

PHC6068 Biostatistical Computing: Course Project

(Deadline: December 12, 2016)

Instruction: (i) Choose any three questions from the first four to complete; and (ii) complete Question 5.

1. (Gelfand and Smith, 1990, JASA) For a population of 30 rats, the weight y_{ij} of rat i is observed at age x_j and is associated with the model

$$Y_{ij} \sim N(\alpha_i + \beta_i(x_j - \bar{x}), \sigma_c^2).$$

- (i) Give the Gibbs sampler associated with this model and the prior

$$\alpha_i \sim N(\alpha_c, \sigma_\alpha^2), \quad \beta_i \sim N(\beta_c, \sigma_\beta^2),$$

with almost flat hyperpriors

$$\alpha_c, \beta_c \sim N(0, 1000), \quad \sigma_c^{-2}, \sigma_\alpha^{-2}, \sigma_\beta^{-2} \sim \text{Gamma}(0.01, 0.01).$$

- (ii) Simulate 10 datasets from the model and analyze the simulated data using the Gibbs sampler.
2. (Yu, Liang, Chatterjee and Ciampa, 2011, Biostatistics) Implement the stochastic approximation Monte Carlo (SAMC) algorithm for evaluating p -values of the two-sample t -test using a simulated dataset.
3. (Liang, Song and Yu, 2013, JASA) Implement the stochastic approximation Monte Carlo (SAMC) algorithm for variable selection from high-dimensional generalized linear models.
4. (Liang, Cheng and Lin, 2014, JASA) Given a set of cities, the traveling salesman problem (TSP) is to find the shortest tour which goes through each city once and only once. Implement the simulated annealing and simulated stochastic approximation annealing algorithms for a traveling salesman problem for which 100 cities are uniformly distributed on a square region of length 100 miles. Compare the performance of the two algorithms and find the average length of the shortest tour (over 100 datasets).
5. Write a short course paper (around 15 pages): Using Monte Carlo methods to solve a problem in your study or research.