

TURING

图灵原版数学·统计学系列 31

CAMBRIDGE

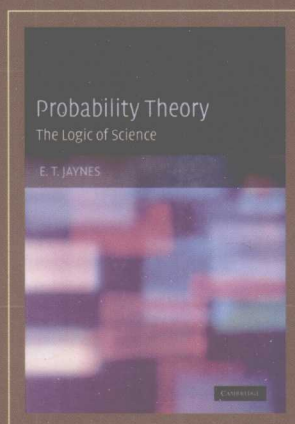
Probability Theory

The Logic of Science

概率论沉思录

(英文版)

[美] E. T. Jaynes 著



人民邮电出版社
POSTS & TELECOM PRESS

Probability Theory The Logic of Science

概率论沉思录 (英文版)

“这是几十年来最重要的一部概率论著作。它解决了许多长期困扰我的问题。概率、统计、模式识别、数据分析、机器学习、数据挖掘……只要你的工作涉及不完全和不确定信息的处理，就应该仔细研读本书。它将大大改变你思考问题的方式。”

——Kevin S. Van Horn, 资深计算机技术和概率统计专家

“本书广受欢迎。读者会在书中发现很多引人深思的内容，不仅涉及日常实践，更深入统计和概率理论本身。无论对于统计学者还是各应用领域的科技工作者，本书都是必读之作。”

——美国《数学评论》

“这不是一本普通的教材。它全面、彻底地阐述了统计中的贝叶斯方法。书中有上百个例子，足够让你透彻理解其中的理论和应用。每个对统计问题或统计应用感兴趣的人都应该仔细研读。”

——SIAM News

这是一部奇书。它是著名数学物理学家Jaynes的遗作，凝聚了他对概率论长达40年的深刻思考。原版出版后产生了巨大影响，深受众多专家和学者的好评，并获得Amazon网上书店读者全五星评价。

在书中，作者在H. Jeffreys、R. T. Cox、C. E. Shannon和G. Pólya等数学大师思想的基础上继续探索，将概率论置于更大的背景下考察，提出将概率推断作为整个科学的逻辑基础，以适应实际科学研究中对象往往都是信息不完全或者不确定的这一难题，从而超越了传统的概率论，也超越了传统的数理逻辑思维定式。

本书将概率和统计推断融合起来，用新颖的观点生动地描述了概率论/贝叶斯理论在物理学、数学、化学、生物学、经济学和社会学等领域的广泛应用，弥补了其他概率和统计教材的不足。不仅适合概率和统计专业人士阅读，也是需要应用统计推断的各领域科技工作者的必读之作。

E. T. Jaynes (1922–1998) 已故著名数学家和物理学家。生前曾任华盛顿大学圣路易斯分校和斯坦福大学教授。他因为提出了热动力学的最大熵原理(1957年)和量子光学的Jaynes-Cummings模型(1963年)而闻名于世。此后的几十年，他一直在探求将概率和统计推断作为整个科学的逻辑基础这一重大课题，其成果和心得最终凝结为本书。

CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org

本书相关信息请访问：图灵网站 <http://www.turingbook.com>

读者/作者热线：(010)88593802

反馈/投稿/推荐信箱：contact@turingbook.com



分类建议：数学/概率

人民邮电出版社网址 www.ptpress.com.cn

ISBN 978-7-115-19536-4



9 787115 195364 >

ISBN 978-7-115-19536-4/O1

定价：99.00 元

TURING

图灵原版数学·统计学系列

Probability Theory

The Logic of Science

概率论沉思录

(英文版)

[美] E. T. Jaynes 著

人民邮电出版社
北 京

图书在版编目 (CIP) 数据

概率论沉思录 = Probability Theory: The Logic of Science: 英文/ (美) 杰恩斯 (Jaynes, E. T.) 著.

—北京: 人民邮电出版社, 2009.4

(图灵原版数学·统计学系列)

ISBN 978-7-115-19536-4

I. 概… II. 杰… III. 概率论—教材—英文 IV. O211

中国版本图书馆CIP数据核字 (2008) 第203456号

内 容 提 要

本书将概率和统计推断融合在一起, 用新的观点生动地描述了概率论在物理学、数学、经济学、化学和生物学等领域中的广泛应用, 尤其是它阐述了贝叶斯理论的丰富应用, 弥补了其他概率和统计教材的不足. 全书分为两大部分. 第一部分包括10章内容, 讲解抽样理论、假设检验、参数估计等概率论的原理及其初等应用; 第二部分包括12章内容, 讲解概率论的高级应用, 如在物理测量、通信理论中的应用. 本书还附有大量习题, 内容全面, 体例完整.

本书内容不局限于某一特定领域, 适合涉及数据分析的各领域工作者阅读, 也可作为高年级本科生和研究生相关课程的教材.

图灵原版数学·统计学系列

概率论沉思录 (英文版)

-
- ◆ 著 [美] E. T. Jaynes 著
责任编辑 明永玲
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号
邮编 100061 电子函件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京隆昌伟业印刷有限公司印刷
 - ◆ 开本: 700×1000 1/16
印张: 47.25
字数: 753千字 2009年4月第1版
印数: 1-2 500册 2009年4月北京第1次印刷
著作权合同登记号 图字: 01-2008-5171号
ISBN 978-7-115-19536-4/O1
-

定价: 99.00元

读者服务热线: (010) 88593802 印装质量热线: (010) 67129223

反盗版热线: (010) 67171154

人民邮电出版社图灵公司自成立以来,引进并出版了大量世界级大师的经典著作,深受读者好评,在国内数学界引起了强烈反响,被誉为国内高端数学类图书的“旗舰出版单位”之一。这些作品都是世界级大师的心血力作,大师们包括数学界的一代宗师哈代(G. H. Hardy)、当代最杰出的数学家Peter D. Lax、20世纪日本最伟大的数学家小平邦彦、20世纪最伟大的概率学家威廉·费勒、天才青年数学家陶哲轩……

图灵数学·统计学丛书

序号	作者	书名
32	Alan Tucker	应用组合数学(第5版)
31	Douglas C.Montgomery	实验设计与分析(第6版)
30	Stuart A.Klugman, Harry H.Panjer, Gordon E.Willmot	损失模型:从数据到决策(第2版)
29	Peter D.Lax	线性代数及其应用(第2版)
28	Hamdy A.Taha	运筹学导论:高级篇(第8版)
27	G.H.Hardy, J.E.Littlewood, G.Pólya	不等式(第2版)
26	陶哲轩(Terence Tao)	陶哲轩实分析
25	John Willard Milnor	从微分观点看拓扑(双语版)
24	Garrett Birkhoff, Saunders Mac Lane	近世代数概论(第5版)
23	G.H.Hardy, Edward Maitland Wright	数论导引(第5版)
22	Hamdy A.Taha	运筹学导论:初级篇(第8版)
21	小平邦彦(Kunihiko Kodaira)	微积分入门II:多元微积分
20	小平邦彦(Kunihiko Kodaira)	微积分入门I:一元微积分
19	Morris W.Hirsch, Stephen Smale, Robert Devaney	微分方程、动力系统与混沌导论(第2版)
18	William Feller	概率论及其应用(第2卷·第2版)
17	Sheldon Ross	应用随机过程:概率模型导论(第9版)
16	Charalambos D.Aliprantis, Owen Burkinshaw	实分析习题集(第2版)
15	Kenneth Falconer	分形几何——数学基础及其应用(第2版)
14	Samuel Karlin, Howard M.Taylor	随机过程初级教程(第2版)
13	Gary Chartrand, Ping Zhang	图论导引
12	David C.Lay	线性代数及其应用(第3版修订版)
11	Sheldon M.Ross	统计模拟(第4版)
10	James W.Demmel	应用数值线性代数
9	Morris H.DeGroot, Mark J.Schervish	概率统计
8	Sheldon M.Ross	概率论基础教程(第7版)
7	Lloyd N.Trefethen, David Bau, III	数值线性代数
6	Alison Etheridge	金融数学教程
5	William Feller	概率论及其应用(第3版)
4	W.J.Conover	实用非参数统计(第3版)
3	Michael Spivak	流形上的微积分(双语版)
2	K.W.Morton, D.F.Mayers	偏微分方程数值解(第2版)
1	Martin W.Baxter, Andrew J.O.Rennie	金融数学:衍生品定价引论

即将出版

蔡瑞胸(Ruey S. Tsay)
Geof H.Givens, Jennifer A.Hoeting
Sheldon Axler

金融时间序列分析(第2版)
统计计算
线性代数(第2版)

版 权 声 明

Probability Theory: The Logic of Science, first edition (ISBN 978-0-521-59271-0) by E. T. Jaynes, first published by Cambridge University Press 2003.

All rights reserved.

This reprint edition for the People's Republic of China is published by arrangement with the Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

© Cambridge University Press & POSTS & TELECOM PRESS 2009.

This book is in copyright. No reproduction of any part may take place without the written permission of Cambridge University Press and POSTS & TELECOM PRESS .

This edition is for sale in the People's Republic of China (excluding Hong Kong SAR, Macau SAR and Taiwan Province) only.

本书由剑桥大学出版社授权人民邮电出版社出版。本书任何部分之文字及图片，未经出版者书面许可，不得用任何方式节录或翻印。此版本仅限在中华人民共和国境内（不包括香港、澳门特别行政区及台湾省）销售。

Dedicated to the memory of
Sir Harold Jeffreys, who
saw the truth and preserved it.

Editor's foreword

E. T. Jaynes died April 30, 1998. Before his death he asked me to finish and publish his book on probability theory. I struggled with this for some time, because there is no doubt in my mind that Jaynes wanted this book finished. Unfortunately, most of the later chapters, Jaynes' intended volume 2 on applications, were either missing or incomplete, and some of the early chapters also had missing pieces. I could have written these latter chapters and filled in the missing pieces, but if I did so, the work would no longer be Jaynes'; rather, it would be a Jaynes–Bretthorst hybrid with no way to tell which material came from which author. In the end, I decided the missing chapters would have to stay missing – the work would remain Jaynes'.

There were a number of missing pieces of varying length that Jaynes had marked by inserting the phrase 'MUCH MORE COMING'. I could have left these comments in the text, but they were ugly and they made the book look very incomplete. Jaynes intended this book to serve as both a reference and a text book. Consequently, there are question boxes (Exercises) scattered throughout most chapters. In the end, I decided to replace the 'MUCH MORE COMING' comments by introducing 'Editor's' Exercises. If you answer these questions, you will have filled in the missing material.

Jaynes wanted to include a series of computer programs that implemented some of the calculations in the book. I had originally intended to include these programs. But, as time went on, it became increasingly obvious that many of the programs were not available, and the ones that were were written in a particularly obscure form of BASIC (it was the programs that were obscure, not the BASIC). Consequently, I removed the references to these programs and, where necessary, inserted a few sentences to direct people to the necessary software tools to implement the calculations.

Numerous references were missing and had to be supplied. Usually the information available, a last name and date, was sufficient to find one or more probable references. When there were several good candidates, and I was unable to determine which Jaynes intended, I included multiple references and modified the citation. Sometimes the information was so vague that no good candidates were available. Fortunately, I was able to remove the citation with no detrimental effect. To enable readers to distinguish between cited works and other published sources, Jaynes' original annotated bibliography has been split into two sections: a Reference list and a Bibliography.

Finally, while I am the most obvious person who has worked on getting this book into publication, I am not the only person to do so. Some of Jaynes' closest friends have assisted me in completing this work. These include Tom Grandy, Ray Smith, Tom Lored, Myron Tribus and John Skilling, and I would like to thank them for their assistance. I would also like to thank Joe Ackerman for allowing me to take the time necessary to get this work published.

G. Larry Bretthorst

Preface

The following material is addressed to readers who are already familiar with applied mathematics, at the advanced undergraduate level or preferably higher, and with some field, such as physics, chemistry, biology, geology, medicine, economics, sociology, engineering, operations research, etc., where inference is needed.¹ A previous acquaintance with probability and statistics is not necessary; indeed, a certain amount of innocence in this area may be desirable, because there will be less to unlearn.

We are concerned with probability theory and all of its conventional mathematics, but now viewed in a wider context than that of the standard textbooks. Every chapter after the first has ‘new’ (i.e. not previously published) results that we think will be found interesting and useful. Many of our applications lie outside the scope of conventional probability theory as currently taught. But we think that the results will speak for themselves, and that something like the theory expounded here will become the conventional probability theory of the future.

History

The present form of this work is the result of an evolutionary growth over many years. My interest in probability theory was stimulated first by reading the work of Harold Jeffreys (1939) and realizing that his viewpoint makes all the problems of theoretical physics appear in a very different light. But then, in quick succession, discovery of the work of R. T. Cox (1946), Shannon (1948) and Pólya (1954) opened up new worlds of thought, whose exploration has occupied my mind for some 40 years. In this much larger and permanent world of rational thinking in general, the current problems of theoretical physics appeared as only details of temporary interest.

The actual writing started as notes for a series of lectures given at Stanford University in 1956, expounding the then new and exciting work of George Pólya on ‘Mathematics and Plausible Reasoning’. He dissected our intuitive ‘common sense’ into a set of elementary qualitative desiderata and showed that mathematicians had been using them all along to

¹ By ‘inference’ we mean simply: deductive reasoning whenever enough information is at hand to permit it; inductive or plausible reasoning when – as is almost invariably the case in real problems – the necessary information is not available. But if a problem can be solved by deductive reasoning, probability theory is not needed for it; thus our topic is the optimal processing of incomplete information.

guide the early stages of discovery, which necessarily precede the finding of a rigorous proof. The results were much like those of James Bernoulli's *Art of Conjecture* (1713), developed analytically by Laplace in the late 18th century; but Pólya thought the resemblance to be only qualitative.

However, Pólya demonstrated this qualitative agreement in such complete, exhaustive detail as to suggest that there must be more to it. Fortunately, the consistency theorems of R. T. Cox were enough to clinch matters; when one added Pólya's qualitative conditions to them the result was a proof that, if degrees of plausibility are represented by real numbers, then there is a uniquely determined set of quantitative rules for conducting inference. That is, any other rules whose results conflict with them will necessarily violate an elementary – and nearly inescapable – desideratum of rationality or consistency.

But the final result was just the standard rules of probability theory, given already by Daniel Bernoulli and Laplace; so why all the fuss? The important new feature was that these rules were now seen as uniquely valid principles of logic in general, making no reference to 'chance' or 'random variables'; so their range of application is vastly greater than had been supposed in the conventional probability theory that was developed in the early 20th century. As a result, the imaginary distinction between 'probability theory' and 'statistical inference' disappears, and the field achieves not only logical unity and simplicity, but far greater technical power and flexibility in applications.

In the writer's lectures, the emphasis was therefore on the quantitative formulation of Pólya's viewpoint, so it could be used for general problems of scientific inference, almost all of which arise out of incomplete information rather than 'randomness'. Some personal reminiscences about George Pólya and this start of the work are in Chapter 5.

Once the development of applications started, the work of Harold Jeffreys, who had seen so much of it intuitively and seemed to anticipate every problem I would encounter, became again the central focus of attention. My debt to him is only partially indicated by the dedication of this book to his memory. Further comments about his work and its influence on mine are scattered about in several chapters.

In the years 1957–1970 the lectures were repeated, with steadily increasing content, at many other universities and research laboratories.² In this growth it became clear gradually that the outstanding difficulties of conventional 'statistical inference' are easily understood and overcome. But the rules which now took their place were quite subtle conceptually, and it required some deep thinking to see how to apply them correctly. Past difficulties, which had led to rejection of Laplace's work, were seen finally as only misapplications, arising usually from failure to define the problem unambiguously or to appreciate the cogency of seemingly trivial side information, and easy to correct once this is recognized. The various relations between our 'extended logic' approach and the usual 'random variable' one appear in almost every chapter, in many different forms.

² Some of the material in the early chapters was issued in 1958 by the Socony-Mobil Oil Company as Number 4 in their series 'Colloquium Lectures in Pure and Applied Science'.

Eventually, the material grew to far more than could be presented in a short series of lectures, and the work evolved out of the pedagogical phase; with the clearing up of old difficulties accomplished, we found ourselves in possession of a powerful tool for dealing with new problems. Since about 1970 the accretion has continued at the same pace, but fed instead by the research activity of the writer and his colleagues. We hope that the final result has retained enough of its hybrid origins to be usable either as a textbook or as a reference work; indeed, several generations of students have carried away earlier versions of our notes, and in turn taught it to their students.

In view of the above, we repeat the sentence that Charles Darwin wrote in the Introduction to his *Origin of Species*: 'I hope that I may be excused for entering on these personal details, as I give them to show that I have not been hasty in coming to a decision.' But it might be thought that work done 30 years ago would be obsolete today. Fortunately, the work of Jeffreys, Pólya and Cox was of a fundamental, timeless character whose truth does not change and whose importance grows with time. Their perception about the nature of inference, which was merely curious 30 years ago, is very important in a half-dozen different areas of science today; and it will be crucially important in all areas 100 years hence.

Foundations

From many years of experience with its applications in hundreds of real problems, our views on the foundations of probability theory have evolved into something quite complex, which cannot be described in any such simplistic terms as 'pro-this' or 'anti-that'. For example, our system of probability could hardly be more different from that of Kolmogorov, in style, philosophy, and purpose. What we consider to be fully half of probability theory as it is needed in current applications – the principles for assigning probabilities by logical analysis of incomplete information – is *not* present at all in the Kolmogorov system.

Yet, when all is said and done, we find ourselves, to our own surprise, in agreement with Kolmogorov and in disagreement with his critics, on nearly all technical issues. As noted in Appendix A, each of his axioms turns out to be, for all practical purposes, derivable from the Pólya–Cox desiderata of rationality and consistency. In short, we regard our system of probability as not contradicting Kolmogorov's; but rather seeking a deeper logical foundation that permits its extension in the directions that are needed for modern applications. In this endeavor, many problems have been solved, and those still unsolved appear where we should naturally expect them: in breaking into new ground.

As another example, it appears at first glance to everyone that we are in very close agreement with the de Finetti system of probability. Indeed, the writer believed this for some time. Yet when all is said and done we find, to our own surprise, that little more than a loose philosophical agreement remains; on many technical issues we disagree strongly with de Finetti. It appears to us that his way of treating infinite sets has opened up a Pandora's box of useless and unnecessary paradoxes; nonconglomerability and finite additivity are examples discussed in Chapter 15.

Infinite-set paradoxing has become a morbid infection that is today spreading in a way that threatens the very life of probability theory, and it requires immediate surgical removal. In our system, after this surgery, such paradoxes are avoided automatically; they cannot arise from correct application of our basic rules, because those rules admit only finite sets and infinite sets that arise as well-defined and well-behaved limits of finite sets. The paradoxing was caused by (1) jumping directly into an infinite set without specifying any limiting process to define its properties; and then (2) asking questions whose answers depend on how the limit was approached.

For example, the question: 'What is the probability that an integer is even?' can have any answer we please in $(0, 1)$, depending on what limiting process is used to define the 'set of all integers' (just as a conditionally convergent series can be made to converge to any number we please, depending on the order in which we arrange the terms).

In our view, an infinite set cannot be said to possess any 'existence' and mathematical properties at all – at least, in probability theory – until we have specified the limiting process that is to generate it from a finite set. In other words, we sail under the banner of Gauss, Kronecker, and Poincaré rather than Cantor, Hilbert, and Bourbaki. We hope that readers who are shocked by this will study the indictment of Bourbakism by the mathematician Morris Kline (1980), and then bear with us long enough to see the advantages of our approach. Examples appear in almost every chapter.

Comparisons

For many years, there has been controversy over 'frequentist' versus 'Bayesian' methods of inference, in which the writer has been an outspoken partisan on the Bayesian side. The record of this up to 1981 is given in an earlier book (Jaynes, 1983). In these old works there was a strong tendency, on both sides, to argue on the level of philosophy or ideology. We can now hold ourselves somewhat aloof from this, because, thanks to recent work, there is no longer any need to appeal to such arguments. We are now in possession of proven theorems and masses of worked-out numerical examples. As a result, the superiority of Bayesian methods is now a thoroughly demonstrated fact in a hundred different areas. One can argue with a philosophy; it is not so easy to argue with a computer printout, which says to us: 'Independently of all your philosophy, here are the facts of actual performance.' We point this out in some detail whenever there is a substantial difference in the final results. Thus we continue to argue vigorously for the Bayesian methods; but we ask the reader to note that our arguments now proceed by citing facts rather than proclaiming a philosophical or ideological position.

However, neither the Bayesian nor the frequentist approach is universally applicable, so in the present, more general, work we take a broader view of things. Our theme is simply: *probability theory as extended logic*. The 'new' perception amounts to the recognition that the mathematical rules of probability theory are not merely rules for calculating frequencies of 'random variables'; they are also the unique consistent rules for conducting inference (i.e. plausible reasoning) of any kind, and we shall apply them in full generality to that end.

It is true that all ‘Bayesian’ calculations are included automatically as particular cases of our rules; but so are all ‘frequentist’ calculations. Nevertheless, our basic rules are broader than either of these, and in many applications our calculations do not fit into either category.

To explain the situation as we see it presently: The traditional ‘frequentist’ methods which use only sampling distributions are usable and useful in many particularly simple, idealized problems; however, they represent the most proscribed special cases of probability theory, because they presuppose conditions (independent repetitions of a ‘random experiment’ but no relevant prior information) that are hardly ever met in real problems. This approach is quite inadequate for the current needs of science.

In addition, frequentist methods provide no technical means to eliminate nuisance parameters or to take prior information into account, no way even to use all the information in the data when sufficient or ancillary statistics do not exist. Lacking the necessary theoretical principles, they force one to ‘choose a statistic’ from intuition rather than from probability theory, and then to invent *ad hoc* devices (such as unbiased estimators, confidence intervals, tail-area significance tests) not contained in the rules of probability theory. Each of these is usable within the small domain for which it was invented but, as Cox’s theorems guarantee, such arbitrary devices always generate inconsistencies or absurd results when applied to extreme cases; we shall see dozens of examples.

All of these defects are corrected by use of Bayesian methods, which are adequate for what we might call ‘well-developed’ problems of inference. As Harold Jeffreys demonstrated, they have a superb analytical apparatus, able to deal effortlessly with the technical problems on which frequentist methods fail. They determine the optimal estimators and algorithms automatically, while taking into account prior information and making proper allowance for nuisance parameters, and, being exact, they do not break down – but continue to yield reasonable results – in extreme cases. Therefore they enable us to solve problems of far greater complexity than can be discussed at all in frequentist terms. One of our main purposes is to show how all this capability was contained already in the simple product and sum rules of probability theory interpreted as extended logic, with no need for – indeed, no room for – any *ad hoc* devices.

Before Bayesian methods can be used, a problem must be developed beyond the ‘exploratory phase’ to the point where it has enough structure to determine all the needed apparatus (a model, sample space, hypothesis space, prior probabilities, sampling distribution). Almost all scientific problems pass through an initial exploratory phase in which we have need for inference, but the frequentist assumptions are invalid and the Bayesian apparatus is not yet available. Indeed, some of them never evolve out of the exploratory phase. Problems at this level call for more primitive means of assigning probabilities directly out of our incomplete information.

For this purpose, the Principle of maximum entropy has at present the clearest theoretical justification and is the most highly developed computationally, with an analytical apparatus as powerful and versatile as the Bayesian one. To apply it we must define a sample space, but do not need any model or sampling distribution. In effect, entropy maximization creates a model for us out of our data, which proves to be optimal by so many different

criteria³ that it is hard to imagine circumstances where one would not want to use it in a problem where we have a sample space but no model.

Bayesian and maximum entropy methods differ in another respect. Both procedures yield the optimal inferences from the information that went into them, but we may choose a model for Bayesian analysis; this amounts to expressing some prior knowledge – or some working hypothesis – about the phenomenon being observed. Usually, such hypotheses extend beyond what is directly observable in the data, and in that sense we might say that Bayesian methods are – or at least may be – speculative. If the extra hypotheses are true, then we expect that the Bayesian results will improve on maximum entropy; if they are false, the Bayesian inferences will likely be worse.

On the other hand, maximum entropy is a nonspeculative procedure, in the sense that it invokes no hypotheses beyond the sample space and the evidence that is in the available data. Thus it predicts only observable facts (functions of future or past observations) rather than values of parameters which may exist only in our imagination. It is just for that reason that maximum entropy is the appropriate (safest) tool when we have very little knowledge beyond the raw data; it protects us against drawing conclusions not warranted by the data. But when the information is extremely vague, it may be difficult to define any appropriate sample space, and one may wonder whether still more primitive principles than maximum entropy can be found. There is room for much new creative thought here.

For the present, there are many important and highly nontrivial applications where Maximum Entropy is the only tool we need. Part 2 of this work considers them in detail; usually, they require more technical knowledge of the subject-matter area than do the more general applications studied in Part 1. All of presently known statistical mechanics, for example, is included in this, as are the highly successful Maximum Entropy spectrum analysis and image reconstruction algorithms in current use. However, we think that in the future the latter two applications will evolve into the Bayesian phase, as we become more aware of the appropriate models and hypothesis spaces which enable us to incorporate more prior information.

We are conscious of having so many theoretical points to explain that we fail to present as many practical worked-out numerical examples as we should. Fortunately, three recent books largely make up this deficiency, and should be considered as adjuncts to the present work: *Bayesian Spectrum Analysis and Parameter Estimation* (Bretthorst, 1988), *Maximum Entropy in Action* (Buck and Macaulay, 1991), and *Data Analysis – A Bayesian Tutorial* (Sivia, 1996), are written from a viewpoint essentially identical to ours and present a wealth of real problems carried through to numerical solutions. Of course, these works do not contain nearly as much theoretical explanation as does the present one. Also, the Proceedings

³ These concern efficient information handling; for example, (1) the model created is the simplest one that captures all the information in the constraints (Chapter 11); (2) it is the unique model for which the constraints would have been sufficient statistics (Chapter 8); (3) if viewed as constructing a sampling distribution for subsequent Bayesian inference from new data D , the only property of the measurement errors in D that are used in that subsequent inference are the ones about which that sampling distribution contained some definite prior information (Chapter 7). Thus the formalism automatically takes into account all the information we have, but avoids assuming information that we do not have. This contrasts sharply with orthodox methods, where one does not think in terms of information at all, and in general violates both of these desiderata.

volumes of the various annual MAXENT workshops since 1981 consider a great variety of useful applications.

Mental activity

As one would expect already from Pólya's examples, probability theory as extended logic reproduces many aspects of human mental activity, sometimes in surprising and even disturbing detail. In Chapter 5 we find our equations exhibiting the phenomenon of a person who tells the truth and is not believed, even though the disbelievers are reasoning consistently. The theory explains why and under what circumstances this will happen.

The equations also reproduce a more complicated phenomenon, divergence of opinions. One might expect that open discussion of public issues would tend to bring about a general consensus. On the contrary, we observe repeatedly that when some controversial issue has been discussed vigorously for a few years, society becomes polarized into two opposite extreme camps; it is almost impossible to find anyone who retains a moderate view. Probability theory as logic shows how two persons, given the same information, may have their opinions driven in opposite directions by it, and what must be done to avoid this.

In such respects, it is clear that probability theory is telling us something about the way our own minds operate when we form intuitive judgments, of which we may not have been consciously aware. Some may feel uncomfortable at these revelations; others may see in them useful tools for psychological, sociological, or legal research.

What is 'safe'?

We are not concerned here only with abstract issues of mathematics and logic. One of the main practical messages of this work is the great effect of prior information on the conclusions that one should draw from a given data set. Currently, much discussed issues, such as environmental hazards or the toxicity of a food additive, cannot be judged rationally if one looks only at the current data and ignores the prior information that scientists have about the phenomenon. This can lead one to overestimate or underestimate the danger.

A common error, when judging the effects of radioactivity or the toxicity of some substance, is to assume a linear response model without threshold (i.e. without a dose rate below which there is no ill effect). Presumably there is no threshold effect for cumulative poisons like heavy metal ions (mercury, lead), which are eliminated only very slowly, if at all. But for virtually every organic substance (such as saccharin or cyclamates), the existence of a finite metabolic rate means that there must exist a finite threshold dose rate, below which the substance is decomposed, eliminated, or chemically altered so rapidly that it causes no ill effects. If this were not true, the human race could never have survived to the present time, in view of all the things we have been eating.

Indeed, every mouthful of food you and I have ever taken contained many billions of kinds of complex molecules whose structure and physiological effects have never been determined – and many millions of which would be toxic or fatal in large doses. We cannot

doubt that we are daily ingesting thousands of substances that are far more dangerous than saccharin – but in amounts that are safe, because they are far below the various thresholds of toxicity. At present, there are hardly any substances, except some common drugs, for which we actually know the threshold.

Therefore, the goal of inference in this field should be to estimate not only the slope of the response curve, but, *far more importantly*, to decide whether there is evidence for a threshold; and, if there is, to estimate its magnitude (the ‘maximum safe dose’). For example, to tell us that a sugar substitute can produce a barely detectable incidence of cancer in doses 1000 times greater than would ever be encountered in practice, is hardly an argument against using the substitute; indeed, the fact that it is necessary to go to kilodoses in order to detect any ill effects at all, is rather conclusive evidence, not of the danger, but of the *safety*, of a tested substance. A similar overdose of sugar would be far more dangerous, leading not to barely detectable harmful effects, but to sure, immediate death by diabetic coma; yet nobody has proposed to ban the use of sugar in food.

Kilodose effects are irrelevant because we do not take kilodoses; in the case of a sugar substitute the important question is: What are the threshold doses for toxicity of a sugar substitute and for sugar, compared with the normal doses? If that of a sugar substitute is higher, then the rational conclusion would be that the substitute is actually safer than sugar, as a food ingredient. To analyze one’s data in terms of a model which does not allow even the possibility of a threshold effect is to prejudge the issue in a way that can lead to false conclusions, however good the data. If we hope to detect any phenomenon, we must use a model that at least allows the *possibility* that it may exist.

We emphasize this in the Preface because false conclusions of just this kind are now not only causing major economic waste, but also creating unnecessary dangers to public health and safety. Society has only finite resources to deal with such problems, so any effort expended on imaginary dangers means that real dangers are going unattended. Even worse, the error is incorrectible by the currently most used data analysis procedures; a false premise built into a model which is never questioned cannot be removed by any amount of new data. Use of models which correctly represent the prior information that scientists have about the mechanism at work can prevent such folly in the future.

Such considerations are not the only reasons why prior information is essential in inference; the progress of science itself is at stake. To see this, note a corollary to the preceding paragraph: that new data that we insist on analyzing in terms of old ideas (that is, old models which are not questioned) *cannot lead us out of the old ideas*. However many data we record and analyze, we may just keep repeating the same old errors, missing the same crucially important things that the experiment was competent to find. That is what ignoring prior information can do to us; no amount of analyzing coin tossing data by a stochastic model could have led us to the discovery of Newtonian mechanics, which alone determines those data.

Old data, when seen in the light of new ideas, can give us an entirely new insight into a phenomenon; we have an impressive recent example of this in the Bayesian spectrum analysis of nuclear magnetic resonance data, which enables us to make accurate quantitative determinations of phenomena which were not accessible to observation at all with the