Probability for Data Science – Master's Degree in Data Science – A.Y. 2023/2024
Academic Staff: Francesca Collet, Paolo Dai Pra

PROJECT 2: BINARY SIGNAL RECOVERY

Abstract. The scope of the present project is to illustrate how to recover a binary signal from noisy observations using Markov Chain Monte Carlo techniques.

Binary signal recovery via maximum likelihood estimate

Let $X \in \mathbb{R}^{m\times d}$ be a random sensing matrix with i.i.d. entries sampled from $\mathrm{N}(0,1)$. Let $\xi \in \mathbb{R}^m$ be a noise vector, independent of X, with i.i.d. entries sampled from $\mathrm{N}(0,1)$.
Take $\Theta = \{0,1\}^d$ (signal space) and let $\theta \in \Theta$ (signal) be chosen uniformly at random and be independent of the pair (X,ξ).
The measurement vector $y \in \mathbb{R}^m$ is generated as

$$y = X\theta + \xi.$$

We want to recover the unknown vector θ using Markov Chain Monte Carlo techniques, given the observations (X, y). We are interested in the case when d is large. We recover θ by finding the maximum likelihood estimate.

In the present setting, the maximum likelihood estimate of θ is given by the value $\hat{\theta} \in \Theta$ that maximizes the likelihood function

$$\mathcal{L}(X,y;\theta) = \frac{\exp\{-\frac{1}{2}\,(y - X\theta)^{\mathsf{T}}(y - X\theta)\}}{(2\pi)^{m/2}},$$

given the observations (X, y). Here the superscript T represents the transpose operation. We can equivalently cast the question in the form of a minimization problem. Indeed, *the maximum likelihood estimate of θ is given by the value $\hat{\theta} \in \Theta$ that minimizes the function*

$$\mathcal{H}(X,y;\theta) = (y - X\theta)^{\mathsf{T}}(y - X\theta),$$

given the observations (X, y).

Metropolis-Hastings algorithm

Let $\beta > 0$ be a fixed real parameter. We construct the Metropolis-Hastings (discrete-time) Markov chain on the state space Θ, with stationary distribution

$$\pi_\beta(\theta) = \frac{e^{-\beta\mathcal{H}(X,y;\theta)}}{Z_\beta}, \qquad \text{with} \qquad Z_\beta = \sum_{\theta\in\Theta} e^{-\beta\mathcal{H}(X,y;\theta)}.$$

Observe that the probability distribution π_β concentrates on the maximum likelihood estimate as $\beta \to +\infty$. Therefore, if we choose β sufficiently large and we run the chain for a large number N of steps, we can take the state visited at time N as the maximum likelihood estimate $\hat{\theta}$.

The following algorithm produces the first N steps $\theta_1, \ldots, \theta_N$ of the Metropolis-Hasting chain on Θ.

Input: value of the parameter β;
number of steps N;
initial state $\bar{\theta} \in \Theta$;

Output: trajectory of the Metropolis-Hastings chain starting at $\bar{\theta}$;

Procedure

Step 1. Set $\theta_0 = \bar{\theta}$.

Step 2. For $t = 1, 2, \ldots, N-1$:

1. pick i uniformly at random in $\{1, 2, \ldots, d\}$;

2. let the proposed state be $\theta^* \in \Theta$, with entries

$$\theta^*(j) = \begin{cases} \theta_{t-1}(j) & \text{if } j \neq i \\ 1 - \theta_{t-1}(j) & \text{if } j = i \end{cases} \qquad (j = 1, 2, \ldots, d);$$

3. set

$$\theta_t = \begin{cases} \theta^* & \text{with probability } \min\left\{1, \dfrac{e^{-\beta\mathcal{H}(X,y;\theta^*)}}{e^{-\beta\mathcal{H}(X,y;\theta_{t-1})}}\right\} \\[2ex] \theta_{t-1} & \text{with probability } 1 - \min\left\{1, \dfrac{e^{-\beta\mathcal{H}(X,y;\theta^*)}}{e^{-\beta\mathcal{H}(X,y;\theta_{t-1})}}\right\}. \end{cases}$$

Project

By implementing the Metropolis-Hastings algorithm above, we determine an estimate $\hat{\theta}$ of a signal $\theta \in \Theta$ for any given realization of (X, y). To check the quality of our estimate, we analyze the mean squared error

$$\mathcal{E} = E\left((\hat{\theta} - \theta)^{\mathsf{T}}(\hat{\theta} - \theta)\right),$$

where the expectation is over θ and (X, y), for different values of m (number of measurements). Fix $d = 10$. For every $1 \leq m \leq 15$, compute the mean squared error. Plot $\mathcal{E}$ as a function of m and comment on the characteristics of your plot. What is the minimum value of $\frac{m}{d}$ required to reliably recover θ?

Remark. The mean squared error $\mathcal{E}$ can be estimated by exploiting the law of large numbers. Let M denote the number of independent realizations of (θ, X, y). Moreover, let $\hat{\theta}^{(j)}$ be the maximum likelihood estimate of the j-th signal $\theta^{(j)}$, obtained by the j-th run of the Metropolis-Hastings algorithm, given $(X^{(j)}, y^{(j)})$. If M is *sufficiently large* (use M of order 10^4), then we have the approximation

$$\mathcal{E} \approx \frac{1}{M}\sum_{j=1}^{M}\left(\hat{\theta}^{(j)} - \theta^{(j)}\right)^{\mathsf{T}}\left(\hat{\theta}^{(j)} - \theta^{(j)}\right).$$

References

[1] Levin D.A. and Peres Y., *Markov chains and mixing times*, Volume 107, American Mathematical Society, 2017

[2] Ross S.M., *Simulation*, Academic Press, 2006