Markov Chains and Algorithmic Applications

Fall 2023

Mini project: Signal recovery using MCMC

In this project, we study the problem of recovering a signal from noisy observations using Markov Chain Monte Carlo (MCMC) techniques. Let $X \in \mathbb{R}^{m \times d}$ be a random sensing matrix with its entries sampled i.i.d. from $\mathcal{N}(0,1)$. Let $\xi \in \mathbb{R}^m$ be a noise vector sampled from $\mathcal{N}\left(0, \sigma^2 I_m\right)$ and independent of X. Let $\Theta \subseteq \mathbb{R}^d$ be a finite set. Let θ be sampled from Θ uniformly at random and independent of (X, ξ). The measurement y is generated as:

$$y = X\theta + \xi. \tag{1}$$

We want to recover the unknown vector θ using MCMC given the observations (X, y). We are interested in the regime where d is large. We recover θ by finding the maximum likelihood estimate (MLE). In the given model, MLE is given by

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \text{Prob}\{y|\theta, X\}. \tag{2}$$

1 Part 1: Questions

1.1 Question 1: Optimizing over binary hypercube

For this question, assume $\Theta = \{0,1\}^d$.

1. If $\sigma = 0$, what is the minimum number of measurements m required to recover θ with probability 1? Give an algorithm for doing so and comment on its computational complexity with respect to d.
   Answer: When $\sigma = 0$, one sample is sufficient to recover θ, since with probability 1 there would be only a single $\theta \in \{0,1\}^d$ that satisfy $y = X\theta$. To recover θ, we can do an exhaustive search over $\{0,1\}^d$ and its complexity is $O(2^d)$.

2. What is $\text{Prob}\{y|\theta, X\}$? What is the minimization problem to be solved to find MLE?
   Answer:
   $$\text{Prob}\{y|\theta, X\} = \frac{1}{(2\pi\sigma^2)^{m/2}} e^{-\|y - X\theta\|^2}.$$
   $$\hat{\theta} = \arg\max_{\theta \in \Theta} \text{Prob}\{y|\theta, X\} = \arg\min_{\theta \in \Theta} \|y - X\theta\|^2.$$

3. Let β be the inverse temperature. Write down a probability distribution π_β that concentrates on MLE as $\beta \to \infty$.
   Answer:
   $$\pi_\beta(\theta) = \frac{1}{Z_\beta} e^{-\beta\|y - X\theta\|^2},$$
   where $\mathbb{Z}_\beta = \sum_{\theta \in \Theta} e^{-\beta\|y - X\theta\|^2}$

4. Design a Metropolis-Hastings (MH) algorithm to draw samples from the probability distribution π_β. Clearly specify the state space, the base chain, and the expression for the acceptance probabilities of your algorithm. What is the computational complexity of computing the acceptance probabilities with respect to m and d?
Answer: The following is one possible MH algorithm:

   (a) Initialize with a random configuration θ^0

   (b) For t from 0 to $N-1$:

      i. Pick i uniformly at random in $\{1, 2, \cdots, d\}$

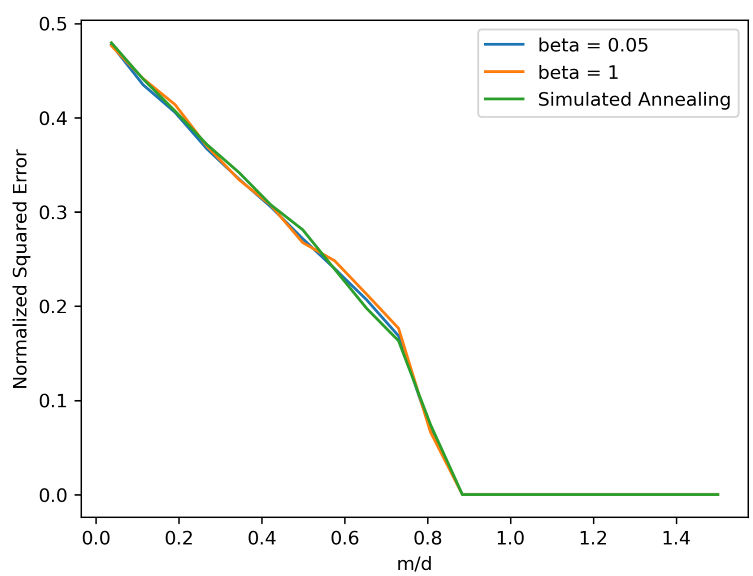
      ii. Let the proposed state be

$$\theta_j^* = \begin{cases} \theta_j^t & \text{for } j \neq i \\ 1 - \theta_j^t & \text{for } j = i \end{cases}$$

      iii. Accept θ^* as θ^{t+1} with probability $\min\left\{1, e^{-\beta(\|y - X\theta^*\|^2 - \left\|y - X\theta^t\right\|^2)}\right\}$. Set $\theta^t = \theta^{t-1}$ with remaining probability

   Observing that θ^* and θ^t differ only at one coordinate, we just need $O(m)$ computations to compute $\|y - X\theta^*\|^2$ given $\left\|y - X\theta^t\right\|^2$. Hence, the complexity of computing the acceptance probability is $O(m)$.

5. Implement the designed MH scheme to compute an estimate of $\hat{\theta}$ for a given realization of θ and X. Estimate the mean squared error $\frac{2}{d}\mathbb{E}\left\|\hat{\theta} - \theta\right\|^2$ for different values of m, where the expectation is over θ and X. Experiment with different values of β and also try the Simulated Annealing technique. What works best for your algorithm?
Answer: We expect the plots of MSE for different β and simulated annealing. The above MH algorithm does not show much difference with different values of β and simulated annealing.



6. Plot $\frac{2}{d}\mathbb{E}\left\|\hat{\theta} - \theta\right\|^2$ as a function of m. Comment on the characteristics of your plot. What is the minimum value of $\frac{m}{d}$ required to reliably recover θ?
   Answer: MSE decreases almost linearly with m initially. There is a small elbow at $\frac{m}{d} \approx 0.75$. The minimum m/d required for reliable recovery is approximately 0.9. We expect a plot similar to the previous one.

1.2 Question 2: Recovering a sparse, binary signal

Now, let $\Theta = \left\{\theta \in \{0,1\}^d : \|\theta\|_0 = s\right\}$, where $\|\theta\|_0$ gives the number of non-zero components of θ. Assume that $s \ll d$. Again, we study MH algorithms for finding MLE.

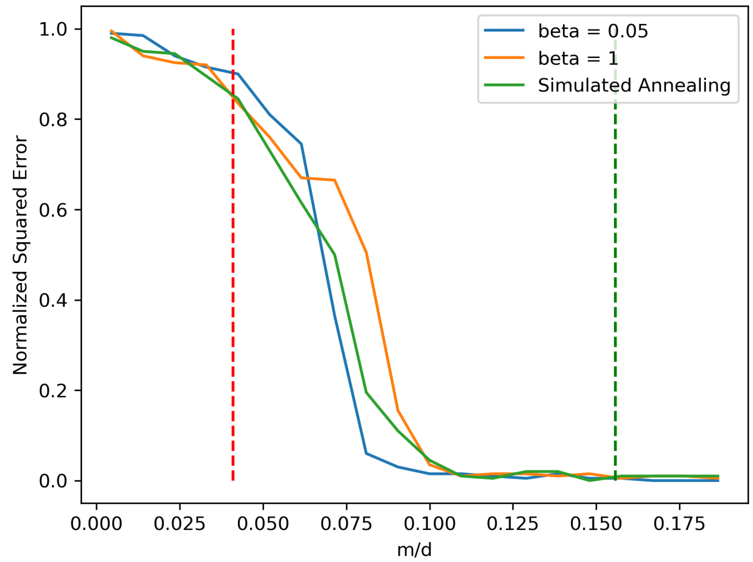
1. With θ being the current state of the MH algorithm, consider the following base chain:

   $(*)$ Choose two coordinates i and j independently and uniformly at random. Swap the i^{th} and j^{th} coordinates of θ to obtain the next state.

   Do you think that this base chain is suitable for the current problem? What issue do you see when $s \ll d$?
   Answer: This base chain is not suitable for recovering sparse signals. With this base chain, there is a high probability that the proposed state is the same as the current state. Hence the Markov chain takes a longer time to move to a new state, which degrades the convergence rate.

2. How would you modify the base chain $(*)$ to improve the convergence? You should aim to find a base chain that enables the MH algorithm to recover θ with $m = O(s\log(d))$ measurements.
   Answer: One possibility is the following base chain: Choose i uniformly at random among the coordinates with value 1 and j from the coordinates with value 0. Swap the i^{th} and j^{th} coordinates of θ to obtain the next state.
   This base chain is guaranteed to propose a state different from the current state at each step, and hence we expect better convergence. Numerical experiments show the same.

3. Plot $\frac{1}{2s}\mathbb{E}\left\|\hat{\theta} - \theta\right\|^2$ as a function of m for the MH algorithm with the base chain you proposed. Comment on the characteristics of your plot. What is the minimum $\frac{m}{d}$ required to reliably recover θ?
   Answer: We observe that the MSE falls sharply to 0 after a certain value of m/d. This threshold of m/d is of the order $\frac{s}{d}\log(d)$. Hence the minimum m required is of the order $s\log(d)$.



1.3 Question 3: Recovering sparse signal from 1-bit measurements

For this question assume that $\Theta = \left\{\theta \in \{0,1\}^d : \|\theta\|_0 = s\right\}$, and the measurement is generated as

$$y = \text{sign}\left(X\theta + \xi\right), \tag{3}$$

where the sign function acts component-wise on vectors and is given by

$$\text{sign}(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ -1 & \text{if } x < 0. \end{cases}$$

1. Formulate the optimization problem to find the MLE in this case. Write down a probability distribution π_β that concentrates on MLE as $\beta \to \infty$.
   Answer: We see that

   $$\text{Prob}\{y_i = \pm 1 | X, \theta\} = \text{Prob}\{\xi_i \leq \pm (X\theta)_i\}.$$

   Therefore,

   $$\text{Prob}\{y|X,\theta\} = \prod_{i=1}^{m} \Phi\left\{\frac{y_i(X\theta)_i}{\sigma}\right\},$$

   where Φ is the CDF of standard normal distribution. The optimization problem to solve MLE is

   $$\begin{aligned}\hat{\theta} &= \arg\max_{\theta\in\Theta} \text{Prob}\{y|\theta, X\}, \\ &= \arg\max_{\theta\in\Theta} \prod_{i=1}^{m} \Phi\left\{\frac{y_i(X\theta)_i}{\sigma}\right\}, \\ &= \arg\min_{\theta\in\Theta} - \sum_{i=1}^{m} \log \Phi\left\{\frac{y_i(X\theta)_i}{\sigma}\right\}.\end{aligned}$$

We can choose the distribution π_β to be

$$\pi_\beta = \frac{e^{\beta \sum_{i=1}^{m} \log \Phi\left\{\frac{y_i (X\theta)_i}{\sigma}\right\}}}{Z_\beta},$$

where $\mathbb{Z}_\beta = \sum_{\theta \in \Theta} e^{\beta \sum_{i=1}^{m} \log \Phi\left\{\frac{y_i (X\theta)_i}{\sigma}\right\}}$.

2. Design and implement an MH algorithm to estimate the derived MLE. Plot $\frac{1}{2s}\mathbb{E}\left\|\hat{\theta} - \theta\right\|^2$ as a function of m.
Answer: We can use the same MH algorithm used for the previous problem (1.2.2), with acceptance probability given by

$$\min\left\{1, e^{\beta \sum_{i=1}^{m} \log \frac{\Phi\left\{\frac{y_i (X\theta^*)_i}{\sigma}\right\}}{\Phi\left\{\frac{y_i (X\theta^t)_i}{\sigma}\right\}}}\right\}.$$

The MSE vs m/d plot is shown below. We observe that there is no sharp fall in MSE. However, the number of measurements m required for reliable recovery is much less than d.