Stat 8678 - SAS Programming & Data Analysis

Chi-Kuang Yeh

2025-10-27

Table of contents

Preface

Description

This course covers programming using the SAS statistical software package, and it provides an introduction to data analysis stressing the implementation using SAS.

Topics include two main parts:

1) **SAS Programming**: data management and manipulation, basic procedures, macro programming;
2) **Data Analysis**: descriptive statistical analysis, one- and two-sample inference, basic categorical data analysis, regression analysis, and other selected topics.

Prerequisites

MATH 4544/6544, or equivalent.

Instructor

Chi-Kuang Yeh, I am an Assistant Professor in the Department of Mathematics and Statistics, Georgia State University.

- Office: Suite 1407, 25 Park Place.
- Email: cyeh@gsu.edu.

Office Hour

TBA

Grade Distribution

- TBA

Assignment

☐ TBA

Midterm

☐ TBA

Topics and Corresponding Lectures

Those chapters are based on the lecture notes. This part will be updated frequently.

Topic	Lecture Covered
Introduction to SAS and modules	1–

Recommended Textbooks

- Statistics 480: Introduction to SAS, The Pennsylvania State University.
- SAS Training, SAS Institute.
- SAS Resources, University of California, Los Angeles.

1 Introduction

The posterior distribution is obtained from the prior distribution and sampling model via *Bayes' rule*:

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{\int_{\Theta} p(y \mid \tilde{\theta})p(\tilde{\theta})d\tilde{\theta}}.$$

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
1 + 1
```

```
[1] 2
```

1.1 Why Bayesian?

2 Course Topics and Schedule

Week	Topics	Key Concepts / Readings	Computing Focus
1	Introduction to Bayesian Thinking	Bayesian vs. Frequentist paradigms; Prior, likelihood, posterior	Review of R basics and reproducible workflows
2	Bayesian Inference for Simple Models	Conjugate priors, Beta-Binomial, Normal-Normal, Poisson-Gamma	Simulating posteriors, visualization
3	Prior Elicitation and Sensitivity	Informative vs. noninformative priors, Jeffreys prior	Prior sensitivity plots
4	Monte Carlo Integration	Law of large numbers, sampling-based inference	Random sampling and Monte Carlo approximation
5	Markov Chain Monte Carlo (MCMC)	Metropolis-Hastings, Gibbs sampler	Implementing MCMC in R
6	Convergence Diagnostics	Trace plots, autocorrelation, Gelman–Rubin statistic	`coda`, `rstan`, and `bayesplot` packages
7	Hierarchical Bayesian Models	Partial pooling, shrinkage, multilevel structures	`rstanarm` / `brms`
8	Midterm Project: Bayesian Linear Regression	Posterior inference for regression, model selection	`brms`, `rstanarm`, custom Gibbs samplers
9	Bayesian Model Comparison	Bayes factors, BIC, DIC, WAIC, LOO	Practical comparison via cross-validation
10	Model Checking and Diagnostics	Posterior predictive checks, residual analysis	`pp_check` in `brms`
11	Advanced Computation	Hamiltonian Monte Carlo (HMC), Variational Inference	Using `Stan` and `CmdStanR`
12	Bayesian Decision Theory	Utility functions, decision rules, loss minimization	Simple decision problems in R

Week	Topics	Key Concepts / Readings	Computing Focus
13	Modern Bayesian Methods	Approximate Bayesian computation (ABC), Bayesian neural networks	Examples via `rstan` or `tensorflow-probability`
14	Student Project Presentations	Applications and case studies	Full workflow demonstration in R

Interesting Article:

- Goligher, E.C., Harhay, M.O. (2023). What Is the Point of Bayesian Analysis?, American Journal of Respiratory and Critical Care Medicine, 209, 485–487.

References

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.