Computational Statistics
Chapter 5: Monte Carlo Markov Chains

1. 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 $\epsilon \sim \mathcal{N}(0, \sigma^2)$ to generate proposals $x^* = x^{(t-1)} + \epsilon$. Draw the sample paths and autocorrelation functions (use function `acf`) for various values of σ^2. After you have found a chain with good mixing, draw a histogram of the sampled values, together with the target density.

2. 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

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

   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 β 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 (use function `acf`).

   (e) Compute the posterior expectation of β and its simulation standard error.

3. 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 $f(\theta_1) \sim G(\alpha_{01}, \beta_{01})$, $f(\theta_2) \sim G(\alpha_{02}, \beta_{02})$, and $k \sim \mathcal{U}(n)$, a discrete uniform distribution in $\{1, \dots, n\}$, where n is the number of years. We want to approximate the posterior distributions of θ_1, θ_2 and k using the Gibbs sampling algorithm.

(a) Plot the data.

(b) Show that

$$f(\theta_1, \theta_2, k|\mathbf{x}) \propto$$
$$\theta_1^{\alpha_{01}-1} e^{-\beta_{01}\theta_1} \theta_2^{\alpha_{02}-1} e^{-\beta_{02}\theta_2} \prod_{i=1}^{k} e^{-\theta_1} \theta_1^{x_i} \prod_{i=k+1}^{n} e^{-\theta_2} \theta_2^{x_i}$$

and

$$\theta_1|\mathbf{x}, k \sim G(\alpha_{01} + \sum_{i=1}^{k} x_i, \beta_{01} + k)$$
$$\theta_2|\mathbf{x}, k \sim G(\alpha_{02} + \sum_{i=k+1}^{n} x_i, \beta_{02} + n - k)$$
$$f(k|\mathbf{x}, \theta_1, \theta_2) \propto e^{k(\theta_2-\theta_1)} \left(\frac{\theta_1}{\theta_2}\right)^{\sum_{i=1}^{k} x_i}.$$

(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 function `acf`).

(e) Estimate the conditional expectations of each parameter with the associated simulation standard error using the batch means method.

4. The file `investment.txt` contains 15 yearly observations of U.S. investment data for the period 1968-1982. The variables are

- `Year` = Date,
- `GNP` = Nominal GNP,
- `Invest` = Nominal Investment,
- `CPI` = Consumer price index,
- `Interest` = Interest rate,
- `Inflation` = rate of inflation computed as the percentage change in the CPI.

(a) We consider the linear regression model $\mathbf{y} \sim N_n(\mathbf{X}\beta, \sigma^2 I_n)$ with the dependent variable `Invest/(10*CPI)` and, as covariates, the time trend (a vector of integers from 1 to 15), `GNP/(10*CPI)`, `Interest` and `Inflation`. Plot the data and compute the least-squares estimate of the parameters.

(b) We assume a normal prior for β: $\beta \sim N(\beta_0, \mathbf{B}_0)$ and an inverse Gamma prior for σ^2: $\sigma^2 \sim IG(\alpha_0/2, \delta_0/2)$. The prior density of σ^2 is, thus,

$$f(\sigma^2|\alpha_0, \delta_0) = \frac{(\delta_0/2)^{\alpha_0/2}}{\Gamma(\alpha_0/2)} \frac{1}{(\sigma^2)^{\alpha_0/2+1}} e^{-\delta_0/(2\sigma^2)} I(\sigma^2 > 0).$$

Show that

$$\beta|\sigma^2, \mathbf{y} \sim N(\overline{\beta}, \mathbf{B}_1) \quad \text{and} \quad \sigma^2|\beta, \mathbf{y} \sim IG(\alpha_1/2, \delta_1/2),$$

with

$$\begin{aligned}
\mathbf{B}_1 &= (\sigma^{-2}\mathbf{X}^T\mathbf{X} + \mathbf{B}_0^{-1})^{-1} \\
\overline{\beta} &= \mathbf{B}_1(\sigma^{-2}\mathbf{X}^T\mathbf{y} + \mathbf{B}_0^{-1}\beta_0) \\
\alpha_1 &= \alpha_0 + n \\
\delta_1 &= \delta_0 + (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta).
\end{aligned}$$

(c) Program a Gibbs sampler to simulate the posterior distribution of parameters $\boldsymbol{\beta}$ and σ. (To simulate from the multivariate normal distribution, you can use function `mvrnorm` in package `MASS`).

(d) Use the Gibbs sampler with different parameters of the prior probability distributions. To determine the parameters of the inverse Gamma distribution, you may use the following equations that relate the shape parameter α and the rate parameter δ to the mean μ and the variance V of the inverse Gamma distribution:

$$\alpha = \mu^2/V + 2, \quad \delta = \mu(\mu^2/V + 1).$$

(e) Write a function that generates samples from the posterior predictive distribution of $y_0 = x_0^T\beta + \sigma u$, with $u \sim N(0, 1)$ and from the posterior distribution of $E(y_0) = x_0^T\beta$.