problem set no. 5 — due on monday 12/10 before noon

Problem 1. Gibbs Sampling in the Bivariate Normal

Consider a single observation (y_1, y_2) from a bivariate normally distributed population with unkown mean $\theta = (\theta_1, \theta_2)$ and known covariance $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. With a flat prior on θ the posterior distribution is:

$$(\theta_1, \theta_2) \sim N((y_1, y_2), \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$$

Although it is simple to sample from a multivariate distribution in practice, we will use it to demonstrate properties of the Gibbs sampler.

a) Derive the conditional distributions $\theta_1|\theta_2, y$ and $\theta_2|\theta_1, y$.

b) Assume $(y_1, y_2) = (0.5, 0.9)$. For each of $\rho = 0.3, 0.6$, and 0.9 use Gibbs sampling to draw 1000 samples from the posterior. Initialize (θ_1, θ_2) to (y_1, y_2). In addition, for each condition:

1. Plot the samples in a scatter plot.
2. Plot the traceplot for θ_1.
3. Compute the lag-10 autocorrelation and effective sample size of θ_1 (you may find the effectiveSize and acf functions in the "coda" package helpful).

c) Comment on your results to part b). In particular, what is the relationship between ρ and the efficiency of the sampler. Why does this make sense?

Problem 2. Metropolis Sampling in the Bivariate Normal

Consider the same setup as Problem 1, where a single observation (y_1, y_2) is obtained from a bivariate normally distributed population with unknown mean $\theta = (\theta_1, \theta_2)$ and known covariance $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. With a flat prior on θ the posterior distribution is:

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Although it is simple to sample from a multivariate distribution in practice, we will use it to demonstrate properties of the Metropolis sampler.

Assume that we observe $(y_1, y_2) = (1, 1)$ and $\rho = 0.9$. Implement a Metropolis algorithm using the proposal distribution

$$\begin{pmatrix} \theta_1^* \\ \theta_2^* \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1^{(s)} \\ \theta_2^{(s)} \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

to draw $N = 10,000$ samples from the posterior. For each value of $\sigma^2 = 0.1, 0.2, 0.5, 1, 2, 4, 6, 9$:

a) Plot the traceplot and acf for θ_1. Observe and comment on the chain's mixing behavior. (To save trees, submit your plots for σ^2 values $0.1, 2, 9$ only).

b) Compute the effective sample size of the θ_1 chain, and plot them as a function of σ^2.

c) Compute the acceptance ratio of the θ_1 chain, and plot them as a function of σ^2.

d) Based on your results from part a), b) and c), comment on the choice of σ^2 in terms of the efficiency of the sampler.

e) Now consider a different proposal distribution:

$$\begin{pmatrix} \theta_1^* \\ \theta_2^* \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1^{(s)} \\ \theta_2^{(s)} \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

Repeat part b) and c) using this new proposal distribution, and compare your new results to the old. What does it tell you in terms of the efficiency of the sampler?