Bayesian Statistics (STA306)
Final Project

24/05/2024–07/06/2024

1 Topics

Choose one of the following two options (each containing a dataset and a model) to perform Bayesian analysis.

1.1 Option I

The data is a poly-pharmacy dataset which is available from "polypharm" in the R package "aplore3". The set contains data on 500 subjects studied over 7 years. The response is whether the subject is taking drugs from 3 or more different groups. We consider the covariates, Gender $= 1$ if male and 0 if female, Race $= 0$ if subject is white and 1 otherwise, Age, and the following binary indicators for the number of outpatient mental health visits, $\text{MHV}_1 = 1$ if $1 \leq \text{MHV} \leq 5$, $\text{MHV}_2 = 1$ if $6 \leq \text{MHV} \leq 14$ and $\text{MHV}_3 = 1$ if $\text{MHV} \geq 15$. Let INPTMHV $= 0$ if there were no inpatient mental health visits and 1 otherwise. We consider a logistic random intercept model of the form:

$$\begin{aligned}\text{logit}(\mu_{ij}) = \quad & \beta_0 + \beta_1 \text{Gender}_i + \beta_2 \text{Race}_i + \beta_3 \text{Age}_{ij} + \\ & \beta_4 \text{MHV}_{1ij} + \beta_5 \text{MHV}_{2ij} + \beta_6 \text{MHV}_{3ij} + \beta_7 \text{INPTMHV}_{ij} + u_i,\end{aligned}$$

for $i = 1, \ldots, 500$, $j = 1, \ldots, 7$, where $u_i \sim N(0, e^{2\xi})$. The priors can be $\beta_j \sim N(0, \sigma_\beta^2)$ and $\xi \sim N(0, \sigma_\xi^2)$ with $\sigma_\beta^2 = \sigma_\xi^2 = 100$.

1.2 Option II

Here we consider daily observations of the weekday exchange rates of the US Dollar against the British Pound from October 1, 1981, to June 28, 1985. The data is available from the dataset "Garch" in the R package "Ecdat". The number of responses is $n = 945$. We consider the stochastic volatility model widely used in describing financial time series, which is an example of a non-linear state-space model. The mean-corrected return at time t, $\{y_t\}$, are computed from the exchange rates $\{r_t\}$ as

$$y_t = 100 \times \left\{ \log(r_t / r_{t-1}) - \frac{1}{n} \sum_{i=1}^{n} \log(r_i / r_{i-1}) \right\}.$$

We assume y_t are generated from a zero-mean Gaussian distribution with a variance stochastically evolving over time, and the unobserved log volatility b_t is modelled as an AR(1) process with Gaussian disturbances. That is,

$$\begin{aligned}
&y_t \sim N\left(0, \exp(\lambda + \sigma b_t)\right), \text{ for } t = 1, \ldots, n, \\
&b_1 \sim N(0, 1/(1-\phi^2)), \\
&b_{t+1} \sim N(\phi b_t, 1), \text{ for } t = 2, \ldots, n.
\end{aligned}$$

where $\lambda \in \mathbb{R}$, $\sigma > 0$ and $0 < \phi < 1$, b_t is assumed to follow a stationary process with b_1 drawn from the stationary distribution. We transform the constrained parameters to $\mathbb{R}$ by letting $\sigma = \exp(\alpha)$ and $\phi = \frac{\exp(\psi)}{\exp(\psi)+1}$, where $\alpha, \psi \in \mathbb{R}$. The priors can be $\alpha \sim N(0, \sigma_\alpha^2)$, $\lambda \sim N(0, \sigma_\lambda^2)$ and $\psi \sim N(0, \sigma_\psi^2)$ with $\sigma_\alpha^2 = \sigma_\lambda^2 = \sigma_\psi^2 = 100$.

2 Guidelines on Data Analysis

What you are required to do is to derive the posterior distribution of the model that you choose, use at least one computational method taught in this course to infer the posterior distribution and summarize your inference. Optional further analysis can include comparing different computational methods, model checking, model selection, sensitivity analysis with respect to different model and prior choices, etc.

3 Guidelines on Presentation

The presentation will take place in-class on 07/06/2024. Each group will have 20 mins in total with 15mins for presenting and 5mins for Q&A. No report is demanded for this project.