Chapter 2

马氏链的德布林理论

2.1 离散时间马尔科夫过程（马氏链）的基本概念

一个马氏链定义在可数的状态空间 $\mathbb{S}$ 上，由一列取值在 $\mathbb{S}$ 上的随机变量 $\{X_n : n \geq 0\}$ 构成。我们要求 X_n 满足所谓的马氏性，也就是说，存在一个矩阵[1]$\mathbf{P}$，其各项都为非负且每行之和都为 1，对于所有的 $n \geq 0$ 和 $\{i_0, i_1, \ldots, i_n, j\} \subseteq \mathbb{S}$，

$$\mathbb{P}(X_{n+1} \mid X_0 = i_0, \ldots, X_n = i_n) = (\mathbf{P})_{i_n, j}, \tag{2.1}$$

或者，如果对条件概率熟悉的话，可以等价地写成

$$\underbrace{\mathbb{P}(X_{n+1} = j \mid \overbrace{X_0, \ldots, X_n}^{\text{生成一个 } \sigma\text{-代数}})}_{\sigma(X_1, \ldots, X_n) \text{ 上的函数}} = \underbrace{(\mathbf{P})_{X_n, j}}_{\text{只是 } X_n \text{ 的函数}}.$$

也就是说，每一步跳跃的结果，只取决于起跳时的状态，而与达到起跳状态的历史无关。

如果无特殊要求，我们假定 $\mathbb{S}$ 是 $\{1, 2, \ldots, n\}$ 或者 $\{1, 2, \ldots\} = \mathbb{Z}_+$。

2.1.1 马氏链的存在性

上面提到的矩阵 $\mathbf{P}$ 称为转移概率矩阵。

是不是每个非负的，每行之和为 1 的正方形矩阵（$n \times n$ 或者 $\infty \times \infty$）都定义一个马氏链？是否 X_0 的分布可以任意？

[1]这里我们容许无穷维矩阵。

我们下面给一个构造性的肯定回答。这个构造建立在一个基本假定：在某概率空间，存在一列相互独立的随机变量 $\{U_n : n \geq 0\}$ 均匀分布在 $[0,1)$ 上。(这个是测度论的基本结论，在此不证，但是这是一个需要证明的结果!)

我们令 X_0 为 U_0 的函数，(X_0, X_1) 为 (U_0, U_1) 的函数，⋯⋯，这样定义 $(X_0, X_1, \ldots, X_n)$ 的联合分布。

1. X_0 可以有一个任意的*初始分布* μ：$\mu(i) = a_i$（$i \in S$）。这样，定义 $\alpha_0 = 0$，$\alpha_i = a_1 + \cdots + a_i$（$i \in S$），最后 X_0 依赖于 U_0：如果 $U_0 \in [\alpha_{i-1}, \alpha_i)$，则 $X_0 = i$。

2. (X_0, X_1) 也取决于 μ，但是也依赖于 $\mathbf{P}$：对于任意 $i \in S$ 我们定义 $\beta_{i0} = 0$ 且对于 $j \in S$，$\beta_{ij} = (\mathbf{P})_{i1} + \cdots + (\mathbf{P})_{ij}$。则 (X_0, X_1) 依赖于 (U_0, U_1)：
$$U_0 \in [\alpha_{i-1}, \alpha_i) \& U_1 \in [\beta_{j-1}, \beta_j) \implies X_0 = i \& X_1 = j.$$

3. (X_0, X_2, X_3) 以至 $(X_0, X_1, \ldots, X_n)$ 的定义：练习。

可以验证，这样定义的 $\{X_n\}$ 满足(2.1)。

2.1.2 转移概率和概率向量

如果 X_0 的初始分布为 $\mu(i) = \mu_i$，则
$$\mathbb{P}(X_0 = i, X_1 = j) = \mu_i (\mathbf{P})_{ij},$$
并且通过对马氏性作归纳法，
$$\begin{aligned}
&\mathbb{P}(X_0 = i, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = j) \\
&= \mathbb{P}(X_0 = i, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}) (\mathbf{P})_{i_{n-1}, j} \\
&= \cdots \\
&= (\mathbf{P})_{i, i_1} (\mathbf{P})_{i_1, i_2} \cdots (\mathbf{P})_{i_{n-1}, j}.
\end{aligned}$$
我们可以把 $\mathbf{P}$ 视作一个矩阵，于是由矩阵乘法计算可以定义 $\mathbf{P}^n$。因为我们的状态空间是 $\mathbb{S} = \{1, 2, \ldots\}$，很自然地我们用行向量表示 $\mathbb{S}$ 上的概率分布。比如，X_0 的初始分布是 μ，我们可以把 μ 理解为一个行向量 $\mu = (\mu_1, \mu_2, \ldots) = (\mu(1), \mu(2), \ldots)$，于是 X_1 的分布也可以用行向量 $\mu\mathbf{P}$ 表示（$\mathbb{P}(X_1 = j) = \sum_{i \in \mathbb{S}} \mu_i (\mathbf{P})_{ij} = (\mu\mathbf{P})_j$）。类似地，$X_n = j$ 的概率为
$$\sum_{i_0 \in S} \cdots \sum_{i_{n-1} \in S} \mathbb{P}(X_0 = i, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = j) = (\mu\mathbf{P}^n)_j.$$

如果一个向量 v 表示一个 S 值随机变量的分布，则它在 $\ell^1(\mathbb{S})$ 这个巴拿赫空间里，且其 ℓ_1 范数为 1。对于任意向量 $v \in \ell^1(\mathbb{S})$（有可能没有概率意义），如果 $\mathbf{P}$ 是一个 S 上概率转移矩阵，则我们都有不等式（我们省略 $\ell^1(\mathbb{S})$ 为 ℓ^1 空间。这在状态空间无限是是完全正确的，有限情况也不会有任何问题）

$$\|v\mathbf{P}\|_1 = \sum_{j\in S}\left|\sum_{i\in S} v_i(\mathbf{P})_{ij}\right| \le \sum_{i\in S}\left(\sum_{j\in S}|v_i|(\mathbf{P})_{ij}\right) = \|v\|_1.$$

所以 $\mathbf{P}$ 是 ℓ^1 空间上的有界线性算子，其范数不大于 1。

2.1.3 转移概率和转移函数

我们把状态空间上的函数 $f : S \to \mathbb{R}$ 视作列向量 $f = (f_1, f_2, \ldots)^T = (f(1), f(2), \ldots)^T$。这样如果 X_0 的分布函数是 μ，且我们把它理解为行向量，则 $\mathbb{E}[f(X_0)] = \mu f$，也就是 μ 和 f 这两个向量的（标量）积[2]。如果指定初始条件 $X_0 = i$，则在时间 n 时，$f(X_n)$ 的条件期望为

$$\mathbb{E}[f(X_n) \mid X_0 = i] = \sum_{j\in S} f(j)\mathbb{P}(X_n = j \mid X_0 = j) = \sum_{j\in S} f_j(\mathbf{P}^n)_{ij} = (\mathbf{P}^n f)_i,$$

这里 $\mathbf{P}$ 是 $|S|$ 维矩阵，而 f 为 $|S|$ 维列向量，所以其乘积为列向量。类似地，根据马氏性我们有

$$\begin{aligned}
&\mathbb{E}[f(X_n) \mid X_0 = i_0, \ldots, X_m = i_m] \\
&= \sum_{j\in S} f(j)\mathbb{P}(X_n = j \mid X_0 = i_0, \ldots, X_m = i_m) \\
&= \sum_{j\in S} f_j \underbrace{(\mathbf{P}^{n-m})_{i_m,j}}_{\mathbb{P}(X_n=j\mid X_0=i_0,\ldots,X_m=i_m)} = (\mathbf{P}^{n-m} f)_{i_m},
\end{aligned}$$

[2]为了让这个乘积有意义，如果 $S = \mathbb{N}$，则我们要求或者 f 是非负函数，于是根据单调收敛定理

$$\sum_{i=1}^{\infty} \mu_i f_i = \lim_{n\to\infty} \sum_{i=1}^{n} \mu_i f_i$$

存在，或者要求 f 是有界函数，于是根据勒贝格控制收敛定理（$g(x) \equiv \sup|f(i)|$）

$$\sum_{i=1}^{\infty} \mu_i f_i = \lim_{n\to\infty} \sum_{i=1}^{n} \mu_i f_i$$

存在。

和等价的

$$\mathbb{E}[f(X_n) \mid X_0, \ldots, X_m] = (\mathbf{P}^{n-m} f)_{X_m}, \tag{2.2}$$

以及，如果 X_0 的初始分布是 μ，则

$$\mathbb{E}[f(X_n)] = \sum_{i \in S} \mu_i \mathbb{E}[f(X_n) \mid X_0 = i] = \sum_{i \in S} \mu_i (\mathbf{P} f)_i = \mu \mathbf{P}^n f.$$

关于任何在 S 上的（非负或有界）函数 f 上都可以定义 ℓ^∞ 范数，也就是（这里直接把 f 看作一个列向量，$f_i = f(i)$）

$$\|f\|_\infty = \sup_{i \in S} |f_i|.$$

作为练习，我们可以自行证明对于任意（非负或有界）f 和任意转移概率矩阵 $\mathbf{P}$

$$\|\mathbf{P} f\|_\infty \le \|f\|_\infty.$$

2.1.4 对于概率函数的马氏性

令 $F : S^{n+1} \to \mathbb{R}$ 为一个（非负或者有界）的函数，则对任意 m，我们可以定义随机变量 $F(x_m, \ldots, x_{m+n})$。且有

$$\begin{aligned}
&\mathbb{E}[F(x_m, \ldots, x_{m+n}] \mid X_0 = i_0, \ldots, X_m = i_m] \\
&= \sum_{j_1 \in S} \cdots \sum_{j_n \in S} F(i_m, j_1, \ldots, j_n) \\
&\qquad \times \mathbb{P}(X_{m+1} = j_1, \ldots, X_{m+n} = j_m \mid X_0 = i_0, \ldots, X_m = i_m) \\
&= \sum_{j_1 \in S} \cdots \sum_{j_n \in S} F(i_m, j_1, \ldots, j_n) \underbrace{(\mathbf{P})_{i_m, j_1} (\mathbf{P})_{j_1, j_2} \cdots (\mathbf{P})_{j_{n-1}, j_n}}_{= \mathbb{P}(X_1 = j_1, \ldots, X_n = j_m \mid X_0 = i_m)} \\
&= \mathbb{E}[F(X_0, \ldots, X_n] \mid X_0 = i_m].
\end{aligned}$$

2.2 德布林（Doeblin）理论

当一个马尔科夫过程走过很多步后，X_n 会不会收敛到一个与 X_0 的初始分布无关的平稳分布？也就是说，是否存在一个 S 上的分布 π，使得 $X_n \xrightarrow{d} \pi$，无论 X_0 的分布 μ 是什么？

本节我们考虑一般的马尔科夫过程，但是假定的条件往往在有限状态情况下才自然出现。

2.2.1 德布林基本定理

定理 2.2.1. 假设 $\mathbf{P}$ 为 S 上的概率转移矩阵，且存在一个状态 $j_0 \in S$，使得对所有状态 i，$(\mathbf{P})_{i,j_0} > 0$，也就是说，存在 $\epsilon > 0$ 使得 $(\mathbf{P})_{i,j_0} \geq \epsilon$。则 $\mathbf{P}$ 存在唯一的平稳分布π，使得 $\pi_{j_0} \geq \epsilon$，且对于任意初始分布 μ，

$$\|\mu\mathbf{P}^n - \pi\|_1 \leq 2(1-\epsilon)^n, \quad n \geq 0. \tag{2.3}$$

证明. 不失一般性，我们假定 $j_0 = 1$。

这个证明需要两个引理。一个是，对于所有 $|S| = N$ 维向量 ρ，

$$\sum_{j\in S}(\rho\mathbf{P})_j = \sum_{i\in S}\rho_i.$$

这个等是容易验证。另外一个是，对于各项和为 0 的 N 维向量 ρ（也就是 $\rho_1 + \cdots + \rho_N = 0$），如果 $\mathbf{P}$ 满足定理2.2.1中的假设，则对每个 $n \geq 1$，

$$\|\rho\mathbf{P}^n\|_1 \leq (1-\epsilon)^n\|\rho\|_1.$$

要证明这个结果，我们只要证明 $n = 1$ 的情况即可，然后由数学归纳法可以推广到所有 n 的情况。如果我们定义 Q_{ij}（$i, j = 1, \ldots, N$）为如果 $j \neq 1$，则 $Q_{ij} = (\mathbf{P})_{ij}$，否则 $Q_{i1} = (\mathbf{P})_{i1} - \epsilon$。我们注意到 Q_{ij} 非负，且对任意 i，$Q_{i1} + \cdots + Q_{i,N-1} = 1 - \epsilon$。如果 $\sum_{i\in S}\rho_i = 0$，则容易看出对所有 $j \in S$，$(\rho\mathbf{P})_j = (\rho Q)_j$，于是 $\|\rho\mathbf{P}\|_1 = \|\rho Q\|_1$，于是只要证明 $\|\rho Q\|_1 \leq (1-\epsilon)\|\rho\|_1$。令 $\tilde{Q} = (1-\epsilon)^{-1}Q$。我们知道 $\tilde{Q}$ 是一个转移概率矩阵（各项非负，每行各项为 1）。所以，$\|\rho\tilde{Q}\|_1 \leq \|\rho\|_1$。而这个不等式只要两边各乘以 $(1-\epsilon)$，就得到了我们想要的结果。

这时我们发现，$\mathbf{P}$ 可以认为是子空间 $\{\rho \in \ell^1 : \sum_{i\in\mathbb{S}}\rho_i = 0\}$ 上的算子，而且在这个子空间上它是一个压缩映射。

对于任意初始分布 μ，我们考虑 $\rho = \mu - \mu\mathbf{P}$。我们有

$$\sum_{i\in S}\rho_i = \sum_{i\in S}\mu_i - \sum_{i\in S}(\mu\mathbf{P})_i = 1 - 1 = 0,$$

以及

$$\|\rho\|_1 \leq \|\mu\|_1 + \|\mu\mathbf{P}\|_1 = 1 + 1 = 2.$$

于是，由上面两个技术引理，$\|\rho\mathbf{P}^n\|_1 \leq (1-\epsilon)^n\|\rho\|_1 \leq 2(1-\epsilon)^n$。我们得到：

$$\mu, \quad \mu\mathbf{P} = \mu - \rho, \quad \mu\mathbf{P}^2 = \mu - (\rho + \rho\mathbf{P}), \quad \ldots,$$
$$\ldots, \quad \mu\mathbf{P}^n = \mu - (\rho + \rho\mathbf{P} + \cdots + \rho\mathbf{P}^{n-1}), \quad \ldots$$

收敛到一个极限 π。易知这个 π 是一个概率向量，且 $\pi\mathbf{P} = \pi$ 所以它是一个平稳概率向量。我们有

$$\|\mu\mathbf{P}^n - \pi\|_1 = \|\sum_{i=n}^{\infty} \rho\mathbf{P}^i\|_1 \le \sum_{i=n}^{\infty} \|\pi\mathbf{P}^i\|_1 \le \epsilon^{-1}(1-\epsilon)^n.$$

如果向得到更好的估计(2.3)，我们需要另外的方法：对任意 $m > n$，

$$\|\mu\mathbf{P}^n - \mu\mathbf{P}^m\|_1 = \| \underbrace{(\mu - \mu\mathbf{P}^{m-n})}_{\text{各项和为 } 0\text{，}\ell_1 \text{ 范数 } \le 2} \mathbf{P}^n\|_1 \le 2(1-\epsilon)^n,$$

然后令 $m \to \infty$。

如果存在另一个平稳概率向量 π'，则用上面的思路，我们可以证明 $\|(\pi - \pi')\mathbf{P}\|_1 \le 2(1-\epsilon)^n$，也就是 $\pi\mathbf{P}^n - \pi'\mathbf{P}^n$ 趋近于 0。但是，因为 π 与 π' 都是平稳概率向量，所以 $\pi\mathbf{P}^n - \pi'\mathbf{P}^n = \pi - \pi'$。我们得到 $\pi = \pi'$，也就是说，平稳概率向量是唯一的。

最后，因为 $\pi_1 = \sum_{j\in\mathbb{S}} \pi_j(\mathbf{P})_{j1} \ge \sum_{j\in\mathbb{S}} \pi_j\epsilon = \epsilon$，证明完毕。 □

2.2.2 两个推广

首先，如果 $\mathbf{P}$ 没有正好完全为正数的一列，但是它的某一个方幂，比如 $\mathbf{P}^M$，满足这个性质，则定理2.2.1仍然近似成立：假设 $\mathbf{P}^M$ 的某行元素全都 $\ge \epsilon$，则 $\mathbf{P}$ 也存在唯一的平稳概率向量 π，且对任意概率向量 μ，

$$\|\mu\mathbf{P}^n - \pi\|_1 \le 2(1-\epsilon)^{\lfloor n/M \rfloor}. \tag{2.4}$$

欲证明这个公式，我们不妨先考虑特殊情况，当 $n = mM$。这时，我们只考虑 $\{X_n\}$ 的子序列 $\{X_0, X_M, X_{2M}, \dots, X_{mM}, \dots\}$。这个子序列也构成一个马氏链，并且其转移概率矩阵为 $\mathbf{P}^M$（证明作为练习）。这样的话，根据定理2.2.1，这个子序列（子马氏链）存在唯一的平稳概率向量 π，并且不等式(2.4)成立。剩下的就是证明不等式对 $n = mM + r$（$1 \le r < M$）成立，于是这个 π 也是原马氏链的平稳概率向量（留作练习）。

有些很简单的马氏链，却没有平稳概率向量。往往这是由这个状态空间的周期性引起的。比如，在 $\mathbb{Z}/4\mathbb{Z}$ 上的对称随机游动（见图2.1）如果从 $X_0 = 1$ 出发，则不管过多长时间，在奇数时间，X_n 只能在 $2, 4$，而在偶数时间，X_n 只能在 $1, 3$。不可能有一个 π 同时（近似）满足这两个特征。

但是，直觉上我们又都认可这个简单的随机游动有一个极限：经过很长时间后，X_n 应该均匀分布在这 4 个状态上。数学上怎么刻画呢？

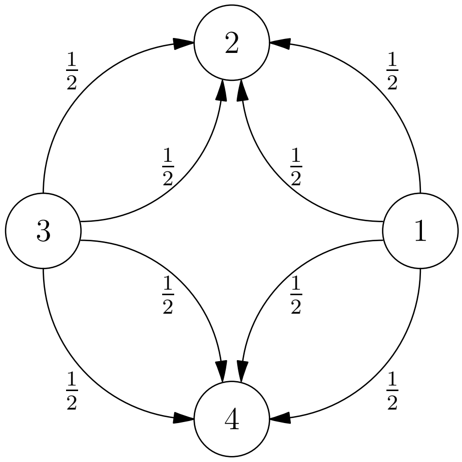

图 2.1: 在 $S = \{1, 2, 3, 4\}$ 上的马氏链，转移概率矩阵为 $\begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}$，可以看做一个环上的对称随机游动。

在这样的马氏链上，我们不再有固定时间的极限分布，但是，仍然有随时间平均的极限分布。

定义

$$\mathbf{A}_n = \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{P}^m. \tag{2.5}$$

则 $\mathbf{A}_n$ 也满足转移概率矩阵的特征，并且它的项有如下意义：

$$(\mathbf{A})_{ij} = \frac{1}{n} \sum_{m=0}^{n-1} \mathbb{P}(X_m = j \mid X_0 = i) = \mathbb{E}\left[\frac{1}{n} \sum_{m=0}^{n-1} 1_{\{j\}}(X_m) \middle| X_0 = i \right].$$

如果 X_0 的初始分布是 μ，则 $(\mu \mathbf{A}_m)_j$ 是 $X_0, X_1, \ldots, X_n, \ldots$ 在从时间 0 到时间 $n-1$ 这段时间，停留在状态 j 的平均总时间再除以 n。

定理 2.2.2. *假设 $\mathbf{P}$ 是 $\mathbb{S}$ 上的转移概率矩阵，并且存在一个正整数 $M > 1$ 和一个状态 j_0，使得对于所有 $i \in \mathbb{S}$，$(\mathbf{A}_M)_{i,j_0} > 0$，也就是说，存在 $\epsilon > 0$ 使得所有 $i \in \mathbb{S}$，$(\mathbf{A}_M)_{i,j_0} \geq \epsilon$。这样的话，存在一个唯一的 $\mathbf{P}$ 的平稳概率向量 π，并且满足 $(\pi)_{j_0} \geq \epsilon$ 以及对于任意初始分布 μ*

$$\|\mu \mathbf{A}_n - \pi\|_1 \leq \frac{M-1}{n\epsilon}.$$

证明. 首先，我们证明存在唯一一个概率向量 π 使得 $\pi \mathbf{P} = \pi$。如果把 $\mathbf{A}_M$ 看作一个转移概率矩阵，则存在唯一的概率向量 π，使得 $\pi \mathbf{A}_M = \pi$。如

果 $\pi\mathbf{P}=\pi$，则这是我们想要的唯一的 π。否则，令 $\pi\mathbf{P}=\pi'$，则 $\pi'\mathbf{A}_M=\pi\mathbf{P}\mathbf{A}_M=\pi\mathbf{A}_M\mathbf{P}=\pi\mathbf{P}=\pi'$，所以 π' 也是 $\mathbf{A}_M$ 的平稳概率向量，与 π 的唯一性矛盾。

下面证明从任意初始概率分布 μ，$\mu\mathbf{A}_n$ 都收敛到 π。这个证明需要一个窍门：对所有概率向量 μ，

$$\|\mu\mathbf{A}_n\mathbf{A}_m-\mu\mathbf{A}_n\|_1\le\frac{m-1}{n},\quad m,n\ge 1. \tag{2.6}$$

证明如下：

$$\|\mu\mathbf{A}_n\mathbf{A}_m-\mu\mathbf{A}_n\|_1\le\frac{1}{m}\left\|\sum_{k=0}^{m-1}(\mu\mathbf{A}_n\mathbf{P}^k-\mu\mathbf{A}_n)\right\|_1\le\frac{1}{m}\sum_{k=0}^{m-1}\|\mu\mathbf{A}_n\mathbf{P}^k-\mu\mathbf{A}_n\|_1,$$

而对于 $k=1,\ldots,m-1$，

$$\|\mu\mathbf{A}_n\mathbf{P}^k-\mu\mathbf{A}_n\|_1=\frac{1}{n}\|\underbrace{\mu\mathbf{P}^{n+k}+\cdots+\mu\mathbf{P}^n}_{k\text{ 个}}-\underbrace{\mu-\mu\mathbf{P}-\cdots-\mu\mathbf{P}^{k-1}}_{k\text{ 个}}\|_1\le\frac{2k}{n},$$

所以(2.6)得证。这样，取 $m=M$，我们有

$$\begin{aligned}\|\mu\mathbf{A}_n-\pi\|_1&\le\|\mu\mathbf{A}_n-\mu\mathbf{A}_n\mathbf{A}_M\|_1+\|\mu\mathbf{A}_n\mathbf{A}_M-\pi\|_1\\&\le\frac{M-1}{n}+\|(\mu\mathbf{A}_n-\pi)\mathbf{A}_M\|_1\\&\le\frac{M-1}{n}+(1-\epsilon)\|\mu\mathbf{A}_n-\pi\|_1.\end{aligned}$$

于是，我们证明本定理。 □

2.3 遍历论初步

2.3.1 平均遍历定理

对于马氏链 $\{X_n:n\ge 0\}$，令

$$\bar{T}_j^{(n)}=\frac{1}{n}\sum_{m=0}^{n-1}1_{\{j\}}(X_m)$$

为在时刻 n 之前，马氏链在状态 j 停留的时间。对于任意初始分布 μ，我们有 $\mathbb{E}[\bar{T}_j^{(n)}]=(\mu\mathbf{A}_n)_j$。于是，定理2.2.2可以理解为当 $n\to\infty$ 时，$\mathbb{E}[\bar{T}_j^{(n)}]\to\pi_j$，我们要证明更强的结论：随机变量 $\bar{T}_j^{(n)}$ 本身收敛到 π_j。这样的结果，也就是对于任意初始条件，某变量的长时间平均等于一个确定值，通常叫做遍历论结果。具体地说，我们证明当 $n\to\infty$ 时，$\bar{T}_j^{(n)}\to\pi_j$ 均方收敛。

定理 2.3.1. 在和定理2.2.2相同的假设下，

$$\sup_{j\in\mathbb{S}}\mathbb{E}\left[(\bar{T}_j^{(n)}-\pi_j)^2\right]\le\frac{2(M-1)}{n\epsilon},\quad n\ge 1. \tag{2.7}$$

并且，对于任意 $\mathbb{S}$ 上的有界函数 f，（令 $\pi f=\sum_{i\in\mathbb{S}}\pi_i f(i)$）

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{m=0}^{n-1}f(X_m)-\pi f\right)^2\right]\le\frac{2(M-1)\|f\|_\infty^2}{n\epsilon}. \tag{2.8}$$

证明. 我们先直接证明(2.8)，因为(2.7)是(2.8)当 $f(k)=\delta_{jk}$ 时的特例。

这里有界函数可以视为 $\ell^\infty=\ell^\infty(\mathbb{S})$ 中的列向量。定义 $\bar{f}\in\ell^\infty$ 为 $\bar{f}_i=f_i-\pi f$。则对于任意随机变量 X，其函数 $f(X)-\pi f$（这里 πf 是一个常数）可以写成 $\bar{f}(X)$。于是

$$\begin{aligned}\left(\frac{1}{n}\sum_{m=0}^{n-1}f(X_m)-\pi f\right)^2&=\frac{1}{n^2}\left(\sum_{m=0}^{n-1}\bar{f}(X_m)\right)^2\\&=\frac{2}{n^2}\sum_{0\le k\le l<n}\bar{f}(X_k)\bar{f}(X_l)-\frac{1}{n^2}\sum_{m=0}^{n-1}\bar{f}(X_m)^2 \qquad (2.9)\\&\le\frac{2}{n^2}\sum_{0\le k\le l<n}\bar{f}(X_k)\bar{f}(X_l).\end{aligned}$$

取均值，我们有

$$\begin{aligned}\mathbb{E}\left[\sum_{0\le k\le l<n}\bar{f}(X_k)\bar{f}(X_l)\right]&=\sum_{k=0}^{n-1}\mathbb{E}\left[\bar{f}(X_k)\sum_{m=0}^{n-k-1}\bar{f}(X_{k+m})\right]\\&=\sum_{k=0}^{n-1}\mathbb{E}\left[\bar{f}(X_k)\sum_{m=0}^{n-k-1}\underbrace{\mathbb{E}\left[\bar{f}(X_{k+m})\middle|X_k\right]}_{\text{先求对 }X_k\text{ 的条件概率}}\right] \qquad (2.10)\\\text{（这里用(2.2)）}\quad&=\sum_{k=0}^{n-1}\mathbb{E}\left[\bar{f}(X_k)\sum_{m=0}^{n-k-1}(\mathbf{P}^m\bar{f})_{X_k}\right]\\&=\sum_{k=0}^{n-1}(n-k)\mathbb{E}\left[\bar{f}(X_k)(\mathbf{A}_{n-k}\bar{f})_{X_k}\right].\end{aligned}$$

我们可以验证，对于一般的 $f\in\ell^\infty$ 与 $\pi\in\ell^1$ 且 $\|\pi\|=1$，都有 $\|\bar{f}\|_\infty\le 2\|f\|_\infty$[3]。（证明留作练习）。令 X_k 的分布为 ν，则总有

$$|\bar{f}(X_k)|\le\|\bar{f}\|_\infty\|\nu\|_1\le 2\|f\|_\infty$$

[3] Stroock 在教科书里写道 $\|\bar{f}\|_u\le\|f\|_u$（他的 $\|\cdot\|_u$ 等价于我们的 $\|\cdot\|_\infty$）。但是，这应该是错的。举例：$\mathbb{S}=\{1,2\}$，$f=(1,-1)^T\in\ell^\infty(\mathbb{S})$，$\pi=(0,1)$。这样，$\bar{f}=(2,0)^T$，于是 $\|\bar{f}\|_\infty=2\|f\|_\infty$

于是，

$$\begin{aligned}|\mathbb{E}[\bar{f}(X_k)(\mathbf{A}_{n-k}\bar{f})_{X_k}| &\le 2\|f\|_\infty \mathbb{E}[\mathbf{A}_{n-k}\bar{f})_{X_k}] = 2\|f\|_\infty \nu(\mathbf{A}_{n-k}\bar{f})\\ &= 2\|f\|_\infty \nu \underbrace{(\nu\mathbf{A}_{n-k}-\pi)}_{\|\cdot\|_1 \le \frac{M-1}{n\epsilon}} f\\ &\le 2\|f\|_\infty][\frac{M-1}{(n-k)\epsilon}\|f\|_\infty]\end{aligned}$$

把这个估计带入(2.10)再带入(2.9)，我们得到(2.8)。如果限制 f 为 $f_i=\delta_{ij}$，则得到特殊结论(2.7)。[4] □

2.3.2 返回时

和随机游动模型中一样，我们也可以定义一般马氏链中状态 j 的返回时 $\rho_j=\inf\{n>0: X_n=j\}$。并且，我们可以推广到 $\rho_j^{(m)}$ （$m\ge 0$）：令 $\rho_j^{(0)}\equiv 0$ 而对于 $m>1$，$\rho^{(m)}(j)=\inf\{n>\rho_j^{(m-1)}: X_n=j\}$。并且，如果对状态有 $\mathbb{P}(\rho_j<\infty \mid X_0=j)=1$，我们就说 j 是常返的，反之我们称它是瞬时的。下面我们讨论一些结果。其中有些在随机游动中已经证明过。

首先，我们注意到 $\{\rho_j>n\}$ 这个事件是可以由 $X_0,\ldots,X_n$ 决定的（准确地说，这个事件包括在由 $X_0,\ldots,X_n$ 生成的 σ 代数中）：

$$1_{(n,\infty]}(\rho_j)=F_{n,j}(X_0,\ldots,X_n),$$

这里 $F_{n,j}$ 是一个可测函数

$$F_{n,j}(i_0,\ldots,i_n)=\begin{cases}1, & \text{所有 } i_m\ne j \ (m=0,\ldots,n) \ ,\\ 0, & \text{不然的话.}\end{cases}$$

更进一步，对于所有 m，事件 $\rho_j^{(m)}>n$ 也被 $X_0,\ldots,X_n$ 决定。（证明做习题，对 m 做数学归纳法。）

定理 2.3.2. 对于任意 $m\in\mathbb{Z}_+$ 和 $(i,j)\in\mathbb{S}^2$，

1.

$$\mathbb{P}(\rho^{(m)}<\infty \mid X_0=i)=\mathbb{P}(\rho_j<\infty \mid X_0=i)\mathbb{P}(\rho_j<\infty \mid X_0=j)^{m-1}.$$

特别地，如果 j 是常返的，则对所有 m，$\mathbb{P}(\rho_j^{(m)}<\infty \mid X_0=j)=1$。

[4]注意到，虽然特殊结论中 $\|f\|_\infty=1$，但是因为对这种特殊形式的 f，对任意概率分布 π 有 $\|\bar{f}\|_\infty\le\|f\|_\infty$，我们可以把(2.7)右边的常数 2 缩小为 1。

2. 如果 j 是常返的，则在 $X_0=j$ 的条件下，$\{\rho_j^{(m)}-\rho_j^{(m-1)}:m\geq 1\}$ 是一列相互独立的同分布随机变量，其分布等同于 ρ_j。

证明. 我们需要 $\{X_n\}$ 对于概率函数的的马氏性（见2.1.4节），以及单调收敛定理。对第1部分，我们只要证明

$$\begin{aligned}
&\mathbb{P}(\rho_j^{(m)}<\infty \mid X_0=i)\\
&=\sum_{n=1}^{\infty}\mathbb{P}(\rho_j^{(m-1)}=n\&\rho_j^{(m)}<\infty \mid X_0=i)\\
&=\sum_{n=1}^{\infty}\lim_{N\to\infty}\mathbb{E}\left[1-F_{N,j}(X_n,\ldots,X_{n+N}),\rho_j^{(m-1)}=n\Big|X_0=i\right]\\
&=\sum_{n=1}^{\infty}\lim_{N\to\infty}\mathbb{E}\left[1-F_{N,j}(X_n,\ldots,X_{n+N})\Big|\rho_j^{(m-1)}=n,X_0=j\right]\\
&\qquad\qquad\times\mathbb{P}(\rho_j^{(m-1)}=n\mid X_0=i)\\
&=\sum_{n=1}^{\infty}\lim_{N\to\infty}\mathbb{E}\left[1-F_{N,j}(X_0,\ldots,X_N)\mid X_0=j\right]\mathbb{P}(\rho_j^{(m-1)}=n\mid X_0=i)\\
&=\sum_{n=1}^{\infty}\lim_{N\to\infty}\mathbb{P}(\rho_j\leq N\mid X_0=j)\mathbb{P}(\rho_j^{(m-1)}=n\mid X_0=i)\\
&=\mathbb{P}(\rho_j\leq M\mid X_0=j)\mathbb{P}(\rho_j^{(m-1)}<\infty\mid X_0=i).
\end{aligned}$$

所以第1部分可由数学归纳法证得。至于第2部分，我们只需要证明等式

$$\mathbb{P}(\rho_j^{(m+1)}>n+n_m\mid X_0=j,\rho_j^{(1)}=n_1,\ldots,\rho_j^{(m)}=n_m)=\mathbb{P}(\rho_j>n\mid X_0=j)$$

即可。上述等式源于马氏性：

$$\begin{aligned}
\mathbb{E}[F_{n,j}(X_{n_m},\ldots,X_{n_m+n})\mid X_0=j,\rho_j^{(1)}=n_1,\ldots,\rho_j^{(m)}=n_m]&\overset{n_m\text{ 作为时间起点}}{=}\\
\mathbb{E}[F_{n,j}(X_0,\ldots,X_n)\mid X_0=j]=\mathbb{P}(\rho_j>n\mid X_0=j).&
\end{aligned}$$

□

定义 $T_j=\sum_{m=0}^{\infty}1_{\{j\}}(X_m)$ 为马氏链 $\{X_n\}$ 在状态 j 停留的总时间。显然有

$$\mathbb{P}(T_j>m\mid X_0=i)=\begin{cases}\mathbb{P}(\rho_j^{(m)}<\infty\mid X_0=j), & i=j,\\ \mathbb{P}(\rho_j^{(m+1)}<\infty\mid X_0=j), & i\neq j.\end{cases}$$

所以，根据定理2.3.2第1部分，我们有

$$\mathbb{E}[T_j \mid X_0 = i] = \delta_{i,j} + \frac{\mathbb{P}(\rho_j < \infty \mid X_0 = i)}{\mathbb{P}(\rho_j = \infty \mid X_0 = j)},$$

并且

$$\begin{aligned} \mathbb{E}[T_j \mid X_0 = j] = \infty &\iff \mathbb{P}(T_j = \infty \mid X_0 = j) = 1, \\ \mathbb{E}[T_j \mid X_0 = j] < \infty &\iff \mathbb{P}(T_j < \infty \mid X_0 = j) = 1. \end{aligned} \tag{2.11}$$

（推导过程类似1.2.2小节里对于随机游动这个特殊情况的讨论。）于是，j 是常返态，当且仅当 $\mathbb{E}[T_j \mid X_0 = j] = \infty$。如果我们假定如下条件：对于某 $M > 1$ 与 $j_0 \in \mathbb{S}$，任取 $i \in \mathbb{S}$，都有 $(\mathbf{A}_M)_{i,j_0} \geq \epsilon$，其中 $\epsilon > 0$ 为一个常数，而 $\mathbf{A}_n = n^{-1}(\mathbf{P}^0 + \mathbf{P}^1 + \cdots + \mathbf{P}^{n-1})$，其中 $\mathbf{P}$ 为此马氏链的概率转移矩阵。（这也是定理2.2.2的假设条件。）则根据定理2.2.2的结论，我们有 $(\mathbf{A}_n)_{j_0,j_0} \to \pi_{j_0} > 0$，于是

$$\mathbb{E}[T_{j_0} \mid X_0 = j_0] = \sum_{m=0}^{\infty} (\mathbf{P}^m)_{j_0,j_0} = \lim_{n\to\infty} n(\mathbf{A}_n)_{j_0,j_0} = \infty.$$

也就是说，定理2.2.2保证了 j_0 是常返态。我们还可以证明更多结果如下。我们首先引入记号从 i 可到达 j，

$$i \to j, \quad i, j \in \mathbb{S}$$

代表存在 $m \geq 1$ 使得 $(\mathbf{P}^m)_{i,j} > 0$，也就是说，马氏链从 i 出发，可以经过有限步走到 j。

定理 2.3.3. *假定和定理2.2.2一样的条件，也就是：对于某 M 和某 $\epsilon > 0$，存在 $j_0 \in \mathbb{S}$，对所有 $i \in \mathbb{S}$，都有 $(\mathbf{A}_M)_{i,j_0} \geq \epsilon$。则 j 是常返态当且仅当 $j_0 \to j$。并且，如果 $j_0 \to j$，则对于任意 $p \in (0, \infty)$，$\mathbb{E}[\rho_j^p \mid X_0 = j] < \infty$。*

证明. $j_0 \not\to j$ 等价于 $\mathbb{P}(\rho_j = \infty \mid X_0 = j_0) = 1$，并且因为 $(\mathbf{A}_M)_{j,j_0} \geq \epsilon$，就有至少一个 $m \in \{1, \ldots, M-1\}$ 使得 $(\mathbf{P}^m)_{j,j_0} > 0$。于是我们有

$$\begin{aligned} &\mathbb{P}(\rho_j^{(m)} = \infty \mid X_0 = j) \\ \geq\ &\mathbb{P}(X_{m+1}, X_{m+2}, \ldots \text{ 都不等于 } j \,\&\, X_m = j_0 \mid X_0 = j) \\ =\ &\lim_{N\to\infty} \mathbb{E}[F_{N,j}(X_m, \ldots, X_{m+N}), X_m = j_0 \mid X_0 = j] \\ =\ &\lim_{N\to\infty} \mathbb{E}[F_{N,j}(X_m, \ldots, X_{m+N}) \mid X_m = j_0, X_0 = j]\mathbb{P}(X_m = j_0 \mid X_0 = j) \\ =\ &\lim_{N\to\infty} \mathbb{E}[F_{N,j}(X_0, \ldots, X_N) \mid X_0 = j_0]\mathbb{P}(X_m = j_0 \mid X_0 = j) \\ =\ &\mathbb{P}(\rho_j = \infty \mid X_0 = j_0)(\mathbf{P}^m)_{j,j_0} > 0. \end{aligned}$$

根据定义，j 是个瞬时状态。

另一方面，如果 $j_0 \to j$，就存在 $m = m(j) > 0$ 使得 $(\mathbf{P}^m)_{j_0,j} > 0$。于是对任意 $i \in \mathbb{S}$

$$\begin{aligned}(\mathbf{A}_{m+M})_{i,j} &= \frac{1}{m+M}\sum_{l=0}^{m+M-1}(\mathbf{P}^l)_{ij} \geq \frac{1}{m+M}\sum_{l=0}^{M-1}(\mathbf{P}^l)_{i,j_0}(\mathbf{P}^m)_{j_0,j}\\ &= \frac{M}{m+M}\underbrace{(\mathbf{A}_M)_{i,j_0}}_{\geq\epsilon}(\mathbf{P}^m)_{j_0,j} \geq \frac{M\epsilon}{m+M}(\mathbf{P}^m)_{j_0,j} \geq 0.\end{aligned}$$

所以，也存在 $M' = M'(j) = M+m$ 和 $\epsilon' = \epsilon'(j) = (M\epsilon/M'(j))(\mathbf{P}^{m(j)})_{j_0,j} > 0$，使得对于所有 $i \in \mathbb{S}$，$(\mathbf{A}_{M'})_{i,j} \geq \epsilon'$。由之前已有结果，我们知道 j 是常返的。

最后，我们证明 $\mathbb{E}[\rho_j = n \mid X_0 = j]$ 当 $n \to \infty$ 时指数衰减，于是就能证明对所有 $p > 0$，$\mathbb{E}[\rho_j^p \mid X_0 = j] < \infty$。

对于任意 $n \in \mathbb{Z}_+$ 和 $i \in \mathbb{S}$，令 $M' = M'(i)$ 与 $\epsilon' = \epsilon(i)$ 如上面定义，且 $u(n,i) = \mathbb{P}(\rho_j > nM' \mid X_0 = i)$。我们有

$$\begin{aligned}&u(n+1,i)\\ =&\sum_{k\in\mathbb{S}}\mathbb{P}(\rho_j > (n+1)M' \& X_{nM'} = k \mid X_0 = i)\\ =&\sum_{k\in\mathbb{S}}\mathbb{E}\left[F_{M',j}(X_{nM'},\ldots,X_{(n+1)M'}), \rho_j > nM' \& X_{nM'} = k \middle| X_0 = i\right]\\ =&\sum_{k\in\mathbb{S}}\mathbb{E}\left[F_{M',j}(X_{nM'},\ldots,X_{(n+1)M'}) \middle| \rho_j > nM' \& X_{nM'} = k, X_0 = i\right]\\ &\qquad\times\mathbb{P}(\rho_j > nM' \& X_{nM'} = k \mid X_0 = i)\\ =&\sum_{k\in\mathbb{S}}\mathbb{E}\left[F_{M',j}(X_0,\ldots,X_{M'}) \middle| X_0 = k\right]\mathbb{P}(\rho_j > nM' \& X_{nM'} = k \mid X_0 = i)\\ =&\sum_{k\in\mathbb{S}}\mathbb{P}(\rho_j > M' \mid X_0 = k)\mathbb{P}(\rho_j > nM' \& X_{nM'} = k \mid X_0 = i)\\ =&\sum_{k\in\mathbb{S}}u(1,k)\underbrace{\mathbb{P}(\rho_j > nM' \& X_{nM'} = k \mid X_0 = i)}_{\text{把所有的 } X_{nM'} = k \text{ 按照 } k \text{ 都加起来为 } U_{n,i}}\\ \leq& U\cdot u(n,i), \quad U = \max_{k\in\mathbb{S}}u(1,k).\end{aligned}$$

下面我们证明 $U \leq 1 - \epsilon''$，这里 $\epsilon'' = \epsilon'/M'$。这是因为，对任意 k，

$$\mathbb{P}(\rho_j \leq M' \mid X_0 = k) \geq (\mathbf{P}^{M'})_{k,j} \geq \frac{1}{M'}(\mathbf{A}_{M'})_{k,j} \geq \frac{\epsilon'}{M'}.$$

这样, 我们得到 $u(u,j)\le 1-\epsilon''$, 且由上面的递推不等式, $u(n,j)\le(1-\epsilon'')^n$。

接下来我们结束本定理最后部分的证明。这部分证明只是分析推导，没什么概率意味了：

$$\begin{aligned}
\mathbb{E}[\rho_j^p \mid X_0=j] &= \sum_{n=1}^{\infty} n^p \mathbb{P}(\rho_j=n\mid X_0=j)\\
&\le \sum_{m=1}^{\infty}(mM')^p \underbrace{\sum_{n=(m-1)M'+1}^{mM'} \mathbb{P}(\rho_j=n\mid X_0=j)}_{\le \mathbb{P}(\rho_i>(m-1)M'|X_0=j)}\\
&\le (M')^p\sum_{m=1}^{\infty} m^p \mathbb{P}(\rho_j>(m-1)M'\mid X_0=j)\\
&\le (M')^p\sum_{m=1}^{\infty} m^p(1-\epsilon'')^{m-1}<\infty.
\end{aligned}$$

□

2.3.3 平稳概率分布 π 的表示

假定和定理2.2.2一样的条件，也就是：对于某 M 和某 $\epsilon>0$，存在 $j_0\in\mathbb{S}$，对所有 $i\in\mathbb{S}$，都有 $(\mathbf{A}_n)_{i,j_0}\ge\epsilon$。我们已知此马氏链存在平稳概率分布 π 存在。在本节，我们在同样的假设下，证明更进一步的结果：

$$\pi_j=\frac{1}{\mathbb{E}[\rho_j\mid X_0=j]}. \tag{2.12}$$

在证明此结论前，我们首先明确一下：当 j 为常返态时，由定理2.3.3，我们知道 $\mathbb{E}[\rho_j\mid X_0=j]<\infty$，所以 $\pi_j>0$。相反地，如果 j 是瞬时态，则显然 $\mathbb{E}[\rho_j\mid X_0=j]=\infty$，也就意味着 $\pi_j=0$。

对于常返态 j，证明的思路大体是：假设 $X_0=j$。既然 $\rho_j=\rho_j^{(1)}$，$\rho_j^{(2)}-\rho_j^{(1)}$，$\rho_j^{(3)}-\rho_j^{(2)}$ 等等都是独立同分布的随机变量（定理2.3.2第2部分），则由大数定律，当 $n\to\infty$，$\rho_j^{(n)}\sim n\mathbb{E}[\rho_j\mid X_0=j]$，而在从时间 0 到时间 $n\mathbb{E}[\rho_j\mid X_0=j]$ 这段时间里，马氏链到达 j 状态一共（约）n 次。另一方面，这段时间，马氏链到达 j 状态的次数（的均值）是

$$\begin{aligned}
\sum_{i=0}^{n\mathbb{E}[\rho_j|X_0=j]} \mathbb{E}[X_i=j\mid X_0=j] = \bar{T}_j^{(n\mathbb{E}[\rho_j|X_0=j])} &= \sum_{i=0}^{n\mathbb{E}[\rho_j|X_0=j]} (\mathbf{P}^i)_{jj}\\
&\approx n\mathbb{E}[\rho_j\mid X_0=j](A_{n\mathbb{E}[\rho_j|X_0=j]})_{jj}.
\end{aligned}$$

于是，我们有，当 $n\to\infty$ 时，

$$n \approx n\mathbb{E}[\rho_j \mid X_0=j](A_{n\mathbb{E}[\rho_j|X_0=j]})_{jj} \approx n\mathbb{E}[\rho_j \mid X_0=j]\pi_j,$$

就大概地得到了所证结论。下面我们给出严格证明。为了叙述方便，令 $r_j = \mathbb{E}[\rho_j \mid X_0=j]$。

我们要证明

$$\mathbb{P}\left(\lim_{n\to\infty}\bar{T}_j^{(n)} = \frac{1}{r_j}\middle| X_0=j\right)=1, \tag{2.13}$$

也就是说，$\bar{T}_j^{(n)}$ 在 $X_0=j$ 这个条件下，几乎处处收敛到 r_j^{-1}。这个结果的一个推论就是[5]

$$\pi_j \overset{\text{定理2.3.1}}{=} \lim_{n\to\infty}(\mathbf{A}_n)_{jj} = \lim_{n\to\infty}\mathbb{E}\left[\bar{T}_j^{(n)} \mid X_0=j\right] \overset{\text{公式(2.13)}}{=} \frac{1}{r_j}.$$

（最后一个等式是由勒贝格控制收敛定理保证的：$|\bar{T}_j^{(n)}|\le 1$。）

证明. 公式2.13的证明

- 首先我们假定 $j_0 \not\to j$。这时 j 是个瞬时态，于是由(2.11)，$\mathbb{P}(T_j<\infty \mid X_0=j)=1$。根据定义，$\bar{T}_j^{(n)}\le T_j/n$，所以我们有 $\mathbb{P}(\lim_{n\to\infty}\bar{T}_j^{(n)}=0 \mid X_0=j)=1$。另一方面，如果 j 是个瞬时态，则 $r_j=\mathbb{E}[\rho_j \mid X_0=j]=\infty$。所以在这个情况下，证明完成。

- 不然，$j_0\to j$。这时 j 是个常返态。利用 $\mathbb{E}[\rho_j^4 \mid X_0=j]<\infty$ 这个性质，以及在 $X_0=j$ 这个条件下，$\{\rho_j^{(m)}-\rho_j^{(m-1)} : m\ge 1\}$ 是相互独立的与 ρ_j 同分布的随机变量这个性质，我们可以用强大数率得到

$$\begin{aligned}&\mathbb{P}\left(\lim_{m\to\infty}\frac{\rho_j^{(m)}}{m}=r_j\middle| X_0=j\right)=1\\ \iff\ &\mathbb{P}\left(\lim_{m\to\infty}\frac{m}{\rho_j^{(m)}}=\frac{1}{r_j}\middle| X_0=j\right)=1,\end{aligned}$$

  也就是说，我们已经对 $n=\rho_j^{(m)}$ 这种特殊情况证明了(2.13)。下面我们推广到一般 n。对于 $n\in[\rho_j^{(m)},\rho_j^{(m+1)})$，定义 $f(n)=m/\rho_j^{(m)}$ 和 $g(n)=m/\rho_j^{(m+1)}$。由上面的收敛性，我们有

$$\mathbb{P}(\lim_{n\to\infty}f(n)=r_j^{-1} \mid X_0=j)=1,\quad \mathbb{P}(\lim_{n\to\infty}g(n)=r_j^{-1} \mid X_0=j)=1.$$

[5] (2.13)和定理2.3.1结合起来，我们发现，在 $X_0=j$ 条件下，$\bar{T}_j^{(n)}$ 既在均方意义下，也在几乎处处意义下，收敛到 $\pi_j=r_j^{-1}$。想一想：这个结论在一般初始条件下成立么？

因为对所有 n，$g(n) \le \bar{T}_j^{(n)} \le f(n)$，所以我们在 $j_0 \to j$ 情况下证明(2.13)。

$\square$