STATISTICS AND TRUTH

Putting Chance to Work

second edition

This page is intentionally left blank

C. Radhakrishna Rao

Pennsylvania State University, USA

STATISTICS AND TRUTH

Putting Chance to Work

second edition

World Scientific

Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 912805

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Rao, C. Radhakrishna (Calyampudi Radhakrishna), 1920--

Statistics and truth : putting chance to work / C. Radhakrishna

Rao. -- 2nd enl. ed.

p. cm.

Includes bibliographical references and index.

ISBN 9810231113 (alk. paper)

1. Mathematical statistics. I. Title

QA276.16.R36 1997

519.5--dc21

97-10349

CIP

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

First published 1997

Reprinted 1999

Copyright © 1997 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

Printed in Singapore.

For instilling in me
the quest for knowledge

I owe to
my mother

A. Laxmikanthamma

who, in my younger days, woke me up
everyday at four in the morning and
lit the oil lamp for me to study in
the quiet hours of the morning
when the mind is fresh

*Knowledge is what we know
Also, what we know we do not know.
We discover what we do not know
Essentially by what we know.*

Thus knowledge expands.

*With more knowledge we come to know
More of what we do not know.*

Thus knowledge expands endlessly.

* * *

*All knowledge is, in final analysis, history.
All sciences are, in the abstract, mathematics.
All judgements are, in their rationale, statistics.*

Foreword

Beginning this year, CSIR has instituted a distinguished lectureship series. The objective here is to invite eminent scientists from India and abroad for delivering a series of three lectures on topics of their choice. The lectures, known as the CSIR Distinguished Lectures, were to be delivered in different locations of the country. The first of this series has been dedicated to the memory of the mathematics genius Srinivasa Ramanujan.

It augurs well that this first set of lectures (CSIR Ramanujan lectures) has begun with those of Prof. C. Radhakrishna Rao, National Professor (and currently Eberly Professor of Statistics, Penn State University), a distinguished scientist in the international statistics scene.

The lectures were delivered at the National Physical Laboratory in Delhi, at the Central Leather Research Institute in Madras and in the Indian Statistical Institute at Calcutta and were widely appreciated by professional statisticians, by physicists, chemists and biological scientists, by students of different age groups and by professionals and administrators. The scope of these lectures was wide and pervaded many areas of human activities, both scientific and administrative.

By arranging to have the lectures published now, CSIR hopes that a wider community of scientists the world over will be able to derive the benefit of the expertise of renowned men like Prof. Rao.

I express my appreciation of the efforts of Dr. Y.R. Sarma for having edited and brought out the volume quickly.

New Delhi,
December 31, 1987

A.P. MITRA
Director-General
Council of Scientific &
Industrial Research

***Uncertain
knowledge***

+

***Knowledge of the
amount of
uncertainty in it***

=

***Usable
knowledge***

Preface

I consider it a great honour to be called upon to deliver the Ramanujan Memorial Lectures under the auspices of the CSIR (Council of Scientific & Industrial Research). I would like to thank Dr. A.P. Mitra, Director General of the CSIR, for giving me this honour and opportunity to participate in the Ramanujan centenary celebrations.

I gave three lectures, the first one in Delhi, the second in Calcutta and the third in Madras as scheduled, which I have written up in four chapters for publication. In the beginning of each lecture I have said a few words about the life and work of Ramanujan, the rare mathematical genius who was a legendary figure in my younger days. This is to draw the attention of the younger generation to the achievements of Ramanujan, and to emphasize the need to reform our educational system and reorganize our research institutes to encourage creativity and original thinking among the students.

When I was a student, statistics was in its infancy and I have closely watched its evolution over the last 50 years as an independent discipline of great importance and a powerful tool in acquiring knowledge in any field of enquiry. The reason for such phenomenal developments is not difficult to seek.

Statistics as a method of learning from experience and decision making under uncertainty must have been practiced from the beginning of mankind. But the inductive reasoning involved in these processes has never been codified due to the uncertain nature of the conclusions drawn from given data or information. The breakthrough occurred only in the beginning of the present century with the realization that inductive reasoning can be made precise by specifying the amount of uncertainty involved in the conclusions drawn. This paved the way for working out an optimum course of action involving minimum risk, in any given uncertain situation, by a purely deductive process. Once this mechanism was made available, the flood gates

opened and there was no end to the number of applications impatiently awaiting methods which could really deliver the goods.

From the time of Aristotle to the middle of the 19th century, chance was considered by scientists as well as philosophers to be an indication of our ignorance which makes prediction impossible. It is now recognized that chance is inherent in all natural phenomena, and the only way of understanding nature and making optimum predictions (with minimum loss) is to study the laws (or the inner structure) of chance and formulate appropriate rules of action. Chance may appear as an obstructor and an irritant in our daily life, but chance can also help and create. We have now learnt to put chance to work for the benefit of mankind.

I have chosen to speak on the foundations, modern developments and future of statistics, because of my involvement with statistics over the last 45 years as a teacher, research worker and consultant in statistics, and as an administrator managing the academic affairs of a large organization devoted to statistics. I grew up in a period of intensive developments in the history of modern statistics.

As a student I specialized in mathematics - the logic of deducing consequences from given premises. Later I studied statistics - a rational approach to learning from experience and the logic of identifying the premises given the consequences. I have come to realize the importance of both in all human endeavours whether it is in the advancement of natural knowledge or in the efficient management of our daily chores. I believe:

All knowledge is, in final analysis, history.

All sciences are, in the abstract, mathematics.

All judgements are, in their rationale, statistics.

The title of my Lectures, *Statistics and Truth*, and their general theme are somewhat similar to those of *Probability, Statistics and Truth*, the collected lectures of R. von Mises published several

years ago. Since the latter book appeared, there have been new developments in our thinking and attitude towards chance. We have reconciled with the "Dice-Playing God" and learnt to plan our lives to keep in resonance with uncertainties around us. We have begun to understand and accept the beneficial role of chance in situations beyond our control or extremely complex to deal with. To emphasize this I have chosen the subtitle, *Putting Chance to Work*.

Dr. Joshi, the Director of the National Physical Laboratory, reminded me what Thomas Huxley is reported to have said, that a man of science past sixty does more harm than good. *Statistically* it may be so. As we grow old we tend to stick to our past ideas and try to propagate them. This may not be good for science. Science advances by change, by the introduction of new ideas. These can arise only from uninhibited young minds capable of conceiving what may appear to be impossible but which may contain a nucleus for revolutionary change. But I am trying to follow Lord Rayleigh who was an active scientist throughout his long life. At the age of sixty-seven (which is exactly my present age), when asked by his son, who is also a famous physicist, to comment on Huxley's remark, Rayleigh responded:

That may be if he undertakes to criticize the work of younger men, but I do not see why it need be so if he sticks to things he is conversant with.

However J.B.S. Haldane used to say that Indian scientists are polite and they do not criticize the work of each other, which is not good for the progress of science.

It gives me great pleasure to thank Dr. Y.R.K. Sarma of the Indian Statistical Institute for the generous help he has given in editing the Ramanujan Memorial Lectures I gave at various places in the form of a book and looking after its publication.

Calcutta

December 12, 1987

C.R. Rao

*He who accepts statistics indiscriminately
will often be duped unnecessarily. But
he who distrusts statistics indiscriminately
will often be ignorant unnecessarily.*

Preface to the Second Edition

The first edition of **Statistics and Truth: Putting Chance to Work** was based on three lectures on the history and development of statistics I gave during Ramanujan Centenary Celebrations in 1987. The topics covered in each lecture were reproduced in a more detailed form as separate sections of the book. The present edition differs from the first in many respects.

The material which appears under lectures 1, 2 and 3 in the first edition is completely reorganized and expanded to a sequence of five chapters to provide a coherent account of the development of statistics from its origin as collection and compilation of data for administrative purposes to a fullfledged separate scientific discipline of study and research. The relevance of statistics in all scientific investigations and decision making is demonstrated through a number of examples. Finally a completely new chapter (sixth) on the public understanding of statistics, which is of general interest, is added.

Chapter 1 deals with the concepts of randomness, chaos and chance, all of which play an important role in investigating and explaining natural phenomena. The role of random numbers in confidential transactions, in generating unbiased information and in solving problems involving complex computations is emphasized. Some thoughts are expressed on creativity in arts and science. Chapter 2 introduces the deductive and inductive methods used in the creation of new knowledge. It also demonstrates how quantification of uncertainty has led to optimum decision making,

Statistics has a long antiquity but a short history. Chapters 3 and 4 trace the development of statistics from the notches made by the primitive man to keep an account of his cattle to a powerful logical tool for extracting information from numbers or given data and drawing inferences under uncertainty. The need to have clean data free from faking, contamination or editing of any sort is stressed, and some methods are described to detect such defects in data.

Chapter 5 deals with the ubiquity of statistics as an inevitable tool in search of truth in any investigation whether it is for unravelling the mysteries of nature or for taking optimum decisions in daily life or for settling disputes in courts of law.

We are all living in an information age, and much of the information is transmitted in a quantitative form such as the following. The crime rate this year has gone down by 10% compared to last year. There is 30% chance of rain tomorrow. The Dow Jones index of stock market prices has gained 50 points. Every fourth child born is Chinese. The percentage of people approving President's foreign policy is 57 with a margin of error of 4 percentage points. You lose 8 years of your life if you remain unmarried. What do these numbers mean to the general public? What information is there in these numbers to help individuals in making right decisions to improve the quality of their lives. An attempt is made in Chapter 6 of the new edition to emphasize the need for public understanding of statistics, and what we can learn from numbers to be efficient citizens, as emphasized by H.G. Wells:

Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.

At the beginning of each lecture delivered in 1987, some aspects of Ramanujan's life and work were mentioned. All these biographical details are put together as a connected essay on Ramanujan in the Appendix to the book.

University Park,
March 31, 1997

C.R. Rao

Contents

CHAPTER 1 Uncertainty, Randomness and Creation of New Knowledge	1
1. Uncertainty, and its quantification	1
2. Randomness and random numbers	3
3. From determinism to order in disorder	17
4. Randomness and creativity	21
References	25
APPENDIX: Discussion	26
A.1 Chance and chaos	26
A.2 Creativity	28
A.3 Chance and necessity	34
A.4 Ambiguity	37
A.5 Are decimal digits in π random?	39
CHAPTER 2 Taming of Uncertainty - Evolution of Statistics	41
1. Early history: statistics as data	41
2. Taming of uncertainty	49
3. Future of statistics	60
CHAPTER 3 Principles and Strategies of Data Analysis: Cross Examination of Data	63
1. Historical developments in data analysis	63
2. Cross-examination of data	70
3. Meta analysis	87
4. Inferential data analysis and concluding remarks	89
References	92

CHAPTER 4 Weighted Distributions - Data with Built-in Bias 95

1. Specification	95
2. Truncation	96
3. Weighted distributions	99
4. P.p.s. sampling	101
5. Weighted binomial distribution: empirical theorems	102
6. Alcoholism, family size and birth order	110
7. Waiting time paradox	116
8. Damage models	117
References	119

CHAPTER 5 Statistics: An Inevitable Instrument in Search of Truth 121

1. Statistics and truth	121
2. Some examples	129
2.1 Shakespeare's poem: an ode to statistics	129
2.2 Disputed authorship: the Federalist papers	131
2.3 Kautilya and the Arthaśāstra	133
2.4 Dating of publications	133
2.5 Seriation of Plato's works	134
2.6 Filiation of manuscripts	134
2.7 The language tree	135
2.8 Geological time scale	136
2.9 Common breeding ground of eels	137
2.10 Are acquired characteristics inherited?	138
2.11 The importance of being left-handed	139
2.12 Circadian rhythm	143
2.13 Disputed paternity	145
2.14 Salt in statistics	145
2.15 Economy in blood testing	147

2.16 Machine building factories to increase food production	148
2.17 The missing decimal numbers	150
2.18 The Rhesus factor: a study in scientific research	152
2.19 Family size, birth order and I.Q.	154
References	155

CHAPTER 6 Public Understanding of Statistics: Learning from Numbers 157

1. Science for all	157
2. Data, information and knowledge	158
3. Information revolution and understanding of statistics	162
4. Mournful numbers	166
5. Weather forecasting	168
6. Public opinion polls	169
7. Superstition and psychosomatic processes	171
8. Statistics and the law	173
9. ESP and amazing coincidences	177
10. Spreading statistical numeracy	178
11. Statistics as a key technology	179
References	180

APPENDIX

Srinivasa Ramanujan - a rare phenomenon	181
---	-----

Index	187
-------	-----