Extracted from:

Bayesian Logical Data Analysis for the Physical Sciences

by Phil Gregory

Preface

The goal of science is to unlock nature's secrets. This involves the identification and understanding of nature's observable structures or patterns. Our understanding comes through the development of theoretical models which are capable of explaining the existing observations as well as making testable predictions. The focus of this book is on what happens at the interface between the predictions of scientific models and the data from the latest experiments. The data are always limited in accuracy and incomplete (we always want more), so we are unable to employ deductive reasoning to prove or disprove the theory. How do we proceed to extend our theoretical framework of understanding in the face of this? Fortunately, a variety of sophisticated mathematical and computational approaches have been developed to help us through this interface, which go under the general heading of statistical inference. Statistical inference provides a means for assessing the plausibility of one or more competing models, and estimating the model parameters and their uncertainties. These topics are commonly referred to as `data analysis' in the jargon of most physicists.

We are currently in the throes of a major paradigm shift in our understanding of statistical inference based on a powerful generalization of Aristotelian logic. For historical reasons, it is referred to as Bayesian Probability Theory or Bayesian statistic. To get a taste of how significant this development is, consider the following: probabilities are commonly quantified by a real number between 0 and 1. The end-points, corresponding to absolutely false and absolutely true, are simply the extreme limits of this infinity of real numbers. Deductive logic, which is based on axiomatic knowledge, corresponds to these two extremes of 0 and 1. Ask any mathematician or physicist how important deductive logic is to their discipline! Now try to imagine what you might achieve with a theory of extended logic that encompassed the whole range from 0 to 1. This is exactly what is needed in science and real life where we never know anything is absolutely true or false. Of course, the field of probability has been around for years, but what is new is the appreciation that the rules of probability are not merely rules for manipulating random variables. They are now recognized as uniquely valid principles of logic, for conducting inference about any proposition or hypothesis of interest. It is thus a mathematical theory that encompasses both inductive and deductive logic. Ordinary deductive logic is just a special case in the idealized limit of complete information.

The reader should be warned that most books on Bayesian statistics do not make the connection between probability theory and logic. This connection, which is captured in the book[1] by physicist E. T. Jaynes, *Probability Theory---the Logic of Science*, is particularly appealing because of the unifying principles it provides for scientific reasoning.

What are the important consequences of this development? We are only beginning to see the tip of the iceberg. Already we have seen that for data with a high signal-to-noise ratio, a Bayesian analysis can frequently yield many orders of magnitude improvement in model parameter estimation, through the incorporation of relevant prior information about the signal model. For several dramatic demonstrations of this point, have a look at the first four sections of Chapter 13. It also provides a more powerful way of assessing competing theories at the forefront of science by quantifying Occam's razor, and sheds a new light on systematic errors (e.g. Section 3.11). For some problems, a Bayesian analysis may simply lead to a familiar statistic. Even in this situation it often provides a powerful new insight concerning the interpretation of the statistic. But most importantly, Bayesian analysis provides an elegantly simple and rational approach for answering any scientific question for a given state of information.

This textbook is based on a measurement theory course which is aimed at providing first year graduate students in the physical sciences with the tools to help them design, simulate and analyze experimental data. The material is presented at a mathematical level that should make it accessible to physical science undergraduates in their final two years. Each chapter begins with an overview and most end with a summary. The book contains a large number of problems, worked examples and 132 illustrations.

The Bayesian paradigm is becoming very visible at international meetings of physicists and astronomers (e.g. *Statistical Challenges in Modern Astronomy III*, edited by E. D. Feigelson and G. J. Babu 2002). However, the majority of scientists are still not at home with the topic and much of the current scientific literature still employs the conventional "frequentist" statistical paradigm. This book is an attempt to help the new student to make the transition while at the same time exposing them in Chapters 5,6 and 7 to some of the essential ideas of the frequentist statistical paradigm that will allow them to comprehend much of the current and earlier literature and interface with his or her research supervisor. This also provides an opportunity to compare and

---

[1] Early versions of this much celebrated work by Jaynes have been in circulation since at least 1988. The book was finally submitted for publication in 2002, four years after his death, through the efforts of his former student G. L. Bretthorst. The book is published by Cambridge University Press (Jaynes 2003, Edited by G.L. Bretthorst).

contrast the two different approaches to statistical inference. No previous background in statistics is required; in fact, Chapter 6 is entitled “What is a Statistic?” For the reader seeking an abridged version of Bayesian inference, Chapter 3 provides a stand-alone introduction on the ``How-To of Bayesian Inference."

The book begins with a look at the role of statistical inference in the scientific method and the fundamental ideas behind Bayesian Probability Theory (BPT). We next consider how to encode a given state of information into the form of a probability distribution, for use as a prior or likelihood function in Bayes' theorem. We demonstrate why the Gaussian distribution arises in nature so frequently from a study of the Central Limit Theorem and gain powerful new insight into the role of the Gaussian distribution in data analysis from the Maximum Entropy Principle. We also learn how a quantified Occam's razor is automatically incorporated into any Bayesian model comparison and come to understand it at a very fundamental level.

Starting from Bayes' theorem, we learn how to obtain unique and optimal solutions to any well-posed inference problem. With this as a foundation, many common analysis techniques such as linear and nonlinear model fitting are developed and their limitations appreciated.

The Bayesian solution to a problem is often very simple in principle, however, the calculations require integrals over the model parameter space which can be very time consuming if there are a large number of parameters. Fortunately, the last decade has seen remarkable developments in practical algorithms for performing Bayesian calculations. Chapter 12 provides an introduction to the very powerful Markov Chain Monte Carlo (MCMC) algorithms, and demonstrates an application of a new automated MCMC algorithm to the detection of extrasolar planets.

Although the primary emphasis is on the role of probability theory in inference, there is also focus on an understanding of how to simulate the measurement process. This includes learning how to generate pseudo-random numbers with an arbitrary distribution (in Chapter 5). Any linear measurement process can be modeled as a convolution of nature's signal with the measurement point spread function, a process most easily dealt with using the convolution theorem of Fourier analysis. Because of the importance of this material, I have included Appendix B on the Discrete Fourier Transform (DFT), the Fast Fourier Transform (FFT), Convolution and Weiner filtering. We consider the limitations of the DFT and learn about the need to zero pad in convolution to avoid aliasing. From the Nyquist Sampling Theorem we learn how to minimally sample the signal without losing information and what prefiltering of the signal is required to prevent aliasing.

In Chapter 13, we apply probability theory to spectral analysis problems and gain a new insight into the role of the DFT, and explore a Bayesian revolution in spectral analysis. We also learn that with non-uniform data sampling, the effective bandwidth (the largest spectral window free of aliases) can be made much wider than for uniform sampling. The final chapter is devoted to Bayesian inference when our prior information leads us to model the probability of the data with a Poisson distribution.

Software support

The material in this book is designed to empower the reader in his or her search to unlock nature's secrets. To do this efficiently, one needs both an understanding of the principles of extended logic, and an efficient computing environment for visualizing and mathematically manipulating the data. All of the course assignments involve the use of a computer. An increasing number of my students are exploiting the power of integrated platforms for programming, symbolic mathematical computations, and visualizing tools. Since the majority of my students opted to use *Mathematica* for their assignments, I adopted *Mathematica* as a default computing environment for the course. There are a number of examples in this book employing *Mathematica* commands, although the book has been designed to be complete without reference to these *Mathematica* examples. In addition, I have developed a *Mathematica* tutorial to support this book, specifically intended to help students and professional scientists with no previous experience with *Mathematica* to efficiently exploit it for data analysis problems. This tutorial also contains many worked examples.

In any scientific endeavor, a great deal of effort is expended in graphically displaying the results for presentation and publication. To simplify this aspect of the problem, the *Mathematica* tutorial provides a large range of easy to use templates for publication-quality plotting.

It used to be the case that interpretative languages were not as useful as compiled languages such as C and Fortran for numerically intensive computations. The last few years have seen dramatic improvements in the speed of *Mathematica.* Wolfram Research now claims [2] that for most of Mathematica's numerical analysis functionality (e.g., data analysis, matrix operations, numerical differential equation solvers and graphics) Mathematica 5 operates on a par [3] with Fortran or MATLAB code. In the authors experience, the time required to develop and test programs with *Mathematica* is approximately 20 times shorter than the time required to write and debug the same program in Fortran or C, so the efficiency gain is truly remarkable.

Acknowledgements

Most of the Bayesian material presented in this book I have learned from the works of Ed Jaynes, Larry Bretthorst, Tom Loredo, Steve Gull, John Skilling, Myron Tribus, Devinder Sivia, Jim Berger and many others from the international community devoted to the study of Bayesian

[2] http://www.wolfram.com/products/mathematica/newin5/performance/numericallinear.html

[3] Look up *Mathematica* gigaNumerics on the web.

inference. On a personal note, I encountered Bayesian inference one day in 1989 when I found a monograph lying on the floor of the men's washroom entitled “Bayesian Spectrum Analysis and Parameter Estimation” by Larry Bretthorst. I was so enthralled with the book that I didn't even try to find out whose it was for several weeks. Larry's book led me to the work of his Ph.D. supervisor, Edwin T. Jaynes. I became hooked on this simple, elegant and powerful approach to scientific inference. For me, it was a breath of fresh air providing a logical framework for tackling any statistical inference question in an optimal way in contrast to the recipe or cookbook approach of conventional statistical analysis.

I would also like to acknowledge the proof reading and suggestions made by many students who were exposed to early versions of this manuscript, in particular, Iva Cheung for her very careful proof reading of the final draft. Finally, I am really grateful to my partner, Jackie, and our children, Rene, Neil, Erin, Melanie, Ted, and Laura, for their encouragement over the many years it took to complete this book.