

**Key concepts:**

- *Markov* 链;
- *Chapman-Kolmogorov* 方程;
- 不变分布。

**4.1 离散时间 Markov 链**

**Definition 4.1 (离散时间 Markov 链)** 设  $E$  是一个可数集合, 不妨为  $\mathbb{N}^+ = \{1, 2, \dots\}$ 。称状态空间为  $E$  的离散时间随机过程  $\{X_n, n = 0, 1, 2, \dots\}$  为一个(离散时间) *Markov* 链, 如果对于任意  $n \geq 0$ , 任意状态  $i_0, i_1, \dots, i_{n-1}, i, j \in E$ , 有 *Markov* 性:

$$\begin{aligned} P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} \\ = P\{X_{n+1} = j | X_n = i\} \end{aligned} \quad (4.1)$$

成立。称  $P(n, i; n+1, j)$  为时刻  $n$  处于状态  $i$ , 时刻  $n+1$  转移到状态  $j$  的**转移函数**。若它与  $n$  无关, 则记为  $P_{ij}$ , 此时称  $\{X_n\}$  为一个时齐 *Markov* 链, 矩阵  $P = (P_{ij})_{i, j \in E}$  称为一步转移矩阵。我们之后都考虑时齐 *Markov* 链。

**注1.** *Markov* 性(4.1)等价于

$$\begin{aligned} P\{X_{n+1} = j, X_{n-1} = i_{n-1}, \dots, X_0 = i_0 | X_n = i\} \\ = P\{X_{n+1} = j | X_n = i\} \cdot P\{X_{n-1} = i_{n-1}, \dots, X_0 = i_0 | X_n = i\} \end{aligned} \quad (4.2)$$

即验证:

$$P(C | BA) = P(C | B) \Leftrightarrow P(CA | B) = P(C | B)P(A | B)$$

其中 $A$ 表示“过去”， $B$ 表示“现在”， $C$ 表示“未来”，用条件概率的初等定义即可验证。

(4.2)说明 Markov性等价于在已知“现在”的条件下，“过去”和“未来”是独立的。

**注2.** Markov性(4.1)还等价于

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = P(X_0 = i_0) \prod_{k=1}^n P(k-1, i_{k-1}; k, i_k). \quad (4.3)$$

(4.3)有的地方称为链式法则(chain rule)。可以看出，Markov链的有限维联合分布由初始分布和转移函数决定，在时齐情形，

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = P(X_0 = i_0) \prod_{k=1}^n P_{i_{k-1}i_k}.$$

**注3.** 若矩阵 $P = (P_{ij})_{i,j \in E}$ 满足：

$$\begin{aligned} P_{ij} &\geq 0, \quad i, j \in E \\ \sum_{j \in E} P_{ij} &= 1, \quad \forall i \in E \end{aligned}$$

则称 $P$ 为 $E$ 上的转移矩阵 (transition matrix)/随机矩阵(stochastic matrix, 注意和random matrix的区分)。称 $P_{ij}$ 为从状态 $i$ 到 $j$ 的转移概率。

## 4.2 例子

**Example 4.2 (一维简单随机游走)** 考虑一个在 $\mathbb{Z}$ 上运动的粒子，它每一次等可能地向左或向右移动一步，即

$$P_{i,i+1} = P_{i,i-1} = \frac{1}{2}, \quad \forall i \in \mathbb{Z}.$$

粒子在 $\mathbb{Z}$ 上的位置为一个随机过程，称为一维简单随机游走。

我们也可以用独立地抛一枚公平的硬币的随机试验来刻画这个模型：每次抛到正面则让粒子往右走(位置加1)，抛到反面则让粒子往左走(位置减1)。

我们还可以用随机变量序列来进行刻画，假设 $\xi_1, \xi_2, \dots$ 独立同分布于

$$P(\xi_1 = 1) = P(\xi_1 = -1) = \frac{1}{2}.$$

假设 $S_0$ 为取整数值的随机变量，且它独立于 $\xi_1, \xi_2, \dots$ ，令

$$S_n := S_0 + \xi_1 + \dots + \xi_n = S_{n-1} + \xi_n.$$

那么 $S_n$ 为从 $S_0$ 出发的一维简单随机游走。

$\xi_n$ 可以服从更一般的分布，此时， $S_n$ 称为一般随机游走。

**Example 4.3 (Ehrenfest模型)** 在统计热力学的研究中，一个经典的模型是Ehrenfest模型，用来模拟气体分子在两个容器中的扩散过程。考虑甲乙两个容器，其中总共的气体分子为 $N$ 。假设单位时间内，有且只有一个分子在甲乙两个容器之间扩散，扩散的分子是随机选取的。

记 $n$ 时间后甲容器内的气体分子数为 $X_n$ ，则 $\{X_n\}$ 是一个状态空间为 $E = \{0, 1, \dots, N\}$ 的离散时间Markov链，转移概率为

$$P(X_{n+1} = j | X_n = i) = \begin{cases} 1 - \frac{i}{N}, & j = i + 1 \\ \frac{i}{N}, & j = i - 1 \\ 0, & |j - i| \neq 1 \end{cases}$$

直观上，如果一开始甲容器里没有气体分子，经过自由扩散，两个容器中的气体分子数量应该会趋于一致。这本质上是热力学第二定律描述的熵增现象，说明在封闭系统中，微观随机行为会导致宏观熵的增加。

**Example 4.4 (图上的随机游走)** 设 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ 为一个图，对于节点 $i, j \in \mathcal{V}$ ，如果 $(i, j) \in \mathcal{E}$ ，则称节点 $i, j$ 相邻，记为 $i \sim j$ 。

考虑一个位置位于图上节点的粒子，它每次随机地移动到相邻节点上。记 $X_n$ 为它在时刻 $n$ 所处的位置，则 $\{X_n\}$ 是一个状态空间为 $\mathcal{V}$ 的离散时间Markov链，转移概率为

$$P_{ij} = \begin{cases} \frac{1}{\deg(i)} & \text{if } j \sim i, \\ 0 & \text{otherwise.} \end{cases}$$

其中 $\deg(i)$ 为节点 $i$ 的度(degree)，表示节点 $i$ 相邻的节点的数目。

### 4.3 Chapman-Kolmogorov 方程

记时齐Markov链的 $n$ 步转移概率为

$$P_{ij}^{(n)} = P\{X_{n+m} = j | X_m = i\}, \quad n \geq 0, \quad i, j \geq 0.$$

**Proposition 4.5 (Chapman-Kolmogorov 方程)** 对任意 $i, j \in E, m, n \geq 0$ ,

$$P_{ij}^{(n+m)} = \sum_{k \in E} P_{ik}^{(n)} P_{kj}^{(m)}$$

**Proof:**

$$\begin{aligned} P_{ij}^{(n+m)} &= P\{X_{n+m} = j | X_0 = i\} \\ &= \sum_{k \in E} P\{X_{n+m} = j, X_n = k | X_0 = i\} \\ &= \sum_{k \in E} P\{X_{n+m} = j | X_n = k, X_0 = i\} P\{X_n = k | X_0 = i\} \\ &= \sum_{k \in E} P_{kj}^{(m)} P_{ik}^{(n)}. \end{aligned}$$

■

记 $P^{(n)}$ 为 $n$ 步转移矩阵, 则由Chapman-Kolmogorov 方程

$$P^{(n+m)} = P^{(n)} \cdot P^{(m)},$$

其中 $\cdot$ 为矩阵乘法, 那么

$$P^{(n)} = P \cdot P^{(n-1)} = P \cdot P \cdot P^{(n-2)} = \dots = P^n.$$

**Example 4.6 (两状态的Markov链 I)**

设离散时间 *Markov* 链的样本空间只有两个状态。一步转移概率矩阵为

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}$$

其中 $\alpha, \beta \in (0, 1)$ 。我们想要计算它的 $n$ 步转移矩阵，由Chapman-Kolmogorov方程，只需要计算 $P^n$ 。

做特征值分解， $P$ 的两个特征值为

$$\lambda_0 = 1, \quad \lambda_1 = 1 - \alpha - \beta$$

可得

$$P = \frac{1}{\alpha + \beta} \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 - \alpha - \beta \end{pmatrix} \begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix}$$

那么

$$\begin{aligned} P^n &= \frac{1}{\alpha + \beta} \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (1 - \alpha - \beta)^n \end{pmatrix} \begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix} \\ &= \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1 - \alpha - \beta)^n}{\alpha + \beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix} \end{aligned}$$

注意到，随着 $n \rightarrow \infty$ ， $n$ 步转移概率存在极限

$$\begin{aligned} \lim_{n \rightarrow \infty} P_{00}^{(n)} &= \lim_{n \rightarrow \infty} P_{10}^{(n)} = \frac{\beta}{\alpha + \beta} \\ \lim_{n \rightarrow \infty} P_{11}^{(n)} &= \lim_{n \rightarrow \infty} P_{01}^{(n)} = \frac{\alpha}{\alpha + \beta} \end{aligned}$$

这是我们之后要学习的遍历性。

思考：如果 $|1 - \alpha - \beta| = 1$ ，即 $\alpha = \beta = 1$ 时会发生什么？

## 4.4 不变分布

之前我们关心Markov链的转移概率，但 $X_n$ 作为一个随机变量，我们也关心它分布的变化。

设 $\{X_n\}$ 是一个转移矩阵为 $P$ 的Markov链，初始分布为 $X_0 \sim \mu$ ，则 $X_1$ 的分布列为

$$\begin{aligned} P(X_1 = j) &= \sum_{k \in E} P(X_0 = k)P(X_1 = j | X_0 = k) \\ &= \sum_{k \in E} \mu_k P_{kj}, \quad j \in E \end{aligned}$$

把分布 $\mu$ 视为状态空间 $E$ 上的行向量, 那么 $X_1$ 的分布 $\mu_1$ 为 $\mu P$ . 自然地,  $X_n$ 的分布为 $\mu P^{(n)} = \mu P^n$ .

**Definition 4.7 (不变测度)** 设 $\pi = \{\pi_i, i \in E\}$ 为 $E$ 上的测度, 若 $\pi$ 满足不变方程

$$\pi_j = \sum_{k \in E} \pi_k P_{kj}, \quad \forall j \in E$$

则称 $\pi$ 为 $P$ 的不变测度(*invariant measure*). 进一步, 若 $\sum_{i \in E} \pi_i = 1$ , 则称 $\pi$ 为不变分布. 也称 $\pi$ 为以 $P$ 为转移矩阵的Markov链的不变测度.

**注1.** 显然,  $\pi_i = 0, \forall i \in E$ 是一个平凡的不变测度. 若 $\pi$ 是非平凡的, 且 $\sum_{i \in E} \pi_i < \infty$ 的不变测度, 那么我们总可以将其归一化得到不变分布

$$\frac{\pi_i}{\sum_{k \in E} \pi_k}, \quad i \in E$$

**注2.** 设 $\pi$ 为不变分布, 如果 $X_0 \sim \pi$ , 那么

$$(X_0, X_1, \dots, X_n) \stackrel{d}{=} (X_m, X_{m+1}, \dots, X_{m+n}), \quad \forall n, m \geq 1. \quad (4.4)$$

满足(4.4)的随机过程称为平稳过程(Stationary Process), 即任何时刻 $m$ 作为起点, 相同时间间隔 $n$ 随机过程的有限维分布都是相同的. 因此不变分布 $\pi$ 也被称为平稳分布(Stationary Distribution).

**注3.** 若 $E$ 是有限的, 把 $\pi$ 视为行向量, 不变分布 $\pi$ 就是矩阵 $\mathbf{P}$ 的特征值为1的左特征向量, 于是我们可以通过解线性方程组

$$\pi = \pi P$$

来求不变分布.

**注4.** 一般而言, Markov链可以没有不变分布, 例如一维简单随机游动(思考一下); 也可以有多个不变分布, 例如以单位矩阵为转移矩阵的Markov链, 或者一个非平凡的例子: 转移矩阵为

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

的Markov链，它的不变分布为

$$\left(\frac{\alpha}{2}, \frac{\alpha}{2}, \frac{\beta}{2}, \frac{\beta}{2}\right), \quad \alpha, \beta \geq 0, \alpha + \beta = 1$$

有无穷多个。

Markov链研究中的一个基本问题是确定所有的不变分布。有下列几个自然的问题：

1. 不变分布在什么条件下存在？
2. 假设不变分布存在，那么它在什么条件下是唯一的？在什么条件下  $\mu \mathbf{P}^n$  收敛到不变分布？
3. 如果  $\mu \mathbf{P}^n$  收敛到不变分布，那么收敛速度有多快？

我们在后面的学习中会回答这些问题。

#### Example 4.8 (两状态的Markov链 II)

考虑状态空间为  $\{0, 1\}$ ，转移矩阵为

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}$$

的Markov链，它的不变分布  $\pi$  满足：

$$\pi_0(1 - \alpha) + \pi_1\beta = \pi_0$$

$$\pi_0\alpha + \pi_1(1 - \beta) = \pi_1$$

结合  $\pi_0 + \pi_1 = 1$ ，可以解得

$$\pi_0 = \frac{\beta}{\alpha + \beta}, \pi_1 = \frac{\alpha}{\alpha + \beta}.$$

思考：回顾转移概率的极限：

$$\lim_{n \rightarrow \infty} P^{(n)} = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} = \begin{pmatrix} \pi & \pi \\ \pi & \pi \end{pmatrix}.$$

和不变分布之间有什么关系呢？