भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Stochastic Processes

Ergodic Theorem, Markov Chain Monte Carlo Methods

Karthik P. N.
Assistant Professor, Department of AI
Email: pnkarthik@ai.iith.ac.in

22 April 2025

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Recap of Important Results

- Period, transience, positive recurrence, and null recurrence are class properties

Recap of Important Results

- Period, transience, positive recurrence, and null recurrence are class properties
- Open communicating classes are transient.
  Closed communicating classes are recurrent.
  Finite closed communicating classes are positive recurrent.

Recap of Important Results

- Period, transience, positive recurrence, and null recurrence are class properties
- Open communicating classes are transient.
  Closed communicating classes are recurrent.
  Finite closed communicating classes are positive recurrent.
- The following are equivalent.

Recap of Important Results

- Period, transience, positive recurrence, and null recurrence are class properties
- Open communicating classes are transient.
  Closed communicating classes are recurrent.
  Finite closed communicating classes are positive recurrent.
- The following are equivalent.
  - State x is transient.

Recap of Important Results

- Period, transience, positive recurrence, and null recurrence are class properties
- Open communicating classes are transient.
  Closed communicating classes are recurrent.
  Finite closed communicating classes are positive recurrent.
- The following are equivalent.
  - State x is transient.
  - $\mathbb{E}[N_x \mid X_0 = x] = \sum_{n\in\mathbb{N}} P^n_{x,x} < +\infty$

Recap of Important Results

- Period, transience, positive recurrence, and null recurrence are class properties
- Open communicating classes are transient.
  Closed communicating classes are recurrent.
  Finite closed communicating classes are positive recurrent.
- The following are equivalent.
  - State x is transient.
  - $\mathbb{E}[N_x \mid X_0 = x] = \sum_{n\in\mathbb{N}} P^n_{x,x} < +\infty$
  - $f_{xx} = \mathbb{P}(\tau_x^{(1)} < +\infty \mid X_0 = x) = \sum_{n\in\mathbb{N}} f_{xx}^{(n)} < 1, \qquad f_{xx}^{(n)} = \mathbb{P}(\tau_x^{(1)} = n \mid X_0 = x).$

Recap of Important Results

- Period, transience, positive recurrence, and null recurrence are class properties
- Open communicating classes are transient.
  Closed communicating classes are recurrent.
  Finite closed communicating classes are positive recurrent.
- The following are equivalent.
  - State x is transient.
  - $\mathbb{E}[N_x \mid X_0 = x] = \sum_{n \in \mathbb{N}} P^n_{x,x} < +\infty$
  - $f_{xx} = \mathbb{P}(\tau_x^{(1)} < +\infty \mid X_0 = x) = \sum_{n \in \mathbb{N}} f_{xx}^{(n)} < 1, \qquad f_{xx}^{(n)} = \mathbb{P}(\tau_x^{(1)} = n \mid X_0 = x).$
- The following are equivalent.

Recap of Important Results

- Period, transience, positive recurrence, and null recurrence are class properties
- Open communicating classes are transient.
  Closed communicating classes are recurrent.
  Finite closed communicating classes are positive recurrent.
- The following are equivalent.
  - State x is transient.
  - $\mathbb{E}[N_x \mid X_0 = x] = \sum_{n \in \mathbb{N}} P^n_{x,x} < +\infty$
  - $f_{xx} = \mathbb{P}(\tau_x^{(1)} < +\infty \mid X_0 = x) = \sum_{n \in \mathbb{N}} f_{xx}^{(n)} < 1, \qquad f_{xx}^{(n)} = \mathbb{P}(\tau_x^{(1)} = n \mid X_0 = x).$
- The following are equivalent.
  - State x is recurrent.

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Recap of Important Results

- Period, transience, positive recurrence, and null recurrence are class properties
- Open communicating classes are transient.
  Closed communicating classes are recurrent.
  Finite closed communicating classes are positive recurrent.
- The following are equivalent.
  - State x is transient.
  - $\mathbb{E}[N_x \mid X_0 = x] = \sum_{n\in\mathbb{N}} P^n_{x,x} < +\infty$
  - $f_{xx} = \mathbb{P}(\tau_x^{(1)} < +\infty \mid X_0 = x) = \sum_{n\in\mathbb{N}} f_{xx}^{(n)} < 1,$ $\qquad f_{xx}^{(n)} = \mathbb{P}(\tau_x^{(1)} = n \mid X_0 = x).$
- The following are equivalent.
  - State x is recurrent.
  - $\mathbb{E}[N_x \mid X_0 = x] = \sum_{n\in\mathbb{N}} P^n_{x,x} = +\infty$

Recap of Important Results

- Period, transience, positive recurrence, and null recurrence are class properties
- Open communicating classes are transient.
  Closed communicating classes are recurrent.
  Finite closed communicating classes are positive recurrent.
- The following are equivalent.
  - State x is transient.
  - $\mathbb{E}[N_x \mid X_0 = x] = \sum_{n\in\mathbb{N}} P^n_{x,x} < +\infty$
  - $f_{xx} = \mathbb{P}(\tau_x^{(1)} < +\infty \mid X_0 = x) = \sum_{n\in\mathbb{N}} f_{xx}^{(n)} < 1,$ $\quad f_{xx}^{(n)} = \mathbb{P}(\tau_x^{(1)} = n \mid X_0 = x).$
- The following are equivalent.
  - State x is recurrent.
  - $\mathbb{E}[N_x \mid X_0 = x] = \sum_{n\in\mathbb{N}} P^n_{x,x} = +\infty$
  - $f_{xx} = \mathbb{P}(\tau_x^{(1)} < +\infty \mid X_0 = x) = \sum_{n\in\mathbb{N}} f_{xx}^{(n)} = 1,$ $\quad f_{xx}^{(n)} = \mathbb{P}(\tau_x^{(1)} = n \mid X_0 = x).$

Recap of Important Results

- If x is transient, then

$$\lim_{n\to\infty} P^n_{x,x} = 0, \qquad \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P^k_{x,x} = 0.$$

- If x is recurrent, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P^k_{x,x} = \frac{1}{\mu_{xx}}.$$

Recap of Important Results

- If x is transient, then

$$\lim_{n\to\infty} P^n_{x,x} = 0, \qquad \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} P^k_{x,x} = 0.$$

- If x is recurrent, then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} P^k_{x,x} = \frac{1}{\mu_{xx}}.$$

  In addition, if x is aperiodic, then

$$\lim_{n\to\infty} P^n_{x,x} = \frac{1}{\mu_{xx}}.$$

- For an irreducible Markov chain,

  Unique stationary distribution exists $\iff$ Markov chain is positive recurrent.

  Furthermore, in this case, $\pi(x) = \frac{1}{\mu_{xx}} > 0$ for all x.

Recap of Important Results

- If $x \longleftrightarrow y$, and y is recurrent, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P^k_{x,y} = \frac{1}{\mu_{yy}}.$$

Recap of Important Results

- If $x \longleftrightarrow y$, and y is recurrent, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P^k_{x,y} = \frac{1}{\mu_{yy}}.$$

In addition, if y is aperiodic, then

$$\lim_{n\to\infty} P^n_{x,y} = \frac{1}{\mu_{yy}}.$$

Recap of Important Results

- If $x \longleftrightarrow y$, and y is recurrent, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P^k_{x,y} = \frac{1}{\mu_{yy}}.$$

In addition, if y is aperiodic, then

$$\lim_{n\to\infty} P^n_{x,y} = \frac{1}{\mu_{yy}}.$$

In addition, if y is positive recurrent, then

$$\lim_{n\to\infty} P^n_{x,y} = \frac{1}{\mu_{yy}} = \pi(y).$$

Ergodicity and Convergence to Stationary Distribution

Theorem (Ergodicity and Convergence to Stationary Distribution)

Consider a time-homogeneous DTMC $\{X_n\}_{n=0}^{\infty}$ on a discrete state space $\mathcal{X}$ with TPM P.
Assume that $X_0 = x$, and for each $n \in \mathbb{N}$, let π_n denote the PMF of X_n.
If P is ergodic with associated stationary distribution π, then

$$\lim_{n\to\infty} d_{\mathrm{TV}}(\pi_n, \pi) = \lim_{n\to\infty} \frac{1}{2} \|\pi_n - \pi\|_1 = 0,$$

where d_{TV} denotes the total variation distance.

Ergodic Theorem

Ergodic Theorem

Theorem (Ergodic Theorem)

For an ergodic Markov chain $\{X_n\}_{n=0}^{\infty}$ on a finite state space $\mathcal{X}$ with TPM P, stationary distribution π, and any starting state $X_0 = x \in \mathcal{X}$,

$$\frac{N_y(n)}{n} \xrightarrow{\text{a.s.}} \pi(y) \qquad \forall y \in \mathcal{X},$$

where $N_y(n)$ is the number of times the Markov chain visits state y up to time n.
More generally, for any bounded, continuous function $f : \mathcal{X} \to \mathbb{R}$ and for any starting state $X_0 = x \in \mathcal{X}$,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{\text{a.s.}} \sum_{y \in \mathcal{X}} f(y)\,\pi(y).$$

Proof of Theorem

- Proof of first part of the theorem is an exercise (see Homework 6)

Proof of Theorem

- Proof of first part of the theorem is an exercise (see Homework 6)
- To prove the second part of the theorem, observe that

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) = \sum_{y\in\mathcal{X}} f(x)\,\frac{N_y(n)}{n}$$

The result then follows from part 1

Markov Chain Monte Carlo Methods

Objective

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.

Objective

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

Objective

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)

Objective

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.,

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

- 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

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

- 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 $X_0 = x$

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:

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

$$\mathbb{P}(Y = y \mid X_{n-1} = x) = Q_{x,y}$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

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
$$X_n = \begin{cases} y, & \text{with probability } A_{x,y}, \\ X_{n-1}, & \text{with probability} 1 - A_{x,y}. \end{cases}$$

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!

Proof of Claim

- For any $x, y \in \mathcal{X}$, we have

$$P_{x,y} = \mathbb{P}(X_n = y \mid X_{n-1} = x)$$

Proof of Claim

- For any $x, y \in \mathcal{X}$, we have

$$P_{x,y} \quad = \quad \mathbb{P}(X_n = y \mid X_{n-1} = x)$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Proof of Claim

- For any $x, y \in \mathcal{X}$, we have

$$P_{x,y} \quad = \quad \mathbb{P}(X_n = y \mid X_{n-1} = x) \quad = \quad \begin{cases} Q_{x,y}\, A_{x,y}, & x \neq y, \\ 1 - \sum_{z \neq x} Q_{x,z}\, A_{x,z}, & x = y. \end{cases}$$

- It then follows that for all $x, y \in \mathcal{X}$,

$$\pi(x)\, P_{x,y} \quad =$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Proof of Claim

- For any $x, y \in \mathcal{X}$, we have

$$P_{x,y} \quad = \quad \mathbb{P}(X_n = y \mid X_{n-1} = x) \quad = \quad \begin{cases} Q_{x,y}\, A_{x,y}, & x \neq y, \\ 1 - \sum_{z \neq x} Q_{x,z}\, A_{x,z}, & x = y. \end{cases}$$

- It then follows that for all $x, y \in \mathcal{X}$,

$$\pi(x)\, P_{x,y} \quad =$$

Proof of Claim

- For any $x, y \in \mathcal{X}$, we have

$$P_{x,y} \quad = \quad \mathbb{P}(X_n = y \mid X_{n-1} = x) \quad = \quad \begin{cases} Q_{x,y}\, A_{x,y}, & x \neq y, \\ 1 - \sum_{z \neq x} Q_{x,z}\, A_{x,z}, & x = y. \end{cases}$$

- It then follows that for all $x, y \in \mathcal{X}$,

$$\pi(x)\, P_{x,y} \quad = \quad \Big(\pi(x)\, Q_{x,y}\, \pi(y)\, Q_{y,x}\Big) \cdot \frac{S_{x,y}}{\pi(x)\, Q_{x,y} + \pi(y)\, Q_{y,x}} \quad = \quad \pi(y)\, P_{y,x}.$$

- Therefore, $\pi(x) = \sum_y \pi(y)\, P_{x,y}$ for all x, thus proving that π is a stationary distribution

References

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.