Statistical Computing and Simulation

Spring 2024

Assignment 5, Due June 11/2024

1. First, simulate 100 observations from a mixed distribution of N(−2,1) and N(2,1), each with probability 0.5. Then, use at least 3 density estimating methods to smooth the observations. You need to specify the parameters in the smoothing methods, and compare the results.

2. Crime data are available in many countries and we can use them the explore whether there are hot spots and/or peak seasons. Explore the reported cases of stolen motorcycles provided by Taipei City and evaluate which month(s) has the highest reported cases of stolen motorcycles (via density estimation methods). (Bonus: Explore if there are hot spots in stolen motorcycles.)

3. Let x be 100 equally spaced points on $[0,2\pi]$ and generate random sample $y_i = \sin x_i + \varepsilon_i$ with $\varepsilon_i \sim N(0,0.09)$. Apply at least 3 linear smoothers and compare the differences, with respect to mean squares error (i.e., bias^2 and variance) from 1,000 simulation runs.

4. Use “MCMCregress” in the module MCMCpack to obtain MCMC estimation of regression analysis. Duplicate the analysis in the lecture notes and apply the MCMC on the “bikes.csv” data. Compare your results with the regular simple linear regression.

5. We will apply Bayesian computing (Normal + Normal → Normal) to construct Taiwan’s life tables, use Taiwan’s mortality data in 2020. Try different prior distributions and compare your analysis results to the official abridged life tables. For example, you may treat the official life table as the prior. Also, you need to specify the parameters used.

6. You can use “MCMClogit” in the module MCMCpack to obtain MCMC estimation of logistic regression analysis. Conduct the logistic regression via the “glm” and MCMC, using the data “birthwt”, and comment on the results you found. (Note: data("birthwt", package = "MASS"))