CHAPTERS IN SCIENTIFIC COMPUTING (67714)

EXERCISE 1 - **MCMC**

(1) **Periodicity and Reducibility.** Consider the following transition matrix

$$T = \begin{pmatrix} \epsilon & 1-\epsilon \\ 1-\epsilon & \epsilon \end{pmatrix}.$$

Show that for $\epsilon = 0$ there is no unique stationary distribution, where as for $\epsilon > 0$ there is. Show that the following transition matrix

$$T = \begin{pmatrix} 1 & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & 1 \end{pmatrix},$$

is also not ergodic. Solve both problems by computing the eigenvalues of the matrices. Explain the results in terms of periodicity and reducibility.

(2) **Gibbs Sampling.** In class we showed that this process obeys the detailed-balance (DB) condition. Show explicitly (and not using the DB condition) that it maintains the distribution P, that defines the process, invariant.

(3) **Local Minima and Mixing Time.** Consider the following distribution, resulting from a double well energy potential

$$P(x) \propto e^{-(x-1)^2/T} + e^{-(x+1)^2/T},$$

where T is some temperature parameter. This function has two maxima (at $x = -1$ and $x = 1$). In this problem we can assume a continuous variable x or x that lies on a grid with spacing smaller than one. We want to generate samples from

Use a metropolis sampling with proposal distribution that suggests moving to the right and the left with equal probability; in case you choose a discrete implementation, one grid point to the left and right with probabilities half each, or in the continuous setting, suggests $x^n + \xi$ where ξ is a Gaussian random variable with small variance and zero mean. Show the at low temperatures the observable function

$$\mu_1 = \int xp(x)dx,$$

converges slowly and that this is not the case for

$$\mu_2 = \int x^2 p(x)dx.$$

(replace the integral by a sum for the discrete case). Explain why we get these results. Suggest another proposal move to cure this slow convergence. Measure the convergence rate by plotting the intermediate averages of the observables and by computing the autocorrelation of x^n.

(4) **Generalizing Metropolis.** The idea behind metropolis accept/reject step is to correct the transition distribution such that it obeys the detailed balance condition. Show that the following generalization also achieves the detailed balance condition:

$$x^{n+1} = \begin{cases} x^* & \text{if } P(x^*) < P(x^n) \text{ and } u < \alpha(x^n, x^*)\frac{P(x^*)}{P(x^n)}, \quad q \sim U([0,1]) \\ x^* & \text{if } P(x^*) > P(x^n) \text{ and } u < \alpha(x^n, x^*), \quad q \sim U([0,1]) \\ x^n & \text{otherwise} \end{cases}$$

where $0 \le \alpha(x,y) = \alpha(y,x) \le 1$ and $\alpha(x,y)P(y)/P(x) \le 1$. The Metropolis correction is obtained when $\alpha(x,y) = 1$. Show that also the following correction, known as Boltzmann acceptance,

$$x^{n+1} = \begin{cases} x^* & \text{if } u < \frac{P(x^*)}{P(x^*)+P(x^n)}, \quad q \sim U([0,1]) \\ x^n & \text{otherwise, i.e., } u < \frac{P(x^n)}{P(x^*)+P(x^n)} \end{cases}$$

obeys the detailed-balance condition by showing it corresponds to an admissible $\alpha(x, y)$ function.

(5) **Ising Model.** This model in 2D is given by

$$E(\mathbf{x}) = -\sum_{i,j} x_{i,j}\big(x_{i-1,j-1} + x_{i+1,j-1} + x_{i-1,j+1} + x_{i+1,j+1}\big),$$

and

$$P(\mathbf{x}) = e^{-E(\mathbf{x})/T},$$

where $\mathbf{x} \in \{-1, 1\}^{N\times N}$. We set $T = 2.27$ (critical temperature). Sample this model using the following three methods for computing the observable defined at the bottom.

(a) Importance sampling (this is not MCMC but an independent sampling scheme). Come up with some distribution, over $\mathbf{x} \in \{-1, 1\}^{N\times N}$ that you know to sample from and use it to compute averages with respect to P. What is the ratio between your distribution and the true distribution, does the logarithm of this ratio has a dependency on N ?

(b) Use Metropolis sampling that scans the grid and proposes a change in a single spin each step.

(c) Use the Wolff method to sample P.

Compute the correlation times when running the two MCMC options and plot the weights you get by the first sampling scheme. Plot teh intermediate averages of each scheme as well. Which strategy is most effective for estimating the magnetism observable,

$$\mu_0 = \sum_{\mathbf{x}} \Big(\sum_{i,j} x_{i,j}\Big) P(\mathbf{x}).$$

given that you know what it should be.