Bayesian book ideas

How to use this book

- Bayesian analysis is the most intuitive a theory of statistics, with an applicability that is wider than current frequentist methods. It is however easy to get lost in the relatively advanced maths on which the subject is based. The maths poses a sufficient challenge to those wanting to learn Bayesian methodology, that many students give up after attempting to get to grips with the theory.

As such, It is hoped that throughout this book that the emphasis will be on intuitive explanations of theory, supported by practical examples which provide the reader with context.

- Associated with each marker within the theory, there is a corresponding video(s). These will not merely be reiterations of the text of thus book, but are intended to complement.

- In general for each new piece of theory, this book will provide an example of its use. The videos will also provide a different set of examples, aimed at building further intuition for the theory.

- This book is practical. We will make use of the open source software BUGS to introduce the student to the practicalities of undertaking Bayesian data analysis.

- Problem sets will mainly be focussed on analysis of real data. The data is introduced in the text, and by video. Answers to the problems are also available; this is a book intended for the student to build up practical experience and test their knowledge.

Intro chapter

- why do Bayesian analysis? Extoll the virtues of this type of statistics. Best predictions: Nate Silver predictions. Easy interpretation of results. Probabilities over confidence levels. Best way of updating predictions in light of new data.

- The subjectivity of objectivity: (perhaps title of first chapter). The idea that subjectivity is everywhere. When we write a linear regression model down, we are making a subjective judgement that this is the type of process likely to generate observed data. (Careful of the CEF argument from Mostly Harmless Econometrics here). Implicit vs explicit declaration of subjective information.

- Why are posterior distributions better than priors? They incorporate both prior and data, to produce a result which has a lower variance than the prior. Page 37 in Gelman

- Jeffrey's prior intuition: BUGS book chap 5 (PDF)- flipping a coin. Why is a uniform prior actually very informative? Uniform prior over theta suggests a distribution for theta-squared (probability of getting two heads) which is skewed towards zero. This is because squaring has the effect of taking a point in (theta,p(theta)) space and pushing it towards the origin in (theta-squared,p(theta-squared)) space. This bunching up of points towards the origin causes the skew. The skew will be very much more to the origin if we were talking about theta-power-k space, where k is greater than 2. Therefore being uniform over choice of theta implies very much non-uniformity over choice of theta-squared. Ie having a uniform prior for the probability of a head implies a prior for the probability of getting two heads which is very skewed towards zero. This makes sense because all values of theta(prob of heads) less than about 0.7 (sqrt 0.5) give the probability of a heads which is less than a half. We would like a prior which made no such presumptions. (The only fixed points in the transformation are zero and 1. Could calculate the distance moved by points as a function of theta - ie maximise theta minus theta-squared - biggest distance moved is by half.) Jeffrey's prior beta(0.5,0.5) has this property. It is peaked at zero and one - positions of theta that correspond to lowest variance. This makes intuitive sense because these points have the lowest variance (zero), so any likelihood which is peaked near there should be given extra weight. Intuitively, since the variance is maximised near the centre we should give less weight to values of the likelihood which are peaked there; a range of values of theta could result in the same likelihood being obtained. However, near the edge, especially if we have lots of throws, there are only a small range of values for theta that could give that likelihood. Therefore we give more weight to these types of likelihood since they are unlikely to come from other values of theta. If number of heads achieved is a half the posterior should be wider than for heads close (in fraction) to 1. In video perhaps mention comparison with WLS. Cf pages 89 and 86 from dogs book - why is Jeffrey's prior better than this? Imagine using uniform prior and getting 9/10 heads - what would the posterior look like? How would the prior and posterior look in theta-power-10 space? Is that sensible? Why having a prior that is peaked near points of curvature is better? Why does this align likelihood with prior? Also look up reference priors.

Blurb

Bayesian analysis is the most intuitive theory of statistics, with an applicability that is wider than currently covered by ‘traditional’ methods. It is however easy to get lost in the relatively advanced maths on which the subject is based. The maths poses such a sufficient challenge to those wanting to learn Bayesian methodology that many students give up after attempting to get to grips with the theory. This is especially common for students of the social sciences; almost everyone has a general idea of the concepts inherent in Bayesian statistics, but very few actually practice it.

Current literature aimed at students generally falls into one of two categories: most often the texts are very mathematical (examples being, “Bayesian Data Analysis” by Gelman, “Bayesian Analysis for the Social Sciences” by Jackman, and Koop’s “Bayesian Econometric Methods”), with limited practical examples, limiting their appeal to students; alternatively the texts are overly basic (examples include, “Bayes’ rule: A Tutorial Introduction to Bayesian Analysis” by Stone), or fail to provide a sufficiently broad introduction to the field (for example, “Doing Bayesian Data Analysis” by Kruschke).

It is the view of the author that a book can be written which places an emphasis on the intuition *not* the maths behind Bayesian theory, with a focus on practical examples, without compromising on the breadth of material covered. With this in mind, the book would follow the following principles:

1. The student should be able to come away from reading a chapter, and feel that they would be able to apply the theory learned to real world data; the central aim of the book is to provide a *practical* guide to modern Bayesian statistical techniques. As such, the book will provide data-based examples using the open source BUGS software.
2. The level of mathematics will be kept as simple as possible. There will of course be occasions when it is necessary to introduce mathematical concepts which are *prima facie* difficult to understand. In these circumstances, a very low level of prior knowledge will be assumed, and an emphasis will be placed on understanding the intuition behind the equations.
3. Whenever new theory is introduced, it will be supplemented with one or more examples.
4. New concepts will be explained through the use of videos. Bayesian theory is full of intuition, even though it is sometimes difficult to convey this through traditional static media. Videos will not merely be a recapitulation of the theory and examples explained in the text. Wherever possible the material in the videos will be unique, and complementary to the text.
5. After reading the book the student should be able to read the majority of modern research papers in the social sciences which make use of Bayesian statistics, and be able to understand the theory contained within them. They should also have an idea as to how to go about replicating the research practically.
6. The chapters will be as self-contained as possible. This principle is guided by the fact that students will often use the text for reference with specific practical goals in mind. Of course, the book will be most effective when read in order, but attempts will be made to limit the reliance of each chapter on those preceding it.
7. Problem sets will be focused on analysis of real data. The data will be introduced in the text, and by video.

The text produced would be an ‘essential textbook’, which when followed will provide a comprehensive self-contained introduction to the subject. Whilst the time taken to cover the material contained within the book would be institution-dependent, the course covered would span over one or two semesters. The book would be usable as a source of information for undergraduates, graduates and professionals in the field. It would also have appeal as a ‘supplementary book’ since it would be full of examples, and due to the self-contained nature of chapters. I would anticipate that the length of the book would be of order 300 pages. It is my belief that whilst more pages provide more room for explaining theory, students are perhaps discouraged by lengthy texts.

Chapter list

- The purpose of this book, and how best to use it.

- The subjectivity of the seemingly objective

- The purpose of statistics: population vs sample

- Classical statistics: an introduction to frequentist perspective

- Enter Bayes. Short biographical introduction

- The Bayesian central dogma: the Bayesian formula - Venn diagram + possibly another example of its intuition

- An introduction to the Bayesian inference process

- Where do Bayesian and frequentist views fit into our definition of statistics

- What is probability? The added flexibility of Bayesian views of probability compared to classical theory

- Explicit vs implicit subjectivity: an answer to the common frequentist complaint

- What are the tangible (non-academic) reasons for using Bayesian theory? It offers the best predictions, allows easiest interpretation of results, testing of models against existing empirical and intuitive

- Next two chapters introduce reader bit wise to the elements of the Bayesian formula

- When a probability isn't a probability: Likelihood

- An introduction to likelihood, and reason d'être

- MLE: a frequentist theory. Maximising a probability, which isn't a probability: likelihood

- Log likelihood

- Examples of MLE estimation

- Estimating the uncertainty of estimates: intuition behind the CRLB

- How to specify likelihood: subjectivity of choice

- The stalwarts of a good Bayesian: Prior, likelihood and posterior distributions

- Why is it important to know your distributions?

- Normal

- Beta

- Gamma

- etc

- The difficult denominator: the probability of data (think of better title)

- How the probability of the data is really p(data given choice of model)

- Why is it difficult?

- The nuts and bolts of Bayesian analysis

- Revisiting the central dogma: the posterior as the first goal of Bayesian analysis

- The posterior as trade-off between the likelihood (data) and prior information