## Bayesian book chapter list

1. The purpose of the book, and how best to use it
   1. The goal of this book: after reading this book, the student should be able to understand the majority of the Bayesian statistical methods that are used in modern applied social sciences research papers. Furthermore they should have an idea about how to go about how practically to go about recapitulating their results.
   2. How to find the videos associated with the relevant material.
   3. How to obtain BUGs/WinBugs. Note this will be a *very* comprehensive guide to getting the software to working starting from how to install the software.
   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 is devoted to developing an understanding of the central formula in Bayesian statistics. The first chapter will explain the purpose of Bayesian statistics, and highlight its differences with classical statistics; culminating 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 final chapter of this section will introduce the student to the last component necessary to build 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 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 Bayesian central 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 address the purpose of statistics?
   8. What is a probability? The flexibility of the Bayesian notion opposed 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, continuing with using classical statistics appears to be the path of least resistance.
   12. Chapter conclusion: 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 its intuition.
      6. Example continued: Estimating the imprecision of estimated probability that an individual has blood type B.
   5. How to choose a likelihood appropriate to the 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. The prior: representing your pre-investigation knowledge of the phenomena in question.
   1. The goal of this chapter: introduction to the concept of a prior; an introduction as to 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 as to how to write down a prior in a given situation. This will only be a short section, as it is better explained later on after more theory has been covered.
   5. Objective Bayesian priors. What if we have no idea about the value of a model’s parameters? This concept will be only explained in brief, as the purpose isn’t to go into depth about how to carry out Objective Bayesian inference. It is more to put the student at ease about the supposed subjectivity of Bayesian methods.
   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 of 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 mean? This is a short introduction to the meaning of 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. Example made more complex. The addition of more parameters to the previously simple probability example to highlight the increase in complexity of evaluating the denominator, especially in the presence of a high degree of uncertainty surrounding the priors.
   6. The difficulty made easy. Modern Bayesian (computational) methods ignore 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’, to the reader.
   7. Chapter summary. Description of the fact 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 reason behind this section is to gain experience with using Bayesian theory on datasets, and also to allow the student to be able to read and understand a significant proportion of the texts and papers written using this type of statistical analysis. It will also provide an introduction as to how Bayesian theory can be used for forecasting.

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 each of these distributions in real life 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 evocation of the Central Limit Theorem means that it is convenient to think about the likelihood for an observation as being normally distributed.
      2. Conjugate to itself.
   4. The Binomial distribution.
      1. For trials where a result can either be a success or a failure.
   5. The beta distribution.
      1. A distribution ideal for representing beliefs about probabilities.
   6. Poisson distribution.
      1. Useful representation for discrete data, where there is a constant probability of an event occurring.
   7. The gamma and inverse gamma.
      1. A way of representing non-negative parameters, which in theory are bounded by positive infinity. In particular it is useful for representing prior knowledge about variances.
   8. Inverse chi-squared.
      1. The resultant posterior from assuming an inverse gamma as a prior, with a normal likelihood.
   9. Chapter summary: we should now understand the use of a number of distributions. It is NOT necessary to understand how the distributions work mathematically, just to be aware of them, and the types of situation where they can be used.
   10. Problem set introduction.
2. Conjugate priors and their uses in Bayesian data analysis.
   1. The goal of this chapter: The aim of this chapter is to explain the use of conjugate distributions as being idealisations which can help illuminate how more realistic (and usually complex) models may work.
   2. What is meant by a conjugate prior?
   3. Why are conjugate priors convenient? Avoid actually having to do the maths if you know the rules, and can simply ‘write out’ the posterior density.
   4. How useful are conjugate priors in real life data examples? Pretty useful, as even though the set of distributions they span is relatively restrictive, it is still possible to get insight into more complex situations using these simple building blocks.
   5. A table (across two pages) which details likelihood functions along with the corresponding priors, and resultant posterior in terms of the parameters of the aforementioned.
   6. 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 different 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 thought to be less dependent on experimenter preconceptions.
   7. 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 on 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.
   8. 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.
   9. 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.
   10. Chapter summary: This chapter has been the first real introduction as to how to find posterior densities in practice. After reading this chapter this should allow the student to read, and hopefully be able to digest, a number of papers which make use of 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 will be covered in the second half of the book.
   11. Problem set introduction.
3. Expressing uncertainty on parameters.
   1. The goal of this chapter: This will be a short chapter which provides 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 here.
   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. 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 should now understand the ways in which parameter uncertainty is expressed in Bayesian theory, as well as its improvement over classical statistical methods.
4. Objective Bayesian data analysis.
   1. The goal of this chapter: Considerable attention in modern literature is around attempts to make 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 so-called *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 a little more focussed on the philosophy and theory behind Bayesian analysis, but will continue to be grounded in data-based examples.
   2. The uniform prior. On first glances it appears that this seeming 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 intuition behind the 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 having to worry about the 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. Chapter summary: The student should be aware of the principles and intuition behind modern Objective Bayesian theory. The reader should also have learned that if there is sufficient data, then the prior is of less important, and Bayesian theory is less susceptible to the claim that it is ‘subjective’. Essentially the same the adages apply as in classical statistics; one shouldn’t make inferences if there is sufficient data.
   10. Problem set introduction.
5. How to forecast in Bayesian statistics
   1. The goal of this chapter: This chapter will be short, but I think is quite essential. Often it is simply assumed that students can apply the Bayesian formula, and that it is self-evident how this can be used for forecasting. I think that this subject is sufficiently important that it merits its own section, which will take the student firstly through the principals of making good forecasts, and then through the Bayesian approach.
   2. How do we forecast in typical statistics? What are the principals that make a model able to forecast well? Model parsimony, preventing overfitting.
   3. How do Bayesian ‘do’ forecasting? How do these methodologies prevent some of the issues prevalent in classical statistics?
   4. Find the expected value of y given the past values of y.
   5. 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.
   6. Forecasting for more general models. Derivation of general marginal distribution on the values of the variable to be predicted.
   7. Example: following on from Example 3 in the previous chapter. Prediction of the probability of men and women having a certain disease.
   8. Forecasting election results (following on from 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.
   9. Chapter summary: after reading this chapter the student should have a grasp as to how the posterior density can be used to create forecasts (with errors of forecast) for a number of situations.
   10. 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 the previous part, 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. I also believe that, 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 is aimed as a bridging chapter; linking together the previous part on analytical Bayesian theory with the computational chapters which follow it.
   2. Real life doesn’t deal solely with conjugate distributions.
   3. Real life doesn’t just deal with analytic distributions.
      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 the 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. An introduction to BUGs.
   1. The goal of this chapter: The idea behind this chapter to provide a comprehensive introduction to the use of WinBUGS; how to set up simple simulations, how to use the results of the simulation in R to visualise the posterior and make predictions. After this chapter the book will use BUGS to run simulations, and produce the visualisations of the simulations.
   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 do prediction using BUGS and R.
   6. Chapter summary: the reader should be able to understand how to run basic simulations in BUGS, and how to use and manipulate the results in R.
   7. Problem set introduction.
3. Computational Bayes introduction part 1: Grid approximations.
   1. The goal of this chapter: This chapter has the aim of introducing the student to the most easy to understand form of numerical Bayesian technique.
   2. Discretising a continuous distribution. The Bayesian formula for discrete random variables.
   3. Multiplying together a discretised likelihood and a prior to get a discrete posterior. Approaches the continuous distribution as the number of points tends to infinity.
   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 inherent in using the discretisation: the method quickly becomes untenable when dealing with multiparameter models.
   8. Chapter summary: the first step towards a modern application of Bayesian theory.
   9. Problem set introduction.
4. Computational Bayes introduction part 2: the Metropolis-Hastings algorithm.
   1. The goal of this chapter: This chapter aims to provide a very accessible introduction to this powerful algorithm, and explain the intuition behind the MH simulation technique. It will also provide plenty of examples with using BUGS/R.
   2. The denominator as a source of the nuisance in Bayesian analysis.
   3. How to forget about the denominator: motivating 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 a spot which is higher in a frequency that is proportionate to its 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 a number of variables. The impracticality of the grid-approximation approach when compared to Metropolis-Hastings.
   9. The issues with Metropolis-Hastings. The relatively long burn in and simulation period required to begin with before one starts to sample from the posterior. Its inability to deal with particular posterior distributions in an efficient manner.
   10. Chapter summary: the main point about this chapter should be that the reader is able to understand how MCMC works for the case of the MH algorithm. They should also be able to use it to sample from the posterior distribution.
   11. Problem set introduction.
5. Computational Bayes introduction part 3: 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 the issue with this is that if the proposal distribution is too narrow, then the sampler could get stuck on a mountain range or desert; alternatively it could completely miss out large swathes of landscape due to taking steps that are too far. 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 all steps 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 all known.
   5. The issue with Gibbs that it is necessary to know the conditional distribution.
   6. Chapter summary: the benefits of each of the three types of sampler. Grid approximation for simplicity, MH for general problems, Gibbs for problems where we can analytically derive the conditional distribution.
   7. Problem set introduction.

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

The first part of section part will introduce the reader to means as to how to test hypotheses, and evaluate a model’s fit; providing the student with the tools which are required to evaluate regression models introduced next. The benefits of regression models in Bayesian frameworks in part come as a result of the use of hierarchical models; meriting a section describing the concept of a hierarchical model. The next part of this section will introduce the reader to some of the quirks and benefits of regression analysis with Bayesian statistics, making ample use of the chapter before it on hierarchical models.

1. Hypothesis testing and evaluating model fit.
   1. The goal of this chapter: This chapter will explain state-of-the-art techniques (which thankfully are full of intuition) to testing hypotheses about parameters, as well as evaluating the fit of a Bayesian model. It will start out by outlining classical approaches to this, and explaining their shortcoming.
   2. The classical approach: null vs alternative.
   3. The subjectivity inherent with classical hypothesis testing.
   4. Bayesian approaches. Is the parameter in the HDI?
   5. ROPE.
   6. Bayesian null vs alternative hypothesis testing.
   7. Classical approaches to model fit. R-squared, AIC and BIC. The lack of adequacy of these methodologies for a Bayesian framework.
   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. A comparison between models by computing the ratio of P(data). The issue with this approach (the Bayes factor problem).
   10. Simulation as a method of evaluating model fit.
   11. 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?
   12. 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.
   13. Sensitivity analysis. The idea of fitting several different likelihoods and priors to the data, and seeing whether the results which are obtained vary significantly.
   14. Expected deviance.
   15. 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.
   16. Problem set introduction.
2. Hierarchical models. The aim of this chapter is 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.
3. Linear regression models. 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 main benefits of a Bayesian approach to regression are: the ability to test a model in a much more extensive (and informative) manner, and the ability to use hierarchical models.

**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 plenty 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 become more important in the social sciences.

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. 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 is 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, and reverse jump MCMC.