SCIENTIA SINICA Mathematica

论 文



Markov 链大偏差与矩阵谱

献给严士健教授 90 华诞

陈金文

清华大学数学科学系, 北京 100084 E-mail: chenjw@mail.tsinghua.edu.cn

收稿日期: 2019-01-13; 接受日期: 2019-08-20; 网络出版日期: 2019-12-30

国家自然科学基金 (批准号: 11671226) 资助项目

摘要 本文研究可数状态 Markov 链的大偏差及相关问题,通过一种比较与控制的方法,探讨减弱对不可约性要求的可能性.这种途径还让我们从不同角度刻画了大偏差速率函数,得到它们与过程平稳分布之间的一些本质联系.特别地,本文证明有限状态链一定满足大偏差原理,其速率函数的 (局部)极小值与链的转移矩阵的主特征值和特征向量密切相关.

关键词 Markov 过程 大偏差 熵

MSC (2010) 主题分类 60J10, 60F10, 28Dxx

1 引言

Markov 过程的大偏差理论是现代 Markov 过程理论中的重要内容. Donsker 和 Varadhan [1-4] 及 Wentzell [5-8]、Freidlin 和 Wentzell [9] 分别开创了两个方向的研究. Donsker 和 Varadhan 的奠基性工作 [1-4] 从遍历性角度研究了过程的大偏差行为, 而 Freidlin 和 Wentzell 的工作则主要研究过程轨道的大偏差问题. 他们的工作一方面搭建了 Markov 过程大偏差研究的基本框架, 获得了过程长时间演化行为的一系列深刻结果; 另一方面也建立了这一理论与其他许多分支的实质联系, 包括在统计、遍历理论、随机控制、渐近积分、偏微分方程、特征值理论、统计物理和信息论等领域中的应用, 其中, 大偏差速率函数也成了具有独立研究意义的重要对象, 它不但是研究系统随机涨落的主要工具, 也是刻画系统许多关键指标的重要桥梁. 迄今为止, Markov 过程大偏差理论和方法得到了广泛的研究和应用, 一直是一个非常活跃的领域. 关于大偏差的一般理论、方法和应用, 可参见文献 [10,11].

Donsker 和 Varadhan 的经典工作主要研究了 Markov 过程经验均值、经验测度和经验过程的大偏差问题. 周期性对此已不产生影响, 但不可约性却是一个重要的影响因素. 为了得到大偏差原理, Donsker 和 Varadhan [1-4] 对过程作了较强的不可约假设, 这些假设在离散时间和可数状态情形下可表述如下:

英文引用格式: Chen J W. Large deviations for Markov chains and spectrum of matrices (in Chinese). Sci Sin Math, 2020, 50: 31–46, doi: 10.1360/N012019-00019

- (1) 一步转移概率 $\pi_{i,j} > 0, \forall i, j;$
- (2) 存在函数 $u \ge 1$, 使得 $\forall a, \{i : u_i/(\Pi u)_i \le a\}$ 有限.

这些假设强于通常的不可约性和遍历性. 在这些假定下, Donsker 和 Varadhan [1-4] 给出了上述三种水平的大偏差原理, 也给出了相应速率函数的经典形式, 尤其是在经验测度和经验过程的情形下, 相应的速率函数与熵的概念具有密切联系, 不但有着重要的概率含义, 也具有重要的物理意义和其他应用. 以下就离散时间 Markov 链的情形简要介绍有关记号和结论, 主要考虑经验测度的大偏差.

设 $\{X_n, n \geq 0\}$ 是定义在概率空间 (Ω, \mathcal{F}, P) 上取值于距离可测空间 (E, \mathcal{E}) 中的时齐 Markov 链, 初始分布为 μ_0 ,转移函数为 Π . $M_1(E)$ 为 (E, \mathcal{E}) 上的概率测度全体, 赋予弱拓扑. 定义链的经验测度为

$$L_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k},\tag{1.1}$$

其中 δ_x 是 E 上质量集中在 x 处的 Dirac 测度. 以 μ_n 表示 L_n 在 P 之下的概率分布, 则 μ_n 是 $M_1(E)$ 上的概率测度. 称 $\{L_n, n \geq 1\}$ 在 P 之下 (或等价地 $\{\mu_n, n \geq 1\}$) 满足大偏差原理 (large deviation principle, LDP), 若存在 $M_1(E)$ 上的非负函数 I, 满足

- (i) I 下半连续;
- (ii) 对 $M_1(E)$ 的任意开子集 G,

$$\liminf_{n \to \infty} \frac{1}{n} \log \mu_n(G) = \liminf_{n \to \infty} \frac{1}{n} \log P(L_n \in G)$$

$$\geqslant -\inf_{\mu \in G} I(\mu); \tag{1.2}$$

(iii) 对 $M_1(E)$ 的任意闭子集 F,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mu_n(F) = \limsup_{n \to \infty} \frac{1}{n} \log P(L_n \in F)$$

$$\leq -\inf_{\mu \in F} I(\mu). \tag{1.3}$$

此时称 I 为相应的大偏差速率函数. 若 $\forall a \geq 0$, 水平集

$$\Phi_a \triangleq \{ \mu \in M_1(E), I(\mu) \leqslant a \}$$

是紧集, 则称 I 是紧的 (或好的). Donsker 和 Varadlhan [1-4] 给出了 $\{\mu_n, n \ge 1\}$ 大偏差速率函数的下述经典形式:

$$I(\mu) = -\inf_{u \in \mathcal{U}_1} \int \log \frac{\Pi u}{u} d\mu, \tag{1.4}$$

其中

$$U_1 = \{u : u \in E \perp f \ F \ T \ M, \ L \ u \ge 1\}.$$

当 $\{X_n, n \ge 0\}$ 独立同分布时, I 就是熟知的、如下定义的相对熵:

$$h(\mu;\nu) = \begin{cases} \int \log \frac{d\mu}{d\nu} d\mu, & \text{ if } \mu \ll \nu, \\ +\infty, & \text{ if } \Pi. \end{cases}$$
 (1.5)

上述 I 函数具有很多重要性质和应用,即使没有证明大偏差原理,I 仍然扮演着重要的角色,例如,当 Π 是 Feller 转移函数时,I 刻画了过程的全部极限状态,确切地说, $I(\mu)=0$ 当且仅当 $\mu\Pi=\mu$,即 μ 是 Π 的平稳分布. 又例如,Donsker 和 Varadhan [12] 应用 I 给出了分析中 Dirichlet 特征值的推广形式的变分公式. 虽然后续研究工作减弱了对不可约性和遍历性的要求,速率函数也与上述 I 函数有所不同,但速率函数的性质仍是一个重要的研究课题. 此外,(1.4) 是一种变分形式,这在很大程度上限制了它的应用—不难想象,如果能给出 I 的简洁的显式表达,它将在诸如算法设计等实际问题中起到关键作用.

本文的主要目的之一是试图利用比较的方法,尤其是通过与独立同分布随机变量序列的比较来研究 Markov 过程经验测度的大偏差,采用这一途径的好处之一是可以避免直接讨论不可约性和遍历性.事实上,我们的结果对过程的要求较弱,本质上减弱了对不可约性和遍历性的要求,见下节定理 2.1 及其后的推论和注. 我们的研究将涉及过程经验序对的大偏差问题. Ellis [13] 就对经验序对的大偏差进行过研究. Dembo 和 Zeitouni 的文献 [11] 中也包含了经验序对的一些基本结论. 他们的结果都表明,经验序对大偏差的速率函数将以某种非变分形式出现,见下一节. 第 3 节将研究速率函数的局部极小值和局部极小值点的刻画,这在模拟退火和统计物理的局部能量等问题的研究中具有重要应用. 我们将看到这种刻画与 Perron-Frobenius 特征理论密切相关.

为了能实施这些想法, 我们将以可数状态情形为例进行研究. 此时转移函数 Π 将以转移矩阵的形式给出. 不失一般性, 我们将假设 $\{X_n,n\geq 0\}$ 取值于 $E=\mathbb{Z}_+=\{0,1,2,\ldots\}$ 中, 在其上赋离散拓扑和相应的 Borel σ - 代数 \mathcal{E} . 第 2 节将研究 $\{X_n,n\geq 0\}$ 的经验序对的大偏差问题, 由此应用压缩原理即可得到经验测度的大偏差原理. 本节的主要结果是利用满足大偏差原理的过程来控制我们所研究的过程, 从而得到后者的大偏差原理. 一个主要工作实际上是对大偏差理论中的 Varadhan 积分引理作了非平凡推广. 第 3 节主要研究大偏差速率函数的性质, 确切地说, 主要研究速率函数的极小值和极小值点, 包括全局最小值和局部极小值. 研究表明它们与转移矩阵的主特征值和相应的正特征向量密切相关. 将这些结论应用于有限状态情形时, 我们得到了较完整的系列结果.

2 经验序对与经验测度的大偏差

为了便于比较不同的过程, 我们将通过在轨道空间 $(\Omega, \mathcal{F}) = (E^{\mathbb{Z}_+}, \mathcal{E}^{\mathbb{Z}_+})$ 上定义坐标过程来构造 Markov 序列 $\{X_n, n \geq 0\}$, 即对 $\omega = (\omega_n, n \geq 0) \in \Omega$, 定义

$$X_n(\omega) = \omega_n, \quad n \geqslant 0.$$

取 $\mathcal{F}_n = \sigma\{X_k, k \leq n\}$, 对于 E 上的转移概率矩阵 $\Pi = (\pi_{ij}, i, j \in E)$, 转移函数自然定义为

$$\pi(i,B) = \sum_{j \in B} \pi_{ij}, \quad B \in \mathcal{E}.$$

对于 E 上任意概率测度 α , 由下式定义 \mathcal{F} 上的概率测度 $P_{\alpha,\Pi}$:

$$P_{\alpha,\Pi}(B_n \times E^{Z^+}) = \int \alpha(di_0) \int \pi(i_0, di_1) \cdots \int 1_{B_n} \pi(i_{n-1}, di_n), \quad B_n \in \mathcal{E}^n, \quad n \geqslant 1,$$

则 $\{X_n, n \ge 0\}$ 在 $P_{\alpha,\Pi}$ 之下是时齐的 Markov 链, 具有初始分布 α 和转移矩阵 Π . 有时也称 Π 为一 Markov 链.

如前所述, 为了证明 $\{L_n, n \ge 1\}$ 的大偏差原理, 需要用到 $\{X_n, n \ge 0\}$ 的经验序对, 定义为

$$L_n^{(2)} = L_n^{(2)}(\omega, \cdot) = \frac{1}{n} \sum_{k=1}^n \delta_{(X_{k-1}, X_k)}(\cdot).$$

注意到

$$Y_0 = (X_0, X_0), \quad Y_n = (X_{n-1}, X_n) \quad (n \geqslant 1)$$

是 $E \times E$ 上的时齐 Markov 链, 具有初始分布 $\beta(ij) = \alpha_i \delta_{ij}$ 和转移矩阵 $\hat{\Pi} = (\hat{\pi}((i_1, j_1), (i_2, j_2)))$, 其中

$$\hat{\pi}((i_1, j_1), (i_2, j_2)) = \delta_{j_1}(i_2)\pi_{i_2, j_2}.$$

设 $\Gamma = (\gamma_{ij})$ 也是 E 上的转移矩阵, 若 $\gamma_{ij} = 0 \Rightarrow \pi_{ij} = 0$, 则记 $\Pi \ll \Gamma$. 此时定义

$$\psi(i,j) = \begin{cases} \frac{\pi_{ij}}{\gamma_{ij}}, & \text{\'at } \gamma_{ij} > 0, \\ 0, & \text{\'at } \gamma_{ij} = 0, \end{cases}$$
 (2.1)

则易知对 E 上任意概率分布 α ,

$$\frac{dP_{\alpha,\Pi}}{dP_{\alpha,\Gamma}}\Big|_{\mathcal{F}_n} = \prod_{i=1}^n \psi(X_{i-1}X_i). \tag{2.2}$$

对于 $E \times E$ 上的概率测度 μ , 分别以 $\mu^{(1)}$ 和 $\mu^{(2)}$ 表示它的第一和第二边缘分布, 并定义条件分布

$$\mu(\cdot \mid i) = \begin{cases} \frac{\mu(i, \cdot)}{\mu^{(1)}(i)}, & \text{\'at } \mu^{(1)}(i) > 0, \\ \nu^*, & \text{\'at } \mu^{(1)}(i) = 0, \end{cases}$$

其中 ν^* 是 E 上任一固定的概率测度. 本节的主要结论如下:

定理 2.1 假设 Π 是 E 上的转移矩阵, α 是 E 上的概率分布. 若存在 E 上的转移矩阵 $\Gamma = (\gamma_{ij})$, 使得

- (1) $\{L_n^{(2)}, n \ge 1\}$ 在 $P_{\alpha,\Gamma}$ 之下满足大偏差原理, 具有紧的速率函数 J_{Γ} ;
- (2) $\Pi \ll \Gamma$, 且存在非降函数 g, 满足 $g(M) \to \infty$ $(M \to \infty)$, 使得

$$c \triangleq \limsup_{n \to \infty} \frac{1}{n} \log E_{\alpha,\Pi} g^n \left(\left(\frac{d P_{\alpha,\Pi}}{d P_{\alpha,\Gamma}} \Big|_{F_{\alpha}} \right)^{1/n} \right) < \infty, \tag{2.3}$$

则 $\{L_n^{(2)}, n \ge 1\}$ 在 $P_{\alpha,\Pi}$ 之下满足大偏差原理, 具有速率函数

其中 $\log 0 = -\infty$.

为证明定理,需要用到下述 Varadhan 积分引理.

定理 2.2 (Varadhan 积分引理 [11]) 设 E 是正则拓扑空间, 赋 Borel σ - 代数 \mathcal{E} , { μ_{ϵ} , $\epsilon > 0$ } 是其上的一族概率测度, 当 $\epsilon \to 0$ 时满足大偏差原理, 具有紧的速率函数 I, 则

(1) 对任意下半连续函数 φ : $E \to \mathbb{R}$ 和 E 的任意开子集 G, 有

$$\liminf_{\epsilon \to 0} \epsilon \log \int_{G} e^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \geqslant \sup_{x \in G} [\varphi(x) - I(x)]; \tag{2.5}$$

(2) 对任意上半连续函数 φ : $E \to \mathbb{R}$, 若

$$\lim_{M \to \infty} \lim_{\epsilon \to 0} \epsilon \log \int_{\varphi \geqslant M} e^{\epsilon^{-1} \varphi} d\mu_{\epsilon} = -\infty, \tag{2.6}$$

则对 E 的任意闭子集 F, 有

$$\limsup_{\epsilon \to 0} \epsilon \log \int_{F} e^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \leqslant \sup_{x \in F} [\varphi(x) - I(x)]. \tag{2.7}$$

我们需要上述引理的下述推广形式, 它的证明是本节的主要技术性工作.

命题 2.1 在上述引理中, 若允许 φ 取值于 $[-\infty, \infty]$, 即允许 φ 取 $\pm \infty$, 则相应地有

$$\liminf_{\epsilon \to 0} \epsilon \log \int_{G} e^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \geqslant -\inf_{x \in G} I_{\varphi}(x), \quad \forall \ \mathcal{H} \ G \subset E$$
 (2.8)

和

$$\limsup_{\epsilon \to 0} \epsilon \log \int_{F} e^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \leqslant -\inf_{x \in F} I_{\varphi}(x), \quad \forall \ \ \overrightarrow{\exists} \ \ F \subset E,$$
 (2.9)

其中

$$I_{\varphi}(x) = \begin{cases} I(x) - \varphi(x), & \text{若 } I(x) < \infty, \\ +\infty, & \text{否则.} \end{cases}$$
 (2.10)

若 φ 是连续的, 则 (2.8) 和 (2.9) 同时成立且 I_{φ} 是紧的.

证明 下界 (2.8) 的证明较易, 因为要证它, 只需证明

$$\liminf_{\epsilon \to 0} \epsilon \log \int_{G} e^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \geqslant -I_{\varphi}(x), \quad \forall x \in G.$$

当 $\varphi(x) = -\infty$ 或 $I(x) = +\infty$ 时, 上式显然成立, 因为此时 $I_{\varphi}(x) = +\infty$, 当 $\varphi(x) > -\infty$ 时, 证明完全与原始结论的证明一样 (参见文献 [11, 引理 4.4.3 和练习 4.3.11]).

为证上界, 先设 φ 有上界, $\varphi \leq M$, 此时 $I_{\varphi} = I - \varphi$. 对任意 $k \geq 1$, 定义 $\varphi_k = \varphi \vee \frac{1}{k}$, 则 φ_k 有界. 因此,

$$\limsup_{\epsilon \to 0} \epsilon \log \int_{F} e^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \leqslant \limsup_{\epsilon \to 0} \epsilon \log \int_{F} e^{\epsilon^{-1} \varphi_{k}} d\mu_{\epsilon} \leqslant I_{F}^{(k)}, \tag{2.11}$$

其中

$$I_F^{(k)} = \sup_{x \in F} [\varphi_k(x) - I(x)].$$

取子列 {k'}, 使得

$$\limsup_{k} I_F^{(k)} = \lim_{k'} I_F^{(k')}.$$

对任意常数 A > 0, 记 $\Psi_I(A) = \{x : I(x) \leq A\}$, 则

$$I_F^{(k)} = I_{F \cap \Psi_I(A)}^{(k)} \vee I_{F \cap [\Psi_I(A)]^c}^{(k)}. \tag{2.12}$$

对每个 k', 取 $x_{k'} \in F \cap \Psi_I(A)$, 使得

$$I_{F \cap \Psi_I(A)}^{(k')} \leqslant \varphi_{k'}(x_{k'}) - I(x_{k'}) + \frac{1}{k'}$$

由于 $\Psi_I(A)$ 是紧的, 从而, $F \cap \Psi_I(A)$ 也是紧的. 因此存在 $\{x_{k'}\}$ 的收敛子列, 不妨假设就是 $\{x_{k'}\}$ 本身. 设 $x_{k'} \to x_{\infty}$, 则 $x_{\infty} \in F \cap \Psi_I(A)$, 且

$$\limsup_{k' \to \infty} I_{F \cap \Psi_I(A)}^{(k')} \leqslant \limsup_{k' \to \infty} \left[\varphi_{k'}(x_{k'}) - I(x_{k'}) + \frac{1}{k'} \right]
\leqslant \varphi(x_{\infty}) - I(x_{\infty})
\leqslant I_F \equiv \sup_{x \in F} [\varphi(x) - I(x)].$$
(2.13)

另一方面, 当 k' 充分大时,

$$I_{F\cap[\Psi_I(A)]^c}^{(k')} \leqslant M - A.$$

从而,

$$\lim_{k' \to \infty} I_F^{(k')} \leqslant I_F \lor (M - A).$$

$$\lim_{k' \to \infty} I_F^{(k')} \leqslant I_F.$$

结合 (2.8) 即得所需结论.

一般情形下, 对任意 M > 0, 定义 $\varphi_M = \varphi \wedge M$, 则

$$\begin{split} \limsup_{\epsilon \to 0} \epsilon \log \int_{F} \mathrm{e}^{\epsilon^{-1} \varphi} d\mu_{\epsilon} &\leqslant \left[\limsup_{\epsilon \to 0} \epsilon \log \int_{F \cap \{\varphi < M\}} \mathrm{e}^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \right] \vee \left[\limsup_{\epsilon \to 0} \epsilon \log \int_{F \cap \{\varphi \geqslant M\}} \mathrm{e}^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \right] \\ &\leqslant \left[\limsup_{\epsilon \to 0} \epsilon \log \int_{F} \mathrm{e}^{\epsilon^{-1} \varphi_{M}} d\mu_{\epsilon} \right] \vee \left[\limsup_{\epsilon \to 0} \epsilon \log \int_{\{\varphi \geqslant M\}} \mathrm{e}^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \right] \\ &\leqslant \left(\sup_{x \in F} [\varphi_{M}(x) - I(x)] \right) \vee \left(\limsup_{\epsilon \to 0} \epsilon \log \int_{\{\varphi \geqslant M\}} \mathrm{e}^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \right). \end{split} \tag{2.14}$$

若在 F 上, $I \equiv +\infty$, 则由 (2.6) 知, 当 $M \to \infty$ 时,

$$(2.14) 右端 = (-\infty) \lor \left(\limsup_{\epsilon \to 0} \epsilon \log \int_{\{\varphi \geqslant M\}} e^{\epsilon^{-1}\varphi} d\mu_{\epsilon} \right) \to -\infty = -\inf_{F} I_{\varphi}.$$

若存在 $x \in F$ 使得 $I(x) < \infty$, 则

$$(2.14) \ \text{右遠} = \sup_{x \in F: I(x) < \infty} [\varphi_M(x) - I(x)] \lor \left(\limsup_{\epsilon \to 0} \epsilon \log \int_{\{\varphi \geqslant M\}} e^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \right)$$

$$\leq \sup_{x \in F: I(x) < \infty} [\varphi(x) - I(x)] \lor \left(\limsup_{\epsilon \to 0} \epsilon \log \int_{\{\varphi \geqslant M\}} e^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \right)$$

$$= \left[-\inf_{x \in F: I(x) < \infty} I_{\varphi}(x) \right] \lor \left(\limsup_{\epsilon \to 0} \epsilon \log \int_{\{\varphi \geqslant M\}} e^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \right)$$

$$\rightarrow -\inf_{x\in F} I_{\varphi}(x) \quad (M\to\infty).$$

最后证明 φ 连续时 I_{φ} 的紧性. 为此首先注意到 I_{φ} 是下半连续的. 其次, $\forall L>0$, 由 (2.6) 可以取 M 足够大, 使得

$$\limsup_{\epsilon \to 0} \epsilon \log \int_{\varphi \geqslant M} \mathrm{e}^{\epsilon^{-1} \varphi} d\mu_{\epsilon} < -L.$$

而由于 $\{\varphi > M\}$ 是开的, 因此,

$$\limsup_{\epsilon \to 0} \epsilon \log \int_{\varphi \geqslant M} \mathrm{e}^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \geqslant \liminf_{\epsilon \to 0} \epsilon \log \int_{\varphi > M} \mathrm{e}^{\epsilon^{-1} \varphi} d\mu_{\epsilon} \geqslant -\inf_{\varphi > M} I_{\varphi}.$$

上述两式表明, $\varphi(x) > M$ 时 $I_{\varphi}(x) > L$, 从而,

$${I_{\varphi} \leqslant L} = {I_{\varphi} \leqslant L, \varphi \leqslant M} \subset {I \leqslant L + M}.$$

上式右端是紧集, 因而 I_{φ} 是紧的.

定理 2.1 的证明 注意到在定理条件下, $0 \le \psi(i,j) < \infty$, 因此, 若记 $\varphi(ij) = \log \psi(ij)$ (约定 $\log 0 = -\infty$), 则 $\varphi \in [-\infty, \infty]$, 且

$$P_{\alpha,\Pi}(L_n^2 \in A) = \int_{L_n^2 \in A} e^{n \int \varphi dL_n^2} dP_{\alpha,\Gamma}.$$
 (2.15)

因为 E 可数, 在散拓扑下 φ 连续, 由命题 2.1 中的下界 (2.8) 知, 对任意开集 $G \subset M_1(E)$, 有

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathcal{P}_{\alpha,\Pi}(L_n^{(2)} \in G) \geqslant -\inf_{\mu \in G} J_{\Pi}(\mu). \tag{2.16}$$

为证上界, 由命题 2.1 知, 只需验证条件 (2.6). 注意到由定理的假设 (2), 有

$$\Delta(n) \equiv \frac{1}{n} \log \int_{\int \varphi dL_{n}^{(2)} \geqslant M} e^{n \int \varphi dL_{n}^{(2)}} dP_{\alpha,\Gamma}
= \frac{1}{n} \log P_{\alpha,\Pi} \left[\frac{1}{n} \log \left(\frac{dP_{\alpha,\Pi}}{dP_{\alpha,\Gamma}} \Big|_{\mathcal{F}_{n}} \right) \geqslant M \right]
\leqslant \frac{1}{n} \log P_{\alpha,\Pi} \left[g^{n} \left(\left(\frac{dP_{\alpha,\Pi}}{dP_{\alpha,\Gamma}} \Big|_{\mathcal{F}_{n}} \right)^{1/n} \right) \geqslant e^{n \log g(e^{M})} \right]
\leqslant -\log g(e^{M}) + \frac{1}{n} \log E_{\alpha,\Pi} g^{n} \left(\left(\frac{dP_{\alpha,\Pi}}{dP_{\alpha,\Gamma}} \Big|_{\mathcal{F}} \right)^{1/n} \right).$$
(2.17)

因此,

$$\lim_{M \to \infty} \limsup_{n \to \infty} \Delta(n) = -\infty.$$

从而, 由命题 2.1 中的上界 (2.9) 得, 对任意的闭集 $F \subset M_1(E)$, 有

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathcal{P}_{\alpha,\Pi}(L_n^{(2)} \in F) \leqslant -\inf_{\mu \in F} J_{\Pi}(\mu). \tag{2.18}$$

证毕.

以下是两个便于实际使用的推论. 第一个推论虽然平凡, 但提供了对 Π 的一种特殊的、简单的控制方式.

推论 2.1 若存在 E 上的转移矩阵 $\Gamma = (\gamma_{ij})$, 使得

- (1) $\{L_n^{(2)}, n \ge 1\}$ 在 $P_{\alpha,\Gamma}$ 之下满足大偏差原理, 具有紧的速率函数 J_{Γ} ;
- (2) 存在常数 c > 0, 使得 $\pi_{ij} \leq c\gamma_{ij}$, $\forall i, j$,

则定理 2.1 的结论成立.

证明 只需注意到若取 q(x) = x, 则定理 2.1 的条件满足即可.

第二个推论采用的是一种特殊的控制过程,即独立同分布序列,在有限状态等情形下特别有效.

推论 2.2 若存在 E 上的概率测度 γ , 使得当 $\gamma_{ij}=\gamma_{j}$ 时定理 2.1(2) 满足, 则在 $P_{\alpha,\Pi}$ 之下, $L_{n}^{(2)}$ 满足大偏差原理, 具有紧的速率函数

$$J_{\Pi}(\mu) = \begin{cases} \sum_{i=0}^{\infty} \nu_i h(\mu(\cdot \mid i) \mid \pi_{i\cdot}), & \tilde{\pi} \mid \mu^{(1)} = \mu^{(2)} = \nu, \\ +\infty, & \tilde{\pi} \downarrow \downarrow, \end{cases}$$
(2.19)

其中

$$h(\mu \mid \nu) = \begin{cases} \int \log \frac{d\mu}{d\nu} d\mu, & \stackrel{\text{Z}}{=} \mu \ll \nu, \\ +\infty, & \text{否则} \end{cases}$$
 (2.20)

是概率测度 μ 关于 ν 的相对熵, $\mu(\cdot \mid i)$ 代表给定 $\mu_i^{(1)}$ 之下的条件分布. 特别地, 若存在常数 c>0 使得

$$p_{ij} \leqslant c\gamma_i, \quad \forall i, j,$$

则上述结论成立.

证明 只需证明 $L_n^{(2)}$ 在 $P_{\alpha,\Gamma}$ 之下满足大偏差原理, 再证明在 $P_{\alpha,\Pi}$ 之下大偏差速率函数具有 (2.19) 中的形式即可.

为证明 $L_n^{(2)}$ 在 $P_{\alpha,\Gamma}$ 之下满足大偏差原理, 我们需要验证此时的 $\hat{\Gamma}$ 满足文献 [11] 中的下述假设 (U) 即可:

(U) 存在整数 $0 < l \le N$ 和常数 $M \ge 1$, 使得对任意 (i', j') 和 (i'', j''), 有

$$\hat{\gamma}^{(l)}((i',j'),\cdot) \leqslant \frac{M}{N} \sum_{k=1}^{N} \hat{\gamma}^{(k)}((i'',j''),\cdot).$$

为此, 回顾二维链 $Y_n = (X_{n-1}, X_n)$ 及其转移矩阵 $\hat{\Pi}$ 的定义. 若一维链 X_n 的转移矩阵为

$$\Gamma = (\gamma_{ij}), \quad \gamma_{ij} = \gamma_j,$$

则 Y_n 的转移矩阵为

$$\hat{\Gamma} = (\hat{\gamma}((i_1j_1), (i_2j_2))),$$

其中

$$\hat{\gamma}((i_1j_1),(i_2j_2)) = \delta_{j_1}(i_2)\gamma_{j_2}.$$

于是, 对任意 (i', j')、(i'', j'') 和 (i.j), 二步转移概率

$$\hat{\gamma}^{(2)}((i'j'),(ij)) = \gamma_i \gamma_i = \hat{\gamma}^{(2)}((i''j''),(ij)). \tag{2.21}$$

由此易知取 l = M = N = 2 时, 假设 (U) 满足, 从而, 由文献 [11, 定理 6.5.4、习题 6.5.9 和推论 6.5.10] 知, $L_n^{(2)}$ 在 $P_{\alpha,\Gamma}$ 之下满足大偏差原理, 具有紧的速率函数

$$J_{\Gamma}(\mu) = \sup \int \log \frac{d\sigma}{d\sigma \hat{\Gamma}} d\mu, \qquad (2.22)$$

其中上确界是对 $E \times E$ 上满足 $\log \frac{d\sigma\hat{\Gamma}}{d\sigma}$ 有界的所有概率测度 σ 取的. 由文献 [11, 定理 6.5.12] 知, 当 $\mu^{(1)} \neq \mu^{(2)}$ 时, $J_{\Gamma}(\mu) = \infty$; 而当 $\mu^{(1)} = \mu^{(2)}$ 时,

$$J_{\Gamma}(\mu) = \sup_{\alpha} \sum_{i,j} \mu_{ij} \log \frac{\alpha_{ij}}{(\alpha \hat{\Gamma})_{ij}}$$

$$= \sup_{\alpha} \sum_{i,j} \mu_{ij} \log \frac{\alpha_{ij}}{\sum_{(st)} \alpha_{st} \hat{\gamma}((st)(ij))}$$

$$= \sup_{\alpha} \sum_{i,j} \mu_{ij} \log \frac{\alpha_{ij}}{\sum_{(st)} \alpha_{st} \delta_{t}(i) \gamma_{j}} = \sup_{\alpha} \sum_{i,j} \mu_{ij} \log \frac{\alpha_{ij}}{\alpha_{i}^{(2)} \gamma_{j}}$$

$$= \sup_{\alpha} \left[\sum_{i} \mu_{i}^{(1)} \sum_{j} \mu(j \mid i) \log \frac{\mu(j \mid i)}{\gamma_{j}} - \sum_{i} \mu_{i}^{(1)} \sum_{j} \log \frac{\mu(j \mid i)}{\alpha_{ij} / \alpha_{i}^{(2)}} \mu(j \mid i) \right]$$

$$= \sum_{i} \mu_{i}^{(1)} h(\mu(\cdot \mid i), \gamma) - \inf_{\alpha} \sum_{i} \mu_{i}^{(1)} h\left(\mu(\cdot \mid i), \frac{\alpha_{i}}{\alpha_{i}^{(2)}}\right). \tag{2.23}$$

易验证 $h(\mu(\cdot \mid i), \alpha_i / \alpha_i^{(2)}) \ge 0$, 且当 $\alpha = \mu$ 时, $h(\mu(\cdot \mid i), \alpha_i / \alpha_i^{(2)}) = 0$. 因此,

$$J_{\Gamma}(\mu) = \sum_{i} \mu_{i}^{(1)} h(\mu(\cdot \mid i), \gamma).$$

于是,

$$J_{\Pi}(\mu) = J_{\Gamma}(\mu) - \int \log \pi d\mu$$

$$= J_{\Gamma}(\mu) - \sum_{ij} \mu_{i,j} \log \frac{\pi_{i,j}}{\gamma_{j}}$$

$$= \sum_{i} \mu_{i}^{(1)} \left[\sum_{j} \mu(j \mid i) \log \frac{\mu(j \mid i)}{\gamma_{j}} - \sum_{j} \mu(j \mid i) \log \frac{\pi_{i,j}}{\gamma_{j}} \right]$$

$$= \sum_{i} \mu_{i}^{(1)} \sum_{j} \mu(j \mid i) \log \frac{\mu(j \mid i)}{\pi_{i,j}}$$

$$= \sum_{i} \mu_{i}^{(1)} h(\mu(\cdot \mid i), \pi_{i}). \tag{2.24}$$

定理证毕.

由定理 2.1 和下述压缩原理即可得到经验测度的大偏差原理.

定理 2.3 (压缩原理 [11]) 设 X 和 Y 是两个 Hausdorff 拓扑向量空间, $\{\mu_{\epsilon}, \epsilon > 0\}$ 是 X 上的一族概率测度, f 是 X 到 Y 的连续函数. 若当 $\epsilon \to 0$ 时, μ_{ϵ} 满足大偏差原理, 具有紧的速率函数 J, 则 $\nu_{\epsilon} \triangleq \mu_{\epsilon} \circ f^{-1}$ 在 Y 上也满足大偏差原理, 具有紧的速率函数

$$I(y) = \inf\{J(x), x \in X, f(x) = y\}.$$

推论 2.3 在上述记号下, 若定理 2.1 的假设成立, 则 $\{L_n, n \ge 1\}$ 在 $P_{\alpha,\Pi}$ 之下满足大偏差原理, 具有速率函数

$$I_{\Pi}(\nu) = \inf_{\mu \in M_1(E \times E): \mu^{(1)} = \nu} J_{\Pi}(\mu).$$

3 速率函数

3.1 速率函数的零点

如第 1 节所叙述的, 本节主要研究第 2 节中得到的大偏差速率函数的极小值和极小值点. 由于大偏差原理成立, 因此,

$$\inf_{M_1(E\times E)}J_\Pi=\inf_{M_1(E)}I_\Pi=0.$$

此时的极小值点就是速率函数的零点,它们刻画了系统的极限状态,在大偏差研究中占有特殊地位,因此确定它们具有重要意义. 形如 (2.19) 的 J_{Π} 是经验序对大偏差速率函数的典型形式,我们先研究其零点.

引理 3.1 设 J_{Π} 由 (2.19) 给出. 若 $J_{\Pi}(\mu) = 0$, 则 $\mu^{(1)} = \mu^{(2)}$ 是 Π 的平稳分布.

证明 由定义知, 若 $J_{\Pi}(\mu) = 0$, 则必有 $\mu^{(1)} = \mu^{(2)}$, 且 $\mu_i^{(1)} h(\mu(\cdot \mid i), \pi_{i\cdot}) = 0$, $\forall i$. 因此, 当 $\mu_i^{(1)} > 0$ 时, $h(\mu(\cdot \mid i), \pi_{i\cdot}) = 0$, 这意味着 $\mu(\cdot \mid i) = \pi_{i\cdot}$, 从而,

$$\mu_{i,j} = \mu_i^{(1)} \pi_{i,j}, \quad \forall j;$$

而当 $\mu_i^{(1)} = 0$ 时上式显然也成立. 于是, 对任意 j, 上式两边对 i 求和得

$$\mu_j^{(1)} = \mu_j^{(2)} = \sum_i \mu_i^{(1)} \pi_{i,j},$$

这表明 $\mu^{(1)} = \mu^{(2)}$ 是 Π 的平稳分布.

下一定理刻画上述 J_{Π} 的零点. 定理结论表明, J_{Π} 的零点恰好是 $\hat{\Pi}$ 的平稳分布, 可由 Π 的平稳分布得到. 先引进一个记号. 对于 $\mu \in M_1(E \times E)$, 若 $\mu^{(1)} = \mu^{(2)}$, 定义 $\hat{\mu} \in M_1(E \times E)$ 如下:

$$\hat{\mu}_{i,j} = \mu_i^{(1)} \pi_{i,j}, \quad (i,j) \in E \times E.$$
 (3.1)

П

定理 3.1 在上述记号下,以下结论等价:

- (1) $J_{\Pi}(\hat{\mu}) = 0$;
- (2) $\hat{\mu}$ 是 $\hat{\Pi}$ 的平稳分布;
- (3) $\mu^{(1)} = \mu^{(2)}$ 是 Π 的平稳分布.

证明 (1) 若 $J_{\Pi}(\hat{\mu}) = 0$, 首先证明 $\hat{\mu}^{(2)} = \hat{\mu}^{(1)} = \mu^{(1)} = \mu^{(2)}$. 第一个等号由 $J_{\Pi}(\hat{\mu})$ 的定义和 $J_{\Pi}(\hat{\mu}) = 0$ 得到, 第三个等号由 $\hat{\mu}$ 的定义中对 $\hat{\mu}$ 的要求得到. 只需证明第二个等号. 由 J_{Π} 和 $\hat{\mu}$ 的定义知.

$$0 = J_{\Pi}(\hat{\mu}) = \sum_{i} \hat{\mu}_{i}^{(1)} h(\hat{\mu}(\cdot \mid i), \pi_{i}.)$$
$$= \sum_{i} \hat{\mu}_{i}^{(1)} \sum_{i} \frac{\hat{\mu}_{i,j}}{\hat{\mu}_{i}^{(1)}} \log \frac{\hat{\mu}_{i,j}/\hat{\mu}_{i}^{(1)}}{\pi_{i,j}}$$

$$= \sum_{i} \hat{\mu}_{i}^{(1)} \sum_{j} \frac{\hat{\mu}_{i,j}}{\hat{\mu}_{i}^{(1)}} \log \frac{\mu_{i}^{(1)} \pi_{i,j} / \hat{\mu}_{i}^{(1)}}{\pi_{i,j}}$$

$$= \sum_{i} \hat{\mu}_{i}^{(1)} \sum_{j} \frac{\mu_{i}^{(1)} \pi_{i,j}}{\hat{\mu}_{i}^{(1)}} \log \frac{\mu_{i}^{(1)} \pi_{i,j} / \hat{\mu}_{i}^{(1)}}{\pi_{i,j}}$$

$$= \sum_{i} \mu_{i}^{(1)} \log \frac{\mu_{i}^{(1)}}{\hat{\mu}_{i}^{(1)}} = h(\mu^{(1)}, \hat{\mu}^{(1)}). \tag{3.2}$$

因此, $\hat{\mu}^{(1)} = \mu^{(1)}$. 于是, 对任意的 (i, j), 有

$$(\hat{\mu}\hat{\Pi})(i,j) = \sum_{(s,t)} \hat{\mu}_{s,t} \hat{\Pi}((s,t),(i,j)) = \sum_{(s,t)} \hat{\mu}_{s,t} \delta_t(i) \pi_{i,j}$$

$$= \sum_s \hat{\mu}_{s,i} \pi_{i,j} = \hat{\mu}_i^{(2)} \pi_{i,j} = \mu_i^{(1)} \pi_{i,j} = \hat{\mu}_{i,j},$$
(3.3)

即 $\hat{\mu}$ 是 $\hat{\Pi}$ - 平稳的.

(2) 现在设 $\hat{\mu}$ 是 $\hat{\Pi}$ - 平稳的. 对任意的(i,j), 有

$$\begin{split} \mu_i^{(1)} \pi_{i,j} &= \hat{\mu}_{i,j} = (\hat{\mu} \hat{\Pi})(i,j) \\ &= \sum_{(s,t)} \hat{\mu}_{s,t} \hat{\pi}((s,t),(i,j)) \\ &= \sum_{(s,t)} \mu_s^{(1)} \pi_{s,t} \delta_t(i) \pi_{i,j} \\ &= \left[\sum_s \mu_s^{(1)} \pi_{s,i} \right] \pi_{i,j} = (\mu^{(1)} \Pi)_i \pi_{i,j}, \end{split}$$

于是, 对任意的 i, 取 j 使得 $\pi_{i,j} > 0$, 即得 $\mu_i^{(1)} = (\mu^{(1)}\Pi)_i$, 因此, $\mu^{(1)}$ 是 Π - 平稳的.

(3) 最后, 设 $\mu^{(1)} = \mu^{(2)}$ 是 Π- 平稳的. 首先由 $\hat{\mu}$ 的定义易知, $\hat{\mu}^{(1)} = \mu^{(1)}$. 事实上, 也有 $\hat{\mu}^{(2)} = \mu^{(2)}$, 这是因为

$$\hat{\mu}_j^{(2)} = \sum_i \hat{\mu}_{i,j} = \sum_i \mu_i^{(1)} \pi_{i,j} = \mu_j^{(1)} = \mu_j^{(2)}.$$

因此,

$$\hat{\mu}_{i,j} = \mu_i^{(1)} \pi_{i,j} = \hat{\mu}_i^{(1)} \pi_{i,j},$$

从而, 对任意的 i, 有 $\hat{\mu}_i^{(1)} h(\hat{\mu}(\cdot \mid i), \pi_i \cdot) = 0$. 于是, 由 J_{Π} 的定义即知, $J_{\Pi}(\hat{\mu}) = 0$.

接下来刻画 I_{Π} 的零点.

定理 3.2 $I_{\Pi}(\nu) = 0$ 当且仅当 ν 是 Π - 平稳的.

证明 若 ν 是 Π - 平稳的, 定义 $\mu = \nu \times \nu \in M_1(E \times E)$, 则 $\mu^{(1)} = \mu^{(2)} = \nu$ 是 Π - 平稳的. 因此, 由定理 3.1 知, $J_{\Pi}(\hat{\mu}) = 0$, 且 $\hat{\mu}^{(1)} = \mu^{(1)} = \nu$. 于是, 由 I_{Π} 的定义即知 $I_{\Pi}(\nu) = 0$.

另一方面, 若 $I_{\Pi}(\nu) = 0$, 则存在 $\mu_n \in M_1(E \times E)$, $\mu_n^{(1)} = \nu$ 且 $J_{\Pi}(\mu_n) \to 0$ $(n \to \infty)$. 由 J_{Π} 的紧性知, 存在子列 $\{\mu_{n_k}, k \ge 1\}$ 使得 $\mu_{n_k} \Rightarrow$ 某 $\mu \in M_1(E \times E)$, 显然 $\mu^{(1)} = \nu$, 且由 J_{Π} 的下半连续性知 $J_{\Pi}(\mu) = 0$. 由引理 3.1 知, $\mu^{(1)} = \nu$ 是 Π - 平稳的.

3.2 速率函数的局部极小值

零点是速率函数的最小值点, 上面的结果表明它们通常是过程的平稳分布. 那么速率函数的局部极小值和极小值点代表了什么呢? 下面以有限状态情形为例进行探讨.

设 Π 是有限状态空间 E 上的 Markov 链, $\hat{\Pi}$ 、 L_n 和 $L_n^{(2)}$ 等如前定义. 显然推论 2.1 的条件满足, 因此有下述定理:

定理 3.3 对任意初始分布 $\alpha \in M_1(E)$, $\{L_n^{(2)}, n \geq 1\}$ 在 $P_{\alpha,\Pi}$ 之下满足大偏差原理, 具有凸的紧速率函数 J_{Π} , 它不依赖于 α , 具有 (2.19) 中的形式. $\{L_n, n \geq 1\}$ 也满足大偏差原理, 具有凸的紧速率函数 I_{Π} .

由于 E 有限, Π 一定存在平稳分布, 因此, 定理 3.2 表明 0 是 I_{Π} 在 E 上的最小值, 而 Π 的平稳分布则是最小值点. 现在设 $E_0 \subset E$, 考虑极小值

$$\lambda_0 = \inf_{\nu(E_0)=1} I_{\Pi}(\nu)$$

及其极小值点的刻画. 仍由于 E 有限, 这一极小值一定能取到, 即一定存在 E_0 上的概率测度 ν , 使得 $I_{\Pi}(\nu) = \lambda_0$. 我们将证明极小值 λ_0 和极小值点 ν 与 Π 在 E_0 上的谱性质密切相关. 为此先引进一些记号. 以 Π_{E_0} 表示 Π 在 E_0 上的限制, $\rho_0 = \rho(\Pi_{E_0})$ 是 Π_{E_0} 的谱半径, 并假定 $\rho_0 > 0$. 由非负矩阵理论 (参见文献 [14,15]) 知, Π_{E_0} 存在相应于 ρ_0 的非负非零 (右) 特征向量 r. 记

$$E_0^{r,+} = \{ i \in E_0, r_i > 0 \},\$$

则 r 实际上也是 $\Pi_{E_0^{r,+}}$ 的相应于 ρ_0 的 (Ξ) 特征向量. 因此, 仍由矩阵理论知, $\rho(E_0^{r,+})=\rho_0$, 从而, $\Pi_{E_0^{r,+}}$ 存在相应于 ρ_0 的非负非零 (Ξ) 特征向量 l. 于是, 若定义

$$\bar{\nu}_i = l_i r_i, \quad i \in E_0^{r,+},$$

则 $\bar{\nu}$ 是 $E_0^{r,+}$ 上一个非平凡测度, 不妨设它已归一化成 $E_0^{r,+}$ 上的概率测度.

注意到我们关注的是 $\inf_{\nu(E_0)=1} I_{\Pi}(\nu)$, 而 $E_0^{r,+} \subset E_0$, 因此我们将需要下述引理:

引理 3.2 设 $E_1 \subset E_2 \subset E$, 对于 k = 1, 2 及 $E_k \times E_k$ 上的概率测度 μ , 定义

$$J_k(\mu) = \inf_{u: u_{i,j} > 0, i, j \in E_k} \int_{E_k \times E_k} \log \frac{\hat{\Pi}_{E_k} u}{u} d\mu,$$

则当 $\mu(E_1 \times E_1) = 1$ 时, $J_1(\mu) = J_2(\mu)$.

证明 一方面, 由于 $\mu(E_1 \times E_1) = 1$, 因此,

$$J_{2}(\mu) = -\inf_{u_{i,j}>0, i, j \in E_{2}} \sum_{i,j \in E_{1}} \log \left[\frac{\sum_{s,t \in E_{1}} \hat{\pi}((i,j),(s,t)) u_{s,t}}{u_{i,j}} + \frac{\sum_{s,t \in E_{2}-E_{1}} \hat{\pi}((i,j),(s,t)) u_{s,t}}{u_{i,j}} \right] \mu_{i,j}$$

$$\leq -\inf_{u_{i,j}>0, i, j \in E_{1}} \sum_{i,j \in E_{1}} \log \left[\frac{\sum_{s,t \in E_{1}} \hat{\pi}((i,j),(s,t)) u_{s,t}}{u_{i,j}} \right] \mu_{i,j}$$

$$= J_{1}(\mu);$$

另一方面, 任意给定 $E_1 \times E_1$ 上的严格正函数 u, 即 $u_{i,j} > 0$, $\forall i,j \in E_1$, 对任意 $\epsilon > 0$, 将 u 如下延拓 成 $E_2 \times E_2$ 上的严格正函数 $u^{(\epsilon)}$:

$$u_{i,j}^{(\epsilon)} = \begin{cases} u_{i,j}, & \text{若 } i, j \in E_1, \\ \epsilon, & \text{否则}, \end{cases}$$

则由 J2 的定义知,

$$J_{2}(\mu) \geqslant -\int_{E_{1} \times E_{1}} \log \frac{\hat{\Pi}_{E_{2}} u^{(\epsilon)}}{u^{(\epsilon)}} d\mu$$

$$= -\sum_{i,j \in E_{1}} \log \left[\frac{\sum_{s,t \in E_{1}} \hat{\pi}((i,j),(s,t)) u_{s,t}}{u_{i,j}} + \frac{\epsilon \sum_{s \neq t \in E_{2} - E_{1}} \hat{\pi}((i,j),(s,t))}{u_{s,t}} \right] \mu_{i,j}$$

$$= -\sum_{i,j \in E_{1}} \log \left[\frac{\hat{\Pi}_{E_{1}} u}{u} + \frac{\epsilon \sum_{s \neq t \in E_{2} - E_{1}} \hat{\pi}((i,j),(s,t))}{u_{i,j}} \right] \mu_{i,j}.$$

$$J_2(\mu) \geqslant -\int_{E_1 \times E_1} \log \frac{\hat{\Pi}_{E_1} u}{u} d\mu.$$

再由 u 的任意性知,

$$J_2(\mu) \geqslant J_1(\mu).$$

综合上述结果即得 $J_1(\mu) = J_2(\mu)$.

现在可以叙述这一部分的主要定理.

定理 3.4 若 $E_0 \subset E$, $\rho_0 = \rho(\Pi_{E_0})$ 是 Π_{E_0} 的谱半径, 则

$$\inf_{\nu(E_0)=1} I_{\Pi}(\nu) = -\log \rho_0.$$

进一步, 若 r 是 Π_{E_0} 相应于 ρ_0 的任一右特征向量, $E_0^{r,+} = \{i \in E_0 : r_i > 0\}$, l 是 $\Pi_{E_0^{r,+}}$ 相应于 ρ_0 的 非负非零左特征向量, $\bar{\nu}$ 如上定义, 则

$$I_{\Pi}(\bar{\nu}) = -\log \rho_0 = \inf_{\nu(E_0)=1} I_{\Pi}(\nu).$$

证明 首先注意到, 由于推论 2.1 中的 Γ 满足假设 (U), 因此, 由文献 [11, 推论 6.5.10] 知,

$$J_{\Gamma}(\mu) = -\inf_{\{u: u_{i,j} > 0, i, j \in E\}} \int \log \frac{\hat{\Gamma}u}{u} d\mu.$$

从而,由(2.4)知,

$$J_{\Pi}(\mu) = \begin{cases} -\inf_{\{u: u_{i,j} > 0, i, j \in E\}} \int \log \frac{\hat{\Pi}u}{u} d\mu, & \tilde{\Xi} \ \mu^{(1)} = \mu^{(2)} = \nu \ \text{\mathbb{L}} \ J_{\Gamma}(\mu) < \infty, \\ +\infty, & \tilde{\Xi} \text{\mathbb{M}}. \end{cases}$$
(3.4)

在 $E_0^{r,+}$ 上定义矩阵 $r\Pi$ 如下:

$$_{r}\pi_{i,j} = \frac{\pi_{i,j}r_{j}}{\rho_{0}r_{i}},$$
 (3.5)

则 $_r\Pi$ 是 $E_0^{r,+}$ 上的转移概率矩阵. 由 (3.4) 易验证, 若 $\mu(E_0^{r,+}\times E_0^{r,+})=1$ 且 $\mu^{(1)}=\mu^{(2)}$, 则由引理 3.2 可知,

$$J_{r\Pi}(\mu) = -\inf_{\{u: u(i,j) > 0, (i,j) \in E_0^{r,+} \times E_0^{r,+}\}} \int \log \frac{\hat{\Pi}(ur)}{ur} d\mu + \log \rho_0$$

$$= -\inf_{\{u: u(i,j) > 0, (i,j) \in E \times E\}} \int \log \frac{\hat{\Pi}u}{u} d\mu + \log \rho_0$$

= $J_{\Pi}(\mu) + \log \rho_0$, (3.6)

其中 $(ur)(i,j) = u(i,j)r_j$ 在 $E_0^{r,+} \times E_0^{r,+}$ 上严格正. 于是, 相应地, 若 $\nu(E_0^{r,+}) = 1$, 则

$$I_{r\Pi}(\nu) = I_{\Pi}(\nu) + \log \rho_0 \geqslant 0. \tag{3.7}$$

由于 $I_{r\Pi} \ge 0$, 因此,

$$\inf_{\nu(E_0)=1} I_{\Pi}(\nu) \geqslant -\log \rho_0.$$

另一方面, 易验证由 l 和 r 定义的 $\bar{\nu}$ 是 $_r\Pi$ 的平稳分布. 因此, 由定理 3.3 知, $I_{r\Pi}(\bar{\nu})=0$. 于是, $I_{\Pi}(\bar{\nu})=-\log\rho_0$, 从而,

$$\inf_{\nu(E_0)=1} I_{\Pi}(\nu) = -\log \rho_0 = I_{\Pi}(\bar{\nu}).$$

证毕.

关于极小值点的唯一性, 我们有下述定理:

定理 3.5 设 E_0 、 ρ_0 、r 和 $E_0^{r,+}$ 如定理 3.3 后面所述. 若 $E_0^{r,+} = E_0$,则 $I_{\Pi}(\nu) = -\log \rho_0$, $\nu(E_0) = 1$ 有唯一解当且仅当 Π_{E_0} 对应于 ρ_0 的非负非零左特征向量 l 唯一 (不计常数因子). 特别地,若 Π_{E_0} 不可约,则 I_{Π} 在 E_0 上的极小值点 $\bar{\nu}$ 唯一,它由 Π_{E_0} 相应于 ρ_0 的唯一左特征向量 l 和唯一右特征向量 r 决定.

证明 设 l 唯一. 若 $\nu(E_0) = \bar{\nu}(E_0) = 1$, 且 $I_{\Pi}(\nu) = I_{\Pi}(\bar{\nu}) = -\log \rho_0$, 则由定理 3.3 的证明知, ν 和 $\bar{\nu}$ 都是 $_r\Pi$ 的平稳分布. 因此, $\forall j \in E_0$,

$$\sum_{i \in F_0} \nu_i \frac{\pi_{i,j} r_j}{\rho_0 r_i} = \nu_j.$$

这表明 $(\nu_i/r_i, i \in E_0)$ 是相应于 ρ_0 的非负非零左特征向量. 同理, $(\bar{\nu}_i/r_i, i \in E_0)$ 也是相应于 ρ_0 的非负非零左特征向量. 由唯一性即知, $\nu = \bar{\nu}$.

现在设 $I_{\Pi}(\nu) = -\log \rho_0$ ($\nu(E_0) = 1$) 的解唯一, 则 $_r\Pi$ 的平稳分布唯一. 若 l 和 \bar{l} 都是相应于 ρ_0 的非负非零左特征向量, 则易验证 $lr \triangleq (l_i r_i, i \in E_0)$ 和 $\bar{l}r \triangleq (\bar{l}_i r_i, i \in E_0)$ 都是 $_r\Pi$ 的非零不变测度. 于是, $c_l lr$ 和 $c_{\bar{l}} \bar{l}r$ 都是 $_r\Pi$ 的平稳分布. 由唯一性知, $c_l lr = c_{\bar{l}} \bar{l}r$. 这表明 l 和 \bar{l} 相差常数因子.

若 Π_{E_0} 不可约, 则由经典的 Perron-Frobenius 理论知, Π_{E_0} 存在唯一的相应于 ρ_0 的左特征向量 l 和右特征向量 r, 它们在 E_0 上严格正. 此时, 由 (3.5) 定义的转移矩阵 $_r\Pi$ 在 E_0 上不可约, 具有唯一的平稳分布 $\bar{\nu}$, 它是 I_Π 的极小值点. 若 $\hat{\nu}$ 也是 I_Π 的极小值点, 即 $I_\Pi(\hat{\nu}) = -\log \rho_0$, 则由 (3.7) 知, $I_{r\Pi}(\hat{\nu}) = 0$. 从而, 由定理 3.2 知, $\hat{\nu}$ 是 $_r\Pi$ 的平稳分布. 由 $_r\Pi$ - 平稳分布的唯一性知, $\hat{\nu} = \bar{\nu}$.

注 3.1 若 Π_{E_0} 可约, 则 I_{Π} 在 E_0 上的极小值点可以唯一, 也可以不唯一. 下面是不唯一的例子. **例 3.1** 考虑如下转移矩阵:

$$\Pi = \begin{bmatrix}
1 & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 \\
\frac{1}{2} & 0 & 0 & \frac{1}{2} \\
\frac{1}{2} & 0 & 0 & \frac{1}{2}
\end{bmatrix}$$
(3.8)

及其子矩阵

$$\Pi_{0} = \Pi_{E_{0}} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \end{bmatrix},$$
(3.9)

其谱半径 $\rho_0 = 1/2$, 具有形如 $r = (a,b,b)^{\rm T}$ (a,b>0) 的正的右特征向量和形如 $l = (c,0,d)^{\rm T}$ (c,d>0) 的非负非零左特征向量. 取 $l = (1,0,1)^{\rm T}$, 若取 $r = (1,1,1)^{\rm T}$, 则得 $I_{\rm II}$ 在 E_0 上的一个极小值点 $\bar{\nu} = (1/2,0,1/2)^{\rm T}$. 若换 r 为 $(1,2,2)^{\rm T}$, 则得极小值点 $\bar{\nu} = (1/3,0,2/3)^{\rm T}$.

以下是极小值点唯一的例子.

例 3.2 考虑如下转移矩阵:

$$\Pi = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & \frac{1}{2} & \frac{1}{2} \\
0 & \frac{1}{2} & 0 & \frac{1}{2}
\end{bmatrix}$$
(3.10)

及其子矩阵

$$\Pi_{0} = \Pi_{E_{0}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix},$$
(3.11)

则 $\rho_0 = 1$, 它有唯一的严格正的右特征向量 $r = (1,1,1)^{\rm T}$ 和唯一的非负非零左特征向量 l = (1,0,0). 从而极小值点唯一, 即为 l.

参考文献 -

- 1 Donsker M D, Varadhan S R S. Asymptotic evaluation of certain Markov process expectations for large time, I. Comm Pure Appl Math, 1975, 28: 1–47
- 2 Donsker M D, Varadhan S R S. Asymptotic evaluation of certain Markov process expectations for large time, II. Comm Pure Appl Math, 1975, 28: 279–301
- 3 Donsker M D, Varadhan S R S. Asymptotic evaluation of certain Markov process expectations for large time, III. Comm Pure Appl Math, 1976, 29: 389–461
- 4 Donsker M D, Varadhan S R S. Asymptotic evaluation of certain Markov process expectations for large time, IV. Comm Pure Appl Math, 1983, 36: 183–212
- 5 Ventsel' A D. Rough limit theorems on large deviations for Markov stochastic processes, I. Theory Probab Appl, 1977, 21: 227–242
- 6 Ventsel' A D. Rough limit theorems on large deviations for Markov stochastic processes, II. Theory Probab Appl, 1977, 21: 499–512
- 7 Ventsel' A D. Rough limit theorems on large deviations for Markov stochastic processes, III. Theory Probab Appl, 1979, 24: 675–692
- 8 Ventsel' A D. Rough limit theorems on large deviations for Markov stochastic processes, IV. Theory Probab Appl, 1983, 27: 215–234
- 9 Freidlin M I, Wentzell A D. Random Perturbations of Dynamical Systems. New York: Springer, 1984

- 10 Deuschel J, Stroock D W. Large Deviations. Boston: Academic Press, 1989
- 11 Dembo A, Zeitouni O. Large Deviations Techniques and Applications. Boston: Jones and Bartlett, 1998
- 12 Donsker M D, Varadhan S R S. On the principal eigenvalue of second-order elliptic differential operators. Comm Pure Appl Math, 1976, 29: 595–621
- 13 Ellis R S. Large deviations for the empirical measure of a Markov chain with an application to the multivariate empirical measure. Ann Probab, 1988, 16: 1496–1508
- 14 Seneta E. Nonnegative Matrices and Markov Chains. New York: Springer, 1981
- 15 Schaefer Helmut H. Banach Lattices and Positive Operators. New York-Heidelberg: Springer, 1974

Large deviations for Markov chains and spectrum of matrices

Jinwen Chen

Abstract This paper deals with large deviations and related problems for Markov chains with countable state spaces. By comparing with some dominating chains that satisfy the large deviation principle, we weaken the requirement for irreducibility to get the large deviation principle. In this way we can also characterize the stationary distributions of the chains via the large deviation rate functions. In particular, we prove that any finite Markov chain satisfies the large deviation principle. The global and local minimums of the rate function are shown to be closely related to the principal eigenvalues and the corresponding left and right eigenfunctions of the transition matrix.

Keywords Markov process, large deviation, entropy MSC(2010) 60J10, 60F10, 28Dxx doi: 10.1360/N012019-00019