**Bayesian book – proposal**

**Proposed title:** “A student’s guide to Bayesian statistics using BUGS”

**Author name:** Ben Lambert

**Date:** 13th June 2014

**Type of book:** essential textbook/supplementary book aimed at undergraduates.

**Method of use:** Each chapter roughly covers a week’s worth of material. The book (and supporting videos) would be used for students as a ‘hard copy’ reference to any theory covered in lectures. The problem sets will be a mixture of theory questions, and data-led examples; although will be slanted towards the latter.

**Book synopsis:**

**Topic:** A comprehensive guide to applied Bayesian statistics. The book is meant to be accessible to undergraduate social scientists, but will be useful for graduate students.

**Sales pitch:** Provide the first fully-accessible comprehensive guide to Bayesian statistics for undergraduate social scientists. The aim is to produce a book similar to Andy Field’s guide to classical statistics.

**Blurb:** Bayesian analysis is the most intuitive theory of statistics. However, a lot of people are discouraged from using it because they do not understand the relatively advanced maths behind it. This book aims to change that.

This book aims to provide an accessible guide to modern applications of Bayesian statistics, with the amount of mathematics kept to a minimum. Whenever new theory is introduced, it is supplemented by use of videos, and a number of practical examples. R and BUGS are used to take the reader through practical applications of Bayesian theory, and all code is given in the text.

The text is aimed at social scientists at the undergraduate level, although it will likely be useful for graduate students wanting to brush up on their knowledge of Bayesian statistics.

**Book maxims:** The book will be created according to the following principles in order to avoid falling into the traps of the existing literature.   
  
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 instead on understanding the intuition behind the equations. The level of maths required is High School calculus.

3. 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.  
  
4.    Whenever new theory is introduced, it will be supplemented with one or more examples. An emphasis will be placed on making explanations of theory accessible, yet succinct, with preference given to visual guides (ie diagrams and figures).  
  
5. 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.  
  
6. Problem sets will be focused on analysis of real data. The data will be introduced in the text, and by video.

**Length:** The book will be around 350 pages in length.

**Pedagogical features**

**Videos:** The videos will be aimed at providing some intuition behind the (relatively) advanced theory, as well as giving practical examples of applications of the theory. It is my belief that a barrier to application of Bayesian theory is that students’ progress falters when they have to wade through lines and lines of mathematics in a book. It is my experience that whilst lists of equations may appear unapproachable, there is *always* intuition behind the objects themselves. If students are taught this they are more likely to engage with the material, and also to remember it!

I have used videos in the past to teach econometrics and statistics, and have made lecture courses for undergraduate and graduate level study on YouTube. In order to see a video I have made in the past to explain a theoretical concept, (in this case three different estimators), see:

[**https://www.youtube.com/watch?v=1SchyQ77VFg**](https://www.youtube.com/watch?v=1SchyQ77VFg)

Furthermore to support this proposal I have produced two videos which might become part of the accompanying materials for this book. The first video is a description of the syllabus which will be covered, and the second is an explanation of the intuition behind Gibbs sampling.

Syllabus:

Gibbs sampling:

The data for problem sets will be introduced in videos, which accompany the sections at the end of each chapter. Rather than focus on generating a large range of problem questions, the book will aim to produce a smaller number of relevant data-led examples. The data will be downloadable from a companion website (most likely my personal website since it will allow for me to update material at a faster rate). For an example of a problem set introduction in video form, see this video I created for an undergraduate course on econometrics:

[**https://www.youtube.com/watch?v=jCOLMmlD90A**](https://www.youtube.com/watch?v=jCOLMmlD90A)

**Chapter objectives and summaries:** The book will be goal-orientated, with each chapter having a certain intended learning outcome described briefly at its beginning. There will also be a ‘progress arrow’ describing the importance of the chapter in terms of the overall goal of the book; gaining a practical understanding of Bayesian statistics. The chapters will be summarised at their end, and an explanation of the next chapter’s focus provided.

**Case studies/ examples:** The theory will be brought to life via numerous case studies. These would ideally be coloured differently to the rest of the text so that the student, if so wishing, can simply read the theory without having to read all the case studies.

**Diagrams:** An emphasis will be placed on making explanations of theory as visual as is possible. It is my belief that students are discouraged from reading long prose, and find it easier to learn visually.

**Code snippets:** The reader will be provided with code snippets (which will be short) in the chapters themselves. The full code will either be provided by the companion website, or at the end of the chapters/book.

**Market**

**What courses would the book be used for?**

The book would be used as an accompanying source for fairly advanced undergraduate courses on Bayesian statistics, for social scientists. It could potentially also be used as a source for introductory graduate courses on the subject.

**Are such courses core/compulsory?**

For social scientists Bayesian statistics is increasingly becoming more important. Whilst not currently ‘core’ for undergraduates, courses on the subject are certainly of high relevance for graduate students. However, it is my belief that as the popularity of Bayesian statistics increases, courses on this subject will become more central to undergraduate syllabuses. Bayesian statistics is already ubiquitous across research in the sciences, and it is, in my view, likely that this trend will be replicated in the social sciences.

**Course length**

I think where these courses are taught, the length of time taken to cover the material is most likely 2 semesters. The book will complement this by roughly metering one chapter per week of term.

**Market size estimation**

University enrolment across the UK (according to wikipedia) is approximately 2.5m. If we were to assume that 1/10 students study social sciences (I think a fairly conservative estimate), then this would mean that there are approximately 250,000 students in these subjects. If we then assume that half of these students are not at a sufficiently-advanced level to merit the use of a book on Bayesian statistics, then this leaves a market size of 125,000. Obviously, not all of the lecturers of these students will wish to use this book, and there are competing titles out there already, but this provides a basic guide to the likely numbers involved in the UK.

It is hoped that the book will also be useful for the wider audience of students abroad. The videos, which will be available on YouTube, will also act as a marketing tool to increase the notoriety of the book, outside of the UK’s borders.

**Competing Titles**

The main competition in this area is in my view from the following books:

*‘Bayesian Data Analysis’ by Gelman et al.*

This is a good book, with plenty of examples, although is overly long, too advanced for undergraduates, and its first few chapters are full of maths! I know a number of intelligent individuals who have been discouraged from using the book after reading the first few chapters. I think nonetheless this book has been quite successful, and I believe is often used as a central tome for graduate courses on the subject.

*‘Doing Bayesian Data Analysis’ by Kruschke*

A good book; full with graphs and examples. However, in my view, the book is overly long, and the prose at times is verbose. The main problem with the book is though I wouldn’t consider its coverage of a number of topics as suitably modern. In particular the sections on hypothesis testing, regression, and its lack of coverage of objective priors, I consider strong weaknesses. I think that whilst the book provides a good explanation of Bayesian statistics, it doesn’t convince the reader sufficiently of its benefits vs classical statistics.

*‘Bayesian Analysis for the Social Sciences’ by Jackman*

This title of this text seems suggestive of fulfilling a role which I would see the proposed book as occupying. However, a first glance into the book will convince anyone not extremely-mathematics-orientated to avoid it. As such, I believe that this guide is inadequate to cater to the needs of the ‘everyday’ social scientist.

The proposed text will differ from the aforementioned titles by being both accessible to social scientists, and comprehensive in its coverage. I think that it would likely occupy a middle ground between the Gelman and Kruschke texts; having their strengths (Gelman in its coverage, Kruschke in its accessibility), hopefully without their respective weaknesses. I also believe that the addition of videos alongside the text will automatically make the text more accessible than those which have gone before it.

**Writing plan**

**6-8 months:** A draft ofthe first six chapters of the book will be completed.

**18 months:** The first draft of the book will be written, with each of the chapters finished in terms of theory, and the majority of examples.

**18 months – 2 years:** The videos which accompany the book will be produced, and sources of data for the problem setslocated. These will then be paired up with the text to make a contiguous and self-consistent body of work.