## Bayesian book chapter list

The following details the proposed outline of the book. Each number 1-21 pertains to a particular chapter, with the letters representing chapter subsections. Furthermore the book is split into parts, each of which is intended to cover a relatively broad aspect of Bayesian statistics.

1. The purpose of the book, and how best to use it
   1. The goal of this book: after reading this book, the student will be able to understand most of the Bayesian statistical methods that are used in modern applied social sciences research papers. Additionally, the student will feel capable of applying these techniques to recapitulate the results of these research papers.
   2. How to find the videos associated with the relevant material.
   3. How to install and operate BUGs/WinBugs.
   4. How to use the problem sets. Where to find the videos associated with each of these datasets.

**Part I: Understanding the Bayesian formula**

This section will be devoted to developing an understanding of the central formula in Bayesian statistics. The first chapter will explain the purpose of Bayesian statistics; highlighting its differences with classical statistics, and it will end with an introduction of the Bayesian formula. The second chapter (of this section) will explain the first part of the Bayesian formula: the likelihood. The third chapter explains the second part of the Bayesian formula: the prior distribution. The fourth and final chapter of this section will introduce the student to the last component needed to construct the posterior distribution: the denominator of the Bayesian formula.

1. The subjective and the seemingly objective: An introduction to classical and Bayesian statistics.
   1. The goal of this chapter: Introduce and motivate the purpose of statistics in general; classical statistics will be compared and contrasted with Bayesian statistics; the tangible (non-academic) benefits of Bayesian methods will be highlighted; finally, the Bayesian formula will be introduced, defining each of its elements.
   2. The purpose of statistics.
   3. The world according to classical statistics.
   4. The central Bayesian dogma: The Bayesian formula (along with a short biography of Bayes – the tragedy of posthumous success).
   5. An introduction to the Bayesian inference process: Choose a model for data (specify a likelihood); What do you know about the situation? (specify a prior). These two choices together result in ‘updated’ knowledge about the situation (the posterior distribution).
   6. The intuition behind the Bayesian formula.
   7. How do classical and Bayesian theories differ in their approach to statistical inference?
   8. What is a probability? The flexibility of the Bayesian approach in contrast to the classical (frequentist) view.
   9. Explicit vs implicit subjectivity: the danger of the word ‘objective’. This would be a critique of the notion of objectivity commonly thought to hold in frequentist statistics.
   10. What are the tangible (non-academic) benefits of Bayesian theory vs classical statistics? It yields the best predictions, allows easier interpretation of results, and is more intuitive.
   11. Why don’t more people use Bayesian statistics? Existing literature is very heavy with the advanced mathematics behind Bayesian theory. For many students, continued use of classical statistics appears to be the path of least resistance.
   12. Chapter summary. Summary and where we are in understanding the Bayesian formula.
   13. An introduction to the material of the next few chapters: introducing the elements of the Bayesian formula.
   14. Problem set introduction.
2. Choosing an appropriate model for the data: specifying a likelihood.
   1. The goal of this chapter: introduce the concept of likelihood; explain how to choose a likelihood; explain the idea behind maximum likelihood estimation.
   2. What is a likelihood?
   3. Why likelihood is not a probability.
   4. An aside to a classical method: maximum likelihood estimation.
      1. Maximising the ‘probability’ of obtaining the sample from a population.
      2. Example: what’s the probability someone has blood type B? Writing down likelihood.
      3. Why do we maximise log likelihood?
      4. Example continued: Estimating the probability that an individual has blood type B.
      5. How to estimate uncertainty of maximum likelihood estimates? The Cramer-Rao Lower Bound and the intuition as to why this can be used to estimate the variance of estimators.
      6. Example continued: Estimating the imprecision of estimated probability that an individual has blood type B.
   5. How to choose a likelihood appropriate to a given situation.
   6. The subjectivity of model choice.
   7. Chapter summary. An update of where we are in part I: understanding the Bayesian formula.
   8. Problem set introduction.
3. Representing your pre-investigation knowledge of the phenomena in question: the prior.
   1. The goal of this chapter: introduce the concept of a prior; explain how to specify a prior; a short section on objective Bayesian priors.
   2. Introduction to priors. A way of representing current knowledge of the situation.
   3. Example: probability an individual has blood type B. What values would we consider plausible beforehand? How we can express our uncertainty in a prior.
   4. How to specify a prior. A guide to writing down a prior in a given situation. This section will only be brief, as this material is better explained later on after more advanced theory has been covered.
   5. Objective Bayesian priors. How to proceed if we have no prior knowledge about a model’s parameters? This section will be brief, and designed to reassure students.
   6. ‘Good’ model robustness. The predictions and inferences from a good model should be robust to changes in the prior. The prior becomes less important the more data is available.
   7. Chapter summary. An update on where we are in part I: understanding the Bayesian formula.
   8. Problem set introduction.
4. The denominator.
   1. The goal of this chapter. Explain the meaning of the denominator term of the Bayesian formula, and its inherent complexity.
   2. What does the denominator represent? A short introduction to the denominator as the probability of the data (given a choice of model). The denominator as a weighted average. The denominator as a nuisance normalising constant.
   3. The difficulty with the denominator. Why is the denominator often hard to evaluate?
   4. Example: the probability an individual has blood type B. How to evaluate the probability of obtaining the data.
   5. Extension of previous example. The addition of more parameters to the simple example to highlight the complexity associated with evaluating the denominator, especially in the presence of a high degree of uncertainty surrounding the priors.
   6. Dealing with the difficulty. Modern Bayesian (computational) methods ignore the denominator, and can still make exact simulations from the posterior distribution. This will not be a full introduction to, for example, Metropolis-Hastings, but is intended to convey the message, ‘not all hope is lost’.
   7. Chapter summary. Stress that we have now covered all the parts of the Bayesian formula, and are now ready to start applying it.
   8. Problem set introduction.

**Part II: A practical guide to doing (and understanding) analytical Bayesian analysis**

This part of the book will introduce the student to the practicalities of doing analytical Bayesian statistics. The section will focus on analytical Bayesian data analysis: conjugate priors, calculation of posteriors, and prediction of dependent variables. The aim of this section is to gain practical experience of applying Bayesian theory to datasets, and also to enable the student to read and understand texts and papers written using Bayesian statistical analysis. An introduction to using Bayesian theory for forecasting will also be included.

1. An introduction to well known (and frequently used) probability distributions.
   1. The goal of this chapter. As well as introducing the distributions this chapter will aim to explain the importance of each of these distributions in real life applications of Bayesian statistics.
   2. The uniform distribution: A basic way to indicate a lack of knowledge about a process.
      1. Using this distribution to represent beliefs about a probability.
   3. The Normal distribution.
      1. Often invoking the Central Limit Theorem means that it is convenient to think about the likelihood for an observation as being normally distributed.
   4. The Binomial distribution.
      1. For trials with a binary output.
   5. The beta distribution.
      1. A distribution ideal for representing beliefs about probabilities.
   6. Poisson distribution.
      1. Representation for discrete data, where there is a constant probability of an event occurring.
   7. The gamma and inverse gamma distributions.
      1. A way of representing non-negative parameters, which are in theory bounded by positive infinity. In particular it is useful for representing prior knowledge about variances.
   8. Inverse chi-squared distribution.
      1. The resultant posterior from assuming an inverse gamma as a prior, with a normal likelihood.
   9. Chapter summary. The reader will now have a familiarity with several distributions, and an appreciation of when to use particular distributions. It will not be necessary that the reader has a complete understanding of the mathematics behind these distributions.
   10. Problem set introduction.
2. Conjugate priors and their uses in Bayesian data analysis.
   1. The goal of this chapter: To explain the use of conjugate distributions as idealisations which illustrate how more realistic (and typically more complex) models may work.
   2. What is meant by a conjugate prior?
   3. Why are conjugate priors convenient and useful? By knowing the rules, it is possible to write out the posterior distribution without having to do the maths.
   4. A table (across two pages) which details likelihood functions along with the corresponding priors, and resultant posterior in terms of the parameters of the distributions covered in the previous chapter.
   5. Example 1: Single parameter unknown: Calculating the posterior for the case of a binomial/bernoulli likelihood, and beta prior. What is the probability that a randomly selected individual has a disease?
      1. How does choice of parameters in the beta prior affect the posterior?
      2. How does more data influence the posterior? Decreases the importance of the prior. Hence if there is sufficient data then Bayesian analysis can be considered less dependent on experimenter preconceptions.
   6. Example 2: Single parameter unknown: Finding the posterior distribution for the poll result in an election. Assume normal likelihood for the mean of the result, and a normal prior.
      1. Normal self-conjugancy. The normal posterior precision as a sum of the precision of the prior and the data precision.
      2. Increase uncertainty in parameter – what is the effect on the posterior?
      3. Increase the amount of data, and the resulting increase in robustness of the posterior to differing priors.
   7. Example 3: multi-parameter posterior – two binomial proportions. The probability of having a certain disease across both males and females.
      1. Derivation of the overall likelihood for the system.
      2. Derivation of the overall posterior.
   8. Example 4: multi-parameter posterior – uncertainty over both the mean and variance in the election prediction example above. Using a gamma distribution as a prior for the variance.
      1. Calculation of the posterior in its analytic form.
      2. Comparison vs the posterior from example 2. The posterior from Example 2 as a limit under infinite precision in the variance of the distribution.
      3. Visual depiction of the marginal distributions across the two parameters.
   9. Chapter summary. This chapter has been the first real introduction as to how to find posterior densities in practice. After reading this chapter the student will be able to understand most papers which use analytical Bayesian statistics. However, questions remain: how can the posterior be used practically? Also what happens when the prior and likelihood functions are not conjugate? These issues will be covered in the second half of the book.
   10. Problem set introduction.
3. Expressing uncertainty in parameters.
   1. The goal of this chapter: This short chapter will provide the reader with a comprehensive overview of the ways in which parameter uncertainty is expressed in Bayesian theory.
   2. What do we mean by uncertainty in a parameter’s value? Do parameters actually have a point value? A comparison of classical vs Bayesian viewpoints.
   3. The classical confidence interval. An explanation of the issues inherent with this concept, that it really has nothing to do with confidence.
   4. The highest posterior density (HDI). An intuitive (better) alternative to classical regions of uncertainty. The issue with the HDI being that it can lead to non-contiguous regions being selected.
   5. The central posterior interval. An improvement on HDI.
   6. Chapter summary. The reader will now understand how parameter uncertainty is expressed in Bayesian theory, and the improvement this gives over classical statistical methods.
4. Objective Bayesian data analysis.
   1. The goal of this chapter: Considerable attention now centres on making Bayesian analysis, ‘as objective as possible’. I anticipate that this will be an area of increasing importance in the coming years. As such, I would like to devote a part of this book to the principles of *Objective Bayesian data analysis.* Individual chapters will introduce the reader to the Jeffrey’s prior, reference priors and Zellner’s G-priors. This section will necessarily be more focussed on the philosophy and theory behind Bayesian analysis, but will continue to be grounded in data-driven examples.
   2. The uniform prior. At first glances this seemingly uninformative assumption is a good basis for making a non-subjective form of Bayesian prior. However, appearances can be misleading. Use example of probability that an individual has a disease: a uniform prior for the probability here actually implies that the probability of two people selected at random having the disease is quite small.
   3. The Jeffrey’s prior as a potential solution.
   4. The intuitive explanation of Jeffrey’s prior. It produces a prior which is most ‘in line’ with the likelihood, and hence allows the data to ‘speak for itself’.
   5. Proper vs improper priors. The uniform prior as an improper prior. When do we need to worry about improperness?
   6. An introduction to reference priors.
   7. An introduction to Zellner’s G-priors. The idea here is to allow the investigator (often in linear regression models) to specify weak assumptions about the parameters, without worrying about specifics of parameter cross-correlations.
   8. The falsity of the notion of objectivity. Priors are only important if there isn’t a lot of data. In that case they actually allow one to make inferences about a situation where they would not be able to in classical theory.
   9. Empirical Bayes: a way of using data to formulate a prior.
   10. Chapter summary. The student will now be familiar with the principles and intuition underpinning modern Objective Bayesian theory. The reader will also have learned that if there is sufficient data, then the prior is less important, and Bayesian theory is less susceptible to the claim that it is ‘subjective’. Essentially the same adages apply as in classical statistics; one shouldn’t make inferences if there is sufficient data.
   11. Problem set introduction.
5. How to forecast in Bayesian statistics.
   1. The goal of this chapter: This chapter will be short, but is important. Often it is assumed that students will be able to apply the Bayesian formula, and use it for forecasting. I think that this subject is sufficiently important to merit its own section. This section will take the student through the principals of making good forecasts, and through the Bayesian approach.
   2. How do we forecast in classical statistics? What principals make a model able to forecast well? Model parsimony, preventing overfitting.
   3. How are Bayesian techniques used for forecasting? How do they resolve some of the shortcomings associated with classical statistics?
   4. Example: forecasting with a binary dependent variable: what is the probability that a person is cured of a disease after they have taken a drug, given the results of the previous trials?
   5. Forecasting for more general models. Derivation of general marginal distribution on the values of the variable being predicted.
   6. Example: continuation of example 3 in the previous chapter. Prediction of the probability of men and women having a certain disease.
   7. Forecasting election results (continuation of examples 2 and 4 in the previous chapter) with:
      1. Uncertainty over the mean of a normal likelihood.
      2. Uncertainty over both the mean and variance of a normal likelihood.
   8. Chapter summary. On completing this chapter the student will have a good grasp of how the posterior density can be used to create forecasts (with errors of forecast) for a number of situations.
   9. Problem set introduction.

**Part III: A practical guide to doing (and understanding) real life Bayesian analysis**

This section will focus on real life applications of Bayesian analysis; straying away from the comfort and analytical (over-) simplicity of the distributions and assumptions used in part II, by introducing the student to computational Bayesian analysis using BUGs; introducing the student to posterior simulation algorithms. I believe that explanations of the algorithms on which modern Bayesian analysis is built are often where existing texts falter. Fortunately, these algorithms are intuitive, and hence can be explained in a manner which can be understood by anyone.

1. Leaving conjugate priors behind: MCMC.
   1. The goal of this chapter: This chapter will be a bridging chapter; linking part II on analytical Bayesian theory with the computational chapters that follow.
   2. Conjugate distributions are not always realistic assumptions.
   3. Analytic distributions are not necessarily realistic either.
      1. The need to express bi-modal beliefs that may not be expressible in a closed form distribution. Example: previous studies have indicated that the probability that an individual has a disease is either ¼ or ¾. If we believe that one or other of these is correct, then this might motivate the use of a bi-modal distribution.
   4. How to choose a prior distribution that summarises all relevant information about the situation?
   5. All is not lost however when leaving the analytical framework behind: the ability to forget about the denominator. Just concentrate on the landscape swept out by the numerator in parameter space.
   6. An introduction to MCMC. The concept of sampling from the posterior distribution, and using sample statistics to characterise it. The analogy that this is like *in silico* flipping of a coin, and using these flips to understand the probability that a head is obtained, as well as the variance of outcomes.
   7. Chapter summary. The central aim of this chapter is explaining to the reader that the world is more complicated than that which is contained within conjugate distributions. It also explains to the students the central point of numerical Bayes; the idea of using the numerator as a window through which to view the posterior distribution.
   8. Problem set introduction.
2. Computational Bayes I: Grid approximations.
   1. The goal of this chapter: This chapter will introduce the simplest form of numerical Bayesian technique.
   2. Discretising a continuous distribution. The Bayesian formula for discrete random variables.
   3. Combining a discretised likelihood and a prior to get a discrete posterior. Demonstrate that the discrete posterior approaches the exact solution as the number of discretisation points becomes infinite
   4. Approximating the denominator of Bayes’ formula using the discretisation.
   5. Examples 1-4 revisited. The use of grid approximation to sample from the posterior distribution.
   6. Forecasting using the grid approximation.
   7. The issues associated with discretisation: the method quickly becomes computationally prohibitive when dealing with multiparameter models.
   8. Chapter summary. A first step towards a modern application of Bayesian theory.
   9. Problem set introduction.
3. Computational Bayes II: the Metropolis-Hastings algorithm.
   1. The goal of this chapter: This chapter will provide an accessible introduction to this powerful algorithm, and explain the intuition behind the MH simulation technique. It will be supplemented with lots of examples with in R.
   2. The denominator as a source of the nuisance in Bayesian analysis.
   3. How to forget about the denominator?This motivates the use of the Metropolis-Hastings algorithm.
   4. The analogy of walking around a landscape. Moving to a new spot probabilistically if it is lower, and moving to a higher spot definitely. The idea is that you will visit positions at a rate proportional to their height.
   5. How to pick the distance to step? The importance of a proposal distribution.
   6. The ‘burn-in’ period. The necessity to leave the simulator to settle down so that it does not become ‘stuck’ in a particular aspect of the landscape.
   7. Examples 1-4 revisited. Now using Metropolis-Hastings.
   8. Example 5: a system with several variables. The impracticality of the grid-approximation approach when compared to Metropolis-Hastings.
   9. Issues associated with Metropolis-Hastings. The relatively long burn-in and simulation period required before one can sample from the posterior; its inability to deal with particular posterior distributions in an efficient manner; the need to tune the sampler to make it efficient.
   10. Chapter summary. The reader will understand how MCMC works for the case of the MH algorithm. They will also be able to use it to sample from the posterior distribution.
   11. Problem set introduction.
4. Computational Bayes III: the Gibbs sampler.
   1. The goal of this chapter: An introduction to another powerful algorithm, which is in general more efficient than MH. Explanation that this is the workhorse behind BUGS.
   2. The Gibbs sampler as a subset of the Metropolis-Hastings algorithm. Alternatively, vice versa as well!
   3. The intuition behind the Gibbs algorithm. Imagine wanting to sample the heights of all the points on earth as a posterior distribution. One could do a random walk a la MH, but if the proposal distribution is too narrow, then the sampler could get stuck on a mountain range or desert; alternatively large swathes of landscape may be ignored if the step size is too large. This means it can take require a long burn-in, and may or may not be representative of the underlying landscape. The Gibbs algorithm works by cutting through the earth along lines of latitude or longitude, and allowing all steps along this conditional distribution. Since arbitrary step lengths are allowed, the algorithm is quicker, and there is no need to fine tune it to the degree of MH.
   4. The Gibbs sampler for simple examples, where the conditional distributions of parameters are known.
   5. The main issue with Gibbs: the conditional distribution must be known.
   6. Chapter summary. The benefits of the three types of sampler. Grid approximation for simplicity, MH for general problems, Gibbs for problems for which the conditional distribution has an analytical representation.
   7. Problem set introduction.
5. Gibbs sampling in practice: An introduction to BUGS.
   1. The goal of this chapter: The aim of this chapter is to provide a comprehensive introduction to the use of WinBUGS; how to set up simple simulations, and how to analyse simulation results in R. From this chapter onwards BUGS will be used to run simulations wherever possible.
   2. Running WinBUGS as a standalone, or calling it from R using R2WINBUGS.
   3. How to run a simple WinBUGS example.
   4. How to visualise the results using R.
   5. How to make predictions using BUGS and R?
   6. Chapter summary. The reader will be able to run basic simulations in BUGS, and know how to use and manipulate the results in R.
   7. Problem set introduction.

**Part IV: Regression analysis and hierarchical models**

The first chapter of section part will explain how to test hypotheses, and evaluate a model’s fit; useful tools to evaluate models which will be introduced in the next sections. The final chapter of this part will introduce regression models in a Bayesian framework; emphasising the benefits of hierarchical models in these models.

1. Hypothesis testing I: The Bayesian equivalent of classical tests.
   1. The goal of this chapter: Classical hypothesis tests will be introduced, and their issues extensively explained. The Bayesian equivalents of these tests will also be introduced.
   2. The classical approach: null vs alternative.
   3. The subjectivity inherent with classical hypothesis testing.
   4. Bayesian null vs alternative hypothesis testing.
   5. The classical confidence interval, and its problems.
   6. Bayesian approaches to confidence. Is the parameter in the HDI?
   7. The region of practical equivalence (ROPE).
   8. How should we judge model fit? There is no simple answer, but it should be to do with what we hope to achieve by estimating a model in the first place.
   9. Classical approaches to model fit. R-squared, AIC and BIC. The lack of adequacy of these methodologies for a Bayesian framework. The BIC as a Bayesian purist’s way of judging model fit.
   10. Chapter summary. The student will now be aware of the classical approaches to hypothesis testing, and their associated problems, as well as the Bayesian equivalents.
   11. Problem set introduction.
2. Hypothesis testing II: The pure Bayesian approach to hypothesis testing.
   1. The goal of this chapter: The modern Bayesian approaches to hypothesis testing will be covered, and their benefits emphasised relative to classical approaches.
   2. A comparison between models by computing the ratio of P(data).
   3. Jeffrey’s scale as an arbitrary way of selecting between models, but nonetheless a frequently used methodology.
   4. Simulation as a method of evaluating model fit.
   5. Posterior predictive probabilities. Generate new data from your model by simulating from the posterior. Compare the ‘generated’ data with the actual data. Do these look similar?
   6. Test statistics in Bayesian theory as a means of evaluating model fit.
      1. This section will necessarily be quite substantial, and full of examples in order to convey the flexibility and comprehensiveness of this approach.
      2. The chi-squared measure of model fit.
      3. Graphical means of testing a model’s fit.
   7. Sensitivity analysis. The idea of fitting several different likelihoods and priors to the data, and seeing whether the results which are obtained vary significantly.
   8. Expected deviance.
   9. Chapter summary. The reader should now have a very good understanding of the transparent methods used to evaluate a Bayesian model. This should allow them to read, and critique many applications of Bayesian models, since this is often where applications fall down.
   10. Problem set introduction.
3. Hierarchical models.
   1. The goal of this chapter: Introduce the reader to the concept of ‘hierarchical’ models, where uncertainty in parameters’ uncertainty is taken into account, and these chains of priors for a logical chain which is the backbone of modern Bayesian statistics. There will also be a discussion as to when to ‘stop’ iterating upwards with hyperpriors.
   2. What is meant by a hierarchical model?
   3. What are the benefits of a hierarchical vs a simpler model?
   4. An example of the inadequacy of a non-hierarchical model.
   5. Diagrams which explain parameter dependencies, and translation of these into BUGS code.
   6. What is the appropriate level to stop iterating upwards with hyperpriors?
   7. Another example of the use of hyperpriors. The diminishing returns to increasingly abstract hyperpriors.
   8. Chapter summary. The benefit of hierarchical models, and when to use them.
   9. Problem set introduction.
4. Linear regression models.
   1. The goal of this chapter. For a causal reader, the advantage of Bayesian approaches to econometrics is not made clear. An example of a text which fails in this regard is Cooper’s text on Bayesian econometrics. On first glances, since the Bayesian point estimates of parameters are often not considerably different to those achieved in classical theory, the benefits are not immediately obvious. The benefits of Bayesian linear regression will be extolled in this chapter: the ability to test a model in a much more extensive (and informative) manner, and the simple extension to use hierarchical models.
   2. The formulation of regression models in a Bayesian framework.
   3. The choice of priors for the parameters.
   4. The similarity between classical and Bayesian linear regression results.
   5. The difference between classical and Bayesian linear regression results.
   6. The power of Bayesian linear regression over frequentist approaches:
      1. Testing the results of the analysis in a robust way.
      2. Simple extension to hierarchical models.
   7. Chapter summary. The student will now be aware of the benefits of Bayesian approaches to regression.
   8. Problem set introduction.

**Part V: GLM, Bayesian Decision Theory, and New advances in simulation.** This part will deal with relatively advanced (but important) topics in Bayesian theory. The topics will all be individually explained in a very applied manner, with a lot of examples. Whereas it is essential that the reader learns about GLM, the chapters on Bayesian Decision Theory and ‘Advanced simulation’ methodologies are non-essential, but I anticipate will become more important in later editions, as these topics grow in importance in the social sciences. These chapters will be relatively short, and will give only short introductions to the two subfields.

1. Generalised linear models. This chapter will provide a description of how to apply Bayesian theory to models with a non-linear link function: logistic/probit regression, poisson regressions, multinomial models. Shrinkage estimators will also be introduced. This chapter will make extensive use of the chapter on hierarchical models, and will be very example-led; making extensive use of BUGS and R.
2. Bayesian decision theory. This chapter will be a brief introduction to using Bayesian theory to make decisions under uncertainty. This area is very much growing and will become more important in years to come. It is more complex that just Bayesian statistics, but it can still be explained in a relatively simple manner. The hope is that this chapter will make the student aware of the application of this topic.
3. Advanced simulation techniques for Bayesian theorists. This chapter will describe the modern techniques to speed up convergence and simulation from a posterior, covering topics such as simulated annealing, approximate Bayesian computation, reverse jump MCMC, and Sequential Monte Carlo. As well as explaining (in brief) these theories, this section will provide the reader with a number of references for further reading.