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.

2. What is $\text{Prob}\{y|\theta, X\}$? What is the minimization problem to be solved to find MLE?

3. Let β be the inverse temperature. Write down a probability distribution π_β that concentrates on MLE as $\beta \to \infty$.

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?

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?

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?

You may use $d \in [2000, 5000]$, Total Markov chain steps $= 100d$, $\sigma = 1$.

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 $s \ll d$?

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.

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?

You may use $d \in [2000, 5000]$, $s = d/100$, Total Markov chain steps $= 100d$, $\sigma = 1$.

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 $\beta \to \infty$.

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.

   You may use $d \in [500, 2000]$, $s = d/100$, Total Markov chain steps $= 100d$, $\sigma = 1$.

2 Part 2: Competition

During the competition, to take place on **Tuesday, Dec 19, 08:15 AM**, we will ask you to change your algorithm, so as to solve a slightly different problem. We will still specify in which format exactly we expect you to send us your solution(s).
Your tasks for this second part of the project: Be ready for it!

3 Deadlines

- You should work in teams of 3. Please send to Anand, Nicolas, and Olivier an email with the composition of your team. Please do this by **Thursday, Nov 23, 23h59**.
- We expect a report from each team with approx 4-5 pages with answers to questions in Part 1. Please include only the graphs that matter! The report is due on **Friday, Dec 15, 23h59**.