# Lab session: Markov Chain Monte Carlo

April 28, 2025

## 1 Introduction

So far in DTMCs, we were given a Markov chain $(X_n)_{n\in\mathbb{N}}$ on a finite state space $\mathcal{S}$ and we then studied the existence and uniqueness of its stationary distribution and convergence to it. In this lab, we will consider the reverse problem :

*Given a probability distribution $\boldsymbol{\pi}$ on a sample space $\mathcal{S}$, can we construct a Markov chain $(X_t)_{t\geq 0}$ such that $\boldsymbol{\pi}$ is a stationary distribution?*

If, in addition, the chain is irreducible and aperiodic, then by the convergence theorem, we know that the distribution $\boldsymbol{\pi}_t$ of $X_t$ converges to $\boldsymbol{\pi}$. Hence, if we run the chain for long enough, the state of the chain is asymptotically distributed as $\boldsymbol{\pi}$. In other words, we can sample a random element of $\mathcal{S}$ according to the prescribed distribution $\boldsymbol{\pi}$ by emulating it through a suitable Markov chain. This method of sampling is called Markov chain Monte Carlo (MCMC).

The algorithm we focus on is the well-known **Metropolis-Hastings algorithm** for MCMC sampling.

## 2 The Metropolis-Hastings algorithm

Let $\mathbf{P}$ be a transition matrix of a Markov chain on a finite state space $\mathcal{S} = \{1, 2, \ldots, m\}$, $m \in \mathbb{N}^*$. Let $\boldsymbol{\pi}$ be a probability distribution on $\mathcal{S}$, which is not necessarily a stationary distribution for $\mathbf{P}$. Our goal is to design a Markov chain on $\mathcal{S}$ that has $\boldsymbol{\pi}$ as its stationary distribution.

The idea is to construct a new Markov chain that uses $\mathbf{P}$ to propose transitions, and then decides whether to accept or reject these proposals using a suitable acceptance probability.

Fix an $m \times m$ matrix $\boldsymbol{A}$ with entries from $[0, 1]$. Consider a Markov chain $(X_t)_{t\geq 0}$ on $\mathcal{S}$ defined as follows :

- **Generation Step :** Suppose the current state is $X_t = a \in \mathcal{S}$. Generate a candidate state $b \in \mathcal{S}$ according to the proposal distribution $\mathbf{P}(a, \cdot)$, where $\boldsymbol{P}(a, \cdot)$ is the $a$-th row vector of $\mathbf{P}$.
- **Rejection Step :** Flip an independent coin with success probability $\boldsymbol{A}(a, b)$, where $\boldsymbol{A}(a, b)$ is the value of $\boldsymbol{A}$ at its $a$-th row and $b$-th column :
  - If the coin flip is successful (i.e., with probability $\boldsymbol{A}(a, b)$), accept the proposed move and set $X_{t+1} = b$.
  - Otherwise (with probability $1 - \boldsymbol{A}(a, b)$), reject the move and remain at the current state: $X_{t+1} = a$.

Here, the entry $\boldsymbol{A}(a, b)$ is called the **acceptance probability** of the move from $a$ to $b$.

1. Let $\mathbf{Q}$ denote the matrix of the chain $(X_n)_{n\in\mathbb{N}}$ described above. Show that :

$$\mathbf{Q}(a,b) = \begin{cases} \mathbf{P}(a,b)\mathbf{A}(a,b) & \text{if } b \neq a, \\ 1 - \sum\limits_{c\in\mathcal{S}, c\neq a} \mathbf{P}(a,c)\mathbf{A}(a,c) & \text{if } b = a. \end{cases}$$

2. Show that if $\boldsymbol{\pi}(x)\mathbf{Q}(x,y) = \boldsymbol{\pi}(y)\mathbf{Q}(y,x)$ for all $x, y \in \mathcal{S}$ such that $x \neq y$, then $\boldsymbol{\pi}\mathbf{Q} = \boldsymbol{\pi}$.

3. Deduce that if

$$\boldsymbol{\pi}(x)\mathbf{P}(x,y)\mathbf{A}(x,y) = \boldsymbol{\pi}(y)\mathbf{P}(y,x)\mathbf{A}(y,x), \quad \forall x, y \in \mathcal{S}, x \neq y,$$

then $\boldsymbol{\pi}$ is a stationary distribution for $(X_n)_{n\in\mathbb{N}}$.

4. We are also interested in fast convergence of the Markov chain. Thus, we want to choose the acceptance probability $\mathbf{A}(a,b) \in [0,1]$ as large as possible for each $a, b \in \mathcal{S}$. Show that the following choice

$$\mathbf{A}(x,y) = \min\left(\frac{\boldsymbol{\pi}(y)\mathbf{P}(y,x)}{\boldsymbol{\pi}(x)\mathbf{P}(x,y)}, 1\right)$$

for all $x, y \in \mathcal{S}$, $x \neq y$, satisfies the condition in the previous question, and each $\mathbf{A}(x,y)$ is maximized for all $x \neq y$.

(iv) Let $(Y_t)_{t\geq 0}$ be a random walk on the 5-wheel graph $G = (V, E)$ as shown in Figure 8. Show that

$$\pi = \left(\frac{6}{20}, \frac{5}{20}, \frac{2}{20}, \frac{1}{20}, \frac{4}{20}, \frac{3}{20}\right)$$

is the unique stationary distribution of $Y_t$. Apply the Metropolis-Hastings algorithm derived in (i)-(iii) above to modify $Y_t$ to obtain a new Markov chain $X_t$ on $V$ that converges to Uniform$(V)$ in distribution.

We claim that if

$$\boldsymbol{\pi}(x)Q(x,y) = \boldsymbol{\pi}(y)Q(y,x) \quad \forall x, y \in \mathcal{S},\ x \neq y$$

then $\boldsymbol{\pi}Q = \boldsymbol{\pi}$, i.e., $\boldsymbol{\pi}$ is a stationary distribution of the chain.

Now, suppose that

$$\boldsymbol{\pi}(x)P(x,y)A(x,y) = \boldsymbol{\pi}(y)P(y,x)A(y,x) \quad \forall x, y \in \mathcal{S},\ x \neq y$$

then the detailed balance condition holds for $Q$, hence $\boldsymbol{\pi}$ is a stationary distribution for $(X_t)_{t\geq 0}$.

# (iii) Choice of Acceptance Probability

Since we want the Markov chain to converge quickly, we choose the acceptance probability $A(a,b) \in [0,1]$ to be as large as possible for all $a, b \in \mathcal{S}$. The following choice:

$$A(x,y) = \min\left(1, \frac{\boldsymbol{\pi}(y)P(y,x)}{\boldsymbol{\pi}(x)P(x,y)}\right) = \frac{\boldsymbol{\pi}(y)P(y,x)}{\boldsymbol{\pi}(x)P(x,y)} \wedge 1 \quad \forall x, y \in \mathcal{S},\ x \neq y$$

satisfies the condition in (ii) and ensures that each $A(x,y)$ is maximized under the constraint of symmetry in detailed balance.

# (iv) Application: Random Walk on the 5-Wheel Graph

Let $(Y_t)_{t\geq 0}$ be a random walk on the 5-wheel graph $G = (V, E)$ as shown in Figure 8. The state space is $V = \{1, 2, 3, 4, 5, 6\}$.

The unique stationary distribution of $(Y_t)$ is:

$$\boldsymbol{\pi} = \left[\frac{3}{20}, \frac{3}{20}, \frac{3}{20}, \frac{3}{20}, \frac{3}{20}, \frac{5}{20}\right]$$

We apply the Metropolis-Hastings algorithm derived above to modify $(Y_t)$ into a new Markov chain $(X_t)$ on $V$ such that the stationary distribution is uniform on $V$, i.e.,

$$\boldsymbol{\pi}_{\text{target}} = \left[\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right]$$

Let $P(x, y)$ be the transition probabilities of $(Y_t)$. Define acceptance probabilities:

$$A(x, y) = \min\left(1, \frac{\boldsymbol{\pi}_{\text{target}}(y)P(y, x)}{\boldsymbol{\pi}_{\text{target}}(x)P(x, y)}\right)$$

Then, define the transition matrix $Q$ for $(X_t)$ using the formula in part (i). The resulting chain will converge in distribution to $\boldsymbol{\pi}_{\text{target}} = \text{Uniform}(V)$.

# Learning Objectives

By the end of this lab, you should be able to :

- Understand the relevance of Markov chains in sampling.
- Implement the Metropolis-Hastings algorithm in Python.
- Analyze and interpret the behavior of MCMC samplers.
- Apply the algorithm to both continuous and discrete probability distributions.

# Part 1: Conceptual Questions

**Q1.** Define a Markov chain and explain what is meant by the stationary distribution.

**Q2.** Why do we use MCMC methods in Bayesian inference? What problems do they help to solve?

**Q3.** Describe each step of the Metropolis-Hastings algorithm. What role does the proposal distribution play?

**Q4.** What could happen if the proposal distribution's variance is too large or too small?

**Q5.** How would you evaluate if your Markov chain has converged?

**Q6.** For the Gaussian example:

- Plot the histogram of samples and overlay the target PDF.
- What does the trace plot reveal about the sampler?
- How does changing the proposal standard deviation affect the sampling?

**Q7.** For the Zipf distribution:

- Compare the sampled histogram with the true PMF.
- What difficulties arise when sampling from heavy-tailed or discrete distributions?

# Extension Questions (Optional)

**E1.** Modify the sampler to work on a 2D target distribution (e.g., a bivariate Gaussian). What changes are necessary?

**E2.** Implement a different discrete proposal strategy for the Zipf sampler (e.g., geometric steps).

**E3.** Compare acceptance rates across different samplers. What rate seems to balance convergence and exploration?

**References :** MATH 171 Stochastic Processes Lecture Notes (Hanbaek Lyu).