Contents

Preface

The roots of this book span more than 40 years of using, teaching, and contributing to the development of Bayesian methods. Forty years ago, several excellent (but limited) books about Bayesian statistics were already available, and subsequently, additional books have been published. Consequently, I have avoided until now contributing another book on this topic, because I did not want to repackage information that was already in other books. So what happened that led me to write the current book? There are two related answers to this question. First, an accumulation of new material provides for a powerful, relatively simple way to do Bayesian statistical analyses for cases that heretofore did not have a full Bayesian solution. These new methods are not designed for some obscure statistical problems; instead they deal with very basic procedures that occur regularly in scientific experiments and in engineering applications. Second, there is a need for a book that informs experimental scientists that fundamental problems arise with orthodox statistical analyses that distort the scientific process. There is a mismatch between the needs of experimental scientists in making inferences from data and the properties of standard tools of orthodox or classical statistics. This mismatch has led to the problem that most researchers generally do not fully understand what the classical methods are actually doing, and importantly, what they are not doing. I believe that Bayesian statistics is a better match to the needs of experimental scientists. Moreover, in my opinion, Bayesian statistics is a more rigorous inferential system, and it can be made relatively easy to understand and easy to implement with freely available software. Bayesian statistical tools will enable experimental scientists to make valid inferences without evoking arbitrary, one-off rules, and these tools are a natural fit to the process of scientific reasoning.

Although there are Bayesian methods for many of the statistical procedures that were developed originally from an orthodox relative-frequency perspective, one area has been a tough problem to reformulate in the Bayesian framework. That area is distribution-free statistics. This limitation has been particularly unfortunate, because these methods were among my favorite aspects of classical statistics. These statistics are based on a minimalistic approach that avoids making the distributional assumptions used in conventional (parametric) statistics. The classic term for this type of statistics was *nonparametric statistics*. Lindley (1972) observed that nonparametric or distribution-free statistics was a topic on which the Bayesian approach was embarrassingly silent. That observation was made a long time ago, but until recently, there still were no Bayesian analyses for simple rank-based nonparametric methods. What developed instead was a branch of Bayesian theory that has come to be called *Bayesian nonparametric models* (Ferguson, 1973; Ghosh & Ramamoorthi, 2002; Müller & Quintana, 2004; Müller et al., 2015; Walker et al., 1999). But these models, which have a misleading label in my opinion, are not minimalistic methods that are free of distributional assumptions. Instead they are complex explorations of parameter spaces that have an infinite number of dimensions. This approach also makes numerous assumptions. Most importantly, these theoretical explorations are not easy to understand and communicate to experimental scientists. If the statistical analysis is too mathematical and too complex, then experimental scientists will be wary about using it. As an experimental scientist, I share that concern about using methods that skeptical professionals in my academic discipline would have trouble understanding in principle. Instead I was seeking a simple Bayesian analysis for the existing classical distribution-free methods, such as the Wilcoxon signed-rank statistic. Heretofore I tried to make headway on developing a Bayesian counterpart for those methods, but I was unable to solve the problem. Thus the gap in the Bayesian toolkit that Lindley identified in 1972 was still applicable in the twenty-first century except for some recent developments. Fortunately, a full Bayesian analysis can now be done for many of the frequency-theory distribution-free statistics. Consequently there is good reason for a new book on Bayesian statistics that provides a general introduction to the topic with a particular focus on distribution-free, nonparametric methods. But before continuing with a discussion of the recent developments in Bayesian methods, let me first clarify what the classic field of nonparametric statistics is and why this approach is so valuable.

For a parametric statistical analysis of interval- or ratio-scale data, assumptions are made about the properties of the population distributions from which the data were sampled. The Gaussian (or the normal distribution) is a widely assumed model for the configuration of the values in the population. The normal distribution has a mathematical structure that describes precisely the shape of the population variate. The central-limit theorem provides a rationale for why the distribution of the sample mean should (in the limit) be distributed as a normal distribution. Consequently the parametric Gaussian model for the sample mean is reasonable for a single experimental condition. But most research requires multiple conditions. With more than one condition, the statistical inference is also well described by the parametric model if the population variates are identical except for a difference in their means. For example, even if the population variates are distributed as a non-Gaussian distribution, such as the Weibull distribution, the statistical inference about the difference in the mean between two Weibull-distributed variates is well described by the parametric model, provided that the shape and scale properties of the Weibull distributions are equal. Yet if the characteristics of the two population Weibull-distributed variates differ in terms of their shape and scale parameters, even if the population means are equal, then huge errors result from using the parametric model. For example, I have seen errors of 59 percent or more in the predicted quantile points. This example is not far fetched, because real processes in nature are complex and often contain mixtures of different probabilistic processes. Although statistical tests exist for examining whether there is a violation of the assumed distribution, the detection of a mixture is challenging (Chechile, 2013; Congdon, 2001). Moreover, a statistical test that fails to reject a hypothesized distribution is not conclusive evidence in itself, especially for studies with a small sample size. Furthermore, most statistical analyses make an assumption about the equality of the error variance across separate conditions. This assumption is almost certainly not correct, and it results in underestimating the magnitude of the uncertainty of the estimates for the population parameters (Box & Tiao,1973; Chechile, 1979).

In his classic book on nonparametric statistics, Siegel (1956) observed that data in the behavioral sciences frequently are not consistent with parametric assumptions. I should add that other sciences also have similar concerns about the parametric statistical assumptions—even engineering and the physical sciences. For example, in chapter 7, an engineering application is described where the parametric analysis requires stipulating the population distribution of strength values for a material as a step towards estimating the reliability of the material when used under stress. There are four reasonable distributions for the assumed model of the strength of the material in the reliability-engineering literature (Chaudhary, Kumar, & Tomer, 2017; Kundu & Raqab, 2009; Rao, 2014; Surles & Padgett, 2001). But if the same data are examined with these assumed distributions, then there are four very different estimates for the reliability! Consequently the distributional assumption matters. Yet in chapter 7 a Bayesian, distribution-free analysis is used for that same data to arrive at a direct answer about the product reliability. Thus there is a decided advantage for a statistical method that does not need to assume any distribution in the first place (i.e., a distribution-free, nonparametric solution). Such a distribution-free approach has an additional advantage of being robust with respect to the contaminating effects of outliers and mixtures. In chapter 7 cases are described where aberrant values and unusual distributions that differ from the standard probability models have little effect on the nonparametric analysis, but these data do dramatically distort the statistical inference for parametric analyses. Thus distribution-free statistics is a potent tool for research scientists. Also there are many cases where the data are categorical rather than interval. For these studies, the parametric model is not even possible.

Nonparametric statistics developed entirely in the context of orthodox sampling theory (or what is called the *frequentist approach*). Bayesian statisticians reject the frequentist system of statistical inference for many reasons that will be explained more completely in chapters 2 and 4, but there is no objection to the idea of having a Bayesian version for the sampling-theory distribution-free methods. Yet this extension of Bayesian methodology has been difficult to achieve for two important reasons. First, in the Bayesian approach, it is necessary to identify an unknown population parameter that underlies the probabilistic variation of the sample value for the distribution-free statistic. Second, the likelihood function for obtaining the observed sample value of the statistic has to be found for any possible value of the unknown parameter. Not knowing the likelihood function makes it impossible to use popular Markov chain Monte Carlo methods. Fortunately, these problems were recently overcome, so it is possible to have a Bayesian analysis for some of the highly successful distribution-free statistics from sampling theory (Chechile, 2018a, 2019). Particularly appealing aspects of the Bayesian solution are that it is very simple to implement and the results are easy to interpret. In most cases, the results can be obtained after a few calculations on an ordinary calculator plus a few lines of R code on a computer. Thus these new methods represent an exciting new step in the development of Bayesian methodology.

The book is designed as the basis for an introductory course on Bayesian statistics with a focus on distribution-free methods. Even though there are novel ideas in this book that would be of interest to statisticians, the book is constructed for a more general audience that has had some previous statistical training. Specifically, it is assumed that the reader has already taken an introductory statistics course but now wants to learn about Bayesian statistics. The reader should also have a rudimentary understanding of mathematical sets. Furthermore, it would be helpful if the reader has an elementary understanding (however limited) of calculus, but he or she will not need to perform calculus-based calculations. The level of the book is thus designed either for the advanced undergraduate student or for the beginning graduate student in any academic discipline that is engaged in empirical research. Someone with no background in statistics whatsoever should read first an introductory book, such the text by Berry (1996).

Apart from the issue of the reader's educational background, the book is best suited for the subset of readers who are primarily interested in empirical studies where data have been collected from an experiment. Many academic disciplines fit this model of research. Yet other disciplines almost exclusively use multivariate methods, such as multiple regression. This book is not an ideal resource for this group, because the topic of multiple regression is not covered; although chapter 8 does deal with the analysis of rank-based correlation between variables. The primary focus for the book is instead on statistical questions arising from the analysis of data collected as part of an experimental study. This approach is routine in the behavioral and biological sciences, but it is not limited to those fields. Investigators in many other areas also might employ experiments with two or more conditions or groups that are compared to address research questions. For these researchers, the material in the book provides them with many (perhaps most) of the methods that they will need for a statistical analysis from a Bayesian distribution-free framework.

Having taught advanced statistics to both upper-level undergraduate students and beginning graduate students, I found that there are many preliminary ideas about probability and statistics that need to be covered before tackling the more advanced issues of statistical inference. In some cases, the students are simply rusty and need to relearn some information. In other cases, the prior learning either has gaps or demonstrates mistaken concepts that need to be cleaned up before going on to advanced statistics. Consequently, this book begins by examining some essential concepts from probability theory, and it gradually eases into topics more directly related to performing a Bayesian analysis.

All the chapters include exercises. These exercises, in my opinion, are essential for helping the reader achieve a good understanding of the information. Statistics is learned more completely by doing problems and exercises.

The book is organized into two parts. Part I has four chapters that cover elementary probability theory, the binomial model, the multinomial model, and methods for comparing different experimental conditions or groups. The focus in part I is on distribution-free methods that apply for categorical data. Part II of the book also has four chapters. These chapters focus on distribution-free statistics that are based on having ranked data. Three of these chapters are concerned with data from experimental studies. The final chapter deals with rank-based correlational methods, but even in that chapter, applications from experimental science are highlighted.

Considerable attention in this book is directed toward understanding the theoretical and philosophical foundations of statistics, as well as the linkage of statistics to practices in experimental science. Statistics is all too often poorly understood and poorly taught. As a result, many researchers have numerous misconceptions. The beauty of Bayesian statistics is it is an internally coherent system of scientific inference that can be proved from probability theory. In contradistinction to this tight inferential system, frequentist statistical practices often violate the theoretical premises on which the method was originally based. Too often the procedures of frequentist statistics are *one-off* decisions or rules that are not rigorously grounded in probability theory. Moreover, these rules do not provide the information that scientists are seeking. Thus, frequentist practitioners conducting scientific research routinely make incorrect statements that originated from their misconceptions of statistics. To make matters worse, frequentist procedures have influenced some experimental researchers to focus on evaluating trivial scientific issues. Consequently several chapters in this book offer an extensive critique of frequentist statistical methods as well as demonstrations of why those methods can be improved by employing the Bayesian approach. If students and researchers better understand the flaws of frequentist methods and the advantages of the Bayesian distribution-free approach, then the state of experimental science will be substantially improved. To better communicate these methodological ideas, the writing style in the book deliberately minimizes mathematical formalism (when possible) and relies instead on verbal discussions and examples. Footnotes and chapter appendices provide additional detail for the more advanced reader.

Finally I would like to acknowledge some of the people who provided help in a variety of ways toward the completion of this writing project. First, a special note of acknowledgment is due to my former professor and mentor, Donald L. Meyer (1932–2013), who introduced me to the beauty of Bayesian statistics. Without Don's influence, it is unlikely that I would have written this book many years later. I also wish to thank many of the students in my advanced statistics courses as well as my graduate-student collaborators who had an indirect hand in my development toward understanding the needs and the research problems of a wide variety of experimental scientists. I also want to thank Daniel Barch for his help on R programming and for his part in our collaborative work that is highlighted in chapter 8. I also wish to thank the four anonymous reviewers of this book for their many insightful comments and criticisms that helped improve the book. Finally, I especially want to thank my wife, Jeannette, who patiently read and thoughtfully commented on all the drafts of the book.

Richard A. Chechile