Stat202C Project no. 2 (15 points)

Due date: May 17 Thursday 11:50pm on CCLE

I, Exact sampling of the Ising/Potts model with coupled Markov Chains.

We consider Ising model in an n x n lattice (n=64, or you may try n= 128 if you have a fast computer) with 4-nearest neighbor. The state X is a binary image defined on the lattice and the variable X_s at each site or pixel s takes value in $\{0, 1\}$. The model is

$$\pi(X) = \frac{1}{Z}\exp\{\beta \sum_{<s,t>} 1(X_s = X_t)\}$$

We simulate two Markov Chains with the Gibbs sampler:

- MC1 starts with all sites being 1 (call it the white chain) and its state is denoted by X^1;
- MC2 starts with all sites being 0 (call it the black chain) and its state is denoted by X^2.

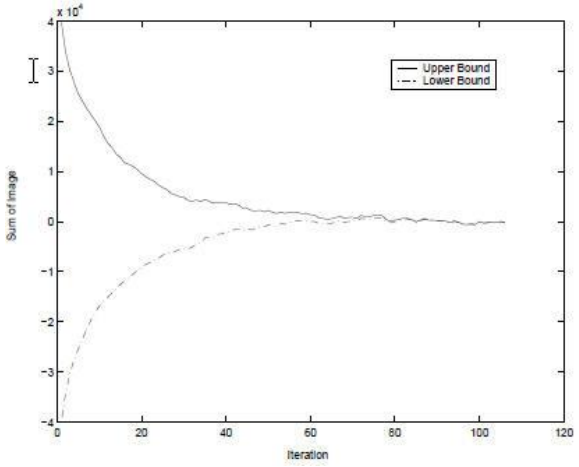
At each step, the Gibbs sampler picks up a site s in both images, and calculates the conditional probabilities, which only depends on its 4 nearest neighbor denoted by ∂s.

$$\pi(X_s^1 | X_{\partial s}^1) \quad \text{and} \quad \pi(X_s^2 | X_{\partial s}^2)$$

It updates the variables X_s^1 and X_s^2 according to the above two conditional probabilities, and shares the same random number r = rand[0,1]. The two Markov Chains are said to be "*coupled*".

1, Prove that $X_s^1 \geq X_s^2, \forall s$ in any time. That is, the white chain is always above the black chain.

2, When the two chains meet each other, i.e. $X_s^1 = X_s^2 \ \forall s$ after many sweeps, they are said to "coalesce". They will stay in the same state forever as they are driven by the same random number at each step. We denote the coalescence time (number of sweeps) by τ. The images after time τ are said to be *exact samples* from the Ising model.



Plot the two chain states (using their total sum $\sum_s X_s^1$ and $\sum_s X_s^2$) over the sweeps as it is shown above and show the image when the two chains coalesce.

Try values $\beta = 0.5,\ 0.65,\ 0.75,\ 0.83,\ 0.84,\ 0.85,\ 0.9,\ 1.0.$

3, Plot the curve of τ versus β (using the parameters above) to see a critical slowing-down around 0.84, which the physicists call "phase transition".

II. Hamiltonian Monte Carlo and Langevin Monte Carlo

This exercise investigates the properties of HMC and LMC by sampling from 2D distributions.

II.1: Sampling from a Non-Isotropic Gaussian

Consider the target distribution $(X, Y) \sim N(0, \phi)$ where $\Phi = \begin{pmatrix} 1 & 0.9998 \\ 0.9998 & 1 \end{pmatrix}$

a) What is the energy function of the target distribution? Make a contour plot of the energy function.
b) Suppose you are sampling from target distribution with the identity matrix I_2 as the momentum covariance. What is the max step size ε^* and minimum number of Leapfrog steps L^* needed to obtain an independent sample from the target distribution in a single HMC iteration?
c) What is the ideal choice of the momentum covariance matrix Σ_{ideal}? What is the max step size ε^*_{ideal} and minimum number of Leapfrog steps L^*_{ideal} needed to obtain independent samples when using the ideal momentum covariance?
d) For each of the methods listed below, start a chain from the state (X, Y) = (0, -10), run 1,000*K burn-in iterations, and 10,000*K sampling iterations from the target distribution (K is needed for fair comparison between methods and is given). For each method, visualize the burn-in path, plot the values of the X and Y coordinates over iterations of the sampling phase, and calculate the ESS of the final X and Y samples. For the plots and ESS calculation, use one sample point per K iterations. Comment on differences between the results. ε^* and L^* refer to the answers from b).
    a. Direct sampling from $N(0, \phi)$, K = 1
    1. Metropolis-Hastings with Gaussian Proposal $N(0, (\varepsilon^*)^2)$, $K = L^*$
    2. Random walk: HMC with $p \sim N(0, I_2)$, $\varepsilon = \varepsilon^*$, $L = L^*$, K = 1
    3. HMC with $p \sim N(0, I_2)$, $\varepsilon = \varepsilon^*$, $L = L^*/2$, K = 1
    4. LMC with $p \sim N(0, I_2)$, $\varepsilon = \varepsilon^*$, $K = L^*$
    5. HMC with $p \sim N(0, \Sigma_{ideal})$, $\varepsilon = \varepsilon^*_{ideal}$, $L = L^*_{ideal}$, K = 1 (answers from part c above)
    6. LMC with $p \sim N(0, \Sigma_{ideal})$, $\varepsilon = \varepsilon^*_{ideal}$, $K = L^*_{ideal}$

II.2: Sampling from a "Banana" Distribution

Consider the posterior distribution of $\theta = (\theta_1, \theta_2)$ with prior $\theta \sim N(0, I_2)$ and $Y \mid \theta \sim N(\theta_1 + (\theta_2)^2, 2)$. Dataset with 100 samples of Y are provided on CCLE.

a) What is the energy function of the posterior density $P(\theta|Y)$? Make a contour plot of the energy function.
b) In new settings, the step size ε^* can be found by tuning the acceptance rate. For a grid of ε, run 2000 LMC iterations with momentum covariance I_2 starting from the origin. Choose a step size with rejection rates between 10%-35%, and report the value.
c) For each of the methods listed below, start a chain from the state $(\theta_1, \theta_2) = (0, 0)$, run 1,000*K burn-in iterations, and 10,000*K sampling iterations from the target distribution (K is needed for fair comparison between methods and is given). For each method, plot the values of the θ_1 and θ_2 coordinates over iterations of the sampling phase, and calculate the ESS of the final θ_1 and θ_2 samples. For the plots and ESS calculation, use one sample point per K iterations. For the HMC methods 4 and 5, visualize the Leapfrog steps of an accepted path and a rejected path. Comment on differences between the results. ε^* refers to the value from b above).
    1. Metropolis-Hastings with Gaussian Proposal $N(0, (\varepsilon^*)^2)$, K = 25
    2. Random walk: LMC with $p \sim N(0, I_2)$, $\varepsilon = \varepsilon^*$, K = 1
    3. LMC with $p \sim N(0, I_2)$, $\varepsilon = \varepsilon^*$, K = 25
    4. HMC with $p \sim N(0, I_2)$, $\varepsilon = \varepsilon^*$, L = 5, K = 1
    5. HMC with $p \sim N(0, I_2)$, $\varepsilon = \varepsilon^*$, L = 25, K = 1.