

# Computational Statistics

## Chapter 5: Monte Carlo Markov Chains

1. The dataset `coal.dat` is a time series of the numbers of disasters at coal mines annually between 1851 and 1962. We assume that the number of coal mining disasters follows a Poisson distribution  $\mathcal{P}(\theta_1)$  until some unknown change point  $k$ , and then a Poisson distribution  $\mathcal{P}(\theta_2)$ . The prior distributions are  $\pi(\theta_1) \sim G(\alpha_{01}, \beta_{01})$ ,  $\pi(\theta_2) \sim G(\alpha_{02}, \beta_{02})$  and  $k \sim \mathcal{U}(0, 1)$ , where  $G(\alpha, \beta)$  denotes the Gamma distribution with shape parameter  $\alpha$  and rate  $\beta$ . We want to approximate the posterior distributions of  $\theta_1$ ,  $\theta_2$  and  $k$  using the Gibbs sampling algorithm.

- (a) Plot the data.
- (b) Show that

$$\pi(\theta_1, \theta_2, k | \mathbf{x}) \propto$$

$$\theta_1^{\alpha_{01}-1} e^{-\beta_{01}\theta_1} \theta_2^{\alpha_{02}-1} e^{-\beta_{02}\theta_2} \prod_{i=1}^k e^{-\theta_1} \theta_1^{x_i} \prod_{i=k+1}^n e^{-\theta_2} \theta_2^{x_i}$$

and

$$\theta_1 | \mathbf{x}, k \sim G(\alpha_{01} + \sum_{i=1}^k x_i, \beta_{01} + k)$$

$$\theta_2 | \mathbf{x}, k \sim G(\alpha_{02} + \sum_{i=k+1}^n x_i, \beta_{01} + n - k)$$

$$\pi(k | y, \theta_1, \theta_2) \propto e^{k(\theta_2 - \theta_1)} \left( \frac{\theta_1}{\theta_2} \right)^{\sum_{i=1}^k y_i}.$$

- (c) Construct a Gibbs algorithm for this problem. Apply it to the data with  $\alpha_{01} = \alpha_{02} = 0.5$  and  $\beta_{01} = \beta_{02} = 1$ .
- (d) For each parameter, plot the sample path, the histogram of simulated values and the autocorrelation function (use the function `acf`).
- (e) Estimate the conditional expectations of each parameter with the associated standard error using the batch means method.

2. Construct a random walk MH sampler to generate a sample of 10,000 observations from the Laplace distribution,

$$f(x) = \frac{1}{2}e^{-|x|}, \quad -\infty < x < +\infty.$$

Use  $u \sim \mathcal{N}(0, \sigma^2)$  to generate proposals  $y = x^{(t-1)} + u$ . Compare the mean, variances, and autocorrelations of the generated series for various values of  $\sigma^2$ .

3. Consider the model

$$y_i = \beta x_i + u_i, \quad u_i \sim \mathcal{N}(0, 1), \quad i = 1, \dots, n,$$

with the gamma prior distribution  $\beta \sim G(2, 1)$ ,  $\beta > 0$ .

- (a) Verify that the posterior distribution is

$$\pi(\beta | y_1, \dots, y_n) \propto \beta \exp(-\beta) \exp \left[ -\frac{1}{2} \sum_{i=1}^n (y_i - \beta x_i)^2 \right] \mathbb{1}_{(0, +\infty)}(\beta).$$

Note that this distribution does not have a standard form.

- (b) Construct an MH algorithm to sample from this distribution with an independence kernel, where the kernel is the prior distribution.
- (c) Generate a dataset by choosing  $n = 50$ ,  $x_i$  from  $\mathcal{N}(0, 1)$  and a value of  $\beta$  from its prior distribution. Apply the MH algorithm to this dataset.
- (d) Plot the sample path, the histogram of simulated values and the autocorrelation function.
- (e) Compute the posterior expectation of  $\beta$  and its standard error.