well into the nineteenth century, but the concept definitely existed long before Voltaire whose charming story on the subject is entitled “*L'homme a quarante écus*” deals with a person who has forty minted pieces of money, which was the average wealth of a Frenchman at that time. The point is that someone had made a fairly reliable estimate of the total national wealth and of the total population, thereby reaching the estimate of average wealth and actually showing the possibility of a statistical approach to the whole subject. That is, the “average men” who figure in the literature of the nineteenth century depend upon a statistical attitude towards humanity. This naturally is to be expected from the long tradition of study which lies back of Sully, Bodin, Turgot, and the entire school of physiocrats; it also implies the rise of a new type of people who liked to think in this manner peculiar to the French bourgeoisie.

Let us look now at the development of the need for a wider kind of statistics than that necessary only for budget and taxation. In the year 1542, for example, we find that at Antwerp wagers are being laid against the sex of on unborn child. A merchant would undertake to pay 30 livres if the offspring were a girl, whereas in gratitude, he would receive from the mother 48 livres if a son were born. Or there may be a wager that the exchange rate would be at a 2 % premium or discount; there would be other wagers which would deal with the failure or success of a certain standing crop or the safe homecoming of a given ship. These look like gambles, but you will see at once that these are a primitive type of insurance. The thirty livres go towards the girl's dowry, whereas the son could earn his keep and the 48 livres besides. Insurance of cargoes goes right back to Roman and Grecian times. Nevertheless, with statistics undeveloped, and a very poor control of the subject as well as very few cases coning up, the merchant who undertook this enterprise was virtually a gambler. Briefly, if you insure one man for 7 million rupees, you are running a far greater risk than if you insure 7000 men for 1000 rupees each. The fuller development comes in the year 1836, when the Belgian A. Quetelet published his book on *Social Physics*. Quetelet, to who the term “average man” is due, was the first person to break down any modern census figures. He found among other things that crime depended to a regularly predictable extent upon the economic level of the class under consideration. The poorer the class, the greater the incidence of crime, independently of the locality. It was also found that mortality depended with a tremendous regularity upon professions. If a man has to work for his living in a military profession in a period of constant warfare, he certainly risks his life in an open manner. But it was not realised that forcing a certain worker to work with lead—as for example types of glazing—or phosphorus and sulphur as in matchmaking, or mercury as for felt hats, amounted to sentencing that worker to die a few years earlier, or in the last case, to lunacy as well. I should like to point out that this discovery of Quetelet led in the long run to the discrediting of statistics as a science, whereas Quetelet himself believed that he had discovered new scientific laws whose inexorable character really frightened him. What he had discovered were not laws of nature but actually laws of a certain particular type of industrial society.

Such discoveries are used nowadays in building up the basis of such a gigantic financial development as the insurance business, those importance I need not explain to anyone in Bombay who can walk down a single mile of Hornby Road and read the signboards around him. Mortality tables were first made privately by the actuaries themselves in a very crude fashion and from the late seventeenth century onwards are available from census figures. The differences of mortality in different occupations are fully recognised and allowed for in the premium charged by the insurance company. But insurance is no longer speculation or gamble for the simple reason that the data from millions of policies are available and that millions of people get themselves insured so that the statistical average can apply very well. In fact, the data from insurance companies have allowed tremendous advances to be made in the science of demography, so that we can predict years in advance the approximate amount of the population as well as its structure by age groups. In the USA, for example, manufacturers pay a great deal of attention to this. It was known before the war that the number of children born was decreasing due to falling birth rate and the predictions based on this warned manufacturers of school textbooks and children's clothes to allow for so many million articles less annually. Such forecasts are not only possible but absolutely essential in a country where mass production is the rule. In other words, this type of statistics is also a guide to action, provided *it pays the right people to act upon it*. On the other hand, if it be discovered that conditions led to the shortage of certain commodities, the action generally taken is that of attempting to secure a corner in those commodities, regardless of its effect on the lives of the consumers; this happened, for example, in the Bengal famine. It is difficult to commend such action upon any basis, statistical or otherwise. I take just one more example that of Lotka's findings from the data accumulated by the Metropolitan Life Insurance Company of New York. Among other things, he showed that the general USA death rate was decreasing slowly, but with absolute regularity, up to the year 1918 when a sudden rise occurred because of the epidemic of Spanish influenza. But after that epidemic was over, the death rate fell again succeeding years not to the straight line on which it had been declining but another straight line decidedly lower than but parallel to the first. The conclusion was that the Spanish influenza had killed those with the least powers of resistance whether physical or financial and those that survived were under the circumstances of contemporary society, fitter to live. Possibly, some genius may arise here to prove a similar beneficial action for the Bengal famine.

People were not willing to face up to the idea that the structure of society might itself force certain classes to criminality, which was really a result marked out in Quetelet's work. Towards the end of the last century, eminent Italian criminologists following a great tradition founded by the jurist Cesare Beccaria began to investigate the criminal's circumstances. Among them, Lambroso, Ferraro, and Mantegazza studied the physique of convicted criminals only to discover that there was a criminal type. Criminals had abnormalities of vision, asymmetric skulls, and various types of impediments. Ergo, there was a criminal type, the criminal could not help himself, and his nature was crooked because his head was crooked. This kind of research (which would not be taken as accurate by current standards) attracted a great deal of attention because it drew emphasis away from the economic question. No one asked why the man's head had become deformed, whether poverty in childhood had anything to do with it, nor even whether the confusion of cause and effect might not have existed in such researches. Later on, these same scientists undertook to examine in the same way men of superior intellect, supposedly valuable members of society such as artists, musicians, and even professors of mathematics. Being honest men, they had to publish their findings which were, alas, that the heads of these geniuses were also not symmetrical and they too suffered from a considerable number of impediments. Nevertheless, society chose not to pay any attention to this, preferring to drop the entire subject in a quiet but discreet fashion.

At this stage, statistics becomes the joke that it is often mistaken to be. No statistician today can deliver a popular lecture without quoting that British statesman who gave vent to a classical utterance "Gentlemen—There are three kinds of lies; lies, damned lies, and statistics". The idea is that one can prove anything he likes by reference to the appropriate set of statistics. I once made a collection of jokes of this sort about statistics of which one or two may be given here. At a medical conference, an eminent child specialist gave the conclusion of 15 years of painstaking research to prove that the first day of life was the most dangerous as having the highest recorded average mortality of 28.2 %. His rival jumped up with the remark that this research was all bosh and that the last day of life was much more dangerous because then the mortality was 100 %. In another case, two very learned people were having an intellectual tea with an even more intellectual conversation which ran somewhat as follows: "My dear colleague—I find that the latest statistics show men graduates of our colleges as having 1.4 children each, whereas the lady graduates have 3.7 children each. What does this prove?" The answer came immediately, "Obviously this shows that women have more children than men". One could go on like this for ever, but I only want to make it clear to you that statistics had fallen into considerable disrepute.

From this stage, it had to be rescued by the need for application to branches of science in which experiment could not be refined beyond a certain level. In physics of the classical type, one can measure more accurately, or make pure alloys or refine the experiment almost indefinitely. In modern physics on the contrary, electrons, cosmic rays, and such new discoveries behave in a highly individualistic manner. In fact, they behave like biological specimens, that is to say the output of a certain type of grain planted in a field or the blood content of a given strain of mice, or fish in a certain lake. In this case, you can only observe but not refine the observation. Nevertheless, some method is needed for drawing accurate conclusions, that is to say a guide for immediate action. It was a biological science that developed statistics itself as a basic and even a fundamental science. The first stop in the modern direction came in the year 1908 from the mind of an able mathematician employed by the great breweries of Messrs. Guinness. They had to find from experiment in small plots which variety of barley would give the greatest yield and what types of fertilisers would increase this yield most economically. Now, it is impossible to count every grain of barley and experimenting with all their land year after year for different varieties world not only be very costly but would also not give valid result for the simple reason that rainfall conditions also differ from year to year. Finally, the ground is not uniform so that soil variation has also to be taken into account along with the possible action of insects, weeds, and other causes that might affect the grain. Nevertheless, the method was worked out and can give good results using not more than a couple of dozen plots each the size of a small room provided the plots are selected in certain random unbiased fashion. The methods were later worked up with far more detail, accuracy, penetration, and insight by R.A. Fisher who is today the great name in modern statistics. These methods are now applicable not only to biology and to cosmic rays but to sociology, archaeology, and even financial questions. Statistics may now be regarded as a science rather than as a joke or a laborious but painful method of description. With a small sample, often less than 5 % of the whole, we can draw conclusions about the entire aggregate and in addition say what the error of our estimate happens to be. This point has not always been grasped by statisticians of the older school, primarily economists, who use new methods mechanically without realising that statistics even of the most improved type cannot be a substitute for intelligence. For example, I recall an economic survey of Poona City undertaken in the years 1937–1938 which chose one house out of every fifteen for its sample without attention to appropriate randomisation and without testing for bias. The conclusions were published only in the year 1945 by which time the findings had been completely invalidated by the pressure of war, by the influx of military and other new Government establishments, and by the construction of tremendous new factories for which the working population was based upon Poona. The only excuse for the sampling survey of this sort is its rapidity as well as accuracy and its estimate of error; all three were absent in the case cited.

Let me give you a concrete example of what a precise scientific prediction is like and then show you that such a prediction can be made also by using statistical methods. Just a 100 years ago, the Newtonian theory of gravitation had began to be suspected because the outermost planet then known, namely Uranus, did not follow the path predicted for it by Newton's laws. Uranus had been discovered by Herschel (then a professional musician) an amateur astronomer who prepared his own lenses and telescopes. The question now was as to whether the Newtonian theory with its inverse square law of gravitation had to be modified or whether there existed another planet whose pull could account for the discrepancies between theory and observation as regards the movement of Uranus through the sky. Two unknown but ambitious young men independently undertook this task of explaining the discrepancy on the hypothesis of an unknown planet. Of these, Adams had the misfortune to send his conclusions to the Astronomer Royal who quietly filed them away, the other Urbain Jean-Jacques Leverrier wrote to the astronomer at Berlin Dr. Galle to the effect that if Dr. Galle would point his telescope to a part of the sky where he had recently chartered his stars he would find within the field of that telescope a new heavenly body, a star that moved, in fact a planet; and on 23 September 1846, Galle had the stirring experience given to so few of finding a new planet swim into his ken. As Arago put it to the French Academy of Sciences in reporting on the great discovery of the young French astronomer, M. Leverrier had not to see his planet with a telescope; he saw it at the end of his pen. This is the classical example of a prediction in science which was spectacular as well as fully confirmed by experiment. I could give you many such in pure science where statistics was the tool of analysis, but these will not be as spectacular as the one that I shall now call to your attention. In the USA, a journal called the Literary Digest had started the custom of taking straw votes among its readers by asking them to fill out certain types of coupons on various questions of interest, and this had led them to a quite successful study of prediction in events such as elections. In 1936, they announced their intention of taking another poll of this sort for the forthcoming presidential election. The editor was annoyed to read the assertion from a young public opinion expert, Dr. Gallup, that the *Literary Digest* would show a vote of 56 % against Roosevelt and 44 % for him, while the real facts would show an overwhelming majority of votes cast for Roosevelt and a still larger percentage of votes in the electoral college under the peculiar American system of presidential elections. What annoyed the *Literary Digest* most was that this assertion was made six weeks before their survey actually started. Yet when they had finished counting well over two million of their returns, they did announce the conclusion that Gallup had said they would announce while the election results again proved Gallup's own forecast completely while driving the *Literary Digest* to ruin. All that Gallup had done was to follow the method laid down by Fisher. He had counted, with trained observers, a small percentage of the total population of the USA in which were represented all groups in every locality according to their appropriate strength. That is, in the sample Dr. Gallup chose at random, he made certain that working class voters, voters of the professional class such as doctors and lawyers, religious groups such as Protestants of various denominations, Roman Catholics, Jews, and racial minorities were all properly represented and picked at random from local directories without personal knowledge. Predictions of this sort enable us to say that statistics has a claim to be more than a joke, in fact to be a very respectable science.

In modern statistics, the estimate of the error is a specially important point, for our statistics no longer deals only with averages and percentages but also with the amount of variation. An important function of statistics is that of "costing" or giving the amount of information in a certain sample which in effect is equivalent to showing how sharply the mechanism of observation can focus upon a given problem. The statistician, if consulted *before* the observations are taken and supplied with some minimum data about the nature of the population, can say how best to allocate the energy available for observation, or to what degree of accuracy information may be obtained from a given amount of available resources. The results have to be expressed in terms of probability which has misled many people. When we say that a certain result is significant in that there is only a chance in 20 or one in a hundred of being exceeded, we are not speaking in terms of the race course but are actually using the same kind of reasoning that you use, almost instinctively, but actually because of long experience, when you allow for a given amount of time to catch a bus or a train or for a certain letter to reach a certain correspondent. Naturally, the allowance made depends not only upon your previous experience of such happenings but also upon the importance you may happen to attach to the outcome of the particular event, whether it is a matter of routine or a matter of importance to which a cash value can be attached, or a matter of life and death. What has not been grasped is that in all cases of this sort where noticeable variation necessarily occurs—and this means in all cases where scientific observation has to be repeated—one can never get an infallible answer but only an answer that is likely to be right in almost all cases. Statistics tries to give you a definite estimate to how often you are likely to be right *in the long run*.

Nevertheless, the fact still remains that very few people are interested in statistics itself as a pure science and those people have no voice in the affairs of the world. We still have the habit of taking conclusions that are pleasing to us and ignoring the rest. For example, Gallup was able to say that the vast majority of the people who voted for Roosevelt did not approve of Roosevelt's ideas about the Supreme Court reform in spite of which Roosevelt took his third-term election an a mandate for driving the Supreme Court to acquiescence. If it comes to that, Gallup does not make his living by forecasting elections but by predicting the popularity of a certain commercial product say a new soap or a new brand of coffee or a particular kind of advertising programme for the business people of the USA. He is, inevitably, subservient to the interests of the business community, and his election forecasts are more a sort of advertising for the superior accuracy of the methods he uses. In the matter of these public opinion surveys, I might point out in them two types of results, qualitative and quantitative. Gallup's result are quantitative, and for this, a precise, accurate statistical mechanism is indispensible. But others, for example, the great anthropologist B. Malinowski, used a totally different approach to reach qualitative results. Malinowski had made quite remarkably acute observations on living conditions upon the Trobriana islanders using his own Western education as a background against which to measure the mentality of the primitive people being studied. Then, he turned this method as well the background he himself acquired by his studies over to observing the British public, and his inquiries were directed towards asking all kinds of people through the medium of trained impartial observers as to why they did or did not do certain things and noting down the results verbatim. This showed, for example, that the football pool which is virtually a swindle in Great Britain is nevertheless popular only because it is the sole method by which a member of the British working class has any chance of clearing enough money to rise out of that class. He investigated questions as to why pubs (places where alcoholic drinks are served) are so popular and what time of the day or week they were specially popular, why people do not vote in spite of the franchise, and so on. The method still continues in Great Britain under the name of Mass Observation, but its findings have raised the expected opposition and antipathy, and the mass observers are often regarded as gratuitous snoopers in spite of the fact that the Ministry of Information found it very useful to avail itself of their service on questions such as that of morale and rationing which are after all qualitative rather then quantitative questions. In India, we could use this type of qualitative analysis to find for example not how many Muslims wanted Pakistan (which would need a sampling survey) but what kind of Pakistan was meant by what particular type or types of Muslims. I might add that the question cannot be settled in any other way except such observations, for the electorate is not a random survey sample of the total Muslim population and victory at the polls says virtually nothing about the actual desire of the masses about which every politician can speak interminably. The great example of such observation is of course the work of the supreme realist of our times, Vladimir Illyich Ulianov, better known as Lenin. He kept his pulse so accurately not only on the voiced but even on the unspoken desire of the masses that he was able to guide an entire revolution in its most critical period and through most unfavourable circumstances to a successful consummation. With him, we reach the stage not only of observing society but also of changing it. But if I go any further into his achievements, I shall be preaching Bolshevism in the sacred precincts of Bombay House and so must stop here.

Chapter 4
A Report to the JRD Tata Trust

In 1945, shortly before he joined the TIFR, Kosambi received a grant of Rs. 1800 for a 6-month period from the Tata Trust. His "completion report," reproduced below, reveals another side of DDK and the multiple scholarly interests that he pursued in parallel.

There are six projects described here. The first deals with his manuscript on path geometry; *this was eventually submitted to Marston Morse at the Institute for Advanced Study in Princeton, but was never published. The second is the Kosmagraph project that probably owes its genesis to the ideas on computing machines that DDK outlined in* [DDK36]. *Although initiated around the same time as the ENIAC project in the USA, this was on a far smaller scale and far less successful; it is not clear if a working model was ever actually tested. DDK's interest in computing machines stayed alive for some years* [36] *and in the 1960s TIFR would eventually construct the TIFRAC, an indigenous calculating machine. The* joint paper *mentioned in the report was almost surely never published. Projects 3 and 4 dealt with Kosambi's interest in applying statistics to real data, and Project 5 reflects his concerns with social issues. The final Project 6 was a consequence of his learning Sanskrit in order to "give an opinion upon points concerning ancient Indian mathematics." DDK undertook a translation of* Bhartrhari, *and this resulted in two works,* The Satakatrayam of Bhartrhari *and* The Southern Archetype of Epigrams Ascribed to Bhartrhari [60].

On May 10, 1945, I received from the JRD Tata Trust the sum of Rs. 1800 as a research grant, for 6 months from April 1945. A word about my situation at that time and the use I have made or propose to make of the money may clarify the position and enable the Trustees to judge whether or not the grant is being utilized as they had originally intended.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_4

At the end of 1944, I had (apart from my usual small mathematical research papers which need only paper, pencil, leisure, and postage to publish) the following major projects in hand: (1) The preparation of a book on path geometry. (2) The design and construction of a universal calculating machine, called the Kosmagraph. (3) Statistical analysis of cancer and tuberculosis data. (4) Statistical work on ancient Indian coinage. (5) Investigation of living conditions among the poorest class of workers (Manga) in Poona and other accessible cities of India. (6) The edition of poetry ascribed to a Sanskrit poet *Bhartrhari*.

From 1945, however, most of these projects had to be kept in abeyance for I had had ten attacks of fever from August 1944 to May 1945. These, with the stomach trouble and neuralgia which I have had almost continuously since 1935, made it difficult for me to do much mechanical work alone.

Hiring assistants was out of the question, as I had run completely out of funds. My trained students had necessarily to take the good jobs offered them, and at the time the grant was received, I had only one trained assistant whom I could afford to pay: a Sanskritist Sastri who helped with Project 6 on a part-time basis. The one bright student left was Mr. A.B. Magdum, who has suffered from recurrent attacks of appendicitis (and chronic poverty), and who had, to leave for Sangli two days before the grant was actually received. He has not been able to work for more than 1 month since May 1945.

The first attempted use of the grant, therefore, was in helping able mathematicians who were in great need. The offer was accordingly made to Dr. S. Minakshisundaram of Andhra University and Mr. K. Chandrasekharan of Madras. They could have done research at Poona for the summer months with profit to all, particularly as Prof. H.C. Chow, the Tata Public Relations scholar from China was expected to come to Poona to work with me. But by May 10, Minakshisundaram was under medical treatment at Madras, unfit to travel, Chandrasekharan had to await the results of his attempt to secure continuation of his temporary lectureship at the Presidency College, Madras. Finally, Prof. Chow was recalled suddenly to China, never having been able to do any work at Poona. Just about this time, I was appointed to a professorship at the newly founded Tata Institute of Fundamental Research. This gave me mental as well as financial relief, as my relations with the Fergusson College authorities had not been of the happiest. The actual situation as regards the six research projects I have mentioned is as follows:

1. The book on path-spaces is completed in typescript, though minor additions and revisions are inevitable before going to press. The assistance of Mr. V. Seetharaman, a former research student of mine now a lecturer at Annamalai University, would have been very useful, enabling me to round out certain results in the book and to take his doctorate. But he was doubtful about the wishes of God and his vice-chancellor in this matter; as I have no influence with either of the two, I had to finish up the work as best I could and leave him to his own devices. Very little money was actually used for the book, and that came from the generous invitation tendered by Prof. H.J. Bhabha, then at Bangalore, to lecture on the topic. As for the publication, I believe no funds will be required, unless we decide to publish

the work locally; there is a good chance of getting the work published in the USA or in England, the only case for local publication being that the book would then be more easily available to our students.

2. The Kosmagraph is finished, and a working model being improved at St. Xavier's College. The total outlay for workshop charges, electric motors, cathode-ray oscillographs, valve tubes, etc., would have exceeded the total amount of the Tata grant. But the St. Xavier's authorities stood the expense of these items, as Fr. Rafael has collaborated in the work. My total expenses from the grant have been a nominal honorarium of Rs. 250/- to K.B. McCabe, the third collaborator, and another of Rs. 50 to Salvador D'Souza, head mechanic at the St. Xavier's workshop. Both have deserved far more, end the work of McCabe in particular seems to me to be beyond recompense.

   A joint paper is being made ready for publication, though it will be some months before all the points are checked.
3. Prof. R.A. Fisher thought that my ideas on blood groups and cancer were worth following through, but I have been unable to get the necessary technical collaboration. The Tata Memorial Hospital staff have their hands full, and in addition, they have the aid of the statisticians for their own experiments, so that my function does not go beyond an occasional consultation, which costs me nothing. Tuberculosis data is expected at any time from UMT Sanatorium, Arogyavaram, which had asked me to work out a better index for the measurement for TB than the one now in use. I did give them a new provisional index, but that was on the basis of data from 50 patients from group III, whereas at least ten times the number would be really necessary. Some portion of the grant has been earmarked for this purpose, as I shall need the services of a tabulator, probably Mr. Magdum, and in addition shall have to travel to Arogyavaram again to see the index applied in practice. I have paid Rs. 150/- to Dr. J. Frimodt-Møller of the UMT Sanatorium.
4. For the work on ancient Indian coinage, I hoped to be able to borrow the services of Dr. S. Paramasivan, Archaeological Chemist at the Government Museum, Madras, for 4 to 6 months. This has proved to be impossible, and all work done in this connection has been done by me. This is mostly the work of classifying and weighing coins accurately, preliminary to applying modern statistical methods to the data, and I have concentrated on hoards of punch-marked coins. The analysis of 690 coins of the Khandesh Hoard with the BBRAS is almost complete and bears out my former (published) conclusions about such coins. The most substantial new discovery has been that of an entirely new type of punch-marked coin, and I hope to publish a paper on this find shortly. It may be of interest to note that the same hoard was first studied by an expert of the Prince of Wales Museum, who was able to classify only 218 of the 690 coins, not always correctly; the new discovery belongs to his "unclassified" group, which would ordinarily have been scattered as of no importance.
5. Working class conditions in Poona are being studied with the assistance of Messrs. T.G. Kulkarni, A.T. Patil, A.D. Taskar—three former students of mine now studying for the MSc degree at Bombay University. In addition, I hire occasional tabulators to help in the work. The data was available from figures collected by Prof.

S.R. Bhagwat, well known for his work on adult education, civic and rural uplift. But tabulation and checking have taken Rs. 450/- over 5 months, and another couple of hundred rupees at least should be necessary before the work can be completed. This money is divided between the students and tabulators, all of whom work very hard. So far, food figures for 200 of the 300 families I expect to study have been tabulated and checked. The rest will be covered in a month, after which we shall start field investigation (from home to home) of the same families. The conclusion so far reached is that the lack of purchasing power is more serious than any real shortage in rationing among the Manga workers, but this will be amplified later. We hope to publish the findings in detail.

6. Work on the edition of *Bhartrhari* was taken up as a sideline for two reasons: (A) I have often been asked to give an opinion upon points concerning ancient Indian mathematics, which I could not do without studying Sanskrit. I also find that the editions now extant of such technical works are not satisfactory, so that I should need general editorial experience before doing anything in ancient mathematical texts. (B) The particular problem of *Bhartrhari* was taken up at the suggestion of the late Dr. V.S. Sukthankar, famed for his great (though incomplete) edition of the Mahabharata. Dr. Sukthankar believed that I could edit *Bhartrhari* and that the work would be of value. As a matter of fact, his methods, and even more my own, have had to be based upon good statistical practice, stratified sampling. For this work, I need one copyist at least (at present Mr K.V. Krishnamoorthi), and one or more people to write the critical apparatus (now being written by D.V. Navarane). The expenses of borrowing, comparing, and copying MSS with checking collation sheets have been heavy, about 805 rupees, but results seen to have justified the outlay. One edition based on two MSS and a commentary has already appeared at the Anandāśrama, the total outlay for this would not exceed Rs. 75/-, and help from the JRD Tata Trust has been acknowledged in the Preface. A second edition of a different recension has gone to press under the auspices of the Bharatiya Vidya Bhavan, Bombay. This is costlier, as it is based upon 22 sources with four commentaries. The final *editio princeps* (based on 42 MSS plus 14 commentaries and study of another 150 MSS) should take another year to prepare, though the main task of collecting and analyzing sources is over, even the final stage of writing the variants down and preparing a press copy being half finished.

   In this connection, I have sent for publication (before the grant was thought of) a statistical paper (accepted by the Journal of the American Oriental Society) on Sanskrit studies and hope to have another of a slightly different type on the comparative (numerical) study of *Bhartrhari* versions, of which I have been able to establish eight, with not less than four others which certainly exist, but cannot be established on the basis of available numerical data, which is derived from the close study of about 250 MSS from all over the world. A third paper on the authorship of the *Satakas* has been sent for publication to the J. Or. Research, Madras, by invitation.

   As for Indian mathematical texts, I hope to edit the Aryābhatiya, which should take another 2 years of work. The India Office Librarian has already sent me MSS of this work, and I have others at my disposal here. My copyists will undertake this as soon as the essentials of the *Bhartrhari* work are dealt with.

Abstract of Expenditure.

Project	(1)	Nil
"	(2)	250./- + 50./-
"	(3)	150./-
"	(4)	Nil
"	(5)	450./-
"	(6)	805./-
Total		1705./-

MV/10-10-1945.

Part I
References

1. J.D. Bernal, Nature **211**, 1024 (1966).
2. The bibliography that now appears on pages xv–xix of this volume is a listing of the complete set of the papers of DDK that are of a mathematical nature. The list has been compiled in part from incomplete sources in the biography by Chintamani Deshmukh [DDK-JK] as well as web listings. In addition to the papers listed, many of his essays relate to scientific issues, but these are not included here.
3. *The Oxford India Kosambi*, ed. by B.D. Chattopadhyaya (Oxford University Press, New Delhi, 2009); *Combined Methods in Indology & Other Writings: Collected Essays*, D.D. Kosambi, Compiled, edited and introduced by Brajadulal Chattopadhyaya (Oxford University Press, New Delhi, 2005); *Indian Numismatics*, ed. by D.D. Kosambi (Orient Longman, Hyderabad, 1981); *Exasperating Essays*, ed. by D.D. Kosambi, (Peoples Publishing House, New Delhi, 1957).
4. *The many careers of D.D. Kosambi: Critical essays*, ed. by D.N. Jha (Leftword, Delhi, 2011); *Damodar Dharmanand Kosambi* (in Hindi), ed. by R.S. Sharma (SAHMAT, New Delhi 2010).
5. R. Ramaswamy, Integrating mathematics and history: the scholarship of D.D. Kosambi. Econ. Polit. Weekly **47**, 58–62 (2012). Reproduced in [35].
6. S.G. Dani, Kosambi the Mathematician. Reson. J. Sci. Educ. 514–528 (June 2011). This issue of the journal is dedicated to D.D. Kosambi and contains several articles that discuss the scientific contributions of DDK as well as two essays on his life and historical work.
7. R. Narasimha, Kosambi and proper orthogonal decomposition. Reson. J. Sci. Educ. 574–581 (June 2011).
8. C.K. Raju, Kosambi the Mathematician. Econ. Polit. Weekly **54**, 38 (2009).
9. P. Antonelli, R. Ingarden, M. Matsumoto, *The Theory of Sprays and Finsler Spaces with Applications in Physics and Biology* (Kluwer Academic Publishers, Amsterdam, 1993).
10. K. Karhunen, Über lineare Methoden in der Wahrscheinlichkeitsrechnung. Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys. **37**, 1–79 (1947); M. Loève, Fonctions aleatoires de seconde ordre. C. R. Acad. Sci. **220**, 295 (1945) and related papers.
11. K.K. Vinod, Kosambi and the genetic mapping function. Reson. J. Sci. Educ. 540–50 (2011).
12. Starting with [DDK3], Kosambi developed the idea in a number of papers, including [DDK5, DDK6, DDK8] and [DDK18] and so on. In the 1950's he was on the editorial board of the Japanese journal, Tensor (New Series) wherein he published [DDK55], possibly his final paper on the topic.
13. 'Atomic Energy for India', the text of a talk by DDK to the Rotary Club of Poona, on July 25, 1960 was published in the posthumous volume, *Science, Society, and Peace* (The Academy of Political and Social Studies, Pune, 1967, reprinted by People's Publishing House, 1995).
14. A. Weil, (1992). *The Apprenticeship of a Mathematician* (Birkhäuser, Basel).

15. T. Vijayaraghavan (1902–1955) was a Founding Fellow of the IASc, being elected in 1934. He did his Ph.D. under the supervision of G. H. Hardy in Cambridge. From Dacca he moved to Waltair, and eventually became the founding director of the Ramanujan Institute of Mathematics in Madras.
16. The number theorist S. Chowla (1907–1995) moved to the US in 1947 after a career at Delhi, Benaras, Waltair and Lahore in undivided India.
17. In 1940, Weil was in military prison in Bonne-Nouvelle for refusing to take part in the war as a conscientious objector (since his true *dharma* was the pursuit of mathematics and not war, he said) when he proved an analogue of the Riemann hypothesis (for the zeta function of curves over finite fields). He did discuss the Riemann hypothesis with T. Vijayaraghavan, who is supposed to have said that if he could have six months—undisturbed and undistracted—in a prison, he could have a crack at solving the RH. See Ref. [14], and M. Raynaud, André Weil and the foundations of algebraic geometry. Notices of the AMS, **46**, 864 (1999).
18. The historical spellings of some city names have been retained.
19. S.S. Chern, Bulletin des Sciences Mathematiques **63**, 206.
20. In a letter to Bhabha on 8th July 1946 (TIFR Archives, D-2004-387-5) DDK says, "Of the Chinese and more particularly of our Visiting Professor Chern there is no news; the difficulty here is unquestionably that of permission to leave China and passports. A letter from Jawaharlal Nehru would have helped, and in fact he has written that he is in full sympathy with my project; but he can't do anything further [with important work keeping] him much too busy for lesser affairs like ours. However, I have every hope of getting co-operation from him as well as from China—in due course".
21. The Kosambi–Bhabha correspondence has been made available through the kind courtesy of the TIFR archives. There are a large number of letters that are presently being catalogued and edited. The letter of 8 July, 1946 (TIFR Archives, D-2004-387-5) and 21 November, 1946 (TIFR Archives, D-2004-387-10) were both written while Bhabha was in England, and DDK was the Acting Director of the TIFR. Some letters have been reproduced in [35].
22. In 2010, when Louise J. (Mrs. Marston) Morse was nearly 100 years old, the Morse archives were searched one last time. However, she was not able to locate this manuscript or any reference to it.
23. Of the 110 or so Fellows appointed in 1934 and 1935, about two thirds were from the south of India or worked there and Raman might have had greater familiarity with their work or their reputation.
24. The University of Madras announced the Ramanujan Memorial Prize for "the best thesis based on original contributions submitted by an Indian (or one domiciled in India) on some definite branch of mathematics, applied or pure" in 1933. The prize was awarded in 1934, as reported in Nature **135**, 28–28 (1935).
25. 'A Chapter In The History Of Indian Science', an unpublished essay by DDK is a damning indictment of Raman's role in suppressing creativity in Indian science. The first half of the essay was published anonymously in People's War (the Organ of the Communist Party of India) on 22 July 1945 [51].
26. K.A.N., Metrology of punch-marked coins. Curr. Sci. **7**, 345–346 (1941). This might have been the historian of South India, K.A. Nilakantha Sastry (R. Thapar, Private Communication).
27. D.D. Kosambi, A Note on two hoards of punch marked coins found at Taxila. New India Antiquary **3**, 156–157 (1940) and On the study and metrology of silver punch marked coins. New India Antiquary **4**, 1–35 and 49–76 (1941).
28. 'Adventure into the Unknown', in *Current Trends in Indian Philosophy*, ed. by K. Satchidananda Murty, K. Ramakrishna Rao (Asia Publishing House, Bombay, 1972).

29. H.M. Edwards, *Riemann's Zeta Function* (Academic Press, New York, 1974). I am grateful to Prof. S.K. Dani for several technical suggestions regarding this section.
30. These results need elementary methods of complex analysis that can be found in standard textbooks.
31. M.V. Berry, Private Communication. In November 1960, Carl Siegel sent a very negative review of the paper in confidence to Professor Chandrasekharan at the TIFR, and this letter (TIFR Archives, D-2004-00387-145) also quotes the opinion of A. Selberg. Both mathematicians were then at the Institute for Advanced Study, Princeton.
32. M. du Sautoy, *The Music of the Primes* (Harper Perennial, London, 2004); K. Sabbagh, *Dr. Riemann's Zeros* (Atlantic Books, NY, 2003); J. Derbyshire, *Prime Obsession: Bernhard Riemann and the Greatest Unsolved Problem in Mathematics* (Plume Books, New York, 2004); D. Rockmore, *Stalking the Riemann Hypothesis* (Vintage, New York, 2005).
33. A. Odlyzko, Private Communication.
34. I. Chowdhury, Fundamental research self-reliance and internationalism: the evolution of the TIFR, 1945–1947, in *Science and Modern India: An Institutional History, c.1784-1947*, Project of History of Science, Philosophy and Culture in Indian Civilization, vol. XV, Part 4, ed. by Uma Das Gupta (Pearson, New Delhi, 2010). See also I. Chowdhury, A. Dasgupta, *A Masterful Spirit: Homi J. Bhabha (1909-1966)* (Penguin Books India, 2010). A copy of the letter sent by Bhabha to DDK on 18th May, 1945, the same day that the Council met, is available in the TIFR Archives, D-2004-00387-1.
35. Meera Kosambi (ed.), *Unsettling the Past: Unknown Aspects and Scholarly Assessments of D. D. Kosambi* (Permanent Black, New Delhi, 2012).
36. I. Chowdhury, D.D. Kosambi: A Most Unusual Scholar. Mid Day (Bangalore), 11 Feb 2008.
37. Professor M. Raizuddin Siddiqi who originally headed the committee decided to emigrate to Pakistan to take up a position in Peshawar at that time. He indicated to the Policy Committee of the AMS that Kosambi should take his place.
38. The original delegation for the meeting that was held from August 27–29, 1950 in New York was to have consisted of Profs. M.R. Siddiqi, D.D. Kosambi and T. Vijayaraghavan, none of whom could attend (see the several letters in the TIFR Archives, D-2004-389-2 onwards). In addition to Chandrasekharan and Minakshisundaram, Prof. S.S. Pillai of Madras University, who was going to the Institute for Advanced Study in Princeton was also deputed to attend the conference. Tragically, he left too late to attend the meeting, taking the TWA Flight 903 from Bombay to New York that crashed outside Cairo on 31 August, killing all on board. An Interim Committee of the IMU was set up at this meeting, and DDK was invited to serve on it for a period of two years.
39. In addition to K. Chandrasekharan and K.G. Ramanathan, the other two signatories were M.S. Narasimhan and C.S. Seshadri. This letter (TIFR Archives, D-2004-00390-4) dated 23 August, 1960, alleges that Kosambi, in correspondence with eminent mathematicians outside India, in addition to the RH also claimed a proof of Fermat's last theorem (FLT), namely that the equation $x^n + y^n = z^n$ has no solutions with x, y, z being positive integers, if n is larger than 2. This is odd; other than in some notes he made in 1926 when he was an undergraduate at Harvard and in the epilogue of the essay "Adventure into the Unknown", there is no mention of the FLT elsewhere among Kosambi's other papers. This theorem was proved conclusively only in 1995 by Andrew Wiles (see Simon Singh, *Fermat's Last Theorem* (Fourth Estate, London, 2002)).
40. R. Ramaswamy, *Adventures into the Unknown: Essays by D.D. Kosambi* (Three Essays Collective, Gurgaon, 2016).
41. The essay 'On Statistics' is Chapter 3 of this volume. See the concluding line: *But if I go any further into his achievements, I shall be preaching Bolshevism in the sacred portals of Bombay House and so must stop here.*

42. As noted by Robert Anderson in *Nucleus and Nation: Scientists, International Networks, and Power in India* (Chicago University Press, 2010), some of Kosambi's "Marxist way of seeing things" appealed to Homi Bhabha in the initial days of their association.
43. This is one of three essays on solar energy that were first reprinted in *Science, Society and Peace* (The Academy of Political and Social Studies, Pune, 1986), as well as now in [35].
44. The history of the Bourbaki collective has been written about extensively by Maurice Mashaal, *Bourbaki: A Secret Society of Mathematicians* (American Mathematical Society, Providence, 2006) as well as others, including Liliane Beaulieu [49]. While the eventual name chosen by the group was Nicolas, the original Kosambi paper [DDK2] cites D. Bourbaki.
45. L. Beaulieu, Bourbaki's art of memory. Osiris **14**, 291 (1999).
46. See http://tinyurl.com/q2medrh and the linked pages.
47. N. Bourbaki, *Éléments de Mathématique, Book 1: Théorie des ensembles: Fascicule de Resultats* (Hermann, Paris, 1939).
48. R.P. Boas Jr., Bourbaki and me. Math. Intell. **8**, 84 (1986).
49. L. Beaulieu, *Nicolas Bourbaki: History and Legend, 1934-1956* (Springer, Berlin, 2006).
50. D.D. Kosambi, The function of leadership in a mass movement; The Cawnpore Road. Fergusson Coll. Mag. 1–7 (1939). Ahriman is the destructive spirit in Zoroastrian mythology.
51. 'The Raman Effect', Peoples War, 22 July 1945, by An Indian Scientist. An editorial note adds: The writer of this article who prefers to be anonymous is a versatile Indian Scientist whose original work in Mathematics is well-appreciated in foreign countries especially in America and Great Britain. We hope to be able to publish many more articles like this on popular science subjects from his pen. – Editor.
52. D.D. Kosambi (with Miss Sushila Gokhale), 'Progress in the production and consumption of textile goods in India. J. Indian Merchants' Chamber (Bombay), January, pp. 11–15 (1943). There is also an unpublished essay, 'Notes on the Marxian Theory of Value', where DDK signs off as Vidyārthi.
53. As informed by Meera Kosambi. However, Divyabhanusinh Chavda, who was a student of DDK's at this time, maintains that according to DDK, the S was for 'Stupid'.
54. D. D. Kosambi, *Prime Numbers*. The manuscript of this book, that was apparently mailed to his publishers shortly before DDK's death in June 1966, has not been traced.
55. H.W.O. Pétard, A contribution to the mathematical theory of big game hunting. Am. Math. Monthly **45**, 446 (1938).
56. In 1958 DDK authored an article in collaboration with U.V.R. Rao, analysing the statistical defects underlying para-psychological experiments [DDK58]. This paper was subsequently commented upon by A.W. Joseph who pointed out an error in analysis as well as in the conclusions, ending with "The above comments do not detract from the valuable experiments in card–shuffling made by the authors, but it is suggested that there is little weight left in their criticism of the ESP investigations.". See A.W. Joseph, A note on the paper by D.D. Kosambi and U.V. Ramamohan Rao on 'The efficiency of randomization by card–shuffling'. J. R. Soc. Stat. **122**, 373–74 (1959).
57. In 'Artless innocents and ivory-tower sophisticates: Some personalities on the Indian mathematical scene'. Curr. Sci. **85**, 526–537 (2003), M.S. Raghunathan recalls a conversation with André Weil in 1966 or 1967, when he (Weil) says of DDK, "… Let me tell you this: he was one of the finest intellects to come out of your country." In his autobiography [14], Weil has this to say: "I appointed Kosambi for the following year. He was a young man with an original turn of mind, fresh from Harvard where he had begun to take an interest in differential geometry. I had met him in Benares (now Varanasi) where he had found a temporary position". Weil was a little over a year older than DDK.
58. G.D. Birkhoff, Mathematics at Harvard in the 1940's. Proc. Am. Philos. Soc. **137**, 268–272 (1993).

59. ‘Steps in Science’, in *Science and Human Progress: Essays in Honour of Late Prof. D.D. Kosambi, Scientist, Indologist, and Humanist* (Popular Prakashan, Mumbai, 1974).
60. DDK edited the following three books on the poetry of *Bhartrhari*: (a) *The Satakatrayam of Bhartrhari with the Comm. of Ramarsi*, ed. by D.D. Kosambi, K.V. Krishnamoorthi Sharma (Anandasrama Sanskrit Series, No.127, Poona, 1945), (b) *The Southern Archetype of Epigrams Ascribed to Bhartrhari* (Bharatiya Vidya Series 9, Bombay, 1946) and (c) *The Epigrams Attributed to Bhartrhari* (Singhi Jain Series 23, Bombay, 1948).

Part II
Select Publications of D.D. Kosambi

Chapter 5
Precessions of an Elliptical Orbit

D.D. Kosambi, Banaras Hindu University

This was the first of Kosambi's published papers and is a largely pedagogical exercise started while he was a student at Harvard, and polished up after his return to India. In this early work, DDK displays a sophisticated ability to integrate many strands of thought, to generalize observations from one context to another, while retaining sufficient rigor. The mention of quasiperiodic motion suggests that some of these topics were possibly covered in the special course on the many-body problem that was given by the mathematician G.D. Birkhoff. Kosambi's biographer reports [DDK-JK] *that Birkhoff counted DDK among the better students at Harvard and allowed him to take this course.*

(Notes on: Vibrating Strings; Planetary Orbits; The Raman Effect.)

I. Any textbook of hydrodynamics will give the following equations for the motion of an infinite cylinder through a perfect incompressible fluid, at rest at infinity:

$$\begin{aligned}\left(M+M'\right)\dot{\xi}+k\rho\eta &= X \qquad (5.1)\\ \left(M+M'\right)\ddot{\eta}-k\rho\dot{\xi} &= Y\,.\end{aligned}$$

$$\begin{aligned}\left(M+M'\right)\frac{dU}{dt} &= P \qquad (5.2)\\ \left(M+M'\right)U\frac{d\psi}{dt} &= k\rho U+Q\,.\end{aligned}$$

Here, M is the mass of an unit length of the cylinder, M' that of the fluid displaced thereby; ξ, η coordinates of the central axis with respect to axes fixed in space, the whole motion being in a direction perpendicular to the length of the cylinder. The

Published in the Indian Journal of Physics **5**, 359–64 (1930). Reprinted with permission.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_5

constant of circulation about the body is denoted by k; density of the fluid medium by ρ; components of the external forces per unit length by X, Y. In Eq. (5.2), P, Q are components of the external force along the tangent and normal to the path of the center; ψ the angle made by the direction of the velocity U, with a fixed direction. (cf. Lamb, Hydrodynamics, 5th ed., p. 76.)

The important characteristic of these equations is that the total energy of the motion represented is exactly the same as when there is no circulation ($k = 0$). In fact, the force of circulation is perpendicular to the velocity, and so does no work. These equations may be made also to represent the case of an electron in a magnetic field, a Foucault pendulum, and even the restricted problem of three bodies, as by the equations of Hill. We processed to consider special cases.

II. If a vibrating string be set in motion by plucking it in the middle, most of its motion will be represented by

$$y = A \sin bx \sin bct. \tag{5.3}$$

Seen lengthwise, the string is approximately our infinite cylinder attracted to the center with a force proportional directly to the distance. Due to natural unevenness of the apparatus, the actual path will be a flat ellipse rather than a straight line, and so circulation of the air will be set up. Our Eq. (5.1) become:

$$\left.\begin{array}{l} \ddot{\xi} + 2\nu\dot{\eta} + \lambda^2\xi = 0 \\ \ddot{\eta} - 2\nu\dot{\xi} + \lambda^2\eta = 0 \end{array}\right\} \quad \nu = \frac{k\rho}{2(M + M')}, \quad \lambda^2 = \frac{\text{force at unit distance}}{(M + M')}. \tag{5.4}$$

These may be formally integrated by an ingenious device due to Bronwich (Proceedings of the London Math. Soc. Series 2, Vol. XIII, p. 225). Multiplying the second by $i = \sqrt{-1}$ and adding to the first, we have:

$$\ddot{z} - 2\nu\lambda\dot{z} + \lambda^2 z = 0\,, \quad z = \xi + i\eta\,, \quad i = \sqrt{-1} \tag{5.4a}$$

$$\text{whence} \qquad z = e^{\nu i t}\left(Ae^{ipt} + Be^{-ipt}\right) \tag{5.5}$$

where $p = \sqrt{\nu^2 + \lambda^2}$.
With the initial conditions $z' = 0$, $z = a$, when $t = 0$, the motion follows an hypocycloid tangent to $|z| = \frac{a\nu}{p}$ with cusps on the circle $|z| = a$.

With another set of initial conditions, say $\xi_0 = \eta_0 = \eta_0 = 0\ \dot{\xi}_0 = \nu$, when $t = 0$, we obtain:

$$z = \frac{\nu}{p} e^{i\nu t} \sin pt \tag{5.6}$$

In polar coordinates, $z = re^{i\theta}$

$$\theta = \nu t\,, \qquad r = \frac{\nu}{p}\sin pt \quad \text{or} \quad r = \frac{\nu}{p}\sin\frac{p\theta}{\nu} \tag{5.6a}$$

The path is thus a rosette, described by a casual observer as a rotating ellipse, as also in the gyroscopic pendulum. We see that a new period has appeared, that of precession:

$$T = 2\frac{\pi}{\nu} = \frac{4\pi(M+M')}{k\rho} \tag{5.7}$$

Actually if one tightens the heavy string of a banjo and plucks it in the middle, the whole motion seems a blurred region to naked eye, and its boundaries, instead of narrowing down uniformly to rest because of air resistance, are seen expanding again, after a little while. The period of this expansion would be just a half of the above. An ink dot on the string apparently follows the "rotating ellipse," and the period of a full rotation would then be T. An experiment seems to be called for with brilliant points and photographic observations.

III. The theory of relativity has still to account for the observed precessions of perihelion in our planets, especially Venus. We can consider the atmosphere of the planet and other light surrounding debris such as the rings of Saturn as constituting a circulatory effect with respect to the space occupied by the planet, since the rotation of the planet is sure to set this matter into motion. Unfortunately, the infinite cylinder cannot yield numerical results applicable to the planetary case, and the three-dimensional analysis presents difficulties. However, a criterion may be obtained as to the direction of rotation of the planet, i.e., as to the circulation setup. Notice in passing that Venus is known to have a dense atmosphere though the question of its period of rotation about its axis is still a point of some doubt.

Instead of taking Eq. (5.1) with attraction inversely as the square of the distance, we assume the divergence from the normal state to be small and apply Eq. (5.2).

$$\text{Let} \quad \left.\begin{aligned} U &= U_1 + U_2 \\ \psi &= \psi_1 + \epsilon \end{aligned}\right\} U_1,\, \psi_1,\ \text{as in the Keplerian condition.} \tag{5.8}$$

then

$$(M+M')\frac{dU}{dt} = P = (M+M')\frac{dU_1}{dt} \qquad\qquad \therefore U_2 = 0$$

and

$$(M+M')U\frac{d\psi}{dt} = k\rho U + Q$$

and

$$(M + M')U_1 \frac{d\psi}{dt} = Q$$

whence

$$\frac{de}{dt} = \frac{k\rho}{M + M'} \quad \text{and} \quad e = \int_0^t \frac{K\rho dt}{M + M'}$$

which leads to the result that ϵ in a complete period increases by $\frac{k\rho T}{M+M'}$. If the direction from which ψ is measured be taken as the radius vector to the perihelion, the orbit being nearly circular, the perihelion is retarded by this amount if k be positive. The actual direction of procession depends only on the direction of circulation. An advance would signify that the directions of circulation and of description of the orbit were opposite; a retardation would mean that they were identical. Finally, this precession when small might also be taken as a slight change of period of the moving body: advance of perihelion as an increase of period, retardation as a decrease. In the string, the circulation is set up by the vibration itself, and a retardation should always occur.

IV. The solar system is but a step from the atomic model. The two-dimensional fluid motion of our cylinder, of a vortex, of an electric charge in an electromagnetic field is indistinguishable. Thus, the change of period just pointed out in the case of a planet might well be seen as an extra line of the spectrum, if all the electrons have a sole direction in describing the orbit; as two lines, symmetrically displaced from the ordinary line, if the constituent electrons have both senses of description. And indeed, this is the usual explanation of the well-known Zeeman effect, after Larmor. In the Raman effect, however, a number of electrons are excited by a light wave and send out a certain number of extra waves not yet satisfactorily explained (the Smekal jump is most unattractive). We might extend our analysis, and replace the constant of circulation k by a periodic function of the time which represents the change of electromagnetic intensity. The phenomenon and the problem will not be discussed in the present note, though I hope a suggestion will not be ill received. Quantization and the critical value method of Schrödinger should be used to obtain the proper numbers corresponding to the new periods caused by the disturbing function. Secondly, the asymmetry of the Raman effect has also to be considered. The Zeeman effect might be observed, for purposes of comparison, in a magnetic field wherein the intensity has a period comparable to that of the light waves used in the more recent discovery. Let it finally be noticed that while electrodynamically unsound models have usually been employed for purposes of illustration and deduction, equations of our type will be applicable wherever a stable periodic motion of any given system is acted on by "non-energic" forces, here forces perpendicular to the displacement. As for the mathematical justification of the assumption that in an arbitrary system, a number of stable recurrent—if not periodic—motions exist, and the reader is referred to modern dynamical theorists, such as Birkhoff.

HARVARD UNIVERSITY, SEPTEMBER, 1927.
BANARAS HINDU UNIVERSITY, APRIL, 1930.

Note: The Kármán solution involving rotational motion of the fluid about the infinite cylinder has been neglected; it should most certainly be taken into account in any experiments with vibrating strings.

Chapter 6
On a Generalization of the Second Theorem of Bourbaki

D.D. Kosambi, The Muslim University, Aligarh

The name of Bourbaki first appears in the published literature in this paper. As described by André Weil, the prank was designed to deflate a senior colleague's ego, presumably by demonstrating the greater familiarity that Kosambi had with modern methods and the then current literature. The rest of the paper is serious enough and points to Kosambi's interests in differential geometry which were to preoccupy him for the next twenty years. Schnirelmann's work continues to be of interest.

In a paper under publication [1], I have discussed the existence of covariant derivatives and proved that there are infinitely many parallelisms connected with the paths:

$$\ddot{x}^i + \alpha^i(x, \dot{x}, t) = 0 \tag{6.1}$$

These parallelisms are defined by

$$D(u)^i = \dot{u}^i + \gamma^i_k u^k + \epsilon^i \tag{6.2}$$

where

$$\dot{x}^k \gamma^i_k(x, \dot{x}, t) + \epsilon^i(x, \dot{x}, t) = \alpha^i$$

One of these, for which

$$\gamma^i_k = \frac{1}{2}\alpha^i_{;k} \tag{6.3}$$

Published in the Bulletin of the Academy of Sciences, U.P. **1**, 145–47 (1931). Reprinted with permission of the National Academy of Sciences, India. The journal is also known in the literature as the Bulletin of the Academy of Sciences of the United Provinces of Agra and Oudh, Allahabad, India.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_6

is the fundamental parallelism; with this, a covariant derivative independent of the direction exists only for the symmetric affine connection:

$$\alpha^i = \Gamma^i_{jr}\dot{x}^j\,\dot{x}^k\,, \qquad \Gamma^i_{jr} = \Gamma^i_{rj} \tag{6.4}$$

I was not aware that a little-known Russian author, D. Bourbaki, who died of acute lead poisoning during the revolution, had anticipated part of these results and pointed out a way to their extension. I shall not go into details here, for an excellent résumé and critique has been published recently by Lusternik and Schnirelmann [2]. But it will be clear to geometers acquainted with last-named paper that I merely proceed by discarding all three of the '*Vysokoblagodaren*' axioms. With our notations, this means that a vector field $u^i(x)$ will have a covariant derivative $u^i_{|r}$ independent of direction, such that:

$$u^i_{|r}\dot{x}^r = D(u)^i \tag{6.5}$$

We have, therefore:

$$u^i_{lr} = \frac{\partial u^i}{\partial x^r} + \gamma^i_{kr}u^k + \epsilon^i_r \tag{6.6}$$

where $\quad \gamma^i_{kr}\dot{x}^r = \gamma^i_k$ and $\epsilon^i_r\dot{x}^r = \epsilon^i$.

It follows, with the notation of my first paper, that:

$$\begin{aligned} \gamma^i_{kr} &= \gamma^i_{k;r} \quad \text{independent of } \dot{x} \\ \epsilon^i_r &= \epsilon^i_{;r} \end{aligned}$$

That is:

$$\alpha^i_{;r} - \dot{x}^r\left[\gamma^i_{kr} + \gamma^i_{kr}\right] = \phi^i_r(x)$$

Thus for the most general α^i, we can have at best:

$$\alpha^i = \gamma^i_{kr}\,\dot{x}^k\dot{x}^r + \phi^i_r\dot{x}^r \tag{6.7}$$

For the principal parallelism, $\alpha^i_{;k} = 2\gamma^i_k$. This gives my former result. For the general γ^i_k, linear in $\dot{x}$, (6.7) gives the most general form of the α's and hence of the paths.

It will be noted that the $\phi^i_r(x)$ are precisely the ϵ^i_r. Furthermore, an important consequence of this generalization is the inclusion of Cartan's torsion, which is given by:

$$\Omega^i_{kr} = \gamma^i_{kr} - \gamma^i_{rk} = \gamma^i_{k;r} - \gamma^i_{r;k} \tag{6.8}$$

The second is the most general form of torsion, for all possible parallelisms. The quantities

$$\epsilon^i_r = \alpha^i_{;r} - \left[\gamma^i_{kr} + \gamma^i_{rk}\right]\dot{x}^k$$

are used in the new unitary field theories to denote the electromagnetic components of the forces deforming the hyperspace E_4.

References

1. D.D. Kosambi, Modern differential geometries. Ind. J. Phys. (to appear).
2. *Topologicheskie Metody v Variatsionnykh Zadachakh*. Math.-mech. Forschungsinstitut, Moskau, 1930, pp. 69–73. I am indebted to Dr. A. Weil for this important reference, and for permission to use his private reprint. I understand that Schnirelmann's work is shortly to be published in German, and this will undoubtedly fill a considerable gap in the existing literature. It is also highly desirable that Bourbaki's posthumous papers, at present lodged with the Leningrad Academy, should be published in full. Unofficial reports claim that Bourbaki was shot after the Minkhii Znak affair with other members of the 'Russko-Angliskii Slovar.'

Chapter 7
Parallelism and Path-Spaces

D.D. Kosambi, Fergusson College, Poona

After DDK returned to India, he kept up his contacts with well-known European mathematicians such as T. Levi-Civita and É. Cartan, who also communicated his papers to professional journals. This paper appears to have been sent to Élie Cartan prior to publication, and the ensuing correspondence resulted in this paper by Kosambi and a note by Cartan being published back-to-back in Zeitschrift (see the next Chapter). Along with a later paper by S.S. Chern in the Bulletin des Sciences Mathématiques, **63**, *206–212 (1939) these papers lay the foundations of the Kosambi-Cartan-Chern theory.*

1. This paper is devoted to the geometrical study of an arbitrary system of second-order differential equations of the form:

$$\ddot{x}^i + \alpha^i(x, \dot{x}, t) = 0 \qquad (i = 1, 2, \ldots, n). \tag{I}$$

The integral curves of (I) are assumed to be such that in some continuous n-dimensional region of the space (x^i) they possess the property of convexity: One and only one such curve—which we shall call *path*—passes through any two points of the region. The parameter t can be considered as an additional time-like coordinate, or an arbitrary arc-like parameter, the latter point of view being rarely stressed. Besides the tensor summation convention, I use the following notation for partial differentiation:

$$\frac{\partial A_{::}}{\partial x^k} = A_{::,k}\,, \quad \frac{\partial A_{::}}{\partial \dot{x}^k} = A_{::;k}\,.$$

Published in Mathematische Zeitschrift 37, 613–18 (1933).

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_7

In a previous memoir[1] (the knowledge of which is not assumed here) I attempted to investigate the possibility of deducing the system (I) from a variational principle

$$\delta \int f(x, \dot{x}, t) dt = 0 \tag{7.1}$$

this is equivalent to finding a metric for the path-space. It is seen at once that the integrand f must be a solution of

$$-\alpha^i f_{;i;j} + \dot{x}^i f_{,i;j} + \frac{\partial}{\partial t} f_{;j} - f_{,j} = 0\,. \tag{7.2}$$

These equations as they stand are tensor-invariant, but without simple or even geometrically interpretable compatibility conditions. If, however, an additional condition is imposed on the effect that f be a constant over the paths, i.e., $\frac{df}{dt} = 0$ the system (7.2) reduces at once to

$$\begin{aligned} -\alpha^i f_{;i} + \dot{x}^i f_{,i} + \frac{\partial f}{\partial t} &= 0\,, \\ \frac{1}{2} \alpha^i_{;j} f_{;i} - f_{,j} &= 0\,. \end{aligned} \tag{7.3}$$

The compatibility conditions for this first-order system are easily worked out, and if $f_{,k}$ be eliminated wherever they occur, we obtain the following differential invariants as coefficients in successive equations:

$$\begin{aligned} \epsilon^i &= \alpha^i - \frac{1}{2} \alpha^i_{;k} \dot{x}^k\,, \\ -2P^i_j &= \alpha^\sigma \alpha^i_{;\sigma;j} - \dot{x}^\sigma_{,\sigma;j} - \frac{1}{2} \alpha^\sigma_{;j} \alpha^i_{;\sigma} - \frac{\partial}{\partial t} \alpha^i_{;j} + 2\alpha^i_{,j}\,, \\ 3R^i_{jk} &= P^i_{j;k} - P^i_{k;j}\,. \end{aligned}$$

The ordinary Riemann-Christoffel curvature tensor is $R^i_{jk;l}$. M.E. Cartan[2] has been kind enough to point out that one differential invariant, namely, $\alpha^i_{;j;k;l}$ does not so appear, but this may be regarded essentially as $\epsilon^i_{;j;k}$. To see the geometrical bearing of these invariants, it is necessary to develop the concept of parallelism for our spaces, and this is the main purpose of this work.

[1] D.D. Kosambi, *The existence of a metric and the inverse variational problem*, Bull. U.P. Acad. of Science, vol. 2. The main ideas of the investigation were set forth in a lecture to the Aligarh Mathematical Seminar on March 5, 1931. Some of the results of this paper have also been given in a note in the Rendiconti R. Accad. Dei Lincei 16 (1932), S. 410–415.

[2] In a personal letter, an extract of which is published after this paper. Mathematische Zeitschrift. 37.

The fundamental ideas are sufficient to develop the full use of a parallelism that the paths should be autoparallel lines, and that the operator of the parallelism should have the tensorial character. From these, it is shown that a non-distributive parallelism results, but that a distributive biparallelism can be obtained by omitting an additive vector. If, on the other hand, we keep consistently to the point of view that the inverse variational problem is essential, then the parallelism must give an invariantive form to the equations of variation of the paths. And that too leads to much the same result. These methods show no way of obtaining an additional index for a given tensor, in a fashion analogous to covariant differentiation in the Ricci calculus. But it is seen that since only the biderivate can be defined for the general tensor, this is not a fundamental question.

In the first paper referred to, there appeared a curious result in the guise of the theorem:

A necessary and sufficient condition that the integral of any twice differentiable function $\varphi(f)$ be stationary over the extremals of $\delta\int f\,dt = 0$ is that the integrand f be a constant along the extremals, i.e., $\frac{df}{dt} = 0$.

This gives the same reduction as in the system (7.3). To account for it, it is necessary to develop the analog of the equations of Killing. The metric f can be regarded as an invariant of a certain fundamental group, and any function thereof will naturally be an invariant also. The second-order equations given by Davis[3] in his treatment of the inverse variational problem can be deduced from (7.2) by requiring the metric to be a relative invariant in place of an absolute invariant, which implies that $\frac{df}{dt} = \lambda f$. The arbitrary constant λ is then eliminated from the resulting compatibility conditions.[4]

2. In the metric case, where the vanishing of the first variation gives the usual system of Euler's equations, or (as we then assume) the equivalent system (I), the important part played by the second variation of the integrand is well known; this leads to the equations of variation, which we have to consider in order to restrict the end points of the integral in (7.1) to lie between two conjugate foci on the extremal. On the other hand, in the metric as well as in the non-metric case, the equations of variation can be obtained by taking

$$\bar{x}^i = x^i + u^i\delta\tau$$

where $\delta\tau$ is an infinitesimal, and $\bar{x}^i$ and x^i are assumed to lie on nearby paths or extremals. Substituting in (I) and neglecting higher powers of $\delta\tau$, we have

$$\ddot{u}^i + \dot{u}^k\alpha^i_{;k} + u^k\alpha^i_{,k} = 0\,. \tag{II}$$

[3] D.R. Davis in the Bull. Amer. Math. Soc. (1929), pp. 371–380. The equations given there would seen to be necessary but not sufficient.

[4] For a detailed bibliography of the subject, and in particular for references to the numerous papers of Berwald, I refer the reader to the article of Koschmieder, Jahresber. d. D.M.V. **40**, pp.109–132. Other papers related to the present investigation are D.R. Davis, l. c. and Trans. Amer. Math. Soc. 33, p. 246 and J. Douglas, Ann. of Math. (II) 29.

It will be seen that these are equivalent to the equations of second variation in the metric case, provided f is non-trivial, i.e.,

$$\Delta \equiv |f_{;k;l}| \neq 0\,.$$

We shall take (I) *and* (II) *as fundamental systems of equations for affine path-spaces and proceed to investigate the possibility of deducing both from a unifying basis of parallelism.*

3. The *derivate* $D(u)^i$ of a set of n quantities shall be defined along a given curve as

$$D(u)^i = \dot{u}^i + \beta^i(x, \dot{x}, t)\,. \tag{7.4}$$

We neglect here the possibilities of derivates containing differential coefficients of x and u of higher order than the first. Even so, a more general form might seem to be

$$D(u)^i = f^i(x, \dot{x}, u, \dot{u}, t).$$

But if the derivate is to be fundamental, we must deduce (I) from

$$f^i(x, \dot{x}, \dot{x}, \ddot{x}, t) = 0\,.$$

This implies that the equations,

$$f(x, \dot{x}, u, \dot{u}, t) = 0$$

thought of as equations in five variables, are soluble for $\dot{u}$. Hence rather than attempt an investigation of all possible linear or functional combinations of the f^i, we might postulate a derivate as in (7.4). Since an invariantive form of equations is to be desired, we restrict the derivate to be such that on a non-singular transformation of coordinates, the transformed derivate also vanishes with the original, at least when the set derived forms the components of a contravariant vector. That is to say,

$$\overline{D}(u)^i = F^i_j D(u)^j\,. \tag{7.5}$$

The coefficients F^i_j must be non-singular for $D(u) = 0$, and their determinant $F = |F^i_j| \neq 0$. The simplest possible assumption that gives the result desired is that the derivate of a contravariant vector is itself a contravariant vector. The coefficients F^i_j are then $\frac{\partial \bar{x}^i}{\partial x^j}$, functions of the transformation itself, and the nonvanishing of F is equivalent to the non-singularity of the transformation.

The assumption is more restrictive than others that can be made, but it leads directly to the tensor invariance of all our fundamental equations.

The parallel displacement of a vector or even a set of n functions along a given curve will be said by definition to take place when and only when the derivate along the curve vanishes. The vector curvature of any curve will be defined as the derivate $D(\dot{x})$ along the curve itself. Finally, the paths are to be autoparallel or curves of zero vector curvature:

$$D(\dot{x}^i) \equiv \dot{x}^i + \alpha^i(x, \dot{x}, t) = 0\,. \tag{7.6}$$

This gives at once a restriction on the form of the derivate,

$$\beta^i(x, \dot{x}, \dot{x}, t) = \alpha^i(x, \dot{x}, t)\,. \tag{7.7}$$

4. Since the Eq. (II) are fundamental in the metric K-space, it should be also possible to reduce them by an application of the principle of derivation to a tensor-invariant form. To this end, we shall assume $u = \frac{x-x}{\delta\tau}$ to be a contravariant vector, which is displaced parallel along the path that is the *base* of the variation. The Eq. (II) must reduce to the form

$$D^2(u)^i \equiv D[D(u)]^i = \varphi^i(x, \dot{x}, u, t)\,. \tag{7.8}$$

To compute φ^i equate the two forms of (II).

$$\begin{aligned}\ddot{u}^i + \dot{u}^k\alpha^i_{;k} + u^k\alpha^i_{,k} \equiv \ddot{u}^i - \alpha^k\beta^i_{;k} + \dot{x}^k\beta^i_{,k} + \dot{u}^k\frac{\partial\beta^i}{\partial u^k}\\ + \frac{\partial\beta^i}{\partial t} + \beta^i(x, \dot{x}, \dot{u} + \beta, t) - \varphi^i = 0\,.\end{aligned}$$

It follows immediately by observing the fashion in which $\dot{u}$ enters the identity that

$$\begin{gathered}\beta^i(x, \dot{x}, u, t) = \gamma^i_k u^k + \epsilon^i\\ \gamma^i_k = \gamma^i_k(x, \dot{x}, t), \quad \epsilon^i = \epsilon^i(x, \dot{x}, t)\,,\\ \frac{1}{2}\alpha^i_{;k} = \gamma^i_k\,.\end{gathered} \tag{7.9}$$

Furthermore, recalling the previous identity (7.7) we have our complete formula for the derivate;

$$\begin{aligned}\epsilon^i &= \alpha^i - \frac{1}{2}\dot{x}^k\alpha^i_{;k}\,,\\ D(u)^i &= \dot{u}^i + \frac{1}{2}u^k\alpha^i_{;k} + \left[\alpha^i - \frac{1}{2}\dot{x}^k\alpha^i_{;k}\right].\end{aligned} \tag{7.10}$$

Note that the residual coefficients ϵ^i vanish identically if and only if $\alpha^i(\lambda\dot{x}) = \lambda^2\alpha^i(\dot{x})$. Whenever the residual coefficients are zero, we have the derivate of the null vector also vanishes, and the operation of the derivate becomes distributive;

$$D(u+v) - \{D(u) + D(v)\} = -\epsilon^i = 0\,. \tag{7.11}$$

The vanishing of ϵ^i is seen to be a necessary as well as a sufficient condition for the last. Postulating a distributive law a priori for the derivate would have greatly restricted the spaces for which the operation had a meaning.

5. The Eq. (II) or (7.8) can now be written in the invariantive form

$$D^2(u)^i = u^r S^i_r + D(\epsilon)^i$$

where

$$S^i_r = \gamma^i_k \gamma^k_r - \alpha^i_{,r} - \gamma^i_{r;k}\alpha^k + \dot{x}^k \gamma^i_{r,k} + \frac{\partial \gamma^i_r}{\partial t} = P^i_r\,. \tag{7.11}$$

It can be seen that S^i_r is a mixed tensor. For if $D(u)^i$ is to be a vector with u^i then ϵ^i must be the components of a vector, which we call the residual vector. And we have assumed that u is a vector, and $D(u)$, $D^2(u)$ are then vectors also. The chief usefulness of our equations is seen to be in their normal form, that is, when the equations can be reduced by means of some change of coordinates that is non-singular and bring (7.11) to

$$\ddot{u}^i = P^i_r u^r\,. \tag{7.12}$$

This implies that we can make

$$\ddot{u} = D^2(u) - D(\epsilon)$$

along the path that forms the base. The transformation must therefore be that particular one which makes

$$\gamma^i_k = \frac{1}{2}\alpha^i_{;k} = 0$$

along the base. We can thus state a theorem.

A necessary and sufficient condition for the reduction of the equations of variation to the normal form is the existence of a non-singular transformation of coordinates for which $\gamma^i_k = \alpha^i_{;k} = 0$ along the given base.[5]

Thus, we see the need for what amounts to an extended theorem of Fermi for our K-spaces. This is proved in a later section though when the residual vector is identically null, the proof for symmetric affine connections is immediately extensible.

[5] Note that for other parallelisms, where $\gamma^i_k \neq \frac{1}{2}\alpha^i_{;k}$, we must have transformations that make both γ^i_k and $\alpha^i_{;k}$ vanish simultaneously on the base, for reduction to the normal form; this is the general necessary and sufficient condition.

In the normal form, the Eq. (II) give the geodesic deviation of Levi-Cività. Successive roots of $|u^i_j| = 0$ give conjugate foci on the base. $|u^i_j|$ is the determinant of n independent solution of (7.12), or of (II). Dynamical stability for systems that are given by (I) would mean that u can be made arbitrarily small for all values of the parameter by choosing reasonably small values of u and $\dot{u}$ initially. But here again, the various definitions of stability will have to be considered separately.

By taking $v(x)$ to be a vector field, one can consider the existence of a covariant derivative $v^i_{|r}$

$$\begin{aligned} D(v)^i &= v^i_{|r}\dot{x} = \dot{x}^r[v^i_{,r} + \gamma^i_{kr}v^k + \epsilon^i_r]\,, \\ \dot{x}^r\gamma^i_{kr} &= \frac{1}{2}\alpha^i_{;k}\,, \quad \dot{x}^r\epsilon^i_r = \alpha^i - \frac{1}{2}\dot{x}^k\alpha^i_{;k}\,. \end{aligned} \tag{7.13}$$

To be of any use at all the covariant derivative must be independent on the direction of derivation, which gives

$$\alpha^i = \dot{x}^k\dot{x}^j\Gamma^i_{jk}(x,t)\,, \quad \Gamma^i_{jk} = \Gamma^i_{kj} \tag{7.14}$$

for the most general K-spaces in which a proper covariant derivative exists: the symmetric affine connections.

If, however, we use the general parallelism γ^i_k and ϵ^i, where $\epsilon^i = \alpha^i - \dot{x}^k\gamma^i_k$ $2\gamma^i_k \neq \alpha^i_{;k}$ we can still get valid results; these are seen to be:

$$\begin{aligned} \gamma^i_k =& \dot{x}^r\gamma^i_{kr}(x,t) \\ \epsilon^i_r(x,t) =& \alpha^i_{;k} - [\gamma^i_{kr} + \gamma^i_{rk}]\dot{x}^k \end{aligned} \tag{7.15}$$

i.:

$$\alpha^i = \gamma^i_{rk}\dot{x}^r\dot{x}^k + \epsilon^i_r\dot{x}^r\,.$$

These are the most general parallelisms admitting a covariant derivative independent of direction. And it is shown that the torsion tensor is given by

$$\Omega^i_{jk} = \gamma^i_{jk} - \gamma^i_{kj} = \gamma^i_{j;k} - \gamma^i_{k;j}\,. \tag{7.16}$$

6. As ϵ^i is a vector on the assumption that $D(u)$ is always a vector with u, we can have a restricted derivate, or the "biderivate"

$$\mathcal{D}(u)^i = \dot{u}^i + \frac{1}{2}\alpha^i_{;k}u^k\,. \tag{7.17}$$

The operation so defined becomes distributive and is also a vector with u. We get correspondingly bipaths and a biparallelism and the reduction of the equations of variation is simplified, though for the canonical form, the necessary and sufficient condition reads as before.

The same results are seen to be true even for a more general tensor analysis, of the sort suggested in the first paper. For instance, let the vector law of transformation be

$$\bar{u}^i = F^i_j(x, \dot{x}, t)u^j \,, \qquad F = |F^i_j| \neq 0 \,. \tag{7.18}$$

where the coefficients F^i_j are functions of the transformation, as well as of any particular curve along which the vector may be defined. We again assume that the derivate $D(u)$ is a vector with u and that the paths are given by $D(\dot{x}) \equiv \ddot{x}^i + \alpha^i(x, \dot{x}, t) = 0$ as before. Then the identities following must hold for all vectors u.

$$\begin{aligned} D(u)^i &= \dot{u}^i + \beta^i(x, \dot{x}, u, t) \,, \\ \overline{D}(\bar{u})^i &= F^i_j D(u)^j \,, \\ \dot{u}^j F^i_j + u^j \frac{d}{dt} F^i_j + \beta^i(\bar{x}, \dot{\bar{x}}, u^j F^i_j, t) &\equiv F^i_j[\dot{u}^j + \beta^j(x, \dot{x}, u, t)] \,. \end{aligned} \tag{7.19}$$

The following results are then read off by inspection:

(a) The β^i are linear in u:

$$\beta^i(x, \dot{x}, u, t) = u^k \gamma^i_k(x, \dot{x}, t) + \epsilon^i(x, \dot{x}, t) \,.$$

(b) The residual coefficients ϵ^i still form a vector:

$$\begin{aligned} \epsilon^i &= \alpha^i - \gamma^i_k \dot{x}^k \\ \bar{\epsilon}^i &= F^i_j \epsilon^j \,. \end{aligned}$$

(c) The law of transformation for γ^i_k is

$$\bar{\gamma}^i_k F^k_j + \frac{d}{dt} F^i_j = F^i_k \gamma^k_j \,.$$

We have again a vector biderivate. The complete determination of the γ^i_k requires further conditions, which we can impose as before, with the same results.

The most general transformation laws, which involve the vector itself, are not feasible, for

$$\bar{u}^i = u^i F^i_j(x, \dot{x}, u, t)$$

implies that

$$\overline{D}(\bar{u})^i = D(u)^j F^i_j(x, \dot{x}, D(u), t) \,.$$

To solve the resulting equations, we ought to assume that

$$\ddot{x}^i + \alpha^i(x, \dot{x}, t) = 0$$

when $D(\dot{x}) = 0$ but without the identical relationship.

The analysis becomes complicated, and as yet, I see no elegant presentation, or even any particularly interesting results that can be included here.

7. To revert to the discussion of reduction to a normal form of the Eq. (II), we note first of all that the standard proof by Eisenhart[6] of the theorem of Fermi can be extended as it stands to the case $\epsilon^i = 0$. Furthermore, the formula (7.19)c shows that the $\bar{\gamma}^i_k$ will all vanish provided only

$$\gamma^i_k \overline{F}^k_j + \frac{d}{dt}\overline{F}^i_j = 0\,. \tag{7.20}$$

If this is scrutinized accurately, one can see at once that a sufficient condition for the reducibility in question is the existence of n independent families of vectors $\lambda^i_{(j)}$ that are all *biparallel* along the base and equal to the coefficients of the inverse transformation to F^i_j. We must have

$$\begin{aligned} \lambda^i_{(j)} &= \overline{F}^i_j \qquad \overline{F}^k_j F^i_k = \delta^i_j = \overline{\lambda}^i_{(j)} \\ \mathcal{D}(\lambda^i_{(j)}) &= 0\,. \end{aligned} \tag{7.21}$$

For the ordinary laws of point transformation, this can always be done, as is seen from the work of Eisenhart cited, formula (25.10), where such a transformation is built up for any curve as base; the process can be reproduced merely by using biparallelism of the K-space in question. Stability thus comes to discussing the roots of the characteristic equation

$$|\lambda\delta^i_j - S^i_j| = 0\,. \tag{7.22}$$

If these are all real and negative, we have a transformation for our normal coordinates and finite oscillations in these.

$S^i_j = P^i_j$ is a mixed tensor, and the roots of (7.22) will therefore be invariants under point transformation.

With the general $F^i_j(x, \dot{x}, t)$ for transformation coefficients, reduction is not always possible, as there may not be a transformation corresponding to a given set of coefficients even along a curve; $\dot{x}$ is not in general a vector. The condition (7.20) and its interpretation are unchanged. Matrix laws of combination apply to the coefficients F, when two or more transformations are performed in succession. Lastly, a covariant biderivate can be defined;

$$\mathcal{D}u_i = \dot{u}_i - \gamma^k_i u_k\,. \tag{7.23}$$

[6] See Eisenhart, *Non-Riemannian Geometry*, Am. Math. Soc. Coll. (1927), pp. 64–67.

The coefficients of covariance being φ^i_j

$$\bar{u}_i = \varphi^j_i u_j \, .$$

For self-consistence and simplicity $\overline{F}^i_j = \varphi^i_j$ can be assumed, using the upper index for summation, though this is merely a sufficient condition. Biderivation for tensors of any rank can be defined by analogy with the usual formulae for covariant differentiation, giving always a tensor of the same type as the original.

8. The foregoing deductions can be motivated by considering groups of deformations of the space. If the infinitesimal transformation of the group be

$$\bar{x}^i = x^i + \xi^i \delta\tau \tag{7.24}$$

invariants of the group of the form $f(x, \dot{x})$ will be given by

$$\delta f \equiv \delta\tau \left[\xi^i f_{,i} + \dot{\xi}^i f_{;i}\right] = 0$$

we demand that the transformation be parallel in the path-space:

$$\dot{\xi}^i + \gamma^i_k \xi^k + \epsilon^i = 0 \, . \tag{7.25}$$

This gives us at once

$$\xi^i \left[f_{,i} - \gamma^k_i f_{;k}\right] - \epsilon^i f_{;i} = 0 \, . \tag{7.26}$$

Sufficient conditions for invariance are

$$\epsilon^i \frac{\partial f}{\partial \dot{x}^i} = 0 \, , \tag{7.27a}$$

$$\frac{\partial f}{\partial x^i} - \gamma^k_i \frac{\partial f}{\partial \dot{x}^k} = 0 \, . \tag{7.27b}$$

In addition to this, if f contains the parameter t explicitly, we can expand the group by adding to (7.24)

$$\bar{t} = t + \delta\tau \, .$$

By this, we shall say that t is an *affine parameter* for the path-space. There is then added an extra term $\frac{\partial f}{\partial t}$ to (7.26) and (7.27) becomes

$$f_{,i} - \gamma^k_i f_{;k} = 0 \, , \qquad \epsilon^i f_{;i} - \frac{\partial f}{\partial t} = 0. \tag{7.28}$$

If the parallel transformations are $n + 1$ in number, it is seen that (7.28) is necessary as well as sufficient conditions for invariance. If the streamlines of the transformation defined by the congruence $\dot{x}^i = \xi^i$ are paths of the space, as would be expected from (7.25) we see from the relation $\epsilon^i = \alpha^i - \gamma^i \dot{x}^k$ and from the condition for invariance that

$$\frac{df}{dt} \equiv -\alpha^i \frac{\partial f}{\partial \dot{x}^i} + \dot{x}^i \frac{\partial f}{\partial x^i} + \frac{\partial f}{\partial t} = 0\,, \tag{7.29}$$

i.e., the function $f(x, \dot{x}, t)$ is constant along the path-streamline. The $n + 1$ transformations imply that every path can be made a streamline in some sufficiently restricted n dimensional manifold of the space: We should expect the same result from (7.28). This indeed is seen at once to be true by eliminating γ^i_k. And this property of the invariant is independent of the particular γ^i_k—or parallelism—chosen. Our reduction of the equations of variation gave us $\gamma^i_k = \frac{1}{2}\alpha^i_{;k}$, the transformation being one which carried paths into paths.[7]

9. If then, we are to deduce our geometry from some fundamental group to which the space is subjected, and the space has one or more metrics attached to it, we should expect the metric $f(x, \dot{x}, t)$ to be an invariant of the group, and the paths to be the geodesics of the metric, i.e., extremals of the variational principle

$$\delta \int f(x, \dot{x}, t)dt = 0\,.$$

This implies that f be a solution of

$$\delta_i f \equiv \alpha^j \frac{\partial^2 f}{\partial \dot{x}^i \partial x^j} - \dot{x}^j \frac{\partial^2 f}{\partial \dot{x}^i \partial x^j} - \frac{\partial^2 f}{\partial \dot{x}^i \partial t} + \frac{\partial f}{\partial x^i} = 0 \tag{7.30}$$

such that

$$\Delta \equiv \left| \frac{\partial^2 f}{\partial \dot{x}^i \partial \dot{x}^j} \right| \neq 0\,.$$

Now the actual condition of invariance $\frac{df}{dt} = 0$ along the paths reduces this system to one of the first orders;

$$\frac{1}{2}\frac{\partial \alpha^i}{\partial \dot{x}^j}\frac{\partial f}{\partial \dot{x}^i} - \frac{\partial f}{\partial x^j} = 0. \tag{7.31}$$

[7] And the group, if solutions of the equation of variation define one as would be expected, will be valid in that neighborhood of a point within which no conjugate focus exists on any extremal through the point.

The parallelism is then determined at once[8] as $\gamma^i_k = \frac{1}{2}\alpha^i_{;k}$. The same reduction can be obtained by imposing on f a property of the invariants that any $\varphi(f)$ be also a possible metric with f.

As is well known from Hamilton's principle, or as can be proved directly, any solution of (7.30) has the property

$$\frac{d}{dt}\left[\dot{x}^i \frac{\partial f}{\partial \dot{x}^i} - f\right] + \frac{\partial f}{\partial t} = 0\,. \tag{7.32}$$

If f does not contain explicitly the parameter t, then $\dot{x}^i f_{;i} - f$ is a constant along the paths.

Whence any $f(x, \dot{x})$, homogeneous of any degree except one in $\dot{x}$, and a solution of (7.30) is also a solution of (7.31), and constant along a path-extremal. There will in general be a finite number of independent solutions if Eq. (7.31) are compatible with $\frac{df}{dt} = 0$, or else none. And just as all invariants of the group can be expressed as a function of a finite number of them, so also any invariant metric will be expressible in terms of these fundamental solutions, the analogy being complete. Of course, a metric proper would need certain other conditions to determine it completely.

Differential invariants of the space, including the two curvature tensors, appear as coefficients in successive compatibility conditions of (7.28) or (7.31) depending on the choice of parallelism.

(Eingegangen am 2. Dezember 1932.)

[8] Or else extra compatibility conditions are introduced $(\alpha^i_{;k} - \gamma^i_k) f_{;i} = 0$.

Chapter 8
Observations sur le mémoire précédent

par Élie Cartan, (à Paris)
(Extrait d'une lettre à M. D. D. Kosambi.)

As mentioned in footnote 2 in the previous paper, in private correspondence, Élie Cartan made certain observations on the manuscript that DDK had sent him. There seems to have been a lively exchange of letters and ideas between Kosambi and Cartan, as the next paper also testifies. The previous paper, this note, and a later paper, by S.S. Chern in the Bulletin des Sciences Mathématiques, **63**, *206 (1939) form the basis of the Kosambi–Cartan–Chern or KCC theory. Kosambi's contribution here is the short explanatory paragraph in English at the end. Although the paper is in French, it seems more appropriate to include it here, given the close relationship it bears to the preceding paper, rather than in Sect. III.*

…J'admets donc que dans les équations

$$\ddot{x}^i + \alpha^i(x, \dot{x}, t) = 0 \tag{8.1}$$

t est un paramètre *imposé* (par exemple le temps). S'il en est ainsi, on peut formuler de deux manières différentes le problème à résoudre:

(A) *Trouver les propriétés géométriques qu'on peut attacher au système* (8.1) *et qui ont un caractère intrinsèque par rapport au groupe infini des transformations*

$$(x^i)' = F^i(x)\,, \quad t' = t\,.$$

Published in Mathematische Zeitschrift **37**, 619–22 (1933).

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_8

(B) *Trouver les propriétés géométriques qu'on peut attacher au système* (8.1) *et qui ont un caractère intrinsèque par rapport au groupe infini des transformations*

$$(x^i)' = F^i(x, t)\,, \quad t' = t\,.$$

Dans le cas (A), qui est celui que vous semblez envisager, on a affaire à un espace et à un temps (qui a une signification absolue).

Dans le cas (B), on a affaire à un *espace-temps*, le temps ayant encore une signification absolue.

Du reste, toute notion géométrique fournie par le problème (B) conserve un sens dans le problème (A); mais la réciproque n'est pas vraie.

Il est remarquable que dans le problème (A) vous avez en somme trouvé *presque* tous les tenseurs fondamentaux, et cela par un procédé très élégant. Il n'y en a qu'un que vous n'ayez pas obtenu; c'est le tenseur $\alpha^i_{;j;k;l}$ (qui existe dans les deux problèmes (A) et (B)). Vous n'avez pas non plus obtenu les procédés les plus généraux de dérivation covariante des tenseurs (formation par dérivation de nouveaux tenseurs à partir d'un tenseur donné), car il y en a d'autres que la dérivation par rapport aux $\dot{x}^h$. A mon avis, pour résoudre complètement le problème, il faut considérer l'espace à $2n + 1$ dimensions des $x, \dot{x}$ et t et considérer dans cet espace un transport parallèle à partir du point $(x, \dot{x}, t)$, en faisant varier les $x, \dot{x}$ et t *sans que δx soit égal à $\dot{x}\delta t$*. Admettons a priori que la différentielle covariante d'un vecteur X^i soit

$$DX^i = dX^i + \left[\gamma^i_k dt + \gamma^i_{kh}(dx^h - \dot{x}^h dt)\right] X^k \quad (\gamma^i_{jk} = \gamma^i_{kj})$$

et plaçons-nous dans le cas du problème général (B). Les quantités $\omega^i_d = dx^i - x^i dt$ déterminent manifestement un vecteur (d est un symbole de différentiation indéterminée). Introduisons deux symboles d et δ échangeables entre eux et formons $D\omega^i_\delta - \Delta\omega^i_d$. On obtient, en supposant $\gamma^i_{rh} = \gamma^i_{hr}$,

$$\begin{aligned} D\omega^i_\delta - \Delta\omega^i_d = dt &\left[\delta\dot{x}^i + \alpha^i \delta t + \gamma^i_h(\delta x^h - \dot{x}^h \delta t)\right] \\ &- \delta t\left[d\dot{x}^i + \alpha^i dt + \gamma^i_h(dx^h - \dot{x}^h dt)\right]. \end{aligned} \tag{8.2}$$

Posons

$$\tilde{\omega}^i_d = d\dot{x}^i + \alpha^i dt + \gamma^i_h(dx^h - \dot{x}^h dt)\,.$$

Comme le système d'equations $\tilde{\omega}^i_d = \omega^i_d = 0$ n'est autre que le système donné (8.1), il a un caractère invariant, et la formule (8.2) montre que la forme $\tilde{\omega}^i$ a le caractère d'un vecteur. Formons alors

$$\begin{aligned} D\tilde{\omega}^i_\delta - \Delta\tilde{\omega}^i_d = (\alpha^i_{;k} - 2\gamma^i_k)&[\tilde{\omega}^k_d\delta t - \tilde{\omega}^k_\delta dt] \\ &+ (\gamma^i_{hk} - \gamma^i_{h;k})(\omega^h_\delta\tilde{\omega}^k_d - \omega^h_\delta\tilde{\omega}^k_\delta) \\ &+ A^i_k(\omega^k_d\delta t - \omega^k_\delta dt) + A^i_{kh}(\omega^k_d\omega^h_\delta - \omega^k_\delta\omega^h_d)\,. \end{aligned} \tag{8.3}$$

On aura une détermination intrinsèque de la différentiation covariante en annulant les deux premiers coefficients:

$$\gamma^i_k = \frac{1}{2}\alpha^i_{;k}\,, \quad \gamma^i_{hk} = \frac{1}{2}\alpha^i_{;h;k}\,. \tag{8.4}$$

Les autres coefficients donneront alors vos deux tenseurs P^i_j et R^i_{jk}, qui jouent ainsi le rôle du *tenseur de torsion*. Quant au tenseur de courbure, il se calcule comme d'habitude et fait intervenir

1. le tenseur $P^i_{k;h}$;
2. le tenseur que vous avez indiqué $R^i_{hk;l}$;
3. le tenseur $\alpha^i_{;j;k;h}$.

Enfin si les composantes X^i d'un vecteur sont des fonctions de x, $\dot{x}$ et t, on a, en posant

$$DX^i = X^i_{|0}dt + X^i_{|k}(dx^k - \dot{x}^k dt) + X^i_{;k}\left[d\dot{x}^k + \alpha^k dt + \frac{1}{2}\alpha^k_{;h}(dx^h - \dot{x}^h dt)\right]\,,$$

les tenseurs derivés $X^i_{|0}$, $X^i_{|k}$ et $X^i_{;k}$, avec

$$\begin{aligned} X^i_{|0} =& \frac{\partial X^i}{\partial t} + \dot{x}^k X^i_{,k} - \alpha^k X^i_{;k} + \frac{1}{2}\alpha^i_{;k}X^k\,; \\ X^i_{|k} =& X^i_{,k} + \frac{1}{2}\alpha^r_{;k}X^i_{;r} + \frac{1}{2}\alpha^i_{;k;r}X^r\,. \end{aligned}$$

La différentiation covariante obtenue correspond à votre biparallélisme, seulement vous supposez que dans votre transport les x, $\dot{x}$, t varient de manière à satisfaire toujours à $dx^i - \dot{x}^i dt \equiv 0$.

Dans le problème (B), les espaces à courbore nulle sont ceux pour lesquels les fonctions sont réductibles à de simples fonctions des x^i et de t (les $\dot{x}^i$ n'entrant plus). Les espaces à courbure nulle et torsion nulle se réduisent à l'espace linéaire ($\alpha^i = 0$). ... Pour l'espace quelconque, si l'on considère la trajectoire d'un point mobile d'une manière quelconque ($dx^i - \dot{x}^i dt = 0$), ce point n'a pas de vitesse intrinsèque, comme vous le remarquez vous-même, puisqu'on peut à chaque instant faire une transformation permise annulant les $\dot{x}^i$; mais il a une accélération intrinsèque $\ddot{x}^i + \alpha^i$ et des accélérations covariantes de différents ordres.

Dans le problème (A), aux tenseurs fondamentaux de torsion et de courbure, s'ajoute le vecteur $\dot{x}^i$ (coefficient de dt dans ω^i) et le vecteur ε^i (coefficient de dt dans $\tilde{\omega}^i$). On a du reste

$$D\dot{x}^i - \tilde{\omega}^i = -\varepsilon^i\,dt - \varepsilon^i_{;k}(dx^k - \dot{x}^k\,dt)\,, \qquad \dot{x}^i_{|k} = -\varepsilon^i_{;k}\,.$$

Pour que l'espace soit linéaire, il ne suffit plus que la courbure et la torsion s'annulent, mais il faut et il suffit qu'en outre le vecteur ε^i soit nul (done $P^i_j = 0$,

$\varepsilon^i = 0$, $\alpha^i_{;j;k;h} = 0$). Ici, l'accélération covariante (dérivée covariante de la vitesse d'un point mobile) est $\ddot{x}^i + \alpha^i - \varepsilon^i$. Les espaces à connexion affine ordinaire sont ceus pour lesquels les deux tenseurs ε^i et $\alpha^i_{;j;k;h}$ sont identiquement nuls.

Enfin, on peut dans les problèmes (A) et (B) interpréter la condition $\alpha^i_{;j;k;h} = 0$. Remarquons pour cela qu'un vecteur est défini non seulement par ses composantes X^i, mais encore par son origine $(x, \dot{x}, t)$; on pourrait dire que l'élément linéaire $(x, \dot{x})$ est *l'élément d'appui* d'un vecteur. Cela posé, supposons qu'on transporte le vecteur par parallélisme *à temps constant* ($dt = 0$) en faisant varier les x^i infinitement peu; si l'on obtient la même variation des X^i en changeant les paramètres directeurs de l'élément linéaire d'appui, c'est que le tenseur $\alpha^i_{;j;k;l}$ est nul et réciproquement. On pourrait naturellment varier ces considérations...

Note:

To compute the curvature tensor, the "usual method" referred to by M. Cartan is that of parallel displacement about an infinitesimal circuit. In essence, this is equivalent to computing the second covariant derivative of any vector

$$(\Delta D - D\Delta)X^i$$

This can be written as

$$X^k \left\{ A^i_{hk}(\omega^h_d \delta t - \omega^h_\delta dt) + B^i_{rhk}\omega^r_\delta \omega^h_d + C^i_{hrk}(\omega^h_d \tilde{\omega^r_\delta} - \omega^h_\delta \tilde{\omega}^r_d) \right\}$$

The coefficients are seen to be tensors, and their actual values come out to be

$$A^i_{hk} = P^i_{h;k} + R^i_{hk}\,, \qquad B^i_{rhk} = R^i_{rh;k}\,, \qquad C^i_{hrk} = \frac{1}{2}\alpha^i_{;h;r;k}\,.$$

The slight difference in the introduction of the tensor R^i_{jk} does not affect, of course, M. Cartan's deduction, since this is only $\frac{1}{3}(P^i_{j;k} - P^i_{k;j})$.

(D.D. Kosambi.)

(Eingegangen am 2. Dezember 1932.)

Chapter 9
The Tensor Analysis of Partial Differential Equations

This paper, read at the tenth conference of the Indian Mathematical Society, Lucknow, in 1938, came to the attention of mathematicians at the University of Hokkaido in Sapporo, Japan. A Japanese translation was published in the journal Tensor *(and as it happens, a few weeks earlier than this paper). DDK was subsequently invited to the editorial board of the journal and he published a paper* [DDK55] *in* Tensor (New Series) *in 1954. Both journals were published by Akitsugu Kawaguchi (1902–1984) who founded the Tensor Society and the journal Tensor in 1938 in Sapporo. Publications were initially in Japanese, but after World War II, the new series of Tensor published papers in English as well. The journal is now housed at the Kawaguchi Research Institute in Chigasaki, Japan.*

After the comprehensive works of Bortolotti [1] on partial differential equations of the second order from the differential geometer's point of view, and the equally comprehensive memoir of Kawaguchi and Hombu [2] on systems of higher order, the present note serves only to show that slightly different results can be obtained by keeping to the point of view that I have used in my former papers [3]. The method, in particular, is to handle such systems as obeying the following postulates and for

Published in Journal of the Indian Mathematical Society **3**, 249–53 (1939). Reprinted with permission.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_9

the special transformation groups under which the postulates hold: (1) the system of equations transforms according to the tensor law; (2) the equations of variation of the given system are also tensorial when the variation itself is a vector; and (3) there exists at least one operator which is vectorial in character and corresponds to total differentiation with respect to one of the independent variables.

To illustrate this, let us consider the second-order system

$$\frac{\partial^2 x^i}{\partial u^\alpha \partial u^\beta} + H^i_{\alpha\beta}(u, x, p^r_\nu) = 0\,; \quad p^i_\alpha = \frac{\partial x^i}{\partial u^\alpha}\,. \tag{9.1}$$

Here, the Latin indices refer to the coordinates x and range over values $1, \ldots, n$; the Greek indices refer to the parameters u and have the values $1, \ldots, m$. The functions $H^i_{\alpha\beta}$ must have the transformation law

$$\begin{aligned} -H'^i_{\alpha\beta} = -H^r_{\nu\delta} \frac{\partial x'^i}{\partial x^r} \frac{\partial u^\nu}{\partial u'^\alpha} \frac{\partial u^\delta}{\partial u'^\beta} + \frac{\partial^2 x'^i}{\partial x^r \partial x^s}\, p^r_\delta p^s_\nu \frac{\partial u^\delta}{\partial u'^\alpha} \frac{\partial u^\nu}{\partial u'^\beta} \\ + \frac{\partial^2 u^\nu}{\partial u'^\alpha \partial u'^\beta} \frac{\partial x'^i}{\partial x^r}\, p^r_\nu\,, \end{aligned} \tag{9.2}$$

under the group

$$x'^i = F^i(x^1, \ldots, x^n)\,; \quad u'^\alpha = \phi^\alpha(u^1, \ldots, u^m)\,. \tag{9.3}$$

But we can speak of x-transformations or u-transformations alone and of x-tensors or u-tensors accordingly. *Tensor* will mean, unless specialised, a geometric object which has the proper law of transformation for both sorts of indices. It is assumed that the conditions of integrability of the partial differential equations are identically satisfied, but no direct use will be made of them. We introduce the non-tensorial operator of differentiation with respect to u, viz.

$$\partial_a \equiv \frac{\partial}{\partial u^a} + p^i_a \frac{\partial}{\partial x^i} - H^i_{\alpha\beta} \frac{\partial}{\partial p^i_\beta}\,. \tag{9.4}$$

It follows that for an x-vector λ^i, a vectorial operator must be of the type $D_\alpha \lambda^i \equiv \partial_\alpha \lambda^i + \gamma^i_{\alpha r} \lambda^r$. But this will not do for any tensor with Greek indices. Therefore, the $\gamma^i_{\alpha j}$ must behave like covariant u-vectors, and an additional term will have to enter the D_α. We may, therefore, take the general operator to be of the form

$$\left.\begin{aligned} D_\alpha T{:}{:}{:} = \partial_\alpha T{:}{:}{:} + \gamma^i_{\alpha r} T^{r}{:}{:} - \gamma^r_{\alpha j} T_r{:}{:} \\ + \Gamma^\nu_{\alpha\rho} T{:}{:}^{\rho} - \Gamma^\rho_{\alpha\sigma} T{:}{:}_\rho \end{aligned}\right\} \tag{9.5}$$

The laws of transformation for the two sets of coefficients must be as follows:

$$\begin{aligned}\gamma^{s}_{\sigma r}\frac{\partial x'^{i}}{\partial x^{s}}\frac{\partial u^{\sigma}}{\partial u'^{\alpha}} &= \gamma'^{i}_{\alpha s}\frac{\partial x'^{s}}{\partial x^{r}} + p^{s}_{\sigma}\frac{\partial^{2} x'^{i}}{\partial x^{r}\partial x^{s}}\frac{\partial u^{\sigma}}{\partial u'^{\alpha}}\,.\\ \Gamma'^{\sigma}_{\alpha\beta}\frac{\partial u^{\nu}}{\partial u'^{\sigma}} &= \Gamma^{\nu}_{\sigma\tau}\frac{\partial u^{\sigma}}{\partial u'^{\alpha}}\frac{\partial u^{\tau}}{\partial u'^{\beta}} + \frac{\partial^{2} u^{\nu}}{\partial u'^{\alpha}\partial u'^{\beta}}\end{aligned} \tag{9.6}$$

Let the coordinates x^i now undergo a vector variation represented by $x^i = x^i + \varepsilon\lambda^i$. Neglecting the coefficients of ε^2 and higher powers as usual, we have the equations of variation

$$\partial_\alpha\partial_\beta\lambda^i + \partial_\nu\lambda^r\frac{\partial H^i_{\alpha\beta}}{\partial p^r_\nu} + \lambda^r\frac{\partial H^i_{\alpha\beta}}{\partial x^r} = 0\,. \tag{9.7}$$

These can at once be put in the invariantive form

$$D_\alpha D_\beta\lambda^i + D_\nu\lambda^r T^{i\nu}_{r\alpha\beta} + \lambda^r P^i_{r\alpha\beta} = 0\,, \tag{9.8}$$

where

$$T^{i\nu}_{j\alpha\beta} = \delta^i_j\Gamma^\nu_{\alpha\beta} - \gamma^i_{\alpha j}\delta^\nu_\beta - \gamma^i_{\beta j}\delta^\nu_\alpha + \frac{\partial H^i_{\alpha\beta}}{\partial p^j_\nu}\,; \tag{9.8a}$$

$$P^i_{j\alpha\beta} = \frac{\partial H^i_{\alpha\beta}}{\partial x^j} - \frac{\partial H^i_{\alpha\beta}}{\partial p^r_\sigma}\gamma^r_{\sigma j} + \gamma^r_{\beta j}\gamma^i_{\alpha r} - \partial_\beta\gamma^i_{\alpha j}\,. \tag{9.8b}$$

In my former work, the coefficients $\gamma^i_{\alpha j}$ were determined by taking the tensor corresponding to the first of these as equal to zero. This can no longer be done here, as it would be too restrictive and would not even then serve to determine both $\gamma^i_{\alpha j}$ and $\Gamma^\nu_{\alpha\beta}$. $\Gamma^\nu_{\alpha\beta}$ need not be symmetrical in the subscripts, but the anti-symmetrical part (torsion) will be indeterminate; one may therefore assume $\Gamma^\nu_{\alpha\beta}$ to be symmetric. It is also clear that the derivatives $\frac{\partial^2\gamma^i_{\alpha j}}{\partial p^k_\beta\partial p^l_\nu}$ and $\frac{\partial\Gamma^\nu_{\alpha\beta}}{\partial p^i_\sigma}$ are the components of tensors. The operator $\frac{\partial}{\partial p^i_\alpha}$ is also tensorial in character, adding a covariant x-component and a contravariant u-component.

The problem is now to determine the coefficients $\gamma^i_{\alpha j}$ and $\Gamma^\nu_{\alpha\beta}$ in a fashion which has some claim to be called intrinsic. Let us assume that all possible contractions of $T^{i\nu}_{j\alpha\beta}$ vanish. This gives

$$\begin{aligned}T^{r\nu}_{r\alpha\beta} &\equiv n\Gamma^\nu_{\alpha\beta} - \delta^\nu_\beta\gamma^r_{\alpha r} - \delta^\nu_\alpha\gamma^r_{\beta r} + \frac{\partial H^r_{\alpha\beta}}{\partial p^r_\nu} = 0\\ T^{r\nu}_{r\alpha\nu} &\equiv n\Gamma^\nu_{\alpha\nu} - (m+1)\gamma^r_{\alpha r} + \delta^\nu_\alpha\gamma^r_{\beta r} + \frac{\partial H^r_{\alpha\nu}}{\partial p^r_\nu} = 0\,, \quad \text{etc.}\end{aligned} \tag{9.9}$$

The general solutions must then have the form:

$$\gamma^i_{\alpha j} = \frac{1}{m+1}\left[\Gamma^\sigma_{\alpha\sigma}\delta^i_j + \frac{\partial H^i_{\alpha\sigma}}{\partial p^j_\sigma}\right];$$
$$\Gamma^\nu_{\alpha\beta} = -\frac{1}{n}\frac{\partial H^r_{\alpha\beta}}{\partial p^r_\nu} + \delta^\nu_\beta\tau_\alpha + \delta^\nu_\alpha\tau_\beta\,. \tag{9.10}$$

The τ_α that enter into the second of these must have the law of transformation

$$\tau'_\alpha = \tau_\nu\frac{\partial u^\nu}{\partial u'^\alpha} - \frac{1}{n}\frac{\partial^2 x'^j}{\partial x^r\partial x^s}\frac{\partial x^r}{\partial x'^j}p^s_\nu\frac{\partial u^\nu}{\partial u'^\alpha}\,, \tag{9.11}$$

but are otherwise arbitrary so far as the present argument is concerned. This shows, in the first place, that *the logical types of connection for our systems are not affine but projective*; however, we shall not pursue this further.

There exists another set of the equations of variation, namely those obtained by giving a vector variation to u. These are

$$p^i_\nu D_\alpha D_\beta\mu^\nu + D_\nu\mu^\sigma Q^{i\nu}_{\alpha\beta\sigma} + \mu^\sigma R^i_{\alpha\beta\sigma} = 0\,, \tag{9.12}$$

if $u^\alpha = u^\alpha + \eta\mu^\alpha$, η being an infinitesimal, where

$$Q^{i\nu}_{\alpha\beta\sigma} = \frac{\partial H^i_{\alpha\beta}}{\partial p^j_\nu}p^j_\sigma - H^i_{\alpha\sigma}\delta^\nu_\beta - H^i_{\beta\sigma}\delta^\nu_\alpha + \Gamma^\nu_{\alpha\beta}p^i_\sigma - p^i_\rho\Gamma^\rho_{\alpha\sigma}\delta^\nu_\beta - p^i_\rho\Gamma^\rho_{\beta\sigma}\delta^\nu_\alpha$$
$$R^i_{\alpha\beta\sigma} = p^i_\nu(\Gamma^\nu_{\delta\sigma}\Gamma^\delta_{\alpha\beta} - \Gamma^\nu_{\alpha\delta}\Gamma^\delta_{\beta\sigma} - \partial_\beta\Gamma^\nu_{\alpha\sigma}) - \frac{\partial H^i_{\alpha\beta}}{\partial u^\sigma} - \Gamma^\delta_{\nu\sigma}Q^{i\nu}_{\alpha\beta\delta}\,. \tag{9.13}$$

The first of these might also be used to determine the connection $\Gamma^\nu_{\alpha\beta}$. But it is clear, from the manner in which p^i_ν enters, that this cannot be done without differentiation. In particular, inasmuch as $\frac{\partial\Gamma^\nu_{\alpha\beta}}{\partial p^j_\delta}$ is a tensor, we might set $\frac{\partial Q^{i\sigma}_{\alpha\beta\sigma}}{\partial p^i_\nu} + p^i_\sigma\frac{\partial\Gamma^\sigma_{\alpha\beta}}{\partial p^i_\nu} = 0$, which leads at once to

$$\Gamma^\nu_{\alpha\beta} = \frac{1}{n}\left[\frac{\partial^2 H^i_{\alpha\beta}}{\partial p^j_\alpha\partial p^i_\nu}p^j_\sigma - \frac{\partial H^i_{\alpha\beta}}{\partial p^i_\nu}\right]. \tag{9.14}$$

This connection involves the *second* partial derivatives of $H^i_{\alpha\beta}$. With the first derivatives alone, we cannot go beyond (9.10). Since the difference of two sets of γ's is a tensor, the differential invariants of the space may be calculated for any one connection and are then obtained for any other by the use of this tensor difference.

There is another operator besides D_ν, and $\partial/\partial p^i_\alpha$, but it may be obtained by alternating the pair thus:

$$\left[\frac{\partial}{\partial p^j_\nu} D_\nu T^{i\alpha} - D_\nu \frac{\partial}{\partial p^j_\nu} T^{i\alpha}\right] = m \frac{\partial T^{i\alpha}}{\partial x^j} + \frac{\partial T^{i\alpha}}{\partial p^r_\beta}\left[\gamma^r_{\beta j} - \gamma^\nu_{\nu\beta}\delta^r_j - \frac{\partial H^r_{\nu\beta}}{\partial p^j_\nu}\right] + T^{r\alpha}\frac{\partial \gamma^i_{\nu r}}{\partial p^j_\nu} + T^{i\beta}\frac{\partial \Gamma^\alpha_{\nu\beta}}{\partial p^j_\nu} . \tag{9.15}$$

Discarding the additive tensorial terms and making use of (9.10), we get the simplified operator

$$\nabla_j T^{i\alpha} = \frac{\partial T^{i\alpha}}{\partial x^j} - \gamma^r_{\beta j}\frac{\partial T^{i\alpha}}{\partial p^r_\beta} + \frac{1}{m} T^{r\alpha}\frac{\partial \gamma^i_{\nu r}}{\partial p^j_\nu} . \tag{9.16}$$

The differential invariants of the space are to be obtained by alternating the three operators given, as usual.

The usefulness of the foregoing discussion lies in its adaptability to the case of differential equations of higher order. Let, for instant, such a system be given by

$$\frac{\partial^{q+1} x^i}{\partial u^{\alpha_1}\partial u^{\alpha_2}\cdots\partial u^{\alpha_{q+1}}} + H^l_{\alpha_1\alpha_2,\cdots\alpha_{q+1}}(u, x, p^i_{\alpha,\cdots}, p^i_{\alpha_1\cdots\alpha_q}) = 0 . \tag{9.17}$$

The operator D_ν will be of the same type as before. The connection coefficients can again be determined from the two sets of equations of variation, in particular, by contraction of the corresponding coefficients of the varied equations. The remaining coefficients of the equations of variation given the "primary" differential invariants of the system. A new differential operator is obtained by alternating and contracting $\partial/\partial p^i_{\alpha_1\cdots\alpha_q}$ and D_ν. This will give an operator with one covariant Latin index and $q-1$ contravariant Greek indices, viz. $\nabla_i^{\alpha_1\cdots\alpha_{q-1}}$. Alternation and contraction of this $\nabla_i^{\alpha_1\cdots\alpha_{q-1}}$ with D_ν will again get rid of another Greek index and give a second operator $\nabla_i^{\alpha_1\cdots\alpha_{q-2}}$. This can be continued till no Greek indices are left, and we obtain the operator which corresponds to the purely Latin index covariant derivative ∇_i. Further alternations will give only differential invariants. At each stage, additive tensorial terms can be discarded to obtain a reduced operator.

The complete set of differential invariants is not worked out here in view of the memoirs already cited. But it must consist, for the greater part, of those that enter into the equations of variation p^i_α, and such others are to be obtained from these by the application of the tensorial operators.

References

1. E. Bortolotti, Rendiconti della Reale Accademia Nazionale dei Lincei, **23**, 16–21; 104–110; 175–180 (1936).
2. A. Kawaguchi, H. Hombu, Journal of the Faculty of Science. Hokkaido Imperial University **1**(6), 21–62 (1937).
3. D.D. Kosambi, *Quarterly Journal of Mathematics*, (Oxford), **6** (1935), 1–12 and **7** (1936), 97–104.

Chapter 10
A Statistical Study of the Weights of Old Indian Punch-Marked Coins

D.D. Kosambi, Fergusson College, Poona

This paper marks the beginning of DDK's foray into numismatics. As Kosambi's biographer states [DDK-JK], *"Coins, being means of financial transaction, are true indicators of the kind of regime the kings of those specific periods ruled with. Kosambi tried to glean historically important information by studying them and was successful to a large extent. He collected hundreds of ancient coins cleaned them very carefully and weighed each of them accurately on the sensitive balance in the chemical laboratory of the Fergusson College. He then noted the minute differences in their weights due to usage and drew their graphs. Applying statistical tests to this data he successfully drew conclusions regarding the exchange rate, the period when the coins were cast, etc., that could stand up to scientific tests. These graphs and the punch-marks of the mint and the traders' guilds impressed on the coins helped him draw inferences about the state of affairs of that land, in that specific period."*

The punch-marks on old silver coins found in Indian have presented an unsolved riddle which has been attacked by a classification of the *obverse* marks. The efforts of Messrs. Durgā Prasād,[1] Walsh,[2] and Allan,[3] in this direction will be valuable to future scholars, but as yet lead to no conclusion. The first two have paid some attention to the reverse marks also, while the third sometimes ignores them; the reason for this partiality to the obverse is that a group of five marks occurs systematically there, while the reverse may be blank or contain from one to sixteen marks.

Published in Current Science (Bangalore), **9**, 312–14 (1940). Reprinted with permission.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_10

Table 10.1

	x	0	1	2	3	4	5	6	7	8	9	10
Square	n	224	128	132	85	64	46	21	25	10	9	8
	m	53.26	52.93	52.74	52.47	52.53	52.17	52.03	51.67	51.40	51.47	51.01
Round	n	58	34	29	28	25	10	13	8	9	3	3
	m	53.35	52.84	52.75	51.90	52.29	51.67	51.82	52.23	51.23	50.10	51.20

The most important qualities of the coins in the ancient days were undoubtedly the weight and the composition. The latter has received very little attention, a coin or two being sampled from each new lot. The former is given as a rule, for every coin, but the statistical study of coin group by weight does not seem to have been attempted[4]. The resulting confusion as to what standard of weight actually existed can be seen by consulting any of the above works; even Rapson[5] found documentary evidence too self-contradictory for use.

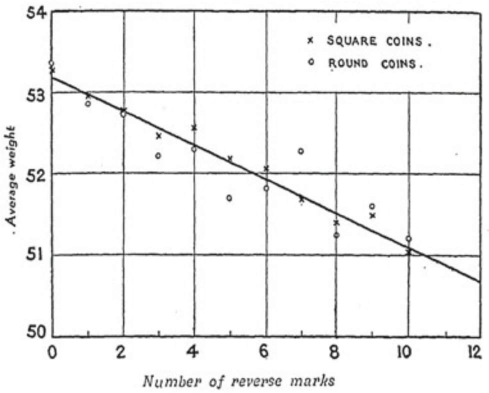
For the basis of a preliminary study, I took Walsh's memoir on two Taxila hoards as fundamental. The work is full of oversights and mistakes, as I have shown in a note to be published in the *New Indian Antiquary*. Nevertheless, it is the only sizeable mass of data available to me, and I take all figures from Appendix XI, with the hope that no error of any importance enters into the weighing. Excluding the 33 long bar coins which approximate to Persian sigloi, and the 79 min coins, all the rest, to a total of 1059 coins which seem meant to represent the same amount of metal, average 52.45 grains in weight. The 162 later coins (App. XII) of a single coinage average 52.72 grains. But the standardization of weights was not the same as is shown by applying the z test to the variances of the two lots.

But even the main hoard of 1059 *kārṣāpana* is not homogeneous. So, I classified them by the number of reverse marks and found the following data, in which the 64 double obverse coins have been omitted.

In Table 10.1, n is the number of coins with the number x of reverse marks given at the column head, and m the average weight in grains. One coin in the square 10-reverse mark class has been omitted, because it has a decidedly different history from that of the rest.[6] There exists coins with as many as 16 reverse marks, but counting the number of marks becomes difficult, and the total not tabulated being 15 square coins and 7 round; the table given below will represent substantially the most reliable portion of the data available to us.

It is seen at once that there is a regular drop in average weight with increase in the number of reverse marks. In fact, for the square coins, the linear regression can be fitted accurately enough by eye and is found on calculation to give the formula: $y = 53.22 - 0.212x$, where y is the average weight in grains and x the number of reverse marks. For round coins, the fit is not so good, though still satisfactory, the regression being $y = 53.1 - 0.214x$, that is particularly the same line servers for both (Fig. 10.1). The second result concerns the number of coins in each group. For simplicity, taking the sum y of both round and square with a given number x of reverse marks, the drop in number is exponential (Fig. 10.2). That is, the regression is given

Fig. 10.1



by $y = 283.86\,e^{-x/3}$. This was obtained by taking the logarithm of the number of coins with each x and fitting a linear regression. The divergence between the formula and the observed number is not significant by the χ^2 test, and the calculation obtained from the above table serves also for the omitted coins, giving, for $x = 0$ to 16, a value of χ^2 with P near 0.2, on the whole, a just tolerable fit. These two results are quite startling. They show that the reverse marks—irregular as they might appear—were not distributed at random, for had they been so distributed, we should have obtained a Poisson's distribution or something of the sort for the number of coins as a function of x, and the linear regression for weight would not have fitted so well. The only hypothesis that can account for our results is that *the reverse marks are checking marks stamped on by contemporary regulators or controllers of currency, at regular intervals*.

If accepted, this means that among the obverse marks, there might exist some symbols that specify the date of issue of the coins. This would, possibly, account for the fifth variable symbol found on the obverse. Even now, we have a sixty-year cycle with a name for each year, and there certainly existed an older 12-year cycle, still extant in Chines and Tibetan tradition, which was converted into a sixty-year affair by associating twelve years with each of the five elements. This could account for one or two of the five obverse marks. One observes mark is fixed: the sun symbol. If it is not votive, it might be a symbol of the metal itself. The next commonest mark is some form of the wheel, with (usually) six points of varying design. This *ṣaḍaracakra* is, in my opinion, not to be interpreted as a symbol of any deity, but as representative of the issuing authority, the *cakravartin* or kind. The form of the points of the wheel, with perhaps one of the extra symbols, might be the ruler's personal monogram. This is borne out by the fact that in the few cases where the six-pointed wheel does not occur, we invariably get (with two exceptions) small homo-signs in their place (Durgā Prasād, p. 41). That is, when the issue was not authorized by a king, it was authorized by a council of some sort.

Leaving these doubtful conjectures, we can use groupings by obverse marks for the purpose of weight analysis and compatibility tests, in particular the *t* test and the *z* test. I shall publish my results on this elsewhere.

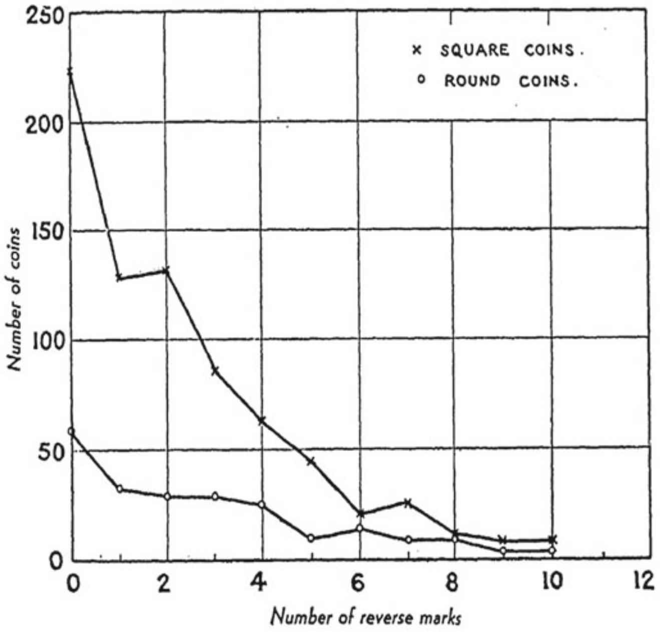


Fig. 10.2

Even in modern times, a certain amount of currency will be lost each year due to damage, hoarding, melting down, etc. This should, in stable times, be proportional to the actual number of coins in circulation. But when the coin does not represent full value in metal content, being just a token coin, with a rigorous control of weight by the examiners of currency, the formula for the number of coins surviving t years after issue would be given by

$$y = ae^{-bt}\left(\frac{1}{\sigma\sqrt{2\pi}}\int_{m_1-r}^{m_1+r} e^{\frac{(x-m)^2}{2\sigma^2}}\,dx\right)$$

where $m = m_1 - tm_2$; $\sigma^2 = \sigma_1^2 + t\sigma_2^2$.

Here, a is a constant of integration, essentially the number minted. The legal weight, as also the average of freshly minted coins, is taken as m_1, the variance at the mint as σ_1^2. The average loss of weight per year is m_2, and the variance of this annual loss, σ_2^2. The legal remedy, i.e., the weight by which a coin may exceed or fall below the legal standard, is called r in the formula.

When the coin is a source of metal, the first factor would account for most of the currency in circulation, particularly as the variances with modern technique of minting are very small. But with a token coin, and in any case after the passage of a greater number of years, the second factor would begin to dominate and the coins withdrawn rapidly from circulation by those who check the currency. The phenomenon is similar to that often seen in biology, where a gene or culture of bacteria shows exponential *growth* till a threshold value is reached; when the situation changes entirely, the growth makes its own surroundings lethal, and further growth is either inhibited, or the whole of the variate vanishes altogether.

References

1. *Journal and Proceedings of the Asiatic Society of Bengal*, New Series (1934), **30**, Numismatic Number.
2. *Memoirs of the Archeological Survey of India* (1939), No. 59.
3. *Catalogue of Indian Coins in the British Museum* (Ancient India, 1930).
4. Ibid., Andhras, W. Kṣatrapas (1908), p. clxxvii et sec.
5. A.S. Hemmy, J. R. Asiatic Soc. (1937), pp. 1–26 must be dismissed as mere trifling with an important subject.
6. One coin in the 3-mark round lot should also have been so omitted, bringing the mean to 52·20, which would have fitted much better.

Chapter 11
A Bivariate Extension of Fisher's Z-Test

D.D. Kosambi, Fergusson College, Poona

DDK had great regard for R.A. Fisher and had studied his work extensively. There appears to have been some correspondence between them, although this has not been documented. A few years later (see Chap. 4), DDK was to write that "R.A. Fisher thought that my ideas on blood groups and cancer were worth following through", so it appears that Fisher appreciated DDK's applications of his methods and techniques (see for example also Chaps. 14, 17 and 18).

A normal distribution in k variates $x_1, x_2, \ldots, x_k$, each with expectation (population mean) zero is defined by the probability density $c \exp -\phi/2$, where c is always to be understood as a constant so chosen as to make the total probability equal to unity, and ϕ is a positive definite homogeneous quadratic form in the variates, i.e.:

$$\begin{aligned}\rho =&\frac{1}{\sigma(2\pi)^{\frac{k}{2}}}\int_R \cdots \int e^{-\phi/2} dx_1 \cdots dx_k\,; \qquad (11.1)\\ \phi =&\sigma^{ij} x_i x_j\,.\\ \sigma^{ij} =&\sigma^{ji}\,;\ \sigma^{ir}\sigma_{rj} = \delta^i_j\,;\ \sigma^2 = |\sigma_{ij}|\,.\end{aligned}$$

Here, we use the tensor summation convention for repeated indices, and the integral is to be taken as extended over that portion of the k-space in which the variates are to lie. The coefficients σ_{ij} are to be formed by taking the normalized co-factors of the corresponding element in $\|\sigma^{ij}\|$, as usual. Alternatively, we can write $\sigma^{ij} = \frac{\partial \log \sigma^2}{\partial \sigma_{ij}}$. The form ϕ being definite, the determinant σ^2 does not vanish, and there is no theoretical difficulty in finding either σ^{ij} or σ_{ij}, the matrix of the other coefficients being given.

Published in Current Science (Bangalore), **10**, 191–92 (1941). Reprinted with permission.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_11

Suppose now that a sample of n observations be taken from such a population, the j-th sample value of the variate x_1 being x_{ij}. Then, it is known that the best [1] estimates of σ_{ij} are given by

$$S_{ij} = \frac{1}{n-1}\sum_{r=1}^{n}(x_{ir} - \bar{x}_i)(x_{jr} - \bar{x}_j) \tag{11.2}$$

where

$$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$$

The best [1] estimate σ^2 is $s^2 = |s_{ij}|$ and of σ^{ij}, the corresponding normalized cofactors, s^{ij}.

It is well known that the quantities s_{ij} are the sample variances when $i = j$, and the sample correlations multiplied by the corresponding standard deviations when $i \neq j$. Again, s^2, the determinant of the sampling coefficients, has a strong claim to be considered as the generalized variance of the multivariate sample. The ratio of two such variances chosen from the same populations would be independent of a linear homogeneous transformation of the co-ordinates, and also of the population parameters. It is natural to ask whether the distribution of this ratio, or rather of its logarithm, has anything in common with Fisher's z, so that the z tables could be used without further ado. The answer is negative in general, but it is the purpose of this note to point out the fact that for a bivariate population ($k = 2$), such an extension is valid.

2. Following the methods given by Uspensky [2], it is a comparatively simple matter to find the distribution of S, where

$$S^2 = \det\left\{\sum_{i=1}^{n}(x_{ir} - \bar{x}_i)(x_{jr} - \bar{x}_j)\right\}\ ; \quad i, j = 1, 2. \tag{11.3}$$

It is to be noted that $s^2 = S^2/(n-1)^2$. By a distribution, we mean the probability that $S^2 < t^2$, the derivative of this with respect to t being then the probability density, which is sometimes called the "distribution" by statistical writers.

For convenience of notation, let the two variates be x and y. Then, $\phi = ax^2 + 2bxy + cy^2$. But as we mean ultimately to consider the *ratio* of two generalized variances, which is a function independent of linear homogeneous transformations, we might as well consider the transformation to have been performed in advance which brings ϕ to its canonical form: For a positive definite form, $\phi = x^2 + y^2$. The required distribution is then given by

$$p(t) = \frac{1}{(2\pi)^n}\int_R \cdots \int e^{-\frac{1}{2}(x_1^2+\cdots+x_n^2+y_1^{+}\cdots+y_n^2)}dx_1\cdots dx_n dy_1\cdots dy_n \tag{11.4}$$

where the region of integration R is defined by the inequality:

$$S^2 \equiv \sum(x_i - \bar{x}) \sum(y_i - \bar{y}) - \left\{\sum(x_i - \bar{x})(y_i - \bar{y})\right\}^2 < t^2\,; \tag{11.5}$$

with

$$\begin{aligned}\bar{x} =& \frac{1}{n}\sum x_i\,,\\ \bar{y} =& \frac{1}{n}\sum y_i\,.\end{aligned}$$

The variates x and y have the sampling values $x_1, \ldots, x_n$; $y_1, \ldots, y_n$, which are independent, being chosen at random by hypothesis, and the formulae (11.4–11.5) are then self-evident.

For the reduction of the integral, the treatment by Uspensky for the distribution of the correlation coefficient is rigorous and can be carried out step by step. Choosing the new variables of integration as the means $\bar{x}$, $\bar{y}$ and $n-1$ each of the differences $x_i - \bar{x}$, $y_i - \bar{y}$, and performing a suitable linear homogeneous transformation, the integral in (11.4) is reduced to a similar one with $n-1$ in place of n, the usual loss of a degree of freedom for measuring from the sample mean. A second transformation and one integration will reduce the integral further to

$$\begin{aligned}p(t) =& c\int_R \ldots \int e^{-\frac{1}{2}(w_1^2+\ldots+w_{n-1}^2+\xi_1^2+\ldots+\xi_{n-2}^2)}\,dw_1 \ldots dw_{n-1}d\xi_1 \ldots d\xi_{n-2} \\ R :& (w_1^2 + \cdots + w_{n-1}^2)(\xi_1^2 + \cdots + \xi_{n-2}^2) \;<\; t^2\end{aligned} \tag{11.6}$$

But we have the two classical formulae of integration:

$$\begin{aligned}(a):&\quad \int_0^\infty e^{-x^2-\frac{a^2}{x^2}}\,dx = \frac{\sqrt{\pi}}{2}e^{-2a}\\ (b):&\quad \int \cdots \int_{x_1^2+\cdots+x_r^2<a} e^{-\frac{1}{2}(x_1^2+\cdots+x_r^2)}\\ &\quad F(x_1^2 + \cdots + x_r^2)dx_1 \cdots dx_r\\ &\quad = \frac{\pi^{r/2}}{\Gamma(r/2)}\int_0^a e^{-\frac{u}{2}}u^{\frac{r}{2}-1}F(u)du\end{aligned} \tag{11.7}$$

These allow us at once to write down dp/dt in the form:

$$\frac{dp}{dt} = ce^{-t}t^{n-3}\;:\quad \text{range}\;\; t = 0 - \infty\,. \tag{11.8}$$

This is, again, of the form of the integrand for the incomplete gamma function, and so, if we wish to find the distribution of the ratio of two independent sampling observations of S^2, we can proceed as usual. But it is clear that the exponent is not the usual number of degrees of freedom. In fact, the degrees of freedom, as is to be seen by comparing exponents with those in the usual formula, are now $2n - 4$. Thus, we must use $(2n - 4)^2$ as the divisor for S^2 in place of $(n - 1)^2$. Finally, a last correction is necessary for the fact that we have used $S^2 < t^2$ in place of the usual distribution, which would be the probability $S^2 < t$. All of this, however, is now quite obvious, and the result can be summed up in a theorem:

If two independent samples of n, n' specimens are taken at random from a bivariate normal population, then the quantity

$$\begin{aligned} z =& \frac{1}{4} \log \frac{S^2}{S'^2} + \frac{1}{2} \log \frac{n' - 2}{n - 2} \\ =& \log \left\{ \sqrt{\frac{S}{n-2}} \Big/ \sqrt{\frac{S'}{n'-2}} \right\} . \end{aligned} \tag{11.9}$$

has the same distribution as Fisher's z for a single variate, with the degrees of freedom $2n - 4$, $2n' - 4$.

The distribution was known (Wilks [3], 478), but the adjustment for the proper number of degrees of freedom, and the possibility of using Fisher's tables, have apparently been overlooked. The rule is quite as simple as for a single variate. In the usual notation, we calculate the quantity $s_x^2 s_y^2 (1 - r^2)$ multiplied by the correction factor

$$(n - 1)^2 / 4(n - 2)^2 \, .$$

and take a *quarter* instead of a half of the natural logarithm of the ratio of two such sampling observations. Then, enter Fisher's tables of z as usual, but with the degrees of freedom $2n - 4$ instead of $n - 1$.

3. The results of the preceding section are not extensible to $k \geq 3$. The integral do not reduce so easily, at least by any known formulae. For example, the case $k = 3$ can be solved completely if an explicit formula for the integral from zero to infinity of $\exp -(x + a^2/x^2)$ is found. But it does not seem possible that this would allow a rigorous use to be made of the z tables.

It would be interesting to see the extended Z-test for $k = 2$ used for analysis of variance: say for plot experiments with two simultaneous crops sown on each plot. The test is open to the same criticisms leveled against the Z-test for one variate, in that it does not take the mean values into account, but tests directly on the basis of the observed variances, the hypothesis that both samples might have been drawn from the same normal population. For tests also taking the mean values into account, as in Student's t test, we have the T^2 of Hotelling and its generalizations. But for a bivariate population, the test suggested here is surely more complete than the usual method of testing the variances s_x^2, s_y^2 individually along with the correlation coefficient τ.

References

1. J.L. Coolidge, *Theory of probability* (Oxford, 1924), p. 82.
2. J.V. Uspensky, *Introduction to the Mathematical Theory of Probability* (1937), p. 332, et.seq.
3. S.S. Wilks, Certain generalizations in the analysis of variance. Biometrika **24**, 471–494 (1932).

Chapter 12
The Effect of Circulation Upon the Weight of Metal Currency

D.D. Kosambi, Fergusson College, Poona

As DDK says in his autobiographical note (see Chap. 2*), he took up this problem as a way of learning statistics. Since examination marks provided poor quality data, he turned to the statistical study of punch-marked coins. He noted that* "not all coins issued at the same time are used in exactly the same manner. Therefore, the effect of circulation is to decrease the average weight but also to increase the variation". *A side effect was that he became more aware of the sociology of the process: these years at Fergusson College were to see his interest in the other aspects grow, as he brought experimental tools, the careful weighing of thousands of coins, to make numismatics* "a science rather than a branch of epigraphy and archaeology".

In contrast to the physical sciences, the social sciences allow, even now, the detection of quite important effects with the aid of comparatively simple apparatus and a certain amount of knowledge of modern statistical technique. The historical evidence of the demand for currency shown by the loss of weight of coins still in active circulation comes under this head. The same methods may be applied to hoards deposited in ancient times and recovered intact, thus giving the foundations of numismatics as a science.

The normal law of weight distribution may be assumed to hold for a set of coins honestly minted to a fixed legal standard in large numbers. The population mean may be taken as the supposed legal weight, the variance could be estimated by taking the number of rejections at the mint beyond the fixed "legal remedy" by which the coin is allowed to differ from legal weight. Supposing the minted weight distribution to be represented by I in Fig. 12.1 (and ignoring the absorption of the coinage), the effect of circulation will be to lower the mean and to increase the variance, as in II. Further circulation changes the curve to III, where only the heavier half has been drawn. Deviations from normality will become more strongly marked, and the currency will tend to disappear from circulation. While the general case can be brought under

Published in Current Science (Bangalore), **11**, 227–231 (1942). Reprinted with permission.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_12

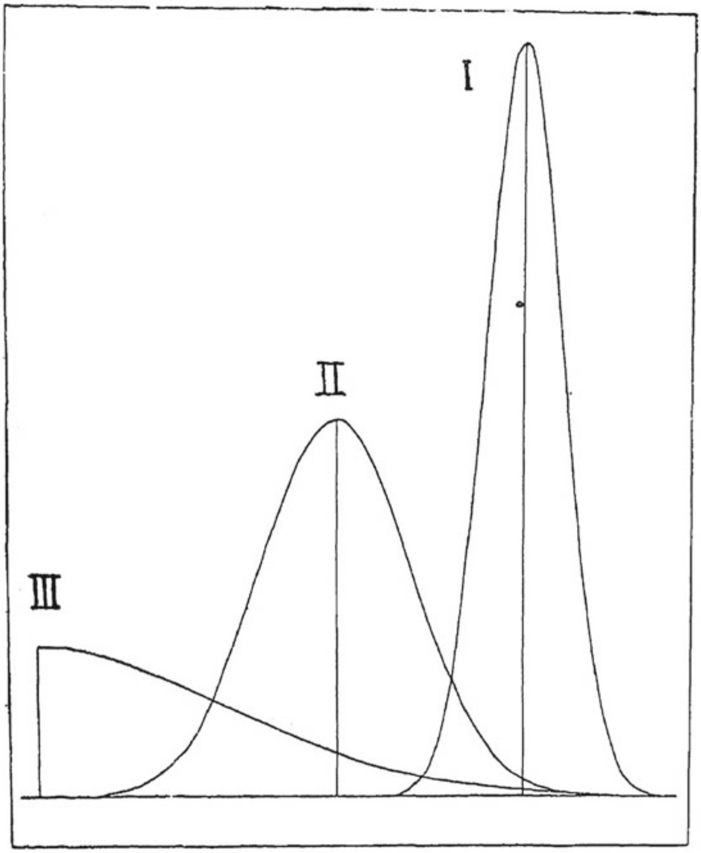


Fig. 12.1 Effect of circulation on weight

the "homogeneous random process" [1] which is so universal in application as to qualify for a law of nature, it suffices for comparatively short periods of time to take the average weight as a linear function of the date.

This theory was applied to a statistical analysis [2] of the earlier Taxila hoard (deposited *circa* 317 B.C.), but work on other ancient hoards of interest was prohibited by lack of access to the material and by the honoured custom of scattering most such material *unweighed* after a perfunctory study. So, the validity of the theory is here proved on modern coins from active circulation [3], as a control measure. During March and April 1942, I gathered from some stores in Poona, from the great marketplace (*maṇṇại*), and when not otherwise available, from the day's take over the counter of a local bank as many specimens as my finances permitted and my energy sufficed to weigh. These were stripped of the pieces whose date was illegible, or which were severely damaged by accident, or which did not ring true for the higher denominations. Experience shows that, as regards weight, coins of the latter two classes invariably differ in a marked fashion from the rest of their annual group for the first, there was no choice. The effect of the two latter discards is to decrease the variance within a year, so that the goodness of fit is actually reduced by this process and the theory stands confirmed even under the most unfavourable circumstances. The date on worn specimens could probably be restored by means of an examination of the crystal structure formed at the time of stamping, but I was unable to devise any method with the apparatus at hand. The pieces were taken as they stood; for the other currency, modern specimens, minted in 1936 and after 1939k, were in overwhelmingly large proportion, and subsamples had to be taken to reduce the numbers. The final selections were classified according to the date of the issue and each coin weighed to a tenth of a milligram. The time of the weighing was reduced by using a chainomatic analytical balance of Indian manufacture; the error of the (new) instrument was rather high—$\frac{1}{2}$ mg—but decreased with use. Proper checks were taken regularly, and the fourth place of decimals ignored in the statistical work; all means would have to be increased by half a milligram and Sheppard's corrections necessary for the variances of the data were to be used for purposes of estimation. The final stage was the statistical analysis of the weights by the methods of Fisher [4].

With larger samples, the estimates of composition and even of the actual weight and its variance would be more accurate; reliable information could be gained as to the proportion of counterfeits, mint-defective, dumb, and accidentally damaged coins in circulation. The variation between localities and local needs can also be estimated by the allocation of the properly randomized samples to various regions. Finally, the residuals after fitting the regressions would be of great use in correlating the wave of various denominations to show the extent to which one type was supplementing another and enable a scientific distribution of currency to be made. Any method of currency control based on science, not on the flat of authority, would have to consider these matters seriously. As for the weights of a larger sample, the analytical balances will no longer be necessary; a histogram can be run off directly by setting the mint's automatic weighting machines in series and counting the number of coins not rejected at each step.

A look at the tables of analysis of variance shows at once that the results of my observations are highly favourable to the theory. Where deviations from the linear regression become significant, they are immediately explicable. The pies being not current in Poona bazaars had to be imported from Benares where they are gathered from the shops before Hindu holidays by the frugal pious, distributed to beggars, and revert to the shops immediately after. This can hardly be called active circulation; as an aside, be it noted that in places like Benares simple bits of copper can be and are still used to substitute for the lower currency: for Benares, the Butwal "pice", almost any ancient coin in most of the purely agrarian districts of India.

The Poona pice fall into two classes, the weight of the denomination having been materially reduced in 1907, apparently to 75 grains. In fact, all pice of my 1906 sample fall into either the 4-g or the 6-g group, without a single specimen of 5 g; the mean for this year is very significantly lighter by the t test than for previous years, heavier than for succeeding years; the variance by the Z-test is significantly

Table 12.1 Analyses of variance. Regressions given only where significant. Unit: one milligram; y = weight in milligrams, x = date in years

Source	d.f.	Sum square	Mean sq.	F
Æ Pies (Benares) 1912–1939; $y - 1599.55 = 1.955(x - 1929.12)$;				
Regression	1	43015	43015	36.66***
Deviations	23	61528	2675.13	2.28**
Within a year	198	232300	1173.23	$r = 0.357$
Total	222	336843	1517.31	1.29*
Æ Pies (superseded) 1835–1906				
Regression	1	35969	35969	$(5.95)^{-1}$
Deviations	27	7133371	264198.92	1.234
Within a year	99	21195723	214098.21	$r = -0.0356$
Total	127	28365063	223346.95	1.0432
Æ Pies 1907–1941; $y - 4728.86 = 9.903\ (x - 1928.87)$				
Regression	1	8574800	8574800	1663.96***
Deviations	26	201108	7734.94	1.50
Within a year	639	3292918	5153.24	$r = 0.843$
Total	666	12068826	18121.36	3.516**
N Annas 1908–1941; $y - 3803.20 = 6.545\ (x - 1927.70)$				
Regression	1	3250147	3250147	1903.31***
Deviations	26	132110	5081.15	2.975**
Within a year	698	1191923	1707.63	$r = 0.843$
Total	725	4574180	6309.21	3.695***
N 2-Annas 1918–1941; $y - 5759.2 = 8.516\ (x - 1931.99)$				
Regression	1	1890586	1890586	695.86***
Deviations	16	71021	4438.81	1.63**
Within a year	315	855827	2716.91	$r = 0.819$
Total	332	2817434	8486.25	3.12***
AR 4-Annas 1904–1940; $y - 2857.9 = 4.615\ (x - 1928.098)$				
Regression	1	725568	725568	459.70***
Deviations	21	56104	2671.62	1.69
Within a year	224	353551	1578.35	$r = 0.799$
Total	246	1135223	4614.73	2.92***
AR 8-Annas 1905–1941; $y - 5764.83 = 5.949\ (x - 1928.5)$				
Regression	1	259159	259759	139.86***
Deviations	21	31273	1489.19	$(1.2472)^{-1}$
Within a year	43	79865	1857.32	$r = 0.837$
Total	65	370897	5706.11	3.07***

(continued)

Table 12.1 (continued)

Source	d.f.	Sum square	Mean sq.	F
AR Rupees [3] 1903–1920; $y - 11579.86 = 4.16\ (x - 1913.12)$				
Regression	1	15423	15423	674.67***
Deviations	16	1130	70.63	3.0898***
Within a year	2868	65563	22.86	$r = 0.433$
Total	2885	82116	28.463	
AU Sovereigns 1900–1931				
Regression	1	72	72	2.382
Deviations	11	772	70.54	2.333*
Within a year	38	1179	30.23	$r = 0.1885$
Total	51	2027	39.745	1.315

greater than those before or after. This seems to indicate that some of the 1906 pice were minted to the lower weight. Thus, the pre-1907 coins have been withdrawn for the greater part or have otherwise tended to disappear from circulation. Only the unworn specimens have managed to survive, whence neither the regression nor the deviations from it are of any significance. For the nickel one anna coins, the deviations from regression are caused entirely by the oldest issues: Edward VII, 1908–1910. For these, no less than 15 out of a total of 38 had illegibly worn dates, a proportion fourteen times that of the George V issues. The 23 coins retained were, naturally, heavier than the average for their groups, somewhat after the fashion of III in Fig. 12.1. A precisely similar effect is to be seen in the Taxilan coins of more than ten reverse marks. A recalculation of the anna data discarding the Edward VII issues immediately reduces the deviations from linear regression to insignificance, so that the deviations are to be assigned to our mechanism of selection. We can thus state a low of wear for metal currency: *For coins in active circulation, the loss of average weight is proportional to the age. But the oldest coins of a series tend to be above the regression weight and for currency not in active circulation* [5] *or an issues which is superseded, the significance of the regression tends to disappear.*

An even more striking result is that *the correlation coefficient for currency in active circulation over comparable periods of time is independent of the denomination.* Except the pies, the older pice, rupees, and sovereigns all the remaining correlation coefficients do not differ significantly from the population value of $\varrho = 0.838$, estimated by pooling the observed values after Fisher's z transformation [6]. The correlation for the 4-anna bits is somewhat low, but there have been disturbing factors at work here: the 1917–1918 specimens show unusual wear and nickel 4-anna bits (not included in this study) were minted in 1919, 1920, and 1921. In stating such a "law" for currency weights, other things must be equal: minting variances must not be great in comparison with those caused by wear the currency must have been minted over about the same period and must have circulated in the same locality over about the same time. As a matter of fact, 2.886 rupees of 1903–1920 issue sampled at Poona in 1940 gave me a correlation of 43 and deviations from linearity were insufficient to explain this entirely different value. The reason for the difference, however, is very simple. It is known that r^2 is the ratio of sum square due to regression by the total sum square. Our theory requires that the variances increase with age, which means that for coins longer in circulation, the residual sum square takes up a greater proportion of the total, thus depressing the correlation. Even the pice of our sample show a correlation compatible with that of the rupees when calculated only from the 1907–1920 issues in the sample. It is a feature of the data that when the calculations are made from year to year on the basis of the weights, the correlation coefficient is found to increase steadily with the date of the last issue to its maximum value at the end: this holds for all denominations provided the oldest issues do not contain overweight survivors in large proportion and the regression is really significant.

Whereas the samples show that the variances are in general decidedly greater for the older issues, the samples do not allow the question of linear increase of the variance with age to be effectively discussed except for the post-1906 pice. The only method I can see that would test this would be (1) to calculate the linear regression from the sample variances giving each the weight of its degrees of freedom and (2) to apply the χ^2 test, noting that the ratio of the observed to a hypothetical variance should be distributed as χ^2/n. From the total number of degrees of freedom, two have to be subtracted for the fitting. The pice variances only, when all are tested by this method show linear increase with age; on the whole, the pice are statistically the most satisfactory denomination in spite of evidence of heavy corrosion of three specimens by fatty acids—because no one rings them, counterfeits and hoarding are absent, change of hands regular.

Brass $\frac{1}{2}$ annas, annas, and two annas of 1942 issue just reached circulation at the time of the study, so that no disturbing effect was obvious on the rest of the currency, whatever the future may show. The data gives: $\frac{1}{2}$ annas-$n = 53, m = 2.9125$ gm, $s^2 = 786.88$ mgm^2; annas-$n = 38$, $m = 3.8851$ gm, s^2=3934.51 mgm^2; and 2 annas-$n = 22$, $m = 5.8023$ gm, $s = 7773.6\,\text{mg}^2$. The two last fit very well into their respective lines of regression and analysis of variance. It is not likely that the debasement will cause any disturbance due to hoarding, though the rate of wear will naturally change. For the silver alloy had already changed nearly 3 years ago from 11/12 to 6/12 fine, even the nickel of George VI appears to differ from the older composition. Even with the pure metal used for each denomination, including the rupee, the currency would have a value of metal well below its denomination; hence, the change to brass only emphasizes the most universal of all numismatic laws, the inevitable trend towards debasement in time of stress. For our purpose, there is a far more serious effect visible in the samples. The minting since 1939 shows a decided increase in variance, and the occurrence of overweight specimens shows that the only legal remedy (from 1/40 for copper to 1/200 of silver) has been relaxed in practice, whatever the law at present. If this tendency was present in the coins struck during the last Great War (1914–1918), or during the depression years, it is certain to upset the linearity of variance increase, without affecting the law for mean weights. Whether the tendency towards cruder striking of the coins with regard to weight is manifested in the other countries and periods before great changes of structure will also have to be studied with this example in mind.

I am grateful to the kind friends who saved me much of the labour of gathering the samples in an unusually hot summer. Special thanks are due to my geological colleague Prof. K.V. Kelkar for going out of his way to place the facilities of his laboratory at my disposal.

References

1. A. Kolmogoroff, Math. Annalen **104**, 415–458 (1931).
2. D.D. Kosambi, New Indian Antiquary **4**(1), 49 (1941).
3. D.D. Kosambi, Curr. Sci. **10**, 372 (1941).
4. R.A. Fisher, *Statistical Methods for Research Workers*, 7th edn. ex. 42.
5. The gold sovereigns have had almost no circulation, but if just two more specimens, dated 1887, 1897 (and used regularly for worship) are added to the sample accepted, the correlation takes the very highly significant value of 0.64, with very highly significant deviations from regression.
6. R.A. Fisher, *Statistical Methods for Research Workers*, 7th edn. ex. 33.

Chapter 13
A Test of Significance for Multiple Observations

D.D. Kosambi, Fergusson College, Poona

In the years at Fergusson College, Kosambi did extensive statistical measurements and was devising new ways of analysing the data. This work, a precursor to "Statistics in Function Space", was reviewed by Abraham Wald who says "For the purpose of discriminating multivariate normal populations with respect to their mean values, test functions have been introduced and studied by H. Hotelling, R.A. Fisher, P.C. Mahalanobis, R.C. Bose, S.N. Roy and others. If the number of variates as well as the number of populations is greater than two, the application of the test would require the knowledge of certain probability distributions which have not yet been tabulated". DDK proposes a method in this paper to overcome this lack, but falls short of convincing his reviewer Wald, who adds "The author states that F^* has the ordinary F-distribution tabulated by R.A. Fisher and others. It seems to the reviewer that this statement of the author would be correct only if the coefficients $\lambda_1, \dots, \lambda_p$ were chosen independently of the sample. Since $\lambda_1, \dots, \lambda_p$ are functions of the sample values, the sampling distribution of F^* will arise partly from the sampling variation of $\lambda_1, \dots, \lambda_p$ and consequently the distribution of F^* need not be the same as that of F.

1. A test of significant discrimination between two sample groups of multivariate observations can be made by Hotelling's extension [1] of Student's t test; R.A. Fisher's discriminating function [2] based on the multiple correlation coefficient; and the generalised distance [3] of Mahalanobis, Bose and Roy. In addition to these closely related T^2, R^2, and D^2 tests, Wilks [4] has suggested others which would not involve the group means entering into the first three, but these last, as well as

Published in Current Science (Bangalore) **11**, 271–274 (1942), and reviewed in Mathematical Reviews MR0007235 (4,107b) 62.0X by A. Wald. Reprinted with permission.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_13

D^2, necessitate new sets of tables. For the case of two variates, however, it has been shown [5] that the usual analysis of variance can be carried out exactly, using the z-tables of Fisher, provided the degrees of freedom are suitably readjusted.

Here, I propose to extend the Z-test partially to samples drawn from a normally distributed population in $p > 2$ linearly independent variates. I also consider briefly the limiting case in which the number of variates increases beyond an limit, which leads us to the discrimination between samples consisting of sets of whole curves. This has the advantage of theoretical simplicity, in that all finite dimensional normal distributions are special cases, in much the same way as polygonal area rules like Simpson's come under the general $\int y dx$ formula. If accepted, the method would extend the analysis of variance to such material as electrocardiograms, cranial shapes, temperature curves and the like. It is emphasised that the discrimination is performed by the best linear combination of the old variates and not by the characteristic roots as such that appear in the process.

The contents of the opening chapters of Courant–Hilbert, *Methoden der Mathematischen Physik I* (1931), are taken for granted in the deduction.

2. We use the tensor summation convention: a repeated index denotes summation over all values of the index. The variates $1, 2, \ldots, p$ are indicated by Greek indices; sampling values $1, 2, \ldots, n$ of each variate are indicated by an additional Latin index. Thus, $x_{\nu i}$ is the ith sample value of the νth variate. Without loss of generality, the population mean for each variate is taken as zero. The multivariate normal distribution has then the probability density $c \exp(-\phi/2)$ where ϕ is a positive-definite quadratic form in the p variates, c a constant so chosen as to make the total probability over the whole p-space equal to unity.

There exist infinitely many linear homogeneous transformations of the variates reducing ϕ to a sum of squares:

$$\left.\begin{array}{l} \phi = \sigma^{\alpha\beta} x_\alpha x_\beta = \delta^{\alpha\beta} y_\alpha y_\beta; \quad \delta^{\alpha\beta} = 0, \alpha \neq \beta; = 1, \alpha = \beta. \\ y_\alpha = a^\nu_\alpha x_\nu, \ \ |a^\mu_\nu| \neq 0; \ \ \sigma^{\alpha\beta} = \delta^{\mu\nu} a^\alpha_\mu a^\beta_\nu. \end{array}\right\} \tag{13.1}$$

The new variates y are therefore uncorrelated, each with unit variance. The methods of discrimination proposed are that of applying the Z-test in that particular one of the hypothetical y variates for which the observed samples give a maximum value of z. Let this be y_λ. For a sample of n observations, we have:

$$\begin{array}{l} \frac{1}{n} \sum_{i=1}^n y_{\lambda_i} = \bar{y}_\lambda = \bar{x}_\nu a^\nu_\lambda, \ \ \text{where} \ \bar{x}_\nu = \frac{1}{n} \sum_{i=1}^n x_{\nu_i}; \\ \frac{1}{n-1} \sum_{i=1}^n (y_{\lambda_i} - \bar{y}_\lambda)^2 = \frac{1}{n-1} \sum_{i=1}^n \{a^\nu_\lambda (x_{\nu_i} - \bar{x}_\nu)\}^2 = a^\nu_\lambda a^\mu_\nu s_{\mu\nu}; \end{array} \tag{13.2}$$

where $s_{\mu\nu} = s_{\nu\mu} = \frac{1}{n-1} \sum_{i=1}^n (x_{\nu i} - \bar{x}_\nu)(x_{\mu i} - \bar{x}_\mu)$.

The tensors $s_{\mu\nu}$, $s'_{\mu\nu}$ are unbiassed estimates of the normalised cofactors of the population tensor $\sigma^{\alpha\beta}$, calculated from the n, n' random multiple observations, respectively. Nothing is to be assumed known as to the actual values of $\sigma^{\alpha\beta}$ or of the normalising transformation coefficients a^μ_ν.

3. We now take a new vector variable $u^{\alpha} = a^{\alpha}_{\lambda}$, since λ is to be fixed for the problem in hand. The two quadratic forms $s_{\alpha\beta}u^{\alpha}u^{\beta}$, $s'_{\alpha\beta}u^{\alpha}u^{\beta}$ are positive definite because all principal determinants in any sampling matrix $||s_{\alpha\beta}||$ calculated as in (13.2) are Gram determinants, which are positive whenever the p variates are linearly independent. Our special discrimination problem is thus reduced to finding the maximum of $F = s'_{\alpha\beta}u^{\alpha}u^{\beta}/s_{\mu\nu}u^{\mu}u^{\nu}$ or of its reciprocal.

The answer to this is well known. All we need here is the greatest relative characteristic root of the two forms, *i.e.*, of the determinantal equation

$$\text{det.}\quad |s_{\alpha\beta} - \vartheta s'_{\alpha\beta}| = 0, \tag{13.3}$$

or of the reciprocal equation, interchanging s, s'. These roots are all positive. If arranged in descending order of magnitude, they have the minimax property: ϑ^{ν}, $1 < \nu \leq p$, is the smallest value assumed by the maximum of F when the u are subjected to $\nu - 1$ independent linear homogeneous restrictions. Thus, all we have to do here is to put $z = \frac{1}{2}|\log\vartheta|$ for the extreme root, using the z-tables of Fisher with degrees of freedom based on the samples alone, as for the single variate. The distribution of the greatest or of any other characteristic root does not enter into the argument, the ratio of the two hypothetically transformed quadratic forms being always that of two sample variances. What we have obtained is essentially an existence theorem to the effect that the change by means of a suitable linear transformation of coordinates (variates) can give a z-value as great as but no greater than the greatest relative characteristic root of the two sampling tensor matrices. So, the z-tables are to be entered with degrees of freedom one less than the number in the samples, in the absence of any other linear restriction on the variates than that incurred in measuring from the sample mean. It might be possible to use the other roots by compounding probabilities, but it must be kept in mind that the minimax property requires that our transformation coefficients not the variates be sufficiently unrestricted. For example, our method of deduction cannot be called valid for $p = 1$, $p = 2$, as there are then not enough of the a^{μ}_{ν} left free, for a maximum to exist necessarily, after reducing the population form to a sum of squares. Of course, this is immaterial in view of the fact that p=1 is trivial and $p = 2$ settled by means of a special device [5]. In each of these cases, it is true that no *greater* z-discrimination is possible with linear combinations than is indicated by our test.

4. One advantage of the extension is that it holds for any $p > 2$. The ordinary analysis of variance is to be carried out exactly, in view of the fact that any sampling matrix may be broken up into various additive components die to the sources between which one wishes to discriminate. There is the further advantage that in case significant discrimination has been shown, the residual matrix of $||s_{\mu\nu}||$ may be used as the fundamental matrix in Hotelling's T^2 in the same way that the residual estimate of variance is used for Student's t test after the analysis of variance in a single dimension. The disadvantage is that our test would not be so powerful as others in rejecting H_0 when it is false; H_0 here being the null hypothesis that the various sampling tensors are pairwise compatible estimates of the same population tensor.

One method of calculating the extreme root has been given by Fisher (*SMRW* ex. 46.2) who uses divided differences. But Eq. (13.3) also lends itself to approximation for the greatest root by means of root squaring. Where the greatest root is not multiple, the rule can be stated immediately, without going into the very simple proof. We define: $\Delta = |s_{\alpha\beta}|$; $\Delta' = |s'_{\alpha\beta}|$; and Θ is the sum of the p determinants formed by substituting in rotation a single row in $||s_{\alpha\beta}||$ by the corresponding row of $||s'_{\alpha\beta}||$, and Θ' has the same function interchanging s, s'. Finally, let $\Delta_m, \Delta'_m, \Theta_m, \Theta'_m$ be the corresponding functions constructed by squaring (iteration) m times, according to the rule for matrix multiplication, each of the two matrices. Then, an approximate value of z for maximal significance is the greater of

$$\frac{1}{2^{m+1}} \log\left(\frac{\Theta_m}{\Delta_m}\right) \quad \text{or} \quad \frac{1}{2^{m+1}} \log\left(\frac{\Theta'_m}{\Delta'_m}\right). \tag{13.4}$$

Approximation is quite rapid when the greatest root is isolated. For a multiple root, the ratio Θ/Δ must be divided by a factor corresponding to the multiplicity; a similar precaution should also be taken for roots very close together.

5. Still more interesting is the passage to the limit. Suppose we have to deal with silhouettes taken on the profiloscope. One method would be to take some well-defined point such as the ear orifice for the origin and some well-defined line such as that from the origin to the base of the nose as prime vector, and to expand the distance from the origin to the general point of the profile as a Fourier series in terms of the angle from the prime vector. The coordinates would then be the Fourier coefficients; if enough were determined to permit the reproduction of any profile to within the original limits of observation, our test or any suitable multivariate test could be applied directly. Yet this is clearly unsatisfactory in that we are using a finite number of coordinates in an indefinite number of dimensions without knowing anything of those discarded. The argument that professional anthropometrists do this or worse in using a finite number of characters instead of our harmonic analyser, without proving normality of the distribution, does not suffice. So, we take the other form of the passage to the limit represented by integral equations.

We keep the original quadratic form, extended to infinitely many dimensions, and take the coordinates as "Fourier" coefficients associated with expansion in some given set of orthonormal functions defined over $0 \le x \le 1$, which is also to be taken hereafter as the range for all undefined integrations. The probability density will again be represented by $c\exp(-\phi/2)$, with

$$\phi = \int\int K(s,t) f(s) f(t)\, ds\, dt; \quad \bar{f}(s) = \frac{1}{n}\sum_{1}^{n} f_i(s) \tag{13.5}$$

$$S(s,t) = \frac{1}{n-1}\sum_{1}^{n}\{f'_i(s) - \bar{f}(s)\}\{f'_i(t) - \bar{f}(t)\}.$$

These now replace (13.1) and (13.2) in the function space, each multiple observation on the vacate taken to define a function $f(x)$ over $0 \leq x \leq 1$. For significance tests, the reciprocal to $S(s, t)$ is the best estimate of the population kernel $K(s, t)$. An alternative simultaneous visualisation of the space is, as before, the Hilbert space of the coefficients in the orthogonal function expansion of $f_i(x)$. Naturally, it is essential to take the population kernel $K(s, t)$ as positive semi-definite or definite; its characteristic functions form the most convenient orthogonal functions to use for theoretical purposes, which amounts to using a quadratic form with diagonal matrix. If the characteristic orthonormal functions do not form a closed set, as many more are to be adjoined as are necessary for closure, taking the additional coordinates associated with these extra functions to constitute the orthocomplement to the function manifold of $K(s, t)$. In probability integrations, these extra coordinates will be undetermined, hence to be integrated over the whole of the orthocomplement. This allows all kernels to be considered in a proper function space, even the degenerate kernels that actually include the ordinary p-variate normal distribution; conversely, the p-variate case may be considered as associated with a degenerate $K(s, t)$, by ascribing one function of an arbitrary orthonormal set to each coordinate as coefficient. For limits of integration, we use the convenient as well as fashionable terminology of lattice theory, taking $f \smile g$, $f \frown g$, respectively, as the functions whose "Fourier" coefficients are the greater and the lesser of the corresponding coefficients in the expansion of f and g. Thus, the integration can extend from $f \smile g$ to $f \frown g$, and over the whole of the orthocomplement, whenever integration "between" two function limits f, g is to be performed.

6. The trouble with all this is that it has only an appearance of verisimilitude. In a space of infinitely many dimensions, we have as yet failed to define the volume element. If we take the multiple integral over infinitely many dimensions as evaluated by successive iterated integrals in the usual manner, it will be seen that any consistent evaluation making the total probability unity leads in general to zero probability in integrating over any proper sub-manifold of the whole space. One must go much deeper than the intuitive methods of **5.** It is seen that if we merely take the limits increasing the number of dimensions, the "volume" of a hypercube is 0, 1 or ∞—of a hypersphere zero—as the n-dimensional sphere has the volume of $2\pi^{2n} r^n / n\Gamma(n/2) \to 0$ as $n \to \infty$.

This difficulty is surmounted under the hypotheses that the abstract space under consideration has a distance relationship obeying the usual postulates; it is separable, locally compact, with a congruence relation. The two middle ones have to remain assumptions, distance r being defined by $r^2 = \phi(f - g)$, for any two elements f, g. The space has to be restricted to elements for which $K(s, t)$ is a positive-definite kernel. Congruence of two regions may be taken as transformability of one region into the other by some member of a suitably restricted (linear) transformation group, preserving $\phi(f - g)$ and transforming the entire manifold into the entire manifold. Then, a Haar measure [6] and a Lebesgue–Stieltjes integral exist. Unrestricted Hilbert space is not locally compact because no infinite sequence of orthogonal functions can converge in L^2.

It follows that all classical results can be stated and proved again in general abstract spaces, though it is better for our purpose to take kernels of the second (Fredholm) kind for some theorems, which means only the addition of a term $\int f^2 ds$ to the ϕ of (13.5). We may then state such results as follows: *the sum of two normally distributed variates is also normally distributed with mean the sum of the two means and kernel whose (formal) reciprocal is the sum of the two (formal) reciprocals of the given kernels.*

Many fundamental procedures and distributions may be generalised to this space, including some of the more powerful tests considered by Hsu [7]. Not only can the Hotelling–Fisher formulæ [2] be derived from a degenerate population kernel of p degrees of freedom, but a space of sufficiently large (or infinite) number of dimensions would lead to corresponding formulæwith $p = n$, the degrees of freedom within groups. It is clear, however, that the nature of the fundamental abstract space associated with a given population will not be revealed in general by means of the sample taken by a practising statistician; here, I regret my inability to demonstrate with a practical example, for which there is data enough, but no access to the necessary machines: ordinary or cinema integraph, differential analyser, etc. In any case, it is clear that a test which applies independently of dimensionality,[1] without new tables, becomes of importance whether or not more efficient and powerful tests could be devised for the particular unknown population in question. This test is the analogue of (13.3); taking limits, we state it as the problem of locating the extreme characteristic root of $\int \{S[s,t] - \vartheta S'[s,t]\} f \, dt = 0$. By noting that the sample kernels S, S' are degenerate, this can be reduced to a set of linear equations in a finite number of unknowns, whence the existence of a finite number of positive determinate roots follows at once. It is proposed that the extreme root be used as before for the Z-tests; the estimating kernels may still be broken up into additive components, permitting analysis of variance. It would, of course, be convenient to have the distribution of certain sampling functions, as, for example, of $\int\int S^{-1} S' ds \, dt$, where S^{-1} is the reciprocal to $S(s,t)$.

References

1. H. Hoteling, The generalization of student's ratio. Ann. Math. Stat. **2**, 360–378 (1931).
2. R.A. Fisher, *Statistical Methods for Research Workers*, 7th edn. 294—298 (1938).
3. P.C. Mahalanobis, *Proc. Natl. Inst. Sci. India***2**, 49–55 (1936); R.C. Bose, *Sankhyā***2**, 143–154, 379–384 (1936); S.N. Roy, *Ibid.*, 385–396.
4. S.S. Wilks, Certain generalizations in the analysis of variance. Biometrika **24**, 471–494 (1932).
5. D.D. Kosambi, A bivariate extension of Fisher's z-test. Curr. Sci. **10**, 191–492 (1941).
6. S. Banach, *Thèorie de l'Intègrale*, ed. by S. Saks (1933), pp. 204–272.

[1] If the Haar volume of the sphere, $\phi(\ell) \leq r^2$ is cr^k, we have the usual k-dimensional space or its equivalent. But we also get fractional dimensionality when k is non-integral. So, the degenerate kernel need not necessarily lead to the ordinary p-dimensional case. For the existence and construction of point-sets with fractional dimension, see [8].

7. P.L. Hsu, Biometrika **31**, 221–237; Ann. Math. Stat. 9, 231–243 1939. J. London Math. Soc. **16**(1941), 183–194 (1940).
8. F. Hausdorff, Math. Annalen **79**, 157–179 (1919).

Chapter 14
On Valid Tests of Linguistic Hypotheses

D.D. Kosambi, Fergusson College, Poona

Although it is not clear if they knew each other, the linguist Zipf who graduated in 1923 was a near-contemporary of DDK's at Harvard. Zipf went on to study in Berlin and Bonn, returning to Harvard to do his masters and PhD, eventually becoming university lecturer in 1939. DDK is scathing in his criticism of Zipf's law as not being statistically significant. After eighty years and numerous applications, the debate is still on; the validity of this "law" and its applicability are investigated to this day, so it would seem that DDK's reservations on the statistics used had some justification.

It is known that in any connected piece of writing ["language stream"] the number of words used twice is far less than that used only once. The number occurring three times is still less, and the drop continues rapidly. The Harvard philologist George Kingsley ZIPF has proposed a "law" for this, the number of words used n times being, according to him proportional to n^{-2} (**1**, 24; **2**, 40–44). The main purpose of this note to raise serious objections to this inverse square "law". These objections are statistical. I maintain that no such law, whatever the exponent, will do for the data so far given because the fit is not sufficiently good even when the best exponent is taken by calculations on the logarithmic scale. (**1**, 25–26; **2**, 43; **5**, 63). To put this in non-technical language: to every head, there will be one cube-shaped wooden box that fits best, but in general, a rubber cap or a felt hat of the right size will fit better, and the latter is more likely to indicate a contour of the skull.

1. As my attention was first called to the problem by the Old-Kanarese word counts of Mr. M.G. VENKATESAIYA (working under the direction of Mr. C. R. SANKARAN), I shall illustrate the accepted statistical method by an application to his data. K, V, P, denote three works in Haḷagannaḍa, entitled the *Kavirājamārga, Voḍḍārādhane*, and

Published in New Indian Antiquary **5**, 21–24 (1942).

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_14

Table 14.1

Observed				Expected			
Fr.	*K*	*V*	*P*	Totals	*K*	*V*	*P*
1.	3241	2990	1087	7318	3220.6	3041.3	1056.1
2.	270	301	62	633	278.6	263.1	91.3
3.	62	71	19	152	66.9	63.2	21.9
4.	40	45	14	99	43.6	41.1	14.3
5.	29	22	7	58	25.5	24.1	8.4
6.	39	47	18	104	45.8	43.2	15.0
Total	3681	3476	1207	8364	3681.0	3476.9	1207.0

Pampāśatakam, respectively. For purpose of testing it will be necessary to group together the small frequencies at the end, and sufficient to present the counts as follows:

The expected numbers are calculated on the assumption that the three works are uniform in the structure of their language stream, whence it follows that the ratio of the figure in each "expected" cell to the total at the foot of its column must be the same as the corresponding ratio of the marginal totals. The numbers so obtained are rounded off to the first decimal, taking due care to preserve the totals each way. As it is clear that the expected and observed totals will never coincide in practice, some method of calculating the magnitude of the discrepancy and of judging its seriousness is necessary. This, for the care in hand, is Karl PEARSON'S χ^2 test, χ^2 being the sum obtained by squaring the difference between each expectation and observation, and dividing the square by the expected number. This sum is here about 22.25, and inasmuch as ten of the given eighteen entries could have been made at will without disturbing the totals, we enter the tables of χ^2 (to be found in any standard text on statistics, such as R. A. FISHER'S *Statistical Methods for Research Workers*) with 10 degrees of freedom. It is then found that the probability of exceeding this value of χ^2 lies between 0.01 and 0.02. That is, we should, on the hypothesis of uniformity between the three works, expect to obtain such a result not oftener than once in fifty times, but not so rarely as only once in a hundred trials. This is hardly in favour of the hypothesis, though the "level of significance" is to some extent a matter of individual choice, just as the fit of a hat would depend upon the wearer. If P were smaller than 0.05, as it is here, the statistician would take the hypothesis as contradicted, following the standard practice of his trade.

This test is surely more exact than anything suggested by ZIPF (5) or his critics (4), judging from the reference material to which I have access here. If the same test be applied to the data for the *K* and the *V*, it will be found that the two works are compatible, P being not less than about 0.2, which is not at all serious. That is, the *Kavirājamārga* and the *Voḍḍārādhane* follow about the same frequency laws, but the *Pampāśatakam* is decidedly of a different nature. The main cause of the discrepancy lies in words of frequency two, of which the *V* has too many and the *P* far too few.

2. Applying this χ^2 test to ZIPF'S data, we reach the following conclusions: Taking together his numbers for Chinese and Plautian Latin with ELDRIDGE'S for American newspaper English (**1**, 23: **2**, 26–28), the value of χ^2 is enormous and virtually excludes the very notion of uniformly. Of the three, Peiping Chinese and Plautian Latin are closest together, as would be expected from the fact that Eldridge did not count numerals and proper nouns (**2**, 25). We note in passing that the totals as given by ZIPF need two corrections, that for Chinese being 3342 instead of his 3332, and for Eldridge's English, 6001 in place of 6002. Testing the two languages counted by ZIPF, however, we find χ^2 about 40.8, which for 17 degrees of freedom gives a probability of 0.001, almost exactly, about one chance in a thousand that the two languages follow the same frequency law, the discrepancy arising mainly in frequencies 5 and 15.

Finally, the same test applies to any proposed law of frequency, in particular to the inverse square law. For sufficiently extended counts, the expected number of words occurring n times would be given by $6N/(\pi n)^2$, or 0.6079 N/n^2, where N is the total number of distinct words counted. The square of each discrepancy is again divided by the expected number; the ratios are added together for the value of χ^2. It will be found that of all the six sets of counts cited here, the "law" applies best to Chinese. It is again necessary to group together the smaller frequencies at the end (in testing by χ^2 the expected frequency should not in any cell fall much below ten) and for 17 degrees of freedom, I obtain a value of $\chi^2 = 27.17$ whereas the value of P 0.05 is 27.587. The fit, then, is hardly satisfactory; the best that can be said about the proposed law is that the data for Chinese does not contradict it so decisively as that for the remaining languages.

3. To apply these simple tests, little knowledge of statistical theory, none of pure mathematics, is required. The labour involved is trifling when it is considered that final conclusions are to be drawn from data far more laboriously compiled and that their validity is to be tested. It is surprising, therefore, to note that nowhere in the work of ZIPF, nor in the criticisms of JOOS (**4**) nor the argument advanced by an able mathematician like STONE (**5**, 60–61, 63–64) is there any idea of testing goodness of fit or significance. As the U.S.A. are fortunate in possessing many statisticians of eminence, I shall offer a few suggestions here, and leave it to the philologists to work them out, if they see fit to do so.

Table 14.2 Analysis of variance

Source	d.f.	Sum squares	Mean sq.	Ratio
Languages	12	15.424060	1.285338	$(1.0987)^{-1}$
Block t·b. vs. m·r	1	76.736862	–	55.75***
Consonants within a block	8	376.270187	47.033773	33.3046***
Lang. × blocks	12	25.192508	2.099376	1.4866*
Residual	96	135.574263	1.412233	(s.d. 1.18837)
Total	129	631.197880	4.893007	3.4647

None of the inverse exponent laws fit at all well, though each exponent may be said to characterize the sample from which it was calculated just as the best fitting cubical box would characterize a skull. For KAEDING'S data (**2**, 23), the three counts given by ZIPF, as well as the three of Kanarese with which I illustrated the χ^2 test, a type B series derived from the Poisson distribution or one of Neyman's "contagious" distributions (**6**) would be found, to fit far better. But the same series would not do for all the samples any better than the same box or hat for all heads: The statistics would be of a descriptive type, lacking the attractive if fictitious Newtonian simplicity of the inverse square law, supplemented by an appeal to SCHRÖDINGER, HEISENBERG, AND DIRAC (**5**, 61). Another interesting possibility, if a Poissonian or type B series is found to fit well, would be of estimating the passive vocabulary of the stream, words not used at all, by extrapolation; the "maximum-likelihood" formulae for estimating the words of zero frequency from a supposed Poisson distribution can be worked out very easily, but are not given here inasmuch as the said distribution, which is virtually a random distribution, does not fit.

A far more serious matter is that of properly randomized sampling. ZIPF and his followers wish to characterize an entire language, sometimes all languages, by means of their counts. But the total number of words in the respective language streams is always enormous in comparison with the number that can be counted (with obvious exceptions like Anglo-Saxon or Sumerian); therefore, every precaution has to be taken to avoid bias. This, again, is a matter to which the statisticians have devoted a good deal of time; standard methods of randomization exist which might very well be considered *before* the work of counting is begun. It is to be noted that ZIPF'S *scattering point* (**1**, 24) disappears with increased size of the sample, as well as in out test of significance.

Finally, it must be stated that statistics is not just a laborious method of contradicting the pleasing conclusions obtained by the common sense of the philologist. For example, analysis of variance may be applied to the combined data for thirteen languages (**3**, 61, 65) using the percentages given by ZIPF. The conclusions are that the languages are remarkably uniform that there is no difference between the classic and the modern languages and that there is a tremendous difference between the consonants *t d k g p b* on p. 61 and the *m n l r* on 65, whether they be taken in these two blocks or separately. For any two entries in ZIPF'S table, the difference of 3.36 % is to be taken as significant at the 5 % level; for the means between two languages, this should be divided by $\sqrt{10}$, for two consonants, by $\sqrt{13}$. A caution is necessary in that the use of percentages can be objectionable: If all the percentages were taken, every language would have the same total 100. But if the use be allowed in the present case, the information which I give and which does not contradict ZIPF is partially summarized in the following table:

Here, the blocks are the two sets of consonants. It is seen that the languages behave differently in the two sets, but this has not the enormous significance of the difference between consonants.

References

1. G.K. Zipf, *Selected Studies of the Principles of Relative Frequency in Language* (1932).
2. G.K. Zipf, *The psychobiology of Language* (1936).
3. G.K. Zipf, *Harvard Studies in Classical Philology*, vol. XI (1929), pp. 1–95.
4. Martin Joos, *Language*, vol. XII (1936), pp. 196–210.
5. G.K. Zipf, *Language*, vol. XIII (1937), pp. 60–70.
6. Jerzy Neyman, *Annals of Mathematical Statistics*, vol. X (1939), pp. 35–57.

Chapter 15
Statistics in Function Space

D.D. Kosambi, Fergusson College, Poona

Had DDK followed up on this paper in a more systematic manner, and if attribution had been given more properly, the Karhunen–Loève theorem may well have been known more widely as the Kosambi–Karhunen–Loève theorem. This work, which essentially lays the foundation of the proper orthogonal decomposition technique for analyzing random signals, predates the work of Karhunen and Loève by several years, and although published in the Journal of the Indian Mathematical Society, it was reviewed in the more widely circulated Math. Revs. by J.L. Doob who summarized it succinctly: The author discusses statistical problems connected with continuous stochastic processes whose representative functions $x(t)$ *are defined by* $x(t) = \sum_j x_j \phi_j(t)$, *where the* ϕ_j *determine an orthonormal set and* $x_1, x_2, \ldots$ *are mutually independent Gaussian chance variables with vanishing means and variances* $\sigma_1^2, \sigma_2^2, \ldots,$ *respectively. [...]. The samples he considers are functions* $x(t)$ *rather than merely the values of functions* $x(t)$ *at a finite number of points. [...]. Various mechanical and electrical methods are suggested for combining functions x(t), given graphically, as necessitated by this type of statistical approach.*

The mechanical and electrical methods suggested were pursued to some extent in the "Kosmagraph" project (see Chap. 4) and were to inspire him in constructing various computing machines. Beyond summarizing these results in [DDK44]*, however, DDK did not expand on this work.*

The main purpose of this note is to develop statistical methods for discrimination between samples consisting of whole curves.

Journal of the Indian Mathematical Society **7**, 76–88 (1943), and reviewed in Mathematical Reviews MR0009816 (5,207c) 60.0X by J.L. Doob. Reprinted with permission.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_15

We take the observables as simple curves of type $y = f(x)$, the functions $f(x)$ being all single valued, of bounded variation, continuous (though stepwise continuity—as for a sample of histograms—would cause no difficulties), defined on a finite closed interval which may be taken without loss of generality as $0 \leq x \leq 1$ by suitable choice of origin and scale. The methods developed for such curves apply directly to diagrams in polar coordinates $r = f(\theta)$, $0 \leq \theta \leq 2\pi$, by an obvious extension, to suitably restricted surfaces (say crania) or multidimensional varieties. Peano's space-filling curves, Jordan curves of positive area, are naturally excluded.

The problem clearly resolves itself into four components: (1) to define a normal distribution in function space, (2) to deduce useful consequences of such normality, assumed to hold for population of curves, (3) to devise new methods of calculation where necessary, and (4) to examine the generality of the approach.

1. The probability P associated with a multivariate normal distribution is the definite integral, over the proper region, of

$$(2\pi)^{-k/2} e^{-\phi/2} dV, \tag{15.1}$$

where k is the number of variates in which ϕ is a positive definite quadratic form, and dV is the associated volume element. That is, the same transformation that reduces ϕ to a sum of squares makes $dV = dx_1 dx_2 \ldots dx_k$, the whole space being recognizable as an ordinary k-dimensional Euclidean manifold with $\phi = r^2$ as the square of the distance. A continuous function is determined completely by its values on a set of points everywhere dense on (0, 1), say all rational points; a function in general, therefore, has an infinity of coordinates. As approximation is possible by increasing k indefinitely, the first step would be to generalize distance and the quadratic form ϕ.

To this end, we assume the existence of a continuous symmetric kernel $K(s, t)$, positive definite or semi-definite. Then, the distance between any two of our functions $f(x)$, $g(x)$ is given by

$$r(f, g) = \phi(f - g) = \int\int K(s, t)\{f(s) - g(s)\}\{f(t) - g(t)\} ds\, dt. \tag{15.2}$$

The range for all otherwise undefined integrals in s and t is (0, 1) for each variable, here, the unit square. Restricting the population of functions to be such that $K(s, t)$ gives a definite ϕ therein according to the definition of (15.2), we shall have $r(f, g) = 0$ if and only if $f \equiv g$ with respect to the mechanism of observation; it follows, therefore, that r obeys all the basic postulates for distance including the triangular inequality. The normal distribution in function space could be taken as defined by $c \exp(-\phi/2)dV$.

Unfortunately, not all terms of this probability density can be given a meaning that is useful in practice. As $(2\pi)^{-k/2} \to 0$ when $k \to \infty$, c can only be specified by the restriction that the total probability equals unity; also the "volume element" dV may be given a direct mathematical meaning [1], but not one of much real use. To surmount this obstacle, we resort to a choice of independent variates that reduces ϕ to a diagonal form. This amounts to taking $K(s,t)$ in its canonical form [2, 117, 114]

$$K(s,t) = \sum \sigma_i^2 \phi_i(s)\phi_i(t). \tag{15.3}$$

The ϕ_i are the orthonormal characteristic (eigen-) functions of the kernel, σ_i^2 the corresponding characteristic values (=$1/\lambda_i$ in the notation of **2**), all positive with $\sum \sigma_i^4$ convergent [2, 111]. The orthogonal or independent coordinates for any function $f(t)$ are obviously the "Fourier" coefficients $x_1, x_2, \ldots, x_r, \ldots$ with

$$x_r = \int f(t)\phi_r(t)dt, \quad f(t) = \sum_r x_r \phi_r(t). \tag{15.4}$$

As $K(s,t)$ is definite *for the population*, every function served can be so represented. The series will converge uniformly and absolutely [2, 114] for every function that is the $K =$ transform of any piecewise continuous function, a restriction which we place upon the population.

We now define normality in the function space to mean normal distribution in each of the x_i. Without loss of generality, the population mean for the function space and hence for each x_i may be taken as zero. The variances will be σ_i^2. That is, our ϕ has to be taken as generalizing not the k-dimensional population quadratic form, but the one that enters into the characteristic function of the distribution, the Fourier transform. In Euclidean space, both become equal to r^2 when ϕ is expressed as a sum of squares. Our choice for function space is dictated by the implication of (15.4) that $\sum x_i^2$ converges whence $x_i \to 0$ as $i \to \infty$, which can be made to hold *in probability* for random x_i if and only if $\sigma_i^2 \to 0$. If ϕ were to be taken as the population (probability density) quadratic form, or variances would become $1/\sigma_i^2$, the two kernels must be reciprocals, as is seen by application of the Fourier transform to the k-variate distribution. Since we deal here with kernels of the first kind, only one can be properly defined, in general, the other "existing" only in symbolic form as is seen by the fact that except in the degenerate case, the series of squares of characteristic values and the series of squares of their reciprocals cannot both converge simultaneously. To sum up, we may formulate definitions:

Definition 1 Normal distribution in function space will be taken to mean normal distribution for each variate x_i of an (independent) infinite sequence $x_1, x_2, \ldots,$ with all population means zero and variances $\sigma_1^2, \sigma_2^2, \ldots$ These x_i are the "Fourier" coefficients of a random function of the space with reset to the orthonormal characteristic functions $\phi_1(t), \phi_2(t), \ldots$ with characteristic values $\sigma_1^2, \sigma_2^2, \ldots$ belonging to a continuous, symmetric, positive, definite (in the manifold of admissible functions) kernel $K(s,t)$ defined over the unit square $0 \le s \le 1, 0 \le t \le 1$. The kernel $K(s,t)$ thus completely determines the distribution.

2. This definition has the initial advantage of covering all finite-dimensional cases, represented by degenerate kernels where all but a finite number of the variances σ_i^2 vanish. Conversely, it allows approximation by degenerate kernels and application of methods developed for k-variate distributions. These can be used together to prove, for example, that

The sum of two normally distributed function variates is also normally distributed with mean the sum of the population means and kernel the sum of the two given kernels.

This follows directly from the definition since $\exp(-\phi/2)$ is not the probability density but the characteristic function of the distribution. The characteristic function of the sum of two variates is the product of the two characteristic functions. In particular, the mean of a sample of n functions chosen at random from the same normal population will have the population mean and kernel K/n; one degree of freedom will be lost in measuring from the sample mean, and so on. If the same set of orthonormal functions covers both kernels, then we may add corresponding variances as usual; if no, we can at least state that the sum variances do not decrease (**2**, 113 with an obvious correction).

What seems to me to be the most important consequence of our definition rests upon a basic theorem of Kolmogoroff [3]. Using the letter P to indicate the probability and E the expectation of the events bracketed, this may be stated as follows.

Given a random sequence $u_1, u_2, \ldots, u_r, \ldots,$ *the probability of the convergence of the series* $\sum u_r$ *is unity if there exists some random sequence* $v_1, v_2, \ldots, v_r, \ldots$ *such that the three series* $\sum P(u_r \neq v_r)$, $\sum E(v_r)$, $\sum E[(v_r - E(v_r))^2]$ *all converge. If no such sequence exists, the probability for the convergence of* $\sum u_r$ *is zero.*

In our case we take $u_r = v_r = x_r \phi_r(t)$, so that the first two series converge by hypothesis. The third is

$$\sum E(x_r^2 \phi_r^2) = \sum \phi_n^2(t) E(x_n^2) = \sum \sigma_n^2 \phi_n^2(t) = K(t,t) \quad \text{(by } \mathbf{2}\text{, 110)}.$$

The structure of Kolmogoroff's proof shows that in our case convergence and uniform convergence go together. We conclude, therefore, that

A random sequence of the variates $x_1, x_2, \ldots, x_r, \ldots,$ *of our definition represents with unit probability a function of the normally distributed function population.*

This replaces the Riesz–Fischer theorem, proving a 1-1 correspondence between random functions and random sequences of coefficients, in the sense of unit probability. The Riesz–Fischer theorem would require, for unit probability, the convergence of $\sum \sigma_r^2$ and give only convergence in the mean for $\sum x_r \phi_r(t)$.

Let $y_i = f(t_i) = \sum x_r \phi_r(t_i)$ be the ordinate at a fixed point t_i as abscissa. From $E(x_i x_j) = 0$, $E(x_i^2) = \sigma_i^2$, we obtain

$$\left.\begin{aligned} E(y_i^2) &= \sum \sigma_r^2 \phi_r^2(t_i) = K(t_i, t_i) \\ E(y_i y_j) &= \sum \sigma_r^2 \phi_r(t_i) \phi_r(t_j) = K(t_i, t_j) \end{aligned}\right\} \qquad (15.5)$$

The matrix $||E(y_i y_j)|| = ||K(t_i, t_j)||$ of covariances may be regarded as the symbolic product of $||\sigma_r \phi_r(t_i)||$ with the transposed matrix. In case the kernel is degenerate and there are more ordinates $y_i, y_2, \ldots, y_m$ than characteristic functions $\phi_1, \phi_2, \ldots, \phi_k$, it is obvious that the determinant $|E(y_i y_j)|$will vanish. Conversely, if $|K(t_i, t_j)|$ vanishes identically, the kernel is necessarily degenerate as its Fredholm expansion [2, 122] breaks down into the ratio of two polynomials. As the y_i are convergent (in the sense of unit probability) linear combinations of normally distributed variates, we have proved that

For any fixed abscissa t, in a normally distributed population of function, the ordinate is normally distributed with variance $K(t, t)$. The covariance between values of the functions at two points s and t is $K(s, t)$. The distribution of ordinates at k fixed points is multivariate normal and is a proper distribution except when the kernel $K(s, t)$ is itself degenerate with less than k characteristic values.

In this, of course, the nodal points of the entire set of functions, points where $K(t, t) = 0$ in particular, must be avoided when selecting the k points for measuring abscissae. An example would be $\phi_r = \sqrt{2} \sin \pi r t$ and the end points of the interval, $t_1 = 0, t_2 = 1$.

3. The usefulness of the preceding section is manifest, as the population kernel $K(s, t)$ would remain unknown in practice even when the hypothesis of normality is granted. Our theorems enable us to proceed by the methods of the ordinary multivariate normal distribution, measuring ordinates at suitably chosen abscissae. The meteorologist would be justified in working with temperatures taken at noon and midnight, but not necessarily with his maximum and minimum temperatures, which are measured at varying times of the day. The anthropologist's characters and indices would be less justified, than, say, measurements from the ear orifice to the profile at fixed angles from the line joining the orifice to the base of the nose. Coefficients on the harmonic analyzer and regression coefficients in properly chosen orthogonal functions (whether they belong to the kernel or not) are also to be regarded as coordinates in multivariate normal distribution, provided the fitting when done by values at fixed points is done with the same fixed abscissae for each curve.

Given a sample of n curves, $y = f_1(x), f_2(x), \ldots, f_n(x)$, the best estimate of the population mean $\mu(x)$ and the population kernel $K(s, t)$ are given, respectively, by

$$m(t) = \frac{1}{n} \sum f_i(t);\ k(s, t) = \frac{1}{n-1} \sum [f_i(t) - m(t)][f_i(s) - m(s)] \qquad (15.6)$$

as follows obviously from the foregoing. Large sample theory means calculation of these sample functions and therewith the characteristic functions and values [which will approximate those of the population]. Hotelling's $\mathcal{T}^2$, Fisher's discriminating function, and such methods for discrimination would also apply without restriction provided the number of points for taking ordinates did not exceed the number of functions in the samples.

But in many cases the complete curves are recorded automatically with less trouble and more accuracy than for a finite number of observations on the same material. In that case, we could, if the proper machines were available, calculate the sample functions in (15.6) and thereafter the "Student" ratio $t(x)$ or Fisher's $z(x)$ from two given samples for every point of the abscissae $0 \leq x \leq 1$. Corresponding to these or to any other statistics, we shall get a probability $p(x)$ as a function over the unit interval. In methods of discrimination, one may choose a single point, say the point where $p(x)$ takes on its maximum value in the closed unit interval; the corresponding value of x gives the abscissa where the maximum discrimination has been achieved. this, in a way, is the determination of the best [in the obviously restricted sense] linear combination of the unknown coordinates x, the "Fourier" coefficients with respect to the unknown population orthogonal functions. But, again, one is tempted to ask whether something more could not be done, whether one could not calculate or measure a single probability for the whole interval or of any given subinterval, instead of a point probability. What is required is not $p(x)$ but a $P(\alpha, \beta)$ for any given $0 \leq \alpha < \beta \leq 1$, on the basis of the two samples and any given statistic. The most that can be done here is to show that such questions need not be meaningless.

Suppose the kernel to have a single nonnegative characteristic function $\phi(t) \geq 0$ and characteristic value σ^2. We ask for the probability that a sample function $f(t) = x\phi(t)$ lies between the two limits $a(t)$ and $b(t)$ throughout an interval (α, β) in (0, 1). Then, this probability is

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} \exp(\frac{-x}{2\sigma^2})dx, \tag{15.7}$$

if x_2 is the greatest value of x satisfying $x\phi(t) \leq b(t)$ and x_1 the least for $x\phi(t) \geq a(t)$ in (α, β), with $x_2 > x_1$; the probability is zero otherwise. A similar approach is possible for $\phi(t)$ with changes of sign or more than one characteristic function. The general question, for examples of the type chosen for illustration, depends upon the correspondence that can be set up between two different types of function-em lattices, not merely function spaces, with measure and maps upon the unit hypercube in infinitely many dimensions (torus space).

The calculating machines, under the circumstances that now limit my activity, cannot go beyond the stage of design. The fundamental ideas will be made clearly by the two schematic figures appended here in the hope of doing service to some more fortunately situated experimenter. Figure 15.1 shows how $\sigma f_i(x)$ may be drawn by means of templates and a wire passing over a system of alternately fixed and moving pulleys, suggesting by Kelvin's tidal machine. The formulae of (15.6), in particular the most important ones for $m(t)$ and $k(t, t)$, depend upon the operations of addition, summation of the square, subtraction, and division by a chosen factor. For the pulley machine, subtraction is possible by reversing the direction of the ire or rather substituting a moving for a fixed pulley; or by using as template the conjugate curve to the one to be subtracted. Reduction of scale, i.e., division by a given number, will have to be done by a pantagraph, or some such device. Both of these introduce errors, and there is the additional difficulty of getting material for templates that will be stiff enough to stand up under the weights, and smooth enough to allow all the templates to be pulled through on their rack without sticking. For sum squares, and

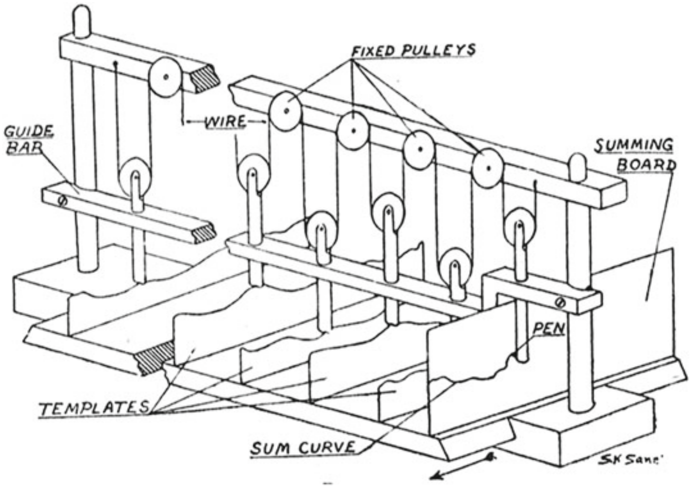


Fig. 15.1

sums of products, the arrangement has to be extended to the measurement of torque and moments, the simple pulley machine being inadequate.

The second instrument is suggested by the high fidelity with which sound is recorded on and reproduced by cinematography. Here, the area under the curve is cut out of standard sheet of paper and scanned by means of movement past a narrow fixed slit. The light that falls on the slit is of uniform intensity throughout, in the first instance. The film is coupled with the template, and both are drawn through with uniform speed. The lens reduces the curve in height, but not in length, and by means of a vertically movable rack, many such curves, say at least a dozen, may be easily recorded on a single film. At the end of each curve template, a standard height is cut out of the template material.

The film is developed and printed as usual and focussed back in all its width through another slit on to a photoelectric cell. The current recorded will be proportional to $\sum f_i(x)$ at each point, standardization being achieved by means of the fixed heights cut out of each template at the end of the curve. The factor $1/n$ or $1/(n-1)$ can also be set, thereby adjusting the primary current, or putting the proper shunt across the current-recording device.

Sums of squares are easily obtained by varying the intensity as well. This is best done by means of a photoelectric cell coupled with the moving template rack. This cell would regulate the current supplied to the light, so that we should have the product of the height of the curve by the intensity as $f_i^2(x)$. The difficulty here, of course, is in the law of darkening, and very much more careful adjustments will have to be made. The same method allows function covariances to be calculated, coupling one set of templates to the photoelectric cell and the other to the slit-rack.

The law of resistance in electric circuits in parallel shows obvious means of calculating harmonic means. For the two-dimensional kernels $K(s,t)$, the best methods

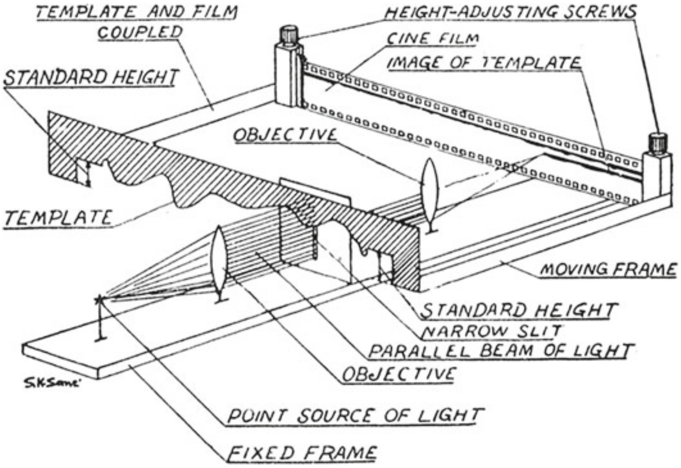


Fig. 15.2

would seem to be those derived from television. Such instruments are now being devised by others for work in a single dimension. If successful, the need for cutting templates would be obviated, with a gain in accuracy.

4. From the purely theoretical point of view, we have ignored many other possibilities. Some of these were mentioned in a former exploratory approach [1]. I give an example to show that theoretical generality is certainly possible, in defining the normal distribution, but that the usual facilities such as the central limit theorem, the chi-square, and other tests used in practice, in short the whole mechanism of everyday statistics is invalidated.

The population is defined by the functions

$$\begin{aligned} \phi_r(t) &= \sqrt{2}\sin \pi r t, \\ f(t) &= \sum 3^{-n/2} a_n \phi_n(t), \quad \text{where } a_n = 0 \text{ or } 2 \end{aligned} \tag{15.8}$$

The kernel $K(s, t)$ is given by

$$K(s, t) = [\sum 3^{-i/2}\phi_i(s)][\sum 3^{-j/2}\phi_j(t)] \tag{15.9}$$

being thus degenerate of the first order. It follows that $r(f, g)$ for two functions defined by sequences a_n, b_n is given by $r = |\sum(b_n - a_n)/3^n|$. If the sequence a_n is regarded as defining a point of Cantor's ternary set [discontinuum], expressed by the same sequence of zeroes and twos in a ternary expansion, it is seen that there is a 1-1 correspondence between points of the set and our population of admissible functions; moreover, the distance between two functions is now the distance between the two corresponding points of the line segment. Now Hausdorff [4] has shown that a measure can be defined over the Cantor set or over any similar set obtained by deletion of a central interval. If each of the surviving pieces is, at each stage, a fraction p of the original, the dimension r of the set is given by $2p^r = 1$, whence the Cantor set is of dimension $\log 2/\log 3$; a trifling extension of Hausdorff's argument will show that when the deletion is not symmetric, the surviving pieces being of fractions p and q at each deletion, the dimension r is given by $p^r + q^r = 1$. What concerns us here is the existence of the outer measure, by means of which we may define our integral of $c\exp(-\phi/2)dV$, where ϕ is the quadratic form defined by means of the kernel $K(s,t)$ of (15.9), and dV is the Hausdorff measure on the Cantor set, extended from the line segment (0, 1) to the entire line $-\infty, +\infty$ by simple translation, along with the coefficients and functions of the space. This shows the possibility of generalizing the normal distribution beyond the needs of the statistician. Let it be noted that in choosing samples of functions from such a space, the gaps might pass unnoticed, because the set of points is perfect though nowhere dense, so that arbitrarily close to every function there could be other functions of the population. In this connection, we might also note the fundamental role of measure and distance, as contrasted to mere one-to-one correspondence. Every point on the line segment (0, 1) may be expressed by means of the two digits 0, 1 in the binary decimal scale; replacing the 1 of the binary by the 2 of the ternary set, we get a one-to-one correspondence, excepting for the points which, in the Cantor discontinuum, have an infinite sequence of 2's. As these doubly represented points are all rational in the binary scale, their totality forms a set of measure zero. But on the continuous line segment (0, 1), extended by translation, our normal statistics can be defined as usual. This is of particular interest in considering such cases as the Kollektiv concept of von Mises, where we usually start by setting up a 1-1 correspondence between the throws of a coin and the binary expansion on (0, 1), which determines the measure a priori without further justification.

It gives me great pleasure to thank Mr. S.K. Sane for his careful execution of the two figures.

References

1. D.D. Kosambi, Curr. Sci. **11**, 271–274 (1942).
2. R. Courant, D. Hilbert, *Methoden d. Mathematischen Physik*, vol. 1, 2nd edn. (Berlin, 1931).
3. A. Kolmogoroff, Math. Ann. **99**, 209–219 (1928).
4. F. Hausdorff, Math. Ann. **79**, 157–179 (1919).

Chapter 16
The Estimation of Map Distances from Recombination Values

D.D. Kosambi, Fergusson College, Poona

The genetic map is a tool to quantify the distance between genes on a chromosome, based on the observed frequency of crossovers during cell division. These have been in use for over a century, from before the structure of DNA or of a gene was known. Kosambi's sole and lasting contribution to genetics built upon the work of J.B.S. Haldane (another polymath whom DDK greatly admired) and gave a formula that improved the estimates provided by the Haldane mapping function, correcting for the interference between genes. These formulas continue to be used in these days.

The paper, written in a charmingly simple and colloquial style, remains fresh and carries the characteristic Kosambi acidity—e.g., "The similarity of this with the velocity-addition formula in the special theory of relativity should not be made the basis of more bad philosophy." *Although Haldane later relocated to India, there is not much information about any further interactions that he and DDK might have had on this problem.*

Suppose three consecutive loci a, b, c of the same linkage group to have the recombination fractions (percentage divided by 100) $(a, b) = y_1$, $(b, c) = y_2$, $(a, c) = y_{12}$. Then it is known that for small values of y_1 and y_2, $y_{12} = y_1 + y_2$ approximately. For slightly larger values, we have a better approximation given by $y_{12} = y_1 + y_2 - y_1 y_2$; for still larger values, the approximation has again to be replaced by $y_{12} = y_1 + y_2 - 2y_1 y_2$. It is desired to obtain one single formula that will cover the entire range 0–$\frac{1}{2}$ of y-values in a reasonably satisfactory manner. This must also correspond to a single-valued, monotonically increasing, continuous function x of y in such a way that the corresponding identity becomes $x_{12} = x_1 + x_2$. The variable x will then be called the map distance corresponding to the given y.

Taking $y = f(x)$, our functional relation, assumed to be independent of the position on and number of the chromosome, must be of the form:

Published in Annals of Eugenics **12**, 172–175 (1944). Reprinted with permission from John Wiley & Sons, Inc.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_16

$$f(x+h) = f(x) + f(h) - pf(x)f(h). \tag{16.1}$$

The evidence that led to the conclusions of the first paragraph indicates that $f(x)/x \to 1$ as $x \to 0$. Also, that the unspecified function p increases from 0 to 2 with increasing x. Transposition and division by h gives

$$\frac{f(x+h) - f(x)}{h} = \frac{f(h)}{h} - pf(x)\frac{f(h)}{h}\,. \tag{16.2}$$

Taking limits as $h \to 0$, and assuming $f(x)$ to possess a derivative, we have

$$f'(x) = 1 - pf(x)\,; \quad \text{or} \quad \frac{dy}{dx} = 1 - py\,. \tag{16.3}$$

So far, we have followed the arguments and derivation of Haldane [4], who then fits an empirical curve from observed data for the X-chromosome, to obtain

$$x = 0.7y - 0.15\log_e(1-2y)\,. \tag{16.4}$$

This fits the observed data reasonably well and seems to fit other data also, to a considerable extent. But this amounts to abandoning (16.3) or taking $p = 0.6/(1-1.4y)$, which does not agree with our hypotheses. At best, (16.4) would indicate the existence of a general formula of the type desired. It is seen that formula (16.4) cannot conveniently be inverted, the usual method of use being by means of a table calculated by Haldane at intervals of 0.01 for those ranges of values of y, where the deviation from Morgan's first formula $y_{12} = y_1 + y_2$ becomes serious. The method would be, then, to find the values of x for given y (by interpolation if necessary), add, and then change back by using the table again.

It seems, however, possible to take one further step directly from the differential equation (16.3), by making a very plausible hypothesis about the unknown function p. This depends in some way on x and must increase steadily so far as known. The simplest such function would be one linear in x and y, and the simplest linear function taking the values 0 and 2 at the two ends of the range is, obviously, $4y$ in view of the fact that no recombination value can exceed 50 %. We thus obtain

$$\frac{dy}{dx} = 1 - 4y^2\,. \tag{16.5}$$

This integrates at once to the very simple solution:

$$2y = \tanh 2x\,; \quad x = \frac{1}{4}\log\frac{1+2y}{1-2y}\,. \tag{16.6}$$

The tables to use are, therefore, those of Fisher and Yates [3] for the transformation of the correlation coefficient, with $2y = r$, $2x = z$. The chief advantage of formula (16.6) is that we obtain a direct combination value

$$y_{12} = \frac{y_1 + y_2}{1 + 4y_1 y_2}\,. \tag{16.7}$$

The similarity of this with the velocity-addition formula in the special theory of relativity should not be made the basis of more bad philosophy.

Formula (16.7) eliminates the use of tables and correction curves. In the examples to be found in our textbooks, and in such other cases for which I have been able to obtain reasonably good data, the formula works at least as well as Haldane's. The use of tables would be necessary in comparing the lengths of two chromosomes, in accurate determination of the position in terms of x of the spindle-fiber attachment, and so forth. A comprehensive recasting of available data on map distances is not possible at present, because I have no access to the necessary bibliographic material and also because a good deal of the data seems to have been estimated by statistically unsatisfactory methods.

For example, the data quoted by Haldane give

$$\begin{aligned}\text{yellow-vermilion-rudimentary} &= 0.345 - 0.241 - 0.429,\\ \text{yellow-vermilion-bar} &= 0.345 - 0.239 - 0.479,\end{aligned}$$

which would indicate that the sum of a given distance to a fixed distance is more when the distance is shorter, contradicting our hypotheses. Similarly, for

$$\begin{aligned}\text{yellow-sable-rudimentary} &= 0.429 - 0.143 - 0.429,\\ \text{yellow-sable-bar} &= 0.429 - 0.138 - 0.479.\end{aligned}$$

These figures are also connected with such questions as the analysis of interference. Bridges and Morgan [1, p. 6] give the recombination percentage between sepia and Minute f as 52.4, which is impossible. The same authors give

$$\text{lethal-iiih-Dichaete-Hairless = 0.177-0.234-0.489}\,,$$

so that $y_{12} > y_1 + y_2$ which can only be explained by discarding the formulae or by emphasizing the paucity of the data and difficulty of locating lethals. Finally, we are given [1, p. 4] Dichaete-spineless recombination fraction as 0.137 from 3030 primary and as 0.153 from 9143 secondary observations, both sets being supposed [1, p. 21] "on an equal footing…in calculating recombination percents." If we restore the original recombination numbers from the given percentages, a rapid calculation gives $\chi^2 = 4.62$ (without Yates's correction) which is significant at the 5 % level for a single degree of freedom, making it unlikely that the two sets of figures locate the same point. This is not surprising, as salivary maps by Bridges and others seem to show, if I am not mistaken, that certain loci refer to whole sections of the chromosome. Under these circumstances, the use of any formula is naturally limited.

A few comments may nevertheless be useful. Various modifications of our most useful hypothesis, $p = 4y$, may be made if required by the evidence. All functions $p = a(x) + yb(x)$ lead to Riccati equations, which may be integrated without much trouble. Another possibility is that of restricting the passage to the limit from Eq. (16.2), obtaining a difference instead of a differential equation. But with $p = 4y$, it will be seen that the leading term in the solution will be of the same type as for the differential equation. One possible use of the last modification would be the derivation of formulae that retain their validity when there is a known minimum length of the chromosome that acts as a crossover quantum. For the present, there seems to be no evidence that would require any definite change of the formulae derived in (16.6) and (16.7).

If y, the recombination value, is to be treated as a probability, the methods of Fisher [2, Chap. XI] show the amount of information about the distance x in a sample of n observations to be

$$I_x = \frac{n(1 - 4y^2)^2}{y(1 - y)} \, . \tag{16.8}$$

It is this, and not n itself, that should be used as a weight in estimating the same x from parallel observations. *Relative to y*, the maximum information about x is obtained at $y = 0.25$, so that a new locus should be estimated from others which give about a quarter of the total number as recombinations. The point of maximum efficiency relative to x is a little farther to the right, so that slightly greater recombination values would do. The problem of efficient estimation of recombination values has been treated directly by Fisher [2, pp. 235–252].

Suppose that between a new gene and a known one n observations give m recombinations; for a second gene, n' and m', between the two reference loci M recombinations occur in N cases. By least squares the "best" values of the recombination fractions y_1, y_2 between the two markers and the gene to be located would be those minimizing

$$w_1\left(y_1 - \frac{m}{n}\right)^2 + w_2\left(y_2 - \frac{m'}{n'}\right)^2 + w_{12}\left(y_{12} - \frac{M}{N}\right)^2 , \tag{16.9}$$

where y_{12} has the value given by (16.7). Here y has to be used in place of x because the distribution is much nearer to normal. For the weights most experimenters would choose $w_i = n_i$, the number of observations, though the proper value would be the amounts of information: $w_i = n_i/y_i(1 - y_i)$, which make (16.9) yield the value of χ^2. For the more efficient maximum-likelihood estimates, we should equate to zero the first partial derivatives of $S\{m_i \log y_i + (n_i - m_i)\log(1 - y_i)\}$, always taking y_{12} as in (16.7).

To illustrate, we simplify still further by taking the recombination value between markers as precisely known (M, N very large). Then $y_{12} = a$, $y_1 = y$, $y_2 = (a - y)/(1 - 4ay)$, and we have

$$\chi^2 = \frac{(ny - m)^2}{ny(1 - y)} + \frac{\{n'(a - y) - m'(1 - 4ay)\}^2}{n'(a - y)\{(1 - a) + y(1 - 4a)\}}\,, \tag{16.10}$$

and

$$\begin{aligned}\log L = \text{const.} &+ m \log y + (n - m) \log(1 - y) - n' \log(1 - 4ay)\\ &+ m' \log(a - y) + (n' - m') \log\{(1 - a) + y(1 - 4a)\}\,.\end{aligned} \tag{16.11}$$

The minimum χ^2 gives an equation of the sixth degree, whereas equating L' to zero gives a quartic:

$$ny^4 - b_1 y^3 + b_2 y^2 - b_3 y + b_4 = 0\,, \tag{16.12}$$

where

$$\begin{aligned} b_1 &= n(a + r + s) + m + n'(r - s) + m'(s - a),\\ b_2 &= n(ar + as + rs) + m(a + r + s) + n'(r - s)(1 + a) + m'(s - a)(1 + r),\\ b_3 &= n(ars) + m(ar + as + rs) + n'a(r - s) + m'r(s - a),\\ b_4 &= m(ars); \quad \text{with} \quad r = \frac{1}{4}a \quad \text{and} \quad s = (1 - a)/(4a - 1). \end{aligned}$$

de Winton and Haldane [5, p. 75] give for *Primula*-II♀ the y-values $PF - FCh - PCh$ corrected as 15.10 – 10.35 – 23.92, whereas our formula (16.7) should give 23.95 for the last, a very good fit; the uncorrected (backcross crossovers) values are 14.52 – 10.83 – 23.10, where the last should have been 23.85 for consistence by (16.7). If we took 23.92 as the fixed value and worked only with the backcross data, we should have $a = 0.2392$, $n = 1253$, $m = 182$, $n' = 2613$, $m' = 283$, which gives

$$y^4 - 18.77314y^3 + 15.2263y^2 + 2.5593y - 0.63951 = 0\,,$$

the root between 0 and 0.5 being 0.146 to the nearest three figures, which is hardly an improvement worth the trouble, the present case being merely an illustration. The standard error is immediately calculated, as usual, by taking the reciprocal of $-L''$ as the variance when the value of y from (16.12) is substituted. For the more general formulae with several y-values determined simultaneously, the best method is to substitute the observed values in the maximum-likelihood equations and proceed by successive approximations.

de Winton and Haldane [5, pp. 96–97] extend our postulates by considering the nature of the coincidence. This amounts to the extra condition that $p(x, y)/2y \to$ const. as $y \to 0$. Taking $p = 8y/(1 + 2y)$, Haldane integrates the differential equation to get

$$12x = \log(1 + 4y) - 4\log(1 - 2y)\,.$$

But $p = 2y/(1 - y)$ also satisfies all conditions to give $6x = 4\log(1 + y) - \log(1 - 2y)$, which gives somewhat better consistency in the values of x, here the sole criterion, as there is no intrinsic unit of map distance. That formula is the most suitable where the distances are additive to within the limits of significance. The data from *Primula*-I may be used for the purposes of comparison [5, p. 98]:

	SB	SG	SL	BG	BL	GL
♀ y	6.25	34.40	37.53	32.14	36.73	3.61
$\bar{x}_0$	6.26	42.20	48.72	38.15	46.93	3.61
$\bar{x}_1$	6.33	46.06	54.03	41.21	51.77	3.61
$\bar{x}_2$	6.27	39.12	44.39	35.72	42.97	3.61
♂ y	11.55	41.03	41.45	35.01	38.86	1.82
$\bar{x}_0$	11.76	57.94	59.24	43.38	47.22	1.82
$\bar{x}_1$	11.92	65.37	67.07	47.45	52.11	1.82
$\bar{x}_2$	11.92	65.37	67.07	47.45	52.11	1.82

where $\bar{x}_0$ is from formula (16.6) of this note, $\bar{x}_1$ is Haldane's revised formula above, and x_2 ours. Others could be devised very easily, as for example by taking the simple value $p(x, y) = 2y + 4y^2$, which satisfies all the conditions to give the very clumsy result

$$10x = 4\tan^{-1}(1 + 2y) + \log(2y^2 + 2y + 1) - 2\log(1 - 2y) - \pi\,;$$

this gives more trouble in the calculation with actually less consistence in the fit for map distances. Besides being less trouble to calculate, the inverse hyperbolic tangent formula has the tremendous advantage of the handy composition rule (16.7), which also allows use in actually fitting crossover values and map distances from the observational data.

References

1. C.B. Bridges, T.H. Morgan, *The Third-Chromosome Group of Mutant Characters of Drosophila melanogaster*, Publication no. 327 (Carnegie Institute of Washington, 1923).
2. R.A. Fisher, *The Design of Experiments* (Oliver and Boyd, Edinburgh, 1937).
3. R.A. Fisher, F. Yates, *Statistical Tables, Table VII* (Oliver and Boyd, Edinburgh, 1938).
4. J.B.S. Haldane, J. Genet. **8**, 299–309 (1919).
5. D. de Winton, J.B.S. Haldane, J. Genet. **31**, 67–100 (1935).

Chapter 17
The Geometric Method in Mathematical Statistics

D.D. Kosambi, Fergusson College, Poona

This paper, where DDK used n-dimensional geometry to derive all the elementary distributions, was among the last that he wrote while he was still at Fergusson College, prior to moving to TIFR. This was also one of the papers written during his anni mirabili*, the war years.*

17.1 Introduction

R.A. Fisher was the first to make use of n-dimensional geometry in the derivation of certain distributions. A direct approach is always possible and is even to be preferred, by the use of the Fourier transform, with or without the transformation theory of positive-definite quadratic forms. Nevertheless, the geometric method offers great advantages in brevity, clarity, and insight. It is in no way inferior in rigor to any other method, and finally, its applications extend to a greater number of the distributions used in small-sample theory than is realized.

17.2 Geometric Preliminaries

In an Euclidean space of $n \geq 1$ dimensions with coordinates $(x_1, x_2, \ldots, x_n)$ for a generic point x, the distance d between two points x and x' is given by

$$d^2 = \sum (x_i - x_i')^2 . \tag{17.1}$$

Published in American Mathematical Monthly **51**, 382–389 (1944). Reprinted with permission from the Mathematical Association of America.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_17

In all summations, unless otherwise specified, the range is over values 1 to n of the index. This definition of distance, combined with the usual methods in use for three-dimensional Euclidean geometry, suffices to derive most of the formulas we need.

The vector $x' - x$ has direction components $a_i = x'_i - x_i$, $i = 1, 2, \ldots, n$, so that the square of its length is $d^2 = \sum a_i^2$ and the direction cosines are $\alpha_i = a_i/d$, with $\sum \alpha_i^2 = 1$. The angle θ between two directions is given by $\cos\theta = \sum \alpha_i \alpha'_i = \sum a_i a'_i / dd'$.

A hyperplane in n-space is represented by a linear equation

$$\sum a_i x_i = y\,. \tag{17.2}$$

The coefficients a_i are the direction components of the normal to the plane. The perpendicular distance from a point x to the plane is $p = \left(\sum a_i x_i - y\right)/\sqrt{\sum a_i^2}$, where the proper sign is to be taken in the square root so as to make this distance positive. When the direction cosines α_i are used in place of a_i in Eq. (17.2), y is itself the perpendicular distance from the origin to the plane. The foot of the perpendicular to the plane from the point x is $\bar{x}$, which is given by

$$\bar{x}_i = x_i - \lambda a_i \ , \quad \text{where} \quad \lambda = \left(\sum a_i x_i - y\right) \Big/ \sum a_i^2\,. \tag{17.3}$$

The hypersphere of radius r centered at the origin is

$$\sum x_i^2 = r^2 \quad \text{or} \quad \sum x_i^2 \le r^2 \tag{17.4}$$

of which the first represents the surface, and the second the volume of the sphere. A plane section of the n-sphere is always an $n - 1$ sphere where we understand by the one-dimensional sphere, the line interval $(-r, r)$, with “surface” $x = \pm r$, the two-dimensional sphere being the circle, and so on. The volume of the n-sphere is $2\pi^{n/2} r^n / n\Gamma(n/2)$. It suffices for our purpose that this volume should be cr^n, which is equivalent to the statement that the volume element for purposes of integration is

$$dV_n = dx_1 dx_2 \cdots dx_n = r^{n-1} f(\theta_1 \theta_2 \cdots \theta_{n-1}) dr d\theta_1 d\theta_2 \cdots d\theta_{n-1}\,. \tag{17.5}$$

The actual form of f for any given n may be derived by simple extension of the three-dimensional spherical polar coordinate formula.

17.3 The Normal Distribution and Euclidean *n*-Space

The most important basic distribution [important because of the "central limit theorem" which shows that it is the limiting form of the distribution of an average from any fairly general type of population] is the normal distribution, where the elementary probability is given by $(1/\sigma\sqrt{2\pi})\exp -(x-\mu)^2/2\sigma^2 dx$. By change of origin and scale, we can always take the population mean μ as zero and the population variance σ^2 as unity; this we shall call standard measure.

For several normally distributed standard independent variables $x_1, \ldots, x_n$, the elementary probabilities are compounded by multiplication to give

$$\begin{aligned} dP =& \frac{1}{\sqrt{2\pi}}\, e^{-x_1^2/2} dx_1 \,\frac{1}{\sqrt{2\pi}}\, e^{-x_2^2/2} dx_2 \cdots \frac{1}{\sqrt{2\pi}}\, e^{-x_n^2/2} dx_n\,, \\ =& \frac{1}{(\sqrt{2\pi})^n}\, e^{-r_n^2/2} dV_n\,. \end{aligned} \tag{17.6}$$

The basis of the geometrical method as applied to mathematical statistics is that this probability density depends only upon r_n^2; i.e., it is isotropic in the n-space. Moreover, this is a characteristic property of the normal distribution. If the elementary probability were $f(x)dx$, and if we should ask ourselves when a relation of type

$$f(x_1)f(x_2)\cdots f(x_n)dx_1dx_2\cdots dx_n = \phi(r_n)dV_n\,, \tag{17.7}$$

would be valid, we should obtain the functional equation

$$f(x_1)f(x_2)\cdots f(x_n) = \phi\left(\sqrt{\sum x_i^2}\right), \tag{17.8}$$

to hold identically in x. Setting $x_2 = \cdots = x_n = 0$, we get $\phi(x_1) = f(0)^{n-1}f(x_1)$. The functional equation is thus equivalent to $f(x)f(y) = cf(\sqrt{x^2+y^2})$, which is known to have no continuous solutions except $f(x) = ae^{bx^2}$. This sort of argument is followed, for example, in the deduction of Maxwell's law in the kinetic theory of gases. If, now, the total range for each x_i is $-\infty, +\infty$, we have $b < 0$, say $b = -1/2\sigma^2$, to make the total probability unity, $a = 1/\sigma\sqrt{2\pi}$. If the range be restricted, other distributions, including the uniform distribution, are possible—a fact that is forgotten by all who use this derivation for theoretical physics.

Not only may we build up the normal distribution in n dimensions from n independent individual distributions, but the process may also be reversed so that one or more dimensions can be cut off as necessary. *Orthogonality is the geometric equivalent of independence in statistics*. We utilize this after performing linear transformations to new sets of orthogonal axes; for our purpose, dealing only with normal distributions in standard measure, rotations alone suffice. The effective number of dimensions in a given problem is called degrees of freedom by R.A. Fisher, who uses the letter n to denote this number. We shall use n for the original number of dimensions, keeping in mind that a random sample of n from an infinite basic population represents such an n-dimensional distribution as in (17.6). The degrees of freedom not in use are to be integrated out. Contrary to dynamical usage, the effective number of degrees of freedom in a statistical problem represents that number of dimensions in which the coordinates are free under the statistical conditions imposed only in the limited sense that a definite probability necessarily attaches to each region of the subspace.

17.4 The Distribution of the Mean and Variance in Samples From normal. The χ^2-Distribution

Let $y = \sum a_i x_i$ be a linear combination of n standard normal variables $x_1, \dots, x_n$. To derive the distribution of y, we rotate so that one axis lies along the normal to the family of hyperplanes $y = \sum a_i x_i$, or with respect to the new axes, $y =$ const., and the remaining $n-1$ lie in the plane $\sum a_i x_i = 0$. Taking $r = y/\sqrt{\sum a_i^2}$, we may split up the distance from the origin to a point x as $r_n^2 = r_1^2 + r_{n-1}^2$. That is, r_1 is the distance from the origin to the foot of the perpendicular to the particular hyperplane of the family $\sum a_i x_i = y$ which passes through the point $x = (x_1, \dots, x_n)$; r_{n-1} is the distance in the plane from the foot of the perpendicular to the point x. The elementary probability of (17.6) may, therefore, be expressed as

$$dP = \frac{1}{(\sqrt{2\pi})^n} e^{-r_n^2/2} dV_n = \frac{1}{\sqrt{2\pi}} e^{-r_1^2/2} dr_1 \cdot \frac{1}{(\sqrt{2\pi})^{n-1}} e^{-r_{n-1}^2/2} dV_{n-1} \,. \tag{17.9}$$

As the r_1 component alone interests us here, we may eliminate the rest by integration over the whole of V_{n-1}. Therefore, r_1 is normally distributed in standard measure. Hence, $y = r_1\sqrt{\sum a_i^2}$ is normal with zero mean and variance $\sum a_i^2$. Replacing $x_i + \mu_i$ by x_i, to pass from standard measure to the general normal distribution, we obtain:

Theorem 1 *If n independent variates $x_1, \dots, x_n$ are normally distributed with means μ_i and variances σ_i^2, any linear combination thereof $\sum a_i x_i$ is also normally distributed with mean $\sum a_i \mu_i$ and variance $\sum a_i \sigma_i^2$.*

On the other hand, we might have integrated out r_1 to concentrate upon r_{n-1}, noting that the resulting distribution in V_{n-1} would be normal, the original n variables being now restricted to lie in a hyperplane, i.e., to obey one identical linear restriction of type $\sum a_i x_i$ = const. As r_{n-1} is to be measured from the foot of the perpendicular from the origin of n-space, formula (17.9) may be applied to give

Theorem 2 *If upon n originally independent standard normal variables a linear restriction $\sum a_i x_i = b$ is imposed, we obtain a normal distribution in $n-1$ variables: $dP = c \exp -r_{n-1}^2/2 \cdot dV_{n-1}$, where*

$$r_{n-1}^2 = \sum (x_i - \bar{x}_i)^2 \ , \ \ \bar{x}_i = ba_i \Big/ \sum a_i^2 \,. \tag{17.10}$$

If we abandon the standard measure, putting $x_i - \mu_i$ for x_i, we should have to replace b by $b + \sum a_i \mu_i$. Thus, $\bar{x}_i$ shifts correspondingly except in the special case where all these separate additions cancel out in each bracket $(x_i - \bar{x}_i)$. To this end, it is necessary and sufficient that

$$\mu_i = \frac{a_i \sum a_i}{\sum a_i^2}\,. \tag{17.11}$$

Since in sampling problems $\mu_i = \mu$, $i = 1, 2, \ldots, n$, we have

Theorem 3 *For random sampling from a normal population, the sole linear restriction independent of the population mean is of type* $(x_1 + \cdots + x_n)/n$ *= constant.*

That is, in general, the population mean being unknown, we may measure from the sample mean $m = \sum x_i/n$ with the loss of a single degree of freedom. Moreover, there is no other linear sample function from which such measurement may be made independent of the population mean. The statistic m has the further advantage that it is normally distributed, according to Theorem 1, with expectation equal to the population mean and variance σ^2/n. Thus, it is *unbiased*, and using results and terminology due to Fisher, since no other estimate of μ can have a smaller variance than σ^2/n, m is the most *efficient* such estimate; finally, as $n \to \infty$, the probability for $m \neq \mu$ tends to zero, so that the estimate is consistent.

For the χ^2 and other tests, we need the distribution of the sum of squares of n normal standard variables, our r_n^2. That is,

$$P(r_n^2 \leq t) = \int_{\sum x_i^2 \leq t} \cdots \int e^{-(x_1^2 + \cdots + x_n^2)/2} dx_1 \cdots dx_n\,. \tag{17.12}$$

This is evaluated at once in spherical polar coordinates, integrating over thin spherical shells centered at the origin. This gives

$$P(r_n^2 \leq t) = \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)(\sqrt{2\pi})^n} \int_0^{\sqrt{t}} e^{-r^2/2} r^{n-1} dr\,. \tag{17.13}$$

Transforming by $u = r^2$, $r = \sqrt{u}$, $dr = du/2\sqrt{u}$, we get

$$P(r_n^2 \leq t) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \int_0^{t} e^{-u/2} u^{n/2-1} du\,. \tag{17.14}$$

Provided the population means is zero, this χ^2 or incomplete gamma function distribution holds even for a change of scale, in particular for $\sum x_i^2/\sigma^2$. The expectation of $u = r_n^2$ is easily calculated by multiplying under the sign of integration by u and integrating to infinity. This amounts to replacing the exponent n by $n+2$, so that the result must be $2 \cdot 2^{n/2}\Gamma(n/2+1)/2^{n/2}\Gamma(n/2) = n$. Therefore, r_n^2/n furnishes, when the population mean is known, an unbiased estimate of σ^2. For unknown population mean, we apply the findings of Theorem 3 to obtain.

Theorem 4 *For random sampling from the same normal population, $\sum(x-m)^2$ with $m = \sum x_i/n$ has the χ^2 distribution with $n-1$ degrees of freedom, its expectation being $(n-1)$.*

To estimate the population variance without bias, we must divide $\sum(x-m)^2$ by $n-1$, and not by n.

17.5 The Student–Fisher *t*-Distribution

The next step is to derive statistics independent of the population variance. The first of these is "Student's" $t = m\sqrt{n}/s = \sqrt{n}(m/\sigma)/(s/\sigma)$, which has the requisite property. We known that m is normally distributed with standard deviation $\sigma/\sqrt{n}$ and that $s^2 = \sum(x-m)^2/(n-1)$ is the best estimate of σ^2 in samples of n. For large samples, a consistent t tends therefore to normality with unit variance; its chief importance, however, lies in its use for small samples.

First, in standard variables, $m\sqrt{n}$ is the distance from the origin to the plane $\sum x_i = mn$. Moreover, $r_{n-1} = s\sqrt{n-1}$ is the distance within the hyperplane from the foot of the perpendicular, as before. Hence, the ratio $m\sqrt{n}/r_{n-1} = t/\sqrt{n-1}$ is the cotangent of the angle made by the radius vector from the origin to a point x of n-space with the normal to a family of parallel hyperplanes. Taking this normal as one axis, the distance from the origin may be integrated out over thin conical shells of angle θ with this direction. In one dimension, two lines can only make the angle 0 or π. In two dimensions, the thin sector has the "volume" element for the "cone" $2rd\theta dr$. In three dimensions, we apply the theorems of Pappus to get the volume element $2\pi r^2 \sin\theta dr d\theta$ over the conical shell, by simple extension to n dimensions, $(n-1)(n-2)cr^{n-1}\sin^{n-2}\theta dr d\theta$. Integrating out the r, the elementary probability is seen to be $c\sin^{n-2}\theta d\theta$, where c is hereafter treated throughout as a generic constant to be determined at the last step by equating the total probability to unity. Putting $u = \cot\theta$, dP transforms to $c(1+u^2)^{-n-2/2}(1+u^2)^{-1}du$. Finally, $u = t/\sqrt{n-1}$, which gives

Theorem 5 *The probability of $m\sqrt{n(n-1)}/\sum[x_i - m]^2 \le t$, where $m = \sum x_i/n$ and the x_i are independent normal variables with zero population mean and an identical variance is*

$$P = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{n-1}\,\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} \int_{-\infty}^{t} \left(1 + \frac{t^2}{n-1}\right)^{-n/2} dt\,. \tag{17.15}$$

Suppose we have two independent random samples of n_1, n_2 members respectively from the same standard normal population. Since $m_1\sqrt{n_1}$, $m_2\sqrt{n_2}$ are distances, independent and normally distributed, Theorem 1 applies to prove that $m_1 - m_2$ is also normal with variance $1/n_1 + 1/n_2$; hence, $(m_1 - m_2)/\sqrt{(1/n_1 + 1/n_2)}$ is a normally distributed variable in standard measure. From this, we get a cotangent provided we can divide by some r_k. This is best done by taking $r_k^2 = r_{n_1-1}^2 + r_{n_2-1}^2$, i.e., by combining the distance in the remaining $n_1 + n_2 - 2$ dimensions, orthogonal to both m_1 and m_2. This gives

Theorem 6 *For two independent samples from the same normal population*

$$\frac{m_1 - m_2}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{m_1 - m_2}{\sqrt{(n_1-1)s_1^2 + (n_2-1)s_1^2}} \times \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

has the t distribution with $n_1 + n_2 - 2$ degrees of freedom. If the mean from the first and the variance from the second are used, $t = m_1\sqrt{n_1}/s_2$ with $n_2 - 1$ degrees of freedom.

In the first portion, standard measure is unnecessary in view of the fact that the expression for t is independent of both μ and σ. The second part follows immediately upon consideration of the fact that the degrees of freedom were associated only with the estimate of variance, for $m\sqrt{n}$ represents just one fixed direction, the movement over the "surface" of the cone being associated with $r_{n-1}^2 = (n-1)s^2$ which gives the degrees of freedom in t.

17.6 Distribution of the Variance Ratio

A most important property of r_n^2 is that it may be broken up into components of the same type. Each component divided by its degrees of freedom gives an estimate of σ^2. The sum of squares is the only analytic nonnegative function of the coordinates that may thus be resolved into additive components without the change of form. This simple algebraic fact is the basis, for example, of dynamical theorems like those of Bertrand and Kelvin on the energy of a system after a certain number of constraints. In statistics, the use made of this resolution into components is called analysis of variance and consists of testing the above independent estimates of σ^2 for compatibility. From the preceding sections, it is clear that the ratio of any two such estimates is independent of both the population parameters μ and σ^2.

Assuming therefore a normal standard population without loss of generality, $s_1^2/s_2^2 = (n_2 - 1)r_{n_1-1}^2/(n_1 - 1)r_{n_2-1}^2$. Now, r_{n_1-1}/r_{n_2-1} is again the cotangent (or tangent) of an angle with this difference: In the t distribution, the numerator was confined to one fixed direction by taking $n_1 = 1$, whereas it is now allowed to move freely through $n_1 - 1$ dimensions independently of the denominator. Accordingly, the n_2 dimensions of the denominator contribute $c'' \sin^{n_2-2}\theta d\theta$ as before, which must be multiplied by $c' \cos^{n_1-2}\theta d\theta$ for the swing of r_{n_1-1} (at right angles). The total elementary probability is, therefore, $c\ \sin^{n_2-2}\theta \cos^{n_1-2}\theta d\theta$. With $u = \cot\ \theta$, we get

Theorem 7 *In two independent random samples of size n_1, n_2 from the same normal populations, the distribution of $F = s_1^2/s_2^2$ is given*

$$P(F \leq t) = \frac{\Gamma\left(\frac{n_1+n_2-2}{2}\right)(n_1-1)^{(n_1-1)/2}(n_2-1)^{(n_2-1)/2}}{\Gamma\left(\frac{n_1-1}{2}\right)\Gamma\left(\frac{n_2-1}{2}\right)} \cdot \int_0^t \frac{F^{(n_1-3)/2}dF}{[(n_2-1)+(n_1-1)F]^{(n_1+n_2-2)/2}}\,. \tag{17.16}$$

The distribution of $z = \log\sqrt{F}$ has special advantages over that of F in the way of asymptotic formulae, and is therefore commonly used for accurate work. In practice, s_1^2, s_2^2 are so labeled as to give $F \geqq 1$, $z \geqq 0$.

17.7 Distribution of the Coefficient of Correlation in Samples from an Uncorrelated Normal Universe

The product-moment correlation (briefly the correlation) coefficient is estimated from a random sample of n pairs $(x_1, y_1)\cdots(x_n, y_n)$ by

$$r = \sum(x_i - \bar{x})(y_i - \bar{y})\Big/\sqrt{\sum(x_i - \bar{x})^2\sum(y_i - \bar{y})^2}\,, \quad \begin{aligned}\bar{x} &= \sum x_i/n\\ \bar{y} &= \sum y_i/n.\end{aligned} \tag{17.17}$$

The basic assumption is that the coordinate pairs are sampled from an uncorrelated bivariate normal population, the elementary probability being $2\pi \exp -(x^2 - y^2)/2$, using standard measure without loss of generality because r as defined above does not depend upon the population parameters.

In this case, we superpose the y-space upon the x-space by rotation, much as two pictures on transparent film may be superposed upon a screen. In this, the x_i and the y_i axis are to coincide for each i, which does not imply any relationship between the generic points x and y. Then, clearly, $r = \cos\theta$, θ being the angle between the two (independent) directions from the points $\bar{x}$, $\bar{y}$ to the points x, y, respectively. The family of hyperplanes $\bar{x}$-const. coincides with the family $\bar{y}$-const. so that we can rotate both spaces simultaneously without destroying the original correspondence and have only to investigate the distribution of $\cos\theta$ in the superposed $n - 1$-space within the hyperplanes.

Here, the x-radius vector traverses the whole space without restriction and hence may be integrated out, and the y-vector has only the restriction of making the angle

θ with the first, so that its radial distance may be integrated out also over thin conical shells as in the two preceding distributions. This gives the total elementary probability $c\sin^{n-3}\theta d\theta$, the exponent being 2 less than the dimensions $n-1$ of the free y-space (under the restriction $\bar{y}$ = const.). This gives for the distribution of r

$$P(r \leq R) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}\right)} \int_{-1}^{R} (1-r^2)^{(n-4)/2} dr\,. \tag{17.18}$$

Our method of derivation shows the intimate relation between t and F, which is that $t = \sqrt{F}$ essentially when $n_1 = 2$. Moreover, the relation between r and t is also immediately evident, given that $r = \cos\theta$ with $t = \sqrt{(n-1)} \cdot \cot\theta$ for a sample of n. Seeing that for t we have one more dimension at the start than for r when generating the hypercone of angle θ, it follows that $t = r\sqrt{(n-2)}/\sqrt{(1-r^2)}$ has the Student distribution with $n-2$ degrees of freedom.

Chapter 18
The Law of Large Numbers

D.D. Kosambi, Tata Institute of Fundamental Research, Bombay

This is probably the first paper that DDK published with the TIFR byline. Although not much more than a pedagogical exercise, the paper was reviewed in Mathematical Reviews by W. Feller who noted that the paper provided "an elementary proof of the weak law of large numbers, stressing the rôle and the implications of the assumptions."

18.1 Introduction

Let $X_1, \ldots, X_n$ be a sequence of random (stochastic) variables with expectations $E(X_r) = m_r$. Then, LLN[1] states that under suitable restrictions upon the X_r, the difference $(X_1 + \cdots + X_n)/n - (m_1 + \cdots + m_n)/n$ may be made arbitrarily small in absolute value with probability arbitrarily close to unity by taking n sufficiently large (U, Chap. 10). The proof is based upon Tshebysheff's Lemma (U, p. 182): *If Xis a positive random variable, then*

$$P\{X \leq t^2 E(X)\} > 1 - \frac{1}{t^2}\,.$$

Thus, if $B_n = E(X_1 - m_1 + \cdots + X_n - m_n)^2$ exists, the lemma applied to the positive random variable $\sum_i^n (X_r - m_r)^2$ gives at once

[1] Abbreviations used are: LLN for the law of large numbers; c.f. for characteristic function; U for J.V. Uspensky, "Introduction to Mathematical Probability" (New York, 1937); Cr. for Harald Cramér, "Random Variables and Probability Distributions" (Cambridge Tract no. 36, Cambridge 1937).

Published in The Mathematics Student **14**, 14–19 (1946). Reprinted with permission.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_18

$$\left|\frac{X_1+\cdots+X_n}{n}-\frac{m_1+\cdots+m_n}{n}\right|\le\varepsilon \quad\text{with}\quad P_n>1-\frac{B_n}{n^2\varepsilon^2}\,.$$

So, for convergence in probability, it suffices that $B_n/n^2\to 0$. On the other hand, if for $r=1,2,\ldots,n$, $\max|X_r-m_r|\le C_n<\infty$, and LLN holds, so that $(1-P_n)\to 0$, we may use the easily derivable inequality: $B_n<n^2C_n^2(1-P_n)+n^2\varepsilon^2P_n$ to prove that $B_n/n^2\to 0$ follows as a necessary condition from LLN under the restriction $C_n^2(1-P_n)\to 0$; hence, in particular, when the variables X_r are uniformly bounded, $C_n\le C$ (U, pp. 185–6). This may be and has been extended in various directions, beginning with the Bienaymé-Tshebysheff inequality (Cr. pp. 21; 38–39). What interests us here is the analysis of the structure of the proof, using only text book methods as far as possible.

In what follows I consider LLN *only* in the particularly useful special case when the variable X_r are all independent.

18.2 The Law of Large Numbers

To cover general types of distributions, we assume at the outset that each X_r has a distribution function $F_r(x)$ which is positive, non-decreasing, with $F_r(-\infty)=0$, $F_r(+\infty)=1$ for all r and $P\{X_r\le x\}=F_r(x)$ for all values of the real variable x. The integral being taken in the sense of Lebesgue Stieltjes, we may assume the existence of $E(X_r)=\int x\,dF_r(x)$ over the entire real line $-\infty\le x\le\infty$. Otherwise LLN has to be given a special meaning for the occasion. We make the further assumption that $\int|x|dF_r(x)$ also exists for each value of the index; this is "reasonable" in that when the distribution is given only discrete values for the observable variable, we have an infinite series replacing the integral and in view of the fact that a preferential order is hardly reconcilable with the intuitive idea of randomness, the series in question would have to converge absolutely, i.e., independently of order, to the same value.

These fundamental assumptions are then also fulfilled for the independent stochastic variable $X_r-E(X_r)$ with which alone LLN is concerned, so that there is no further loss of generality in assuming $E(X_r)=0$ for all r. We are therefore dealing with a sequence of random independent variables $X_1,\ldots,X_n$ such that $\int x dF_r(x)=0$ for all r, and $\int|x|dF_r(x)$ exists. LLN holds for these if and only if $(X_1+\cdots+X_n)/n$ converges in probability to zero. For each n and N, we define non-negative functions of the two variables n and N (of which the first is only a positive integer, the second a continuous variable) as follows:

$$\int_{|x|>N} |x| dF_n(x) = h(n, N) \to 0 \quad \text{as} \quad N \to \infty \quad \text{for each } n \tag{18.1}$$

$$h_0 = \max\ h(r, N), \quad r = 1, 2, \ldots, n\,; \quad H = \sum_1^n h(r, N)\,.$$

$$\int_{|x|<N} |x| dF_n(x) = c(n, N)\,;$$

$$c_0 = \max\ c(r, N)\,,\ r = 1, 2, \ldots, n\,;\ C = \sum_1^n c(r, N)\,.$$

Following a standard device (U, p. 192), the variables X are each split up into two additive stochastic components as follows:

$$X_i = u_i + v_i\,;\ \text{if } |X_i| \leq N\,,\ \ u_i = X_i,\ \ u_i = 0\,; \quad \text{otherwise } u_i = 0,\ v_i = X_i\,. \tag{18.2}$$

We then define $b_i = E(u_i) = -E(v_i)$ and it follows that

$$\begin{aligned} |X_1 + \cdots + X_n| &\leq |u_1 + \cdots + u_n| + |v_1 + \cdots + v_n| \\ &\leq |u_1 - b_1| + \cdots + u_n - b_n| + |b_1 + \cdots + b_n| + |v_1 + \cdots + v_n|\,. \end{aligned}$$

By definition, it also follows that

$$|b_n| = |\int_{|x|>N} x\, dF_n(x)| \leq h(n, N)\,; \quad |b_1 + \cdots + b_n| \leq H(n, N) \tag{18.3a}$$

$$\begin{aligned} P\{v_n \neq 0\} = P\{|X_n| > N\} &= \int_{|x|>N} d\, F_n(x) \\ &= \frac{N}{N} \int_{|x|>N} d\, F_n(x) \leq \frac{1}{N} \int_{|x|>N} |x| d\, F_n(x)\,, \end{aligned} \tag{18.3b}$$

whence $P\{v_n \neq 0\} \leq h(n, N)/N$ and $\Sigma P\{v_r \neq 0\} \leq H(n, N)/N$.
From (18.2) and the above, we have

$$\begin{aligned} &P\left\{\frac{|X_1 + \cdots + X_n|}{n} \leq \varepsilon\right\} \\ &\quad > P\left\{\frac{|u_1 - b_1 + \cdots + u_n - b_n|}{n} \leq \varepsilon - \frac{H(n, N)}{n}\right\} - \frac{H(n, N)}{N}\,. \end{aligned} \tag{18.4}$$

The argument is that $|X_1 + \cdots + X_n| \leq A$ may hold in two mutually exclusive ways: when all the $v_r = 0$, or when at least one $v \neq 0$. The probability for the latter event is allowed for by subtracting the term H/N, which it can never exceed.

The ε in (18.4) may be chosen arbitrarily small, so that H/n must tend to zero with increasing n. For LLN to hold, H/N must also tend to zero with increasing N, for some method of having n, N both $\to \infty$. A third condition has to be satisfied, however. The stochastic variables $u_i - b_i$ are independent with zero expectation each and bounded so that their second moment exists; the second moment of their sum is the sum of their second moments. It is easily proved that, for any random variable, $E(X^2) \geq E[X - E(X)]^2$, so that $E[\sum(u_i - b_i)^2] = B_n \leq \sum E(u_i^2) \leq N \sum E(|u_i|)$, whence $B_n \leq NC$.

Applying LLN in its classical form to $u_i - b_i$, we obtain

$$P\left\{\frac{|u_i - b_i + \cdots + u_n - b_n|}{n} \leq \varepsilon - \frac{H}{n}\right\} > 1 - \frac{NC}{n^2(\varepsilon - H/n)^2}\,. \tag{18.5}$$

That is,

$$P\left\{\frac{|X_1 + \cdots + X_n|}{n} \leq \varepsilon\right\} > 1 - \frac{NC}{(n\varepsilon - H)^2} - \frac{H}{N}\,. \tag{18.6}$$

So, for LLN to hold, the third condition is that $NC/(n\varepsilon - H)^2 \to 0$. In this limit, the quantity $\varepsilon - H/n$ may be ignored, which gives our main result: LLN *holds, $(X_1 + \cdots + X_n)/n$ converging in probability to zero, when $E(X_r) = 0$ and $E(|X_r|)$ exists for all r, if for some manner of approach of n and N to infinity the conditions*

$$\frac{1}{n}\sum_{I}^{n}\int_{|x|>N} |x|dF_r(x) \to 0 \quad ; \quad \frac{1}{N}\sum_{I}^{n}\int_{|x|>N} |x|dF_r(x) \to 0\,, \tag{18.7}$$

$$\frac{N}{n^2}\sum_{I}^{n}\int_{|x|\leq N} |x|dF_r(x) \to 0$$

are all satisfied.

In particular, as a corollary, LLN *holds if* $h(n, N) < G(N) \to 0$. In this case, we need not attempt refinement by asking the condition to hold for all large n in view of the fact that $h(n, N) \to 0$ for each n as $N \to \infty$; so that if the condition holds for $n \geq k$, the larger of the functions $h_0(k, N)$ and $G(N)$ will do for all n. The proof of this corollary is simple: the condition leads at once to $H(n, N) < nG(N)$, $H/n < G(N) \to 0$; $H/N < n\sqrt{G} \cdot \sqrt{G}/N \to 0$ also, if we take $n = N/\sqrt{G} \to \infty$.

One implication of this corollary is that *the absolute expectations $E(|X_r|)$ are bounded*. For,

$$E(|X_n|) \leq N + h_0 < N + G(N) < N + \varepsilon$$

for some N with ε as small as desired, and for all n. The corollary includes special cases as follows:

1. Markoff's: LLN *holds if* $E(|X_r|^{1+\delta})$ *exists and are bounded for all* r *and some* $\delta > 0$. *In this case, again, no loss of generality is caused by taking* $E(X_r) = 0$. *Therefore,*

$$\int |x|^{1+\delta} dF_r(x) < A \quad \text{leads to} \quad h < A/N^{\delta} = G(N)\,.$$

2. Khintchine's LLN *holds if all the* X_r *have the same distribution and* $E(|X|)$ *exists. In this case, for all* n,

$$h = \int_{|x|>N} |x| dF(x) = G(N) \to 0\,,$$

taking $E(X) = 0$ as before. Our result is only a slight and almost obvious refinement of existing deductions. The basic process (due apparently to Markoff) is the division of the variables into two portions, of which one is bounded and the other contains values of negligible probability. The question still remains: What is the function of the *second* moment for the bounded part in the deduction?

18.3 The Law of Large Numbers and the Central Limit Theorem

For a simple random variate X, bounded with $|X| \leq M$ and zero expectation, the probability of X lying outside a given interval centered at the origin can be increased by the increase of the dispersion. Under the conditions of the problem, there is an optimum "most scattered" distribution, independent of the particular and variable limits outside which X may be asked to lie from stage to stage; that is, $X = \pm M$ with probability $\frac{1}{2}$ each will give the greatest possible scattering once for all. Then, $P\{|X| > S\} = 0$ if $S \geq M$ and $P\{|X| > S\} = 1$ if $S < M$. If the value of S be preassigned and not greater than M, we have some choice in limiting the random variable, but the distribution given here is the optimum in that it gives the greatest probability, independently of S.

For two independent variables of the same type, the sum $X_1 + X_2$ can be dispersed in a similar manner. Here, $P\{|X_1 + X_2| > S\} = 1$, $\frac{1}{2}$, or 0 according as $S < M$, $M \leq S < 2M$, or $S \geq 2M$. The first can be obtained in particular by taking one of the variables $\pm M$ with probability $\frac{1}{2}$ each and the other 0 with probability 1. In the last case, no permissible choice of distributions can possibly give any other probability than zero. In the middle case, however, there does exist an optimum, i.e., $X_1, X_2 = \pm M$, $p = \frac{1}{2}$ each, so that the value of the sum would be $\pm 2M$ with $p = \frac{1}{4}$ each and 0 with $p = \frac{1}{2}$. This, incidentally, points out the essential reason for the validity of our LLN, in that the values of the independent stochastic variables with zero expectations cancel out. But it is of basic importance for the optimum to exist, in such addition, that the interval $2S$ be sufficiently large. For the purpose of the preceding section, it suffices to note that S must lie between $(n-1)M$ and nM to enforce the optimum scattering for the sum of n variates. In the variables $u_i - b_i$, S corresponds to $n\varepsilon - H$, and M to N, or rather to $N + b_0$. So, N must be of the same order as $\varepsilon - H/n$, which will not do in view of the fact that ε is arbitrarily small while N has to be taken arbitrarily large.

If we take the distribution with optimum scattering, $X_i = X = \pm M$, with $p = \frac{1}{2}$ each, then each X has the c.f.

$$\frac{1}{2}\left(e^{-int} + e^{+int}\right) = \cos Mt = 1 - 2\sin^2 \frac{1}{2}Mt\,.$$

The c.f. of the sum of n is therefore $(1 - 2\sin^2 \frac{1}{2}Mt)^n$, which, for the average of n tends to $(1 - M^2t^2/2n^2)^n \to \exp(-M^2t^2/2n)$. So the general distribution, bounded by the most scattered case, is thus bounded by a normal distribution with zero mean and variance N^2/n, in substance. If M and N are of the same order, we must have $(N + h_0)^2/n \to 0$ simultaneously with $n\varepsilon - H \sim n(N + h_0)$; but in LLN, it is essential to have $N \to \infty$ unless we restrict ourselves to almost trivial cases. These conditions obviously contradict each other.

The function of the second moment, or any moment higher than the first, is to bridge this gap between the requirements of LLN and the approach of this section, by consideration of the most scattered variables.

Suppose that to our preceding assumptions $E(X) = 0$, $|X| \leq M$, we add the condition $E(X^2) \leq kM$. In the former case, $E(X^2) = M^2$ for the most scattered distribution, so that for $k < M$ (*which we assume hereafter*), we have less concentrated scattering. The extreme limits $\pm M$ can no longer be attained as before with $p = \frac{1}{2}$. The greatest concentrated scattering is, in fact, now given by $pm\sqrt{kM}$; $p = \frac{1}{2}$ each. Nevertheless, $P\{|X| \geq S\}$ need not vanish if we push S beyond $\sqrt{kM}$ so that the actual optimum in this case is much less clear-cut. For $M > S \geq \sqrt{kM}$, the best choice is clearly to take $X = \pm S$ with probability $kM/2S^2$ for each value, and $X = 0$ with probability $1 - kM/S^2$. If the probability is reduced at either of the two extremes, it would be necessary to add an extra value on the same side of zero, or to cut down the probability for the other extreme, in order to preserve the mean value $E(X) = 0$. The optimum for this three-valued distribution is more clearly dependent upon the choice of interval, and the probabilities for the greatest scattering also depend upon S. The c.f. for a single such scattered variable is $1 - (2kM/S^2)\sin^2 \frac{1}{2}St$. For the sum of n, we raise this to the n-th power, for the average of n, we have only to replace t by t/n in raising to nth power. Similarly, for the sum of n, the second moment is $B_n \leq nkM$; for the average, we replace each X by X/n and B_n by $B_n/n^2 \leq kM/n$. So, in getting the distribution of the average of n values, we should not only raise the c.f. to the n-th power but also replace t by t/n and S by $S/\sqrt{n}$. Making these substitutions, we get the most scattered distribution of the average as having the c.f.

$$\left(1-\frac{2kMn}{S^2}\sin^2\frac{St}{2n\sqrt{n}}\right)^n \to \left(1-\frac{kMt^2}{2n^2}\right)^2 \to e^{-kMt^2/2n}\,. \tag{18.8}$$

In spite of its fundamental role, S has cancelled out in the process of deduction, to give a normal distribution for the bounding scattered variate. The mean is zero as before, the variance reduced to kM/n. This must tend to zero if LLN is to hold. For the attack of Sect. 2, we should have to take $M = N + h_0$, $k = c_0$, $S \sim \sqrt{c_0 n M} = n(\varepsilon - h_0)$. The condition would then take on the form $h_0 \to 0$, $c_0(N + h_0)/n \to 0$, which it is possible to fulfill, as, for example, under the assumptions of the preceding section.

That is, the second moment or any moment higher than the first helps only by restricting the admissible variation, the maximum possible scattering. The analysis of this section shows in addition to this a general connection between LLN and the central limit theorem. In fact, we have the limiting values of the bounding distributions as normal distributions.

Chapter 19
Possible Applications of the Functional Calculus

D.D. Kosambi, Tata Institute of Fundamental Research, Bombay

The 34th session of the Indian Science Congress was held in January 1947, with Pandit Jawaharlal Nehru presiding. There had been some correspondence on the manner in which the Science Congress was to be conducted between DDK and Nehru [DDK-JK], *who as vice president in the interim government, was clearly prime minister-in-waiting. The practice that continues to this day, that the prime minister presides over the Science Congress probably dates back to this time. Having joined TIFR in 1945, DDK benefited greatly from the association with Bhabha in the early (and very cordial) years. He was elected Fellow of the Indian National Science Academy in 1946, and in 1947, chosen President of the Mathematics section of the Science Congress. He was also awarded the Bhabha Prize (named for Jehangir Hormusji Bhabha, Homi Bhabha's father).*

In this somewhat didactic article DDK elaborates on his ideas for the proper orthogonal decomposition initiated in his 1944 paper, "Statistics in Function Space" *and also on calculating machines. By this time he had tried his hand at fabricating at least one, his "universal" calculating machine, the* Kosmagraph *discussed in Chap. 4. His biographers mention specifically that DDK did not proofread this paper, and thus it is not in the form that he would have wished. The typesetting was cavalier, particularly the equations. Also, the printed title read* Fundamental *Calculus before DDK corrected it on his copy.*

Published only as part of the Proceedings of the Indian Science Congress. Presidential Address to the Indian Science Congress, Delivered in New Delhi on 6 January, 1947.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_19

1. The concept of a function as developed today is neither simple nor in its definitive form. However, a great deal of what we have to discuss would be illustrated reasonably well by real continuous functions of a single variable. For us, functions are defined as sets of real numbers of which one is known for every value of another set of real numbers which are the values of the independent variable x. It suffices to select the case of a single continuous interval for the range of values of x; there are two essentially distinct cases here, as the interval is finite or infinite and of these again we keep to the first. By suitable choice of origin and scale, we may then take the fundamental x interval as [0, 1], including the endpoints; the function $f(x)$ then takes on a single real value for every value of x in the closed interval, $0 \leq x \leq 1$. Continuity at a point a in the fundamental interval would mean that to every $\varepsilon > 0$ however small, there corresponds another positive number δ such that $|f(x) - f(a)| \leq \varepsilon$ holds for all x with $|x - a| \leq \delta$, in [0, 1]. Naturally, this may be phrased in different ways. We can prove immediately by using the Heine-Borel theorem that since the whole closed interval [0, 1] is covered by intervals δ_1 (the function being continuous at each point there) each of bounded oscillation for $f(x)$, it is so covered by a finite number thereof. That is, a δ exists independent of the location of a in [0, 1], which gives us the notion of continuity and uniform continuity throughout an interval. With the Borel-Lebesgue theory of point sets, we pass to a further generalization labeled absolute continuity, where the definition could read $|f(x) - f(x')| < \varepsilon$ for x, x' on any set in [0, 1] of measure less than δ.

In the same way, our concept of function would change very greatly from that of a simple graph on [0, 1] with which, after all, we are most familiar in practice. We can have functions continuous at every irrational point and discontinuous at every rational point, as for example $f(x) = 0$, x irrational, and $f(x) = 1/q$ for $x = p/q$, where p, q are integers with highest common factor unity. But the opposite case of discontinuity at irrational and continuity at rational points is, according to a theorem of Young, not possible. This and theorems on oscillation [*Schwankung*] show that our definitions, apparently so very general, still limit discontinuity. On the other hand, the apparently simple concept of a function as a graph is still too general for our main purpose. It is "intuitively obvious" that every continuous graph has a tangent except at a discrete set of points—but it is not true! Nondifferentiable functions, beginning from Weierstrass, have been constructed by a procedure which does mathematically in an infinite number of smaller and smaller steps in the plane of the graph what a coiled coil wire filament of a modern electric light bulb does in at most three steps for a space curve. It is known from the time of Peano that a pair of continuous functions $y = f(t), x = g(t)$ can be found giving an unbroken curve in its parametric form which passes through or comes arbitrarily close to every point of a two- dimensional region in the $x - y$ plane. Finally, we follow Euclid in saying that a line, or a curve, has no area, and believe this too holds of the ideally thin curve which is represented by any material graph obtained in practice. If, however, we define the area of any region as the limit of sum of small squares that cover that region and decrease in size to zero, the "area" of a curve should be zero, in the mathematical sense, taking the curve in the sense of Jordan as separating the plane into two parts. But the above example shows what can be proved independently, that there exist Jordan curves of positive area.

There is a very good reason for pointing out all these complications when the idea of a continuous function is simple and capable of being dealt within a rigorous manner. In many of the processes that follow, we shall need mathematical tools of a special kind, namely successive approximations by sequences and series, where the process is infinite. It is then possible that the limiting cases obtained do not belong to the class of functions used in the infinite approximating process, and proofs to be fully valid must extend to the "*exotic*" forms that may arise. For example, before the Lebesgue theory of integration was formulated, it was impossible even to make a proper approach to the theory of convergence of Fourier series; thereafter, still more complicated and specialized integration processes have been introduced to facilitate that discussion in special cases where the Lebesgue integral fails.

2. Restricting the discussion to real single-valued continuous functions of a single variable x on [0, 1], and even to functions of a reasonably simple type, the next step is to visualize a function as a single point in a space of infinitely many dimensions. That is, the function being defined by its value for every x, we take each value of x as representing, in some way, a different dimension. The generality here obtained is necessarily restricted by the continuity of the functions or even otherwise by the result of Young which shows that functions cannot be too highly discontinuous. For, a continuous function is defined everywhere by its values on a set of points $\{x\}$ which is everywhere dense on [0, 1], seeing that the limiting points of the set give the whole of the fundamental interval and that the value of the function coincides with the limit of its values. Thus, the essential number of dimensions is reduced to a denumerable infinity. The ingenious generalization, however, causes some new difficulties of its own. The difference between two dimensions in Euclidean space is qualitative, or if the dimensions are arranged in some particular order and numbered, may be regarded as an integer which could be arbitrarily large for an infinite-dimensional Euclidean space. For our function space, the difference between two values of x has a totally different meaning and is a real number [actually the measure of an interval] never surpassing unity. Naturally, topological questions in function space and those dealing with pure geometry will have a totally different appearance. We may note in passing that though the method is actually of no use for our purpose, a geometry in infinitely many dimensions, utilizing the basic axioms and results of projective geometry such as Desargue's theorem, has been built up even when the number of dimensions is infinite of the same order as the number of points on [0, 1], the dimension being indicated by a coordinate x as in our case.

But we do make use of another basic idea in ordinary geometry. If, in any n-dimensional space with $n \geq 2$ we are given two vectors λ, μ, then the linear combination $a\lambda + b\mu$ is also a vector when a, b are arbitrary real constants not both zero. Moreover, either the linear combination sweeps out a two-dimensional manifold, or a single-dimensional variety; the condition for the latter being that $a\lambda + b\mu$ should vanish identically for some pair of values of a, b not both zero. Except in this latter degenerate case, the arbitrary constants of linear combination a, b can be regarded as coordinates with reference to the two given vectors which define all vector elements of the two-dimensional variety. A similar step can immediately be taken in our function space. We take the identically zero function as the origin, with which any point represented by the generic $f(x)$ defines a vector. Two functions $f(x)$, $g(x)$ are [linearly] independent if and only if the combination $af + bg$ cannot vanish for all x in [0, I] unless a and b are both zero simultaneously. Excluding this degenerate case where every linear combination is a multiple of the same function, we see that $af + bg$ can be taken as the generic function of a two-dimensional subspace of the function space. Again, the real numbers a, b can be regarded as coordinates of the submanifold, with reference to the given functions $f(x)$, $g(x)$. This process is extended to more than two functions of reference. Inasmuch as the fundamental space in all its generality is infinite dimensional, the really interesting case is where an infinite set of functions is used for reference and we have the expansion

$$h(x) = a_1 f_1(x) + a_2 f_2(x) + \cdots + a_n f_n(x) + \cdots .$$

The question then is: In what sense and what type of functions may one regard as represented by the infinite linear combination? The non-degenerate case here is one in which the basic set of functions $\{f_n(x)\}$ is closed with reference to the system of expansion and in the sense of the representation, i.e., the function *zero* can be represented only by the expansion in which all the coefficients a_i vanish simultaneously. The type of representation then involves some manner of determining the coefficients uniquely, and the kind of convergence [which may, for example, be some type of summability, or may be discussed almost everywhere] for which the expansion has some meaning. Here, we are led to the necessity of a process which has no precise analogue in the space of the geometry that we learn, integration.

3. The integral of $f(x)$ over the whole fundamental range [0, 1] is best visualized not as an area, but as an average. The "area between the graph and the x-axis" implies that areas have a sign—which may be admitted—and also that the space is Euclidean with $y = f(x)$ the same sort of variable as x, which is generally not true whatever its appearance on the material surface upon which the graph is traced. If we define the average or mean m as that constant which minimizes $\int [f(x) - m]^2 dx$ over [0, 1], we see at once that whenever $f(x)$ and its square are integrable, the minimum does exist and is actually the Riemann or Lebesgue or some such integral of $f(x)$, depending upon the process used. By integration, then, we shall mean an operation defined over [0, 1] as well as arbitrary subsets thereof, with the following properties: *i.* For every integrable function $f(x)$ any two point sets E, E' in [0, 1],

$$\int f_{E+E'} = \int f_E + \int f_{E'} - \int f_{EE'}$$

where $E + E'$ is the set sum and EE' the crosscut.

ii. For every set fixed in advance and two integrable functions,

$$f(x), g(x), \int [f \pm g] \text{ exists and is equal to } \int f \pm \int g.$$

iii. $\int af = a \int f$, for all constants a.
iv. If $f(x) \geq 0$ and $\int f$ exists, then $\int f \geq 0$; if $f \leq 0$, $\int f \leq 0$.
v. $\int 1$ exists for every permissible E in [0, 1].

Without troubling ourselves here about the completeness and independence of these systems of working postulates, we see that $\int 1$ is a function of E which we may hereafter call the measure of E, and this measure cannot be negative, according to (*iv*). These postulates are trivially satisfied by the value 0 for the integral of every function over any set whatsoever, so that we need one more postulate:
vi. The measure of [0, 1] is unity: $\int 1 = 1$ if E = [0, 1].

This shows at least that [0, 1] cannot be subdivided into a finite number of sets of zero measure; in infinite operations, we need some similar result, or axioms extending (*i*) and (*ii*) for infinitely many operations, but we shall take those for granted when the necessity arises.

It may be noted that we no longer write $\int [\ \ ]dx$, for the integration may depend upon the variations of some function, as in the case of Stieltjes integral, in which case we shall have to write $\int [\ \ ]dg(x)$. With this measure, we may then evaluate the integral as a limit in one or more ways, as the function integrated may allow. For example, for continuous functions, we may divide [0, 1] into smaller and smaller intervals and then show that the two sums formed by multiplying the maximum and minimum, respectively, in each interval by the measure of that interval bound [by (*ii*) and (*iv*)] the integral; these sums are easily shown to converge to a limit which is necessarily the integral in question. So also for a subinterval, or finite set of subintervals of [0, 1]. However, this is not the Riemann integral for the simple reason that nothing in our postulates makes the measure of an interval [except [0, 1]] equal to its length; for that matter, two intervals of equal length need not have the same measure. This can be seen by taking the integral as an ordinary Riemann integral with a nonnegative weight function attached to dx in the integration which may further be visualized as a probability distribution over [0, 1], the integral itself being then the expectation of the function $f(x)$. For the Lebesgue type of integration, we may subdivide [0, 1] into mutually exclusive sets, or at least sets whose pairwise common points are, taken in their totality, a set of measure zeros, and again take limits. The existence of such a limit necessarily depends upon the nature of the function itself, the question for given classes of function [particularly when convergence of series of orthogonal functions is being discussed] being: "What limiting processes should one use to yield the existence of an integral of some sort in this case?"

4. Given the particular class of functions admitting some process of integration and constituting a function space for our purpose, we may introduce a simplification in our coordinate system as proposed at the beginning of the last section. This parallels the use of orthogonal coordinate axes in Euclidean space. In the first place, it is clear that any number of the reference functions, given as linearly independent at the start,

may be replaced by the same number of linearly independent linear combinations thereof, and the new system of reference will do equally well for the space or subspace in question. We use the standard procedure of orthonormalization as follows, for a finite or infinite sequence $\{f, [x]\}$:

$$\phi_1(x) = af_1; \text{ with } \int \phi_1^2 = 1. \quad (a, \alpha, \ldots \text{generic constants}).$$

$$\phi_2(x) = af_1 + bf_2; \int \phi_2^2 = 1, \int \phi_1\phi_2 = 0.$$

$$\phi_3(x) = af_1 + bf_2 + cf_3; \int \phi_3^2 = 1, \int \phi_1\phi_3 = \int \phi_2\phi_3 = 0, \text{ and so on.}$$

Here and hereafter, all unspecified integrals will be taken as extended over [0, 1]. It is then seen that the process leads to the independent linear combinations desired and is uniquely defined, besides being easy to apply. The principal advantage is that where there is an infinite basis for the function manifold, we can determine the coefficients of the expansion of any integrable function, provided the process of term by term integration is justified:

$$f(x) = a_1\phi_1 + a_2\phi_2 + a_3\phi_3 + \cdots + a_n\phi_n + \cdots$$

where multiplying by ϕ_r on each side and integrating gives, taking into account the property of orthonormality, $a_r = \int f\phi_r$.

Moreover, these orthonormal functions 0 have, individually or collectively in finite or infinite number, the averaging property with which we started the section. If we take the integral

$$\int \{f(x) - (a_1\phi_1 + a_2\phi_2 + \cdots + a_n\phi_n)\}^2$$

and ask what choice of coefficients a_i minimizes it, we get again the same answer as for the formal infinite series expansion as before for each ϕ_1 actually present. Such a thing is not true, for example, of the Taylor series, which means expansion in terms of the set $\{1, x, x^2, \ldots, x^n, \ldots\}$ for analytic functions. The minimizing of the squared difference will not, no matter what the fundamental set or interval non-trivially chosen, give the same coefficients for representing a given general type of function if a few of the expansion functions are omitted. Naturally, the convergence for our series of orthonormal functions is also of a different type than for the Taylor series, being associated with the process of integration. In fact, the value of the minimized integral above is

$$\int f^2 - a_1^2 - a_2^2 - a_3^2 - \cdots - a_n^2 \ldots$$

which tells us incidentally that the sum of squares of the coefficients of expansion converges. If now the difference between $\int f^2$ and the sum of squared coefficients converges to zero, the function is properly represented, the series *converging in the mean*. In order to represent all integrable functions in this way, it is necessary and sufficient that the set $\{\phi_n\}$ should be closed, i.e., that no function except the one identically zero should exist with the property $\int f\phi_n = 0$ for all n. Further, the Riesz-Fischer theorem tells us that for a closed set of orthonormal functions and any set of real coefficients $\{a_n\}$ such that $\sum a_n^2$ converges, there exists an integrable-square function to which the series $\sum a_n\phi_n$ actually converges in the mean and which then has the coefficients of expansion a_n.

The orthogonal functions do this work in a quite remarkable way, by changing sign more and more often. This is easily seen for all the polynomials in use, for all Sturm-Liouville type of expansion, functions, and for others defined by the differential equations of mathematical physics. As the convergence of $\sum a_n^2$ implies that a_n tends to zero, and therewith not only $\int f\phi_n$ over [0, 1] but over any subinterval thereof, follows that there must be either an increasingly rapid tendency to zero for ϕ_n, or more and more changes of sign. The full impossibility of the first has not been shown as yet. The functions themselves have a peculiar averaging property, for it is easily shown that over [0, 1] or any subset thereof any infinite sequence of orthonormal functions is summable *in the mean* to zero, by any method of summability in which the maximum weight attached to each function tends uniformly to zero after some stage. At the same time, no orthonormal sequence can converge to any limit function, which means that the sequence cannot be uniformly continuous in the subscript. One consequence is that this type of function space is not locally compact. We can have an infinite sequence of points arbitrarily close to the origin which does not converge, as for example $\varepsilon\phi_1, \varepsilon\phi_2, \ldots, \varepsilon\phi_n, \ldots$ which tends to no limit in the function space for any fixed ε no matter how small. This may be visualized as the non-convergence of points on a small sphere about the origin which are in different orthogonal directions.

5. We have actually ended the preceding section by using the intuitive but undefined concept of distance. For an abstract space of elements $f, g, \ldots$, the distance D, if any exist, is defined by the following postulates:

i. $D(f, g) \geq 0$.
ii. $D(f, f) = 0$, and $D(f, g) = 0$ implies the equivalence of f and g.
iii. $D(f, g) = D(g, f)$.
iv. For any three distinct elements f, g, h, $D(f, g)D(g, h) \geq D(f, h)$.

If, for some suitable arrangement of the three elements, the equality can always be made to hold in the last [triangular] relation, then the space is linear, and the elements can be arranged in some sort of order. We have already pointed to two possible definitions of distance. The first was the "absolute value" $|f - g|$, which is a nonnegative function and whose integral is a pure nonnegative number that satisfies the postulates above provided equivalence in (*ii*) is understood in the proper sense: For Riemann integration, some stepwise discontinuities are permissible in the difference; for Lebesgue integration, one can demand equality "almost everywhere," i.e., at most except over sets of measure zero, when the distance between the two functions vanishes. The other form, which is much nearer to our Euclidean concept, is definable in terms of the expansion coefficients of f, g with respect to some given set of orthonormal functions ϕ_n. If $f \sim \sum a_n\phi_n$ and $g \sim \sum b_n\phi_n$, then

$$D^2 = \sum [a_n - b_n]^2.$$

Here again, either the set of $\{\phi_n\}$ must be closed or the functions that constitute the elements of the space must be such as are expansible in terms of the ϕ_n; otherwise equivalence must be understood as to within the possibilities of such expansion. Actual equality, again, depends upon our generalized integration and measure, so that sets of generalized measure zero may be excluded; in the probability example used, two continuous functions whose distance apart is zero must coincide except over sets where the probability vanishes.

Even the latter type of distance, however, is not general enough for our purpose; besides, the fundamental orthonormal set $\{\phi_n\}$ has to be given in advance. We recall that in a finite number of dimensions a positive definite quadratic form in the coordinates has the same property as the sum of squares when it comes to defining D^2. Proceeding by analogy, we replace the matrix Q of the form by a function of two variables $K(x, y)$ called the kernel; the summation process is then replaced by integration, so that $D(f, g)$ is to be defined by

$$D^2(f, g) = \iint K(x, y)[f(x) - g(x)][f(y) - g(y)]dxdy$$

(*integration over the unit square* $0 \leq x \leq 1, 0 \leq y \leq 1$.) This is again seen to have all the necessary properties provided $\iint Kf(x)f(y)dxdy$ is never negative, vanishing only for functions equivalent to zero, while equivalence is suitably defined for the second postulate. To save space, we shall adopt the notation

$$Xf = \int K(x, y)f(y)dy,$$

(y variable of integration). Thus, regarding $f(x)$ as a vector, it is seen that Xf is a vector of contragradient type, i.e. covariant if f itself be given as contravariant. The square of the distance of f from the origin then becomes

$$\int fXf, \text{ and } D^2(f, g) = \int [f - g]X[f - g].$$

Suppose a set of orthonormal functions be desired that minimizes D^2 according to this present definition between an arbitrary f and any number of terms of the series $\sum a_n\phi_n$. That is, we are to determine a_i so that

$$\iint K(x, y)[f(x) - a_1\phi_1(x) - a_2\phi_2(x) - \ldots][f(y) - a_1\phi_1(y) - a_2\phi_2(y) - \ldots]$$

becomes a minimum. The conditions are

$$\iint \{K(x, y) + K(y, x)\}[f(x)\phi_1(y) - a_1\phi_1(x)\phi_i(y) - a_2\phi_2(x)\phi_i(y) - \ldots] = 0, i = 1, 2$$

These become less awkward if the kernel is symmetric, i.e. $K(x, y) = K(y, x)$, in which case Xf may be defined by integration with respect to either the first or the second variable, the case corresponding to that of the symmetric quadratic form generally used in geometry. Our conditions then become

$$\int fX\phi_i - a_1 \int \phi_1 X\phi_i - a_2 \int \phi_2 X\phi_i + \ldots, i = 1, 2, \ldots$$

These still do not permit direct use of the orthonormality properties. But now suppose that the set of expansion functions is intimately related to the kernel by $X\phi_n - \lambda_n\phi_n$. This would lead immediately to precisely the same determination of the coefficients a_i provided none of the λ_i vanishes, which is guaranteed by the definiteness of the kernel.

Every function $f(x)$ represented in terms of the set $\{\phi_n\}$ associated with a given K is transformed into another by Xf, the former coefficients a_i, being multiplied each with corresponding λ_i. Thus, X is the operator of a linear transformation in the function space. However, this does not represent the identity, for there exists no non-singular kernel which transforms every function of the space into itself. For that matter, the continuous kernels with which we operate transform the space of all integrable functions into the subspace of continuous functions, as can be seen by taking Riemann integrability and stepwise continuous functions for elements of the space. This last cannot be avoided, being concomitant of the integration process, as explained. But we can add the identity to our transformations and then generate a one-parameter Lie group by means of X, the infinitesimal transformation being $X\delta\tau$. That is, we are led to expand the transformation as a series

$$f = f + tXf + \frac{t^2}{2!}X^2 f + \cdots + \frac{t^n}{n!}X^n f + \cdots .$$

The kernels associated with $X^2, \ldots, X^n \ldots$ are the iterates of K, given by

$$K^2 = \int K(x, u)K(u, y), \ldots K^n = \int K(x, u)K^{n-1}(u, y),$$

(u variable of integration). Ignoring the question of convergence, the formal theory is carried through directly. For the characteristic functions ϕ_r and characteristic values λ_r, of K, we get the same ϕ_r and characteristic value λ_r^n for K^n. The full transformation for such a ϕ_r gives

$$\phi_r = \phi_r e^{\lambda_r t}.$$

Finally, several such one-parameter transformations for a group if the defining operators X_i obey the laws

$$X_i X_j - X_j X_i = \sum c^r_{ij} X_r; \quad c^r_{ij} = c^r_{ji};$$

$$\sum_r (c^l_{rj} c^r_{kl} + c^i_{rk} c^r_{lj} + c^i_{rl} c^r_{jk}) = 0.$$

These constants c^i_{jk} are, as usual, the constants of structure of the formal multiparameter Lie group and define it completely as a group. Naturally, the symmetric kernels to which we restrict ourselves in most cases give only Abelian groups with vanishing structure constants. If we admitted complex variables, defined the path of integration suitably and overcame certain difficulties at the endpoints of the fundamental range or boundary, we could use even singular kernels as in the Cauchy contour integral and make the operation X correspond, for example, to a differentiation, which would make the expansion yield such values as $\bar{f} = f(x+t)$.

6. The foregoing treatment is purely formal, as it has not been shown that the necessary relation between the symmetric kernel and the orthonormal expansion set can subsist; or conversely that for a given kernel there exist associated orthonormal functions. This is comparatively simple and straightforward for kernels $K(x, y)$ that are sufficiently small and have a bounded difference quotient in each variable, i.e.,

$$|K(x, y)| \leq 1; |K(x+h, y) - K(x, y)| < hM, M \text{ constant}.$$

The former restriction is not very important, as it follows from the uniform continuity of our kernels in the unit square which gives $|K| \leq G$ and then by taking a new kernel which is the old one divided by its upper bound G. Then, it follows at once that for any function continuous in [0, 1], $Xf, X^2 f, \ldots, X^r f$ are all bounded and of difference quotient bounded uniformly for all values of the index n. By a well-known theorem of Ascoli, such a set of functions has at least one limiting function, which has also the property of continuity. We utilize this as follows: Start with any smooth function $f(x)$, and take $\mu_1 f_1 = Xf$, determining μ_1 so as to make $\int f_1^2 = 1$. Again, $\mu_2 f_2 = X f_1$, normalizing f_2 by proper choice of the characteristic value μ_2. This process leads to a limit function ϕ_1 (taking subsequences if necessary) that does not vanish, and as the kernel was definite by hypothesis, the characteristic values will also converge to a limit λ_1. We start again with a smooth function, and repeat the process, but this time taking the initial function orthogonal to ϕ_1. As symmetry of the kernel gives $\int \phi X = \int f X \phi$, it follows that all the successive transforms are also orthogonal, and we are led to another function ϕ_2 characteristic of the kernel, with a second characteristic value λ_2, the functions ϕ_1, ϕ_2 will be orthonormal, hence distinct, whether the characteristic values are distinct or not. It remains to be pointed out that dividing the original kernel by a constant only means dividing all the characteristic values by the same constant.

For degenerate kernels, symmetric or not, the existence theorem is proved by a matter of simple algebra. A kernel is degenerate if it consists of the sum of products of a finite number of functions. Taking these without loss of generality as orthonormal among themselves, we may take

$$K(x, y) = \sum a_{ij}\phi_i(x)\phi_j(y).$$

Then, the equation $Xf = \lambda f$ has solutions

$$f = b_1\phi_1 + b_2\phi_2 + \cdots + b_r\phi_r,$$

provided the characteristic values are roots of the equation

$$|a_{ij} - \lambda\delta_{ij}| = 0.$$

If the matrix $||a_{ij}||$ was in its diagonal form, then the kernel would be in its simplest representation:

$$K(x, y) = \sum \lambda_i\phi_i(x)\phi_i(y).$$

The question is whether this is possible for the general symmetric kernel, though the algebraic reduction shows that symmetry is not absolutely essential. If the expansion was so possible, we see that the existence of $\iint K^2$ leads to the necessary condition, $\sum \lambda_i^2$ converges. It remains to show that the original infinite series would also converge and to the value $K(x, y)$, under very general conditions.

In the first place, if the double series

$$\sum \lambda_i\phi_i(x)\phi_i(y)$$

converged to any other value [as an integrable function] except $K(x, y)$, then the difference would be treated as a new kernel for which we could show the existence of another characteristic function, which could not be any of the known ϕ_i; hence K would not be definite. Again, it may be noted that $\lambda_n\phi_n(x)$ is the expansion coefficient c_n of $K(x, y)$ in terms of $\sum c_n\phi_n(y)$. Therefore, the series

$$\sum \lambda_i\phi_i(x)\phi_i(y)$$

certainly converges in the mean, in the sense that the difference between the series and $K(x, y)$ squared integrates to zero over the fundamental range. Ordinary convergence, according to the known properties of convergence in the mean, depends then only upon the smoothness of the $\{\phi_n\}$, i.e., ultimately of $K(x, y)$, and in any case, a subsequence of the double series always converges to the proper limit. We are thus left with an expansion

$$K(x, y) = \sum \lambda_i \phi_i(x)\phi_i(y)$$

which will be very useful later. The entire treatment could have been by reduction to the case of degenerate kernels, or by Weierstrass' polynomial approximation theorem applied in two dimensions.

7. We come now to the applications, of which many are already known to the equations of mathematical physics, particularly in the systematic reduction of partial differential equations with given boundary value problems. In particular, we have integral equations of type

$$f(x) = g(x) \int K(x, y)g(y)dy,$$

f known, g to be found. Visualize the right-hand side as $[ItX]g(x)$, so that the inversion is immediately available if the formal expansion

$$[I - tX]^{-1} = I + tX + t^2X^2 + \cdots$$

converges, I being the identity, $If \equiv f$ for all f. In this case, there can be no characteristic function or value with the property $t\lambda_n = 1$, or else the expansion would fail. Now it may happen that the kernel in question is such that there is no characteristic function at all, as for example with the Volterra type differential equations in which $K(x, y) = 0$, $y \geq x$. Alternatively, there exist characteristic values but none is the reciprocal of the particular t chosen in which case we can expand $f(x)$, $g(x)$ in terms of the characteristic functions ϕ_r and determine the coefficients of $g(x)$ as those of $f(x)$ divided by $[1 - t\lambda_n]$.

This need not detain us long, as the discussions are now classic. I come to a comparatively new application to statistical problems where the observed and observable variates are themselves functions defined on [0, 1], and at least stepwise continuous. Now it is known or can be seen from any classical text on the theory that the basic probability distribution in n-space is the normal distribution, and that this depends upon the definition of both distance and volume in the space. That is, we have

$$dP = C \exp(-r_n^2/2)dV_n,$$

the constant C being chosen in such a manner as to make the total probability distribution over the whole n-dimensional infinite continuum equal to unity. This may be generalized in two ways, namely to spaces which allow both distance and volume measure to be defined without being continua such as for example, the Cantor ternary set—which is purely a mathematical fiction at present. It suffices, according to an elegant result of Haar, that the space has distance, a relation which permits volume elements to be superposed one upon the other, that it be separable, and is locally compact. This, incidentally, makes the space "almost finite dimensional" and excludes our Hilbert space. However, the Haar measure of volume is then taken in essence by covering any region of the space by smaller and smaller standard volumes, say spheres of decreasing radius, and taking the lower limit of these *in terms* of the same lower limit for some fixed region like the unit sphere. This differs from the usual procedure in that there we cover by intervals or such elements which in themselves have known measure, in that the measure of a covering element is not taken as known a priori. A practical example is furnished by the measurement of skull volume by shot or seed of decreasingly smaller size. At each stage, one may count the actual number of shot so used and take the ratio for a corresponding count of the number of shot or seed of the same size needed to fill a glass beaker. The ideal limit is then the volume of the skull in terms of the beaker as unit.

For our purpose, this will not do at all, as limiting processes in covering are excluded by the Hilbert space not being locally compact, except when we have a degenerate kernel. However, the kernel itself taken as positive definite and symmetric enables distance to be defined. It need only to be definite for the space of function observed, in the sense that no observable function transforms to zero by means of Xf. Now, we take the kernel in its canonical form

$$\sum \lambda_i \phi_i(x)\phi_i(y),$$

and define normal distribution in the function space as normal distribution in *each* of the coordinates, the coordinates themselves being the coefficients of expansion of the population functions f in terms of the ϕ_n. Here, the population mean has been taken as zero without loss of generality, but we could have taken some function expansible by the ϕ_n as the population mean, in which case the corresponding coefficient would give the population mean for each coordinate. The variances must here be taken equal to the *characteristic values* λ_i, so that the kernel defines the distribution.

Having introduced the concept of probability, the question now arises whether there is a one-to-one correspondence between a random sequence of the coefficients and the population functions f. Given the function, the coefficients are available, but is the converse true for a random sequence of coefficients? A theorem of Kolmogoroff enables us to answer in the affirmative under the conditions of the problem in the sense of unit probability of convergence of the series, whereas the Riesz-Fischer theorem would not hold, without further restriction, and would then give only convergence in the mean with unit probability. As the canonical coordinates are taken as orthogonal, i.e., statistically independent variables, it is possible to define the probability without further difficulty. Theorems on linear combinations of normally distributed variates show us that the values of the population functions at any point u on [0, 1] are normally distributed, those at two or more points are in multivariate normal correlation. The variance at a point x is $K(x, x)$ while $K(x, y)$ is the covariance between functions sampled at two values x, y of the independent variable. The distribution in many variates is proper provided the kernel contains more characteristic functions than the number of distinct points taken, The population, mean function, and the population kernel are approximated, for a given sample of n, by

$$m(x) = -\frac{1}{n}\sum_{i=1}^{n} f_i(x);$$

$$K(x, y) = \frac{1}{n-1}\sum_{i=1}^{n}\{f_i(x) - m(x)\}\{f_i(y) - m(u)\}.$$

One can apply multivariate tests at a finite number of fixed points or replace them by obvious integral tests. For example, Hotelling's T^2 test would mean finding

$$R^2 = \iint S(x, y)[m(x) - \bar{m}(x)][m(y) - \bar{m}(y)],$$

where

$$\int S(x, y)K(x, y) = I,$$

and

$$\exp 2t = cR^2/(1 - R^2),$$

with $c = (n - p + 1)/p$, if $K(x, y)$ degenerate with $p < n$, and $1/n$ otherwise.

8. We now come to the stage where a mathematical laboratory is essential for the realization of all these applications. The cinema intergraph, which measures the light gathered by the [unshaded] area on the film images of the curves, is useful in rapid integration, and therefore in solving integral equations to a rather rough approximation. For ordinary integration, we have the mechanical Bush analyzer of which the basic theory is as old as the planimeter and Kelvin's integrators, but whose practical success is due solely to a remarkable technical advance in the nature of a gearbox without backlash. The new Bush analyzer utilizing an immense number of electronic valves is meant to solve partial differential equations, as are Soviet inventions about which no details have as yet become known. But in these cases, it takes a considerable amount of time to prepare the data for the machine which then solves the problem almost in a flash, or even in the mechanical case with unusual rapidity. What we need is something that will give the sum, differences, the product of functions, and any number of combinations of these; will integrate; and will do all this without retouching the data graphs as obtained in practice. Finally, the calculator will have to operate not only with rectangular graphs but also with those in polar coordinates, and still worse those with one set of lines curved, as are generally obtained by recording on cylindrical drums.

If such instruments exist and are reasonably simple to operate, we can answer questions of great importance in practice. For example, not only will the average temperature curves be available for any range or period, but also it will be possible for us to say whether two samples from two different places differ materially. The question whether two sets of pulse-wave records, say of the effect of a certain illness or drug, differ significantly would be settled by the same methods. Distinguishing between skulls found by the archaeologist or anthropometrician in two different places would be treated as a problem of curves in polar coordinates if one considers profiles alone; our methods, however, are extensible to more than two dimensions, though with more complicated machinery needed in practice. In any case, the method will be much more satisfactory than the extraordinarily cumbrous set of characters and indices in general use today. All of these would be new applications, though the problems have been attacked by older methods without utilizing all the data.

A similar type of problem arises in generalized harmonic analysis, by which is meant not the reproduction of a given curve by means of Fourier series with arbitrary fundamental period but the detection of hidden periods and their intensities from a given graph. We know that a period $\nu/2\pi$ appearing as a term

$$a_\nu \cos \nu x + b_\nu \sin \nu x$$

can be detected by the operation

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-itx} f(x)dx.$$

This is a spectrum function, of the type attained by physicists in energy level calculations, which vanishes for t not equal to any of the ν's and is equal to the coefficient of $\exp i\nu x$ when such a term enters into the composition of $f(x)$. Naturally, we are not concerned with continuous spectra for the present though they can also be treated.

For the graph, it is necessary and for approximation purposes also sufficient—to have the curve decidedly longer than any of its basic periods. Then, $\cos tx$, $\sin tx$ can be generated mechanically and imposed upon $f(x)$, the product being mechanically integrated. This must be done for sufficiently many values of t to give the result as a stepwise continuous function, and periods can be more precisely located by finer shifts in t values, once approximate location has been performed. Some such methods, combined with statistical analysis of the error arising from measurements and recording of the original graph, will be needed to settle the still disputed question of degree and quality of similarity between the sunspot and the magnetic variations cycles. A further application would be to the problem of mixed populations in statistics. Here, we know that the ordinary Fourier transform gives $\exp\{i\mu x - \frac{1}{2}\sigma^2 x^2\}$, where μ is the population mean and σ^2 the population variance. Thus, the first transform of the sum of such probability densities consists of a sum of periodic terms multiplied by factors that tend to zero. These factors must be eliminated experimentally, by multiplying with exponentials $\exp \mu x^2$ till one is found which keeps the graph level at the ends of the range without letting it tend to infinity or to zero asymptotically. Then, a harmonic analysis gives the mean corresponding to that variance. This component of the population is then eliminated by subtraction, and the process repeated, provided the machine allows all this to be done. The method may be of use in determining location of lines in diffuse spectrograms, as for example in fine structure.

References

§ 1, 2, 3. C. Carathéodory, Mengenlehre, in *Reelle Funktionen*, 2nd edn., ed. by F. Haussdorf.

§ 4, 5, 6. R. Courant, D. Hilbert, in *Methoden der Math. Physik*, Band I (2nd ed., 1931).

§ 7. D.D. Kosambi, Statistics in Function Space. J. Ind. Math. Soc. **7**, 76–88 (1943).

Chapter 20
Lie Rings in Path Space

D.D. Kosambi, Tata Institute of Fundamental Research, Bombay

This work, one of two papers that DDK published in the PNAS, was communicated by Oswald Veblen of the Institute for Advanced Study, Princeton on May 8, 1949. The article appeared in print in the July 15 issue of the same year. DDK was visiting the USA during this year and had also visited Princeton where he met Veblen, Einstein, and others.

1. In studying the tensor analysis of the system of differential equations

$$\ddot{x}^i + \alpha^i(x, \dot{x}, t) = 0\,; \qquad i = 1, 2, \ldots, n\,; \quad \dot{x}^i = \frac{dx^i}{dt}, \text{ etc} \tag{20.1}$$

their equations of variation

$$\theta u^i \equiv \ddot{u}^i + \alpha^i_{;r}\dot{u}^r + \alpha^i_{,r}u^r = 0 \tag{20.2}$$

have been found to be a prime tool of exploration. The notation used is, for any function $\varphi(x, \dot{x}, t)$,

$$\varphi_{,r} = \frac{\partial\varphi}{\partial x^r}\,; \quad \varphi_{;k} = \frac{\partial\varphi}{\partial\dot{x}^k}\,; \quad \dot{\varphi} = \frac{d\varphi}{dt} = \frac{\partial\varphi}{\partial t} + \varphi_{,r}\dot{x}^r - \varphi_{;r}\alpha^r \tag{20.3}$$

and the tensor summation convention is followed for indices repeated in subscript and superscript.

These equations of variation (20.2) to be regarded as partial differential equations in u^i are obtained from (20.1) by the "infinitesimal change," $\bar{x}^i = x^i + u^i\delta\tau$, where $\bar{x}$ and x are supposed to be coordinates of points on "nearby" paths. Thus, $\delta x^i = u^i\delta\tau$ and the convention usually adopted is $\delta\dot{x}^i = \dot{u}^i\delta\tau$, which is derived from the assumption $d\delta - \delta d = 0$; this appears in all classical texts without further clarification. The

Published in Proceedings of the National Academy of Sciences of the USA, **35**, 389–94 (1949).

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_20

α^i are assumed to be arbitrarily differentiable, which allows an expansion in which second and higher powers of the parameter $\delta\tau$ are to be neglected.

2. The only meaning (unless we fix the basic path in (20.2)) that can be attached to $\delta x^i = u^i \delta\tau$ is of an infinitesimal transformation of a one-parameter Lie group. But the solutions of $\theta u^i = 0$ are generally functions of $x, \dot{x}, t$, whence the most general such one-parameter group must be taken as operating in the V_{2n+l} of $t, x, \dot{x}$, where $\dot{x}^i$ is to be regarded as a fiber bundle attached to a generic point of the base-space defined by coordinates x^i. The only operator which can be used here must necessarily be of "the first extension," i.e.,

$$X \equiv u^r \frac{\partial}{\partial x^r} + \dot{u}^r \frac{\partial}{\partial \dot{x}^r} \tag{20.4}$$

and the usual infinite series expansion

$$\bar{\psi} = \psi + \tau X\psi + \frac{\tau^2}{2} X^2\psi + \cdots \tag{20.5}$$

can then be obtained, at least symbolically, in the case where the α^i are analytic, to which we restrict our entire discussion. Given several distinct solutions u^i, v^i, w^i, ... of the equations of variation with the associated (extended) operators $X, Y, Z, \ldots$, the question then naturally arises as to the existence of a Lie group, with more than one parameter being formed in some way out of these. For this, the alternant $\{X, Y\} = XY - YX$ must itself be an operator of the first extension with respect to the paths. That is, if

$$XY - YX = \mu^r \frac{\partial}{\partial x^r} + \lambda^r \frac{\partial}{\partial \dot{x}^r}\,, \qquad \text{then} \quad \dot{u}^r - \lambda^r = 0\,. \tag{20.6}$$

Furthermore, the Jacobi condition must also be satisfied:

$$\{X, \{Y, Z\}\} + \{Y, \{Z, X\}\} + \{Z, \{X, Y\}\} = 0 \quad \text{for all} \quad X, Y, Z\,. \tag{20.7}$$

After this, we can see whether the solutions of the equations of variation form a Lie algebra, and over what fields.

Direct calculation gives us

Theorem 1 *Two (extended) operators X, Y associated with vectors u^i, v^i alternate to give one of the same type if the vectors concerned are (each) solutions of the equations of variation $\theta u^i = 0$ or of $u^i_{;j} = 0$. Similarly for the Jacobi condition on three operators.*

The proofs can be shortened considerably by noting another result.

Theorem 2 *The solutions of $\theta u^i = 0$ are just those whose associated operators permute with the linear operator d/dt of (20.3).*

Proof We have

$$X\frac{d}{dt} - \frac{dX}{dt} \equiv \theta u^r \frac{\partial}{\partial \dot{x}^r}\,, \tag{20.8}$$

which suffices for the second theorem. To use this result, we may put $Z = XY - YX$, and note that

$$\frac{dZ}{dt} - Z\frac{d}{dt} = \frac{dXY}{dt} - \frac{dYX}{dt} - XY\frac{d}{dt} + YX\frac{d}{dt} = 0\,. \tag{20.9}$$

The vanishing identically follows from the permutability of d/dt with both X and Y, and then proves that $XY - YX$ is also permutable with d/dt, whence the associated operator for Z must be formed from the solutions of the equations of variation, provided of course it was of the first extension. For solutions of $u^i_{;j} = 0$, the proof is trivial. Similarly for the Jacobi identity. □

Theorem 3 *The solutions of $\theta u^i = 0$ form a vector space which gives a Lie ring over the set of functions φ with $\dot{\varphi} = 0$ (i.e., constant along any path) and a Lie algebra (and therefore defines a Lie group in the analytic case under discussion) over the field of all real constants.*

The latter statement follows from well-known results in Lie groups. For the former case, we have merely to note that φu^i is a solution of the equations of variation with u^i provided $\dot{\varphi} = 0$. However, we only have here $\{\varphi X, Y\} = \varphi\{X, Y\} - (Y\varphi)X$, whence we get a *ring* over the set $\dot{\varphi} = 0$ defined by the vector space of solutions of $\theta u^i = 0$. The basis of the ring is clearly of dimension $2n + 1$. For the Lie group, however, if we take the most general case, the dimension cannot be finite, and the question remains open whether the infinite Lie algebra and group thus obtained are equivalent to E. Cartan's infinite Lie groups.

3. We now consider the subgroups (over the field of all real constants) leaving the base-space of the x as well as the paths invariant; the generators must now satisfy $\partial u^i/\partial t = 0$; $u^i_{;j} = 0$; $\theta u^i = 0$. This leads to

Theorem 4 *The Lie group leaving the base space as well as the paths invariant is of order $\leq n(n + 1)$.*

It suffices to show that the total number of arbitrary parameters in the solutions is finite, $\leq n(n + 1)$ for then these can be specialized to give that number of independent basic solutions and the linearity of the equations allows a general solution to be formed out of linear combinations of these basic solutions. That is, the number of essential parameters being determined, they can be taken to occur in the linear combinations alone.

In this case, we have $\dot{u}^i = u^i_{,r}\dot{x}^r$ and $\ddot{u}^i = u^i_{,j,k}\dot{x}^j\dot{x}^k u^i_{,r}\alpha^r$. Now the equations $\theta u^i = 0$ may be differentiated successively, because of $\partial u^i/\partial t = 0$ to give a succession of homogeneous linear conditions on $u^i_{,j}$, u^k. Differentiating the equations of variation with respect to $\dot{x}^s$ successively gives on the second differentiation an explicit equation for $u^i_{,j,k}$, and thereafter linear homogeneous restrictions in u^i, $u^i_{,r}$. From the original equations $\theta u^i = 0$, and from the first $\dot{x}$ derivative thereof, we may therefore eliminate $u^i_{,j,k}$ and obtain two more linear homogeneous restrictions on $u^i_{,r}$, u_j. The problem therefore is reduced to solving a system of first-order partial differential equations for the variables u^i and $v^i_j = u^i_{,j}$ (adjoining this last differential equation to the system), along with linear homogeneous restrictions upon the variables v^i_j, u^k, to which may be added others derived from the compatibility conditions. In any case, the solution $u^i = v^j_k = 0$ always exists, corresponding to the identity as the sole Lie group for the path-space. But it is well known (cf., E. Goursat, Chap. 1 of "Leçons sur ... équations aux dérivées partielles") that the total number of arbitrary parameters in the general solution is equal to the number of variables which cannot be eliminated from the conditions of compatibility and the linear restrictions, which in no case can exceed the total original number of the variables. Here, that number is n for the u^i plus n^2 for the v^i_j, proving our theorem.

It is easy to see that *the maximum number of parameters for the group may actually be attained*. The simplest example is of $\ddot{x}^i = 0$, the paths being straight lines. The group is then that of the translations with n parameters, plus the linear transformations leaving the origin invariant, of order n^2. In the Riemannian case, for example, as with the equations of Killing, something more is demanded, namely the invariance of a quadratic form as well, whence the maximum order is half the above. For the path-space of straight lines, if we impose, say, a Euclidean metric, the group is then translations plus rotations, order $n + n(n-1)/2 = n(n+1)/2$.

4. Further extensions of the previous results are possible in several directions, e.g.:

The group whose generating vectors satisfy only $u^i_{;j} = 0$, $\theta u^i = 0$, *thus transforming into itself the space* (x, t), *while leaving the absolute parameter t unchanged, has a number of parameters* $\leq n(n+2)$.

The point transformations corresponding to this are Cartan's group B, and their tensor invariants, can be found now in an obvious way. The proof parallels the above step by step.

These processes can be carried out also for systems of ordinary differential equations of higher order, as well for partial differential equations, the sole condition being that they be explicitly soluble for the derivatives of highest total order. The d/dt operator for ordinary differential equations has to be again defined as total differentiation along the paths, whereas for partial differentiation we have as many such operators as there are independent variables. We can then prove as before:

Theorem 5 *The results of Theorems 1–3 are valid also for the systems*

$$\frac{d^{\sigma+1}x^i}{dt^{\sigma+1}} + \alpha^i\left(t, x, \frac{dx^i}{dt}, \ldots, \frac{d^{\sigma}x^i}{dt^{\sigma}}\right) = 0, \tag{20.10}$$

the equations of variation being defined by all those (extended) operators which permute with d/dt, itself defined as:

$$\frac{d}{dt} \equiv \frac{\partial}{\partial t} + \frac{dx^r}{dt}\frac{\partial}{\partial x^r} + \cdots + \frac{d^\sigma x^r}{dt^\sigma}\frac{\partial}{\partial x^{(\sigma-1)^r}} - \alpha^r \frac{\partial}{\partial x^{(\sigma)r}}\,. \tag{20.11}$$

The maximum number of parameters in the group leaving base-space as well as paths invariant cannot then exceed the sum of the first $(\sigma + 1)$ *terms of the series*

$$n\left\{1 + n + \frac{n(n+1)}{2} + \frac{n(n+1)(n+2)}{3} + \cdots\right\}. \tag{20.12}$$

For partial differential equations, we consider only the second-order system:

$$\frac{\partial^2 x^i}{\partial t^2 \partial t^\beta} + H^i_{\alpha\beta}(t, x, p^j_\gamma) = 0\,; \quad \begin{aligned} \alpha, \beta &= 1, \ldots m\,, \\ i, j, \ldots &= 1, \ldots n\,, \\ p^i_\alpha &= \frac{\partial x^i}{\partial t^\alpha}\,. \end{aligned} \tag{20.13}$$

Here, the operators corresponding to d/dt are the set ∂_α defined by

$$\partial_\alpha \equiv \frac{\partial}{\partial t^\alpha} + p^r_\alpha \frac{\partial}{\partial x^r} - H^r_{\alpha\beta}\frac{\partial}{\partial p^r_\beta}\,. \tag{20.14}$$

Theorem 6 *The equations of variation have as solutions those vectors and only those whose associated (extended) operators form the ring that permutes with all operators* ∂_α*. The maximum number of parameters for the group leaving base-space and paths invariant cannot exceed* $n(n + 1)$*, as before.*

The proof is by following the case of ordinary differential equations step by step, and the condition for composition of two operators as well as the Jacobi condition may be derived by direct calculation. The number of parameters is again from considerations of the variation vector u^i being a function of the same number of variables x, as is seen directly from the structure of the equations of variation. However, in this case, there are also equations of variation for the independent variables t^ρ, and it should be made clear that the base-space is of the variables x^i.

For Eqs. (20.1) which are deducible from a metric, i.e., the extremals of a regular problem of the calculus of variations, we have the following formula. If the (inverse) Eulerian equations be abbreviated by $\delta_i f = 0$, then the result of any operation X of the base-space group is:

$$\delta_i(Xf) = X(\delta_i f) + u^r_{,i}\delta_r f + f_{;i;r}\theta u^r\,. \tag{20.15}$$

This shows what would otherwise have been expected:

Theorem 7 *If a metric exists for the paths, it is carried into metric by any transformation of the group preserving base space and paths.*

For the simple case $\ddot{x}^i = 0$, a general metric is any arbitrary function $f(x)$ with the determinant $|f_{;i;j}| \neq 0$, and not containing the x^i at all. This is carried into another of the type by any linear homogeneous transformation, and into itself by any translation. The only possible additive terms are necessarily of type dh/dt, where $h(x, x)$ is homogeneous of degree zero in $\dot{x}$, as can easily be shown.

It is to be noted, in conclusion, that the conditions of Theorem 1 are not necessary. For the alternant of two extended operators associated with vectors u, v to give an extended operator, the precise condition is

$$v^i_{;r}\theta u^r - u^i_{;r}\theta v^r = 0\,. \tag{20.16}$$

Similarly, the Jacobi condition for three such operators is satisfied if and only if

$$u^i_{;r}\{v^r_{;k}\theta w^k - w^r_{;k}\theta v^k\} + (\text{two more terms by cyclic rotation}) = 0\,. \tag{20.17}$$

This shows, in particular, that it even suffices to have one of the two vectors a solution of both $\theta u^i = 0$ and $u^i_{;j} = 0$; in the Jacobi condition, it is again sufficient for one of the three to satisfy both these equations. Thus, the infinitesimal transformations not containing $\dot{x}^i$, and in particular the subgroup leaving both base-space and paths invariant, have a special position.

The second remark is about the possibility of defining a Lie differential operator that carries tensors into others of the same type, but of defining it in a manner that can be carried over to more general classes of transformations, such as those over the entire path-space. To this end, we may note that any such transformation may be regarded either as a change in variables, or a charge of coordinates. *The (infinitesimal) Lie operation gives the difference of the (infinitesimal) changes in any geometric object due first to regarding the transformation as a change of variables, and then as a change of coördinates*. Thus, for the tensor T^i_j of weight p under transformations preserving the base-space, we have, when $u^i_{;j} = 0$,

$$LT^i_j \equiv T^i_{j,r}u^r + T^i_{j;r}u^r - T^r_j\dot{u}^i_{,r} + T^i_r u^r_{,j} + pT^i_j u^r_{,r}\,. \tag{20.18}$$

Moreover, there is the infinite series expansion as in (20.5), and we have the tensor carried over into another of the same sort, independently of any connection that may be assigned to the x-space. The main definition is obviously extensible for more general transformations.

Chapter 21
The Method of Least-Squares

D.D. Kosambi, Poona

The English version of this paper appeared two years after the Chinese "original." During the 1950s and early 1960s, DDK visited China several times on exchange programs. This paper was probably written when he visited the Academia Sinica on an exchange program between India and China as an expert in statistics from TIFR [DDK-JK]. *This was a visit of several months, ample time for DDK to write his paper and have it translated into Chinese.*

This note begins with a discussion of possible metrics in probability spaces associated with independent random variables; the Euclidean metric (in suitable coordinates) turns out to be the only one admissible. The method of least squares is known to be derived from such a concept of distance. In the second section, a unique least-squares solution is derived for general linear systems of equations in abstract spaces even when there may be no proper solution in the usual sense, the two coinciding when the ordinary solution exists. This is of considerable importance for diffusion theory and the integral equations for atomic energy piles. The final section gives a sketch of the extension to general nonlinear systems of equations.

1. We start with a system of measurable sets called "simple events" such that the adjunction of the "compound events" obtained by set addition and set multiplication gives an aggregate of measurable Borel sets constituting a Boolean set algebra. The union $A \cup B$ of two sets is the compound event "A or B"; the intersection $A \cap B$ is the compound event "A and B (simultaneously)"; the operational laws for the dual operations "cap" $= \cap$ and "cup" $= \cup$ being as usual in Boolean algebra, which

Published in the Journal of the Indian Society of Agricultural Statistics **11**, 49–57 (1959). The Chinese version appeared in Advancement in Mathematics **3**, 485–491 (1957). Reprinted with permission.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_21

contains the null set O and the universe I. The probability measure is regulated by the postulates [1]:

$$P(I) = 1. \tag{21.1a}$$

$$P(A) \geq P(B) \text{ if } A \cup B = A, \ i.e., \text{ if } A \supset B. \tag{21.1b}$$

$$\text{If } A \cap B = O, \ P(A \cup B) = P(A) + P(B). \tag{21.1c}$$

Taking $A = I$, $B = O$ in Eq. (21.1c), it follows that $P(O) = O$. With (21.1a) and (21.1b), this gives $O \leq P(A) \leq 1$ for all sets of the ensemble. Finally, seeing that $A \cup B$ is the union of three mutually non-intersecting sets $(A - A \cap B)$, $A \cap B$, $(B - A \cap B)$, we obtain the general result:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)\,. \tag{21.2}$$

This could have been substituted for the third postulate in (21.1a, 21.1b and 21.1c) with the addition of $P(O) = O$. The events are to be regarded as reduced modulo, the ideal of all sets with measure zero. The restriction to Borel sets, though not always necessary, permits infinite repetition of the two operations ∪, ∩.

Definition *Two events A, B such that $A \cap B = O$ are called mutually exclusive. Non-null events $A_1, A_2, \ldots A_n \ldots$ such that $P(A_i \cap A_j \cap A_k \ldots) = P(A_i)P(A_j)$ $P(A_k) \ldots$ for any finite section $i, j, k \ldots$ are called mutually independent events.*

It follows that two mutually exclusive events cannot be mutually independent, nor can two events one of which wholly includes the other; these are the extreme case of zero and unit conditional probability, always omitting from the classification the trivial extremes, O, I. Starting with any simple event of the algebra, we can build an ordered maximal chain of such simple events, with O and I at the two ends, each event of the chain including all preceding members and being included in all that follow, while no other simple event of the algebra outside the chain has this property, with respect to all sets of the particular chain. We consider hereafter *only such Boolean probability algebras whose simple events can be split up into a finite number of maximal chains, every event of each chain being independent of every event in any other chain.*

In the first place, each such chain can be mapped upon the real line segment (0, 1) by the correspondence $A \to [O, P(A)]$. But we need also a map on the whole real axis $-\infty \leq x \leq +\infty$, which is connected with the (0, 1) measure map by a distribution function $F(x)$, which is monotonically non-decreasing, with $F(-\infty) = 0$, $F(+\infty) = 1$. Any set A of the chain can be mapped upon the interval $(-\infty, \alpha)$ on the line such that $F(\alpha) < P(A)$ if $x < \alpha$, while $F(a) = P(A)$. Using one dimension for each such ordered chain, we map the Boolean algebra upon an n-dimensional continuum $(x_1, x_2 \ldots x_n)$, where the image of a simple event is a section from $-\infty$ to $+\infty$ in all dimensions except one, where the section extends only from $-\infty$ to α. The measure image on the unit hypercube is the rectangular parallelepiped of side unity in all except one dimension, where the side is the interval $[O, P(A)]$. Compound events are derived from these by set union and set intersection.

Theorem 1 *If an n-dimensional probability space be associated with a Boolean algebra of events such that each dimension represents a chain of events independent of all the others, and if the space is endowed with a Riemann metric plus a measure function which give a true map upon the unit hypercube, then the metric can only be Euclidean.*

Proof For the Riemann metric, $ds^2 = \Sigma g_{ij} dx_i dx_j$. The measure of any k-dimensional event in the x-space is given, for $1 \leq k \leq n$ by an integral of the form $\int f_k(x_{i_1}, \ldots, x_{i_k})\sqrt{|g_{ij}|}\, dx_{i_1} \ldots dx_{i_k}$. But if the region be the compound event $A_1 \cap A_2 \cap \cdots \cap A_k$, it follows that the integral must break up into a product of k separate integrals for all $k \leq n$. Therefore, any principal minor as well as the whole determinant $|g_{ij}|$ must reduce to a product of diagonal terms: $g_{11}(x_1)$ $g_{22}(x_2) \cdots g_{nn}(x_n)$, and correspondingly for each of its principal minors. The measure function f, essentially the derivative of the distribution, assumed to exist and be continuous, will similarly break up into a product of factors, but that is of lesser interest here. It is clear that the cross terms of the tensor g_{ij} all vanish, with $ds^2 = g_{11}(x_1)dx_1^2 + g_{22}(x_2)dx_2^2 + \cdots + g_{nn}(x_n)dx_n^2$. The g_{ii} are positive from the hypothesis of positive measure for any chain (we need not invoke the positive definition form of the metric here), permitting a transformation of coordinate variables defined by $dx'_r = \sqrt{g_{rr}}\, dx_r$. These are the Euclidean coordinates of the space. □

We have two simple corollaries:—

Corollary 1 *If the space of n random variables be endowed with a Riemann metric and a measure (distribution density) function which permit the original random variables to be replaced by n independent random functions thereof, then the curvature tensor of the original space must vanish, the space being Euclidean.*

The new variables amount simple to a non-singular transformation of coordinates. But there, the space will have the Euclidean coordinates of the preceding theorem; hence its curvature tensor will vanish in both coordinate systems. For the second corollary, we need a topological result, [2] that a Riemann metric exists when the space may be covered by neighborhoods such that each pair of points may be joined by one and only one arc (lying wholly within the neighborhood) of a previously defined class which we may call paths. Then, for any compact portion of the space that can be so covered, a Riemann metric can be assigned whereof the given paths are actually the geodesics. In the present case, we have one compact space, namely the unit hypercube, to which the result may be applied, working back to the original space if the f function is continuous, giving us:

Corollary 2 *If the space of random variables is only endowed with a continuous measure density function, and a set of continuous paths with the property that any two points sufficiently close have a unique path join, then the space also possesses a Riemann metric, hence is Euclidean if the concept of independent random variables is applicable by suitable transformation.*

With Euclidean space, if the compound probability density function of several independent random variables depends only upon the distance, it follows immediately that the distribution of each variate must be normal (Gaussian) [3]. From this to the usual motivation of least squares is only one step, for the best approximation to the population mean from a sample is that which minimizes the sampling variance, which is a sum of squares (the distance, in fact, to a hyperplane), hence the arithmetic mean.

2. We deal throughout with real variables, though the extension to complex or other number systems causes little difficulty. The system of m linear equations in $n < m$ real variables

$$\sum_{j=1}^{n} A_{ij}x_j - y_i = 0\,; \quad i = 1, 2, \ldots, m > n \tag{21.3}$$

has no solution in general. But it has always a least-squares solution minimizing

$$\sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}x_j - y_i \right)^2 , \tag{21.4}$$

thereby, specifying the values of x as solutions of the n equations:

$$\sum_{r=1}^{n} C_{kr}x_r - z_k = 0\,, \quad C_{kr} = \sum_{q=1}^{m} A_{qk}A_{qr}\,, \quad z_k = \sum_{q=1}^{m} A_{qk}y_q\,. \tag{21.5}$$

Here, every free index runs through the values $1, 2, \ldots, n$. The system (21.5) has a unique solution in general, coinciding with the exact solution of (21.3) should those equations be compatible. Clearly, we can take formal passage to the limit to an integral equation of the first kind:

$$\int A(s,t)x(t)dt = y(s)\,, \tag{21.6}$$

and to other general linear systems. This is the work of the present section, regardless of probability considerations.

We begin with a vector space V over the field C of all real numbers, the elements $x, y, \cdots$ being in V and constants $a, b \cdots$ in C give $ax + by + \cdots$ also in V. We further require a symmetric bilinear scalar product $x \cdot y$ as a mapping of $V \times V$ into C, with the properties: $x \cdot y = y \cdot x$, and $x \cdot (ay + bz) = a(x \cdot y) + b(x \cdot z)$. This leads to a quadratic *norm* $x \cdot x$ of which we demand that $x \cdot x = 0$ if and only if $x = 0$, which amounts to reduction of V with respect to elements of zero norm. We shall assume that V is complete with respect to convergence in the norm. The usual condition that the norm be positive is easily imposed, for it must always be of the same sign. If there were two distinct elements x, y with $x \cdot x > 0$, $y \cdot y < 0$, the quadratic in $\lambda : (x + \lambda y) \cdot (x + \lambda y) = 0$ would have real roots, giving an element with vanishing norm, of the form $x + \lambda y$. But this cannot be zero identically, for then $x \cdot x = \lambda^2(y \cdot y)$, which is impossible because the two norms had initially opposite signs. Hence, the norm must always have the same sign, and there is no loss of generality in taking it always positive.

We avoid the trivial cases where V contains only the element 0, or only multiplies of a single element ϕ. Two nonzero elements ϕ, ψ are defined as orthogonal if their scalar product vanishes: $\phi \cdot \psi = 0$, while an element with unit norm (always to be had by multiplication with a suitable constant) is called normal. The assumption is that V has an orthonormal basis $\phi_1, \phi_2, \ldots \phi_n, \ldots$ not necessarily finite, but (by the Hilbert theorem) at most denumerable, and that the Riesz–Fisher theorem applies so that with any convergent Σa_r^2, there always exists a function in V represented by $\Sigma a_n \phi_n$; this is necessary for the completeness of the space, which we have assumed.

To correspond to the matrices in (21.3), we need two-sided linear associative operators $S, T, \ldots$ defined over V, i.e., Tx and $xT \subset V$ for all $x \subset V$; with $(ax + by)T = a(xT) + b(yT)$, $T(ax + by) = a(Tx) + b(Ty)$. For xT, we shall also write T^*x, the *adjoint* of T. This adjoint is governed by the operational rule: $(T^*)^* = T$. If we define the operator product ST by $(ST)x = S(Tx)$, with $x(ST) = (xS)T$, it follows that $STx = S(xT^*) = (xT^*)S^*$, whence $(ST)^* = T^*S^*$, the star operation for the adjoint of these linear operators thus satisfying four of the basic postulates for a C^* algebra in the sense of Gelfand and Neimark. We may write SxT for $S(xT) = ST^*x = xTS^* = T^*xS^*$, according to convenience, without confusion. The scalar product $x \cdot (Ty)$ is similarly abbreviated $xTy = yT^*x$, at will.

Using the orthonormal basis for V, it is seen that the T operation amounts to a linear matrix transformation for the coordinates (Fourier coefficients) of an element. All operations may be visualized and theorems proved by use of the matrix representation. For Hilbert spaces (vector spaces with infinite basis), the argument has to be restricted in general to such operators as may be separated into two additive portions of which one is finite dimensional, the other with arbitrarily small norm. That is, the operators must be *bounded*: $(Tx) \cdot (Tx) \leq M(x \cdot x)$ for all $x \subset V$, M depending only upon T. We shall deal only with non-singular bounded operators, and remark that a symmetric operator such that $T = T^*$ has always a real spectrum. To each T, there correspond always the two symmetric operators TT^* and T^*T, of which the latter is assumed to have a discrete spectrum for our main result.

The entire least-squares procedure rests upon the following [4]

Lemma *The orthonormal portion of V which does not lie in TV is mapped into zero by T^* that is, $T^*(V - TV) = 0$.*

Proof If the transformed space TV is the whole of V, the result is trivially true. If not call $\bar{V}$, the orthonormal component of V not in TV. The generic scalar product of a function in $\bar{V}$ and another in V is $\bar{V}TV = VT^*\bar{V}$. By hypothesis, this scalar product is zero, which means that every element in the whole of V is orthogonal to every element in $T^*\bar{V}$, which is impossible, except when $T^*\bar{V} = 0$, proving the lemma. □

The main least-squares equation now takes the form

$$Tx - y = 0 \tag{21.7}$$

with T, y given, x to be found.

This need not have a solution at all, as Tx lies necessarily in TV, while the given y may have a component outside TV. The norm of the left-hand side is

$$(Tx - y) \cdot (Tx - y) \equiv xT^*Tx - 2(xT^*y) + (y \cdot y)\,. \tag{21.8}$$

To minimize this, give a variation to x, replacing x by $x + \delta x$. Subtracting the original value in (21.8) from the varied value gives

$$2\delta x \cdot (T^*Tx - T^*y) + (T\delta x) \cdot (T\delta x)\,. \tag{21.9}$$

The coefficient of δx is equated to zero for a minimum as usual, for the remainder is positive, while we take the norm of δx as tending to zero. This gives us the following:

Theorem 2 *The least-squares solution of $Tx - y = 0$ is given by*

$$T^*Tx - T^*y = 0\,. \tag{21.10}$$

Our lemma $T^*(V - TV) = 0$ makes the solution possible. Naturally, there are some simple restrictions upon the operator in question. In terms of eigenfunctions and eigenvalues, these give Picard's solution [5] of integral equations of the first kind and the corresponding least-squares solution, which may be subsumed in the following:

Theorem 3 *The least-squares solution of $Tx - y = 0$ exists if and only if $\Sigma(\phi_n T^* y)^2/\lambda_n^4$ converges, where ϕ_n and λ_n^2 are the eigenfunctions and eigenvalues respectively of $T^*T\phi - \lambda^2\phi = 0$. In particular, if the orthonormal set $\{\phi_n\}$ furnish a basis for V, we have the exact solution (for that portion of V in which y lies).*

The proof is as follows: The non-singular operator T^*T leaves the origin invariant in V, hence by continuity maps some portion of V on to some neighborhood of O, in the map space T^*V. We (assume the operator T^*T to have a discrete spectrum, and) expand T^*y in terms of the eigenfunctions. The lemma above says that T^*y cannot be orthogonal to all these eigenfunctions without vanishing identically, while the condition of the theorem merely requires T^*y to lie in the transformed neighborhood of the origin.

The result is independent of the norm. That is, our norm was best taken with respect to the identity, the symmetric operators $xI = Ix = x$ for all $x \subset V$. Any other symmetric operators may be used for the least-squares norming provided $SV = V$, and $xSx = 0$ if and only if $x = 0$. The result is of great use in the solution of integral equations when nothing is known about the closure of the eigenfunctions of the particular kernel.

3. The square sum (21.4) of the linear equation (21.3) amounts to the sum of weighted squares of distances from a generic point (x) to the various hyperplanes. The same idea can be extended, therefore, to nonlinear hypersurfaces. We look for the point or points from which the sum of squares of distances to a given set of (weighted) hypersurfaces is minimum, which is included in the set of points where the distance sum is stationary and which is all we shall investigate without insisting upon a true minimum. The geometric picture tells us that the point sought is common to all the surfaces if they have a common interesection, or that point which lies on the intersection of normals to each of the surfaces. In mathematical notation, let the surfaces be:

$$f_1(x_1, x_2, \ldots, x_n) = c_1,\ f_2(x) = c_2, \ldots f_m(x) = c_m\,; \quad m > n\,. \tag{21.11}$$

The point sought is the solution of the equations:

$$\frac{\partial F}{\partial x} = 0\,, \quad \frac{\partial F}{\partial u} = 0\,, \quad \frac{\partial f}{\partial v} = 0\,, \tag{21.12}$$

where

$$F \equiv \sum_i (x_i - u_i)^2 + (x_i - v_i)^2 + \cdots + \lambda_1 f_1(u) + \lambda_2 f_2(v) + \cdots$$

and the unindexed letters $x, u, v \ldots$ each represent the set of n variables, the index being understood, even in the partial differentiation. This is Lagrange's method of multipliers leading to two sets of equations:

$$x_i = \frac{u_i + v_i + \cdots}{m} \tag{21.13a}$$

$$2(x_i - u_i) - \frac{\lambda_1 \partial f_1}{\partial u_i} = 0\,,\ 2(x_i - v_i) - \frac{\lambda_2 \partial f_2}{\partial v_i} = 0\,, \ldots \tag{21.13b}$$

These lead to compatibility conditions:

$$\frac{\lambda_1 \partial f_1}{\partial u_i} + \frac{\lambda_2 \partial f_2}{\partial v_i} + \cdots = 0\,; \quad i = 1, 2, \ldots, n\,, \tag{21.14}$$

which merely reflects our previous lemma $T^*(V - TV) = 0$ in the total extended space. For linear equations, the process is as before, and for the general case, the extension is fairly clear.

We begin with an abstract vector space V such that $x \subset V$. This V is extended over a variety which was formerly a finite Abelian group and may for the present be taken as an indexed variety. The extension is then indicated by V_α, with variables $u_\alpha \subset V_\alpha$. The α-space has to be compact, with an abstract integral which we shall denote by Σ_α and which has the properties of a Lebesgue–Stieltjes integral, while $\Sigma_\alpha 1 = M$. The scalar product is defined over the extended space as $\Sigma_\alpha (x - u_\alpha) \cdot (x - u_\alpha)$. Finally,

$f(x)$ is a general operator, mapping x into the real field, $f_\alpha(u)$ being the (suitably but completely defined) extended operation, understood as $f_\alpha(u_\alpha)$.

We need further the generalized partial differentiation, which is defined as the infinitesimal operator of the (Abelian) Lie group in the space of $f(x)$ when the base space of x undergoes a translation $x \to x + h$; the Lie group is generated by the usual exponential representation, which leads to a Lie-Taylor series expansion which is the formal representation, and in the analytic case converges to give an exact representation. Our nonlinear functional operators f need not be analytic nor even arbitrarily differentiable, for they may be approximated by such at need; but the f-operators must at least be continuous in the first derivative for the analogue of (21.12) to be valid. By introducing an orthonormal basis and coordinates for V, the partial derivative becomes just the ordinary partial derivative in the coordinates; generically, we represent this by f'. The results are then summed up as follows:

The least-squares solutions of the simultaneous projections $f_\alpha(x) = 0$ are given by $x = (1/M)\Sigma_\alpha u_\alpha$, provided the extended variables u_α satisfy

$$\lambda_\alpha f'_\alpha(u_\alpha) = \left(\frac{1}{M}\right) \Sigma_\alpha u_\alpha - u_\alpha \,. \tag{21.15}$$

The λ_α and u_α being so chosen as to further satisfy $f_\alpha(u_\alpha) = 0$.

References

1. A. Kolmogorov, Grundbegriffe der Wahrscheinlichkeitsrechnung, Ergebnisse d. Mathematik, **2**(3), (1933). (Berlin, the opening sections).
2. D.D. Kosambi, The metric in path-space. Tensor **3**, 67–74 (1954).
3. D.D. Kosambi, The geometric method in mathematical statistics. Am. Math. Mon. **51**, 382–389 (1944).
4. L.H. Loomis, in *Introduction to Abstract Harmonic Analysis* (New York, 1953), p. 27. (for the classical result).
5. R. Courant, D. Hilbert, in *Methoden der mathematischen Physik I*, (Berlin, 1931), pp. 134–36. (E. Picard, in *Rendiconti del Circolo Mathematico di Palermo* **29**, 79–97 (1910)).
6. D.D. Kosambi, An extension of the least-squares method for statistical estimation. Ann. Eugen. **13**, 257–261 (1947).

Chapter 22
An Application of Stochastic Convergence

D.D. Kosambi, Poona

Although DDK was still at the TIFR, Bombay, he chose to publish this paper giving only the Poona address as he did for the other papers published in JISAS. This was the first of the infamous Riemann Hypothesis papers and was reviewed in Mathematical Reviews by the number theorist W.J. LeVeque who was extremely critical of the work: 'The reviewer is unable either to accept this proof or to refute it conclusively. The author must replace verbal descriptions, qualitative comparisons and intuition by precise definitions, equations and inequalities, and rigorous reasoning, if he is to claim to have proved a theorem of the magnitude of the Riemann hypothesis.' Two errors are pointed out in the review, and this does not include an elementary error noted by Berry (private communication) that in Eq. (3.3) all the denominators should be unity.

The approach itself, however, has some points of merit. Odlyzko (private communication) notes that although [Kosambi's] method is highly probabilistic, if it went through, it would provide a proof of the RH in its full strength. If indeed the series (22.1) *converged for every* $\sigma > 1/2$, $\zeta'(s)/\zeta(s)$ *would be analytic in* $\sigma > 1/2$, *aside from s = 1, and we would have the RH. Kosambi's 'proof' is that in the space of his rearrangements of series, this follows with a positive probability, and he really only needs a single rearrangement.*

Published in Journal of the Indian Society of Agricultural Statistics **11**, 58–72 (1959).

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_22

The basic result of this paper, from which the conclusions of Section 3 follow, is the proof that the series

$$\sum_{n\le x}\frac{1}{n^\sigma}-\sum_{p\le x}\frac{\log p}{p^\sigma}\,,\quad x\to\infty\,, \tag{22.1}$$

converges for every real $\sigma > \frac{1}{2}$. Here, n runs through the positive integers and p the primes, in natural order. The convergence of (22.1) for $\sigma \geq 1 + 0$ is known, but not for $\frac{1}{2} < \sigma < 1$. The classical tools for dealing with such convergence problems [1] are inadequate.

The special device employed in this note resembles to a certain extent the use of Lebesgue integration when the integrand oscillates so much that the evaluation of a Riemann integral does not seem feasible. Only, in place of Lebesgue outer measure, we use probability measure. This is more convenient because of the powerful results in probability theory now available.

The following method is adopted for the proof. Terms of the series (22.1) are grouped together for n and p in irregular, non-overlapping, consecutive intervals $\boldsymbol{d}_\nu$ of length d_ν, with a uniquely defined, not necessarily integral, real number x_ν in $\boldsymbol{d}_\nu$. The original series (22.1) is then replaced by

$$\sum_\nu\left\{\frac{d_\nu}{\log x_\nu}-\pi(\boldsymbol{d}_\nu)\right\}\cdot\frac{\log x_\nu}{x_\nu^\sigma}\,,\ \sigma>\frac{1}{2}\,, \tag{22.2}$$

where $\pi(\boldsymbol{d}_\nu)$ is the number of primes in $\boldsymbol{d}_\nu$. Differences between the partial sums of (22.1) and (22.2) are clearly expressible as the sum of terms

$$\left(\frac{d_\nu}{x_\nu^\sigma}-\sum\frac{1}{n^\sigma}\right)-\left(\pi(\boldsymbol{d}_\nu)\,\frac{\log x_\nu}{x_\nu^\sigma}-\sum\frac{\log p}{p^\sigma}\right)\;;\quad n,p\subset(\boldsymbol{d}_\nu)\,. \tag{22.3}$$

Therefore, they may be dissected into components due to the grouping; to the substitution of x_ν for n and p; and those from the partial sums of (22.1) and (22.2) not terminating at the same term.

If a set of covering intervals $\{\boldsymbol{d}_\nu\}$ exists such that (22.2) and the various series and sequences (finite in number) arising from the above differences all converge simultaneously, then clearly the series (22.1) converges.

In what follows, $[\boldsymbol{d}]$ indicates the number of integers in the interval $\boldsymbol{d}$. The prime number theorem is taken for granted in the form: $\pi(0,x)\sim li(x)\sim x/\log x$. P denotes a probability, E the expectation (mean), and V the variance (dispersion) in the sense of probability theory. For stochastic X and scalar λ, we always have $E(\lambda X) = \lambda E(X)$ and $V(\lambda X) = \lambda^2 V(X)$. By a *variate* is meant a stochastic variable, i.e., one that has a probability distribution. The following result due to A. Kolmogoroff [2] is fundamental:

Lemma *K. The stochastic series Σu_n of independent variates $\{u_n\}$ converges with $P = 1$ if there exists another set of independent variates v_n such that the series*

$\Sigma P(u_n \neq v_n)$, $\Sigma E(v_n)$, *and* $\Sigma V(v_n)$ *all converge; otherwise, the convergence probability of* Σu_n *is zero.*

The use of this theorem in the sequel does *not* mean that (22.1) converges with unit probability, for (22.1) is not a stochastic series. The utility of Lemma K lies in showing the existence of a suitable choice of $\boldsymbol{d}$-intervals. That is, a stochastic mechanism of selection may be set up for $\boldsymbol{d}_\nu$, $\nu = 1, 2, \ldots$ so that (22.2) and the series and sequences of its differences with (22.1) all converge, with a positive compound total probability. Therefore, *at least one infinite sequence of covering intervals* $\{\boldsymbol{d}_\nu\}$ *must exist giving simultaneous convergence of all these, and* hence, the series (22.1) converges. The existence theorem need not actually construct a specific set $\{\boldsymbol{d}_\nu\}$, but the logic involved is completely rigorous, having as its basis the fact that a set of positive measure cannot be empty. The proof is not heuristic, as it would have been had the prime numbers been treated as a stochastic sequence because of their irregularity.

In series (22.2) and some of the associated series and sequences, the occurrence of x_ν makes the terms dependent in probability. This is circumvented by setting up comparison series where the terms are independent, and to which lemma K applies. Similarly, $\pi(\boldsymbol{d}_\nu)$ enters into some of the auxiliary series; there, the probabilities required may be assessed by a change of measure in sample space. Thereby, the erratic behavior of the primes in the natural sequence of positive integers, which spoils other proofs, is turned into an asset.

The use of probability methods is easily motivated. If the primes were regularly spaced, $\log p$ apart (22.1), would converge for $\sigma > 0$. On the other hand, suppose that $\pi(n) = li(n)$ exactly, whenever $n = k^a$, k an integer and $\alpha \geq 3$. Further, let these 'pseudo-primes' cluster together at the left-hand end of $k^\alpha \leq n < (k+1)^\alpha$, leaving the rest of the interval void of primes. Then, the corresponding series (22.1) clearly diverges for some $\sigma > \frac{1}{2}$. For the actual primes, the gaps would be much too large even when $a = 2$ (which does not give divergence); it is known [3] that for almost all x and h of order $x^{1/4}$, $\pi(x+h) - \pi(x) \sim h/\log h$. Thus, known facts about primes suffice to exclude regular arrangements that would make (22.1) diverge. The question is really settled by showing that the prime numbers, in suitably defined intervals behave like an *unbiased random sample* from a non-singular probability distribution (or like a von Mises *Kollektiv*). That is, the relative frequencies of intervals containing $0, 1, 2, \ldots$ primes each tend to definite limits as the real line are progressively covered. This is shown in Lemma 1.2, which should be the most useful result of this note.

The problem is *to discover an underlying stochastic population* from which an irregular infinite sequence, specified by a procedure, not by formula, might be drawn as an *unbiased* random sample. The answer can be obtained only in the sense of unit probability. For the sequence of primes considered directly, the question of bias would still remain, i.e., whether they do not form part of the exceptional set of zero probability measure. We consider instead an infinite set of complete coverings defined by choice of the initial point, the situation repeating itself when the initial point moves thorough a single covering interval. Unit probability would mean 'for almost all initial points.' This validates applications of the basic Poisson distribution, which is even more important and useful than the Riemann Hypothesis.

1. This section deals with the mechanism of choice for $\boldsymbol{d}$. Textbook [4] results in probability theory are taken for granted. The real half line $2 \leq x \leq \infty$, on which the integers and primes are marked off, is transformed into $2 \leq y \leq \infty$ by $y = li(x) + c = \int dx/\log x$. Then, to an interval $(a, a + \Delta)$ on the y-line corresponds a unique interval $\boldsymbol{d}$ on the x-line and conversely, with $\Delta = d/\log x$; where x is chosen as that number (not necessarily an integer) lying in $\boldsymbol{d}$, which makes this relationship hold. The mean value theorem for the integral of a monotonic function shows the existence of such an x, which lies properly within $\boldsymbol{d}$. The intervals may be taken to include the left-hand end point, but not the right. An arbitrarily large initial portion of either line may be ignored in discussion of the convergence problem.

Each length $\Delta_\nu, \nu = 1, 2, \ldots$ is taken to have the identical distribution, namely the uniform distribution over $(0, 2)$. Being open on the right, marking off consecutive intervals of the lengths Δ_ν, without gaps, furnishes a complete non-overlapping covering of the y-line. That is, the length is a stochastic variable equivalent to the ν-th independent selection from the uniform distribution; the position of the interval of length Δ_ν is uniquely determined by the particular sample. Hence, the $\boldsymbol{d}$-intervals that correspond by the inverse $li(x)$ transformation give a stochastic covering of the x-line, one complete covering for each such infinite random sample of the Δ's. The number $x_\nu \subset \boldsymbol{d}_\nu$ has been specified above. For the lengths Δ_ν, we have for every ν and any positive integer k:

$$E(\Delta) = 1\,; \quad V(\Delta) = \frac{1}{3}\,; \quad E(\overline{\Delta - 1}^{2k}) = \frac{1}{(2k+1)}\,; \qquad E(\overline{\Delta - 1}^{2k+1} = 0\,. \tag{22.4}$$

The variate y_ν is defined as the sum of the first ν independent, consecutive, non-overlapping Δ-intervals: $y_\nu = \Delta_1 + \Delta_2 + \ldots + \Delta_\nu$; this has the range $(0, 2\nu)$, mean ν, and variance $\nu/3$. Its probability curve is convex upwards, with a single maximum. According to the central limit theorem, the probability distribution of y_ν is approximated efficiently by a normal (Gaussian) distribution with mean ν and standard deviation $\sqrt{\nu}/3$. It follows that the maximum height of the y_ν probability curve is rapidly asymptotic to $1/\sqrt{2\pi\nu/3} = a/\sqrt{\nu}$ and the distribution may be taken as approximately uniform over steps of order less than $\sqrt{\nu}$ in width. The estimates of S. Bernstein [4] may be applied to (22.4) to give:

Lemma 1.1 *The probability is less than* $\exp(-t^2)$ *for each of the inequalities to hold (separately) for all large* ν*:*

$$y_\nu > \nu + t \cdot \sqrt{\frac{2\nu}{3}} \quad \textit{and} \quad y_\nu < \nu - t \cdot \sqrt{\frac{2\nu}{3}}\,. \tag{22.5}$$

Two useful corollaries follow. Taking $t = \sqrt{(3/2)\log\nu}$*,* P *is less than* $\nu^{-3/2}$ *for each of the inequalities*

$$x_\nu > 2\nu \log \nu\,; \quad x_\nu < \frac{\nu}{2}\,, \tag{22.6}$$

for all large x and ν. Secondly, the ratio y_ν/ν converges in probability to unity.

Lemma 1.2 *$\pi(d)$ has a proper frequency distribution over almost all complete coverings, the expectation begin unity, and the variance finite. If consecutive intervals be grouped together k at a time, then the mean and the variance of primes covered are each multiplied by k.*

Proof The prime number theorem and Lemma 1.1 show that the limit as $r \to \infty$ of $\{\pi(\boldsymbol{d}_k) + \pi(\boldsymbol{d}_{k+1}) + \ldots + \pi(\boldsymbol{d}_{k+r})\}/r$ is unity, with unit probability, for every k. But this only gives the general expectation for almost all complete coverings. The limiting distribution may be singular if the primes occurred in maximal clumps separated by sufficiently many voids to restore the average. In the limit, the frequency of intervals with no primes could then be unity, all others zero—and yet no finite limiting variance need exist. Known results on gaps between successive primes (Prachar, [3], p. 154 *ff.*) make this singular case very unlikely, but we need only appeal to the principle of the sieve method. If all multiples of 2, 3, 5, ... are successively struck out, the smallest integer left at each stage is itself the prime to be used in the next deletion, and every prime is reached in this way. The survivors are thus asymptotic to $n(1 - 1/2)(1 - 1/3)\cdots(1 - 1/p)\cdots$ where the product must be suitably terminated. This says precisely that the (suitably bounded) primes act, each with its own probability $1/p$, *independently* (or the probabilities for survival would not be multiplied as above) of each other in the deletion. There is no linear (or even algebraic) relationship between the primes, and any two or more primes have the highest common factor one while $p_k \sim k \cdot \log k$. The theorem of de la Vallée Poussin says that for any arithmetic progression $ar + b$, $r = 1, 2, \ldots$ the primes are asymptotically equally divided between the $\phi(a)$ possible different categories, no matter what a is chosen. □

Therefore, the number of primes 'striking' an integer and the number of integers escaping the sieve ought each to have some sort of asymptotic frequency distribution. Of these, the first is given by Landau's theorem that the relative frequency of integers $< n$ having $k + 1$ prime factors is asymptotic to $e^{-t}t^k/k!$ for $k = 0, 1, 2, \ldots$, with $t = \log\log n$. This is a Poisson distribution, and the value of the parameter t would follow from the prime number theorem, if the distribution were granted.

For the prime survivors, we first take an interval $\boldsymbol{h}$ of y-image μ (fixed), hence of x-length approximately $\mu . \log n$. This is allowed to cover, with a uniform probability, the total range whose image of length $N(n)$ contains the integer n. Then, all deletions from $\boldsymbol{h}$ may be considered as due to primes not exceeding $\sqrt{n + N}$. Of these, the primes smaller than h will cause compulsory deletions, but those between $\mu . \log n$ and $\sqrt{n + N}$ will act as a matter of chance. The survivors of the compulsory deletions are $\approx e^{-\gamma}h/\log h$ (where γ is Euler's constant) which increases beyond any limit. The mean being μ, the probability of the survivors of the first deletion in the interval containing a prime will be $\approx \mu e^\gamma \log h/h$, which tends to zero. The introduction of probability methods is needed because not enough is known about the location of primes for direct calculation of their frequency distribution. This does not distort the actual distribution, particularly as regular arrangements have been disposed of in our preamble. The μ-intervals still belong to a covering and hence do not overlap. It suffices, therefore, to choose any interval at random after giving the initial point of the first μ-interval in the range a uniform distribution over one interval length. The number of primes neglected cannot exceed the maximum covered by a μ-interval of the N-range, which affects neither the distribution nor the convergence of (22.1).

Statements about the number of 'survivors' per interval have to be understood in the sense of unit probability. The arrangement of integers not divisible by $2, 3, 5, \ldots, p_r$ is repeated modulo $N_r = 2, 3, 5, \ldots, p_r$. The theorem of Mertens gives the proportion of numbers prime to N_r as $\sim e^{-\gamma}/\log p_r$. Only the very small primes with $N_r \leq h = \mu \log n$ can have a cyclic effect over an interval length. The less regular effect of the remaining small primes $\leq h$ is most economically described as an independent survival probability $\sim \log p_r/\log h$, because the initial point and length of the range are each of order $\exp(cp)$. By classical probability theorems, the chance of an interval having less than $ah/\log h$ survivors (with suitable $a > 0$) tends to zero. However, every one of the consecutive integers $kN_r + 2$, $kN_r + 3, \ldots, kN_r + P_{r+1} - 1$ has a factor in common with N_r. Inasmuch as N_r is of order n^{μ}, there could be (for small μ) intervals in a range devoid of survivors. These, or intervals with less than any fixed number of survivors, may be ignored in the limiting process as zero probability phenomena.

We have now to consider whether the chances of an integer being a prime or composite are affected by the knowledge that some other integer in the interval is actually a prime or composite. Should r be a prime > 3, then $r + 1$ and $r - 1$ are necessarily composite. Such obligatory dependence is removed by striking out the multiples of $2, 3, \ldots, p < \mu \log n$. Suppose that among the 'first survivors' one is known to be composite. Then its prime factors cannot, by construction, divide any other in the same μ-interval. The chance P' for primality among the rest may at worst have to be P (the original probability) multiplied by $1/(1 - 1/p)$ for every such prime factor. By Landau's theorem above, the average number of prime factors is of order $\log_2 n$; the maximum number of such factors can obviously not exceed $\log n/\log_2 n$. Thus, P would at most have to be multiplied by $e^{\phi(n)}$, where $\log_2 n < 1/\phi(n) < \log n$. The supply of primes which cause deletion, being of order $\sqrt{n/\log n}$, is not materially depleted, so that the argument may be repeated for further numbers found composite. On the other hand, if the known integer be a prime, the probability for the rest is not thereby affected in the same interval, for deletion is caused only by primes $< \sqrt{n}$, approximately. Thus any modification of P is an infinitesimal of higher order, which justifies passage to the limit on the basis of independence in probability among the 'first survivors' within the interval. This is also supported by known sieve theorems (Prachar, [3] Chap. II). Parallel arguments hold *a fortiori* for independence between intervals. As has been shown above, the number of these survivors tends to infinity with n, P tending reciprocally to zero, while the expectation is μ. It follows [5] that the limiting distribution is Poissonian provided there are an unboundedly increasing number of disjoint intervals in the range and $E\{\pi(h)\} \to \mu$ over the separate ranges as $n \to \infty$. These conditions are met by taking $N - n^{3/4}$, though smaller exponents will do as well [3]. The distribution therefore approximates rapidly to the frequencies $e^{-\mu}(1, \mu, \mu^2/2 \cdots)$ for $\pi(h) = 0, 1, 2, \ldots$. If, instead of a fixed length μ, we allowed a uniform distribution over $(0, \mu)$ for the length, simple integration would yield the asymptotic frequency for k primes as $(1 - e^{-\mu} s_k)/\mu$, where S_k is the sum of the first $k + 1$ terms in the Maclaurin expansion of e^{μ}. The mean is now $\mu/2$, and the variance becomes $(\mu/2 + \mu^2/12)$, the second term being the 'Sheppard's correction for grouping' familiar to statisticians.

For any finite number of consecutive N-ranges and fixed μ, the distribution is the weighted average of the component distributions over each range, with the number of μ-intervals in the range as weight. The whole line being thus progressively covered, and this amounts to summability by a *regular* Toeplitz matrix, of the sequence of range-distributions. Therefore, *the distribution over the whole line is the same as the asymptotic distribution over the N-range, namely Poissonian with mean and variance μ.* The other distribution derived naturally holds over the complete real line also. The Poisson distribution being valid for any μ, grouping consecutive intervals together k at a time (every interval belonging to one and only one such grouping) again gives a Poisson distribution with parameter $k\mu$. With uniform distribution of interval length over $(0, \mu)$ the mean and variance of primes will be $\frac{1}{2}k\mu$ and $k(\mu/2 + \mu^2/12)$ respectively; in our special case, k and $4k/3$. Thus, *the $\pi(\boldsymbol{d})$ in consecutive intervals (of a complete covering) grouped together add like independent random variables.* Q.E.D.

The distribution itself is less important than the existence of a non-singular distribution. Each individual $\pi(\boldsymbol{d}_\nu)$ has also some frequency distribution for fixed index ν, which obviously tends to the distribution over a complete covering as the index increases beyond limit.

A more number-theoretic proof of this fundamental lemma would run as follows: Brun's sieve theorem extends (*note* 4, *p.* 52, *th.* 4.7) to: *The number of primes $p \leq N$ for which $p + b_1,\ p + b_2, \ldots, p + b_r,\ 0 < b_1 < b_2 \cdots, b_r$, are also primes is less than $cMN/\log^{r+1} N$, where $M \leq \prod(1 - 1/p)^{-r}$, taken over all primes dividing* $\prod b_i \prod(b_j - b_k) > 0$. Let $p, p + b_i$ be restricted to lie within a single covering interval of length $h \leq \mu \log N$. By the theorem of Mertens, $M < a^r \log^r h$, where a is a constant. The b_i can be chosen at will provided there is no *a priori* restriction to prevent $p + b_i$ being a prime; for example, b_i must be even for $p > 2$. This means precisely that every $p + b_i$ must be a 'first survivor.' If R be the number of choices for any b_i, $R < Bh/\log h$. This follows in the sense of unit limiting probability from the preceding paragraphs, while it is known from purely number-theoretic considerations that no interval of length f can contain more than $Df/\log f$ primes, D constant. The whole set of b's may be specified in $R!/(R - r)!\,r!$ different ways. The number of covering intervals being N/h, the relative frequency of intervals covering $r + 1$ primes cannot exceed the binomial coefficient above, multiplied by $CM/\log^r N$. Therefore, ultimately, *the frequency of intervals having $r + 1 \geq 2$ primes cannot exceed $Ab^r/r!$ (A, b const.)* This suffices to prove the existence of the second and higher moments, but the vital Poisson distribution would require further refinement of the sieve, or probability arguments.

2. The series (22.2) is now written as the difference of the two stochastic series, whose convergence is to be considered separately:

$$\sum_{\nu}(\Delta_\nu - 1)\cdot\frac{\log x_\nu}{x_\nu^\sigma} - \sum_{\nu}[\pi(\boldsymbol{d}_\nu) - 1]\cdot\frac{\log x_\nu}{x_\nu^\sigma};\quad \sigma > \frac{1}{2}. \tag{22.7}$$

Theorem 1 *The series $\Sigma_\nu(\Delta_\nu - 1)\cdot\log x_\nu/x_\nu^\sigma$ converges with unit probability for $\sigma > \frac{1}{2}$ and zero probability for $\sigma < \frac{1}{2}$.*

Proof Step 1. There exists a critical value σ_0 of the exponent σ such that the convergence probabilities for the series under consideration are $P = 0$ for $\sigma < \sigma_0$ and $P > 0$ for $\sigma > \sigma_0$. If any series with a specific choice of $\boldsymbol{d}$ converge for a given σ, it necessarily converges for all greater values of σ; if it diverge, then divergence follows for all lesser values of the exponent. This is a consequence of standard results in the theory [1] of infinite series, noting that the coefficients outside the brackets are ultimately monotonically decreasing and positive. Thus, if the convergence probability be zero for any exponent, it cannot be positive for any lesser value of σ. This enables a Dedekind section to be defined for the values of σ, between the zero and the nonzero probability ranges.

Step 2. The convergence exponent is the same when $\log x_\nu/x_\nu^\sigma$ is replaced by $\log\nu/\nu^\sigma$ in the coefficients. To prove this, we note that by lemma 1.1, an arbitrarily small $\epsilon > 0$ may be chosen, with two suitable sets of positive constants $a, b; \bar{a}, \bar{b}$ such that $\log x_\nu/x_\nu^\sigma$ is bracketed between $a\cdot\log\nu^b/\nu^{\sigma+\epsilon}$ and $\bar{a}\cdot\log\nu^{\bar{b}}/\nu^{\sigma-\epsilon}$. Moreover, the probability P_ν for each bracketing is such that $\prod P_\nu$ converges. Conversely, a similar bracketing of $\log\nu/\nu^\sigma$ by corresponding terms in x_ν is also obviously possible. [In each case, the log terms in the factors may be ignored, as $\log^k z = 0(z^\epsilon)$ for every k and every positive ϵ.] It follows that if $\sigma < \sigma_0$ in the x-series, the convergence probability cannot be positive for the ν-series with the same exponent; similarly for $\sigma > \sigma_0$.

Step 3. Lemma K applies to the series $\sum_\nu(\Delta_\nu - 1)\cdot\log\nu/\nu^\sigma$. The term means are all zero; the variances are $\log^2\nu/3\nu^{2\sigma}$. The critical exponent for the $\nu - series$, and therefore, the x-series also is thus $\sigma = \frac{1}{2}$. Finally, the convergence probability for the x-series with $\sigma > \frac{1}{2}$ is at least $\prod P_\nu$. Inasmuch as $x_\nu \geq 2$, the contribution from any finite number of initial terms remains finite, regardless of probability considerations, and the terms may be omitted without affecting convergence. However, the omission of the corresponding terms in $\prod P_\nu$ brings the probability arbitrarily near to unity. Hence, the probability of convergence for $\sigma > \frac{1}{2}$ must be unity even for the x-series. This completes the proof, though it suffices for our ultimate purpose that $P > 0$ for $\sigma > \frac{1}{2}$. □

Theorem 2 *The series $\sum_\nu[\pi(\boldsymbol{d}_\nu) - 1]\cdot\log x_\nu/x_\nu^\sigma$ has likewise $P > 0$ for convergence when $\sigma > \frac{1}{2}$.*

Proof The existence of a critical exponent which coincides with that of the comparison series $\sum_\nu[\pi(\boldsymbol{d}_\nu) - 1]\cdot\log\nu/\nu^\sigma$ is proved as in the preceding theorem so that we may deal only with the latter series. □

Lemma 1.2 shows that for almost every complete covering $\pi(\boldsymbol{d}) - 1$ has a zero mean (*over the covering*, not for fixed index, ν) and finite variance. Therefore, with probability P_0 arbitrarily close to unity, $V\{\pi(\boldsymbol{d})\} < A^2$ over complete coverings for suitably large A. The sum of any m terms $\pi(\boldsymbol{d}) - 1$ taken at random from the same covering would be less than $A.t.\sqrt{m}$ in absolute value, with $P > 1 - 1/t^2$. Take non-overlapping consecutive blocks of $m = 2^k$ consecutive terms of the covering, with $k = 2, 3, \ldots$ and the understanding that an arbitrary number of initial terms of the series may eventually be omitted at need. The first subscript ν in each block will be equal to the total number of terms in that block. By Abel's lemma [1] and taking $t = k$, the sum of the terms in the comparison series corresponding to the k-th block will be less absolutely than $Ak^2 \log 2/2^{k\cdot\epsilon}$ (where $\epsilon = \sigma - \frac{1}{2} > 0$) with probability $P_k > 1 - 1/k^2$. Hence, the series has a convergence probability not less than $P_0 \prod P_k > 0$.

Choice of consecutive instead of completely random intervals does not vitiate the result. The existence of a distribution for $\pi(\boldsymbol{d})$ was proved as for consecutive intervals; and the selection is uninfluenced at any stage by the actual prime content of any intervals or blocks. In fact, it is known [3] that if m consecutive intervals together cover a stretch of magnitude y^c for any $c > 38/61$, then the (stochastic) block sum under consideration is $0(2m)$, as $y \to \infty$, without any probability condition or exceptional set of integers. This is stronger than what is demanded or yielded by probability considerations. That is, the values assumed by sums of $\pi(\boldsymbol{d}) - 1$ for sufficiently great block lengths cannot be more extreme for consecutive covering intervals than with random choice. Lemma 1.1 says, however, that every block length will be of order arbitrarily close to y with a compound probability given by an infinite product that converges to some $P^* > 0$. Thus, the inequalities can be strengthened, with a convergence probability at the worst multiplied by another factor P^*. The critical exponent of convergence for the series remains $\sigma_0 = \frac{1}{2}$.

Though justified by Lemma 1.2, the multiplication of probabilities P_k is unnecessary. No matter what the joint probabilities, the chance for *all* the grouped sums lying within the absolute limits given above for each is not less than $1 - \Sigma(1 - P_k)$, or than $1 - \Sigma 1/k^2$, which can be brought arbitrarily close to unity by rejection of enough initial terms of the series. The grouping of intervals need not be in geometric progression. It would suffice to combine $(k+1)^\alpha - k^\alpha$ successive intervals at a time with $k = 1, 2, \ldots$ and $\alpha > 3/(2\sigma - 1)$. The process need begin only from some $\nu = \nu_0$. Thus far, only the simple Chebyshev inequality is used, which requires nothing beyond the existence of a distribution with finite second moment. Use of the actual distributions found in Lemma 1.2, with or without the normal approximation given by the Central Limit Theorem, permits still freer groupings. Q.E.D.

An alternative proof of Theorem 2 would run as follows. The correspondence $\boldsymbol{d} \to \pi(\boldsymbol{d})$ maps the space of all permissible complete coverings into the points of an integral infinite-dimensional lattice with coordinates $x_i = \pi(\boldsymbol{d}_i)$. We take the lattice as right-angled and redefine the measure by giving equal weight to every point actually realized. Suppress a suitably large but finite number of initial dimensions altogether. Then, if necessary, trim off just enough peripheral points to make the center of gravity (with equal weights) the unit point $(1, 1, 1, \ldots)$. This can always be done with an arbitrarily small measure of deletion because almost every realizable point of the lattice has $(x_1 + x_2 + \cdots + x_n)/n \to 1$ and a limiting distribution (in the old measure) is approached by all x_i with large i which is also the distribution over the successive x-coordinate values of almost every point. What is left may be further restricted to almost all lattice points of an infinite-dimensional hypercube (with unequal indefinitely increasing sides), of the same center of gravity, and with a rectangular section in any finite number of dimensions. If $x_i = a$ is realizable, then so is every value $0 \le x_i \le a$, by contraction of interval length.

The new measure is defined over this lattice hypercube as the proportion of points lying in any included region to be measured. The total measure of the hypercube is unity, with an induced measure over every subspace which is defined as the relative number of points in the lattice hypercube lying in the cylinder with a region of the subspace as base and sides extended over the entire ortho-complement. Then, the measure over the product-space is the product of the component measures. This change of probability measure amounts to the integration of a positive weight function. It suffices for our purpose that no set of measure unity in the new lattice measure is of measure zero in the original measure. In the lattice measure, the variates $\pi(\boldsymbol{d}_\nu)$ become independent in probability; each has a unit mean because of the center of gravity chosen, and the variance can never be greater than $2 \cdot \log^2 \nu$, whatever the actual distribution. Therefore, Lemma K becomes immediately applicable, and the comparison series converges with uniform lattice measure one for every $\sigma > \frac{1}{2}$; hence, the original series with $P > 0$ in the previous measure.

Theorem 3 *The difference between the series (22.2) and (22.1) may be resolved into the following series and sequences, each of which converges with unit probability for $\sigma > 0$:*

$$\sum_\nu \frac{(d_\nu - [\boldsymbol{d}_\nu])}{x_\nu^\sigma};$$
$$\sum_\nu \frac{\pi(\boldsymbol{d}_\nu) \cdot d_\nu(\sigma \cdot \log x_\nu - 1) - \sigma \cdot d_\nu(d_\nu + 1)}{x_\nu^{1+\sigma}};$$
$$\frac{(d_\nu + 1)}{x_\nu^\sigma}; \quad \pi(\boldsymbol{d}_\nu) \cdot \frac{\log x_\nu}{x_\nu^\sigma}. \tag{22.8}$$

Proof The first of these is due to there being $[\boldsymbol{d}]$ and not d integers in $\boldsymbol{d}$. Now not only $d - [\boldsymbol{d}]$ but the sum of any number of such differences for consecutive non-overlapping intervals $\boldsymbol{d}_\nu$ ranges by definition between -1 and $+1$ without reaching either extreme. By Abel's lemma, the series will converge provided $x_\nu \to \infty$ monotonically, for which the probability is unity. Terms of the second series may be compared with $\log^3 \nu/\nu^{1+\sigma}$; and so it also converges with $P = 1$ for $\sigma > 0$. The two sequences are due to the partial sums of (22.1) and (22.2) not necessarily terminating at the same place; both obviously converge with $P = 1$ for $\sigma > 0$. Q.E.D. □

The three auxiliary theorems lead immediately to:

Theorem 4 *The series (22.1) converges for $\sigma > \frac{1}{2}$.*

Proof The series (22.1) converges if and only if (22.2) and (22.8) converge for at least one choice of consecutive non-overlapping intervals $\boldsymbol{d}$. If no such choice exists, the joint probability for the simultaneous convergence of all the stochastic series and sequences in (22.2) and (22.8) would have to be zero. But the joint probability is positive (in fact arbitrarily close to unity). Q.E.D. □

3. The function $\zeta(s)$ is defined for a complex variable $s = \sigma + it$ with σ, t real, for the half plane $\sigma > 1$ by

$$\zeta(s) = \sum_{1}^{\infty} \frac{1}{n^s} = \prod \frac{1}{(1 - p^{-s})} \,. \tag{22.9}$$

Both the series and the infinite product converge for $\sigma > 1$. The function $\zeta(s)$ thus defined by the series and its analytic continuation has no singularity in the entire finite plane except for the simple pole with unit residue $1/(s-1)$, as is well known.

The zeta-function obeys the functional equation [6]:

$$\zeta(1-s) = 2^{1-s} \pi^{-s} \cos\left(\frac{\pi s}{2}\right) \Gamma(s) \zeta(s) \,. \tag{22.10}$$

The Riemann hypothesis (RH) is the conjecture that all zeros of $\zeta(s)$ not $s = -2, -4, \ldots$ lie on the vertical line $\sigma = \frac{1}{2}$. It is easily seen, directly from the convergence of the infinite product, that no zero can occur in $\sigma > 1$. It is also known from a theorem of G. H. Hardy that an infinity of zeros lie on the line $\sigma = \frac{1}{2}$. Using the functional equation, it would suffice to prove RH if it could be shown that no zero lies in the critical half-strip $\frac{1}{2} < \sigma \leq 1$. To this end, we use a classical lemma of function theory: *Any singularity of an analytic $F(z)$, except isolated simple poles with unit residue, and any zero of $F(z)$ is a singularity of $F(z) + F'(z)/F(z)$.* Only the simple poles $1/(z-a)$ cancel out, but zeros of $F(z)$ now appear as first degree poles because of the second term, the logarithmic derivative. For $F(s) = \zeta(s)$, the fact that $\zeta(s)$ has no finite singularity other than the pole $1/(s-1)$ would mean that the singularities of $\zeta'(s)/\zeta(s) + \zeta(s)$ must be due only to the zeros of $\zeta(s)$.

Formally, differentiation of the logarithm of the infinite product in (22.9) gives, using the series expansion $\log(1-x) = -x - x^2/2 - x^3/3 - \cdots$:

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{v} \frac{\log p}{p^s} + \sum_{p} \frac{\log p}{2p^{2s}} + \sum_{p} \frac{\log p}{3p^{3s}} \cdots \tag{22.11}$$

The expansion is valid for $\sigma > 1$. For $\frac{1}{2} < \sigma$, all the series on the right except the first are together dominated by $2\Sigma \log n/n^{2\sigma} = -2\zeta'(2\sigma)$. Therefore, the discussion by means of $\zeta(s) + \zeta'(s)/\zeta(s)$ reduces to showing that the Dirichlet series

$$\sum \frac{1}{n^s} - \sum \frac{\log p}{p^s}\,, \quad p, n \leq x \to \infty\,, \quad s = \sigma + it\,, \tag{22.12}$$

converges for all $\sigma > \frac{1}{2}$. But we have already shown that (22.1), which is the form assumed by (22.12) on the real axis, converges for all $\sigma > \frac{1}{2}$. Hence, by the known property of such Dirichlet series, (22.12) converges uniformly in any half plane to the right of $\sigma = \frac{1}{2}$. This proves that $\zeta(s) + \zeta'(s)/\zeta(s)$ has no finite singularities for $\sigma > \frac{1}{2}$. Therefore, no zeros of $\zeta(s)$ can occur in $\frac{1}{2} < \sigma$, proving RH:

Theorem 5 *The Riemann zeta-function, defined for* $\sigma > 1$ *as in (22.9), has all its non-trivial zeros on the vertical line* $\sigma = \frac{1}{2}$.

The corresponding theorem for the Dirichlet L-functions is proved in analogous fashion. The consequences are well known [7].

Possible convergence of (22.1) on or to the left of $\sigma = \frac{1}{2}$ would not affect RH because singularities of $\zeta'(s)/\zeta(s)$ occur in any case on the line $\sigma = \frac{1}{2}$ from the second series on the right in (22.11). Moreover, the function $Q(s) = -\Sigma \log p/p^s$ is $\Sigma\mu(n)\zeta'/(ns)/\zeta(ns)$ in $\sigma > 0$, so has poles on $\sigma = \frac{1}{2}$. Hence (22.1), the Dirichlet series for $\zeta(s) + Q(s)$, cannot converge beyond the critical line.

The probability approach allows some conclusions to be drawn quickly without the intermediacy of RH. For example, the Poisson distribution for primes covered by unit-image intervals, and the famous law of the iterated logarithm allows a probability estimate of $|\pi(x) - li(x)|$. This, under the assumption of independence for primes in the given intervals, would exceed with probability arbitrarily close to unity, the magnitude $(1-\delta)\sqrt{2y \cdot \log\log y}$. The probability would be arbitrarily close to zero if the $-\delta$ be replaced by $+\delta$. With $y \sim x/\log x$, it is seen that the original Littlewood result is not quite the best possible.

The zeros of $li(x) - \pi(x)$ appear as recurrence times (on the y-scale) for the equilibrium of a Poisson variate. The distance between consecutive primes amounts to the 'waiting time' on the y-scale and has a distribution given by $dP = e^{-\mu}d\mu$. It follows that *for any* $\phi(n) = o(\log n)$ *and infinitely many primes* p, *the separation from the next prime will exceed* $\phi(p)\log p$. Systematic use of the Poisson distribution would eliminate Theorem 1 altogether, but would not bring out the basic fact that Theorem 2 is independent of any reasonable choice of covering intervals. Finally, RH may be generalized to Dirichlet series whose exponents (our $\log n$) form a complete Abelian semigroup under addition with a basis set of generators, our $\log p$. But no generalization of RH exists if the product corresponding to $\prod(1 - 1/p^s)$ converges in the half plane $\sigma > 0$. This covers the case where the generator basis is finite and should explained the negative Bourbaki–Weil result for Abelian fields.

References

1. Knopp, K. *Theory and Application of Infinite Series* (Trans. London, 1928), particularly pp. 313–15.
2. Kolmogorov, A. Math. Ann. **99**, 309–319 (1928).
3. Prachar, K. *Primzahlverteilung*, Berlin, 1957, pp. 8, 158–64, etc. The position of RH is made clear in various contexts throughout the book. For the Selberg asymptotic result, pp. 323–24; without an exceptional set of integers of measure zero, the best known exponent is 38/61, reducing that given by A. E. Ingham in *Quart. J. Math.*, Oxford, 1937, pp. 255–66.
4. Uspensky, J.V. *Introduction to Mathematical Probability* (New York, 1937), contains the elementary probability theory utilized here. For the Bernstein estimates, *see* pp. 204–206. Also, reference 7 below.
5. Feller, W. *An Introduction to Probability Theory and Its Applications*, vol. 1 (New York, 1950), sections 6.4, 4.5, 11.8 and 17.2; further developments have been promised for the second volume. See also H. Richter: *Wahrscheinlichkeitstheorie* (Berlin, 1956), 336–40. For extensions of the Chebyshev inequality, *ibid.*, 256, though not used here.
6. Titchmarsh, E.C. *Theory of Functions*, Oxford, 1932, p. 153; Chapter IX, pp. 113–16, and p. 260 for other function-theoretic results used here. See also the next reference.
7. Titchmarsh, E.C. *Theory of the Riemann Zeta-function*, Oxford, 1951, Chapters, XIII–XIV; also Prachar as in note 3 above.

Chapter 23
The Sampling Distribution of Primes

D.D. Kosambi, P. O. Deccan Gymkhana, Poona

This paper was communicated to the journal by H. S. Vandiver, number theorist and fellow of the US National Academy of Sciences. The paper was reviewed by J. B. Kelly who felt that the "exposition is rather sketchy; in particular, the reviewer could not follow the proof of the crucial Lemma 4". By the time this paper was submitted to the journal, DDK was no longer at TIFR and also not formally associated with any other academic institution, hence the post office address.

The real half-line $x \geq x_0 \geq 2$, upon which the integers are marked off unit distance apart, is mapped onto $y \geq 0$ by the transformation $y = \int_{x_0}^{x} dt/\log t = \mathrm{li}(x) - \mathrm{li}(x_0)$. Cover the whole of $y \geq 0$ by a sequence of intervals, each of length $u > 0$, fixed. The nth such interval will be $(n-1)u \leq y < nu$, and $\pi_n(u) = \pi(x_0, u; n)$ denotes the number of primes in its x-image. We show that the primes in an arbitrary connected stretch of the y-line have a Poisson distribution in the sense of probability theory, the sequences $\pi_n(u)$ constituting statistical *samples* thereof. Hereafter, take all positions of the initial point (on the y-line) as equally likely and x_0 neither restricted nor specified otherwise.

Textbook results in number theory and probability theory are taken for granted. In particular,

Lemma 1 *The number of primes $p \leq x$ is $\sim \mathrm{li}(x) \sim y$ (for any x_0, as $x \to \infty$). If $\vartheta(x) = \sum \log p$, $p \leq x$, then $\vartheta(x) \sim x$. If p_k be the k th prime in order, starting from $p_1 = 2$, then $p_k \sim k \log k$.*

The first of these is the prime number theorem[1], and the other two are equivalent, as is well known.

Published in Proceedings of the National Academy of Sciences (USA) **49**, 20–23 (1963).

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_23

Lemma 2 *For $p \leq x$, $\prod(1 - 1/p) \sim e^{-\gamma}/\log x$; γ, Euler's constant.*

This is a classical theorem of Mertens[2].

Lemma 3 *If for any set Z of primes, $\prod p = x,\ p \subset Z$, then $\prod(1 - 1/p)^{-1}$ is less than $C \log_2 x,\ \ p \subset Z$, x large.*

Proof The product of $(1 - 1/p)^{-1}$ will be greatest for any given number of primes if the primes are $2, 3, \cdots$ in sequence and all distinct. Then, $\log x = \log \prod p = \sum \log p$ by hypothesis, $p \subset Z$. Lemma 1 says that packing the primes at the beginning of the sequence, max $p \sim \sum \log p$, and here, $\sum \log p = \log x$. By Lemma 2 (the product being not greater than in this case) $\prod(1 - 1/p)^{-1} < C \log_2 x,\ p \subset Z$. Q.E.D. □

Lemma 4 *The proportion of u-intervals for which $\pi(x_0, u, n) \geq 2$ is less than cu^2 for small u, regardless of x_0, if x is large.*

Proof The sieve of Viggo Brun leads to the theorem:[3] The number of primes $p \leq x$ for which $p + b$ is also a prime is $< (cx/\log^2 x) \prod(1 - 1/p)^{-1}, p|b$. The u-intervals containing two or more primes must contain one such pair p, $p + b$ for some $b \leq u \log x$ approximately. Not all b, however, are admissible, as no odd b will do for $p > 2$. The number of admissible b's within the same u-interval is easily seen to be not greater than the number of integers in (the x-image of) the covering interval prime to $N = 2.3 \cdots p$, provided $N \leq u \log x$. Clearly, $p + b$ not a prime to N cannot be a prime except in the interval that begins from $x_0 = 2$, which may be ignored; moreover, such numbers are arranged cyclically modulo N, which, being about the length of the interval on the x-axis, cannot be materially changed in the vicinity of any given x. By Lemmas 2 and 3, the admissible set will contain less than $c'u \log x/\log_3 x$ members, for large x. The bound for $\prod(1 - 1/p)^{-1}$ for primes dividing any b in the interval cannot ultimately be greater than $c'' \log_3 x$. Finally, the total number of covering intervals in the range is $\sim x/u \log x$. The estimate therefore is not in excess of $(cx/\log^2 x)\ (c'u \log x/\log_3 x)\ (c'' \log_3 x)\ (u \log x/x) = \bar{c}u^2$. Q.E.D. □

Lemma 5 *If $f_0, f_1, f_2, \cdots$ be the relative frequencies, $\sum f_i = 1$, with which small u-intervals containing $0, 1, 2, \cdots$ primes occur in a large range of x, then $f_1 = u + o(u)$.*

Proof Corresponding to the theorem cited in the proof of Lemma 4 is an extension by P. Erdős:[4] The number of primes $p \leq x$ for which all the numbers $p + b_1, p + b_2, \cdots, p + b_r, 0 < b_1 < b_2 < \cdots < b_r$ are also primes is less than

$$(cx/\log^{r+1} x) \prod_{p|E}(1 - 1/p)^{-(r+1-\omega(p))}, \quad E = \prod_{i=1}^{r} b_i \prod_{1 \leq i < k \leq r}(b_k - b_i) \qquad (23.1)$$

where $\omega(p)$ is the number of solutions mod p of $m(m + b_1) \ldots (m + b_r) \equiv O(\text{mod} p)$. From this point, the reasoning of the previous lemma holds, except that the number of choices for the set of r b's will not exceed the binomial coefficient ${}^nC_r$, with

$n = c \log x / \log_3 x$ and $\prod p$, $p|E$ cannot exceed $(u \log x)^m$, with $m = r^2(r-1)/2$ (an overestimate which we shall not stop to refine). The upper bound, for small u, is therefore $cu^{r+1}/r!$ for each r, and the same c may be taken throughout, quite obviously. For any u, the contribution of $f_2, f_3, \cdots$ to the expectation (mean value, average) of primes per covering interval may be assessed as not exceeding cu^2e^u. The mean value is $(0.f_0 + 1.f_1 + 2.f_2 + \cdots)$, so that f_0 contributes nothing. Any term from f_2 onward, as assessed above, will contribute $O(u^2)$. The total contribution of those terms will be $O(u^2e^u)$, as may be seen from the upper bounds just given above. Now, the mean value, by the prime number theorem, is exactly u, over the whole y-line, no matter what the x_0. It follows that for small u, $f_1 = u + O(u^2)$. Q.E.D. □

Theorem *With all x_0 equally likely, the probability that exactly r primes will lie in the x-image of $0 \leq y < t$ is $e^{-t}t^r/r!$ (the Poisson distribution, with parameter t).*

Proof Given x_0, there is no question of any probability; the entire sample is completely defined for the whole y-line. But under the present conditions, the irregularity of primes permits the use of the concept "probability" the "event" being $0, 1, 2, \cdots$ primes lying in the interval $0 \leq y < t$. These events are exhaustive and mutually exclusive. The conditions for a Poisson process are given by the following postulates:[5] The probability for one prime in $t \leq y < t + h$ for small h is $h + o(h)$; the probability for more than one prime in the small interval is $o(h)$; and the probability for the small interval being totally void of primes is $1 - h + o(h)$. Lastly, none of these are affected if it is known that k primes have actually occurred in $0 \leq y < t, k = 0, 1, 2 \ldots$.

These postulates are obviously satisfied in view of our lemmas above. Lemma 4 says that the probability (approximated arbitrarily closely by the corresponding frequency) for more than one prime in the small interval is $o(h)$. Lemma 5 gives the probability for a single prime as $h + o(h)$. Since these two cases and that of the h-interval being void of primes are mutually exclusive and exhaustive, the third postulate is satisfied. Finally, the lemmas hold regardless of x_0 and t, over the whole of the y-line, $y > t$. Moreover, the number of primes known to have occurred in $0 \leq y < t$ does not in any way affect the frequencies or probabilities or permit x_0 to be determined even approximately. (It is possible to go much further in this direction,) for not even the precise knowledge of the points $t_1, t_2, \cdots$ at which these primes may actually have occurred changes the situation. If it could then be said that there *must* exist a prime in $t \leq y < t + h$, no matter how small the h, it would follow that the $k + 1$ st prime could be located from the positions on the y-line of the first k, for all large primes and some k. This implies a recurrence relation between the primes; no such relation is known, and an algebraic one of any finite degree is demonstrably impossible. There is no finite upper bound for the gap between consecutive primes on the y-line[6] and no known positive lower bound. On the other hand, it is known that subsequences of primes (of positive density) exist[7] for which the y-distance between consecutive primes is dense over a certain positive range, whose precise termini are not known. This shows the impossibility of using any but probability methods. Q.E.D. □

The Poisson distribution of our theorem may be quickly derived as follows. For the argument, allow x to be any point (with equal likelihood) of a range $R(x) \approx x^{\alpha}$, $38/61 < \alpha < 1$. It is known (Ingham, A. E., *Quart. J. Math.*, **8**, 255–266 (1937)) that the prime number theorem holds asymptotically over $R(x)$ as $x \to \infty$. Further, let $I(x)$ be a randomly selected interval within $R(x)$ of y-length t, hence containing $\sim t \log x$ integers regardless of position (since the variation in $\log x$ is negligible over $R(x)$). No matter where $I(x)$ is located, alternate integers in it must be even, four out of every six (regularly arranged) divisible by 2 or 3, etc. This regularity of deletion by the sieve of Eratosthenes extends to all the smallest primes whose product $2.3.5 \cdots p = N \leq t \log x$. About $te^{-\gamma} \log x / \log_3 x = tg(x)$ integers in $I(x)$ will survive. Any p not a factor of N need not be the smallest prime factor of a surviving integer in $I(x)$, and a prime larger than $t \log x$ need not even have a multiple in $I(x)$, so that one of the "survivors" being deleted by any such prime is now a matter of chance with probability $1/p$. By the prime number theorem, the expectation of primes in $I(x)$ is exactly t (in the limit); hence, the compound probability for primality of a "survivor" is asymptotic to $1/g(x)$. Moreover, if some k of these survivors be tested and found composite or prime (without revealing their numerical values), the knowledge does not modify the probability for primality for the rest. In all this, x is merely a background parameter, whose principal use is to furnish relative magnitudes of the various functions involved, as $x \to \infty$.

It follows that if P_r be the probability for precisely r primes in $I(x)$, then in the limit, $P_0 = \lim(1 - 1/g)^{tg} = e^{-t}$. Using textbook definitions and procedures, the limit $P_1 = \lim(1 - 1/g)^{tg-1}(tg)(1/g) = te^{-t}$, and so on, with limit $P_r = e^{-t}t^r/r!$ But any limiting distribution over $R(x)$ as $x \to \infty$ will obviously be the distribution over the entire x-line, here the Poisson distribution with parameter t, as before.

References

1. Prachar, K. in *Primzahlverteilung* (Berlin, 1957). (ch. 3).
2. Hardy, G.H. and E.M. Wright, in *An Introduction to the Theory of Numbers* (Oxford University Press, 1945), Theorem **430**, 349–354.
3. Prachar, K. *op. cit.* (ch. 2, Theorem 4.4).
4. *Ibid.*, Theorem 2.4.7.
5. Feller, W. *An Introduction to Probability Theory and Its Applications* (New York, 1950), vol. 1, p. 366. (*et passim*).
6. For general known results on gaps in the sequence of primes, see Prachar, *op. cit.*, p. 154 ff.
7. Ricci, G. in "Sul pennello di quasi-asintoticità delle differenze di interi primi consecutivi," *Rend. Atti. Accad. Naz. Lincei*, **8**, 192–196 and 347–351 (1954-5).

Chapter 24
Statistical Methods in Number Theory

D.D. Kosambi, Poona

Although no longer affiliated with the TIFR, DDK persisted in his prime obsession, repeating and refining the basic arguments set forth in his earlier papers. The Hungarian mathematician A. Rényi reviewed this paper in Mathematical Reviews and noted that "Neither in this paper nor in his previous paper did the author succeed in proving his hypothesis, nor in deducing from it the Riemann hypothesis". The "Kosambi hypothesis" (see the discussion on pages 7–9) is, according to Rényi "even more difficult than the problem of the validity of the Riemann hypothesis. As a matter of fact, no obvious method exists to prove the author's hypothesis even under the assumption of the Riemann hypothesis." Theorem 1 *of this paper, therefore, is not proven.*

24.1 Basic Results

The function $\zeta(s)$ of a complex variable $s = \sigma + it$ is defined by the Dirichlet series and the Euler product:

$$\zeta(s) = \sum \frac{1}{n^s} = \prod \left(1 - \frac{1}{p^s}\right)^{-1} ; \tag{24.1}$$

n, integer; p, prime; $\sigma > 1$; and by analytic continuation over the rest of the complex domain. It is known [1] that $\zeta(s)$ has no finite singularity except the simple pole $1/(s-1)$. It has no zero in $\sigma > 1$ and only the trivial zeros $s = -2, -4 \ldots$ in $\sigma < 0$. Infinitely many of its zeros lie on the vertical line [2] $s = \frac{1}{2} + it$. The Riemann hypothesis (= RH) is that all non-trivial zeros of $\zeta(s)$ lie on $\sigma = \frac{1}{2}$.

Let $\pi(x)$ be the number of primes $p \leq x$ and $li(x)$ the integral $\int dt/\log t$ over $2 \leq t \leq x$. Then it is further known [3] that the range of variation of $\pi(x) - li(x)$

Published in Journal of the Indian Society of Agricultural Statistics **16**, 126–35 (1964). Reprinted with permission.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_24

must include $\pm x^a$ infinitely often as $x \to \infty$, where a is the greatest abscissa of any zero of $\zeta(s)$. It was proved by J. E. Littlewood [4] that there exists a number $b \geq 0$ such that the value of $\pi(x) - li(x)$ obeys each of the inequalities:

$$\pi(x) - li(x) > b\sqrt{x}\,\frac{\log_3 x}{\log x}\,;$$
$$\pi(x) - li(x) < -b\sqrt{x}\,\frac{\log_3 x}{\log x}\,, \tag{24.2}$$

infinitely often as $x \to \infty$. Here, $\log_2 x = \log(\log x)$ and $\log_3 x = \log(\log_2 x)$, all logs to the base e.

Starting from the initial point $x_0 > 2$ and any fixed but arbitrary $u > 0$, the real half-line $x \geq x_0$ is transformed into $y \geq 0$ and covered by right-open intervals I_n as follows:

$$y = li(x) - li(x_0)\,; \quad I_n : (n-1)u \leq y < nu\,. \tag{24.3}$$

The number of primes in the x-image of I_n is denoted by $\pi_n(u)$ or $\pi(x_0, u; n)$. The prime number theorem [5]: $\pi(x) \sim li(x)$, amounts to $\Sigma\pi_n(u) \sim Nu$, summation over $1 \leq n \leq N \to \infty$. This gives:

Lemma 1 *RH is true if and only if, for every $\varepsilon > 0$ and some $u > 0$:*

$$\sum_1^N \pi(x_0;\ u\,; n) - Nu = O(N^{\frac{1}{2}+\varepsilon})\,. \tag{24.4}$$

It is essential to show that the totality of distinct sequences $\{\pi(x_0, u; n)\}$ is equivalent to the number of points on a continuous line segment. This will enable a suitable measure to be introduced. To this end, the following lemma is essential:

Lemma 2 *There exists at least one $u > 0$ such that the number of distinct sample-sequences $\{\pi_n(u)\}$ obtained by shifting the initial point x_0 through a single covering interval of y-length u can be put into a 1-1 correspondence with the points of $0 \leq t < 1$.*

Proof Suppose that, for some given u and x_0, the same sequence $\{\pi_n(u)\}$ is obtained when the initial point is shifted to the right through a y-distance w. It would then follow that the number of primes gained by any interval at the right is precisely equal to that lost at the left during the shift. Therefore, every w-interval, separated by the y-distance $u - w$ from the next on either side, must contain the same number of primes. Known separation theorems [6] by P. Erdős say that there exist infinitely many gaps between consecutive primes, larger than any preassigned y-length. Hence, these w-intervals must be totally void of any primes. □

The results of Ricci [7] show, on the other hand, that there exist subsequences of primes such that the y-distances between consecutive primes are dense over some nonzero interval $(1-\alpha, 1+\beta)$. If $\alpha = 1$, then take any $u < 1+\beta$. Otherwise, take an integer k so large that $(1-\alpha)/k$ is less than $\alpha+\beta$, and take $(1-\alpha)/k = u$. In either case, the Ricci density theorem shows that w must vanish. Thus, for the chosen u (and there are infinitely many such choices, obviously), there must be as many distinct sequences as points of $0 \le t < u$; this can be projected upon the unit interval $0 \le t < 1$, to complete the proof of the theorem (which holds in fact for all $u > 0$).

Lemma 3 *If $M = M(z)$ be the product of all primes $p \le z$ and γ is Euler's constant, then the number of integer relatively prime to M in any range $A \le n < A + R$ is asymptotic to $Re^{-\gamma}/\log z$ as $z \to \infty$, provided $R/\log z$ is large compared to $2^{\pi(z)}$, where $\pi(z)$ is the number of primes $p \le z$.*

Proof This is essentially the form in which the Sieve of Eratosthenes is to be used. The integers prime to M are cyclically arranged modulo M with symmetry about the middle of any cycle kM to $(k+1)M$. If $A = kM + 1$, the number out of the R consecutive integers, not divisible by any prime $p \le z$, is given by:

$$R - \left[\frac{R}{p_i}\right] + \left[\frac{R}{p_i p_j}\right] - \left[\frac{R}{p_i p_j p_k}\right] + \cdots ; \quad p_i, p_j, p_k \neq . \tag{24.5}$$

The square brackets denote the largest positive integer in the enclosed quotient, or zero. The primes are to run through the complete set $p \le z$. Since no remainder can be as great as unity, the difference when the brackets are removed will not exceed $\frac{1}{2}(1+1)^{\pi(s)}$ in absolute value. For any A, we can regard the result as the sum of difference of two expressions as in (24.5). The asymptotic value of $R\prod(1-1/p)$ which is the value of (24.5) with brackets removed is $Re^{-\gamma}/\log z$ by the classic theorem of Mertens [8]. □

24.2 Lemmas on Measure

Definitions.—A proper frequency distribution is furnished by a set of real numbers $f_i > 0$ such that $\Sigma f_r = 1$. If $A_0, A_1, A_2, \ldots$ be an indexed set of distinct *attributes*, an infinite sequence thereof $A_i A_j A_k \cdots$ (not necessarily all distinct) represents a *sample*, or point is sample-space. A sequence $\{A_r\}$ wherein the limiting frequency with which a particular A_n occurs is, for every n, the f_n above has that distribution. By *probability* is meant a measure function obeying the usual postulates, defined over the whole sample-space or over a subset thereof, such that the total measure of the universe of definition is unity. The probability measure of an *event* (subset of sample-space) is indicated by the letter P. The nth term of a sequence has the designation X_n and $P(X_n = A_j)$ is the probability measure, if it exists, of the set of sample-points where A_j appears as the n-the term of the corresponding sequence. The joint probability of a *compound* event is similarly defined, e.g., $P(X_i = A_j\,;\; X_r = A_k \ldots)$. A sample-sequence is *normal* if every finite combination $A_i A_j A_k \cdots$ occurs with frequency equal to the product of the individual component frequencies. Correspondingly, the events $X_i = A_j,\;\; X_r = A_k \ldots$ are said to be *independent* in probability if for any number of such events the compound probability is the product of the component individual probabilities.

Lemma 4 *Given a set of attributes $A_0, A_1, A_2, \ldots$ and a corresponding proper frequency distribution. Then there exists a mapping whereby: (1) The totality of sample-sequences is mapped in a 1-1 manner onto the right-open unit interval $0 \leq t < 1$. (2) The Lebesgue measure on the map is equivalent to probability measure over the sample-space. (3) Almost all sample-sequences are normal with the given basic frequency distribution and all the events $X_i = A_j$ are independent in probability.*

Proof The actual map is constructed as follows. Divide $(0, 1)$ into right-open subintervals by marking off successive points $t_0 = f_0, t_1 = f_0 + f_1, \ldots, t_i = f_i + t_{i-1}, \ldots$ Then subdivide the subintervals $(0, t_0)$, $(t_0, t_1) \ldots$ in the same manner, *each* in proportion to its total length. And so on, step by step. For the sequence $A_i A_j A_k, \cdots$, take first the subinterval immediately to the left of t_i in the first subdivision. Then in the next subdivision of this selected interval, that to the left of the point marked off with the subscript j; and so on, taking the next stage of subdivision for each successive subscript. The sequence of nesting intervals obviously converges to a single point in $(0, 1)$. Conversely, to each such point there corresponds just one sequence of subscripts, provided a suitable convention is made (to avoid duplication) about sequences terminating in an infinite succession of zeros or of the final index r when the total number of frequencies is finite and equal to $r + 1$. The properties listed follow obviously, with this mapping. □

The well-known theorem of Borel [9]: *almost every number in* $(0, 1)$ *is normal in a "decimal" expansion to any base* becomes a special case of this lemma when the number of attributes is finite, with equal frequencies. The proof, for finite or infinitely many basic frequencies, may be derived from the law of large numbers [10] in probability theory.

Lemma 5 *Given a sample-space where the basic frequencies have a Poisson distribution with parameter u, the sample-sequences are normal, and the attribute A_r assigned the numerical value r. Then almost all points of the sample-space obey the inequalities:*

$$\begin{aligned}-(1+\varepsilon)\sqrt{2Nu\log_2 Nu} \; &< \sum_{1}^{N}(X_i - u)\\ &< (1+\varepsilon)\sqrt{2Nu\log_2 Nu}\,, \qquad (24.6)\end{aligned}$$

with at most a finite number of exceptions as $N \to \infty$, for every $\varepsilon > 0$.

Proof This is the *upper* law of the iterated logarithm, abbreviated $ULIL$. The Poisson distribution has $f_i = e^{-n}u^i/i!$. The standard proof [11] for binomial distributions extends immediately to the Poisson, and hence need not be repeated here. The canonical mapping of Lemma 4 is to be used. □

Lemma 6 *$ULIL$ holds with unit probability for all $\varepsilon > \lambda - 1 > 0$ if the X_i of Lemma 5 have the Poisson distribution with parameter u and a joint distribution such that: (1) The sum of any finite number k of consecutive X_r has the Poisson distribution with parameter ku. (2) The probability that $|\Sigma_1^k(X_r - u)| > \lambda\sqrt{2Nu\log_2 Nu}$ for some $\lambda > 1$ and at least one $k \leq N$ does not exceed the corresponding probability when the X_r have distributions independent in probability; for all large N.*

Proof Lemma 5 does not depend upon any particular mapping, nor is independence necessary, though it suffices. The first Borel–Cantelli lemma [12] upon which $ULIL$ depends does not require independence. The two conditions given here suffice for the textbook proof of Lemma 5 cited, [11] as may be verified by inspection. □

24.3 Applications

In what follows, only the sample-sequences $\{\pi(x_0, u; n)\}$ are considered. The attribute A_k will be taken to have presented itself whenever a member of such a sequence has the value k. Again, X_r is simply the numerical value of the r-th member of such a sequence. Then we have:

Lemma 7 *The sequences $\{\pi_n(u)\}$ have the Poisson frequency distribution with $f_r = e^{-u}u^r/r!$, in the sense of unit probability measure.*

Proof This follows from known [13] results and could be proved again from the following considerations: As the number of trials (integers tested per covering interval) increases, the probability of the event (of a number being prime) tends to zero, but nevertheless the expectation (primes "expected" per interval) tends to u; the probability is unaffected by the results of any number of previous trials, or at worst the change in the probability is an infinitesimal of higher order than P itself. This last point is proved in the next lemma; the rest are obvious. □

There now arise three possibilities:

(A) The sequences $\{\pi_n(u)\}$ are *independent* in probability, in the sense that the actual values of any finite number of X's will not determine x_0, nor affect the probability for a given value of any other X_r to occur. In that case, $ULIL$ of Lemma 5 and therefore Lemma 1 would hold; hence, RH is true. In addition, the *lower* law of the iterated logarithm would also apply, which would enable the Littlewood inequalities (2) to be improved with the bounds replaced by $\pm(1-\varepsilon)\sqrt{2x\log x\log_2 x}$.

(B) The sequences may not be independent, but the effect of any dependence upon the sums of consecutive member may be *compensatory*. That is, deviations in the sums from expectation might be no greater (in probability) than in case A. It suffices if the probability measure of the set $|\Sigma\pi_n(u) - Nu| > a$ (summation over indices 1 to N) does not exceed that in the case A. Then Lemma 5 and $ULIL$ could still hold, but (2) cannot be improved and the Littlewood inequalities might be the best possible.

(C) The effect of dependence (if any) might be *cumulative*. That is, the occurrence of an excess from expectation in either direction, for sums of consecutive $\pi_n(u)$, might imply a similar excess in the same direction somewhere else in the same sequence (positive autocorrelation). In this case, $ULIL$ need not hold, nor RH.

The sieve of Eratosthenes, as will be seen, excludes C.

Lemma 8 *The terms of a sequence* $\{\pi_n(u)\}$ *are asymptotically independent in probability; moreover, the effect (if any) upon sums of consecutive terms of any deviation from independence cannot be cumulative, but at most compensatory.*

Proof In the discussion that follows, consider only such covering intervals as lie in the range $(x/2, x)$. This suffices because:

(i) The prime number theorem (and hence also the Poisson distribution) is asymptotically valid over such lengths of the real half-line; in fact, even over much smaller ranges, $(x, x + x^a)$ if $a > 38/61$, as is known [14].

(ii) The proof and applications of $ULIL$ may be carried through with successive ranges of order (Ab^k, Ab^{k+1}), with any fixed $b > 1$ and $k = 1, 2, 3, \ldots,$ so that there is no loss of generality involved. In the discussion, however, x is only to be regarded as a large background parameter whose sole use is to estimate the relative magnitudes of various arithmetic functions that appear. If x were specified exactly, there would be no question of probability, as everything would be exactly known. □

In the sieve of Eratosthenes, the multiples of 2, 3, 5,... are successively deleted; at each stage, the smallest number left is the next prime to be used in deletion. This way, every prime and only primes are obtained, as a succession of smallest survivors of the deletions. Every integer is deleted by its *smallest* prime factor, and once deleted, so remains regardless of how many other primes divide it. If this division were independent (in the sense of probability theory), there would be nothing left to prove, and alternative A above would be the only one left. Lemma 3, however, says that primes $p \leq h$, where h is the length of an interval I_n hence $h \approx u \log x$ have multiples in every I_r and act independently (with probability $1/p$ each) over the given range, or indeed any range of order not less than $x^{e/\log_2 a}$. Beyond this, it is not possible to go. The larger the primes, the less chance of several of them dividing an integer in the range. If independence in division were present, the Mertens theorem [3] would have given us for the prime number theorem $2e^{-\gamma}x \log x$, instead of $x/\log x$. This is taken by some to show that "probability methods do not apply in prime number theory," but is in fact irrelevant. The independence in probability of the number of primes in various $\pi_n(u)$, specified only by the index, not with *a priori* reference to the number of primes contained nor by knowledge of the initial point x_0, could still be a result of the sieve. The crucial question is: Given that a certain number of primes have actually occurred in a given stretch (i.e., a given number of consecutive I_n), what can then be said of the chances of primality anywhere else as affected or unaffected by this occurrence?

Directly, we are concerned only with deletions by primes $p \leq \sqrt{x}$. The small primes act independently over the range, by Lemma 3, as noted above. The only effect that the occurrence of a composite number can have is that its prime factors will not operate in the immediate neighborhood; for every such inoperative prime, the probability will be *locally* enhanced by a factor $1/(1 - 1/p)$. But a certain number of such dividing primes must become inoperative on the average over any stretch, while the probability for primality and expectation are always given overall by the prime number theorem. This means that unusually many inoperative primes may cause a local enhancement of the probability for primality—unusually few, a lowering of the probability for primality. Otherwise, nothing can be said. For deleting primes between h and $x^{-1/\log_2 z}$, the inoperative primes must be greater than $x^{1/\log_2 z}$ and the local factor can be calculated by packing the maximum possible a number of prime lost as close to $x^{1/\log_2 z}$ from above as possible. The extreme factors are thus easily shown to be bracketed by $(1 \pm \log_2^2 x/\log x)$. For deleting primes not exceeding $x^{1/k}$, $k > 2$ fixed, the loss or gain will be not greater in either direction than $(1 \pm k \log_2 x/\log x)/$. In each case, the sign makes the extreme factors compensatory, while smaller factors in any case cannot be cumulative. Finally, for stretches of length $\sqrt{x}$ or more all the deleting primes have multiples. Unusually many deletions mean unusually many factors higher than $\sqrt{x}$; again, the tendency cannot be cumulative, and the foregoing shows that the probability is changed by very little; asymptotically, not changed at all.

Theorem 1 (RH) *No sample-sequence* $\{\pi(x_0, u; n)\}$ *can lie within the exceptional set of probability measure zero with respect to the ULIL of Lemma 6, for any* $\varepsilon > 0$. *Whence all nontrivial zeros of* $\zeta(s)$ *lie on the vertical line* $s = \frac{1}{2} + it$.

Proof Starting with any x_0 and some fixed u derived from Lemma 2, map all sequences with initial points in I_1 onto (0, 1). The entire coset to be obtained by the displacement of any initial point in I_1 by an integral number of intervals in either direction is also mapped upon the same point of (0, 1). All members of a coset have clearly the same limiting-frequency properties. Probability measure is now taken as Lebesgue measure over the coset map on (0, 1). The probability can be calculated as the Lebesgue integral of the corresponding frequency. Thus, the basic distribution is Poissonian with parameter u, by Lemma 7. The distribution for k consecutive covering intervals amounts to that with covering intervals of length ku and is therefore Poissonian with parameter ku. As for the condition (2) of Lemma 6, we note that the expectation per stretch of length ku on the y-line is ku primes, and that, for large x, it is physically impossible for the deviation from this expectation to be as much as $\sqrt{2Nu \log_2 Nu}$, if k is small enough; e.g., the total stretch covered by ku consecutive intervals does not exceed $\sqrt{x \log x}$. Here, the P is zero hence less than with total independence, as one would expect from the compensatory effect found above. Whatever the extreme actually found in stretches of this order, one can repeat the arguments of Lemma 8. Thus, condition (2) of Lemma 6 is also satisfied by virtue of the non-cumulative effect of the sieve, and for every $\varepsilon > 0$. Therefore, $ULIL$ applies to almost all sample-sequences $\{\pi_n(u)\}$. For large N, the partial sums of the first N terms of any two sample-sequences whose initial points lie within the y-distance u of each other cannot differ by more than $u \log Nu$. Therefore, either all the sample-sequences satisfy $ULIL$, or none. The latter case is excluded, as the measure of the exceptional set would then have to be unity instead of zero. This proves the theorem and the Riemann hypothesis. □

Theorem 2 *The non-trivial zeros of all Dirichlet L-functions likewise lie on the vertical line* $s = \frac{1}{2} + it$.

This is the extended RH or the Piltz conjecture. The result is merely stated without proof, because of the same methods and arguments as above suffice. The consequences of these two theorems are given in books on specialized function theory [2] and advanced number theory [3]. Improvement of the inequalities (2) by the present methods would depend upon $LLIL$ and hence the *second* Borel–Cantelli lemma, which requires independence in probability.

The Poisson distribution of Lemma 2 allows many new results to be obtained directly. For example, *gaps of* y*-length* t *or more between consecutive primes have the distribution function* e^{-1}. However, it should be noted that the Poisson distribution is not essential for RH, which can be proved without any distribution at all, merely by taking the Poissonian as a bounding distribution for estimating the deviations of sums from expectation. Also, the Poisson distribution would not follow directly, granted RH. In other words, Lemma 8 is more important than Lemma 7.

Counter-examples of a fairly complicated nature could be produced which do not affect Lemma 7 nor the inequalities (2) but for which RH is false. These are formed by adding pseudo-primes and by striking out (in suitable stretches) sufficiently "thin" sequences of the primes. Such counter-examples do not affect our arguments because such changes in the series of prime numbers within the positive integers will block sieve deletion, invalidate the Euler product, and destroy unique factorization—all of which are essential to RH (as they are to our present arguments).

The result also indicates that the zeros of $\zeta(s)$ on the vertical line $\sigma = \frac{1}{2}$ should have a distribution of their own, presumably also the Poisson distribution. The proper transformation here must replace the coordinate t on the vertical line by the integral $\int \log t\, dt$ to the upper limit $T/2\pi$. The results will be considered elsewhere.

References

1. E.C. Titchmarsh, *The Theory of Functions* (Oxford, 1932), p. 152.
2. E.C. Titchmarsh *The Zeta-Function of Riemann* (Oxford, 1951), pp. 214–22. Chapters xiii-xiv give the consequences of RH.

3. K. Prachar *Primzahlverteilung*, (Berlin, 1957), pp. 247–55.
4. Littlewood, J.E. Sur la distribution des nombres premiers. *Comptes Rendus*, (Paris, 1914), vol 158, (pp. 72–1869). and note 3 above.
5. K. Prachar, *Loc. cit.,* Chapter ii.
6. K. Prachar, *Ibid*, p. 157. *et seq*.
7. G. Ricci, Rendiconti, Atti. Acad. Naz. Lincei, **8**, 96–192 and 51–347 (1954–55). Prof. P. Erdős was kind enough to inform me that the Ricci statement must be emended to: 'the set of cluster points for the y-distance between consecutive primes is of positive measure'. Even this suffices to prove Lemma 2, except that in this case, not every displacement however small need lead to a different sequence throughout the u-interval. This will be replaced by a set of sub-intervals of the u-interval, whose total length remains positive, and may then be mapped upon (0, 1).
8. G.H. Hardy, E.M. Wright, *An Introduction to the Theory of Numbers*, 2nd edn. (Oxford, 1945), pp. 349–54. Theorem 430.
9. G.H. Hardy, Wright, E.M. *Ibid*., Sections 9.12 and 9.13.
10. W. Feller, *Introduction to Probability Theory and Its Applications*, vol. 1, (New York, 1950), pp. 161–63. proof for finite base only.
11. W. Feller, *Ibid*, pp. 158–59. cf. Feller in Trans. Am. Math. Soc. **54**, 373–402 (1943).
12. W. Feller, *Ibid.*, p. 154.
13. D.D. Kosambi, The sampling distribution of primes. Proc. Nat. Acad. Sci. (U.S.A.), **49**, 20–23 (1963).
14. A.E. Ingham, Quart. J. Math. (Oxford) **8**, 255–66 (1937).

Chapter 25
Probability and Prime Numbers

S. Ducray

This paper, received in July 1964 and marked as being communicated by Sir C.V. Raman, then President of the Indian Academy of Sciences is one of four that DDK wrote under the Ducray pseudonym. One other paper was submitted to the same journal, and two were sent to the Journal of Bombay University. The paper was reviewed in Mathematical Reviews by J. Kubilius, a well-known number theorist and probabalist who noted that he "could not follow the proof of the cardinal Lemma 3."

This note sets up a sample-space connected with the infinite succession of prime integers. The properties of this sample-space cast fresh light upon some fundamental problems of analytic number theory.

Definitions—Let an arbitrary denumerable set of positive real numbers (not necessarily integers) be given: $0 < a_1 < a_2 \cdots$ with $a_n \to \infty$ as $n \to \infty$. For a fixed length $u > 0$, a *covering* of the real half-line $y > 0$ is given by the sequence of intervals $I_1, I_2, \ldots$ where $I_n : \; (n-1)\,u \leq y < nu$. $s_n = s(u, n)$ means that the number of points with co-ordinates $y = a_i$ contained in I_n. Thus, $\{s_n\}$ provides a sample-sequence for the particular covering which begins from $y = 0$. Other sequences similarly obtained by beginning the covering sequence from some other point of $y > 0$ or (what is the same thing) by subtracting the corresponding value from each a_i. Displacement of the initial point through an integral multiple ku of u gives the same sequence begun from the $(k + 1)$st term. We shall say that two sequences are essentially different if they do not coincide after a finite number of terms of one are omitted.

Published in Proceedings of the Indian Academy of Sciences **60**, 159–64 (1964). Reprinted with permission.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_25

Lemma 1 *If the sequence $\{a_i\}$ has the properties: (a) that there are infinitely many gaps $a_{r+1} - a_r > 2u$ and (b) the a_i are Gleichverteilt modulo u, then the number of different sample-sequences obtained by displacement of the initial point through distances not exceeding u can be mapped in a 1–1 manner upon the right-open unit interval* $0 \leq y < 1$.

Proof Suppose that, throughout some displacement $w < u$, the same sequence $\{s_n\}$ is obtained. Then the number of points $y = a_i$ lost to the left by a displaced interval must be precisely equal to that gained from the right, hence the same for all intervals. But the gaps ensure that no matter where the initial point be taken, there are always intervals with zero gain and loss. Hence, the number gained or lost must always be zero throughout the displacement $w > 0$. This means a regular gap of length w in the numbers a_i as reduced modulo u, which contradicts hypothesis b. Therefore, $w = 0$ and there is a different sample-sequence for every point of $(0, u]$, which interval may then be projected upon $(0, 1]$. □

Lemma 2 *If, in Lemma 1, condition b be replaced by requiring only that the set of cluster-points of $\{a_i\}$ modulo u be of positive measure, it still follows that a sub-set of the distinct sample-sequences $\{s_i\}$ obtained by displacement of the initial point through not more than one u-interval may be mapped in a* $1 - 1$ *manner upon* $(0, 1]$.

Proof Now, it may be possible to obtain the same sequence for some positive displacement w, as gaps among the cluster points are permitted. Let w_1 be the limit superior of such displacements, beginning from $y = 0$. If, thereafter, the cluster-points modulo u are dense throughout some sub-interval $(w, w_1 + h)$, there will be a different sample-sequence for every point of this sub-interval, and the theorem is proved. If not, the cluster-points near w can be covered by an interval of arbitrary small length $\varepsilon > 0$, and we proceed to the next cluster-point outside this small sub-interval, say w_2. Again, cover this with an interval $\varepsilon/2$, then $\varepsilon/4$ and so on, if at no stage is a finite interval of density attained for the cluster points. But then the set of cluster-points modulo u must be of measure zero, as their total measure cannot exceed 2ε, arbitrarily small. This proves the lemma by contradicting the hypothesis. □

Definition In what follows, we take $1, 2, \ldots, n, \ldots$ as the positive integers marked off at unit intervals on the half-line $x > 0$. Let $Li(x)$ be the integral $\int dt/\log t$ to the upper limit x. Take $y = Li(x) - Li(x_0)$ for any $x_0 \geq 2$. Our set $\{a_i\}$ is the image-set on this y-line of the prime numbers $x = 2, 3, 5 \ldots p \ldots$. Every covering is always to begin with $y = 0$, but the initial point may be varied by displacement of x_0. Thus, $s_n = s(x_0, u; n)$ is the number of primes included in the x-image of $I_n : (n-1)u \leq y < nu$.

Theorem 1 *There exists at least one* $u > 0$ *such that a subset of the different sample-sequences (of prime images) $\{s_n\}$ obtained by displacement of the initial point continuously through the image of a single covering interval may be mapped in a* $1 - 1$ *manner upon* $(0, 1]$.

Proof Condition a of Lemma 1 is satisfied by the Erdős [1] gap-theorem. On the y-line, there are infinitely many gaps greater than $f(y)$ between the images of consecutive primes, where f tends rather slowly to infinity with y; hence the gaps are greater than any Au, for arbitrary constants A and u. Condition b of Lemma 1 is apparently satisfied by a whole range of u-values, according to a result of Ricci [2]. However, P. Erdős [3] has pointed out that the result actually proven shows only the existence of a positive measure for the set of cluster-points on the y-line (modulo any $u > 0$) for the images of the primes. This in any case satisfies the requirements of Lemma 2, which suffices. □

Hereafter, take u to be one of the particular values under Theorem 1. Then make a *canonical mapping* onto (0, 1] of (almost all) covering sequences with initial points in some one fixed u-interval. Every sequence obtained by displacement of the initial point through any integral number of converting intervals in either direction is to be mapped on the same point of (0, 1]. Every sequence thus mapped upon a given point of (0, 1] has then the same limiting frequency properties. That is, if the number of intervals of the sequence covering $0, 1, 2, \ldots, k, \ldots$ image of primes reaches some limiting proportion f_k for one sequence associated with a point, it does so for every sequence mapped upon that point. *Random choice* of a sequence is defined as follows: first, all points of (0, 1] have an equal chance of being chosen (uniform distribution on the map). Then, for a given point of the map, the actual sequence may be begun from any term whatever, counting that term as s_1, the next as s_2, and so on. This eliminates x_0 altogether from consideration, and we may speak only of properties of the sequences associated with points of the canonical map, using Lebesgue measure on the map for probability.

In this situation, probability concepts apply to the $\{s_n\}$.

Theorem 2 *For all sequences $\{s_n\}$ defined as above, the mean value (expectation) is given by $E(s_n) = u$ for all n.*

Proof This is an immediate and obvious consequence of the prime number theorem, and of the method of choice of the sequences, seeing that s_n can be the number of primes covered by any interval of length u anywhere on the half-line $y > 0$. Here, the prime number theorem is taken for granted, in the form $\pi(x) \sim Li(x)$, where $\pi(x)$ is the number of primes $p \leq x$. □

There arise two cases, according to whether the random variables s_i are independent in the sense of probability theory or not.

Theorem 3 *Should the consecutive s_r of the same sequence (chosen at random, as above) be independent in probability, then the following results are true with unit probability (i.e., for almost all points of the canonical map).*

3.1. The probability for any s_i assuming the value k is given by $P(s_i = k) = e^{-u}u^k/k$ for $k = 0, 1, 2, \ldots$.

3.2. If $S_n = s_1 + s_2 + \cdots s_n$, then for any arbitrary $\varepsilon > 0$,

$$-(1+\varepsilon)\sqrt{2Nu \log\log Nu} < S_N - Nu < (1+\varepsilon)\sqrt{2Nu \log\log Nu}$$

for all except a finite number of values of N.

3.3. If, in 3.2, ε be replaced by $-\varepsilon$, then each of the two inequalities is false infinitely often as $N \to \infty$.

Proof The first of these, namely 3.1, is equivalent to a result published by Kosambi [4] showing the primes on the y-line to be in a Poisson distribution with parameter u. The result is almost obvious under the given conditions.

With the Poisson distribution and complete independence in probability, textbook [5] methods lead immediately to the other two results. Of these, 3.2 is the *upper* law of the iterated logarithm, and 3.3 the *lower* law of the iterated logarithm.

However, independence in probability is not easy to prove for consecutive terms of our sample-sequences. Nevertheless, the sieve of Eratosthenes in its most elementary form enables the most important and useful part of the theorem, namely 3.2, to be carried over. This is best done in two stages: □

Lemma 3 *For large N and each k, $1 \le k \le N$, the probability of k being the first index for which $|S_k - ku| \ge \sqrt{2Nu \log\log Nu}$ cannot exceed the same probability as calculated under the assumption of independence of the s_i as in Theorem 3.*

Proof The main idea is that the sieve of Eratosthenes prevents very large deviations from expectation from accumulating, if it has any effect at all upon independence of $(s_n - u)$ in the sense of probability theory. □

If the position on the x-line were known, the primes in the image of k consecutive u-intervals would be completely determined. As it is, all that can be said is that there exists an unknown background parameter x such that $Nu \sim x/\log x$ for large N. The primes about x on the x-line are the numbers not deleted by the sieve, i.e., the numbers not multiples of any primes $p \le \sqrt{x}$. It is known that a connected stretch of length h on the x-line can contain at most $ch/\log h$ primes, where c is an absolute constant. A length ku on the y-line has an image $\sim ku \log x$. Therefore the probability in Lemma 3 is zero for $k \le C\sqrt{x}(\log\log x/\log x)^{3/2}$, using the asymptotic values for N and x.

The question of independence now appears in the following manner. Given that r primes have in fact occurred in a specific number of consecutive intervals; are the chances of some number m of primes occurring in a pre-assigned number of following intervals increased or decreased thereby, or remain unaffected–with no other information available. The answer is worked out as follows:

For every composite number that occurs, each prime factor $< \sqrt{x}$ cannot act as deleting prime for the corresponding distance on either side. The prime number theorem and the existence of an expectation say that the probability for an integer in the x-image of a single interval being a prime is $1/\log x$, in order. If an unusually large number of primes turn up in a given stretch, this means unusually few composite numbers, unusually few $p \le \sqrt{x}$ inactivated as deleting primes, and so, if the probability for primality is affected at all, a slight decrease therein. In the opposite direction, unusually few primes may mean more than a fair share of deleting primes dropping out of action; hence, possible enhancement of the probability for primality in adjacent stretches. Nothing more can be said, provided of course, that the stretch where the known number of primes have turned up is of x-length less than $\sqrt{x}$ in order. For greater x-lengths, all deleting primes will delete in the stretch. The most that can then be said is that unusually many of these multiply each other when the number of primes left in the stretch is well above expectation; and the opposite when the number of primes covered by the stretch is far below expectation. In neither case can the same phenomenon be expected to continue over the next stretch. So, *the effect of dependence, if any, upon $S_n - nu$ may be compensatory, but never cumulative*. This proves the lemma.

Theorem 4 *If $\pi(x)$ be defined as the number of primes $p \leq x$, then*

$$\pi(x) - Li(x) = O(\sqrt{x \log\log x / \log x}) .$$

Proof With independence in probability, the result 3.2 and the asymptotic values for N and x prove the result immediately. With dependence, the estimates of Lemma 3 still remain valid. This is the key condition for the validity of the *upper* law of the iterated logarithm, which is based upon the first Borel-Cantelli [6] lemma and hence does not require independence (which is only a sufficient condition). Thus, regardless of the validity of 3.1 and 3.3, the result 3.2 still remains true, and the inequalities may at most be strengthened, never weakened [7].

Theorem 3, however, may admit an exceptional set of measure zero, like any such unit-probability result. It remains to show that this must be empty for the particular sample-sequences of primes in covering intervals. Consider two I_n whose coverings do not differ by more than u on the y-scale. The number of integers covered by any such I_n is not greater than $C \log n$ for large n. Therefore, the difference in the number of primes S_N for two different sample-sequences cannot be of greater order than $\log N$. This does not affect the order of magnitude as given in 3.2 which is therefore true for all covering sequences without exception. The result as translated here is thus proved. □

The consequences of Theorem 4 are sufficiently well known to number-theorists and need not be detailed here.

References

1. K. Prachar, *Primzahlverteilung* (Berlin, 1957), p. 157 ff.
2. G. Ricci, Sul pennello di quasi-asintoticità della differenza di interi primi consecutivi. Atti. Accad. naz. Lincei (Rend.), Série **8**, **17**, 96–192 and 51–347 (1954–55).
3. In a private communication of Prof. P. Erdős to Prof. D.D. Kosambi.
4. D.D. Kosambi, The sampling distribution of primes. Proc. Nat. Acad. Sci. (USA) **49**, 20–23 (1963).
5. W. Feller *An Introduction to Probability Theory and its Applications*, vol. 1, (New York, 1950), pp. 61–157. for the laws of large numbers and of the iterated logarithm.
6. W. Feller, *ibid.*, p. 154.
7. S. Ducray, Normal Sequences, to appear in the *J. Uni. Bombay*.

Part III
Select Publications of D.D. Kosambi in Other Languages

DDK's Publications in Other Languages

Kosambi published mainly in English, but also occasionally in French and German. An article of his was translated into Japanese and one article appeared originally in Chinese. The three articles with titles in boldface are reproduced in their entirety.

- In German:
  1. **Affin-geometrische Grundlagen der Einheitlichen Feldtheorie**
     Sitzungsberichten der Preussische Akademie der Wissenschaften, Physikalisch-mathematische klasse **28**, 342–45 (1932)
- In French:
  1. *Geometrie differentielle et calcul des variations*,
     Rendiconti della Reale Accademia Nazionale dei Lincei **16**, 410–15 (1932)
  2. *Les metriques homogenes dans les espaces cosmogoniques*
     Comptes Rendus **206**, 1086–88 (1938)
  3. **Les espaces des paths generalises qu'on peut associer avec un espace de Finsler**
     Comptes Rendus **206**, 1538–1541 (1938)
  4. *Sur la differentiation covariante*
     Comptes Rendus **222**, 211–13 (1946)
  5. *Les invariants differentiels d'un tenseur covariant a deux indices*
     Comptes Rendus **225**, 790–92 (1947)
- In Japanese:
  1. *The tensor analysis of partial differential equations*
     Tensor, **2**, 36–39 (1939)
- In Chinese:
  1. **The method of least-squares**
     Advancement in Mathematics **3**, 485–491 (1957)

Chapter 26
Affin-geometrische Grundlagen der einheitlichen Feldtheorie

Von D.D. Kosambi, in Aligarh (Indien)

DDK's work on path-geometry started in Aligarh with [DDK3] *that was submitted to the Indian Journal of Physics,* [DDK5] *in French and this paper in German were basically expositions designed to present his work to a European audience and also, as it appears, to assert that these ideas were presented at a seminar on 5 March 1931. DDK had an extensive collection of scientific books in German and knew the language well; during his brief stay in Banaras prior to the writing of this paper, he had been teaching German language classes in addition to mathematics* [DDK-JK].

In einer früheren Arbeit[1] habe ich den Versuch gemacht, eine möglichst allgemeine Theorie der affinen Bahngeometrie aufzubauen. Dadurch wird auch die einfachste geometrische Grundlage zu den neueren einheitlichen Feldtheorien[2] geschaffen. Der erste Ansatz zu dieser Auffassung wurde durch die Bemerkung von. P. Straneo gegeben, daß die Autoparallelen von den geodätischen Linien zu unterscheiden sind. In einer rein-affinen Theorie erscheinen in der Tat zwei verschiedene Arten von Parallelismen, die sich aus einer nicht-distributiven Vektorderivierten bzw. aus der entsprechenden distributiven Derivierten ergeben. Dadurch und durch de Annahme der Existenz einer kovarianten Ableitung werden Fünfervektoren überflüssig gemackt, und die Tensoren der einheitlichen Theorie sowie neue bisher physikalisch nich gedeutete Tensoren ergeben sich ohne weiteres.

[1] D.D. Kosambi, Modern Differential Geometries, erschient demnächst in Indian Journal of Physics (1932). Diese Grundzüge dieser Theorie wurden im Mathmatischen Seminar der Aligarh Universität am 5. März 1931 vorgetragen.

[2] P. Straneo, diese Sitzungber., 1931, S. 319–325; Einstein und Mayer, ib., 1931, S. 541–557.

Published in Sitzungsberichten der Preussische Akademie der Wissenschaften, Physikalisch-mathematische klasse **28**, 342–45 (1932).

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_26

1. Es werde ein System von Bahnkurven zugrunde gelegt, die als Lösungen eines allgemeinen Systems von Differentialgleichungen zweiter Ordnung

$$\ddot{x}^i + \alpha^i(x, \dot{x}, t) = 0 \qquad (i = 1, 2, \ldots n) \tag{26.1}$$

gegeben werden. Dabei bedeuten die x Punktkoordinaten, und t einen (will-kürlichen) Bahnparameter.

Die Vektorderivierte $D(u)^i$ längs einer willkürlichen Kurve möge nun folgendermaßen definiert werden:

$$D(u)^i = \dot{u}^i + u^k \gamma^i_k(x, \dot{x}, t) + \varepsilon^i(x, \dot{x}, t), \tag{26.2}$$

wobei:

$$\varepsilon^i = \alpha^i - \gamma^i_k \dot{x}^k. \tag{26.3}$$

Die Parallelverschiebung eines Vektors u^i wird durch $D(u)^i = 0$ erklärt. Infolge (26.3) sind die Bahnkurven (26.1) autoparallele Linien, d. h. $D(\dot{x})^i = 0$.

Da wir verlangen, daß $D(u)^i$ gleichzeitig mit u^i zu einem kontravarianten Vektor wird, lautet das Transformationsgesetz der γ^i_k folgendermaßen:

$$\bar{\gamma}^i_k \frac{\partial \bar{x}^k}{\partial x^j} + \frac{d}{dt}\frac{\partial \bar{x}^i}{\partial x^j} = \gamma^k_j \frac{\partial \bar{x}^i}{\partial x^k}. \tag{26.4}$$

Die γ^i_k werden keiner anderen Beschränkung untworfen. Es folgt unmittelbar, daß ε^i auch ein Vektor ist. Durch das Weglassen der ε^i wird eine distributive Vektorderivierte (die Nebenderivierte) erzeugt:

$$\mathfrak{D}(u)^i = \dot{u}^i + \gamma^i_k u^k \tag{26.5}$$

mit einem entsprechenden Nebenparallelismus.

2. Wir machen jetzt die weitere Annahme, daß jedes Vektorfeld $u^i(x)$ eine von der Kurvenrichtung $\dot{x}^i$ unabhängige kovariante Ablietung $u^i_{|r}$ besitzt, durch welche die Vektorderivierte nach der üblichen Regel erzeugt wird:

$$D(u)^i = u^i_{|r} \dot{x}^r. \tag{26.6}$$

Dafür ist es notwendig und hindreichend, daß:

$$\left.\begin{aligned} u^i_{|r} &= \frac{\partial u^i}{\partial x^r} + \gamma^i_{kr} u^k + \varepsilon^i_r \\ \gamma^i_k &= \gamma^i_{kr} \cdot \dot{x}^r \\ \varepsilon^i &= \varepsilon^i_r \cdot \dot{x}^r \end{aligned}\right\} \tag{26.7}$$

wo die γ^i_{kr} und ε^i_r von den $\dot{x}^i$ unabhängig sein müssen. Daraus ergibt sich:

$$\alpha^i = \gamma^i_{kl}\dot{x}^k\dot{x}^l + \varepsilon^i_r\dot{x}^r. \tag{26.8}$$

Diese sind also die allgemeinsten Übertragungen, die eine richtungsunabhängige kovariente Ableitung ermöglichen.

3. Der x Torsionstensor ist jetzt:

$$\Omega^i_{kl} = \gamma^i_{kl} - \gamma^i_{lk} \tag{26.9}$$

Wir setzen außerdem:

$$2\Gamma^i_{kl} = \gamma^i_{kl} + \gamma^i_{lk}. \tag{26.10}$$

Weitere Tensoren ergeben sich, wie a. a. O.[3] gezeigt wirde, aus Integrabilitätsbedingungen. Zunächst erhält man:

$$S^i_j = -\alpha^i_{,j} + \frac{1}{2}\frac{\partial\alpha^i_{,k}}{\partial\dot{x}^j}\dot{x}^k + \frac{1}{2}\frac{\partial^2\alpha^i}{\partial\dot{x}^j\partial t} - \frac{1}{2}\frac{\partial^2\alpha^i}{\partial\dot{x}^j\partial\dot{x}^k}\alpha^k + \frac{1}{4}\frac{\partial\alpha^i}{\partial\dot{x}^k}\frac{\partial\alpha^k}{\partial\dot{x}^j} \tag{26.11}$$

(wie üblich, wird gesetzt: $f^{\cdots}_{\cdots,k} = \frac{\partial f^{\cdots}_{\cdots}}{\partial x^k}$).

Wird, weiter S^i_{jk} durch

$$3S^i_{jk} = \frac{\partial S^i_k}{\partial\dot{x}^j} - \frac{\partial S^i_j}{\partial\dot{x}^k} \tag{26.12}$$

definiert, so ist

$$R^i_{jkl} = \frac{\partial S^i_{jk}}{\partial\dot{x}^l} \tag{26.13}$$

(bei geeigneter Anordnung der Indizes) der gewöhnliche Riemann-Christoffelsche Krümmungstensor.

In unserem Falle, wo die α^i die Bedingungen (26.8) erfüllen, hat man:

$$S^i_j = K^i_{jkl}\dot{x}^k\dot{x}^l + W^i_{jk}\dot{x}^k + V^i_j. \tag{26.14}$$

Dabei sind:

$$\left.\begin{aligned}
K^i_{jkl} &= \frac{1}{2}(\Gamma^i_{jk,l} + \Gamma^i_{jl,k}) - \Gamma^i_{kl,j} + \frac{1}{2}(\Gamma^i_{kr}\Gamma^r_{jl} + \Gamma^i_{lr}\Gamma^r_{jk}) - \Gamma^i_{jr}\Gamma^r_{kl}\\
W^i_{jk} &= -\varepsilon^i_{k,j} + \frac{1}{2}\varepsilon^i_{j,k} - \varepsilon^r_k\Gamma^i_{jr} + \frac{1}{2}\varepsilon^i_r\Gamma^r_{jk} + \frac{1}{2}\varepsilon^r_j\Gamma^i_{rk} + \frac{\partial\Gamma^i_{jk}}{\partial t}\\
V^i_j &= \frac{1}{2}\frac{\partial\varepsilon^i_j}{\partial t} + \frac{1}{4}\varepsilon^i_r\varepsilon^r_j
\end{aligned}\right\} \tag{26.15}$$

[3] Siehe Anm. 1 auf S. 342.

Der Riemann-Christoffelsche Krümmungstensor hängt nur von K^i_{jkl} ab und wird leicht berechnet. Wir definieren nun den weiteren Tensor:

$$\varepsilon^i_{jk} = \varepsilon^i_{j,k} + \varepsilon^r_j \Gamma^i_{rk} - \varepsilon^i_r \Gamma^r_{jk} \tag{26.16}$$

(das ist nur die gewöhnliche kovariante Ableitung der ε^i_j für den symmetrischen Zussamenhang Γ^i_{jk}).

Dann läßt sich die folgende Identität leicht verifizieren:

$$W^i_{jk} = \frac{1}{2}\varepsilon^i_{jk} - \varepsilon^i_{kj} + \frac{\partial \Gamma^i_{jk}}{\partial t}. \tag{26.17}$$

4. Die Gleichungen der in der neueren Zeit von verschiedenen Autoren vorgeschlagen Feldtheorien können mittels unserer Tensoren ausgedrückt werden, wie wir am Beispiel der Einstein-Mayerschen Theorie erklären wollen, wenn man noch eine Riemannsche Grundform zu Hilfe nimmt, die das Herauf-bzw. Herunterziehen von Indizes ermöglicht; dabei wird ein vierdimensionaler Raum zugrunde gelegt, und es wird auch angenommen, daß der Bahnparameter t nirgends explizit vorkommt. Die jetzt eingeführte Grundform wird mit unserem Parallelismus durch die weitere Annahme in Beziehung gesetzt, daß die geodätischen Linien der Grundform mit den ≫Nebenbahnkurven≪

$$\ddot{x}^i + \gamma^i_{kl}\dot{x}^k\dot{x}^l \equiv \ddot{x}^i + \Gamma^i_{kl}\dot{x}^k\dot{x}^l = 0 \tag{26.18}$$

zusammenfallen. Allgemeiner könnte man aber annehmen, daß die Nebenbahnkurven Extremalen eines nich-entarten Variationsproblemes sind, desen Integrand alsdann als Metrik benutzt werden kann.

Wir setzen also

$$\Gamma^i_{jk} = \begin{Bmatrix} i \\ jk \end{Bmatrix}. \tag{26.19}$$

Der Einstein-Mayersche Tensor F^i_j wird in unserer Schreibweise:

$$-\varrho F^i_j = \varepsilon^i_j. \qquad (\varrho = \text{Konstante}) \tag{26.20}$$

Die Einstein-Mayerschen Gleichungen lauten nun, unter Benutzung der Tensoren W^i_{jk}, V^i_j:

$$\left.\begin{aligned} &(a)\ \ \varepsilon_{ik}\varepsilon_{ki} = 0, \quad \varepsilon_{ik} = g_{ir}\varepsilon^r_k \\ &(b)\ \ W_{ijk} + W_{jki} + W_{kij} = 0, W_{ijk} = g_{ir}W^r_{jk} \\ &(c)\ \ \varepsilon^k_{ik} = 0 \\ &(d)\ \ \frac{\varrho^2}{4}(R^i_j - \frac{1}{2}\delta^i_j R) + (V^i_j - \frac{1}{4}\delta^i_j V) = 0, \quad V = V^k_k. \end{aligned}\right\} \tag{26.21}$$

Wenigstens ein Teil dieser Gleichungen is nun rein-affin. Zunächst is dies der Fall für Gleichung (26.21c). Weiter folgt aus (26.21a):

$$\varepsilon^{k}_{ki} = 0 \tag{26.22}$$

oder, mit Hilfe von (26.17) ausgedrückt und unter Beifügung von (26.21c):

$$W^{k}_{ki} = 0, \quad W^{k}_{ik} = 0, \tag{26.23}$$

wodurch (26.21c) ersetzt werden darf.

In bekannten Spezialfall, wo die absolute Krümmung identisch verschwindet, bekommt man statt (26.21d) die weiteren rein-affinen Gleichungen:

$$\left.\begin{aligned} R_{ij} &= 0 \\ V^{i}_{j} - \frac{1}{4}\delta^{i}_{j} \cdot V &= 0. \end{aligned}\right\} \tag{26.24}$$

In diesem Spezialfall gilt, wie aus (26.23) and (26.24) folgt:

$$S_i = S^{k}_{ki} = 0, \tag{26.25}$$

was aber nicht genügt, um diesen Fall zu charakterisieren.

Wie ersichtlich, bekommt man sämtliche zu benutzende Tensoren durch die Annahme, erstens, der Existenz einer kovarianten Ableitung, die den zugrunde gelegten Parallelismus erzeugt, und, zweitens, der Ableitbarkeit des dazu gehörigen Nebensparallelismus aus einer Grundform. Wie oben bemerkt wurde, hängt die Verallgemeinerung der zweiten Annahme von der Lösung des Umkehrproblemes der Variationsrechnung ab, worüber ich meine Resultate anderswo zu veröffentlichen beabsichtige. Deshalb möge es auch dahingestellt bleiben, ob überhaupt eine rein-affine Feldtheorie möglich sei: dabei wäre auch gewiß der Torsionstensor zu benutzen, der, wie bekannt, in den früheren Theorien von Einstein eine wichtige Rolle spielte.

Chapter 27
Les Espaces des Paths Généralisés Qu'on Peut Associer Avec Un Espace de Finsler

Note de M. Damodar Kosambi, présentée par M. Élie Cartan

This is one of four papers DDK published in Comptes Rendus, of five that appear to have been originally written in French. The other paper was in Rendiconti della Reale Accademia Nazionale dei Lincei. DDK's French connection appears to have been particularly strong, doubtless reinforced by André Weil who remained in contact with Kosambi till late in the 1950's. The geometric methods introduced in the KCC theory have some similarity to the study of geodesics in a Finsler space, and Kosambi pursued these analogies extensively in his series of papers on path spaces.

Soit $\mathbf{K}_n$ un espace des *paths* généralisé défini par les équations

$$\ddot{x}^i + \alpha^i(x, \dot{x}, t) = 0 \qquad (i = 1, 2, \ldots, n; \dot{x} = \frac{dx}{dt}; \ddot{x} = \frac{d^2x}{dt^2}). \tag{27.1}$$

Regardons le paramètre t comme une variable nouvelle, $t = x^0$, et introduisons un autre paramètre s. Les équations (27.1) deviennent

$$x''^i - \frac{x'^i}{x'^0}x''^0 + (x'^0)^2\alpha^i(x, \frac{x'^j}{x'^0}, x^0) = 0 \quad (x'^i = \frac{dx^i}{ds} = \frac{x'^i}{x'^0}, \ldots). \tag{27.2}$$

En prenant une fonction arbitraire $\beta(x, \dot{x}, t)$ avec $x''^0 + (x'^0)^2\beta = 0$ nous aurons le système de $n + 1$ équations

$$\begin{aligned} x''^i + \mathbf{A}^i(x, x') &= 0 \qquad (i = 0, 1, 2, \ldots, n), \\ \mathbf{A}^0 = \beta(x'^0)^2, \quad \mathbf{A}^i &= (\alpha^i + \beta\dot{x}^i)(x'^0)^2 \qquad (i \neq 0). \end{aligned} \tag{27.3}$$

Published in Comptes Rendus **206**, 1538–41 (1938).

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_27

Or, les $\mathbf{A}^i$ étant homogènes de degré 2 en x', et non paramétriques, les équations (27.3) définissent un espace $\mathbf{B}_{n+1}$ du type Berwald-Douglas, dans lequel nous dirons que le $\mathbf{K}_n$ a été plongé. La géométrie du $\mathbf{K}_n$, ne correspond pas à celle du $\mathbf{B}_{n+1}$, parce qu'il n'existe aucune méthode intrinsèque pour associer un tenseur quelconque du $\mathbf{K}_n$ biunivoquemenr à un tenseur du $\mathbf{B}_{n+1}$. Nous nous bornerons donc à la résolution de la question: peut-on déduire les équations (27.3) formellement d'un problème variationnel $\delta\int \mathbf{F}ds = 0$ quand les (27.1) peuvent se déduire de $\delta \int f\,dt = 0$, et *vice versa*?

On aurait cru qu'à la métrique $f[x, \dot{x}, t]$ pour $\mathbf{K}_n$ doit toujours correspondre la métrique $\mathbf{F} = f[x, x^{'i}/x^{'0}, x^0]x^{'0}$ dans $\mathbf{B}_{n+1}$ et réciproquement. Mais cette fonction $\mathbf{F}$ étant homogène de degré 1 en $x^{'}$, $|\partial^2\mathbf{F}/\partial\dot{x}^i\partial\dot{x}^j| = 0$, et les équations covariantes eulériennes de $\delta\int \mathbf{F}ds = 0$ ne peuvent pas être résolues sous la forme contravariante (27.3). Comme d'habitude pour les espaces de Finsler, nous devons remplacer $\mathbf{F}$ par $\mathbf{F}^2$ ou par une autre fonction de $\mathbf{F}$; cela est permis seulement si $\mathbf{F}$ satisfait au système d'équations

$$\frac{\partial \mathbf{F}}{\partial x^i} - \frac{1}{2}\frac{\partial \mathbf{A}^r}{\partial x'^i}\frac{\partial \mathbf{F}}{\partial x'^r} = 0 \quad (i = 0, 1, 2 \ldots, n.) \tag{27.4}$$

Ces équations ont la propriété qu'une fonction quelconque des solutions est encore une solution et, ici, il suffit d'en chercher les solutions homogènes en $x^{'}$. Pour un degré d'homogénéité k non nul, nous pourrons toujours substituer la métrique de Finsler $\mathbf{F}^{1/k}$.

Alors, pour discuter le problème du point de vue de l'espace $\mathbf{K}_n$, nous prenons $\mathbf{F} = \Phi(x, \dot{x}, t)(x^{'0})^k$ où $k = 0$, 1 seulement. Les équations (27.4) prennent la forme

$$\mathrm{D}\Phi = k\beta\Phi, \qquad \Phi_{|i} - \frac{\beta}{2}\Phi_{;i} = \frac{k}{2}\Phi\beta_{;i} \tag{27.5}$$

où nous avons posé

$$\Phi_{;i} = \frac{\partial\Phi}{\partial\dot{x}^i}, \quad \Phi_{,i} = \frac{\partial\Phi}{\partial x^i}, \quad \mathrm{D}\Phi = -\alpha^r\Phi_{;r} + \dot{x}^r\Phi_{,r} + \frac{\partial\Phi}{\partial t}$$
$$\Phi_{|i} = \Phi_{,i} - \frac{1}{2}\alpha^r_{;i}\Phi_{;r}.$$

Pour $k = 1$, nous pouvons éliminer les termes en β et nous trouvons

$$(\mathrm{D}\Phi)_{;i} - 2\Phi_{|i} \equiv \frac{d}{dt}\Phi_{;i} - \Phi_{,i} = 0. \tag{27.6}$$

Cela signifie que Φ satisfait aux équations eulériennes de $\delta\int\Phi dt = 0$, et fournit une métrique pour $\mathbf{K}_n$ si $|\Phi_{i,j}| \neq 0$. Nous avons donc le théorème.

Théorème I. — *Pour que l'espace associé* $\mathbf{B}_{n-1}$ *admette une métrique régulière de Finsler* $F(x, \dot{x})$, *il faut et il suffit qu'une métrique quelconque* Φ *existe pour l'espace* K_n. *La fonction* β, *dans* (27.3), *sera donc uniquement définie par*

$$\beta = \frac{\mathrm{D}\Phi}{\Phi} \tag{27.7}$$

pour chaque métrique Φ du K_n.

Pour le cas $k = 0$, il n'en est pas ainsi. Si $\beta \neq 0$, l'espace K_n n'admet pas la métrique Φ, et nous avons un cas très intéressant. En effet, on obtient facilement le théorème:

Théorème II. —*S'il existe une fonction non identiquement nulle pour laquelle les équations* $D\Phi = \Phi_{i1} - (\beta/2)\Phi = 0$ *ont une solution avec* $\neq 0$, *l'espace* $\mathbf{K}_n$, *métrique ou non, peut être plongé dans* $\mathbf{B}_{n+1}$ *avec une métrique homogène de degré zéro en* x'.

Dans ce cas, pour avoir une métrique de Finsler, on doit répéster le proceédé en plongeant $\mathbf{B}_{n+1}$ dans un $\mathbf{B}_{n+2}$.

Les équations d'intégrabilité de (27.5) sont

$$\begin{gathered}
\big(\mathrm{P}^r_i + \frac{1}{2}\delta^r_i\{\mathrm{D}\beta - \frac{1}{2}\beta^2\}\big)\Phi_{;r} = k\Phi\{\beta_{|i} - \frac{1}{2}D\beta_{;i} - \frac{1}{4}\beta\beta_{;i}\};\\
\big(-2R^r_{ij} + \delta^r_j\beta_{|i} - \delta^r_i\beta_{|j} + \frac{1}{2}\beta\{\delta^r_i - \delta^r_j\beta_{;i}\}\big)\Phi_{;r} = k\Phi\{\beta_{;i|j} - \beta_{;j|i}\}; \qquad (27.8)\\
\mathrm{P}^i_j = -\alpha^i_{;j} + \frac{1}{2}\frac{d}{dt}\alpha^i_{;j} + \frac{1}{4}\alpha^i_{;r}\alpha^r_{;j}; \qquad \mathrm{R}^i_{jk} = \frac{1}{3}(\mathrm{P}^i_{j;k} - \mathrm{P}^i_{k;j}).
\end{gathered}$$

Comme tout $\mathbf{K}_1$, admet une métrique, nous avons un corollaire du théorème I: *Tout* $\mathbf{K}_1$, *peut être plongé dans un* $\mathbf{B}_2$ *avec une métrique de Finsler*.

II y a des différences importantes entre $\mathbf{K}_n$ et $\mathbf{B}_{n+1}$: les équations (27.1) possèdent des solutions à $2n$ constantes arbitraires (à $2n + 2$ pour (27.3)). Les groupes les plus généraux de transformations ponctuelles, pour lesquels (27.1) et (27.3) se transforment par la loi tensorielle ne sont pas les mêmes. Pour les problèmes du calcul des variations, il y a la différence entre l'extremum faible et l'extremum fort. Nous avons suivi les méthodes purement formelles, ayant trouvé dans une autre Note [1] que les deux cas signalés dans les deux théorèmes peuvent effectivement se présenter.

Reference

1. D.D. Kosambi, Compt. Rendus **206**, 1086 (1938). Voir aussi D.D. Kosambi (Oxford), Q. J. Math. **6**, 1–12 (1935). pour les méthodes et notations employées ici.

Chapter 28
最小二乘法 The Method of Least Squares

高善必
（印度）

Kosambi
(India)

During the 1950s and early 1960s, DDK visited China several times on exchange programmes and this paper was probably written when he visited the Academia Sinica on an exchange programme between India and China as an expert in statistics from TIFR [DDK-JK].

The paper, published in Advancement in Mathematics **3**, *485–491 (1957), first appeared in Chinese, and a facsimile of the article is given in the following pages. It is interesting to note that in Chinese, the author's name (see above) is composed of three characters that can be pronounced "gāo-shan-bi" in Mandarin. In all likelihood, DDK would have chosen these himself. The characters individually have the meanings*[1] *"tall/ high", "charitable/ kind", and "surely/ must" (in the sense of necessarily).*

[1] I am grateful to Profs. Zhou Changsong (Hong Kong Baptist University) and Takatsuka Kazuo (Fukui Institute of Fundamental Chemistry, Kyoto) for consultation on this.

R. Ramaswamy (ed.), *D.D. Kosambi*, DOI 10.1007/978-81-322-3676-4_28

最小二乘法[*]

高善必

（印度）

本文第一節開始討論具獨立隨機變數的概率空間中的可能有的度量；證明了歐氏度量(在適當的坐標系下)是惟一可被允許的．而最小二乘法大家知道是從這種距離的概念中導出的．在第二節中對抽象空間中的一般綫性方程系統導出了一組惟一的最小二乘的解，甚至在普通意義下沒有正常的解的情形也仍有這種解，而當普通意義下的解存在時，則這兩種解合一．最後一節給出了對一般非綫性方程系統的推廣的概述．

1. 我們由被稱為“簡單事件”的可測集的一個系統開始使得由集的加法和乘法所得的“複合事件”的集合形成一個包含有一個布里雅 (Boolean) 集代數的波勒爾 (Borel) 可測集的集合．兩集之聚集 $\underline{A} \cup \underline{B}$ 是這個複合事件“$\underline{A}$ 或 $\underline{B}$”；交界 $A \cap B$ 是這個複合事件“$\underline{A}$ 及 $\underline{B}$ (同時存在)”；對偶運算子“帽子” $= \cap$ 及“杯子” $= \cup$ 的運算法則亦如普通的布里雅代數中的一樣，這種代數中包含有零集 $\underline{O}$ 及全集 $\underline{I}$． 概率測度由下面的公理所規定：

a) $P(I) = 1$ (1.1)

b) $P(A) \geqslant P(B)$，如果 $A \cup B = A$，即如果 $A \supset B$

c) 如果 $A \cap B = 0$，$P(A \cup B) = P(A) + P(B)$．

在(1.1) c) 中取 $\underline{A} = 1, \underline{B} = 0$，則得 $P(0) = 0$．與 a) 及 b) 合用，即得 $0 \leqslant P(A) \leqslant 1$ 對於這個集合中的所有的集 A 均成立． 最後，看到 $A \cup B$ 是三個互不相交的集 $(A - A \cap B)$, $A \cap B$, $(B - A \cap B)$ 的聚集，我們得到一般的結果：

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{1.2}$$

這一結果與 $P(0) = 0$ 合併可以用來代替(1.1) c)． 這些事件可以看作對於具有零測度的所有集的理想(Ideal)作為模所化簡過了的． 關於波勒爾集的限制雖然並不總是必要的，但是有了它便可以無限次地重複用 $\cup$, $\cap$ 這兩種運算.

定義： $A \cap B = 0$，則 A, B 這兩事件稱為互相排斥．設 $A_1, A_2, \cdots, A_n, \cdots$ 為非空之集，又對於任何有限部分 $i, j, k \cdots$，有 $P(A_i \cap A_j \cap A_k \cdots) = P(A_i)P(A_j)P(A_k)\cdots$，則 $A_1, A_2, \cdots, A_n, \cdots$ 稱為互相獨立的事件.

由此即得兩個互相排斥的事件既不是互相獨立，也不能一個事件完全包含於另一

[*] 1957 年 3 月 22 日收到，秦元勳譯.

個之內．這些是零的和全部的複合概率的極端情況，常常省去了顯然的極端情況 $\underline{O}$, $\underline{I}$ 的分類．由這個代數的任何簡單事件開始，我們可以作出這類簡單事件的一個有序最大鍊，以 $\underline{O}$ 及 $\underline{I}$ 在兩端，這鍊中每一事件包括所有前面的元素，又被所有後面的元素所包括，而這鍊外的代數的任何其他簡單事件對於這一特別的鍊的所有的事件而言均不具有這種包有和被包的性質．

此後我們只考慮這類的布里雅概率代數，它的簡單事件可以分在有限個最大鍊之中，一個鍊中的任一個事件與另一個鍊中的任一個事件是獨立的．

首先，每一個這樣的鍊可以用對應 $A \to (0, P(A))$ 映到實綫段$(0, 1)$．但我們也需要一個映到全部實軸 $-\infty < x < +\infty$ 上的映像，這映像和$(0, 1)$測度映像之間用一個分佈函數 $F(x)$ 連接起來，$F(x)$ 是單調不減函數，具有 $F(-\infty)=0, F(+\infty)=1$．這鍊中的任何集 A 可以映到實軸上區間 $(-\infty, \alpha)$ 使 $F(x) < P(A)$ 當 $x < \underline{\alpha}$，又 $F(\alpha) = P(A)$．對每一個這種有序鍊用一個一維空間，我們把這個布里雅代數映到 n 維連續統 $(x_1, x_2, \cdots x_n)$ 上，一個簡單事件的映像是一個區域，除了某一維之外，在其他維方向上由 $-\infty$ 到 $+\infty$，而在這特別的一維中則從 $-\infty$ 到 $\underline{\alpha}$．在這個超立方體上的測度映像是直角長方體，它除了一邊外，另外的邊長都是 1，而惟一有一邊是區間 $(0, P(A))$．由這些簡單事件用集的聚合及集的交界導出複合事件．

定理 1： 假定一個 n 維概率空間具有一個事件的布里雅代數使得每一維空間代表事件的一個鍊，每一維與其他維的事件互相獨立，又假定這個空間被賦與一個黎曼測度加上一個測度函數，這個函數在單位超立方形上給出一個眞的映像，則這個測度只能是歐氏的．

證明： 對於黎曼測度，$ds^2 = \Sigma g_{ij}\, dx_i\, dx_j$，在這個 x-空間中的任何 $\underline{k}$ 維事件的測度，對 $1 \leqslant k \leqslant n$，是由 $\int f_k(x_{i_1}, \cdots, x_{i_k}) \sqrt{|g_{ij}|}\, dx_{i_1} \cdots dx_{i_k}$ 形式的一個積分所給出的．但如果這個區域是複合事件 $A_1 \cap A_2 \cap \cdots \cap A_k$ 則積分必須分解爲 $\underline{k}$ 個分開的積分之乘積，對於所有的 $\underline{k} \leqslant \underline{n}$．因此，$(g_{ij})$ 的任何主要子行列式及整個行列式必須化爲對角綫上的項 $g_{11}(x_1), g_{22}(x_2), \cdots, g_{nn}(x_n)$ 的乘積，對於它的主要子行列式也對應地成立．測度函數 $\underline{f}$，實值上是分佈的微分，假定存在並且連續，將要類似地分解爲若干因子的一個乘積，但那在此地的興趣不大．顯然矢量 g_{ij} 的所有交叉項均爲零，故 $ds^2 = g_{11}(x_1)dx_1^2 + g_{22}(x_2)dx_2^2 + \cdots + g_{nn}(x_n)dx_n^2$．由於任何一個鍊的正測度的假定，$g_{ii}$ 是正的(此地我們不需要引用測度是正定形式)，可用一個坐標變換 $dx'_r = \sqrt{g_{rr}}\, dx_r$．這便是空間中的歐氏坐標．

我們有兩個簡單的系：

系 1： 如果一個隨機變數空間中賦以一個黎曼測度及一個測度(分佈密度)函數，它容許用獨立隨機函數來代替原來的隨機變數，則這個原來的空間的曲率矢量必爲零，

亦即此空間爲歐氏的.

這些新坐標也即是一個非奇異的坐標變換而得者. 但這個空間便將具有前述定理的歐氏坐標,因此在兩個坐標系中的曲率矢量都將爲零. 對於第二個系我們需要一個拓撲結果[2], 當空間可由鄰域所蓋滿, 使得每一對點可由一條且僅一條弧所連結(這種弧完全在鄰域內)這種弧屬於我們稱爲道路的預先定下了的族中,則這個空間中有一個黎曼測度存在. 由此,這個空間中的可被蓋住的任何緊緻部分,可給一個黎曼測度,而已給的道路實際上是測地綫. 在我們現在的情形,有一個緊緻空間,即單位超立方體,對它應用這個定理,如果函數 f 是連續的,則作回到原來的空間即得到:

系 2: 假如一個隨機變數空間只賦與一個連續測度密度函數,及一族的連續道路,具有這樣的性質,任何相隔相當近的兩點有惟一的道路相連結,則這個空間也具有一個黎曼測度,如果獨立隨機變數的概率可以應用得上,則這個空間在適當的變換下是歐氏的.

在歐氏空間,如果多個獨立隨機變數的複合概率密度函數只與距離有關,則立刻得出每一個變數的分佈是高斯正則的[3]. 由此地到最小二乘方的普通的動機只不過一步之遙而已, 因爲由一個採樣所得的平均值的最好的近似是使得這個採樣的變異爲最小的,也即是使得一個平方之和爲最小的(事實上,即到一個超平面的距離), 因此也即是算術平均值.

2. 我們完全限制在實變數方面, 雖然很少困難便可推廣到複的或其他數系中去. 具 $\underline{n}<\underline{m}$ 個實變數的 $\underline{m}$ 個綫性方程的系統

$$\sum_{j=1}^{n} A_{ij}\,x_j - y_i = 0; \quad i = 1,\, 2,\, \cdots, \quad m > n \tag{2.1}$$

一般是沒有解的. 但常有一個最小二乘法的解使得

$$\sum_{i=1}^{m}\Bigl(\sum_{j=1}^{n} A_{ij}\,x_j - y_i\Bigr)^2, \tag{2.2}$$

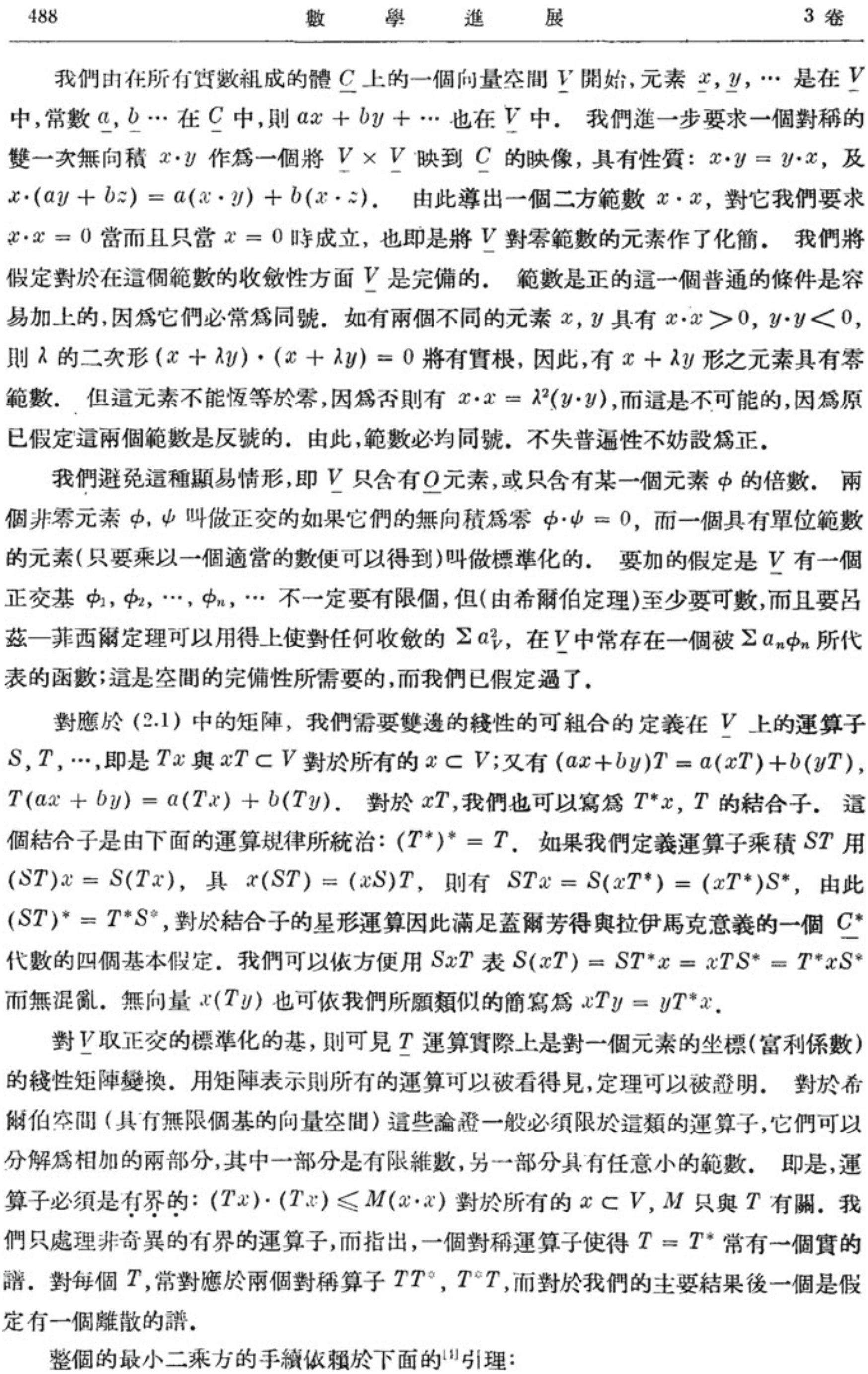
爲最小,因此,以 $\underline{x}$ 記下面的 $\underline{n}$ 個方程

$$\sum_{r=1}^{n} C_{kr}x_r - z_k = 0, \quad C_{kr} = \sum_{q=1}^{m} A_{qk}A_{qr}, \quad z_k = \sum_{q=1}^{m} A_{qk}y_q \tag{2.3}$$

的解. 此地每一個自由的指標是經過 $1, 2, \cdots, n$ 等值的. 一般(2.3)有一組惟一的解,且與(2.1)的準確解重合,如果後者是相容的話. 顯然, 我們可以把這些式子形式地取極限得到第一類的積分方程:

$$\int A(s,\, t)x(t)dt = y(s), \tag{2.4}$$

及變爲其他一般的綫性方程,本節的工作中不考慮概率.



我們由在所有實數組成的體 $\underline{C}$ 上的一個向量空間 $\underline{V}$ 開始，元素 $\underline{x},\underline{y},\cdots$ 是在 $\underline{V}$ 中，常數 $\underline{a},\underline{b}\cdots$ 在 $\underline{C}$ 中，則 $ax+by+\cdots$ 也在 $\underline{V}$ 中．我們進一步要求一個對稱的雙一次無向積 $x\cdot y$ 作爲一個將 $\underline{V}\times\underline{V}$ 映到 $\underline{C}$ 的映像，具有性質：$x\cdot y=y\cdot x$，及 $x\cdot(ay+bz)=a(x\cdot y)+b(x\cdot z)$．由此導出一個二方範數 $x\cdot x$，對它我們要求 $x\cdot x=0$ 當而且只當 $x=0$ 時成立，也即是將 $\underline{V}$ 對零範數的元素作了化簡．我們將假定對於在這個範數的收斂性方面 $\underline{V}$ 是完備的．範數是正的這一個普通的條件是容易加上的，因爲它們必常爲同號．如有兩個不同的元素 x,y 具有 $x\cdot x>0$，$y\cdot y<0$，則 λ 的二次形 $(x+\lambda y)\cdot(x+\lambda y)=0$ 將有實根，因此，有 $x+\lambda y$ 形之元素具有零範數．但這元素不能恆等於零，因爲否則有 $x\cdot x=\lambda^2(y\cdot y)$，而這是不可能的，因爲原已假定這兩個範數是反號的．由此，範數必均同號．不失普遍性不妨設爲正．

我們避免這種顯易情形，即 $\underline{V}$ 只含有 $\underline{O}$ 元素，或只含有某一個元素 ϕ 的倍數．兩個非零元素 ϕ,ψ 叫做正交的如果它們的無向積爲零 $\phi\cdot\psi=0$，而一個具有單位範數的元素（只要乘以一個適當的數便可以得到）叫做標準化的．要加的假定是 $\underline{V}$ 有一個正交基 $\phi_1,\phi_2,\cdots,\phi_n,\cdots$ 不一定要有限個，但（由希爾伯定理）至少要可數，而且要呂茲—菲西爾定理可以用得上使對任何收斂的 Σa_V^2，在 $\underline{V}$ 中常存在一個被 $\Sigma a_n\phi_n$ 所代表的函數；這是空間的完備性所需要的，而我們已假定過了．

對應於 (2.1) 中的矩陣，我們需要雙邊的綫性的可組合的定義在 $\underline{V}$ 上的運算子 $S,T,\cdots$，即是 Tx 與 $xT\subset V$ 對於所有的 $x\subset V$；又有 $(ax+by)T=a(xT)+b(yT)$，$T(ax+by)=a(Tx)+b(Ty)$．對於 xT，我們也可以寫爲 T^*x，T 的結合子．這個結合子是由下面的運算規律所統治：$(T^*)^*=T$．如果我們定義運算子乘積 ST 用 $(ST)x=S(Tx)$，與 $x(ST)=(xS)T$，則有 $STx=S(xT^*)=(xT^*)S^*$，由此 $(ST)^*=T^*S^*$，對於結合子的星形運算因此滿足蓋爾芳得與拉伊馬克意義的一個 $\underline{C}^*$ 代數的四個基本假定．我們可以依方便用 SxT 表 $S(xT)=ST^*x=xTS^*=T^*xS^*$ 而無混亂．無向量 $x(Ty)$ 也可依我們所願類似的簡寫爲 $xTy=yT^*x$．

對 $\underline{V}$ 取正交的標準化的基，則可見 $\underline{T}$ 運算實際上是對一個元素的坐標（富利係數）的綫性矩陣變換．用矩陣表示則所有的運算可以被看得見，定理可以被證明．對於希爾伯空間（具有無限個基的向量空間）這些論證一般必須限於這類的運算子，它們可以分解爲相加的兩部分，其中一部分是有限維數，另一部分具有任意小的範數．即是，運算子必須是有界的：$(Tx)\cdot(Tx)\leqslant M(x\cdot x)$ 對於所有的 $x\subset V$，M 只與 T 有關．我們只處理非奇異的有界的運算子，而指出，一個對稱運算子使得 $T=T^*$ 常有一個實的譜．對每個 T，常對應於兩個對稱算子 TT^*，T^*T，而對於我們的主要結果後一個是假定有一個離散的譜．

整個的最小二乘方的手續依賴於下面的[1]引理：

不在 TV 中的，V 的正交標準化部分由 T^* 映到零集；即是，$T^*(V - TV) = 0$.

證：如果變換後的空間 TV 是 V 的全部，則結論是顯然的. 設稱 $\overline{V}$ 爲不在 TV 中的，V 的正交正則分量. $\overline{V}$ 中的一個函數及 TV 中的另一個函數所成的無向積是 $\overline{V}TV = VT^*\overline{V}$. 由假定，這個無向積是零，也即是說在 V 全體的每一個元素是和 T^*V 中每一個元素正交的，而這是不可能的，除非 $T^*\overline{V} = 0$，便證明了引理.

這個主要的最小二乘方程式現在取這樣的形式

$$Tx - y = 0; \tag{2.5}$$

對 $\underline{T}$，$\underline{y}$ 已給定，$\underline{x}$ 是要求的. 這式不必一定要有一個解，因爲 $\underline{Tx}$ 必須在 $\underline{TV}$ 中，而已給之 $\underline{y}$ 可能有在 $\underline{TV}$ 外面的分量. 左手方的範數是

$$(Tx - y) \cdot (Tx - y) \equiv xT^*Tx - 2(xT^*y) + (y \cdot y) \tag{2.6}$$

要使這數爲極小，對 $\underline{x}$ 給一個變量，用 $\underline{x} + \delta\underline{x}$ 代 $\underline{x}$. 由這個變過了的值減去(2.6)中所給的原有的值便得到

$$2\delta x \cdot (T^*Tx - T^*y) + (T\delta x) \cdot (T\delta x) \tag{2.7}$$

對於極小值而言，如普通一樣 δx 的係數要使之爲零，因爲餘數是正的，當我們將 δx 的範數取趨近於零時. 由此即得：

定理 2: $Tx - y = 0$ 的最小二乘解是由

$$T^*Tx - T^*y = 0 \tag{2.8}$$

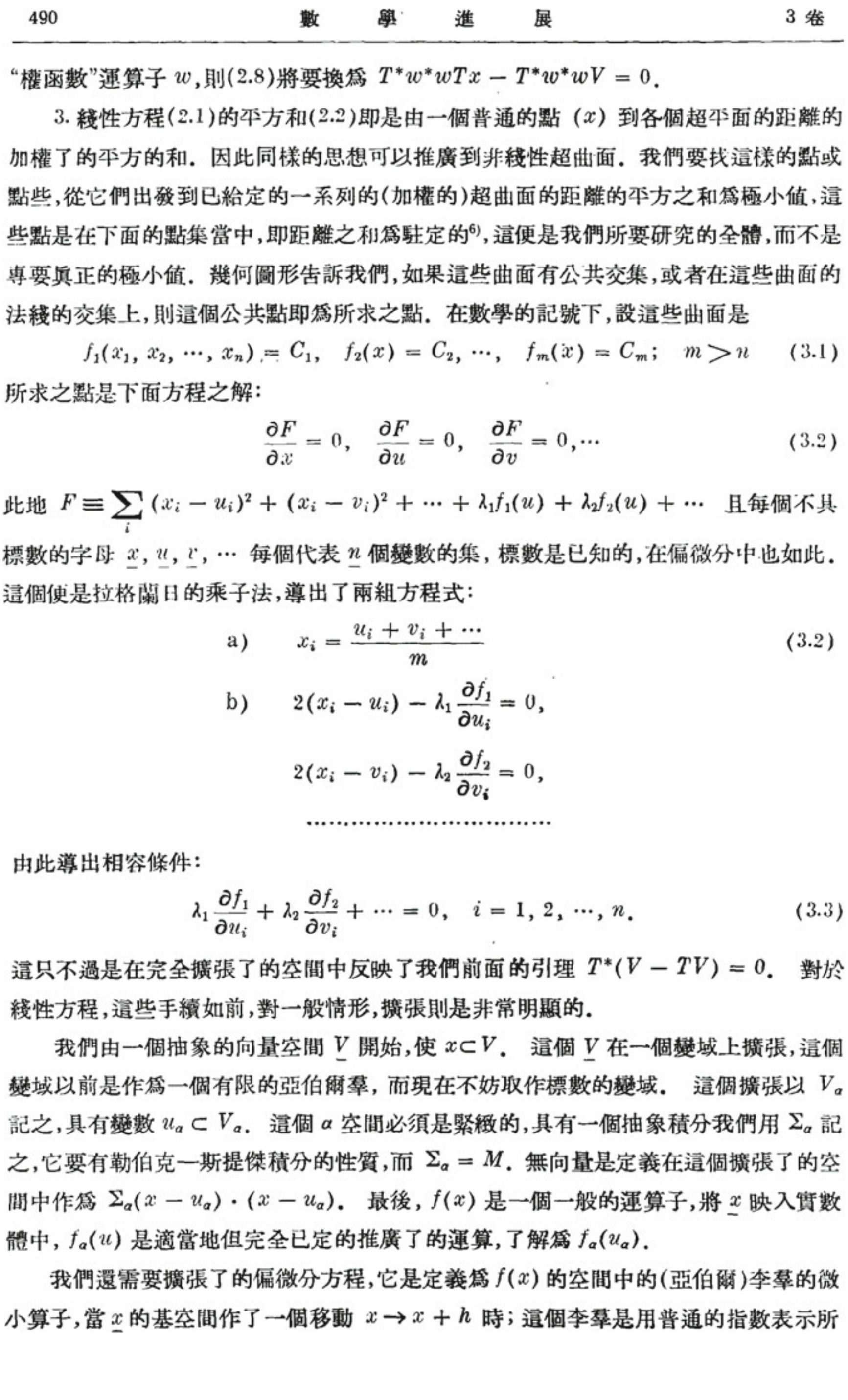
所給出的.

我們的引理 $T^*(V - TV) = 0$ 使得解答存在，因爲它保證 T^*T 的特徵函數在 T^*V 中成一個閉集. 自然，對於問題中的運算子要加上若干簡單的限制. 用固有值和固有函數的語言，這便得到第一類積分方程的畢加的解答[5] 和對應的最小二乘的解. 總結即得：

定理 3: $Tx - y = 0$ 的最小二乘的解存在，當而且只當 $\Sigma(\phi_n T^* y)^2/\lambda_n^4$ 收斂，此地 ϕ_n 及 λ_n^2 是 $T^*T\phi - \lambda^2\phi = 0$ 的固有函數及固有值. 特別，如果 V 具有正交標準化集 $\{\phi_n\}$，則我們有準確的解.

證明如下：非奇異的運算子使 $\underline{V}$ 中的原點不變，因此，由連續性，將 $\underline{V}$ 的某部分映上映像空間 T^*V 的 $\underline{O}$ 之某個鄰近. 我們(假定運算 T^*T 有離散譜，及)用固有函數展開 T^*y. 上面的引理說，T^*y 不可能對所有的這些固有函數正交而又不恆等於零，而本定理之條件只要 T^*y 在原點的變化過的鄰域中. 這一方法在擴散及原子堆中的積分方程中特別有用.

這個結果是與範數無關的. 即是，我們的範數對於恆等變換是取得最好的，恆等變換即是對稱運算子 $xT = Tx = x$ 對於所有 $x \subset V$. 所有的其他的運算子都可以用來作出最小二乘範式，只要 $SV = V$，及 $xSx = 0$ 當而且只當 $x = 0$. 如果要用一個



“權函數”運算子 w，則(2.8)將要換為 $T^*w^*wTx - T^*w^*wV = 0$.

3. 綫性方程(2.1)的平方和(2.2)即是由一個普通的點 (x) 到各個超平面的距離的加權了的平方的和. 因此同様的思想可以推廣到非綫性超曲面. 我們要找這樣的點或點些，從它們出發到已給定的一系列的(加權的)超曲面的距離的平方之和為極小值，這些點是在下面的點集當中，即距離之和為駐定的[6]，這便是我們所要研究的全體，而不是専要眞正的極小值. 幾何圖形告訴我們，如果這些曲面有公共交集，或者在這些曲面的法綫的交集上，則這個公共點即為所求之點. 在數學的記號下，設這些曲面是

$$f_1(x_1, x_2, \cdots, x_n) = C_1, \quad f_2(x) = C_2, \cdots, \quad f_m(x) = C_m; \quad m > n \qquad (3.1)$$

所求之點是下面方程之解:

$$\frac{\partial F}{\partial x} = 0, \quad \frac{\partial F}{\partial u} = 0, \quad \frac{\partial F}{\partial v} = 0, \cdots \qquad (3.2)$$

此地 $F \equiv \sum_i (x_i - u_i)^2 + (x_i - v_i)^2 + \cdots + \lambda_1 f_1(u) + \lambda_2 f_2(u) + \cdots$ 且每個不具標數的字母 $\underline{x}, \underline{u}, \underline{v}, \cdots$ 每個代表 $\underline{n}$ 個變數的集，標數是已知的，在偏微分中也如此. 這個便是拉格蘭日的乘子法，導出了兩組方程式:

$$\text{a)} \qquad x_i = \frac{u_i + v_i + \cdots}{m} \qquad (3.2)$$

$$\text{b)} \qquad 2(x_i - u_i) - \lambda_1 \frac{\partial f_1}{\partial u_i} = 0,$$

$$2(x_i - v_i) - \lambda_2 \frac{\partial f_2}{\partial v_i} = 0,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

由此導出相容條件:

$$\lambda_1 \frac{\partial f_1}{\partial u_i} + \lambda_2 \frac{\partial f_2}{\partial v_i} + \cdots = 0, \quad i = 1, 2, \cdots, n. \qquad (3.3)$$

這只不過是在完全擴張了的空間中反映了我們前面的引理 $T^*(V - TV) = 0$. 對於綫性方程，這些手續如前，對一般情形，擴張則是非常明顯的.

我們由一個抽象的向量空間 $\underline{V}$ 開始，使 $x \subset V$. 這個 $\underline{V}$ 在一個變域上擴張，這個變域以前是作為一個有限的亞伯爾羣，而現在不妨取作標數的變域. 這個擴張以 V_α 記之，具有變數 $u_\alpha \subset V_\alpha$. 這個 α 空間必須是緊緻的，具有一個抽象積分我們用 Σ_α 記之，它要有勒伯克—斯提傑積分的性質，而 $\Sigma_\alpha = M$. 無向量是定義在這個擴張了的空間中作為 $\Sigma_\alpha(x - u_\alpha) \cdot (x - u_\alpha)$. 最後，$f(x)$ 是一個一般的運算子，將 $\underline{x}$ 映入實數體中，$f_\alpha(u)$ 是適當地但完全已定的推廣了的運算，了解為 $f_\alpha(u_\alpha)$.

我們還需要擴張了的偏微分方程，它是定義為 $f(x)$ 的空間中的(亞伯爾)李羣的微小算子，當 $\underline{x}$ 的基空間作了一個移動 $x \to x + h$ 時；這個李羣是用普通的指數表示所

產生,由此導出一個李—泰洛級數展開,它是形式的表示,且在解析的情形收斂並給出一個準確的表示. 我們的非綫性汎涵算子 $\underline{f}$ 既不一定是解析的,甚至不一定可任意微分,因爲他們可以在需要時用這類來近逼;但 $\underline{f}$ 算子必須至少第一次微分要是連續的,爲了使得(3.2)的類似的情形有效的話. 引入一個正交標準化的基及 $\underline{V}$ 中的坐標,偏微分就變成了在這種坐標中的普通的偏微分;普通地,我們用 $\underline{f}'$ 表之. 結果綜述如下:

聯立的投影 $f_a(x) = 0$ 的最小二乘方的解是由 $x = \frac{1}{M}\Sigma_a u_a$ 給出的,只要這些擴張了的變數 u_a 滿足

$$\lambda_a f'_a(u_a) = \frac{1}{M}\Sigma_a u_a - u_a \tag{3.4}$$

此地 λ_a 及 u_a 是如此選取,使更滿足

$$f_a(u_a) = 0.$$

參考文獻

[1] Kolmogorov, A., Grundbegriffe der Wahrscheinlich-Kreisrechnung, *Ergebnisse d. Mathematik*, **2**: 3, Berlin, 1933, the opening sections.

[2] Kosambi, D. D., The metric in path-space; Tensor **3**, 1954, 67—74.

[3] Kosambi, D. D., The geometric method in mathematical statistics, *Amer. Math. Monthly*, **51** (1944), 382—9.

[4] Loomis, L. H., Introduction to abstract harmonic analysis, N.Y. 1953, p. 27, for the classical result.

[5] Courant, R. & Hilbert, D., Methoden der mathematischen Physik I, Berlin, 1931, pp. 134—6; Picard, E., Rendiconti del Circolo Matematico di Palermo, **29** (1910), 79—97.

[6] Kosambi, D. D., An extension of the least-squares method for statistical estimatior, *Annals. of Engenics*, **13** (1947), 257—261.