中国科学: 数学 2020 年 第 50 卷 第 1 期: 101 ～ 118
SCIENTIA SINICA Mathematica

论 文

Markov 链: 遍历性、拟平稳性与不可逆性

献给严士健教授 90 华诞

毛永华

北京师范大学数学科学学院, 北京 100875
E-mail: maoyh@bnu.edu.cn

收稿日期: 2019-03-03; 接受日期: 2019-08-20; 网络出版日期: 2019-11-20
国家自然科学基金 (批准号: 11571043 和 11771047) 资助项目

摘要 本文以首中时 (或回返时) 为脉络, 从三个方面—Markov 链的遍历性、拟平稳分布和不可逆问题—介绍 Markov 链研究的一些最新进展. 这些内容包括: (1) 以首中时的矩给出泛函不等式; (2) 引入修正的回返时判定各种非常返性; (3) 用回返时处理离散时间 Markov 链的泛函不等式; (4) Markov 链首中时的分布表示; (5) 以击中时的矩判定一族遍历的 Markov 过程收敛到平稳分布所产生切断 (cutoff) 现象; (6) 从 Markov 链生命时的分布找到拟平稳分布存在唯一性; (7) 发展 Dirichlet 原理来判定不可逆 Markov 链收敛到平稳分布 "优于" 相应的可逆过程的问题.

关键词 Markov 链 击中时 回返时 平稳性 拟平稳性 切断 非对称性

MSC (2010) 主题分类 60B10, 60J27, 60G20

1 引言

Markov 链自 20 世纪末以来呈现复兴与发展的趋势, 特别是有限 Markov 链, 在 Markov 链 Monte Carlo 算法中的使用, 是其发展的主要动力之一. 例如, 一族 (有限) Markov 链在收敛到平稳分布的过程会发生所谓 "切断" 现象, 对 Markov 链 Monte Carlo 算法产生非常重要的影响.

拟平稳分布 (quasi-stationary distribution, QSD) 是一个经典问题, 而且贯穿整个 Markov 过程的发展. 从经典关于分支过程的 Yaglom 极限定理, 到有限 Markov 链、可数 Markov 链、扩散过程和 Lévy 过程等, 都有丰富的结果, 也有许多待解决的问题.

Markov 过程的定量研究, 特别是收敛速度的估计, 是近二三十年来在 Markov 过程的研究中的巨大进步. 在此过程中, 在一些经典的工具之外, 同时发展了一批新的方法, 如耦合、对偶和泛函不等式等. 但是不得不遗憾地指出, 这些方法更多地 (如果不是绝大多数) 依赖于 Markov 的对称性. 对于非对称的 Markov 过程, 由于方法的局限性, 其研究进展相对缓慢.

英文引用格式: Mao Y H. Markov chains: Ergodicity, quasi-stationarity and asymmetry (in Chinese). Sci Sin Math, 2020, 50: 101–118, doi: 10.1360/N012019-00069

在 Markov 过程研究中, 最基本的是 Markov 链, 即状态空间离散的 Markov 过程, 而其中最核心的概念, 当属首中时或回返时. 我们将从首中时 (回返时) 出发, 串联起 Markov 链研究的三个活跃领域—遍历性、拟平稳性和非对称性. 具体而言, 我们将涉及以下内容:

(1) 遍历性: 包括以飞跃时定义唯一性, 以回返时定义常返性和各种遍历性, 特别对于可逆 Markov 链, 以首中时的各阶矩给出各种遍历性的收敛速度;

(2) 非常返性: 如同遍历性情形, 以修正的回返时, 给出非常返 Markov 链的 ℓ 非常返性和几何 (指数) 非常返性等多种非常返性的定义和判断准则;

(3) 离散时间 Markov 链: 利用回返时的各阶矩, 由相应的泛函不等式给出离散时间 Markov 链的遍历性的定量估计;

(4) 首中时分布: 首中时的分布与相应过程的转移矩阵的谱紧密相连, 但是已有的结论基本针对有限的可逆 Markov 链;

(5) 切断现象: 首中时分布可以用于判定一族 Markov 链在收敛过程中是否产生所谓的切断现象;

(6) 拟平稳分布: 如果把 Markov 链的生命时也作为一个首中时看待, 那么生命时的分布有助于刻画相应最小过程的拟平稳分布;

(7) 非对称性: 以首中时构造的混合时可以判断一般 (非对称) Markov 链较其相应的对称的 Markov 链的某种优越性, 我们的方法是通过给出一些混合时的 Dirichlet 变分公式来作出判定.

设离散时间或连续时间 Markov 链 X_t ($t \in T = \mathbb{Z}_+$ 或 $\mathbb{R}_+$) 具有 (一步) 转移概率矩阵 $P = (p_{ij})$ 或生成元 $Q = (q_{ij})$. 其状态空间 E 为有限的或可数的. 令 $P(n) = P^n = (p_{ij}^{(n)})$ 或 $P(t) = \mathrm{e}^{tQ} = (p_{ij}(t))$. 下面假定 Markov 链是不可约的, 在离散时间还假定链是非周期的.

对状态空间 E 的非空子集 A, 定义首中时

$$\tau_A = \inf\{t \geqslant 0 : X_t \in A\}$$

和回返时

$$\tau_A^+ = \inf\{t \geqslant \text{第一次跳的时刻} : X_t \in A\}.$$

如果 A 是单点集 $\{j\}$, 则将 τ_A 和 τ_A^+ 简记为 τ_j 和 τ_j^+.

2 遍历性

定义生命时 ζ 为 Markov 链无穷次跳的时刻, 则过程非爆炸 (唯一) 当且仅当 $\mathrm{P}_i[\zeta < \infty] = 0$. 由于离散时间 Markov 链跳一次就是一个单位时间, 所以, 爆炸只可能发生于连续时间 Markov 链. 而链常返当且仅当 $\mathrm{P}_j[\tau_j^+ = \infty] = 0$, 所以, 常返链必然非爆炸. 进一步利用回返时的矩可以定义不同的遍历性. 首先是 (普通) 遍历性:

$$\mathrm{E}_j\tau_j^+ < \infty \quad \Leftrightarrow \quad p_{ij}(t) - \pi_j = o(1),$$

其中 $\pi_j = 1/\mathrm{E}_j\tau_j^+$ (离散时间) 或 $\pi_j = 1/q_j\mathrm{E}_j\tau_j^+$ (连续时间).

更高阶的矩蕴涵更快的收敛速度. 令 $\ell = 1, 2, \ldots$, 则 (参见文献 [1,2])

$$\mathrm{E}_j(\tau_j^+)^\ell < \infty \quad \Rightarrow \quad p_{ij}(t) - \pi_j = o(t^{1-\ell}).$$

更经典的是指数阶矩有限与指数遍历性的关系. 链具有指数遍历性, 即存在 $\lambda > 0$ 使得

$$p_{ij}(t) - \pi_j = O(\mathrm{e}^{-\lambda t})$$

当且仅当存在 $\delta > 0$ 使得 $\mathrm{E}_j \mathrm{e}^{\delta \tau_j^+} < \infty$.

而链具有强遍历性:

$$\sup_{i\in E} \sum_{j\in E} |p_{ij}(t) - \pi_j| \to 0$$

当且仅当 $\sup_i \mathrm{E}\tau_j^+ < \infty$. 注意由 $P(t)$ 的半群性质知, 上式等价于存在 $\lambda > 0$ 和 $C < \infty$ 使得

$$\sup_{i\in E} \sum_{j\in E} |p_{ij}(t) - \pi_j| \leqslant C\mathrm{e}^{-\lambda t}.$$

针对不同的遍历性, 特别是指数遍历性和强遍历性, 在相应的矩有限的情形下给出收敛速度的估计是十分迫切而具挑战性的工作.

对于连续时间可逆的 Markov 链, 我们可以引入泛函不等式作出定量的估计. 之所以强调可逆性和连续时间, 是因为此时指数遍历性中的最大收敛速度就是相应的生成元的谱隙, 即 Q 矩阵可视为 $L^2(\pi)$ 中的自伴算子. 这样就可以使用刻画谱隙的 Poincaré 不等式.

设连续时间不可约可逆的 Markov 链 $(X_t)_{t\geqslant 0}$ 具有平稳分布 π, 其转移半群 $P(t)$ 是 Hilbert 空间 $L^2(\pi) := \{f : E \to \mathbb{R} : \pi(f^2) < \infty\}$ 上的压缩自伴算子, 其相应的生成元即 Q 矩阵也可以看作 $L^2(\pi)$ 上的自伴算子 (可能无界).

先来看指数遍历性. 定义指数收敛速度

$$\lambda_1 = -\lim_{t\to\infty} \frac{1}{t} \log |p_{ij}(t) - \pi_j|,$$

与 i 和 j 无关 (由不可约性). 此时, 上述 λ_1 就是 $L^2(\pi)$ 指数收敛速度, 而后者可由经典的 Poincaré 不等式刻画:

$$\lambda_1 = \inf\{D(f,f) : \pi(f) = 0, \pi(f^2) = 1\},$$

其中 D 为 Dirichlet 型 (参见文献 [3, 第 9 章])

$$D(f,f) = \frac{1}{2} \sum_{ij} \pi_i q_{ij} (f_j - f_i)^2. \tag{2.1}$$

其次, 利用分裂技巧将 λ_1 与以有限子集为吸收边界的子 Markov 过程的衰减速度联系起来. 任取非空有限集 H, 并记 $M_H = \sup_{i\in H} \sum_{j\in H^c} q_{ij}$, 定义 H 上的局部 Dirichlet 型:

$$D_H(f) = \frac{1}{2} \int_{H\times H} \pi_H(dx) q(x,dy) (f(x) - f(y))^2,$$
$$\mathcal{D}(D_H) = \{f \in L^2(\pi_H) : D_H(f) < \infty\},$$

其中 $\pi_H(dx) = \pi(dx)/\pi(H)$ 是 π 在 H 上诱导的概率. 同时定义其谱隙为

$$\lambda_1(H) = \inf\{D_H(f) : \pi_H(f) = 0, \pi_H(f^2) = 1\},$$

再令 $\lambda_0(H^c)$ 为过程限制于 H^c 上的 Dirichlet 特征值 (以 H 为吸收边界):

$$\lambda_0(H^c) = \inf\{D(f) : \pi(f^2) = 1, f\,|_H = 0\}.$$

文献 [4] 给出了如下漂亮的结论. 对任一 $H \subset G$ 满足 $0 < \pi(H) < 1$ 和 $M_H < \infty$ 都有判别准则

$$\frac{\lambda_0(H^c)}{\pi(H)} \geqslant \lambda_1 \geqslant \frac{\lambda_1(G)[\lambda_0(H^c)\pi(G) - 2M_H\pi(G^c)]}{2\lambda_1(G) + \pi(G)^2[\lambda_0(H^c) + 2M_H]}. \tag{2.2}$$

从而, $\lambda_1 > 0$ 当且仅当对任一有限的 H, 有 $\lambda_0(H^c) > 0$. 对可逆的跳过程, 有如下的判断准则 (参见文献 [5]).

定理 2.1 对任一有限非空集合 H, 都有

$$\lambda_1 \geqslant \frac{2\lambda_1(H)\lambda_0(H^c)}{\lambda_1(H) + \lambda_0(H^c) + M_H + \sqrt{(\lambda_1(H) + \lambda_0(H^c) + M_H)^2 - 4\lambda_1(H)\lambda_0(H^c)}}.$$

从而, $\lambda_1 > 0$ 当且仅当对任一有限非空的集合 H, 有 $\lambda_0(H^c) > 0$.

最后, 利用首中时的指数矩给出 $\lambda_0(H^c)$ 的下界. 设 $\forall i \in H^c$, $\exists \lambda > 0$ 使得

$$\mathrm{E}_i \mathrm{e}^{\lambda\tau_H} < \infty, \tag{2.3}$$

则 $\lambda_0(H^c) \geqslant \lambda$. 从而给出谱隙的下界估计, 并进一步得到如下的判定准则: 对于可逆跳过程, 指数遍历当且仅当对某个 (从而对所有) 有限非空子集 H, (2.3) 成立. 值得注意的是, 对于不可逆的跳过程, 上述判别准则中充分部分是不正确的, 参见文献 [6] 中的反例.

再来看强遍历性. 定义强遍历收敛速度:

$$\alpha = -\lim_{t\to\infty} \frac{1}{t} \log \sup_i \sum_j |p_{ij}(t) - \pi_j|.$$

众所周知, $\alpha > 0$ 当且仅当 $M := \sup_i \mathrm{E}_i\tau_0 < \infty$. 注意到, $\forall \epsilon < 1/M$, 有 $\sup_i \mathrm{E}_i\mathrm{e}^{\epsilon\tau_0} < \infty$.

定理 2.2 对于不可约的可逆跳过程, 有 $\alpha \geqslant 1/M$.

上述估计依赖于如下的重整化方法.

引理 2.1 给定状态 $0 \in E$, 对任意 i, 有

$$\sum_j |p_{ij}(t) - \pi_j| \leqslant 2\mathrm{P}_i[\tau_0 > t] + \int_0^t \sum_j |p_{0j}(t-s) - \pi_j| d\mathrm{P}_i[\tau_0 \leqslant s].$$

上述重整化方法的要点在于, 关于转移概率出发点 i 的一致性转换到首中时的分布上, 而有 τ_0 的指数矩有限保证了其尾概率指数速度下降. 再使用 Poincaré 不等式, 由首中时的指数矩得到偏差 $|p_{0j}(t-s) - \pi_j|$ 有同样的指数阶速度.

由此, 一个具有挑战性的问题是, 如何将上述重整化方法拓展到非单点集, 从而可以处理更一般的 Markov 过程, 参见文献 [7] 中的相关讨论.

3 非常返性

本节仅考虑离散时间 Markov 链, 这里只假定不可约性, 而不要求非周期性.

Markov 链非常返当且仅当 $f_{jj} := \mathrm{P}_j[\tau_j^+ < \infty] < 1$, 此时有

$$\lim_{n\to\infty} p_{ij}^{(n)} = 0.$$

如同遍历情形, 我们也关心上述收敛的各种速度, 以及相应的判别准则. 为此, 引入修正的回返时

$$\tilde{\tau}_j^+ = \tau_j^+ 1_{\{\tau_j^+ < \infty\}},$$

以此取代原来的回返时 τ_j^+ 的地位.

令 $\ell = 1, 2, \dots$ 若

$$\mathrm{E}_j (\tilde{\tau}_j^+)^\ell < 1,$$

则称链 ℓ 阶非常返; 若

$$\exists\, \kappa > 1, \quad \mathrm{E}_j [\kappa^{\tau_j^+} 1_{\{\tau_j^+ < \infty\}}] < 1, \tag{3.1}$$

则称链指数非常返.

上述定义源于以下的基本分解引理. 记 $f_{ij}^{(m)} = \mathrm{P}_i[\tau_j^+ = m]$, $m = 1, 2, \dots$, 则

$$p_{ij}^{(n)} = \sum_{m=1}^{n} f_{ij}^{(m)} p_{jj}^{(n-m)}.$$

若再记

$$P_{ij}(s) = \sum_{n=0}^{\infty} s^n p_{ij}^{(n)}, \quad F_{ij}(s) = \sum_{m=1}^{\infty} s^m f_{ij}^{(m)},$$

则

$$P_{ij}(s) = \delta_{ij} + F_{ij}(s) P_{jj}(s).$$

为保证可和性, 通常假定 $s \in [0, 1)$. 事实上, 可以延拓 s 的取值范围. 如果 (3.1) 成立, 即

$$F_{jj}(\kappa) = \sum_{m=1}^{\infty} \kappa^m f_{jj}^{(m)} < 1,$$

那么,

$$P_{jj}(\kappa) = (1 - F_{jj}(\kappa))^{-1} < \infty.$$

然后, 利用不可约性, 可以得到对任意 i 和 j, 有 $P_{ij}(\kappa) < \infty$.

利用修正的回返时, 可以给出相应的判别准则, 从而使用 Lyapunov 条件给出等价判定 (参见文献 [8]).

定理 3.1 假定链不可约, 下述论断等价.

(1) 存在某 j 和 $\kappa > 1$ 使得

$$\sum_{n=1}^{\infty} \kappa^n p_{jj}^{(n)} < \infty;$$

(2) 存在某 j 和 $\kappa > 1$ 使得

$$\mathrm{E}_j [\kappa^{\tau_j^+} 1_{\{\tau_j^+ < \infty\}}] < 1;$$

(3) 存在某 j 和 $\kappa > 1$ 使得

$$f_{jj} < 1, \quad \mathrm{E}_j [\kappa^{\tau_j^+} 1_{\{\tau_j^+ < \infty\}}] < \infty;$$

(4) 存在某 j、常数 λ、$b \in (0, 1)$ 和非负函数 $W_j \geqslant 1$ (在某 i_0, 有 $W_{i_0} < \infty$), 满足

$$PW(i) \leqslant \lambda W_i, \quad i \neq j, \quad PW(j) \leqslant b.$$

对于代数式非常返, 有如下的判别准则.

定理 3.2 令 $\ell \in \mathbb{N}$. 假定链不可约, 下述论断等价.

(1) 存在某 j 使得

$$\sum_{n=1}^{\infty} n^{\ell} p_{jj}^{(n)} < \infty;$$

(2) 存在某 j 使得

$$f_{jj} < 1, \quad \mathrm{E}_j[\tilde{\tau}_j^{+}]^{\ell} < \infty;$$

(3) 存在某 j、常数 $d \in (0,\infty)$、$b \in (0,1)$ 和非负函数 $W^{(m)}$, $m = 0,1,\dots,\ell$ (在某 $i_0 \in E$ 有 $W_{x_0}^{(m)} < \infty$), 满足对 $m = 0,1,\dots,\ell$, 有

$$\begin{cases} PW^{(m)}(i) \leqslant W_i^{(m)} - (\ell - m)W_i^{(m+1)}, & i \neq j, \\ W_j^{(m)} \geqslant 1, \quad PW^{(0)}(j) \leqslant d, \quad PW^{(\ell)}(j) \leqslant b, \end{cases}$$

其中 $W^{(\ell+1)} = 0$.

文献 [8] 用上面的思想对一般状态空间的离散时间 Markov 链作了系统的研究. 显然, 上述方法适用于跳过程, 此处不赘述.

4 离散时间 Markov 链的收敛速度

相对于连续时间的 Markov 链, 离散时间的 Markov 链可以使用的泛函工具较少. 一个常用的方法是, 将一步转移概率矩阵 P 转变为 $Q = P - I$, 从而以 Q 为生成元研究相应跳过程, 以期返回得到 P 的有关性质. 以研究可逆过程 P 的谱隙为例, 加以说明. 假定 P 是可逆的, 从而是遍历的, 具有平稳分布 π. 回顾 Q 过程的谱隙:

$$\mathrm{gap}(Q) = \lambda_1(Q) = \inf\{(-Qf, f)_\pi : \pi(f) = 0, \pi(f^2) = 1\}.$$

因此, P 在特征值 1 处的谱隙 $\mathrm{gap}(P) = \mathrm{gap}(Q)$. 如果假定 P 是非负定的 (即 P 的谱是非负的), 那么, $P - \pi$ 的谱半径

$$r(P - \pi) = (1 - \mathrm{gap}(P)). \tag{4.1}$$

这样借助连续时间 Markov 过程 Q 得到离散时间过程的收敛速度. 例如, 对于具有非负定的离散时间 Markov 链 P, 文献 [9] 利用首中时的指数矩

$$\exists\, \kappa > 1, \quad \forall\, i, \quad \mathrm{E}_i \kappa^{\tau_0} < \infty \tag{4.2}$$

得到 $r(P - \pi) \leqslant \kappa^{-1}$.

但是, 没有 P 的非负定性, 上述做法不可行, 即 (4.1) 不成立. 为此, 令 $\tilde{P} = P^2$, 它总是非负定的. 首先要通过 P 的矩条件得到 $\tilde{P}$ 矩条件, 再对 $\tilde{P}$ 使用上述结论, 最后再回到 P 的收敛速度上. 虽然 $r(P^2 - \pi) = r((P - \pi)^2)$, 但是直接处理 P^2 的矩通常是困难的. 因此, 下面关于 P 和 $\tilde{P}$ 的关系是重要的桥梁.

令 $\tilde{\tau}_j^+$ 和 $\tilde{\tau}_j$ 分别是 $\tilde{X}_n(\tilde{P})$ 的回返时和首中时,

$$\tilde{\tau}_j^+=\inf\{n\geqslant 1:\tilde{X}_n=j\},\quad \tilde{\tau}_j=\inf\{n\geqslant 0:\tilde{X}_n=j\}.$$

而 τ_j^+ 和 τ_j 是 Markov 链 $X_n(P)$ 的回返时和首中时. 再令 $f_{ij}^{(n)}=\mathrm{P}_i[\tau_j^+=n]$, $\tilde{f}_{ij}^{(n)}=\mathrm{P}_i[\tilde{\tau}_j^+=n]$, 以及对 $s\geqslant 0$, 定义

$$\tilde{F}_{ij}(s)=\sum_{n=1}^{\infty}s^{2n}\tilde{f}_{ij}^{(n)},\quad F_{ij}^{(0)}(s)=\sum_{n=1}^{\infty}s^{2n}f_{ij}^{(2n)},\quad F_{ij}^{(1)}(s)=\sum_{n=1}^{\infty}s^{2n-1}f_{ij}^{(2n-1)}.\tag{4.3}$$

命题 4.1 对 $0\leqslant s<1$, 有

$$\tilde{F}_{ij}(s)=F_{ij}^{(0)}(s)+F_{ij}^{(1)}(s)F_{jj}^{(1)}(s)[1-F_{jj}^{(0)}(s)]^{-1}.\tag{4.4}$$

由此可以清楚看出 P 和 $\tilde{P}$ 的矩母函数之间的关系, 在此基础上, 结合谱分析, 我们可以得到下面的定理 (参见文献 [7]):

定理 4.1 假设可逆 Markov 链 P 满足对某状态 0, 存在 $\lambda>1$ 使得 $\mathrm{E}_0\lambda^{\tau_0^+}<\infty$. 令

$$\rho=\sup\left\{s\leqslant\lambda:\sum_{n=1}^{\infty}s^{2n}f_{00}^{(2n)}<1\right\},$$

则 $r_0(P)\leqslant\rho^{-1}<1$.

进一步研究离散时间 Markov 链 P 的强遍历收敛速度

$$\alpha(P)=\inf\left\{\epsilon\leqslant 1:\exists C<\infty\ \text{使得}\ \sup_{i\in E}\sum_{j\in E}|p_{ij}^{(n)}-\pi_j|\leqslant C\epsilon^n\right\}.$$

首先, 当 P 为非负定的转移概率矩阵, 有类似于 (4.2) 指数遍历性的简洁结论 (参见文献 [7]).

定理 4.2 假设可逆 Markov 链 P 满足对某状态 0, 有

$$\sup_{i\in E}\mathrm{E}_i\tau_0\leqslant M<\infty,\tag{4.5}$$

那么, $\alpha(P)\leqslant \mathrm{e}^{-1/M}$.

而对于一般的可逆转移概率矩阵 P, 在我们处理一般可逆 Markov 链的指数遍历性速度的基础上, 可以给出其强遍历收敛速度 $\alpha(P)$ 的估计 (参见文献 [7]).

定理 4.3 假设可逆 Markov 链 P 满足对某状态 0, (4.5) 成立, 且

$$\sum_{n\geqslant 1}f_{00}^{(2n-1)}\geqslant\beta^{-1}>0,$$

那么,

$$\alpha(P)\leqslant \mathrm{e}^{-1/(\beta+M+\beta M)}.$$

5 首中时分布

从前面三节已经看到首中时或回返时的矩对于判定常返性、遍历性、强遍历性和非常返性的核心地位. 同时也看到利用不同的矩可以给出可逆 Markov 链的各种收敛速度的估计. 本节将研究首中时

和回返时的分布, 特别是其分布于转移概率矩阵 P (离散时间情形) 或 Q (连续时间情形) 的特征值直接的内在关系. 因为这个原因, 首中时和回返时的分布有很多的应用, 如后两节要涉及的切断现象和拟平稳分布等.

关于首中时的一个 (或许是第一个) 经典而美妙的结论是, 关于生灭过程的首中时的分布用特征值表达, 参见文献 [10, 定理 1.1] 后的有关历史评述.

令

$$G^{(n)} = \begin{pmatrix} -b_0 & b_0 & 0 & 0 & \cdots & 0 & 0 \\ a_1 & -(a_1+b_1) & b_1 & 0 & \cdots & 0 & 0 \\ 0 & a_2 & -(a_2+b_2) & b_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & a_{n-1} & -(a_{n-1}+b_{n-1}) \end{pmatrix} \tag{5.1}$$

为状态空间 $\{0,1,\dots,n-1\}$ 上的生灭过程, 以 n 为吸收点. 令 $\lambda_1^{(n)} < \cdots < \lambda_n^{(n)}$ 是矩阵 $-G^{(n)}$ 的所有 n 个正的特征值, $T_{0,n}$ 为 $G^{(n)}$ 过程从 0 出发首次击中状态 n 的时刻, 则 $T_{0,n}$ 与 n 个独立的具有指数分布的随机变量之和同分布, 每个指数分布的参数为 $\{\lambda_1^{(n)},\dots,\lambda_n^{(n)}\}$, 即

$$\mathrm{E}\mathrm{e}^{-sT_{0,n}} = \prod_{\nu=1}^{n} \frac{\lambda_\nu^{(n)}}{s+\lambda_\nu^{(n)}}, \quad s \geqslant 0. \tag{5.2}$$

上述结论在不同情形下被加以推广. 首先, Miclo [11] 给出了一般有限可逆 Markov 链的首中时分布与转移概率矩阵 (或 Q 矩阵) 的特征值之间的关系. 简言之, 以连续时间 Markov 链为例, 设 Q 是状态空间 $\{1,2,\dots,n\}$ 上的不可约但非保守的 Q 矩阵. 此时, 可加上状态 0 为吸收点, 并记 τ 为过程首达 0 的时刻. 再令 $0 < \lambda_1 < \cdots \leqslant \lambda_n$ 为 $-Q$ 的特征值, 则 τ 分布为

$$\sum_{i=1}^{n} a_i E(\lambda_i,\dots,\lambda_n),$$

其中 $a_1 > 0$, $a_i \geqslant 0$, $\sum_i a_i = 1$, 以及 $E(\lambda_i,\dots,\lambda_n)$ 表示以 $\lambda_i,\dots,\lambda_n$ 为参数的混合指数分布.

与此同时, 在文献 [12] 中, 此结论被推广到具有不同无穷边界的可数生灭过程上. 主要是借助 Courant-Fischer 定理和特征时等式等方法从有限逼近无穷状态空间, 可参见最近的综述文献 [13].

有关一般可逆 Markov 链的首中时 (回返时) 分布如何推广到非对称和可数状态空间上, 是十分具有挑战性的工作. 例如, 上述关于生灭过程的定理在单生过程情形下仍然成立, 那么对于可数状态空间上的单生过程的生命时 (假定有限) 的分布如何描述?

6 切断现象

切断现象是指一族 Markov 过程在收敛于其平稳态时所发生的切断行为. 给定一个遍历的 Markov 链 $P(t) = (p_{ij}(t))$, 具有平稳分布 π. 对给定距离 D, 如果

$$\lim_{t\to\infty} D(P(t),\pi) = 0,$$

则可以称为 D 遍历性. 当研究的对象是一族 Markov 链, 如 $P^{(n)}(t)$ ($n \geqslant 1$) 会发生如下的现象. 假定 Markov 链 $P^{(n)}(t)$, 各有平稳分布 $\pi^{(n)}$. 若存在序列 t_n 使得

$$\lim_{n\to\infty} D(P^{(n)}(ct_n), \pi^{(n)}) = \begin{cases} 0, & \text{若 } c > 1, \\ 1, & \text{若 } c < 1, \end{cases}$$

此现象被 Diaconis 称为切断 (参见文献 [14]). 在研究切断现象时, 通常选择下面三类距离:

(1) 全变差距离: $\sup_i \sum_j |p_{ij}(t) - \pi_j|$;

(2) 分离度 (separation): $\mathrm{sep}(P_i(t), \pi) := \sup_{ij}(1 - p_{ij}(t)/\pi_j)$;

(3) 极大 L^p 距离: $\sup_i[\sum_j \pi_j(1 - p_{ij}(t)/\pi_j)^p]^{1/p}$.

下面将以生灭过程为例讨论分离度下的切断.

研究分离度下的切断现象的重要而有效的工具是最快平稳时. 所谓最快平稳时是指, 独立于过程的最小停时 τ, 使得 X_T 的分布为过程的平稳分布 π.

设 X_t ($t \geqslant 0$) 为状态空间 $E = \{0, 1, \ldots, m\}$ 上的连续时间生灭过程, 具有死率 $a_i > 0$ ($1 \leqslant i \leqslant m$); 生率 $b_i > 0$ ($0 \leqslant i \leqslant m-1$) 且 $a_0 = b_m = 0$. 此过程是遍历的. 令

$$\pi_0 = \frac{1}{Z}, \quad \pi_i = \frac{1}{Z}\frac{b_0 \cdots b_{i-1}}{a_1 \cdots a_i}, \quad 1 \leqslant i \leqslant m, \tag{6.1}$$

其中 Z 使得 $\pi = (\pi_i, 0 \leqslant i \leqslant m)$ 为概率, 即 π 为平稳分布. 令 $\mu(t)$ 为过程在 t 时刻的分布, 则

$$\lim_{t\to\infty} \mathrm{sep}(\mu(t), \pi) = 0.$$

于是, 从 0 出发的连续时间生灭过程 X_t, 存在一个最快平稳时 τ 使得

$$\mathrm{sep}(\mu(t), \pi) = \mathrm{P}[\tau > t],$$

而且 τ 的分布的 Laplace 变换是

$$\mathrm{E}e^{-\lambda\tau} = \prod_{\nu=1}^{m} \frac{\lambda_\nu}{\lambda + \lambda_\nu}, \quad \lambda \geqslant 0,$$

其中 $\lambda_1 < \cdots < \lambda_m$ 为矩阵 $-Q$ 的所有非零特征值. 由此, Diaconis 和 Saloff-Coste[15] 给出了一族离散时间或连续时间的生灭过程发生分离度意义下的切断现象的判别准则, 不过此准则用到了相应过程的所有特征值. 实际上, 我们可以简化证明并给出显式的判别准则 (参见文献 [16]).

定理 6.1 对一族在 $\{0, 1, \ldots, m_n\}$ 上的从 0 出发的连续时间生灭过程 $X_t^{(n)}$, 令 $\tau^{(n)}$ 为其最快强平稳时, 那么此时有分离度下的切断现象当且仅当

$$\frac{(\mathrm{E}\tau^{(n)})^2}{\mathrm{Var}(\tau^{(n)})} \to \infty, \quad \text{或等价地} \quad \frac{(\mathrm{E}\tau^{(n)})^2}{\mathrm{E}(\tau^{(n)})^2} \to 1.$$

在此基础上, 我们将给出显式的判断准则.

定理 6.2 生灭过程有分离度下的切断现象当且仅当 $T^{(n)}\kappa^{(n)} \to \infty$, 其中

$$T^{(n)} = \sum_{\nu=1}^{n} \frac{1}{\lambda_\nu^{(n)}} = \sum_{i=0}^{n-1} \frac{1}{\pi_i^{(n)} b_i^{(n)}} \sum_{j=i+1}^{n} \pi_j^{(n)} \sum_{j=0}^{i} \pi_j^{(n)},$$

$$\kappa^{(n)}=\min_{0\leqslant\ell<m\leqslant n}\left[\left(\sum_{i=0}^{\ell}\mu_i^{(n)}\right)^{-1}+\left(\sum_{i=m}^{n}\mu_i^{(n)}\right)^{-1}\right]\left(\sum_{i=\ell}^{m-1}\frac{1}{\mu_i^{(n)}b_i^{(n)}}\right)^{-1},$$

这里各 $\pi_i^{(n)}$ 和 $\mu_i^{(n)}$ 的定义类同于 (6.1).

对于一般的随机单调过程, 我们利用 Markov 链的构造理论, 给出分离度切断现象的判断准则. 若存在状态 x 满足

$$\mathrm{P}[\tau>t]=1-\frac{\mu_x(t)}{\pi_x}=\mathrm{sep}(\mu(t),\pi)\quad\left(=\sup_y\left(1-\frac{\mu_y(t)}{\pi_y}\right)\right),\quad\forall\, t\geqslant 0,$$

则称 x 为刹车 (halting) 状态. 这样, 为得到强平稳时的分布, 只需要考虑刹车状态的分布. 为此, 需要考虑随机单调过程. 随机矩阵 $P=(p_{ij},0\leqslant i,j\leqslant m)$ 称为随机单调的, 如果对所有 i 和 j $(i\leqslant j)$, 有 $\sum_{\ell\geqslant k}p_{ik}\leqslant\sum_{\ell\geqslant k}p_{jk}$, $\forall\, k$. Markov 链是随机单调的, 如果其转移随机矩阵是随机单调的. 假设 X_t 的时间反转过程 X_t^* 是随机单调的, 并且 X_t 从状态 0 出发 (而过程 X_t^* 从状态 m 出发), 那么状态 m 是一个刹车状态. 令 $\psi_m(\lambda)$ 为转移矩阵 $p_{0m}(t)$ 的预解式, 则

$$\mathrm{E}\mathrm{e}^{-\lambda\tau}=\frac{\lambda\psi_m(\lambda)}{\pi_m},\quad\lambda\geqslant 0.$$

令 $f_{ij}(t)$ 为首返时分布, $g_{ij}(t)=1-f_{ij}(t)$. 再令

$$F_{ij}(\lambda)=\int_0^\infty \mathrm{e}^{-\lambda t}df_{ij}(t),\quad G_{ij}(\lambda)=\int_0^\infty \mathrm{e}^{-\lambda t}g_{ij}(t)dt,$$

则

$$\mathrm{E}\mathrm{e}^{-\lambda\tau}=\frac{F_{0,m}(\lambda)}{\pi_m(\lambda+q_m)G_{m,m}(\lambda)}.$$

随之得到一阶和二阶矩分别为

$$\begin{aligned}\mathrm{E}\tau&=\mathrm{E}_0\tau_m-\frac{\pi_m q_m}{2}\mathrm{E}_m\tau_m^2-\frac{1}{q_m},\\ \mathrm{E}\tau^2&=\pi_m\mathrm{E}_0\tau_m^2+2\pi_m^2\mathrm{E}_0\tau_m E_m\tau_m^2+\frac{2\pi_m}{q_m^2}+\frac12\pi_m^3q_m^2(\mathrm{E}_m\tau_m^2)^2\\ &\quad-\left[\frac13\pi_m^3q_m\mathrm{E}_m\tau_m^3+2\pi_m^2\mathrm{E}_0\tau_m\mathrm{E}_m\tau_m+\pi_m^2\mathrm{E}_m\tau_m^2\right].\end{aligned}$$

因此可以得到一族具有随机单调时间反转过程的分离度的判别准则. 特别是, 生灭过程是可逆的, 本身就是随机单调的, 所以, 通常此方法同样可以得到分离度切断现象的判别准则, 而不需要绕道使用特征值.

下面通过一族生灭过程分离度切断现象的例子感受一下 (参见文献 [16]). 考虑 $\{0,1,\dots\}$ 上的连续时间生灭过程, 具有死 (生) 速率 a_i $(i\geqslant 1)$ 和 b_i $(i\geqslant 0)$. 对每一 $n=1,2,\dots$, 考虑 $\{0,1,\dots,n\}$ 上的遍历的子过程 $X_t^{(n)}$, 即在状态 n 为反射的: $b_i^{(n)}=b_i$, $0\leqslant i\leqslant n-1$ 和 $b_n^{(n)}=0$, 且 $a_i^{(n)}=a_i$, $1\leqslant i\leqslant n$.

推论 6.1 对于如上给出的一列遍历的子生灭过程 $X_t^{(n)}$, 其有分离度的切断现象当且仅当原过程 X_t 非强遍历.

7 拟平稳分布

Markov 过程的拟平稳分布是一种条件化的极限行为, 其存在性、唯一性和其他性质的研究始于 20 世纪 40 年代, 可参见专著 [17]. 另外, Pollett[18] 在其网站上收集了与拟平稳分布相关的文献.

假设 Markov 链 X_t 具有生命时 ζ. 其 (拟) 平稳分布 u 定义为

$$\mathrm{P}_u[X_t = i \mid t < \zeta] = u_i.$$

容易看到, 上述定义等价于存在 $\lambda \geqslant 0$ 使得

$$\sum_i u_i p_{ij}(t) = \mathrm{e}^{-\lambda t} u_j.$$

如果 $\zeta = \infty$ a.e., 则 u 即是平稳分布. 如果 $\zeta < \infty$ a.e., 则 u 称为拟平稳分布, 或者条件平稳分布. 可以看出, 拟平稳分布是一种有条件的平稳分布, 但是却有着不同于平稳分布的独特性. 例如, 一般而言, 有关不可约的遍历的 Markov 链有唯一的平稳分布; 而对不可约 (非常返) Markov 链, 其拟平稳分布或者不存在, 或者唯一, 或者无穷多个. 因此, 其研究更具魅力.

拟平稳分布源于 Yaglom (1947) 对经典分支过程的研究. 假定此时分支机制 ξ 满足

$$m := \mathrm{E}\xi \leqslant 1 \quad \Rightarrow \quad \lim_{n\to\infty} p_{i0}(n) = 1.$$

考虑其极限问题: $\forall i, j \geqslant 1$, 有

$$\lim_{n\to\infty} \mathrm{P}_i[X_n = j \mid \tau_0 > n] = \lim_{n\to\infty} \frac{\mathrm{P}_i[X_n = j, n < \tau_0]}{\mathrm{P}_i[n < \tau_0]} = \lim_{n\to\infty} \frac{p_{ij}(n)}{\sum_{k\geqslant 1} p_{ik}(n)}.$$

此时, 生命时 $\zeta = \tau_0$.

定理 7.1 令 $X_0 = 1$, $X_n = \sum_{k=1}^{X_{n-1}} \xi_{n,k}$. 记

$$f(s) = \mathrm{E}s^{\xi}, \quad g(s) = \lim_{n\to\infty} \mathrm{E}[s^{X_n} \mid X_n > 0], \quad s \in [0,1].$$

假设 $m < 1$, 则 $g(s)$ 存在并满足

$$g(f(s)) = mg(s) + 1 - m.$$

特别地,

$$\lim_{n\to\infty} \mathrm{P}[X_n = i \mid X_n > 0] = b_i > 0, \quad \sum_{i\geqslant 1} b_i = 1.$$

对于有限 Markov 链, 拟平稳分布是 Perron-Frobenius 定理的一个推论. 而对于一般 Markov 过程的拟平稳分布, 有三个基本问题.

- 存在性与唯一性: 首先判定拟平稳分布存在, 而特别的地方是拟平稳分布可以非唯一, 这与平稳分布不一样.
- 吸引域: 从什么样的初分布出发过程的条件极限为给定的拟平稳分布? 特别是 QSD 非唯一的时候, 如何完整刻画不同的吸引域?
- 收敛速度: 从定性到定量的跨越, 因为吸引域的问题, 收敛速度变得更加复杂.

除了分支过程, 生灭过程的有关拟平稳分布的研究结果是较为丰富的. 简要回顾生灭过程的拟平稳分布的结论, 特别是具有不同的无穷边界情形.

若 $R = \infty$, $S = \infty$, 则相应的 Q 过程是唯一的. 当 $a_0 > 0$ 时, 生命时 ζ 即为 -1 的击中时. 此时, 在指数衰减速度 $\lambda > 0$ 的前提下, Van Doorn[19] 刻画了所有的拟平稳分布, 即 $\forall x \in (0, \lambda]$ 都对应一个拟平稳分布. 需要指出在 $a_0 > 0$ 时的衰减速度 $\lambda > 0$ 的判别准则等同于在 $a_0 = 0$ 时过程的指数遍历性.

若 $R=\infty$, $S<\infty$, 则相应的 Q 过程依然唯一. 当 $a_0>0$ 时, Van Doorn [19] 证明了过程存在唯一的拟平稳分布.

那么接下来的问题为, 如果 $R<\infty$, 即相应的 Q 过程非唯一, 会发生什么? 此时, 拟平稳分布的问题是针对最小过程提出的, 即生命时需要考虑到过程的飞跃时. 简言之, 若 $R<\infty$, 无论 $S=\infty$ 或 $S<\infty$, 相应的最小 Q 过程的拟平稳分布唯一.

首先对流出边界情形, 使用对偶方法, 即通过对偶, 流出边界最小生灭过程的拟平稳分布转化为具有流入边界的生灭过程的相应特征函数的有界性问题. 此时其最小过程存在唯一的拟平稳分布 (参见文献 [20]).

对于正则边界的最小过程的拟平稳分布, 我们使用从有限状态空间到可数状态空间的逼近. 考虑下面的以 n 为吸收点的过程:

$$Q^{(n)}=\begin{pmatrix} -b_0 & b_0 & & & \\ a_1 & -(a_1+b_1) & b_1 & & \\ & \ddots & \ddots & \ddots & \\ & & a_{n-1} & -(a_{n-1}+b_{n-1}) & b_{n-1} \end{pmatrix},$$

其极限过程即是最小过程.

令 δ_ν $(\nu\geqslant 1)$ 为最小过程的特征值, 而 $\delta_\nu^{(n)}$ $(1\leqslant\nu\leqslant n)$ 为 $-Q^{(n)}$ 的特征值, 则

$$\delta_\nu^{(n)}\downarrow\delta_\nu,\quad \forall\,\nu\geqslant 1.$$

而令 τ_n 为 n 的首中时, 则 $\tau_n\uparrow\zeta$. 对 $Q^{(n)}$ 过程 $(p_{ij}^{(n)}(t))$, 有拟平稳分布 $u_i^{(n)}$:

$$\sum_{i=0}^{n-1}u_i^{(n)}p_{ij}^{(n)}(t)=\mathrm{e}^{-\delta_1^{(n)}t}u_i^{(n)},$$

其中

$$u_i^{(n)}=\frac{\mu_i(\mathrm{E}_0\mathrm{e}^{\delta_1^{(n)}\tau_i})^{-1}}{\sum_{i=0}^{n-1}\mu_i(\mathrm{E}_0\mathrm{e}^{\delta_1^{(n)}\tau_i})^{-1}}.$$

因为上述有限过程逼近最小过程, 所以使用文献 [12] 中关于正则边界的最小过程的特征值与有限过程的特征值的逼近关系, 从而令 $n\to\infty$ 即得到最小过程的拟平稳分布.

8 非对称性

从以上几节内容来看, 可逆性扮演了非常关键角色. 本节将通过一些重要的指标, 来比较 Markov 链与其对称化 Markov 链. 我们的方法是发展 (非对称) Markov 链的 Dirichlet 原理. 为此, 先回顾经典的 Dirichlet 问题. 考虑有界光滑区域 D 上的 Dirichlet 问题:

$$\text{在 } D \text{ 内 } \Delta f=0,\qquad \text{且在 } \partial D \text{ 上 } f=\phi\in W^{1,2}(D)$$

有唯一解 f_0. 经典的 Dirichlet (或能量) 原理告诉我们,

$$\inf\left\{\int_D|\nabla f|^2dx: f\in W^{1,2}(D)\ \text{且}\ f|_{\partial D}=\phi\right\}$$

在 f_0 达到.

上述 Dirichlet 问题容易推广到可逆扩散过程上, 并由 Pinsky [21] 推广到如下的非对称扩散算子情形. 更一般地, $L = \nabla \cdot a\nabla + b \cdot \nabla$ 为 (非对称) 扩散算子. 此时仍然有相应的 Dirichlet 问题:

$$\text{在 } D \text{ 内 } Lf = 0 \quad \text{且在 } \partial D \text{ 上 } f = \phi \in W^{1,2}(D)$$

有唯一解 f_0. 而此时 Dirichlet 原理转变为

$$\inf_g \sup_h \int \left(\frac{|\nabla g|}{g} - a^{-1}b\right) a \left(\frac{|\nabla g|}{g} - a^{-1}b\right) g^2 dx - \int (\nabla h - a^{-1}b) a (\nabla h - a^{-1}b) g^2 dx$$

在 $g_0 = \sqrt{f_0 f_0^*}$ 和 $h_0 = \frac{1}{2}\log f_0/f_0^*$ 处达到, 其中 f_0^* 是对偶 Dirichlet 问题 $L^* = \nabla \cdot a\nabla - b \cdot \nabla$ 的唯一解.

回到有限 Markov 链. 令 E 是一有限状态空间, $K = (K_{ij})_{i,j\in V}$ 是上面的 (概率) 转移矩阵, 关于概率测度 π 可逆, 即

$$K_{ij} \geqslant 0, \quad \sum_j K_{ij} = 1, \quad \pi_i K_{ij} = \pi_j K_{ji}.$$

再令 P 为另一转移矩阵, 以 π 为平稳分布:

$$\sum_i \pi_i P_{ij} = \pi_j.$$

一般地, 假定 K 是 P 的可逆部分,

$$K_{ij} = \frac{1}{2}[P_{ij} + P_{ij}^*], \quad P_{ij}^* = \frac{\pi_j P_{ji}}{\pi_i}.$$

对可逆 Markov 链 K, 可视之为电网络 (参见文献 [22]). 若 $K_{ij} > 0$, 则连接 i 和 j 为一电导线, 其电阻率为 $a_{ij} = \pi_i p_{ij}$, 而 $r_{ij} = 1/a_{ij}$ 为其电阻. 给定状态 a 和 b, 令 $\tilde{f}$ 为下述方程的解:

$$\begin{cases} Pf(i) = f_i, & i \neq a, b, \\ f(a) = 1, & f(b) = 0, \end{cases}$$

则 $c_{ij} = a_{ij}(\tilde{f}_i - \tilde{f}_j)$ 即为电流. 此即所谓 Ohm 定律. 此时, 可逆 Markov 链的 Dirichlet 原理为

$$\mathrm{Cap}_{a,b} = \inf\left\{\frac{1}{2}\sum_{ij} a_{ij}(f_j - f_i)^2 : f_a = 1, f_b = 0\right\} \tag{8.1}$$

在 $\tilde{f}$ 处达到. 此即所谓电学的 Thompson 原理. 因为对给定状态 a, 有

$$\pi_a \mathrm{P}_a[\tau_a^+ > \tau_b] = \mathrm{Cap}_{a,b},$$

并令 $b = b_n = \{E_n\}$ 满足 $E_n \uparrow E$, 所以有如下的可逆 Markov 链常返性的判断准则: K 常返当且仅当

$$\lim_{n\to\infty} \mathrm{Cap}_{a,b_n} = 0.$$

有关常返性课题, 可参见 Doyle 和 Snell 的书 [22] 或 Chen 的书 [3, 第 7 章].

Gaudillière 和 Landim [23] 将 Thompson 原理推广到不可逆情形. 令 $\tilde{f}^*$ 为下面对偶过程 P^* 的方程的唯一解:

$$\begin{cases} P^* f(i) = f_i, \quad i \neq a, b, \\ f(a) = 1, \quad f(b) = 0. \end{cases}$$

对任意点对 $i \neq j$,

$$\mathrm{Cap}_{a,b} = \inf\{(f, (I-P)(I-K)^{-1}(I-P)^* f)_\pi : f_a = 1, f_b = 0\}$$

在 $(\tilde{f} + \tilde{f}^*)/2$ 处达到. 注意若 P 可逆, 即 $P = K$, 则其中 "inf" 变回 (8.1). 注意到

$$T_{ij} := \mathrm{E}_i \tau_j + \mathrm{E}_j \tau_i = \frac{1}{\mathrm{Cap}_{ij}}.$$

由此, 有了第一个关于 Markov 链与其可逆 Markov 链的比较. 给定点对 $i \neq j$, 令 $T_{ij}(K)$ 和 $T_{ij}(P)$ 分别是 K 和 P 的互通时, 则

$$T_{ij}(P) \leqslant T_{ij}(K).$$

从而, 相应的平均首中时 $T_0 = \frac{1}{2}\sum_{ij} \pi_i \pi_j T_{ij} = \sum_{ij} \pi_i \pi_j \mathrm{E}_i \tau_j$ 满足

$$T_0(P) \leqslant T_0(K).$$

上述结果有一个意外的应用, 可以给出 Aldous 和 Fill 的一个猜想 (参见文献 [24, 第 9 章, 猜想 22]) 的肯定答案 (参见文献 [25]). 令

$$Z_{ij} = \sum_{n=0}^{\infty} [P_{ij}^{(n)} - \pi_j]$$

为 P 的基础矩阵. Aldous 和 Fill 猜测:

$$\mathrm{trace}(Z^2(P^* - P)) \geqslant 0.$$

令 $P_\lambda = \lambda P + (1-\lambda) P^*, \lambda \in [0,1]$, 则此猜想等价于

$$T_0(P_\lambda) \leqslant T_0(P_{1/2}).$$

实际上, 令 $K = 1/2(P + P^*)$, $\Gamma = (\lambda - 1/2)\mathrm{diag}(\pi)(\mathrm{P} - \mathrm{P}^*)$ 和 $P = K + \Gamma$, 可以证明上述猜想.

注意到, Doyle 1994 年在其个人主页上发表的一篇 "小论文" [26] 中, 讨论了 Markov 链的更一般的 Dirichlet 问题. 令 A 为状态空间的一个子集, 且

$$\text{在 } A^c \text{ 上 } Pf = f, \quad \text{而在 } A \text{ 上 } f = \phi$$

有唯一解 $\tilde{f}$. 再令

$$D(f, g) = (f, (I-P)f)_\pi = \frac{1}{2}\sum_{ij} \pi_i P_{ij} (f_i - f_j)(g_i - g_j).$$

对于一般 (不可逆) Markov 链 P, Doyle 证明了

$$D(\tilde{f}, \tilde{f}^*) = \min_{f|_A = \phi} \max_{g|_A = 0} D(f+g, f-g),$$

其中 $\tilde{f}^*$ 是相对应对偶过程 P^* 的 Dirichlet 问题的解.

受此启发, 我们进一步考虑两类 Poisson 方程. 为此, 令 $L_c = I - \mathrm{e}^{-c}P$, 其中势函数 $c \geqslant 0$, 并令 $L = L_0$. 一类是带边界的:

$$\begin{cases} L_c f = \phi, & \text{在 } A^c, \\ f = \xi, & \text{在 } A. \end{cases}$$

另一类是无边界的:

$$Lf = \phi - \pi(\phi),$$

其中 π 是平稳分布.

对相应的对偶过程, 同样令 $L_c^* = I - \mathrm{e}^{-c^*}P^*$ 和 $L^* = L_0^*$. 也有相应的两类对偶 Poisson 方程: 带边界的对偶 Poisson 方程

$$\begin{cases} L_c^* f^* = \phi^*, & \text{在 } A^c, \\ f^* = \xi^*, & \text{在 } A, \end{cases}$$

以及无边界的对偶 Poisson 方程

$$L^* f = \phi^* - \pi(\phi^*).$$

我们将给出关于

$$D_c(f, f^*) = (L_c f, f^*)_\pi = (f, L_c^* f^*)_\pi$$

的变分公式. 首先容易得到带边界 Poisson 方程:

$$\begin{cases} L_c f = \phi, & \text{在 } A^c, \\ f = \xi, & \text{在 } A \end{cases}$$

的解是 $f(x) = \mathrm{E}_x[\sum_{n=0}^{\tau_A - 1} \phi(X_k) \mathrm{e}^{-\sum_{k=0}^{n-1} c(X_k)} + \xi(X_{\tau_A})]$. 其次, 选取适当的函数类 $\mathscr{F}$ 和 $\mathscr{G}$ 可以证明

$$D(f, f^*) = \inf_{h \in \mathscr{F}} \sup_{g \in \mathscr{G}} D(h - g, h + g).$$

我们来看几个应用. 第一个是 Poisson 方程

$$\begin{cases} Lf(i) = f(i) - Pf(i) = 0, & i \neq a, b, \\ f_a = 1, \quad f(b) = 0 \end{cases}$$

有唯一解 $f_i = \mathrm{P}_i[\tau_a < \tau_b]$. 并且对偶过程有 $f_i^* = \mathrm{P}_i[\tau_a^* < \tau_b^*]$. 因为

$$D(f, f^*) = (Lf, f^*)_\pi = \pi_a Lf(a) = \pi_a \mathrm{P}_a[\tau_a^+ > \tau_b],$$

所以回到了 $\mathrm{Cap}_{a,b}$.

其次, 考虑下面 Poisson 方程的唯一解 $(\mathrm{E}_i[\tau_A] : i \in E)$:

$$\begin{cases} Pf(i) = f_i - 1, & i \in A^c, \\ f_i = 0, & i \in A. \end{cases}$$

对此, 我们有下面的定理 (参见文献 [27]):

定理 8.1 令 τ_A 为 A 的首中时, 则

$$\frac{1}{\mathrm{E}_\pi[\tau_A]}=\inf_{f\,|_A=0,\pi(f)=1}\ \sup_{g\,|_A=0,\pi(g)=0} D(f-g,f+g).$$

因为 $T_0(P)=\sum_j \pi_j \mathrm{E}_\pi \tau_j$, 链的平均首中时满足 $T_0(P)\leqslant T_0(K)$.

上述结果可以看作文献 [24, 第 3 章, 命题 41] 对有限可逆 Markov 链相应结论的推广:

$$\frac{1}{\mathrm{E}_\pi[\tau_A]}=\inf_{f\,|_A=0,\pi(f)=1} D(f,f).$$

我们可以考虑势函数为常数的情形, 即 $c_i\equiv\lambda>0$,

$$\begin{cases} Pf(i)=(\lambda+1)f_i, & i\in A^c,\\ f_i=1, & i\in A \end{cases}$$

有唯一解 $(\mathrm{E}_i[\exp(-\lambda\tau_A)]:i\in V)$. 同样可以得到下面的定理:

定理 8.2 令 τ_A 为 A 的首中时, 并对 $\lambda>0$, 令 $D_\lambda(f,g)=D(f,g)+\lambda\pi(fg)$, 那么,

$$\frac{\lambda}{1-\mathrm{E}_\pi[\exp(-\lambda\tau_A)]}=\inf_{f|_A=0,\pi(f)=1}\ \sup_{g|_A=0,\pi(g)=0} D_\lambda(f-g,f+g).$$

此结论即便是对可逆 Markov 链也是全新的:

$$\frac{\lambda}{1-\mathrm{E}_\pi[\exp(-\lambda\tau_A)]}=\inf_{f\,|_A=0,\pi(f)=1} D_\lambda(f,f).$$

事实上, 上面的结果可以导出下面的 Aldous 和 Fill 的猜想的一个加强结论.

推论 8.1 假设 X 是不可约有限 Markov 链, 其转移矩阵是 P, 则对任意 i, 有 $[Z(P^*-P)Z]_{i,i}\geqslant 0$. 特别地, $\mathrm{trace}[Z^2(P^*-P)]\geqslant 0$.

最后来看无边界的 Poisson 方程. 我们指出, 前面的 Dirichlet 变分思想可以应用于 Markov 链的渐近方差. 众所周知, 渐近方差是 Markov 链的中心极限定理的非常重要的指标. 令 X_k 为 Markov 链, 具有转移矩阵 P 和平稳分布 π, 那么, 对 $\pi(g)=0$, 有

$$\sum_{k=0}^{n-1}\frac{g(X_k)}{\sqrt{n}}\Rightarrow N(0,\nu(g,P)),$$

其中

$$\nu(g,P)=\lim_{n\to\infty}\mathrm{Var}_\pi\left[\sum_{k=0}^{n-1}\frac{g(X_k)}{\sqrt{n}}\right].$$

为此, 考虑 Poisson 方程: 给定 $g\in L^2(\pi)$ 满足 $\pi(g)=0$,

$$Pf=f-g$$

有唯一解 $f\in L^2(\pi)$. 从而, $\nu(g,P)=2\pi(fg)-\pi(g^2)$.

由此得到渐近方差的 Dirichlet 变分公式.

定理 8.3 令 $\sigma^2(g,P)=\nu(f,P)+\mathrm{Var}_\pi(g)$, 则

$$[\sigma^2(g,P)]^{-1}=\inf_{\xi\in L^2(\pi),\pi(g\xi)=1}\ \sup_{\eta\in L^2(\pi),\pi(g\eta)=0} D(\xi+\eta,\xi-\eta).$$

若 P 可逆, 则

$$[\sigma^2(g,P)]^{-1}=\inf_{\xi\in L^2(\pi),\pi(g\xi)=1} D(\xi,\xi).$$

此 Dirichlet 变分公式的一个直接应用是将经典的 Peskun 比较定理 [28] 推广到不可逆情形.

定理 8.4 假设 P_1 可逆, P_1 和 P_2 有相同的平稳分布. 若 P_2 在非对角线控制 P_1, 即

$$P_2(i,A\setminus\{i\})\geqslant P_1(i,A\setminus\{i\}),\quad i\in E,\quad A\subset E,$$

则 P_2 的渐近方差小于或等于 P_1 的渐近方差.

经典的 Peskun 定理要求 P_1 和 P_2 都是可逆的 Markov 链.

9 后记

来北京师范大学读博士, 缘于严士健先生 1992 年去吉林大学开的一个短课, 讲的是当时热门的无穷维粒子系统. 于是 1993 年从吉林大学硕士毕业后投考北京师范大学博士, 来北京师范大学后归于陈木法老师麾下, 并一直伴其左右. 其时, 无穷维粒子系统虽然热得发烫, 物极必反, 渐趋强弩之末. 用陈木法老师的话说, 其根本问题还得回到有限维, 甚至于一维. 读博及随后北京师范大学工作期间, 也曾尝试无穷维空间 (如环 (loop) 空间) 和随机分析, 终因天资不逮, 更兼努力不够, 始终如雾里看花, 遂回归 Markov 链. 所幸赶上了所谓 "Markov 链的文艺复兴" (Persi Diaconis 语). 更重要的是, 这是陈老师的根据地, 不愁没饭吃. 聊以自慰, 在这一广袤的根据地上, 占住一个小山包, 开垦出一小片自留地: 在平稳性之外, 研究非常返性和拟平稳性. 特别是最近关于非对称 Markov 链所展开的研究. 戏言是与陈老师 "唱对台戏", 因为陈老师最近的一个得意工作是, 做出复矩阵的可配称性判定, 及其特征值的算法. 其实, 遍历性和非常返性, 对称性和非对称性, 都如同一枚硬币的正反面一样, 是相反相成的.

谨以此文敬祝严士健先生九秩华诞!

参考文献

1 Mao Y H. Algebraic convergence for discrete-time ergodic Markov chains. Sci China Ser A, 2003, 46: 621–630
2 Mao Y H. Ergodic degrees for continuous-time Markov chains. Sci China Ser A, 2004, 47: 161–174
3 Chen M F. From Markov Chains to Non-Equilibrium Particle Systems, 2nd ed. Singapore: World Scientific, 2004
4 Chen M F, Wang F Y. Cheeger's inequalities for general symmetric forms and existence criteria for spectral gap. Ann Probab, 2000, 28: 235–257
5 Mao Y H, Xia L H. Spectral gap for jump processes by decomposition method. Front Math China, 2009, 4: 335–347
6 Kontoyiannis I, Meyn S P. Geometric ergodicity and the spectral gap of non-reversible Markov chains. Probab Theory Related Fields, 2012, 154: 327–339
7 Mao Y H. Convergence rates for reversible Markov chains without the assumption of nonnegative definite matrices. Sci China Math, 2010, 53: 1979–1988
8 Mao Y H, Song Y H. On geometric and algebraic transience for discrete-time Markov chains. Stochastic Process Appl, 2014, 124: 1648–1678
9 Sokal A D, Thomas L E. Exponential convergence to equilibrium for a class of random-walk models. J Stat Phys, 1989, 54: 797–828

10 Fill J A. The passage time distribution for a birth-and-death chain: Strong stationary duality gives a first stochastic proof. J Theoret Probab, 2009, 22: 543–557
11 Miclo L. On absorption times and Dirichlet eigenvalues. ESAIM Probab Stat, 2010, 14: 117–150
12 Gong Y, Mao Y H, Zhang C. Hitting time distributions for denumerable birth and death processes. J Theoret Probab, 2012, 25: 950–980
13 毛永华. 生灭过程的平稳性与拟平稳性. 中国科学: 数学, 2019, 49: 467–484
14 Diaconis P. The cutoff phenomenon in finite Markov chains. Proc Natl Acad Sci USA, 1996, 93: 1659–1664
15 Diaconis P, Saloff-Coste L. Separation cut-offs for birth and death chains. Ann Appl Probab, 2006, 16: 2098–2122
16 Mao Y H, Zhang Y H. Explicit criteria on separation cutoff for birth and death chains. Front Math China, 2014, 9: 881–898
17 Collet P, Martínez S, San Martin J. Quasi-Stationary Distributions. Heidelberg: Springer, 2013
18 Pollett P K. Quasi-stationary distributions: A bibliography. Http://www.maths.uq.edu.au/pkp/papers/qsds.html, 2012
19 Van Doorn E A. Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes. Adv Appl Probab, 1991, 23: 683–700
20 Gao W J, Mao Y H. Quasi-stationary distribution for the birth-death process with exit boundary. J Math Anal Appl, 2015, 427: 114–125
21 Pinsky R G. A generalized Dirichlet principle for second order nonselfadjoint elliptic operators. SIAM J Math Anal, 1988, 19: 204–213
22 Doyle P G, Snell P. Random Walks and Electric Networks. Providence: Amer Math Soc, 1984
23 Gaudillière A, Landim C. A Dirichlet principle for non reversible Markov chains and some recurrence theorems. Probab Theory Related Fields, 2014, 158: 55–89
24 Aldous D J, Fill J A. Reversible Markov chains and random walks on graphs. Https://www.stat.berkeley.edu/~aldous/RWG/book.html, 1995
25 Huang L J, Mao Y H. On some mixing times for nonreversible finite Markov chains. J Appl Probab, 2017, 54: 627–637
26 Doyle P G. Energy for Markov chains. Http://www.math.dartmouth.edu/doyle, 1994
27 Huang L J, Mao Y H. Variational principles of hitting times for non-reversible Markov chains. J Math Anal Appl, 2018, 468: 959–975
28 Peskun P H. Optimum Monte-Carlo sampling using Markov chains. Biometrika, 1973, 60: 607–612

Markov chains: Ergodicity, quasi-stationarity and asymmetry

Yonghua Mao

Abstract Based on the first hitting time or return time, we review the development of Markov chain in the study of stationarity, quasi-stationarity and asymmetry. These topics include: (1) using the moments of the return time to derive the functional inequalities; (2) introducing the modified return time to describe the various transience; (3) obtaining the functional inequalities via the return times; (4) using the eigenvalues to describe the distribution of the hitting times; (5) giving the criteria for the cutoff by the hitting times; (6) obtaining the quasi-stationary distribution through the distribution of the life times; (7) developing the Dirichlet principle to judge which is better between the non-reversible Markov chain and its reversible one.

Keywords **Markov chains, hitting time, return time, stationarity, quasi-stationarity, cut-off, asymmetry**

MSC(2010) **60B10, 60J27, 60G20**

doi: 10.1360/N012019-00069