भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Stochastic Processes

Markov Chain Monte Carlo Methods: Metropolis–Hastings Algorithm, Examples

Karthik P. N.

Assistant Professor, Department of AI

Email: pnkarthik@ai.iith.ac.in

25 April 2025

Recap of MCMC Techniques

Objective

To evaluate a sum of the form

$$I = \sum_{x \in \mathcal{X}} f(x)\, \pi(x),$$

for given functions $f, \pi : \mathcal{X} \to \mathbb{R}$, where π is a probability distribution on $\mathcal{X}$.
Here, $\mathcal{X} \subseteq \mathbb{R}^d$ for some d.

Difficulties:

- Direct sampling from π (via ITT or rejection sampling methods) may be difficult
- High dimensionality: d may be large (e.g., 10000)
- h and/or π may have complex expressions, making it difficult to compute the sum

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Metropolis–Hastings Algorithm

The Metropolis–Hastings MCMC Algorithm

- Originally proposed by Metropolis et al. [MRR+53] and Hastings [Has70]

The Metropolis–Hastings MCMC Algorithm

- Originally proposed by Metropolis et al. [MRR+53] and Hastings [Has70]
- The key idea: construct an ergodic Markov chain whose stationary distribution is π

The Metropolis–Hastings MCMC Algorithm

- Originally proposed by Metropolis et al. [MRR+53] and Hastings [Has70]
- The key idea: construct an ergodic Markov chain whose stationary distribution is π
- If $\{X_n\}_{n=0}^{\infty}$ is this Markov chain, it follows by the ergodic theorem that

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{\text{a.s.}} I.$$

The Metropolis–Hastings MCMC Algorithm

- Originally proposed by Metropolis et al. [MRR+53] and Hastings [Has70]
- The key idea: construct an ergodic Markov chain whose stationary distribution is π
- If $\{X_n\}_{n=0}^{\infty}$ is this Markov chain, it follows by the ergodic theorem that

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \quad \xrightarrow{\text{a.s.}} \quad I.$$

- The method also applies when $\mathcal{X}$ is countable or uncountable

Construction of Markov Chain

- Let Q be any row-stochastic matrix on $\mathcal{X} \times \mathcal{X}$, i.e.,

$$Q_{x,y} > 0, \qquad \sum_{y} Q_{x,y} = 1 \qquad \forall x, y \in \mathcal{X}.$$

Construction of Markov Chain

- Let Q be any row-stochastic matrix on $\mathcal{X} \times \mathcal{X}$, i.e.,

$$Q_{x,y} \geq 0, \qquad \sum_{y} Q_{x,y} = 1 \qquad \forall x, y \in \mathcal{X}.$$

- Let S be a non-negative, symmetric matrix having same size as Q

Construction of Markov Chain

- Let Q be any row-stochastic matrix on $\mathcal{X} \times \mathcal{X}$, i.e.,

$$Q_{x,y} > 0, \qquad \sum_{y} Q_{x,y} = 1 \qquad \forall x, y \in \mathcal{X}.$$

- Let S be a non-negative, symmetric matrix having same size as Q
- Let $A = [A_{x,y}]_{x,y \in \mathcal{X}}$ be defined via

$$A_{x,y} = \frac{S_{x,y}}{1 + \frac{\pi(x)\, Q_{x,y}}{\pi(y)\, Q_{y,x}}}.$$

Construction of Markov Chain

- Let Q be any row-stochastic matrix on $\mathcal{X} \times \mathcal{X}$, i.e.,

$$Q_{x,y} > 0, \qquad \sum_y Q_{x,y} = 1 \qquad \forall x, y \in \mathcal{X}.$$

- Let S be a non-negative, symmetric matrix having same size as Q
- Let $A = [A_{x,y}]_{x,y \in \mathcal{X}}$ be defined via

$$A_{x,y} = \frac{S_{x,y}}{1 + \frac{\pi(x)\, Q_{x,y}}{\pi(y)\, Q_{y,x}}}.$$

- The entries of S are chosen in such a way that $0 < A_{x,y} \leq 1$ for all $x, y \in \mathcal{X}$

Construction of Markov Chain

- Let Q be any row-stochastic matrix on $\mathcal{X} \times \mathcal{X}$, i.e.,

$$Q_{x,y} > 0, \qquad \sum_{y} Q_{x,y} = 1 \qquad \forall x, y \in \mathcal{X}.$$

- Let S be a non-negative, symmetric matrix having same size as Q
- Let $A = [A_{x,y}]_{x,y \in \mathcal{X}}$ be defined via

$$A_{x,y} = \frac{S_{x,y}}{1 + \frac{\pi(x)\, Q_{x,y}}{\pi(y)\, Q_{y,x}}}.$$

- The entries of S are chosen in such a way that $0 < A_{x,y} \leq 1$ for all $x, y \in \mathcal{X}$
- A is called the acceptance matrix

Construction of Markov Chain

- Choose an arbitrary starting state $X_0 = x$
- At each time step $n \in \mathbb{N}$, do the following two tasks:
  - Pick Y such that
  $$\mathbb{P}(Y = y \mid X_{n-1} = x) = Q_{x,y}$$
  - Define X_n as
  $$X_n = \begin{cases} y, & \text{with probability } A_{x,y}, \\ X_{n-1}, & \text{with probability} 1 - A_{x,y}. \end{cases}$$
- **Claim:** The TPM of $\{X_n\}_{n=0}^{\infty}$ will have π as the unique stationary distribution!

Choice of S Matrix in Metropolis–Hastings Algorithm

- The matrix S used in [MRR+53] (call this $S^{(M)}$) is given by

$$S^{(M)}_{x,y} = \begin{cases} 1 + \dfrac{\pi_x Q_{x,y}}{\pi_y Q_{y,x}}, & \dfrac{\pi_y Q_{y,x}}{\pi_x Q_{x,y}} \geq 1, \\[2ex] 1 + \dfrac{\pi_y Q_{y,x}}{\pi_x Q_{x,y}}, & \dfrac{\pi_y Q_{y,x}}{\pi_x Q_{x,y}} < 1. \end{cases}$$

Choice of S Matrix in Metropolis–Hastings Algorithm

- The matrix S used in [MRR+53] (call this $S^{(M)}$) is given by

$$S^{(M)}_{x,y} = \begin{cases} 1 + \frac{\pi_x Q_{x,y}}{\pi_y Q_{y,x}}, & \frac{\pi_y Q_{y,x}}{\pi_x Q_{x,y}} \geq 1, \\ 1 + \frac{\pi_y Q_{y,x}}{\pi_x Q_{x,y}}, & \frac{\pi_y Q_{y,x}}{\pi_x Q_{x,y}} < 1. \end{cases}$$

Choice of S Matrix in Metropolis–Hastings Algorithm

- The matrix S used in [MRR$^+$53] (call this $S^{(M)}$) is given by

$$S^{(M)}_{x,y} = \begin{cases} 1 + \frac{\pi_x\, Q_{x,y}}{\pi_y\, Q_{y,x}}, & \frac{\pi_y\, Q_{y,x}}{\pi_x\, Q_{x,y}} \geq 1, \\ 1 + \frac{\pi_y\, Q_{y,x}}{\pi_x\, Q_{x,y}}, & \frac{\pi_y\, Q_{y,x}}{\pi_x\, Q_{x,y}} < 1. \end{cases}$$

  With $S = S^{(M)}$ as above, the acceptance matrix A becomes

$$A_{x,y} = \begin{cases} 1, & \frac{\pi_y\, Q_{y,x}}{\pi_x\, Q_{x,y}} \geq 1, \\ \frac{\pi_y\, Q_{y,x}}{\pi_x\, Q_{x,y}}, & \frac{\pi_y\, Q_{y,x}}{\pi_x\, Q_{x,y}} < 1 \end{cases} = \min\left\{1, \frac{\pi_y\, Q_{y,x}}{\pi_x\, Q_{x,y}}\right\}.$$

- Barker [Bar65] proposes using $S = S^{(B)}$, where $S^{(B)}_{x,y} = 1$ for all $x, y \in \mathcal{X}$

Choice of S Matrix in Metropolis–Hastings Algorithm

- If Q is symmetric (e.g., every row of Q is uniform), then

$$A_{x,y}^{(M)} = \min\left\{1, \frac{\pi_y}{\pi_x}\right\}, \qquad A_{x,y}^{(B)} = \frac{\pi_y}{\pi_x + \pi_y}.$$

Choice of S Matrix in Metropolis–Hastings Algorithm

- If Q is symmetric (e.g., every row of Q is uniform), then

$$A_{x,y}^{(M)} = \min\left\{1, \frac{\pi_y}{\pi_x}\right\}, \qquad A_{x,y}^{(B)} = \frac{\pi_y}{\pi_x + \pi_y}.$$

- Thus, if $\pi_x = \pi_y$, then $A_{x,y}^{(M)} = 1$, whereas $A_{x,y}^{(B)} = 0.5$
In this case, [MRR+53] suggests picking $X_{n+1} = y$ with probability 1
[Bar65] suggests taking $X_{n+1} = X_n = x$ with probability 0.5 and taking $X_{n+1} = y$ with probability 0.5

References

Anthony Alfred Barker.
Monte carlo calculations of the radial distribution functions for a proton? electron plasma.
Australian Journal of Physics, 18(2):119–134, 1965.

W Keith Hastings.
Monte carlo sampling methods using markov chains and their applications.
Biometrika, 1970.

N Metropolis, A Rosenbluth, MN Rosenbluth, AH Teller, and E Teller.
Perspective on "equation of state calculations by fast computing machines".
J. Chem. Phys, 21:1087–1092, 1953.