Subjective and Objective Bayesian Statistics

Principles, Models, and Applications

Second Edition

S. JAMES PRESS

with contributions by

SIDDHARTHA CHIB

MERLISE CLYDE

GEORGE WOODWORTH

ALAN ZASLAVSKY

WILEY-INTERSCIENCE

A John Wiley & Sons, Inc., Publication

This Page intentionally left blank

Subjective and Objective Bayesian Statistics

Second Edition

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

A complete list of the titles in this series appears at the end of this volume.

Subjective and Objective Bayesian Statistics

Principles, Models, and Applications

Second Edition

S. JAMES PRESS

with contributions by

SIDDHARTHA CHIB

MERLISE CLYDE

GEORGE WOODWORTH

ALAN ZASLAVSKY

WILEY-INTERSCIENCE

A John Wiley & Sons, Inc., Publication

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data is available.

ISBN 0-471-34843-0

10 9 8 7 6 5 4 3 2 1

To my Family
G, D, S, and all the J's

Reason, Observation, and Experience—The Holy Trinity of Science

—Robert G. Ingersoll (1833–1899)

The Reverend Thomas Bayes

This sketch of the person we believe to be Thomas Bayes was created by Rachel Tanur and is reproduced here by permission of her estate.

CONTENTS

This Page intentionally left blank

PREFACE

This second edition is intended to be an introduction to Bayesian statistics for students and research workers who have already been exposed to a good preliminary statistics and probability course, probably from a frequentist viewpoint, but who have had a minimal exposure to Bayesian theory and methods. We assume a mathematical level of sophistication that includes a good calculus course and some matrix algebra, but nothing beyond that. We also assume that our audience includes those who are interested in using Bayesian methods to model real problems, in areas that range across the disciplines.

This second edition is really a new book. It is not merely the first edition with a few changes inserted; it is a completely restructured book with major new chapters and material.

The first edition to this book was completed in 1988. Since then the field of Bayesian statistical science has grown so substantially that it has become necessary to rewrite the story in broader terms to account for the changes that have taken place, both in new methodologies that have been developed since that time, and in new techniques that have emerged for implementing the Bayesian paradigm. Moreover, as the fields of computer science, numerical analysis, artificial intelligence, pattern recognition, and machine learning have also made enormous advances in the intervening years, and because their interfaces with Bayesian statistics have steadily increased, it became important to expand our story to include, at least briefly, some of those important interface topics, such as *data mining tree models* and *Bayesian neural networks*. In addition, as the field of Bayesian statistics has expanded, the applications that have been made using the Bayesian approach to learning from experience and analysis of data now span most of the disciplines in the biological, physical, and social sciences. This second edition attempts to tell the broader story that has developed.

One direction of growth in Bayesian statistics that has occurred in recent years resulted from the contributions made by Geman and Geman (1984), Tanner and Wong (1987), and Gelfand and Smith (1990). These papers proposed a new method, now called *Markov chain Monte Carlo* (or just MCMC), for applying and implementing Bayesian procedures numerically. The new method is computer intensive and involves sampling by computer (so-called Monte Carlo sampling) from the posterior distribution to obtain its properties. Usually, Bayesian modeling procedures result in ratios of multiple integrals to be evaluated numerically. Sometimes these multiple integrals are high dimensional. The results of such Bayesian analysis are wonderful theoretically because they arise from a logical, self-consistent, set of axioms for making judgments and decisions. In the past, however, to evaluate such ratios of high-dimensional multiple integrals numerically it was necessary to carry out tedious numerical computations that were difficult to implement for all but the very computer-knowledgeable researcher. With a computer environment steadily advancing from the early 1980s, and with the arrival of computer software to implement the MCMC methodology, Bayesian procedures could finally be implemented rapidly, and accurately, and without the researcher having to possess a sophisticated understanding of numerical methods.

In another important direction of growth of the field, Bayesian methodology has begun to recognize some of the implications of the important distinction between *subjective* and *objective* prior information. This distinction is both philosophical and mathematical. When information based upon underlying theory or historical data is available (subjective prior information), the Bayesian approach suggests that such information be incorporated into the prior distribution for use in Bayesian analysis. If families of prior distributions are used to capture the prior knowledge, such prior distributions will contain their own parameters (called *hyperparameters*) that will need to be assessed on the basis of the available information. For example, many surveys are carried out on the same topic year after year, so that results obtained in earlier years can be used as a best guess for what is likely to be obtained in a new survey in the current year. Such "best available" information can be incorporated into a *prior distribution*. Such prior distributions are always proper (integrate or sum to one), and so behave well mathematically. A Bayesian analysis using such a prior distribution is called *subjective* Bayesian analysis.

In some situations, however, it is difficult to specify appropriate subjective prior information. For example, at the present time, there is usually very little, if any, prior information about the function of particular sequences of nucleotide base pairs in the DNA structure of the human genome. In such situations it is desirable to have meaningful ways to begin the Bayesian learning updating process. A prior distribution adopted for such a situation is called *objective*, and an analysis based upon such an objective prior distribution is called an *objective* Bayesian analysis. Such analyses serve to provide benchmark statistical inferences based upon having inserted as little prior information as possible, prior to taking data. Objective prior distributions correspond to "knowing little" prior to taking data. When such prior distributions are continuous, it is usually the case that these (improper) prior distributions do not integrate to one (although acceptable *posterior* distributions that correspond to these improper prior distributions must integrate to one). Sometimes, in simple cases, posterior inferences based upon objective prior distributions will result in inferences that correspond to those arrived at by frequentist means. The field has begun to focus on the broader implications of the similarities and differences between subjective and objective types of information. We treat this important topic in this edition, and recognize its importance in the title of the book. In many applications of interest, there is not enough information in a problem for classical inference to be carried out. So some researchers resort to subjective Bayesian inference out of necessity. The subjective Bayesian approach is adopted because it is the most promising way to introduce sufficient additional information into the problem so that a real solution can be found.

In earlier years, it was difficult to take into account uncertainty about which model to choose in a Bayesian analysis of data. Now we are learning how to incorporate such uncertainty into the analysis by using *Bayesian model averaging*. Moreover, we have been learning how to use Bayesian modeling in a *hierarchical* way to represent nested degrees of uncertainty about a problem. A whole new framework for *exploratory factor analysis* has been developed based upon the Bayesian paradigm. These topics are new and are discussed in this edition.

In this edition, for the first time, we will present an extensive listing, by field, of some of the broad-ranging applications that have been made of the Bayesian approach.

As Bayesian statistical science has developed and matured, its principal founders and contributors have become apparent. To record and honor them, in this edition we have included a *Bayesian Hall of Fame*, which we developed by means of a special opinion poll taken among senior Bayesian researchers. Following the table of contents is a collection of the portraits and brief biographies of these most important contributors to the development of the field, and there is an appendix devoted to an explanation of how the members of the Hall of Fame were selected.

The first edition of this book contained eight chapters and four appendices; this edition contains 16 chapters, generally quite different from those in the first edition, and seven appendices. The current coverage reflects not only the addition of new topics and the deletion of some old ones, but also the expansion of some previously covered topics into greater depth, and more domains. In addition, there are solutions to some of the exercises.

This second edition has been designed to be used in a year-long course in Bayesian statistics at the senior undergraduate or graduate level. If the academic year is divided into semesters, Chapters 1–8 can be covered in the first semester and Chapters 9–16 in the second semester. If the academic year is divided into quarters, Chapters 1–5 (Part I) can be covered in the fall quarter, Chapters 6–11 (Parts II and III) in the winter quarter, and Chapters 12–16 (Part IV) in the spring quarter.

Three of the sixteen chapters of this second edition have been written with the assistance of four people: Chapter 6 by Professor Siddhartha Chib of Washington University; Complement A to Chapter 6 by Professor George Woodworth of the University of Iowa; Chapter 13 by Professor Merlise Clyde of Duke University; and Chapter 14 by Professor Alan Zaslavsky of Harvard University. I am very grateful for their help. Much of Appendix 7 was written with the help of my former students, Dr. Thomas Ferryman, Dr. Mahmood Ghamsary, and Ms. Dawn Kummer. I am also grateful to Stephen Quigley of John Wiley and Sons, Inc., who encouraged me to prepare this second edition, and to Heather Haselkorn of Wiley, who helped and

prodded me until it was done. Dr. Judith Tanur helped me to improve the exposition and to minimize the errors in the manuscript. The remaining errors are totally my responsibility. I am grateful to Rachel Tanur for her sketch of Thomas Bayes at the beginning of the book. Her untimely death prevented her from her intention of also sketching the scientists who appear in the Bayesian Hall of Fame. Dr. Linda Penas solved some of our more complex LaTex editorial problems, while Ms. Peggy Franklin typed some of the chapters in LaTex with indefatigable patience and endurance.

S. JAMES PRESS

Oceanside, CA
September, 2002

PREFACE TO THE FIRST EDITION

This book is intended to be an introduction to Bayesian statistics for students and research workers who have already been exposed to a good preliminary statistics and probability course from a classical (frequentist) point of view but who have had minimal exposure to Bayesian theory and methods. We assume a mathematical level of sophistication that includes a good calculus course and some matrix algebra but nothing beyond that. We also assume that our audience includes those who are interested in using Bayesian methods to model real problems in the various scientific disciplines. Such people usually want to understand enough of the foundational principles so that they will (1) feel comfortable using the procedures, (2) have no compunction about recommending solutions based upon these procedures to decision makers, and (3) be intrigued enough to go to referenced sources to seek additional background and understanding. For this reason we have tried to maximize interpretation of theory and have minimized our dependence upon proof of theorems.

The book is organized in two parts of four chapters each; in addition, the back of the book contains appendixes, a bibliography, and separate author and subject indexes. The first part of the book is devoted to theory; the second part is devoted to models and applications. The appendixes provide some biographical material about Thomas Bayes, along with a reproduction of Bayes's original essay.

Chapter I shows that statistical inference and decision making from a Bayesian point of view is based upon a logical, self-consistent system of axioms; it also shows that violation of the guiding principles will lead to "incoherent" behavior, that is, behavior that would lead to economically unsound decisions in a risky situation.

Chapter II covers the basic principles of the subject. Bayes's theorem is presented for both discrete and absolutely continuous random variables.

We discuss Bayesian estimation, hypothesis testing, and decision theory. It is here that we introduce prior distributions, Bayes' factors, the important theorem of de Finetti, the likelihood principle, and predictive distributions.

Chapter III includes various methods for approximating the sometimes complicated posterior distributions that result from applications of the Bayesian paradigm. We present large-sample theory results as well as Laplacian types of approximations of integrals (representing posterior densities). We will show how *importance sampling* as well as *simulation* of distributions can be used for approximation of posterior densities when the dimensions are large. We will also provide a convenient up-to-date summary of the latest Bayesian computer software available for implementation.

Chapter IV shows how prior distributions can be assessed subjectively using a group of experts. The methodology is applied to the problem of using a group of experts on strategic policy to assess a multivariate prior distribution for the probability of nuclear war during the decade of the 1980s.

Chapter V is concerned with Bayesian inference in both the univariate and multivariate regression models. Here we use vague prior distributions, and we apply the notion of predictive distributions to predicting future observations in regression models.

Chapter VI continues discussion of the general linear model begun in Chapter V, only here we show how to carry out Bayesian analysis of variance and covariance in the multivariate case. We will invoke the de Finetti notion of exchangeability (of the population mean vector distributions).

Chapter VII is devoted to the theory and application of Bayesian classification and discrimination procedures. The methodology is illustrated by appylying it to the sample survey problem of second guessing "undecided" respondents.

Chapter VIII presents a case study of how disputed authorship of some of the Federalist papers was resolved by means of a Bayesian analysis.

The book is easily adapted to a one- or two-quarter sequence or to a one-semester, senior level, or graduate course in Bayesian statistics. The first two chapters and the appendixes could easily fill the first quarter, with Chapters III–VIII devoted to the second quarter. In a one-quarter or one-semester course, certain sections or chapters would need to be deleted; which chapters or sections to delete would depend upon the interests of the students and teacher in terms of the balance desired between (1) theory and (2) models and applications.

The book represents an expansion of a series of lectures presented in South Australia in July 1984 at the University of Adelaide. These lectures were jointly sponsored by the Commonwealth Scientific and Industrial Research Organization (CSIRO), Division of Mathematics and Statistics and by the University of Adelaide's Departments of Economics and Statistics. I am grateful to Drs. Graham Constantine, William Davis, and Terry Speed, all of CSIRO, for their stimulating comments on the original lecture material, for their encouragement and support, and for planting the seeds from which this monograph grew. I am grateful to Dr. John Darroch, Dr. Alastair Fischer, Dr. Alan James, Dr. W. N. Venables, and to other participants of the lecture series for their stimulating questions that helped to put the book into perspective. Dr. John Pratt and Dr. S. L. Zabell helped to clarify the issues about de Finetti's theorem in Section 2.9.3, and Dr. S. K. Sinha suggested an example used in Section 2.7.1. Dr. Persi Diaconis and Dr. Richard Jeffrey presented stimulating discussions about randomness, exchangeability, and some of the foundational issues of the subject in a seminar at Stanford University during winter quarter of 1984–1985, a sabbatical year the author spent visiting Stanford University. I am deeply grateful to Drs. Harry Roberts and Arnold Zellner for exposing me to Bayesian ideas. Dr. Stephen Fienberg provided encouragement and advice regarding publishing the manuscript. I am also grateful to Dr. Stephen Fienberg, Dr. Ingram Olkin, and an anonymous publisher's referee for many helpful suggestions for improving the presentation. I am very grateful for suggestions made by Dr. Judith Tanur who read the entire manuscript; to Dr. Ruben Klein who read Chapters I and II; and to Drs. Frederick Mosteller and David Wallace who read Chapter VIII. I also wish to thank graduate students, James Bentley, David Guy, William Kemple, Thomas Lucas, and Hamid Namini whose questions about the material during class prompted me to revise and clarify various issues. Mrs. Peggy Franklin is to be congratulated for her outstanding typing ability and for her forbearance in seeing me through the many iterations that the manuscript underwent. We think we have eliminated most, if not all, errors in the book, but readers could help the author by calling any additional ones they find to his attention.

S. JAMES PRESS

Riverside, California
January, 1989

A BAYESIAN HALL OF FAME

Bayes, Thomas
1701–1761

DeFinetti, Bruno
1906–1985

DeGroot, Morris
1931–1989

Jeffreys, Harold
1891–1989

Lindley, Dennis V.
1923–

Savage, Leonard J.
1917–1971

THOMAS BAYES (1710–1761)
Bayes' Theorem of inverse probability is attributed to Thomas Bayes. Bayes applied the theorem to binomial data in his essay. The essay was published posthumously in 1763. Bayesian statistical science has developed from applications of his theorem.

BRUNO DE FINETTI (1906–1985)
One of the founders of the subjectivist Bayesian approach to probability theory (*Introduction to Probability Theory*, 1970, 2 volumes). He introduced the notion of *exchangeability* in probability as a weaker concept than independence. De Finetti's theorem on exchangeability yields the strong law of large numbers as a special case.

MORRIS DE GROOT (1931–1989)
Morris DeGroot was a student of Jimmie Savage and a leader in the Bayesian statitics movement. He contributed to the field in a wide-ranging spectrum of fields from economic utility and decision theory (*Optimal Statistical Decisions*, 1970) to genetics and the history of statistics. He founded the Carnegie Mellon University Department of Statistics.

HAROLD JEFFFREYS (1891–1989)
Geophysicist, astronomer (*The Earth: Its Origin, History, and Physical Constitution*), philosopher of science, and statistician (*Theory of Probability*, especially the Third Edition of 1961). The 1961 edition showed, using the principle of invariance, how to generate prior probabilities that reflected "knowing little"). Jeffreys believed that probabilities should be common to all (objectivist Bayesian approach). He advocated placing probabilities on hypotheses and proposed a very general theory of hypothesis testing that provided for probabilities that could be placed on hypotheses.

DENNIS V. LINDLEY (1923–)
Dennis Lindley (*Introduction to Probability and Statistics From a Bayesian Viewpoint*, 1965, 2 volumes), has been a stroong leader of the Bayesian movement. Moreover, several of his students have also become leaders of the movement as well as developers of important results in the field. He started as an objectivist Bayesian and became a subjectivist in his later years. Along with A.F.M. Smith he became a strong advocate of hierarchical Bayesian modeling

LEONARD J. (Jimmie) SAVAGE (1917–1971)
Jimmie Savage began his career as a mathematician, but contributed widely to mathematical statistics, probability theory, and economics. He developed an axiomatic basis for Bayesian probability theory and decision-making (*The Foundations of Statistics*, 1954). He held a subjectivist Bayesian view of probability. Along with W. Allen Wallis he founded the Department of Statistics at the University of Chicago. Along with Edwin Hewitt he demonstrated the widely applicable De Finetti theorem involving exchangeability under a broad set of conditions. They demonstrated its uniqueness property.

Subjective and Objective
Bayesian Statistics

Second Edition

PART I

Foundations and Principles

Subjective and Objective Bayesian Statistics: Principles, Models, and Applications
S. james Press

CHAPTER 1

Background

1.1 RATIONALE FOR BAYESIAN INFERENCE AND PRELIMINARY VIEWS OF BAYES' THEOREM

In 1763 an important scientific paper was published in England, authored by a Reformist Presbyterian minister by the name of Thomas Bayes (Bayes, 1763). The implications of the paper indicated how to make statistical inferences that build upon earlier understanding of a phenomenon, and how *formally* to combine that earlier understanding with currently measured data in a way that updates the degree of belief (subjective probability) of the experimenter. The earlier understanding and experience is called the "prior belief" (belief or understanding held prior to observing the current set of data, available either from an experiment or from other sources), and the new belief that results from updating the prior belief is called the "posterior belief" (the belief held after having observed the current data, and having examined those data in light of how well they conform with preconceived notions). This inferential updating process is eponymously called Bayesian inference. The inferential process suggested by Bayes shows us that to find our subjective probability for some event, proposition, or unknown quantity, we need to multiply our prior beliefs about the event by an appropriate summary of the observational data. Thus, Bayesian inference suggests that all formal scientific inference inherently involves two parts, a part that depends upon subjective belief and scientific understanding that the scientist has prior to carrying out an experiment, and a part that depends upon observational data the scientist obtains from the experiment. We present Bayes' theorem compactly here in order to provide an early insight into the development of the book in later chapters.

Briefly, in its most simple form, the form for events (categorical or discrete data), Bayes' theorem or formula asserts that if $P\{A\}$ denotes the probability of an event A, and $P\{B|A\}$ denotes the probability of an event B conditional on knowing A, then:

$$P\{B|A\} = \frac{P\{A|B\}P\{B\}}{P\{A|B\}P\{B\} + P\{A|\bar{B}\}P\{\bar{B}\}},$$

where $\bar{B}$ denotes the complementary event to event B. This simple statement of conditional probability is the basis for all Bayesian analysis. $P(B)$ denotes the prior belief about $B, P\{B|A\}$ denotes the posterior belief about B (once we know A), and $P\{A|B\}$ denotes the model, that is, the process that generates the event A based upon knowing B.

As an example, suppose you take a laboratory test for diabetes. Let A denote the outcome of the test; it is a *positive* outcome if the test finds that you have the tell-tale indicators of diabetes, and it is a *negative* outcome if you do not. But do you really have the disease? Sometimes, although you do not actually have diabetes, the test result is positive because of imperfect characteristics of the laboratory test. Similarly, sometimes when you take the test there is a negative outcome when in fact you do have the disease. Such results are called false positives and false negatives, respectively. Let B denote the event that you actually do have diabetes. You would like to know the chances that you have diabetes in light of your positive test outcome, $P\{B|A\}$. You can check with the laboratory to determine the sensitivity of the test. Suppose you find that when the test is negative, the error rate is 1 percent (false negative error rate), and when the test is positive, its accuracy is 3 percent (the false positive error rate). In terms of the Bayes' formula,

$$\begin{aligned} P\{A = +test|B = diabetes\} &= 1 - P\{\bar{A} = -test|B = diabetes\} \\ &= 1 - 0.01 = 0.099, \end{aligned}$$

and

$$P\{+test|\bar{B} = no\, diabetes\} = \text{probability of a false positive} = 0.03.$$

Bayes' formula then gives:

$$P\{B|+\} = \frac{P\{+|B\}P\{B\}}{P\{+|B\}P\{B\} + P\{+|\bar{B}\}P\{\bar{B}\}}$$

$$P\{diabetes|+\} = \frac{(0.99)P\{B\}}{(0.99)P\{B\} + (0.03)P\{\bar{B}\}}.$$

It only remains to determine $P\{B\}$, the chances of someone having diabetes. Suppose there is no indication that diabetes runs in your family, so the chance of you having diabetes is that of a randomly selected person in the population about your age, say about one chance in one million, that is, $P\{B\} = 10^{-6}$. Substituting in the above formula gives:

$$P\{you\ have\ diabetes|positive\ test\ result\} = 0.003 = 0.3\%.$$

If we are concerned with making inferences about an unknown quantity θ, which is continuous, Bayes' theorem takes the form appropriate for continuous θ:

$$h(\theta|x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n \mid \theta)g(\theta)}{\int f(x_1, \ldots, x_n \mid \theta)g(\theta)d\theta},$$

where $h(\cdot)$ denotes the probability density of the unknown θ subsequent to observing data $(x_1, \ldots, x_n)$ that bear on θ, f denotes the likelihood function of the data, and g denotes the probability density of θ prior to observing any data. The integration is taken over the support of θ. This form of the theorem is still just a statement of conditional probability, as we will see in Chapter 4.

A large international school of scientists (some of whom even preceded Bayes) supported, expanded, and developed Bayesian thinking about science. These include such famous scientists as James Bernoulli, writing in 1713, Pierre Simon de Laplace, writing in 1774, and many nineteenth- and twentieth-century scientists. Today, scientists schooled in the Bayesian approach to scientific inference have been changing the way statistical methodology itself has been developing. Many believe that a paradigm shift has been taking place in the way scientific inference is carried out, away from what is sometimes referred to as classical, or frequentist, statistical inference. Many scientists now recognize the advantages of bringing prior beliefs into the inferential process in a formal way from the start, instead of striving, and almost inevitably failing, to achieve total objectivity, and bringing the prior information into the problem anyway, in surreptitious, or even unconscious ways. Subjectivity may enter the scientific process surreptitiously in the form of seemingly arbitrarily imposed constraints in the introduction of initial and boundary conditions in the arbitrary levels of what should be called a significant result (selecting the "level of significance"), and in the de-emphasizing of certain outlying data points that represent suspicious observations.

Scientists will see that Bayes' theorem gives the degree of a person's belief (that person's subjective probability) about some unknown entity once something about it has been observed (i.e., posterior to collecting data about that entity), and shows that this subjective probability is proportional to the product of two types of information. The first type of information characterizes the data that are observed; this is usually thought of as the objective portion of posterior belief, since it involves the collection of data, and data are generally thought to be objectively determined. (We recognize that we do not really mean that data are objective unless we assume that there were no subjective influences surrounding the data collected.) This so-called *objective information* is summarized in the likelihood function. But the likelihood function is of course almost invariably based upon data that has been influenced by the subjectivity of the observer. Moreover, in small or often in even moderate size samples its structural form is not very well determined. So the likelihood function will almost invariably contain substantial subjective influences and uncertainty.

The second type of information used in Bayesian analysis is the person's degree of belief, the subjective probability about the unknown entity, held prior to observing anything related to it. This belief may be based, at least in part, on things that were observed or learned about this unknown quantity prior to this most recent measurement. Using Bayes' theorem, scientific belief about some phenomenon is formally updated by new measurements, the idea being that we learn about something by modifying what we already believe about it (our prior belief) to obtain a posterior belief after new observations are taken.

While it is well known that for a wide variety of reasons there are always some subjective influences in the research of scientists, and always have been, it is less well known that strong major subjective influences have actually been present in some of the work of the most famous scientists in history (see, for example, Press and Tanur, 2001). The personal beliefs and opinions of these scientists have often very strongly influenced the data they collected and the conclusions they drew from those data. While the phenomena these scientists were investigating were generally truly objective phenomena, external to the human mind, nevertheless, the data collected about these phenomena, and the decisions made relative to these phenomena were often driven by substantial subjectivity. Bayesian analysis, had it been available to these scientists, and had it been used, might have permitted these scientists to distinguish between models whose coexistence has caused controversy about their results even hundreds of years later.

Further, several scientists examining the same set of data from an experiment often develop different interpretations. This phenomenon is not unusual in science. When several scientists interpret the same set of data they rarely have *exactly* the same interpretations. Almost invariably, their own prior beliefs about the underlying phenomenon enter their thinking, as do their individual understanding of how meaningful each data point is. Their conclusions regarding the extent to which the data support the hypothesis will generally reflect a mixture of their prior degree of belief about the hypothesis they are studying, and the observed data.

Thus, we see that whether formal Bayesian inference is actually used in dealing with the data in an experiment, or whether other, nonBayesian methods are used, subjective prior belief is used in one way or another by all good scientists in a natural, and sometimes quite informal, way. Science cannot, and should not, be totally objective, but should and does involve a mixture of both subjective and objective procedures, with the one type of procedure feeding back on the other. As the data show the need for modification of the hypothesis, a new hypothesis is entertained, a new experiment is designed, new data are taken, and what was posterior belief in the earlier experiment becomes the prior belief in the new experiment, because the result of the last experiment is now the best understanding the scientist has of what result to expect in a new experiment. To study the future, scientists must learn from the past, and it is important—indeed inevitable—that the learning process be partly subjective.

During the twentieth century, since the development of methods of Bayesian statistical inference, there have been many exciting new scientific discoveries and developments. Some have been simply of the qualitative type where certain phenomena have been discovered that were not previously known (such as the discovery of the existence of the radiation belts that surround the Earth, the discovery of superconductivity, or the discovery of the double helical structure of DNA), and others have been quantitative, establishing relationships not previously established (such as the discoveries of the dose/effect relationships of certain pharmaceutical drugs, vaccines, and antibiotics that would minimize the chances of contracting various infectious diseases, or maximize the chance of a cure).

Considerable scientific advance is based upon finding important phenomena that are sometimes so shrouded in noise that it is extremely difficult to distinguish the phenomenon of interest from other factors and variables. In such cases, prior information about the process, often based upon previous theory, but sometimes on intuition or even wild guesses, can often be profitably brought to bear to improve the chances of detecting the phenomenon in question. A considerable amount of Bayesian statistical inference procedures that formally admit such prior information in the scientific process of data analysis have had to await the advent of modern computer methods of analysis, an advent that did not really occur until the last couple of decades of the twentieth century. However, since the arrival of real-time interactive computers, computational Bayesian methods such as Markov Chain Monte Carlo (MCMC, see Chapter 6) have been very usefully applied to problems in imaging and other problems in physics and engineering (see the series of books edited by different authors, every year since 1980, Maximum Entropy and Bayesian Methods published by Kluwer), problems of meta-analysis to synthesize results in a field—in biology, medicine, economics, physics, sociology, education, and others—and in a variety of scientific fields (see, for example, Appendix 5).

Subjectivity in science implies that we generally arrive at universal scientific truths by a combination of subjective and objective means. In other words, the methodology we use to discover scientific truths benefits greatly from bringing informed scientific judgment to bear on the hypotheses we formulate, and on the inferences we make from data we collect from experiments designed to test these hypotheses. Informed scientific judgment should not be shunned as a nonobjective, and therefore a poor methodological approach; collateral information about the underlying process should be actively sought so that it can be used to improve understanding of the process being studied. Combining informed knowledge with experimental data will generally improve the accuracy of predictions made about future observations.

Subjectivity is an inherent and required part of statistical inference and the scientific method. It is a *sine qua non* in the process of creating new understanding of nature. It must play a fundamental role in how science is carried out.

However, excessive, informal, untested subjectivity in science is also responsible for some basic errors, misrepresentations, overrepresentations, or scientific beliefs that were later shown to be false, that have occurred in science (see, for example, Press and Tanur, 2001). This author's views of subjectivity in science coincide closely with those of Wolpert (1992, p. 18) who wrote:

> ...the idea of scientific objectivity has only limited value, for the way in which scientific ideas are generated can be highly subjective, and scientists will defend their views vigorously... It is, however, an illusion to think that scientists are unemotional in their attachment to their scientific views—they may fail to give them up even in the face

> of evidence against them . . . scientific theories involve a continual interplay with other scientists and previously acquired knowledge . . . and an explanation which the other scientists would accept.

To illustrate the notion that subjectivity underlies experimental science, in Section 1.2 we use a very simple example involving whether or not a desired effect is observed in an experiment to show that merely observing scientific data and forming a likelihood function can involve considerable subjectivity.

1.2 EXAMPLE: OBSERVING A DESIRED EXPERIMENTAL EFFECT

Let us suppose that 100 observations are collected from an experiment replicated 100 times; there is one observation from each replication. These data are sent to five scientists located in five different parts of the world. All five scientists examine the same data set, that is, the same 100 data points. (Note that for the purposes of this example, the subjectivity involved in deciding what data to collect and in making the observations themselves is eliminated by sending the same "objective" data to all five scientists.) Should we expect all five of the scientists to draw the same conclusions from these data?

The answer to this question is a very definite "no". But how can it be that different observers will probably draw different conclusions from precisely the same data? As has been said above, inferences from the data will be a mixture of both subjective judgment (theorizing) and objective observation (empirical verification). Thus, even though the scientists are all looking at the same observational data, they will come to those same data with differing beliefs about what to expect. Consequently, some scientists will tend to weight certain data points more heavily than others, while different scientists are likely to weight experimental errors of measurement differently from one another. Moreover, if scientists decide to carry out formal checks and statistical tests about whether the phenomenon of interest in the experiment was actually demonstrated (to ask how strongly the claimed experimental result was supported by the data), such tests are likely to have different results for different scientists, because different scientists will bring different assumptions to the choice of statistical test. More broadly, scientists often differ on the mathematical and statistical models they choose to analyse a particular data set, and different models usually generate different conclusions. Different assumptions about these models will very often yield different implications for the same data.

These ideas that scientists can differ about the facts are perhaps startling. Let us return to our 100 observations and five scientists to give a very simple and elementary example, with the assurance that analogous arguments will hold generally for more realistic and more complicated situations.

Let us assume that the purpose of the experiment is to determine the probability that a certain genetic effect will take place in the next generation of a given type of simple organism. The question at issue is whether the effect occurs randomly or is subject to certain genetic laws. If the experiment is carried out many times, inde-