PREFACE

This book has grown over several years from a nucleus consisting of transcripts of tape recordings of a series of lectures given at the Field Research Laboratories of the Socony-Mobil Oil Company in Dallas, Texas, during March, 1958 and June, 1963. The lectures were given also, with gradually increasing content, at Stanford University in 1958, at the University of Minnesota in 1959, at the University of California, Los Angeles in 1960 and 1961, at Purdue University in 1962, at Dartmouth College in 1962 and 1963, at the Standard Oil Company Research Laboratories, Tulsa, in 1963, at the University of Colorado in 1964, at the University of Maryland in 1968, and at Washington University in 1966, 1969, 1970 and 1972. The material of lectures 1-10 and 16-17 was issued by the Socony-Mobil Oil Company as Number 4 in their series, "Colloquium Lectures in Pure and Applied Science", and is reproduced here, with permission, in considerably expanded form.

In editing and adding new material, the informal style of the original presentation has been retained. This and the general format are intended to emphasize that the book is in no sense a textbook or complete treatise, but only a series of informal conversations (necessarily rather one-sided), concerning the foundations of probability theory and how to use it for current applications in physics, chemistry, and engineering. The speaker is simply sharing his views with the audience, and trying to give some more or less convincing arguments in support of them. Often, the trend of a lecture was determined by questions raised from the audience.

The material is addressed primarily to scientists and engineers who are already familiar with applied mathematics and perhaps with certain special uses of probability theory, such as statistical mechanics, communication theory, or data analysis, but who may not have had the time to make an extensive study of modern statistics. Such persons may be appalled, as I was when I commenced serious study of the field in 1950, by the enormous volume of literature dealing with statistical problems, and may despair of ever mastering it--not because it is too advanced, but simply because the field is too large. There is so much diverse and intricate detail that it is almost impossible to locate the underlying principles; and one finally succeeds only to have them dissolve in confusion and controversy, no two authors being in agreement about them.

For such persons, we have good news. Recently, a great simplification and unification of this field has become possible. There is a single very simple set of principles, which can be stated in a few lines and which, when applied to specific problems, will be found to give automatically conventional probability theory, the formalism of equilibrium and nonequilibrium statistical mechanics, the results of communication theory, and the newest methods of statistical inference, which represent a great advance over the "orthodox" methods prevalent in the period 1930-1960.

These principles are summarized on the inside front and back covers of this book. Although at present we are able to give only heuristic (but nevertheless convincing) arguments for their uniqueness, there is no difficulty in demonstrating that they do include the aforementioned applications as special cases. Therefore, whatever changes in viewpoint may come in the future, these principles will retain at least their didactic value, as a concise summary of presently known statistical methods.

Current applications have advanced to the point where the perennial confusion surrounding foundations of probability theory now poses a serious threat to further progress. In particular, we have struggled for over two centuries with concentual problems involving the relation between abstract probability theory and the "real" world. Should we use probability only in the sense of describing frequencies in some "random experiment", or is it legitimate to interpret the mathematical rules in the broader Laplace sense of a "calculus of inductive reasoning?" In my opinion, the time has come when such questions must be settled. Until this is done we cannot hope to resolve the paradoxes of quantum theory and irreversible statistical mechanics, or even to justify the use of probability theory in describing time-dependent phenomena.

Because of conceptual difficulties with Laplace's viewpoint, many attempts have been made to evade the general problem of inductive reasoning, and to develop probability theory from more restrictive postulates concerning limiting frequencies (i.e. the von Mises "collective", etc.). This approach encountered such great mathematical and logical difficulties that it has been almost completely abandoned; but strangely enough, the intolerance of broader views of probability has survived. Thus, today most writers on probability and statistics take the curious position of admitting that a probability cannot be defined merely as a frequency, but still insisting that it must always be interpreted as a frequency in applications.

The theory developed here is more general than in conventional expositions because the rules are derived in a way that makes it clear that neither the notion of probability nor the mathematical rules of probability theory depend on such concepts as random variables, random experiments, or relative frequencies.

According to the viewpoint expounded here, consideration of random experiments is only one particular application (and not even the most important one) of probability theory; the rules apply equally well to general inductive inferences where no random experiment is involved in any way. Indeed, most applications of probability theory can be formulated and carried to completion without ever introducing the notion of a "random variable". This is demonstrated repeatedly in the following text, particularly in Lectures 5, 6, 8, 9, 11 and 18.

In our emphasis on the conceptual, rather than the purely mathematical, problems we are necessarily dealing almost constantly with controversial aspects of probability theory. One of the most difficult problems of principle confronting a person trying to apply the theory (treatment of prior information) has been debated vigorously on philosophical grounds for over a century, without being brought perceptibly nearer solution. In this book (particularly, Lectures 10 and 12) we are able to report some progress in reducing these vague philosophical questions to definite and answerable mathematical ones, and in sufficient generality to cover most current applications. However, we make no claim to have fully resolved the situation, about which L. J. Savage has remarked that "there has seldom been such complete disagreement and breakdown of communication since the Tower of Babel." Indeed, we make no claim to have proved anything at all, in a sense which would be accepted as rigorous by modern mathematician. But the arguments given here are, I believe, sufficiently compelling to justify a claim that we have shifted the burden of proof back to those writers who persist in asserting that the Laplace viewpoint is nonsense, and only the strict frequency interpretation is respectable.

The idea of independent repetitions of a random experiment, which the "frequentist" usually considers essential to the very notion of probability, is from our standpoint only a very special case of an exchangeable sequence. We are able to give, in Lectures 16 and 17, a fairly complete discussion of connections between probability and frequency in this case, via straightforward mathematical deduction without any appeal to arbitrary postulates about the relation between them. Similar, and equally general, connections arise in nearly every other application. As a result, we will claim--not as a theorem, but as an inductive generalization to which no exception has been found--that there is never any need to postulate such relations. <u>All the connections between probability and frequency that are actually used in applications, far from conflicting with the Laplace-Bayes "inductive reasoning" form of probability theory, are derivable as elementary mathematical consequences of that theory.</u>

A word of explanation and apology to mathematicians who may happen on this book not written for them; you will find here no appeal to the notions of Borel fields and Radon-Nikodym derivatives, no use of sets or measure theory other than an occasional "almost everywhere" remark, and no Lebesgue-Stieltjes integrals. I am not opposed to these things, and will gladly use and teach them as soon as I find one specific real application where they are needed. Never having uncovered such a problem, either in my own work or in all the statistical literature known to me, but knowing that their introduction would discourage many from reading this book, I have decided to forego them. From the standpoint of probability theory as I see it, they add little rigor to the subject, but serve rather to generalize and extend it in a direction different from the one we are traveling. We get along happily and without impediment using Riemannian integrals with integrands interpreted, when convenient, in the sense of generalized functions. It is well established that, in fourier analysis, this procedure is actually more powerful and appropriate to the subject than the measure-theoretic approach. I think a good case can be made for the view that this holds also in probability theory.

No author can hope to give proper acknowledgement to all those who have influenced his thinking; the list would be as long as the book. In my own case, however, the greatest debts must be to Sir Harold Jeffreys, R. T. Cox, C. E. Shannon, and G. Polya, the last three for reasons which will be clear from the text. In the case of Jeffreys, I would like to recall the following anecdote.

When, as a student in 1946, I decided that I ought to learn some probability theory, it was pure chance which led me to take the book "Theory of Probability" by Jeffreys, from the library shelf. In reading it, I was puzzled by something which, I am afraid, will also puzzle many who read the present book. Why was he so much on the defensive? It seemed to me that Jeffreys' viewpoint and most of his statements were the most obvious common sense; I could not imagine any sane person disputing them. Why, then, did he feel it necessary to insert so many interludes of argumentation vigorously defending his viewpoint? Wasn't he belaboring a straw man?

This suspicion disappeared quickly a few years later when I consulted another well-known book on probability (Feller, 1950) and began to realize what a fantastic situation exists in this field. The whole approach of Jeffreys was summarily rejected as metaphysical nonsense, without even a description. The author assured us that Jeffreys' methods of estimation, which seemed to me so simple and satisfactory, were completely erroneous, and wrote in glowing terms about the success of a "modern theory", which had abolished all these mistakes.

Naturally, I was eager to learn what was wrong with Jeffreys' methods, why such glaring errors had escaped me, and what the new, improved methods were. But when I tried to find the new methods for handling estimation problems (which Jeffreys could formulate in two or three lines of the most elementary mathematics), I found that the new book did not contain them. On the contrary, the reader was told that these problems did not belong to probability theory at all, but to a new field, statistical inference, which was based on entirely different principles, was very advanced, and should be studied only after one had mastered probability theory and measure theory.

Throwing caution to the winds, I then took down an armful of advanced texts in statistics. Here I found an entirely new vocabulary, new mathematical demonstrations, the most meticulous attention to minutiae which I could not conceive of as being relevant to any application, but no underlying unity of method. I was particularly interested in problems of parameter estimation of the type which arise in extracting signals from noise; but instead of giving a single method applicable to all such problems, as Jeffreys did, the authors would give several different methods for treating each individual problem. They gave demonstrably different results, and the reader was left with nothing to choose between them. An "estimator" ought to be, if possible sufficient, unbiased, efficient, or asymptotically so; but the reader could find neither a clear statement of the relative importance of these, nor any general rule by which an estimator with such properties could be constructed. Instead, the procedure was merely to guess various functions of the sample values on intuitive grounds, and then test them for bias, efficiency, etc.

On the other hand, Jeffreys' method (which was, of course, application of Bayes' theorem in essentially the same way Laplace had used it) told us at once which estimator should be used. It was a revelation to me to read Jeffreys' beautifully clear explanation of why the sample mean is not the best estimate of the population mean except in the special case of Gaussian distribution; in all other cases one should use a weighted average of the sample values. All of a sudden I could see the justification for the physicist's common-sense practice of discarding measurements which show a wide deviation, about which mathematical writers still complain; and a refinement of this in which the theory tells us just how to assign less weight to more widely deviating measurements. But none of this was to be found in the books on statistical inference. Here, for example, authors quote the well-known proposition that for a Cauchy distribution the mean of an arbitrarily large sample is no better estimate of the population mean (by the criterion of efficiency) than a single observation, and that the sample median has a rather good asymptotic efficiency; but they do not offer us any reasonable estimator for the small-sample case. Jeffreys' method determines a definite weighted average estimator which is better than the median for any sample size, and much better in the case of small samples; but I have yet to find an orthodox writer who uses it; or even acknowledges its existence.

Observing these things, I was completely mystified by every author's contemptuous dismissal of Jeffreys' method, which was done invariably on purely philosophical grounds, without letting the reader see how it works in even a single application.

If yousay that method A is better than method B, I think you ought to mean by this, at the very minimum, that there is at least one specific problem where it leads to a better answer; and to prove your point you need only exhibit that problem. But I could not, then or since, find any orthodox writer who had produced any such example, except for one case (noted in Lecture 7) where there was an error in the calculation, and a few others (Lecture 16) based on gross misapplication of the Bayesian method (i.e., taking the solution to one problem and complaining that it is not also the solution to an entirely different problem). Indeed, on working out a few cases for myself, the outcome was invariably the same; whenever there was any appreciable difference in the results, it was Jeffreys' result which clearly agreed best with common sense. Once one understood the mathematical situation, it was easy to invent problems where the orthodox statisticians' methods broke down entirely and gave absurd results; but I was unable to produce any case where Jeffreys' method, properly applied, failed to lead to an intuitively reasonable conclusion.

All this took place just at the time of appearance of Wald's book, "Statistical Decision Functions". It required several years for the full implications of this monumental work to be appreciated, but by now many workers in statistics have recognized the source of, and remedy for, all this confusion.

The details occupy much of Lectures 5, 6, 13, 14; but stated in the briefest terms, the mathematical situation uncovered by Wald showed that in all respects that matter in real applications, Jeffreys was right all along. The most important recent advance in statistics has taken us right back to the methods developed by Bayes, Laplace, and Daniel Bernoulli in the 18'th century, which generations of statisticians have held to be nonsense. For twenty years, the physicist who was fortunate enough to consult Jeffreys' book had at his fingertips statistical methods which were simpler, more general, and more powerful than anything the orthodox statistician had to offer.

Needless to say, the above assertions are not going to be accepted overnight in all quarters. If the reader is puzzled by my repeated lapses into argumentation, I ask him to realize that I am not only trying to be constructive and give a unified method, but I am also trying to answer in a a single volume all the objections to this method which have filled the statistical literature for sixty years. In these sections, I am, in effect, supplying the reader with ammunition which he will need if he tries to discuss these issues with colleagues who have been trained only in the "orthodox" point of view. In this connection, I would like to express my gratitude to two anonymous reviewers of this book who gave valuable suggestions on how to strengthen these arguments, and most of all to a third reviewer, who by his objections reassured me more than any friendly reviewer could possibly have done, that in these sections I am not wasting time and space belaboring a straw man.

The Galileo Strategy. Recently, my attention was called to a remarkable article, "Linguistic Analysis of a Statistical Controversy", by Irwin D. J. Bross (1963),* which attacks Bayesian methods in a way that cries out for a constructive reply. I hope that this book may serve that purpose; and to make it "constructive" we need to recognize that further debate on the philosophical level would be not only fruitless, but it would miss the real issue facing us today. As already noted, we have been debating the philosophical issue for well over a century, and perhaps no great harm will be done if it goes on for another century. But there is a far more important issue which should, and I think can, be settled quickly.

The question of immediate importance is not whether Bayesian methods are 100 per cent perfect, or whether their underlying philosophy is opprobrious, but simply whether, at the present time, they are better or worse than orthodox methods from the standpoint of (a) the actual results they give in practice, (b) the range of problems where they can be applied, and (c) their ease of application.

For example, a large amount of reliability and quality-control testing is needed in modern technology. In some cases, particularly, in the aero-space field, acquisition of a single data point can cost more than the yearly salary of a statistician. Use of statistical methods that fail to extract all the relevant information from a sample, or fail to make use of relevant prior information, is therefore not only illogical; it can lead to staggering economic waste.

As another example, statistical methods are destined for an every-increasing role in helping make fundamental military and governmental policy decisions. In this case, use of methods that fail to make full use of the available information might lead to consequences whose magnitude cannot be measured in economic terms at all. In a very real sense, statistics has become too important to allow its methods to be determined merely by the relative numbers, or aggressiveness, of two parties in a philosophical dispute.

---

*Full references are given in Appendix A.

To assert the superiority of either approach on grounds of some philosophy about the "true meaning of probability" without examining the facts concerning performance in specific cases would be to cast out everything we have learned about scientific methodology and to return to the methods of that learned Doctor of the seventeenth century, who assured the world his theology had proved there could be no moons about Jupiter, and steadfastly refused to look through Galileo's telescope.

Since 1953, I have been making constant routine use of Bayesian methods in statistical problems of physics and engineering, and comparing their results with those obtained by orthodox methods. As a result, I believe that the practical issue can be removed entirely from the realm of ideology, and settled on the level of demonstrable fact. To indicate how this can be done, I offer this book as a small, but revealing glimpse through the Galileo telescope of statistics.

To define our field of view let us start, as did Bross, by quoting the words of J. W. Pratt (1963): "Now that I have ceased pretending to be impartial, I may point out that no connected argument leading to the orthodox methods has ever been advanced. Neyman and Pearson contributed vitally to our understanding by their <u>formulation</u> of statistical problems, but they have never claimed their <u>methods</u> were more than ad hoc procedures with some pleasant properties. Their methods, while extremely ingenious and useful, are not completely satisfactory, let alone uniquely objective and scientific."

Unlike Bross, I am unable to discern any "Neo-Bayesian jargon" or "incongruencies" here; only a clear and accurate statement of fact. But, since a lengthy attack on this statement has been published, it will be of interest to see how it can be defended. Bross particularly objects to the remark that "no connected argument" has been advanced for the orthodox methods, and he specifically brings up the matter of significance tests and confidence intervals. Therefore, we will give particular scrutiny to these topics [as previewed in Jaynes (1973)] when we come to them in the course of the lectures. In fact, we give a quite general proof that these methods, when improved to the maximum possible degree by taking into account all ancillary statistics, lead finally to just the results that could have been derived in three lines by the methods of Laplace.

I am indebted to S. R. Faris, John Heller, L. Massé, S. M. Foulks, and many other workers at the Socony-Mobil Field Research Laboratories in Dallas, for their kind hospitality during my two visits, and for undertaking the monumental task of preparing a typed copy of the lectures, complete with equations, from the tape recordings. Only a person who has done this kind of work can realize how much labor is involved.

Many of the details given in connection with applications in the last half of the book were worked out by, and appear in the doctoral theses of, my students, Perry Vartanian, Steve Heims,

Larry Davis, Ray Nelson, Douglas Scalapino, Baldwin Robertson, Joel Snow, Wm. C. Mitchell, and Charles Tyler.

I am indebted also to several hundred students and colleagues who have attended these lectures at various places. By their penetrating questions they have forced me to think much more carefully about many of the issues raised in these talks.

Finally, it should be emphasized again that in most places the text is a literal transcript of actual lectures, and that expressions used in speaking are not always those one would use in writing. In particular, the term "statistician" is often used as an abbreviation for "statistician of the extreme objectivist school of thought which has dominated the field for several decades." In actual fact, many statisticians are well aware of the points made here, and would find themselves in agreement with my arguments. We have noted the views of Pratt; as other examples one can cite the work of D. V. Lindley (1965) and I. J. Good (1959) who outlines a "neoclassical" theory in which "Bayes' theorem is restored to a primary position from which it had been deposed by the orthodox statisticians..." I hope that statisticians of the neoclassical persuasion (whose numbers are rapidly increasing) will understand, and not be offended by, my use of the term.

E. T. Jaynes
St. Louis, Missouri

June 1973