A. T. BHARUCHA-REID 著

马尔柯夫过程论初步及其应用

上海科学技术出版社

马尔柯夫过程论初步及其应用

A.T. 巴鲁查-赖特　著

杨纪珂　吴立德　译

上海科学技术出版社

内 容 提 要

本书用非测度论的观点阐述马尔柯夫过程的基础理论，及其在不同领域中的应用．全书分三部分．第一部分是理论，共三章，阐述了空间与时间都离散的过程、空间离散时间连续的过程、以及空间与时间都连续的过程的基础理论．第二部分是应用，共六章，介绍了马尔柯夫过程在生物学、物理学、天文学与天体物理学、化学、和运筹学中的应用，着重于随机模型的建立．第三部分包括三个附录，介绍正文中用到的有关母函数和积分变换的知识，以及蒙特卡洛方法．阅读本书要求有初等概率论、数理统计、数学分析、矩阵与微分方程方面的基础知识．可供大专学校数学专业及应用数学专业的学生、研究生以及有关的工程技术人员作参考．

ELEMENTS OF THE THEORY OF
MARKOV PROCESSES
AND THEIR APPLICATIONS
A. T. Bharucha-Reid
McGraw-Hill Book Co. 1960

马尔柯夫过程论初步及其应用

杨纪珂　吴立德　译

上海科学技术出版社出版
（上海瑞金二路 450 号）
新华书店上海发行所发行　上海商务印刷厂印刷

开本 850×1156 1/32　印张 15　字数 400,000
1979 年 2 月第 1 版　1979 年 2 月第 1 次印刷
印数 1—25,000

书号：13119·751　　定价：1.65 元

目　　录

附 录

导　论

在自然科学的任何学科的发展过程中，总有一个时期致力于发展一些数学理论，借以阐明这门学科所涉及的现象所产生的各种观测结果．在这段时期中，定性的与经验的理论被定量的与数学的理论所代替或补充；它们是以某类方程或方程组的形式，表达一种假设的可以用来产生一组理论结果的机制或模型．然后可以把理论的与实验的结果作比较，从而看出所设的模型是否合理；也就是，看出它是否能够说明在模型规定的条件下所得出的实验观测结果．

在立出数学模型时，可选取以下两种方法之一来研究有关的现象．这两种方法，称为确定性的与随机性的(或概率性的)，反映出用数学形式表示的假设机制(或模型)的因果性质．我们所关注的现象究竟是确定性的还是随机性的，这并无深究的必要．我们所关注的只是一个确定性或随机性模型的表述，对它们性质的研究，以及它们说明实验结果的能力．应该指出，在这两种方法中，所涉及的数学方程也许在形式上颇为类似；可是，这两种方法的重大差异在于它们所提出的与企图回答的问题的本质以及对所得结果的解释．为了说明这两种方法，我们在下面为细菌增长现象立出一个简单的确定性模型及其随机模拟．

设 $x(t)$ 是实值的连续函数，它表示时刻 t 时一群体中的细菌数．为了描述群体的增长，我们必须在某种假设机制的基础上立出模型，这种机制刻划了细菌数的变化方式．我们假定：(1) 在时刻 t，群体中有 x 个细菌；(2) 群体的含量只能增加，并且在区间 $(t,\ t+\Delta t)$ 内的增加与时刻 t 时的细菌数成正比．因此得出关系

$$\Delta x(t)=\lambda x(t)\Delta t,\quad \lambda>0,$$

由此导出微分方程

$$\frac{dx(t)}{dt}=\lambda x(t). \tag{0.1}$$

如果假定 $x(0)=x_0>0$, 则方程(0.1)的解是

$$x(t)=x_0\exp(\lambda t). \tag{0.2}$$

在这个简单的模型中，我们并没有对群体中细菌的减少作出任何假定．因此不论从生物学的观点或者从(0.2)看，都很显然，当 $t\to\infty$ 时，群体含量将从 x_0 增至 ∞. 确定性的解(0.2)的显著特点是：它告诉我们，只要初始值 x_0 相同，则在一个给定时刻 $t>0$, 群体含量总是相同的．

其次考虑上述模型的随机模拟．设整数值的随机变量 $X(t)$ 表时刻 t 时一群体中的细菌数，并假定 $X(0)=x_0>0$. 在随机性方法中，我们不是推导 $X(t)$ 的函数方程；而是要找出时刻 t 时群体含量等于 x 的概率的表达式．因此，我们探索的是 $P_x(t)=\mathscr{P}\{X(t)=x\}$.

要立出随机模型，需假定：(1) 如在时刻 t 时群体中有 $x>0$ 个细菌，则在区间 $(t, t+\Delta t)$ 内群体中将增加一个细菌的概率等于 $\lambda x\Delta t+o(\Delta t)$, $\lambda>0$; (2) 在 $(t, t+\Delta t)$ 内，群体中增加两个或更多个细菌的概率为 $o(t)$. 这些假定导出关系

$$P_x(t+\Delta t)=(1-\lambda x\Delta t)P_x(t)+\lambda(x-1)\Delta tP_{x-1}(t)+o(\Delta t).$$

当 $\Delta t\to 0$ 时就得到差分微分方程组

$$\frac{dP_x(t)}{dt}=-\lambda xP_x(t)+\lambda(x-1)P_{x-1}(t),\quad x=x_0,\ x_0+1,\ \cdots. \tag{0.3}$$

因为已经假定 $X(0)=x_0$, 所以解方程(0.3)所用的初始条件是

$$\begin{aligned} P_x(0)&=1, \quad x=x_0,\\ &=0, \quad 其他. \end{aligned} \tag{0.4}$$

对 $x\geqslant x_0$, 方程(0.3)的解为

$$P_x(t)=\mathscr{P}\{X(t)=x\}=\binom{x-1}{x-x_0}\exp(-\lambda x_0t)(1-\exp(-\lambda t))^{x-x_0}, \tag{0.5}$$

为了比较这两个模型，首先需看到，在确定性方法中，群体含量是用一个实值的、并且时间连续的函数来表示的，但在随机性方法中，我们一开始就假定表示群体含量的随机变量是整数值的．观察确定性解(0.2)就能看出，如 λ 与 x_0 被固定，则对应于每一 t 的值，有一个实数 $x(t)$．但从(0.5)看出，如 λ 与 x_0 被固定，则对应于每对 (x, t)，$x\geqslant x_0$，$t\geqslant 0$，存在一个数 $P_x(t)$，$0\leqslant P_x(t)\leqslant 1$，它是在时刻 t 时随机变量取值 x 的概率．值得注意的是，确定性模型在它产生具有概率 1 的结果的意义下是随机性模型的一个特例．

再考虑一下平均的或期望的群体含量来结束这里关于确定性与随机性模型的简短讨论．令 $m(t)=\mathscr{E}\{X(t)\}$．由定义，

$$m(t)=\sum_{x=0}^{\infty} xP_x(t)=x_0\exp(\lambda t). \tag{0.6}$$

因此我们看到，平均群体含量的表达式(0.6)与从确定性模型得出的群体含量的表达式(0.2)相同．鉴于这种对应性，我们不妨说式(0.1)所描述的是平均群体含量，但式(0.3)还把随机的起伏考虑了进去．应当指出，在这里所考虑的两模型之间的对应性一般并不都成立；在本书中将遇到若干情形，其中确定性解不与随机性平均数相同．

上面所考虑的随机性模型就是被称为随机过程的一个例子．J. L. Doob 曾把一个随机过程定义为一个经验过程的数学抽象，这个过程的发展受一些概率规律的控制．必须提出，“随机过程”这一术语是指经验过程的数学抽象、模型或表示，而不是指经验过程本身．在上面所举的例子中，所涉及的经验过程是一个细菌群体的增长，而模型则是概率 $P_x(t)$ 的差分微分方程组，它是以某些假定为依据而推导出来的，这些假定关系到群体发展的概率方式．在上例中，细菌群体可以认为是一个系统，而那些可能有的群体含量 x_0，x_0+1，$\cdots$ 可以认为是这个系统的状态．我们还看到对应于任一状态 x，都有一个由式(0.5)给定的概率 $P_x(t)$；在时刻 t，群体以这个概率处于该状态中．

在过去二十五年中，随机过程的理论有了十分迅速的发展，并且在许多领域中找到了应用. 特别是有一类称为 Markov 链或 Markov 过程的随机过程曾被相当广泛地研究过；它在许多科学与工程技术的分支中以及在其他领域里都有很大的重要性. 在本书中我们将把注意力集中于这一类随机过程.

在定义 Markov 链之前，先考虑经典概率论中讨论过的随机过程，即独立随机变量的一个序列. 令随机变量 $X_1, X_2, \cdots, X_n$ (n 有限或无限) 表 n 次独立试验的结果，并令 $E_1, E_2, \cdots$ 表每次试验的可能后果(系统的状态). 对 $i=1, 2, \cdots, n$，令

$$p_j=\mathscr{P}\{X_i=E_j\}, \quad j=1, 2, \cdots. \tag{0.7}$$

由于试验是独立的，所以观测到一列后果 $E_{j_1}, E_{j_2}, \cdots, E_{j_n}$ 的概率就是

$$\mathscr{P}\{E_{j_1}, E_{j_2}, \cdots, E_{j_n}\}=p_{j_1}p_{j_2}\cdots p_{j_n}. \tag{0.8}$$

即后果序列的概率等于这些后果所对应的各概率之积.

Markov 把经典模型推广为：假定任一次试验的后果只依赖于上一次的后果. 因此，在这里必须引入相应于每一对后果或状态(E_i, E_j)的条件概率 p_{ij}，它定义如下：

$$p_{ij}=\mathscr{P}\{X_{n+1}=E_j \mid X_n=E_i\}, \quad n=0, 1, \cdots. \tag{0.9}$$

除了概率 p_{ij} 之外，还需要知道概率 $q(j)$，它给出初次试验的后果为 E_j 的概率. 在这个比较一般的概型中，后果序列 $E_{j_1}, E_{j_2}, \cdots, E_{j_n}$ 的概率为

$$\mathscr{P}\{E_{j_1}, E_{j_2}, \cdots, E_{j_n}\}=q(j_1)\,p_{j_1j_2}p_{j_2j_3}\cdots p_{j_{n-1}j_n}. \tag{0.10}$$

构成 Markov 链的那些随机变量显然并不相互独立，而是相依的，但其相依性只牵涉到前后一个时间单位. 因此，我们可以作出结论说：Markov 链是从一类经验过程抽象而得，这类过程相应于一些系统，它们的状态按某一概率规律以一定的方式随时刻 n 而变化，其方式为：系统从时刻 n_0 的给定状态 E_i 到时刻 n_0+1 时达于状态 E_j 的概率只依赖于时刻 n_0 时的状态 E_i，而与时刻 n_0 以前系统的状态无关.

第一部分 理 论

第一章 空间与时间都离散的过程

1.1 引 言

本章将介绍具有可数状态的离散 Markov 链的基本理论. 在叙述这一理论时, 将以离散的衍支过程作为具体的参考. 离散的衍支过程的数学模型可看作群体随代增长的一个表示. 取所研究的整数值随机变量 X_n, $n=0, 1, \cdots$ 作为在时刻 n 时或在第 n 世代中群体内个体的数目.

在第 1.2 节中, 我们将把离散的衍支过程作为具有可数状态的离散 Markov 链来考虑, 还给出这类随机过程的基本定义和性质, 以及它们作为衍支过程的解释. 在第 1.3 节中将讨论基本随机变量的矩与累积量.

在衍支随机过程的理论中有一个值得注意的问题, 那就是对某一 n, 确定随机变量 X_n 取值零的条件. 这个问题, 引导出关于衍支过程的基本理论, 将在第 1.4 节中予以考察. 在第 1.5 节里将讨论一个有关的问题, 即求出在群体灭亡前所传世代数的概率分布的问题.

第 1.6 节专门研究极限定理. 考虑几个关于具有可数状态的离散 Markov 链的一般定理以及几个关于衍支过程的特殊的极限定理. 在第 1.7 节中, 我们考虑一个在非负整数上的简单随机游动过程, 并讨论它与衍支过程的关系, 最后, 在第 1.8 节里, 我们考虑 N 维过程, 即群体是由 N 个不同型的个体所组成的那种过程.

1.2 基本定义与性质

A. 离散的衍支过程与 Markov 链　考虑一个随机过程

$\{X_n, n=0, 1, 2, \cdots\}$，那就是一族随机变量，定义在它们能取的全部可能值所构成的空间 $\mathfrak{X}$ 上．空间 $\mathfrak{X}$ 称为过程的状态空间，元素 $x\in\mathfrak{X}$，即 X_n 能取的各个不同的值，称为状态．在本章中，将取状态空间 $\mathfrak{X}$ 为非负整数 $x=0, 1, 2, \cdots$ 的集．因此我们在这一章里所考虑的过程，不论对状态变量 x，或者对时间变量 n，都是离散的．

过程 $\{X_n, n=0, 1, 2, \cdots\}$ 称为简单的离散的衍支过程[1)]，如果下列条件得到满足：

(a) $X_0=x_0=1$.

(b) $p(x)=\mathscr{P}\{X_1=x\}$，且 $\sum\limits_{x=0}^{\infty}p(x)=1$.

(c) 在已知 $X_n=j$ 的条件下，X_{n+1} 的条件分布跟 j 个与 X_1 具有相同概率分布的独立随机变量之和相同．因此，如果把在第 $n+1$ 代中群体里有 k 个个体的概率记为 $p_{n+1}(k)$，则

$$p_{n+1}(k)=\sum_{j=0}^{\infty}p_n(j)\,[p(k)]^{*j}, \tag{1.1}$$

即条件分布是 $p(k)$ 的 j 次自折积[2)].

如同在第 1.1 节里所说，一个离散的衍支过程的数学模型可设想为一群体的随代发展或增长的表示，整数值的随机变量 X_n 则表示群体在第 n 代中所含的个体数．为了研究群体的发展，即研究序列 X_0, X_1, X_2, $\cdots$，需要在已知群体的历史，即已知群体的前 n 代的含量时，求出群体在第 $n+1$ 代中含量取一个给定值的概率的表达式．因此我们要寻求条件概率

$$\mathscr{P}\{X_{n+1}=x_{n+1}\,|\,X_n=x_n,\ X_{n-1}=x_{n-1},\ \cdots,\ X_0=x_0=1\}. \tag{1.2}$$

如果随机过程 $\{X_n, n=0, 1, 2, \cdots\}$ 的构造是这样的：X_{n+1} 的条件概率分布只依赖于 X_n 的值，而与所有更前的值相独立，则称该过程具有 Markov 性质并称之为 Markov 链．更确切地说，

1) 关于衍支过程的抽象表述，请参阅 Otter[49] 与 Urbanik[58]．

2) 关于折积运算的定义，参看附录 A．

$$\mathscr{P}\{X_{n+1}=x_{n+1}|X_n=x_n,\ \cdots,\ X_0=x_0=1\}$$
$$=\mathscr{P}\{X_{n+1}=x_{n+1}|X_n=x_n\}. \tag{1.3}$$

现在记

$$p_{ij}=\mathscr{P}\{X_{n+1}=j|X_n=i\},\ \ i,\ j=0,\ 1,\ 2,\ \cdots. \tag{1.4}$$

由于 p_{ij} 都是条件概率，所以它们满足条件：

$$p_{ij}\geqslant 0,\quad \text{对一切的 } i \text{ 与 } j, \tag{1.5}$$

$$\sum_{j=0}^{\infty} p_{ij}=1,\ \ i=0,\ 1,\ 2,\ \cdots. \tag{1.6}$$

对固定的 i 与 j，这些概率可以理解为：已知在时刻 n 时系统处于状态 i 的条件下，系统在时刻 $n+1$ 时处于状态 j 的条件概率．如果 p_{ij} 只依赖于状态 i 与 j，但不依赖于时刻 n，则称这种条件概率是定常的或平稳的．在本章中我们只限于考虑那些具有平稳概率 p_{ij} 的链．如果对所有的 i 与 j，p_{ij} 为已知，则为了完全确定这过程，只需概率分布 $q(x)$ 就够了，它是

$$q(x)=\mathscr{P}\{X_0=x\},\quad x=0,\ 1,\ \cdots, \tag{1.7}$$

即 $q(x)$ 是时刻 0 时系统处于状态 x 的概率．显然，

$$0\leqslant q(x)\leqslant 1 \quad \text{及} \quad \sum_{x=0}^{\infty} q(x)=1.$$

从以上的讨论可以看出，离散的衍支过程 $\{X_n,\ n=0,1,2,\cdots\}$ 是 Markov 链．因为我们曾假定状态空间 $\mathfrak{X}$ 是非负整数的集，所以可更确切地说，离散的衍支过程是具有可数状态的 Markov 链的一种特殊情况．这一类 Markov 链首先由 Kolmogorov [43][1] 在他关于这个题目的基本论文中提出来的．

B．转移概率与 Markov 矩阵　为了跟 Markov 链理论的术语相一致，我们把条件概率 p_{ij} 称为从状态 i 到状态 j 的一步转移概率，并称 $P=(p_{ij})$ 为转移概率矩阵：

$$P=\begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

[1] 方括弧中的数字就是在章末所列参考文献的编号．

显然，P 是个具非负元素的方阵(无穷阶的方阵，因为链具有可数状态)，这是因为对所有 i 与 j，$p_{ij}\geqslant 0$，并且由于 $\sum\limits_{j=0}^{\infty} p_{ij}=1$，所以各行之和都等于1．凡是满足以上条件的矩阵称为随机矩阵或 Markov 矩阵[1]．一个 Markov 链，可以由一个转移概率矩阵 P，以及一个在时刻零时状态 $x=0, 1, 2, \cdots$ 的概率分布的列向量 $Q=(q(0), q(1), \cdots)$ 完全确定．在简单的离散衍支过程的情形，由于我们假定 $X_0=1$，因此列向量为 $Q=(0, 1, 0, \cdots)$．

在所谓一步转移概率 p_{ij} 以外，还值得考虑高阶的或 n 步的转移概率，记之为 $p_{ij}^{(n)}$．它们表示从状态 i 经 $n(>1)$ 步达于状态 j 的转移概率，或者群体含量从 i 经 n 代达于 j 的变化的概率．n 步转移概率可以递推地定义为

$$p_{ij}^{(1)}=p_{ij}, \quad p_{ij}^{(n+1)}=\sum_{\nu=0}^{\infty} p_{i\nu}^{(n)} p_{\nu j}. \tag{1.8}$$

我们现在要证明，对一切的 k，

$$p_{ij}^{(k)}=\mathscr{P}\{X_{m+k}=j \mid X_m=i\}. \tag{1.9}$$

对于 $k=1$，根据(1.4)，上式成立．现在假定(1.9)对 $k=n+1$ 成立．则有

$$\begin{aligned}
&\mathscr{P}\{X_{m+n+1}=j \mid X_m=i\} \\
&\qquad =\sum_{\nu=0}^{\infty} \mathscr{P}\{X_{m+n}=\nu \mid X_m=i\} \mathscr{P}\{X_{m+n+1}=j \mid X_{m+n}=\nu\} \\
&\qquad =\sum_{\nu=0}^{\infty} p_{i\nu}^{(n)} p_{\nu j}=p_{ij}^{(n+1)}.
\end{aligned}$$

因此，按归纳法，我们已证明了(1.9)对一切 k 都成立．

一般，我们有以下的关系：

$$p_{ij}^{(m+n)}=\sum_{\nu=0}^{\infty} p_{i\nu}^{(m)} p_{\nu j}^{(n)}. \tag{1.10}$$

当 $m=1$ 时，(1.10)就是(1.8)．仍可用归纳法证明(1.10)对一切 m 都成立．(1.10)陈述了以下事实：在从状态 i 以 $m+n$ 步达于状态 j 的过渡中，前 m 步使系统从 i 达于某中间状态 ν，后 n 步再

[1] 如果各列之和也都等于1，则矩阵称为双重随机矩阵．

使系统从 ν 达于 j. 为了得到从状态 i 能够转移到 j 的所有可能的途径，有必要对所有可能的中间状态求和. 现在如把 n 步转移概率的矩阵记为 P^n，即 $P^n=(p_{ij}^{(n)})$，则 (1.10) 写成矩阵形式便是

$$P^{m+n}=P^mP^n. \tag{1.11}$$

方程(1.11)是 Chapman-Kolmogorov 函数方程的矩阵形式. 这个函数方程刻划出 Markov 链的特征，在 Markov 链的理论中具有基本的重要性；并且正是这个方程，建立了 Markov 链与算子半群理论之间的联系[1].

C. 绝对概率　如果当 $n=0$ 时系统处于状态 i 的概率是 $q(i)$，即 $q(i)=\mathscr{P}\{X_0=i\}$，则在时刻 n 时系统处于状态 j 的非条件概率或绝对概率为

$$q^{(n)}(j)=\sum_{i=0}^{\infty}q(i)p_{ij}^{(n)}. \tag{1.12}$$

因此，只要给定 $q(i)$ 以及 n 步转移概率 $p_{ij}^{(n)}$，就可以算出 $q^{(n)}(j)$. 一个特别有意义的问题是当 $n\to\infty$ 时 $q^{(n)}(j)$ 的性态. 在第 1.6 节中我们将研究 $p_{ij}^{(n)}$ 的渐近性态，它能帮助我们去研究绝对概率的渐近性态.

D. 状态的分类　给定一 Markov 链 $\{X_n, n=0, 1, \cdots\}$，它的状态可以用几种方式来分类. 在研究 n 步转移概率 $p_{ij}^{(n)}$ 的渐近性态以及在链的状态的物理解释中，分类具有很大的重要性. 在本节中，我们要考虑状态的不同类别，着重于理论中基本的定义与关系.

定义　状态 i 称为本性的，如果使得 $p_{ij}^{(n)}>0$ 的正整数 n 以及状态 j 的存在，蕴涵着使得 $p_{ji}^{(m)}>0$ 的正整数 m 的存在. 不是本性的状态称为非本性的.

定义　设 i 与 j 是一 Markov 链的两个状态；i 与 j 称为是相通的，如果存在一个 $n\geqslant1$ 以及一个 $m\geqslant1$，使得 $p_{ij}^{(n)}>0$ 并且

[1] 关于从半群理论的观点对 Markov 链的讨论可参阅 Hille 与 Phillips 的著作[33]以及其中所列的参考文献.

$p_{ji}^{(m)}>0$. 如果 i 与 j 是相通的，则记之为 $i\leftrightarrow j$[1]. 不难看出，对相通的状态来说，以下的关系成立：

(a) $i\leftrightarrow j$ 蕴涵 $j\leftrightarrow i$;

(b) $i\leftrightarrow j$ 与 $j\leftrightarrow k$ 蕴涵 $i\leftrightarrow k$;

(c) 如对某个 j, $i\leftrightarrow j$, 则 $i\leftrightarrow i$.

定义 状态集 $S\subset\mathfrak{X}$ 称为是闭的，如果不可能从 S 中的任一状态一步转移到 S 的补集 $\mathfrak{X}-S$ 中的任一状态. 因此对 $i\in S$ 以及 $j\in\mathfrak{X}-S$, $p_{ij}=0$. 如果集 S 只包含一个状态，该状态就称为吸收状态. 很显然，一个状态 i 成为吸收状态的充分必要条件是 $p_{ii}=1$. 如果状态空间 $\mathfrak{X}$ 含有两个或更多个闭集，这链就称为可分解的或可约的[2].

与一可分解的链相联的 Markov 矩阵可以写成分块矩阵的形式；例如

$$P=\begin{bmatrix}P_1 & 0\\ 0 & P_2\end{bmatrix},$$

其中 P_1 与 P_2 表两个 Markov 矩阵，它们描述了在两个闭的状态集的内部的转移. 不是可分解的链或矩阵称为不可分解的或不可约的. 链为不可分解的充分必要条件是每一状态都可从任一其他状态出发而到达之[3].

定义 如果 $i\leftrightarrow i$, 则使得 $p_{ii}^{(n)}>0$ 的整数 n 的集的最大公约数称为状态 i 的周期. 我们把 i 的周期记为 $\omega(i)$, 并称 i 是周期的，具有周期 $\omega(i)$. 如 $i\not\leftrightarrow i$, 则定义 $\omega(i)=0$. 不具周期的状态称为非周期的.

现在我们引入并定义几种概率，它们将在后文中用到. 令

---

[1] 原文记为 $i\rightarrow j$. 由于相通关系是两向的，故改为 $i\leftrightarrow j$, 以便与通常的用法一致. ——译者注

[2] 这里的定义是不妥的. 应改为："一链称为不可约的或不可分解的，如果其中任两状态是相通的. 不是不可约的链称为可约的"，下面判断也不定正确. ——译者注

[3] 这些矩阵性质可以借拓扑学中的脉理论讨论之，参看 Solow[55] 与 Veblen[60].

$$K_{ij}^{(n)}=\mathscr{P}\{X_n=j,\ X_h\neq j,\ 0\leqslant h<n\,|\,X_0=i\}^{1)}. \qquad (1.13)$$

因此 $K_{ij}^{(n)}$ 是在已知系统在时刻零时处于状态 i 的条件下，系统在时刻 n 之前不曾到过 j 而在时刻 n 时处于状态 j 的条件概率. 仿此，令

$$L_{ij}=\mathscr{P}\{X_h=j,\ 对至少一个\ h\geqslant 1\,|\,X_0=i\}. \qquad (1.14)$$

由(1.13)看出

$$L_{ij}=\sum_{n=1}^{\infty}K_{ij}^{(n)}. \qquad (1.15)$$

最后，定义

$$\begin{aligned}Q_{ij}&=\mathscr{P}\{X_h=j,\ 无穷多次\,|\,X_0=i\}\\&=\mathscr{P}\{X_h=j,\ 对无穷多个\ h\ 的值\,|\,X_0=i\}. \qquad (1.16)\end{aligned}$$

现在就利用以上几种概率对 Markov 链的状态进行分类.

定义 一状态 i 称为常返的，如果 $Q_{ii}=1$ 或者 $L_{ii}=1$；称为不常返的(滑过的)，如果 $Q_{ii}=0$ 或者 $L_{ii}<1$.

为了使这个定义是妥贴的，必须证明那些加在 Q_{ii} 与 L_{ii} 上的条件是等价的. 这可以证明如下：令

$$F_n=\mathscr{P}\{X_h=i,\ 对至少\ n\ 个\ h\ 的值\,|\,X_0=i\}.$$

则由(1.14)得 $F_1=L_{ii}$, $F_2=L_{ii}F_1$, $\cdots$, $F_n=L_{ii}F_{n-1}$. 因此 $F_n=(L_{ii})^n$. 所以

$$\begin{aligned}Q_{ii}=\lim_{n\to\infty}F_n=\lim_{n\to\infty}(L_{ii})^n&=0, \quad 如\ L_{ii}<1,\\&=1, \quad 如\ L_{ii}=1.\end{aligned}$$

为了对 $p_{ij}^{(n)}$ 的渐近性态的研究作好准备，我们要介绍并证明一条基本的定理，它建立了 $p_{ij}^{(n)}$ 与 Q_{ij} 之间的一种关系.

定理 1.1 考虑级数

$$\sum_{n=1}^{\infty}p_{ij}^{(n)}. \qquad (1.17)$$

如果(1.17)收敛，则 $Q_{ij}=0$. (如果状态 j 为不常返的，这总是真实的.) 如果(1.17)发散，则 $Q_{ij}>0$，并且 $Q_{jj}=1$. 如果 $i\leftrightarrow j$，则得 $Q_{ij}=Q_{ji}=1$.

---

1) 此处原文误为 $\mathscr{P}\{X_n=j\,|\,X_0=i,\ X_h\neq j,\ i\leqslant h<n\}$. 下面的说明也作了相应的改变. ——译者注

【证明】 考虑事件序列 $\{E_n\}$, $n=1, 2, \cdots$, 其中 E_n 表示在时刻m时系统处于状态 i 而在时刻 $m+n$ 时系统处于状态 j 的事件. 设 $\mathscr{P}\{X_m=i\}>0$, 则

$$\mathscr{P}\{E_n\}=\mathscr{P}\{X_m=i\}p_{ij}^{(n)}.$$

如果(1.17)收敛,则应用 Borel-Cantelli 引理(参看[17]),得

$$\mathscr{P}\{E_n, \text{无穷次多}\}=0,$$

所以 $Q_{ij}=0$.

从 $p_{ij}^{(n)}$ 与 $K_{ij}^{(n)}$ 的定义,对每一 i 与 j, 得

$$p_{ij}^{(n)}=\sum_{\nu=1}^{n} K_{ij}^{(\nu)} p_{jj}^{(n-\nu)}, \tag{1.18}$$

其 $p_{jj}^{(0)}=1$. 如将(1.18)两边都对 n 求和,则得

$$\begin{aligned}\sum_{n=1}^{N} p_{ij}^{(n)} &= \sum_{n=1}^{N} \sum_{\nu=1}^{n} K_{ij}^{(\nu)} p_{jj}^{(n-\nu)} \\ &= \sum_{\nu=1}^{N} K_{ij}^{(\nu)} \sum_{n=\nu}^{N} p_{jj}^{(n-\nu)} \leqslant \sum_{\nu=1}^{N} K_{ij}^{(\nu)}\left(1+\sum_{n=1}^{N} p_{jj}^{(n)}\right).\end{aligned} \tag{1.19}$$

如令 $N\to\infty$, 并使用(1.15),则有

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} \leqslant L_{ij}\left(1+\sum_{n=1}^{\infty} p_{jj}^{(n)}\right). \tag{1.20}$$

如在(1.19)中令 $j=i$, 并把两边都除以 $1+\sum_{n=1}^{N} p_{ii}^{(n)}$, 则得

$$\frac{\sum_{n=1}^{N} p_{ii}^{(n)}}{1+\sum_{n=1}^{N} p_{ii}^{(n)}} \leqslant \sum_{\nu=1}^{N} K_{ii}^{(\nu)}.$$

所以如果(1.17)发散,从(1.20)就推出 $\sum_{n=1}^{\infty} p_{jj}^{(n)}$ 发散,于是得

$$1 \leqslant \sum_{\nu=1}^{\infty} K_{ii}^{(\nu)} = L_{ii}.$$

因此 $L_{ii}=1$, 并且 $Q_{ii}=1$, 因为

$$\begin{aligned} Q_{ii} &= 0, \quad \text{如果 } L_{ii}<1, \\ &= 1, \quad \text{如果 } L_{ii}=1. \end{aligned}$$

这样我们已经证明了