Study Guide

# Bayesian Statistical Methods (BAY)

Semester 2, 2025

Prepared by:
Dr Shuvo Bakar
School of Public Health
The University of Sydney

## Contents

## Bayesian Statistical Methods (BAY)
## Semester 2, 2025

## Contact details

**Dr Shuvo Bakar**

School of Public Health, The University of Sydney

shuvo.bakar@sydney.edu.au
(02) 9036 9262

If you have any general BCA queries, please contact: Jacqueline Vaughan at the BCA Coordinating Office on 02 9562 5076/54 or email bca@sydney.edu.au

## Welcome letter

Welcome to Bayesian Statistical Methods (BAY). The application of Bayesian methods in medicine and health sciences has become increasingly vital as we strive to enhance our understanding of improve diagnostic accuracy, and personalise treatment plans. This unit is designed to introduce students with a background in medicine and health sciences to the principles and practices of Bayesian statistics, providing a practical framework for applying these methods in their respective fields.

This unit is delivered through the eLearning site at the University of Sydney. All course content, excluding required readings, will be uploaded to eLearning, including assignments and supplementary materials. Discussions will take place on the Discussion Board. There is currently an “Introductions” thread on the Discussion Board, please use this thread to introduce yourself to the rest of the class. This unit requires access to the R statistical software, and a detailed installation guide has been uploaded to eLearning. You should organise access to the software as soon as possible.

If you have any questions or issues, please contact me by email. My email address is shuvo.bakar@sydney.edu.au or you can contact me via the Canvas page links. I hope you enjoy the course!

Shuvo Bakar
July 2025

## Background

Bayesian inference is concerned with fitting full probability models to data and summarizing the results as a probability distribution for the parameters that define the model and any other unobserved quantities that can be predicted (prospectively or retrospectively) as a consequence of the model. The Bayesian approach allows the data analyst to make probabilistic statements about both quantities they observe and quantities about which they wish to learn, including model parameters, imputations for missing data and predictions for future observations. In contrast, the standard frequentist approach makes modelling assumptions only about observable quantities and cannot benefit from the explicit use of subjective probability to quantify uncertainty about inferences. Bayesian inference is important in modern statistical practice, including biostatistics. This is in part due to a belief by some statisticians that it has philosophical and logical advantages. It is, however due much more to the ability of the Bayesian modelling process to tackle complex hierarchical problems and, most importantly, the availability of new computational techniques and user-friendly software tied to the ever-increasing processing power of desktop computers.

## Context within the program

This course differs somewhat from many of the other BCA units in the program in that it does not require much in the way of ‘hands on’ analysis or application of formulae (though there is some of this)! It mainly involves working through principles and concepts and applying these to real life situations and problems likely to be encountered in the data analysis and modelling.

In keeping with the above philosophy, the course material is based around extracts from books and published journal articles. The use of eLearning is very important in this course as it provides a guide to the course material and opportunities for discussion and clarification of concepts. You are strongly encouraged to make the most of the Discussion Forums to ask questions about course-related administration, clarify concepts and to understand the relevance of the articles provided.

The **prerequisite** courses are:

Epidemiology (EPI), Mathematical Foundations for Biostatistics (MFB), Principles of Statistical Inference (PSI), Regression Modelling for Biostatistics 1 (RM1)

## Unit summary

This unit introduces Bayesian statistical concepts and methods, with emphasis on practical applications in biostatistics. We begin with a discussion of subjective probability in quantifying uncertainty in the scientific process. Subsequently, the concept of full probability modelling is introduced and developed through single- and multi-parameter models with conjugate prior distributions.

The connection with frequentist approaches is examined in light of the relationship between non-informative and informative prior distributions and their effect on posterior estimates. We discuss the specification of appropriate prior distributions, including the concepts of non-informative and weakly informative priors. We consider the frequentist properties of Bayesian procedures.

This unit explains why Bayesian statistics requires advanced computational techniques, and introduces the class of Markov Chain Monte Carlo (MCMC) algorithms to explore the full posterior distribution. The challenges of and intuition behind the algorithms are shown. Subsequently, we focus on assessing whether the algorithm converged and how to remediate failures. We discuss posterior predictive checks and information criteria as means to assess model fit and compare alternative models.

We illustrate the richness of Bayesian modelling software as a means to estimate hierarchical models and discuss the role of shrinkage. Finally, we revisit the full Bayesian workflow.

This unit is taught exclusively using the R software, using its RStan library and related packages, e.g., brms. All examples and tutorials are provided in R only.

## Workload requirements

The expected workload for this unit is 10-12 hours per week on average, consisting of guided readings, discussion posts, independent study and completion of assessment tasks.

## Learning outcomes

At the completion of this unit students should be able to:

**LO1:** Explain the difference between Bayesian and frequentist concepts of statistical inference.

**LO2:** Demonstrate how to specify and fit simple Bayesian models with appropriate attention to the role of the prior distribution and the data model.

**LO3:** Explain how these generative models can be used for inference, prediction and model criticism.

**LO4:** Demonstrate proficiency in using statistical software packages (R) to specify and fit models, assess model fit, detect and remediate non-convergence, and compare models.

**LO5:** Engage in specifying, checking and interpreting Bayesian statistical analyses in practical problems using effective communication with health and medical investigators.

## Unit content

The unit is divided into 6 modules with 12 lectures, summarised in more detail below.

Each module will involve approximately 2-3 learning activities that may be either interactive, independent or collaborative in nature (no collaborative Assignments). This includes activities:

1. Watching recorded short videos (2 to 5) describing concepts and methods, and

provide reflections based on the videos (visual and auditory learning).

2. Reading module notes or reading describing concepts and methods (linguistic learning).
3. Independently and collaboratively completing tutorial exercises and comparing solutions (solitary and social learning).
4. Participating in live interactive tutorials and discussions (social, visual and auditory learning).
5. Working through provided computer code (linguistic and solitary learning).

The diversity of learning activities are designed to engage the different ways that students learn. Study materials for all Modules are accessed from Canvas. Assignments and supplementary material such as datasets will be available within each Assignment item. Please note that we are not able to post copies of copyright material (journal articles and book extracts)—for these you will have to rely on your home university's library.

## Recommended approaches to study

Students should work through each module systematically, following the module notes and any readings referred to, and working through the accompanying exercises. *You will learn a lot more efficiently if you tackle the exercises systematically as you work through the notes.* You are encouraged to post any content-related questions to eLearning, whether they relate directly to a given exercise, or are a request for clarification or further explanation of an area in the notes. You should also work through all of the computational examples in the notes for yourself on your own computer.

Outline solutions to the exercises in each module (except those to be submitted for assessment, as described below) will be posted online at the midway point of the allocated time period for the module. This is intended to encourage you to attack the exercises independently (or via the eLearning site), and yet not make you wait too long to see the sketch solutions.

## Method of communication with coordinator

Shuvo Bakar is the unit coordinator and instructor for Semester 2 2025. His email is shuvo.bakar@sydney.edu.au

Questions about administrative aspects or course content can be emailed to the coordinator, and when doing so please use "BAY:" in the Subject line of your email to assist in keeping track of our email messages. Coordinator/s will be available to answer questions related to the module notes and practical exercises, and to address any other issues that require clarification. However, please note that instructors are not necessarily available every day of the week and you should expect that it may take a day or so to respond to questions (possibly longer over weekends and during breaks!).

We strongly recommend that you post content-related questions to the Discussions tool in the DES area of BCA's eLearning site. The BCA uses the University of Sydney online Learning Management (eLearning) System (LMS), called **Canvas**. For information on eLearning, see the BCA Introduction to eLearning. Once you have read the instructions and the login advice provided by BCA office staff you may login at: canvas.sydney.edu.au

Module notes, data files and other documents will be made available on eLearning. Assignments and course announcements will likewise be uploaded to eLearning. Communication should generally be via the Discussion Board on eLearning (unless of a personal/confidential nature). You are encouraged to post questions, ideas, suggestions and discussions on eLearning. The Course Coordinator will monitor and respond to communication; however, you are encouraged to answer other students' questions or assist in solving problems (with the exception of assignment question queries, which I will clarify).

## Module descriptions

Throughout the course, we will explore the following key areas:

**Module 1: Bayesian Dreams! Navigating Evidence and Inference:**
Understanding the basics of Bayesian philosophy and how it differs from traditional frequentist approaches. Learning how to update beliefs in light of new data using Bayes' theorem. Exploring directed acyclic graph (DAG) in the Bayesian modelling context, with graphical representations of the probabilistic relationships between variables and parameters.

**Module 2: Chaotics? Prior Problems, Tools, and Computation:**
Exploring the role of prior information and how it influences posterior conclusions. Bayesian context of exact inference and computational techniques for approximating complex posterior distributions such as Markov chain Monte Carlo (MCMC). Generative models with prior and posterior predictive checks.

**Module 3: Bayeswatch! Keeping an Eye on Your Gaussian Model:**
Understanding causation and correlation and how it can be used in Bayesian context by drawing DAG. Nevigating Bayesian hierarchical models with continuous outcome or endpoint variable. Explore the choice for hyper-parameters of prior distributions for Gaussian model. Explore examples in clinical and health research.

**Module 4: Bayeswatch! Keeping an Eye on Your Non-Gaussian Model:**
Extending Bayesian models beyond normality assumptions, where outcome or endpoint variable is binary or counts (i.e., generalised linear models under Bayesian hierarchy), and explanation with DAG. Explore key tactics on the choice of prior distributions.

**Module 5: Clusterphobia? Let Bayes Handle It!:**
Understanding and implementing hierarchical models for complex data structures common in health sciences. Learning the use of latent process modelling in Bayesian hierarchy (i.e., similar to mixed models in frequentist settings) with both Gaussian and non-Gaussian (e.g., binomial) distributions.

**Module 6: Wander into the Wonder! Bayesian Secrets to the Right Sample:**
Bayesian sample size calculations, in particular to aid the sample size selection to design trials. Discussion of adaptations with relevant sample size calculations using Bayesian methods. Bayesian model choice (e.g., Bayes factor, deviance information criterion (DIC), Watanabe-Akaike information criterion (WAIC) and leave-one-out (LOO) cross-validation).

## Course timetable

Semester 2, 2025 starts on Monday 4th August.

| Week | Week Commencing | Module | Topic | Learning Outcomes | Assessment |
|---|---|---|---|---|---|
| 1 | 28 Jul 2025 | Module 1 | Navigating Evidence | LO1, LO5 | |
| 2 | 04 Aug 2025 | Module 1 | Bayesian Inference | LO1, LO2, LO5 | |
| 3 | 11 Aug 2025 | Module 2 | Prior and Posterior | LO2, LO4 | Assignment #1 |
| 4 | 18 Aug 2025 | Module 2 | Generative Models and Tools | LO3, LO4 | |
| 5 | 25 Aug 2025 | Module 3 | Logical Connections | LO1, LO2, LO5 | |
| 6 | 01 Sep 2025 | Module 3 | Prior Tweaks and More | LO1, LO2, LO4, LO5 | ***Assignment #1 - Due 5 September (30%)*** |
| 7 | 08 Sep 2025 | Module 4 | Non-Gaussian | LO1, LO2, LO5 | |
| 8 | 15 Sep 2025 | Module 4 | More on Non-Gaussian | LO2, LO5 | Assignment #2 |
| 9 | 22 Sep 2025 | Module 5 | Cluster Smart with Bayes | LO2, LO4, LO5 | |
| | ***29 Sep 2025*** | | ***Mid Semester Break*** | | |
| 10 | 06 Oct 2025 | Module 5 | Goldilocks | LO2, LO4, LO5 | ***Assignment #2 - Due 6 October (30%)*** |
| 11 | 13 Oct 2025 | Module 6 | Size Matters! | LO1, LO5 | |
| 12 | 20 Oct 2025 | Module 6 | Discover the Perfect Spell! | LO4, LO5 | Assignment #3 |
| 13 | 27 Oct 2025 | | Revision | | |
| 13 | 03 Nov 2025 | | | | ***Assignment #3 – Due 7 November (40%)*** |

## Assessment

The assessment for this unit will involve three written assignments which will be posted on eLearning. Assessments are due by 11:59pm on the due date.

| Assessment name | Assessment type | Coverage | Learning objectives | Weight |
|---|---|---|---|---|
| Assignment 1 | Assignment | Modules 1 & 2 | 1-5 | 30% |
| Assignment 2 | Assignment | Modules 3 & 4 | 1,2,4,5 | 30% |
| Assignment 3 | Assignment | Modules 5 & 6 | 1,2,4,5 | 40% |

Module solutions/guides will be posted on eLearning after the submission date. Individual feedback on assignments will be provided to each student.

Students are expected to monitor eLearning for the posting of assignments, solutions and feedback. Email notifications and other channels of communication will not be used.

Examples and exercises are contained in each module to enable students to ascertain their level of understanding of various topics. These will not form part of the assessment of this course.

In general, you are required to submit your work typed in Word or similar (e.g. using Microsoft's Equation Editor for algebraic work) and we strongly recommend that you become familiar with equation typesetting software such as this. If extensive algebraic work is involved you may submit neatly handwritten work, however please note that marks will potentially be lost if the solution cannot be understood by the markers due to unclear or illegible writing. This handwritten work should be scanned and collated into a single pdf file and submitted via the Canvas site. See the BCA Assessment Guide document for specific guidelines on acceptable standards for assessable work.

Students are encouraged to discuss relevant topics in the Discussion Board. However, please avoid posting questions relating directly to assessable material. These should be emailed to the Unit Coordinator in the first instance.

*Explicit solutions to assessable exercises should not be posted for others to use.* Each student's submitted work must be clearly their own, with anything derived from other students' discussion contributions clearly attributed to the source.

**Submission and academic honesty policy**

All assessment material should be submitted via the relevant Assessment module in Canvas unless otherwise advised. Turnitin plagiarism detection is applied to all submissions. For detailed information, please see the BCA Assessment Guide, which includes links to the Academic Honesty policies at member universities. Please familiarise yourself with the procedures and policies at your home university. You will need to indicate your compliance with the plagiarism guidelines and policy at your home university.

*A special note regarding "contract cheating" sites:* Unfortunately, there have been instances in the past of students using such websites to post assignment questions and receive solutions (usually for a fee). We have arrangements with these sites to identify the student posting questions or accessing the solutions, and such students will be referred to and face disciplinary processes at their home university.

**Use of ChatGPT or other generative AI tools in assessment tasks**

The assessment tasks in this Unit have been designed to be challenging, authentic and complex. Although individual assessment components may provide specific guidance regarding the use of generative AI tools (e.g., ChatGPT), successful completion of these components will require students to critically engage in specific contexts and tasks for which artificial intelligence will provide only limited support and guidance. In all cases, a failure to reference the use of generative AI may constitute student misconduct under the Student Code of Conduct of your University of enrolment. To successfully complete assessment tasks, students will be required to demonstrate detailed comprehension of their written submission independent of AI tools.

**Late submission and extension procedure**

The standard BCA policy for late penalties for submitted work is a 5% deduction from the earned mark for each day the assessment is late, up to a maximum of 10 days (including weekends and public holidays). Extensions are possible, but these need to be applied for (by email) as early as possible. The Unit Coordinator can approve extensions up to three days; for extensions beyond three days, you must apply to your home university, using their standard procedures.

## Learning resources

There is **no** required textbook. The subject notes are self-contained. The subject notes were inspired by following text books:

Gelman A, et al. (2021) Bayesian Data Analysis (3rd Edition)
McElreath R. (2020) Statistical Rethinking (2nd Edition)
Lee P. (2012) Bayesian Statistics: An Introduction (4th Edition)
Lambert B. (2018) A Student's Guide to Bayesian Statistics
Kruschke J. (2015) Doing Bayesian Data Analysis (2nd Edition)
Berry, S. M., Carlin, B. P., Lee, J. J., & Muller, P. (2010). *Bayesian adaptive methods for clinical trials*. CRC press.

## Software requirements and assumed knowledge

For this subject you will need to have access to***:***

**R, including the package brms.** These will in turn pull in RStan. ggplot2 and other libraries. All this software is free and gratis. RStan requires R version 3.4 or later. The subject notes were prepared using R 4.2.1.

The RStan team "strongly" recommends that those who use RStudio use version 1.4 or later. The coordinator does not use RStudio but assumes that students who use RStudio are better off upgrading to 1.4.

Linux users should be aware that R's install.package() occasionally creates conflicts when mixed with the distribution's package manager (apt, rpm, ...). The unit coordinator prefers to have all R packages installed via the linux distribution. Some distributions have limited packages available from the package manager by default. Please see https://cran.r-project.org/bin/linux/ for instructions on adding software repositories. It is plausible that installing everything using install.package() also avoids conflicts.

Note that installing brms **may** require installing the developer tools (Rtools) for R if these are not yet installed on your system. It is more likely that they are missing on MS Windows but even then, often the install.packages() command will pull them in as a dependency. If you don't use linux, please try installing by using via the RStudio interface or by install.packages('brms') **before** attempting instructions from other tutorials found online.

**Please install RStan and verify installation in the first week of the unit, so there is sufficient time to troubleshoot issues. To verify, load the rstan library and then enter the command below in an R session:**

example(stan_model, package = "rstan", run.dontrun = TRUE)

This will spit out strings of text, will take a few minutes to start sampling, and may provide a warning. It should NOT result in an error.

## Required mathematical background

Students who have undertaken the pre-requisites will have the required mathematical background for the course.

## Feedback

Our feedback to you:

The types of feedback you can expect to receive in this unit are:

- Formal individual feedback on submitted exercises assignments
- Feedback from non-assessed online quizzes
- Responses to questions posted on Blackboard

Your feedback to us:

One of the formal ways students have to provide feedback on teaching and their learning experience is through the BCA student evaluations at the end of each unit. The feedback is anonymous and provides the BCA with evidence of aspects that students are satisfied with and areas for improvement.

## Unit changes, including response to recent student evaluation

BAY was last delivered in Semester 2, 2022. Changes were made to BAY in 2025 following suggested changes from a review of the Semester 2, 2022 delivery and are being incorporated into the Semester 2, 2025 delivery. Additional changes have also been made in response to student feedback.

## Acknowledgments

The subject notes were inspired by the above mentioned text books and also the course note prepared by Prof. Lyly Gurrin and Dr Koen Simons in Semester 2, 2022 delivery. We also thank Prof. Armando Teixeira-Pinto (University of Sydney), Prof. John Carlin (University of Melbourne) and Prof. Assoc Prof Lynne Giles (University of Adelaide) for their feedback on the subject notes.