

Final Project

STAT 3303: Bayesian Analysis and Statistical Decision Making

Due: Sunday, April 27, 2025 (submit on Carmen by midnight)

INSTRUCTIONS

This is an exam and should be treated as such. **DO NOT discuss any aspect of this exam with anyone, including MSLC tutors, other than your instructor.** This includes, but is not limited to, not discussing report structure, coding problems, instructor expectations, how long the exam took you, whether you finished it, etc. You are responsible for ensuring that other students do not have access to your exam. You may search online for general coding questions (i.e., “how do I pass through my seed to rjags?”) through websites such as StackExchange. Use of AI for coding or to assist with the write-up is **highly inadvisable** (see below). Any violation of these instructions constitutes academic misconduct and will be reported to the university’s Committee on Academic Misconduct.

PROBLEM STATEMENT

As of December 31, 2024, the total value of all US stocks was over 60 **trillion** dollars.¹ So, predicting the future returns of stock investments is a major financial question. Although investors are commonly cautioned that “past performance is no guarantee of future results”, prior stock returns are commonly used to evaluate whether a company is a good investment.

Although stock returns are continuous, or nearly so, we can approximate the performance of a stock as a weighted coin flip where the stock goes up with probability θ and down with probability $1 - \theta$. (If then, one wishes to model continuous price movements, you can model that as the limit of many coin flips over a short period of time. This method is used, for example, in the derivation of the famous Black-Scholes model of option pricing. For our purposes, simply modeling stock movements as a series of coin flips is sufficient.) Our goal, then, given the results over several periods, is to estimate the true value of θ for each stock given our sample data.

We also want to consider that stock movements are **not** independent. There are factors that affect the entire market, like inflation caused by monetary or trade policy. Further, there are factors that might affect segments of the market, for example durable goods firms may suffer more in an economic downturn than those that produce non-discretionary items like food or medicine. Thus, returns of individual stocks are typically modeled as varying around a common sector mean, which in turn is commonly modeled as varying around a market-wide mean.

You have returns from 30 periods, modeled as 30 coin flips (0 or 1) for 10 stocks each from 5 sectors, for a total of 1500 data points. You may assume for a given stock the returns from one period to the next are iid. The data in the file are as follows:

¹<https://siblisresearch.com/data/us-stock-market-value/>

sector indicates what sector each company belongs to, from 1 to 5.

stock index for the company number within each sector, from 1 to 10.

flip binary indicator where 1 indicates a positive return and 0 a negative return.

ASSIGNMENT:

Write a formal research report as if you were an investment analyst reporting your findings to your supervisor. Your report should begin with a general discussion of the research question, a description of your data (don't forget your exploratory data analysis) including limitations and potential sources of bias. Only then should you move on to describing your model and reaching some conclusions. Be sure to define all variables and interpret model parameters to your audience. You will be graded on the level of professionalism in the report as well as the statistics, so obvious use of AI or any "hallucinations" will be **severely** penalized.

Specific modeling guidelines:

1. You should propose a hierarchical Bayesian model for the probability that each stock has a positive returns in a given period. For your priors, you may use a weakly informative prior.
 2. Specify your model in detail, including conditional independence and prior assumptions (providing just your code is NOT sufficient).
 3. Provide details on model fitting (what were your starting values, how many iterations did your algorithm run, how you diagnosed convergence of the model fitting algorithm, etc.).
 4. Provide interpretations of the results of your statistical analysis in the context of the problem.
 5. You should also give posterior probabilities that each sector is the "best" sector to invest in, as well as posterior probabilities that each stock is "best" within its sector.
-

FORMATTING GUIDELINES:

- Your report should be typed. You should use R Markdown, but DO NOT include any code in the main body of your report.
- Again, carefully proofread and spell check your report. Write in complete sentences and in paragraphs, not bulleted lists.
- Define all mathematical notation in the text of the report.
- Make sure all figures/tables are straightforward to understand, have captions, and are referenced in the text.
- Include commented code in an appendix.

- You may assume that the reader is familiar with Bayesian statistics, but not that they are familiar with the content of STAT 3303. For example, do not refer to specific examples that have been discussed in lecture or homework.

Your report should be **5-6 pages** double-spaced, including figures and tables. Your appendix with code and output does not count toward the six page limit.

Submit your final exam report as a **single PDF file** on Carmen before the deadline.