离散时间 Markov 链的几何遍历速度

宋延红

中南财经政法大学统计与数学学院, 武汉 430073

E-mail: songyh@zuel.edu.cn

收稿日期: 2018-12-11; 接受日期: 2019-07-17; 网络出版日期: 2019-10-23

国家自然科学基金 (批准号: 11501576) 和中南财经政法大学研究生教育创新计划 (批准号: 201821301) 资助项目

摘要 本文研究一般状态空间离散时间几何遍历 Markov 链的收敛速度. 当状态空间中含有原子时, 本文采用与可数状态空间类似的方法, 即采用谱理论法研究几何遍历速度. 对于不含原子的 Markov 链, 先利用 Nummelin 分裂技术构造出含有原子的新 Markov 链, 再建立新 Markov 链与原 Markov 链几何遍历速度的关系, 最后由最小化条件和首次回返时的几何阶矩得到原 Markov 链几何遍历的收敛速度. 进一步, 本文将所得结果应用于研究 Metropolis-Hastings 算法.

关键词 Markov 链 几何遍历 收敛速度

MSC (2010) 主题分类 60J05, 37A30

1 引言

Markov 链的几何遍历具有重要的理论价值和广泛的应用前景, 是概率论中的重要研究课题. 可数状态空间离散时间 Markov 链几何遍历的收敛速度已经得到了广泛的研究, 如文献 [1–3] 等. 一般状态空间中几何遍历速度的研究结果也很丰富, 可参见文献 [4–10] 及其中的参考文献. Sokal 和 Thomas^[3] 利用击中时的矩条件研究了可数状态空间 Markov 链的谱半径, 本文将推广上述结果, 并结合陈木法在专著 [5, 第 9 章] 中的方法, 采用首次回返时的几何阶矩来讨论一般状态空间离散时间 Markov 链的几何遍历速度.

设 $\Phi = \{\Phi_n : n \in \mathbb{Z}_+\}$ 是一般状态空间 X 上的离散时间 Markov 链, 在 X 上赋予可数生成的 σ 代数 $\mathcal{B}(X)$. 令 P_x 和 E_x 分别为给定 $\Phi_0 = x$ 的条件下 Markov 链的条件分布和条件期望, 记

$$P^n(x, A) = P_x\{\Phi_n \in A\}, \quad n \in \mathbb{Z}_+, \quad x \in X, \quad A \in \mathcal{B}(X)$$

为 Markov 链的 n 步转移核, 其作用在非负可测函数 f 和 σ 有限测度 μ 上分别定义为

$$P^n f(x) = \int_X f(y) P^n(x, dy), \quad \mu P^n(A) = \int_X P^n(x, A) \mu(dx).$$

英文引用格式: Song Y H. Convergence rates in geometric ergodicity for discrete-time Markov chains (in Chinese). Sci Sin Math, 2020, 50: 1087–1096, doi: 10.1360/N012018-00291

令概率测度 $\pi \in \mathcal{B}(X)$ 为 Markov 链的平稳分布, 即对任意的 $n \in \mathbb{Z}_+$, 有 $\pi = \pi P^n$, 其按如下方式作用在非负可测函数 f 上:

$$\pi(f) = \int_X f(x) \pi(dx).$$

给定集合 $A \in \mathcal{B}(X)$, A 的首次回返时定义为

$$\tau_A = \inf\{n \geq 1 : \Phi_n \in A\}.$$

对任意的 σ 有限测度 φ , Markov 链 Φ 称为 φ 不可约, 是指若 $\varphi(A) > 0$, 则有 $P_x\{\tau_A < \infty\} > 0, x \in X$. 由文献 [11, 第 4 章] 知, 对于 φ 不可约的 Markov 链, 必存在最大的不可约测度 ψ , 满足任意其他的不可约测度 ν 关于 ψ 是绝对连续的. 记 $\mathcal{B}^+(X) = \{A \in \mathcal{B}(X) : \psi(A) > 0\}$ 为具有正的 ψ 测度的集合, 由于平稳分布 π 与 ψ 相互绝对连续, 因此可认为 $\mathcal{B}^+(X) = \{A \in \mathcal{B}(X) : \pi(A) > 0\}$. 设 $L^2(\pi, X)$ 为常见的实 Hilbert 空间, 其上的内积 $(\cdot, \cdot)_{L^2(\pi, X)}$ 和范数 $\|\cdot\|_{L^2(\pi, X)}$ 分别定义为

$$(f, g)_{L^2(\pi, X)} = \int_X f(x)g(x)\pi(dx), \quad \|f\|_{L^2(\pi, X)} = (f, f)_{L^2(\pi, X)}^{1/2}, \quad f, g \in L^2(\pi, X).$$

若对任意的 $f, g \in L^2(\pi, X)$, 有

$$(Pf, g)_{L^2(\pi, X)} = (f, Pg)_{L^2(\pi, X)},$$

则称 Markov 链 Φ 关于平稳分布 π 可逆. 显然, 可逆 Markov 链的转移核为 $L^2(\pi, X)$ 空间中的自伴算子. 若

$$(Pf, f)_{L^2(\pi, X)} \geq 0, \quad f \in L^2(\pi, X),$$

则称转移核 P 是 $L^2(\pi, X)$ 空间中的非负定算子. 连续时间转移半群总是 $L^2(\pi, X)$ 空间中的非负定算子, 但离散时间转移核却不然, 这就给研究离散时间 Markov 链的收敛速度带来了许多困难, 详细介绍可参见文献 [1-4].

集合 $A \in \mathcal{B}(X)$ 称为细集, 是指存在概率分布 $a = (a_n)_{n \in \mathbb{Z}_+}$, 常数 $\delta > 0$, 以及概率测度 ν 满足

$$\sum_{n=0}^{\infty} a_n P^n(x, \cdot) \geq \delta \nu(\cdot), \quad x \in A.$$

由文献 [11, 定理 5.2.2] 知, 细集并不少见, 对于 ψ 不可约的 Markov 链, $\mathcal{B}^+(X)$ 中的任何集合都包含一个细集. 当概率分布 a 退化时, 上述不等式就转化为存在常数 $k, \delta > 0$, 以及概率测度 ν 使得对任意 $x \in A$, 有

$$P^k(x, \cdot) \geq \delta \nu(\cdot), \tag{1.1}$$

此时称集合 A 为小集. ψ 不可约的 Markov 链 Φ 称为非周期, 是指若对某个小集 $A \in \mathcal{B}^+(X)$, 存在 $n_0 \in \mathbb{Z}_+$ 使得对任意 $n \geq n_0$ 和 $x \in A$, 有 $P^n(x, A) > 0$. 由文献 [12] 知, 在非周期假设下, 每个细集均为小集. 若 (1.1) 中的 $k = 1$, 即

$$P(x, \cdot) \geq \delta \nu(\cdot), \quad x \in A, \tag{1.2}$$

则称集合 A 满足最小化条件. 若 $P(x, \cdot) = \nu(\cdot)$, $x \in A$, 则称集合 A 为原子. 显然, 当 X 为可数状态空间且 Markov 链不可约时, X 中的每个点都是原子.

Markov 链 Φ 称为几何遍历, 是指存在 $\rho < 1$, 对任意满足 $\mu \ll \pi$ 且 $d\mu/d\pi \in L^2(\pi, X)$ 的概率测度 μ , 都有 $C_\mu < \infty$ 使得

$$\|\mu P^n - \pi\|_{\text{Var}} \leq C_\mu \rho^n, \quad n \in \mathbb{Z}_+,$$

其中 $\|\cdot\|_{\text{Var}}$ 是任意符号测度 (如 ν) 的全变差范数, 定义为 $\|\nu\|_{\text{Var}} = \sup_{|f| \leq 1} |\nu(f)|$, 具体介绍可参见文献 [13, 14] 等. 由文献 [11, 定理 15.0.1] 可以得到如下结论, 该结论联系了 Markov 链首次回返时的几何阶矩、漂移 (drift) 条件和几何遍历性.

定理 1.1 假设 Markov 链 Φ 为 ψ 不可约且非周期, 则以下两条等价:

(1) 存在细集 $A \in \mathcal{B}(X)$ 和常数 $\kappa > 1$ 使得

$$\sup_{x \in A} \mathbb{E}_x[\kappa^{\tau_A}] < \infty; \quad (1.3)$$

(2) 存在细集 $A \in \mathcal{B}(X)$ 、常数 $\lambda < 1$ 和 $b < \infty$, 以及可测函数 $V \geq 1$ 使得

$$PV(x) \leq \lambda V(x) + b1_A(x), \quad x \in X. \quad (1.4)$$

进一步, 上述任意一条可推得 Markov 链 Φ 为几何遍历.

上述定理给出了几何遍历的定性刻画, 在此基础上, 文献 [4, 7, 8, 15] 等利用漂移条件 (1.4) 给出了几何遍历的可计算收敛速度. 本文的目的是, 采用谱分析理论和分裂技术, 利用首次回返时的矩条件 (1.3) 研究几何遍历的收敛速度, 主要结果如下:

定理 1.2 对于 ψ 不可约、非周期、可逆且非负定的 Markov 链 Φ , 假设存在集合 $A \in \mathcal{B}(X)$ 和常数 $\kappa > 1$ 使得

$$M := \sup_{x \in A} \mathbb{E}_x[\kappa^{\tau_A}] < \infty. \quad (1.5)$$

(1) 若集合 A 为原子, 则

$$\|\mu P^n - \pi\|_{\text{Var}} \leq \left\| \frac{d\mu}{d\pi} - 1 \right\|_{L^2(\pi, X)} \kappa^{-n}.$$

(2) 若集合 A 满足最小化条件 (1.2), 则

$$\|\mu P^n - \pi\|_{\text{Var}} \leq \left(\left\| \frac{d\mu_1}{d\pi_1} - 1 \right\|_{L^2(\pi_1, X)}^2 + \left\| \frac{d\mu_2}{d\pi_2} - 1 \right\|_{L^2(\pi_2, X)}^2 + 1 \right)^{1/2} K^{-n},$$

其中

$$K = \kappa \wedge (1 - \delta)^{-1/\alpha}, \quad \alpha = \frac{\log((M - \delta\kappa)/(1 - \delta))}{\log \kappa}, \quad (1.6)$$

$$\mu_1 = \delta\mu(A \cap \cdot), \quad \pi_1 = \delta\pi(A \cap \cdot),$$

$$\mu_2 = (1 - \delta)\mu(A \cap \cdot) + \mu(A^c \cap \cdot), \quad \pi_2 = (1 - \delta)\pi(A \cap \cdot) + \pi(A^c \cap \cdot).$$

本文余下内容的结构如下: 第 2 节证明定理 1.2. 第 3 节给出定理 1.2 的应用.

2 定理 1.2 的证明

为了证明定理 1.2, 我们需要做一些准备工作. 对于含有原子的一般状态空间中的 Markov 链, 它的很多理论和研究方法与可数状态空间中的情形类似. 然而, 在一般状态空间中, 原子是很少见的, 这就给研究一般状态空间中的 Markov 链带来了不便. 为了克服此困难, 我们可以先恰当地扩展 Markov 链的概率结构, 人为构造出含有原子的新 Markov 链, 从而采用与可数状态空间类似的方法进行分析, 然后再由新 Markov 链过渡到原 Markov 链, 详细介绍可参见文献 [16, 17], 本文将采用 Nummelin 分裂技术.

假设最小化条件 (1.2) 成立. 接下来将通过依次分裂状态空间 X 和 $\mathcal{B}(X)$ 上的测度以及转移核 P 来构造含有原子的分裂 Markov 链. 首先, 令 $\check{X} = X_0 \cup X_1$, 其中 $X_0 = X \times \{0\}$ 和 $X_1 = X \times \{1\}$ 均可认为是 X 的复制, 并分别赋予 σ 代数 $\mathcal{B}(X_0)$ 和 $\mathcal{B}(X_1)$. 令 $\mathcal{B}(\check{X})$ 是包含集合 $B_0 = B \times \{0\}$ 和 $B_1 = B \times \{1\}$ (其中 $B \in \mathcal{B}(X)$) 的最小 σ 代数. 记 x_i ($i = 0, 1$) 是 $\check{X}$ 中的点, 其中 $x_0 \in X_0$, $x_1 \in X_1$. 接着, 对于 $\mathcal{B}(X)$ 中的任意测度 μ , 定义 $\mathcal{B}(\check{X})$ 中的测度 μ^* 为

$$\mu^*(B_0) = (1 - \delta)\mu(A \cap B) + \mu(A^c \cap B), \quad (2.1)$$

$$\mu^*(B_1) = \delta\mu(A \cap B), \quad (2.2)$$

其中 A 和 δ 分别为最小化条件 (1.2) 中的集合和常数. 需要指出的是,

$$\mu^*(B_0 \cup B_1) = \mu(B), \quad B \in \mathcal{B}(X),$$

即 μ 是 μ^* 的边缘测度. 由于当 $B \subset A^c$ 时, $\mu^*(B_0) = \mu(B)$, 故只有 A 的子集才真正发生分裂. 给定 $x_i \in \check{X}$, 最后定义分裂的转移核 $\check{P}(x_i, \cdot)$ 如下:

$$\begin{aligned} \check{P}(x_0, \cdot) &= P(x, \cdot)^*, \quad x_0 \in X_0 \setminus A_0, \\ \check{P}(x_0, \cdot) &= (1 - \delta)^{-1}(P(x, \cdot)^* - \delta\nu^*(\cdot)), \quad x_0 \in A_0, \end{aligned} \quad (2.3)$$

$$\check{P}(x_1, \cdot) = \nu^*(\cdot), \quad x_1 \in X_1, \quad (2.4)$$

其中 A, δ 和 ν 分别为最小化条件 (1.2) 中的集合、常数和测度. 容易验证, 最小化条件 (1.2) 的成立保证了 (2.3) 总是非负的, 这是假设最小化条件成立的主要原因之一. 记 $\check{\Phi} = \{\check{\Phi}_n : n \in \mathbb{Z}_+\}$ 是以 $(\check{X}, \mathcal{B}(\check{X}))$ 为状态空间、以 $\check{P}$ 为转移核的分裂 Markov 链, 由 (2.4) 知, $\check{\Phi}$ 包含原子 X_1 , 且 $\psi^*(X_1) = \delta\psi(A)$. 由于对任意的 $n \in \mathbb{Z}_+$ 和 $x_i \in \check{X}$, 都有 $\check{P}^n(x_i, X_1 \setminus A_1) = 0$, 即无论从任何点出发都无法到达集合 $X_1 \setminus A_1$, 因此, 本质上集合 A_1 是分裂链 $\check{\Phi}$ 的原子.

Nummelin 分裂技术之所以有着广泛的应用, 不仅是因为新构造出的 Markov 链 $\check{\Phi}$ 包含原子, 更重要的是 Φ 与 $\check{\Phi}$ 之间有着密切的联系. 例如, 由文献 [11, 定理 5.1.3] 和 [11, 性质 10.4.1] 可以分别得到如下的命题 2.1 和 2.2.

命题 2.1 (1) Markov 链 Φ 是 $\check{\Phi}$ 的边缘链, 即对 $\mathcal{B}(X)$ 上的任意初始分布 μ 和任意集合 $B \in \mathcal{B}(X)$, 都有

$$\int_X P^n(x, B)\mu(dx) = \int_{\check{X}} \check{P}^n(x_i, B_0 \cup B_1)\mu^*(dx_i).$$

(2) 若 Markov 链 Φ 为 ψ 不可约且 $\psi(A) > 0$, 则 $\check{\Phi}$ 为 ν^* 不可约, 其中 A 和 ν 分别为最小化条件中的集合和测度; 反之, 若 Markov 链 $\check{\Phi}$ 为 ψ^* 不可约, 则 Φ 为 ψ 不可约.

命题 2.2 若 π 是 Markov 链 Φ 的平稳分布, 则 π^* 是 $\check{\Phi}$ 的平稳分布; 反之, 若 $\check{\pi}$ 是 Markov 链 $\check{\Phi}$ 的平稳分布, 定义

$$\pi(B) = \check{\pi}(B_0 \cup B_1), \quad B \in \mathcal{B}(X),$$

则 π 是 Φ 的平稳分布, 且 $\check{\pi} = \pi^*$.

令 $\check{P}_{x_i}$ 和 $\check{E}_{x_i}$ 分别为给定 $\check{\Phi}_0 = x_i$ 的条件下分裂链的条件分布和条件期望. 由于 A_1 为分裂链的原子, 故对任意的 $x_1 \in A_1$, 可简记 $\check{P}_{A_1} = \check{P}_{x_1}$ 和 $\check{E}_{A_1} = \check{E}_{x_1}$. 定义

$$\check{E}_x = (1 - \delta 1_A(x))\check{E}_{x_0} + \delta 1_A(x)\check{E}_{x_1}, \quad (2.5)$$

显然, 在 $\mathcal{B}(X)$ 上, $\check{E}_x$ 和 E_x 相同. 令

$$\tau_{A_1} = \inf\{n \geq 1 : \check{\Phi}_n \in A_1\}$$

为集合 A_1 的首次回返时, 结合文献 [8, 定理 2.1] (或文献 [4, 引理 A.1]) 和定理 1.1 可以得到如下引理, 该结果表明了 τ_A 与 τ_{A_1} 的几何阶矩之间的关系.

引理 2.1 假设对某集合 $A \in \mathcal{B}(X)$, 条件 (1.2) 和 (1.5) 成立, 则有

$$\check{E}_{A_1}[r^{\tau_{A_1}}] \leq \frac{\delta \check{E}_{A_1}[r^{\tau_A}]}{1 - (1 - \delta) \sup_{x \in A} \check{E}_{x_0}[r^{\tau_A}]},$$

其中常数 $r \in (1, \kappa)$ 满足

$$(1 - \delta) \sup_{x \in A} \check{E}_{x_0}[r^{\tau_A}] < 1.$$

根据上述引理, 我们可以得到 τ_{A_1} 的几何阶矩的估计.

引理 2.2 假设对某集合 $A \in \mathcal{B}(X)$, 条件 (1.2) 和 (1.5) 成立, 则对常数 $r \in (1, \kappa \wedge (1 - \delta)^{-1/\alpha})$, 有

$$\check{E}_{A_1}[r^{\tau_{A_1}}] \leq \frac{\delta r^\beta}{1 - (1 - \delta)r^\alpha},$$

其中 α 如 (1.6) 所示,

$$\beta = \frac{\log((M - (1 - \delta)\kappa)/\delta)}{\log \kappa}.$$

证明 由引理 2.1 知, 我们需要分别估计 $\sup_{x \in A} \check{E}_{x_0}[r^{\tau_A}]$ 和 $\check{E}_{A_1}[r^{\tau_A}]$. 一方面, 由 (2.5) 注意到

$$E_x[r^{\tau_A}] = \check{E}_x[r^{\tau_A}] = (1 - \delta)\check{E}_{x_0}[r^{\tau_A}] + \delta\check{E}_{A_1}[r^{\tau_A}], \quad x \in A, \quad (2.6)$$

故由 Jensen 不等式、(2.6) 和 (1.5) 知, 对任意 $r \in (1, \kappa]$, 有

$$\begin{aligned} \sup_{x \in A} \check{E}_{x_0}[r^{\tau_A}] &\leq \sup_{x \in A} (\check{E}_{x_0}[\kappa^{\tau_A}])^{(\log r)/(\log \kappa)} \\ &= \sup_{x \in A} \left(\frac{E_x[\kappa^{\tau_A}] - \delta \check{E}_{A_1}[\kappa^{\tau_A}]}{1 - \delta} \right)^{(\log r)/(\log \kappa)} \\ &\leq \left(\frac{M - \delta \kappa}{1 - \delta} \right)^{(\log r)/(\log \kappa)} \end{aligned}$$

$$= r^\alpha. \quad (2.7)$$

进而, 对任意 $r \in (1, \kappa \wedge (1 - \delta)^{-1/\alpha})$, 有

$$(1 - \delta) \sup_{x \in A} \check{E}_{x_0}[r^{\tau_A}] \leq (1 - \delta)r^\alpha < 1. \quad (2.8)$$

另一方面, 同理可知, 对任意 $r \in (1, \kappa]$, 有

$$\begin{aligned} \check{E}_{A_1}[r^{\tau_A}] &\leq (\check{E}_{A_1}[\kappa^{\tau_A}])^{(\log r)/(\log \kappa)} \\ &= \left(\frac{E_x[\kappa^{\tau_A}] - (1 - \delta)\check{E}_{x_0}[\kappa^{\tau_A}]}{\delta} \right)^{(\log r)/(\log \kappa)} \\ &\leq \left(\frac{M - (1 - \delta)\kappa}{\delta} \right)^{(\log r)/(\log \kappa)} \\ &= r^\beta. \end{aligned} \quad (2.9)$$

于是, 由引理 2.1 和 (2.7)–(2.9) 便证得所需结论. $\square$

现在, 我们来证明本文的主要结果.

定理 1.2 的证明 (1) 首先, 研究转移核 P^n 的 L^2 几何收敛速度. 记 P_D 为 P 限制在状态空间 $X \setminus A$ 时对应的转移核, 由 P 的可逆性易知, P_D 为 $L^2(\pi, X \setminus A)$ 空间中的自伴算子. 注意到对任意 $x \in X \setminus A$, 有

$$(P_D^n 1)(x) = P_x\{\tau_A > n\} \leq \kappa^{-(n+1)} E_x[\kappa^{\tau_A}], \quad \kappa > 1,$$

且 Kac 公式^[18]

$$\int_A E_x[\kappa^{\tau_A}] \pi(dx) = \kappa \pi(A) + (\kappa - 1) \int_{X \setminus A} E_x[\kappa^{\tau_A}] \pi(dx)$$

成立, 故存在常数 $C < \infty$ 使得

$$\begin{aligned} |(f, P_D^n f)_{L^2(\pi, X \setminus A)}| &\leq (|f|, P_D^n |f|)_{L^2(\pi, X \setminus A)} \\ &\leq \|f\|_\infty^2 \int_{X \setminus A} (P_D^n 1)(x) \pi(dx) \\ &\leq \kappa^{-(n+1)} \|f\|_\infty^2 \int_{X \setminus A} E_x[\kappa^{\tau_A}] \pi(dx) \\ &= \kappa^{-(n+1)} \|f\|_\infty^2 (\kappa - 1)^{-1} \left(\int_A E_x[\kappa^{\tau_A}] \pi(dx) - \kappa \pi(A) \right) \\ &\leq \kappa^{-(n+1)} \|f\|_\infty^2 (\kappa - 1)^{-1} (M \pi(A) - \kappa \pi(A)) \\ &\leq C \kappa^{-(n+1)}, \end{aligned} \quad (2.10)$$

其中 $f \in L^\infty(\pi, X \setminus A)$ 满足 $\|f\|_\infty = \sup_{x \in X \setminus A} |f(x)|$. 由于 $L^\infty(\pi, X \setminus A)$ 在 $L^2(\pi, X \setminus A)$ 中稠密, 由 (2.10) 和文献 [3, 性质 2.5] 知, $\|P_D\|_{L^2(\pi, X \setminus A)} \leq \kappa^{-1}$. 进一步, 由于集合 A 为原子, 根据文献 [3, 引理 3.12] 可得, P 在 $\{f \in L^2(\pi, X) : \pi(f) = 0\}$ 中的谱半径 $r(P)$ 满足

$$r(P) \leq \|P_D\|_{L^2(\pi, X \setminus A)} \leq \kappa^{-1}.$$

于是, 对于非负定的转移核 P , 根据自伴算子的谱映射定理知,

$$\|P^n f - \pi(f)\|_{L^2(\pi, X)} \leq \|f - \pi(f)\|_{L^2(\pi, X)} \kappa^{-n}.$$

接下来, 由转移核 P^n 的 L^2 几何收敛过渡到几何遍历. 由文献 [5, 定理 9.15] 的证明可得, 对任意满足 $\mu \ll \pi$ 且 $d\mu/d\pi \in L^2(\pi, X)$ 的概率测度 μ , 有

$$\begin{aligned} \|\mu P^n - \pi\|_{\text{var}} &= \sup_{|f| \leq 1} |(\mu P^n - \pi)f| = \sup_{|f| \leq 1} \left| \pi \left(\frac{d\mu}{d\pi} P^n f - f \right) \right| \\ &= \sup_{|f| \leq 1} \left| \pi \left(f P^n \left(\frac{d\mu}{d\pi} \right) - f \right) \right| \\ &\leq \left\| P^n \left(\frac{d\mu}{d\pi} - 1 \right) \right\|_{L^1(\pi, X)} \\ &\leq \left\| P^n \left(\frac{d\mu}{d\pi} - 1 \right) \right\|_{L^2(\pi, X)} \\ &\leq \left\| \frac{d\mu}{d\pi} - 1 \right\|_{L^2(\pi, X)} \kappa^{-n}, \end{aligned}$$

其中第三个等号成立是由于 Markov 链具有可逆性, 最后一个不等号成立是利用了 L^2 几何收敛.

(2) 首先研究含有原子 A_1 的分裂 Markov 链 $\check{\Phi}$ 的几何遍历速度. 由命题 2.1 和 2.2 知, 分裂链 $\check{\Phi}$ 为 ν^* 不可约、非周期、关于 π^* 可逆且非负定. 记 $\check{P}_D$ 为 $\check{P}$ 限制在状态空间 $\check{X} \setminus A_1$ 时对应的转移核, 由与 (1) 类似的证明和引理 2.2 知, 对任意常数 $r \in (1, \kappa \wedge (1 - \delta)^{-1/\alpha})$, 存在常数 $\check{C} < \infty$ 使得

$$\begin{aligned} |(\check{f}, \check{P}_D^n \check{f})_{L^2(\pi^*, \check{X} \setminus A_1)}| &\leq r^{-(n+1)} \|\check{f}\|_{\infty}^2 (r-1)^{-1} \left(\int_{A_1} \check{E}_{x_i} [r^{\tau_{A_1}}] \pi^*(dx_i) - r\pi^*(A_1) \right) \\ &= r^{-(n+1)} \|\check{f}\|_{\infty}^2 (r-1)^{-1} (\check{E}_{A_1} [r^{\tau_{A_1}}] \pi^*(A_1) - r\pi^*(A_1)) \\ &\leq r^{-(n+1)} \|\check{f}\|_{\infty}^2 (r-1)^{-1} \delta \pi(A) \left(\frac{\delta r^{\beta}}{1 - (1 - \delta)r^{\alpha}} - r \right) \\ &\leq \check{C} r^{-(n+1)}, \end{aligned}$$

其中 $\check{f} \in L^2(\pi^*, \check{X} \setminus A_1)$, 定义为 $\check{f}(x_i) = f(x)$, $i = 0, 1$, 且 $\|\check{f}\|_{\infty} = \sup_{x_i \in \check{X} \setminus A_1} |\check{f}(x_i)|$, 故对于具有原子 A_1 的转移核 $\check{P}$, 其在 $\{\check{f} \in L^2(\pi^*, \check{X}) : \pi^*(\check{f}) = 0\}$ 中的谱半径 $r(\check{P})$ 满足

$$r(\check{P}) \leq r^{-1}, \quad r \in (1, \kappa \wedge (1 - \delta)^{-1/\alpha}),$$

因此, $r(\check{P}) \leq \kappa^{-1} \vee (1 - \delta)^{1/\alpha} = K^{-1}$. 于是, 同 (1) 的证明可知, 对任意满足 $\mu^* \ll \pi^*$ 且 $d\mu^*/d\pi^* \in L^2(\pi^*, \check{X})$ 的概率测度 μ^* , 有

$$\|\mu^* \check{P}^n - \pi^*\|_{\text{var}} \leq \left\| \frac{d\mu^*}{d\pi^*} - 1 \right\|_{L^2(\pi^*, \check{X})} K^{-n}. \quad (2.11)$$

接下来由分裂 Markov 链过渡到原 Markov 链. 一方面, 由命题 2.1 和 2.2 分别注意到 Markov 链 Φ 是 $\check{\Phi}$ 的边缘链, π 是 π^* 的边缘测度, 故

$$\|\mu P^n - \pi\|_{\text{var}} = \sup_{B \in \mathcal{B}(X)} \left| \int_X P^n(x, B) \mu(dx) - \pi(B) \right|$$

$$\begin{aligned}
 &= \sup_{B \in \mathcal{B}(X)} \left| \int_{\check{X}} \check{P}^n(x_i, B_0 \cup B_1) \mu^*(dx_i) - \pi^*(B_0 \cup B_1) \right| \\
 &\leq \sup_{\check{B} \in \mathcal{B}(\check{X})} \left| \int_{\check{X}} \check{P}^n(x_i, \check{B}) \mu^*(dx_i) - \pi^*(\check{B}) \right| \\
 &= \|\mu^* \check{P}^n - \pi^*\|_{\text{Var}}.
 \end{aligned} \tag{2.12}$$

另一方面, 由分裂 Markov 链的测度定义 (2.1) 和 (2.2) 知,

$$\begin{aligned}
 \left\| \frac{d\mu^*}{d\pi^*} - 1 \right\|_{L^2(\pi^*, \check{X})}^2 &= \int_{\check{X}} \frac{\mu^*(dx_i)^2}{\pi^*(dx_i)} - 1 \\
 &= \int_{X_1} \frac{\mu^*(dx_i)^2}{\pi^*(dx_i)} + \int_{X_0} \frac{\mu^*(dx_i)^2}{\pi^*(dx_i)} - 1 \\
 &= \int_X \frac{(\delta\mu(A \cap dx))^2}{\delta\pi(A \cap dx)} + \int_X \frac{((1-\delta)\mu(A \cap dx) + \mu(A^c \cap dx))^2}{(1-\delta)\pi(A \cap dx) + \pi(A^c \cap dx)} - 1 \\
 &= \int_X \frac{\mu_1(dx)^2}{\pi_1(dx)} + \int_X \frac{\mu_2(dx)^2}{\pi_2(dx)} - 1 \\
 &= \left\| \frac{d\mu_1}{d\pi_1} - 1 \right\|_{L^2(\pi_1, X)}^2 + \left\| \frac{d\mu_2}{d\pi_2} - 1 \right\|_{L^2(\pi_2, X)}^2 + 1.
 \end{aligned} \tag{2.13}$$

于是, 结合 (2.11)–(2.13) 便得到了我们需要的结果. $\square$

3 定理 1.2 的应用

本节将应用定理 1.2 来讨论 Metropolis-Hastings 算法的几何遍历速度. Metropolis 等^[19] 在 1953 年提出了一种构造转移核的方法, Hastings^[20] 随后对之加以推广, 形成了 Metropolis-Hastings 算法. 该算法是一种常见的 Markov 链的 Monte Carlo (MCMC) 算法, 在 Bayes 统计、显著性检验和极大似然估计等方面都有广泛的应用.

设 μ^{Leb} 为 $\mathbb{R}$ 上的 Lebesgue 测度, 给定 $\mathbb{R}$ 上的后验分布 $\pi(x)$, 并由此定义概率测度

$$\pi(dx) = \pi(x) \mu^{\text{Leb}}(dx).$$

设 $Q(x, \cdot)$ 为任意的建议转移核, 其关于 Lebesgue 测度 μ^{Leb} 绝对连续, 概率密度函数记为 $q(x, \cdot)$. Metropolis-Hastings 算法按照如下的转移核 P 进行实施:

$$\begin{aligned}
 P(x, dy) &= q(x, y) \alpha(x, y) \mu^{\text{Leb}}(dy), \quad y \neq x, \\
 P(x, \{x\}) &= \int q(x, y) [1 - \alpha(x, y)] \mu^{\text{Leb}}(dy) + Q(x, \{x\}),
 \end{aligned}$$

其中

$$\alpha(x, y) = \begin{cases} \min \left\{ \frac{\pi(y) q(y, x)}{\pi(x) q(x, y)}, 1 \right\}, & \pi(x) q(x, y) > 0, \\ 1, & \pi(x) q(x, y) = 0. \end{cases}$$

记以 P 为转移核的 Metropolis-Hastings 链为 $\Phi = \{\Phi_n : n \in \mathbb{Z}_+\}$, 上述 $\alpha(x, y)$ 的选取保证了 π 是 Φ 的平稳分布.

本节考虑当 π 服从标准正态分布 $N(0, 1)$ 且 $Q(x, \cdot)$ 服从正态分布 $N(x, 1)$ 时对应的 Metropolis-Hastings 链的几何遍历速度, 关于此问题的详细介绍可参见文献 [4, 7, 8, 21] 等及其中的参考文献. 由文献 [21, 引理 1.1 和 1.2] 知, Metropolis-Hastings 链为 π 不可约、非周期, 且每个满足 $\mu^{\text{Leb}}(A) > 0$ 的紧集 A 均为小集. 根据文献 [4, 引理 3.1] 知, Metropolis-Hastings 链可逆且非负定. 令 $A = [-d, d]$, 则最小化条件 (1.2) 成立, 其中

$$\nu(dx) = c \cdot \exp(-x^2) 1_A(x) \mu^{\text{Leb}}(dx),$$

$$\delta = \sqrt{2} \exp(-d^2) \left[F(\sqrt{2}d) - \frac{1}{2} \right],$$

c 为标准化常数, F 为标准正态分布的分布函数. 取最优常数 $d = 1.1$, 则 $\delta = 0.186$. 进一步, 由文献 [4, 性质 4.1] 和初等计算可得, 存在 $\kappa = 1.024$ 使得

$$\sup_{x \in A} \mathbb{E}_x[\kappa^{\tau_A}] \leq 1.192.$$

于是, 由定理 1.2(2) 计算得几何遍历速度 $K^{-1} = 0.977$.

致谢 作者感谢审稿人对本文初稿提出的宝贵意见, 并感谢毛永华教授对本文提供的有益帮助. 通过与毛永华教授进行相关问题的讨论, 促进了作者对本文的思考与研究.

参考文献

- 1 Mao Y H. Convergence rates for reversible Markov chains without the assumption of nonnegative definite matrices. *Sci China Math*, 2010, 53: 1979–1988
- 2 Mao Y H, Song Y H. Spectral gap and convergence rate for discrete-time Markov chains. *Acta Math Sin Engl Ser*, 2013, 29: 1949–1962
- 3 Sokal A D, Thomas L E. Exponential convergence to equilibrium for a class of random-walk models. *J Stat Phys*, 1989, 54: 797–828
- 4 Baxendale P H. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann Appl Probab*, 2005, 15: 700–738
- 5 Chen M F. *From Markov Chains to Non-Equilibrium Particle Systems*, 2nd ed. Singapore: World Scientific, 2004
- 6 Chen M F. *Eigenvalues, Inequalities and Ergodic Theory*. New York: Springer-Verlag, 2005
- 7 Meyn S P, Tweedie R L. Computable bounds for geometric convergence rates of Markov chains. *Ann Appl Probab*, 1994, 4: 981–1011
- 8 Roberts G O, Tweedie R L. Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Process Appl*, 1999, 80: 211–229
- 9 宋延红. 离散时间 Markov 链在全变差距离下的收敛速度. *中国科学: 数学*, 2015, 45: 819–828
- 10 Wang F Y. *Functional Inequalities, Markov Semigroups and Spectral Theory*. Beijing: Science Press, 2005
- 11 Meyn S P, Tweedie R L. *Markov Chains and Stochastic Stability*. London: Springer-Verlag, 1993
- 12 Meyn S P, Tweedie R L. Stability of Markovian processes II: Continuous-time processes and sampled chains. *Adv Appl Probab*, 1993, 25: 487–517
- 13 Chen M F. Equivalence of exponential ergodicity and L^2 -exponential convergence for Markov chains. *Stochastic Process Appl*, 2000, 87: 281–297
- 14 Roberts G O, Rosenthal J S. Geometric ergodicity and hybrid Markov chains. *Electron Commun Probab*, 1997, 2: 13–25
- 15 Roberts G O, Rosenthal J S. Shift-coupling and convergence rates of ergodic averages. *Comm Statist Stoch Models*, 2007, 13: 147–165
- 16 Nummelin E. A splitting technique for Harris recurrent Markov chains. *Z Wahrscheinlichkeitstheorie Verw Gebiete*, 1978, 43: 309–318
- 17 Athreya K B, Ney P. A new approach to the limit theory of recurrent Markov chains. *Trans Amer Math Soc*, 1978, 245: 493–501
- 18 Cogburn R. A uniform theory for sums of Markov chain transition probabilities. *Ann Probab*, 1975, 3: 191–214

- 19 Metropolis N, Rosenbluth A W, Rosenbluth M N, et al. Equation of state calculations by fast computing machines. *J Chem Phys*, 1953, 21: 1087–1092
- 20 Hastings W K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 1970, 57: 97–109
- 21 Mengersen K L, Tweedie R L. Rates of convergence of the Hastings and Metropolis algorithms. *Ann Statist*, 1996, 24: 101–121

Convergence rates in geometric ergodicity for discrete-time Markov chains

Yanhong Song

Abstract In this paper, we investigate the convergence rates in geometric ergodicity for discrete-time Markov chains on general state spaces. When the state space contains an atom, we study the convergence rates by the spectral theory, which is similar to that on countable state spaces. However, atoms are less frequent. For the general case, we first use the Nummelin splitting technique to construct a new Markov chain which contains an atom, and then establish the connection of the geometrically ergodic rates between the new and the original Markov chains. We also obtain the convergence rate in geometric ergodicity for the original Markov chain by the minorization condition and the geometric moment of the first return time. Moreover, we apply the main result to study the Metropolis-Hastings algorithm.

Keywords Markov chain, geometric ergodicity, convergence rate

MSC(2010) 60J05, 37A30

doi: 10.1360/N012018-00291