Lab 8; MCMC, Metropolis-Hastings & Gibbs

Dietrich von Rosen,
dietrich.von.rosen@slu.se

Exercise 1: Generate 100 observations from a $N(32, 2)$, $Poisson(4)$, Multinomial$(50, 0.3, 0.5, 0.1)$ and $\beta(3.2)$-distribution.

Exercise 2: Let the mixture distribution considered in this exercise be given by

$$\delta N(7, 0.5^2) + (1-\delta)N(10, 0.5^2). \tag{0.1}$$

The goal is to investigate the role of the proposal distribution in a Metropolis-Hastings algorithm designed to simulate from the posterior distribution of the parameter δ. In part (a) you are asked to simulate data from a distribution with δ known. For parts (b)-(d), assume δ is unknown with a Uniform(0,1) prior distribution for δ. For parts (b)-(d), provide an appropriate plot and a table summarizing the output of the algorithm. To facilitate comparisons, use the same number of iterations, random seed, starting values, and burn-in period for all implementations of the algorithm (you may skip (c) and (d))

(a) Simulate 200 realizations from the mixture distribution in (0.1) with $\delta = 0.7$. Draw a histogram of these data.

(b) Implement an independence chain MCMC procedure to simulate from the posterior distribution of δ, using your data from part (a).

(c) Implement a random walk chain with $\delta* = \delta^t + \epsilon, \quad \epsilon \sim Uniform(-1, 1)$.

(d) Compare the estimates and convergence behavior of the three algorithms.

Exercise 3: Stream insects are an effective indicator for monitoring stream ecology. Imagine that at many sites along a stream insects are collected and classified into c classes. The task will be to simulate a site and say something about the distribution of the total number of insects.

Let $Y_1, Y_2, \ldots, Y_c$ denote the counts of insects in each of the c classes and let N denote the random total number of insects collected. The probability that an insect is classified in each class varies randomly from site to site as does the total number of insects collected at each site. Let $P_1, P_2, \ldots, P_c$

denote the class probabilities which depend on known site specific parameters $\alpha_1, \alpha_2, \ldots, \alpha_3$. Let N depend on a site specific known parameter λ. Suppose we are interested to determine a critical level for the total number of insects, i.e. choose a value in the upper tail of the distribution for $\sum_i Y_i$.

Let $c = 3$, choose $\alpha_1 = 10, \alpha_2 = 30, \alpha_3 = 40$ and $\lambda = 80$. The task is to apply Gibbs when data has either been obtained as in Case 1 or as in Case 2.

Case 1: Data; $y_1 = 12, y_2 = 25, y_3 = 75$

Case 2: Data; $y_1 = 30, y_2 = 20, y_3 = 150$

Hints

For $c = 3$ we can establish the following model

$$\begin{aligned} Y_1, Y_2, Y_3 | N = n, P_1 = p_1, P_2 = p_2, P_3 = p_3 &\sim Multinomial(n; p_1, p_2, p_3) \\ P_1, P_2, P_2 &\sim Dirichlet(\alpha_1, \alpha_2, \alpha_3) \\ N &\sim Poisson(\lambda) \end{aligned}$$

Note that the model is overparameterized: $Y_1 + Y_2 + Y_3 = N$ and $P_1 + P_2 + P_3 = 1$. The Gibbs sampling scheme is given by (try to motivate)

$$\begin{aligned} & Y_1, Y_2, Y_3 | N = n, P_1 = p_1, P_2 = p_2, P_3 = p_3 \sim Multinomial(n; p_1, p_2, p_3) \\ & P_1, P_2, P_2 | N = n, Y_1 = y_1, Y_2 = y_2, Y_3 = y_3 \\ & \qquad \sim Dirichlet(y_1 + \alpha_1, y_2 + \alpha_2, n - y_1 - y_2 + \alpha_3) \\ & N | P_1 = p_1, P_2 = p_2, P_3 = p_3, Y_1 = y_1, Y_2 = y_2, Y_3 = y_3 \\ & \qquad \sim Poisson(\lambda(1 - p_1 - p_2)) \end{aligned}$$

To be more precise we can specify the Gibbs algorithm as follows: one complete cycle is given by

$$\begin{aligned} & Y_1^{t+1} | N^t = n^t, P_1^t = p_1^t, P_2^t = p_2^t, P_3^t = p_3^t, Y_2^t = y_2^t \sim Bin(n^t - y_2^t; p_1^t/(1 - p_2^t)), \\ & Y_2^{t+1} | N^t = n^t, P_1^t = p_1^t, P_2^t = p_2^t, P_3^t = p_3^t, Y_1^{t+1} = y_1^{t+1} \sim Bin(n^t - y_1^{t+1}; p_2^t/(1 - p_1^t)), \\ & P_1^{t+1}/(1 - p_2^t) | N^t = n^t, P_2^t = p_2^t, P_3^t = p_3^t, Y_1^{t+1} = y_1^{t+1}, Y_2^{t+1} = y_2^{t+1} \\ & \qquad \sim \beta(y_1^{t+1} + \alpha_1, n^t - y_1^{t+1} - y_2^{t+1} + \alpha_3), \\ & P_2^{t+1}/(1 - p_1^{t+1}) | N^t = n^t, P_1^{t+1} = p_1^{t+1}, P_3^t = p_3^t, Y_1^{t+1} = y_1^{t+1}, Y_2^{t+1} = y_2^{t+1} \\ & \qquad \sim \beta(y_2^{t+1} + \alpha_1, n^t - y_1^{t+1} - y_2^{t+1} + \alpha_3), \\ & N^{t+1} | P_1^{t+1} = p_1^{t+1}, P_2^{t+1} = p_2^{t+1}, Y_1^{t+1} = y_1^{t+1}, Y_2^{t+1} = y_2^{t+1}, \sim Poisson(\lambda(1 - p_1^{t+1} - p_2^{t+1})). \end{aligned}$$