STAT0008: Stochastic Processes

Lecture 4 - Discrete-time Markov Chains

Lecturer: Weichen Zhao — Spring 2026

Key concepts:

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

考虑一个随机序列$\{X_n,\ n=0,1,2,\dots,n\}$，我们关心联合分布

$$P(X_0=i_0,X_1=i_1,\cdots,X_n=i_n)$$

若有独立性，

$$P(X_0=i_0,X_1=i_1,\cdots,X_n=i_n)=P(X_0=i_0)P(X_1=i_1)\cdots P(X_n=i_n)$$

这是我们初等概率论中经常关心的情况，但如果没有独立性，

$$\begin{aligned}&P(X_0=i_0,X_1=i_1,\cdots,X_n=i_n)\\&\qquad=P(X_0=i_0)P(X_1=i_1|X_0=i_0)\cdots P(X_n=i_n|X_{n-1}=i_{n-1},\dots,X_0=i_0)\end{aligned}$$

当序列长度n较大时，条件概率$P(X_n=i_n|X_{n-1}=i_{n-1},\dots,X_0=i_0)$难以研究，一种合理的简化是 Markov 性。

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：

$$P\{X_{n+1}=j|X_n=i,X_{n-1}=i_{n-1},\dots,X_1=i_1,X_0=i_0\}=P\{X_{n+1}=j|X_n=i\} \tag{4.1}$$

成立。称$P(n,i;n+1,j) \triangleq P\{X_{n+1}=j|X_n=i\}$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

$$\begin{aligned}&P\{X_{n+1}=j,X_{n-1}=i_{n-1},\ldots,X_0=i_0|X_n=i\}\\&=P\{X_{n+1}=j|X_n=i\}\cdot P\{X_{n-1}=i_{n-1},\ldots,X_0=i_0|X_n=i\}\end{aligned}\tag{4.2}$$

即验证：

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

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)意味着

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

(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，

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

注3. 若矩阵$P=(P_{ij})_{i,j\in E}$满足：

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

则称P为E上的转移矩阵 (transition matrix)，或者随机矩阵(stochastic matrix)。称P_{ij}为从状态i到j的转移概率。

注4. 算子视角。记 $\mathcal{B}(E)$ 为E上所有有界函数 $f:E\to\mathbb{R}$的集合，可以定义算子

$$\begin{aligned}&P:f\in\mathcal{B}(E)\to P[f]\in\mathcal{B}(E),\\&f(x)\mapsto P[f](x)\triangleq\sum_{j\in E}P_{ij}f(j)=\mathbb{E}[f(X_{n+1})|X_n=i]\end{aligned}$$

下面定理说明了Markov链可以看成是一个“噪声”驱动的动力系统(Dynamical Systems)。

Theorem 4.2 *设 $\{Z_n\}_{n\geq 1}$ 是一个独立同分布的随机变量序列，取值于任意空间 F。设 E 是一个可数空间，$f: E \times F \to E$ 是某个函数。设 X_0 是一个在 E 中取值的随机变量，与 $\{Z_n\}_{n\geq 1}$ 独立。则递推方程*

$$X_{n+1} = f(X_n, Z_{n+1}) \tag{4.4}$$

定义了一个时齐Markov链。

Proof: 对递推关系式 (4.4) 进行迭代可知，对所有 $n \geq 1$，存在一个函数 g_n，使得 $X_n = g_n(X_0, Z_1, \ldots, Z_n)$，因此

$$\begin{aligned}
&P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\
&\qquad = P(f(i, Z_{n+1}) = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\
&\qquad = P(f(i, Z_{n+1}) = j),
\end{aligned}$$

因为事件 $\{X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i\}$ 可由 $X_0, Z_1, \ldots, Z_n$ 表示，因此与 Z_{n+1} 独立。类似地，

$$P(X_{n+1} = j \mid X_n = i) = P(f(i, Z_{n+1}) = j)$$

因此我们得到一个时齐Markov链，而且

$$P_{ij} = P(f(i, Z_1) = j). \tag{4.5}$$

■

注1. 上面的条件可以适当放宽，不再满足$\{Z_n\}$独立，仅要求$n \geqslant 1$时，在给定X_n的条件下，Z_{n+1}和$\{X_0, Z_1, Z_2, \cdots, Z_n\}$独立，即

$$\begin{aligned}
P(Z_{n+1} =& m_{n+1} | X_n = i_n, Z_{n-1} \\
&= m_{n-1}, \cdots, Z_1 = m_1) = P(Z_{n+1} = m_{n+1} | X_n = i_n)
\end{aligned}$$

同样可以证明$\{X_n\}_{n\geqslant 1}$是 Markov 链，且当$\{Z_n\}$同分布时，该链仍然是时齐的，一步转移概率为

$$P_{ij} = P(f(i, Z_1) = j | X_0 = i)$$

注2. 不是所有的时齐Markov链都可以写成由噪声驱动的动力系统形式。参考：Pierre Brémaud, Markov Chains (Second Edition), 2020, Springer. Example 2.1.3

4.2 例子

Example 4.3 (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。

我们可以用随机变量序列驱动来构造随机游走。假设$\xi_1, \xi_2, \ldots$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 $\xi_1, \xi_2, \ldots$，令

$$S_n \triangleq S_0 + \xi_1 + \cdots + \xi_n = S_{n-1} + \xi_n.$$

那么S_n为从S_0出发的一维简单随机游走。

ξ_n可以服从更一般的分布，此时，S_n称为一般随机游走。

Example 4.4 (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$2N$。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 $X_n + N$，则 $\{X_n\}$ 的状态空间为 $E = \{-N, -N+1, \ldots, 0, 1, \ldots, N\}$，可以表示为

$$X_{n+1} = X_n + Z_{n+1}$$

其中驱动噪声Z_n是取值为± 1的随机变量，满足

$$P(Z_{n+1} = 1 | X_n = i) = \frac{N-i}{2N}, \quad P(Z_{n+1} = -1 | X_n = i) = \frac{N+i}{2N}$$

则由定理*4.2*注*1*，$\{X_n\}$为一个离散时间*Markov*链，转移概率为

$$P(X_{n+1} = j | X_n = i) = \begin{cases} \frac{N-i}{2N}, & j = i+1 \\ \frac{N+i}{2N}, & j = i-1 \\ 0, & |j-i| \neq 1 \end{cases}$$

注. 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 4.5 (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

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

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 4.6 (Chapman-Kolmogorov 方程) 对任意$i, j \in E, m, n \geqslant 0$,

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

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

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

Example 4.7 (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$n \to \infty$，n步转移概率存在极限

$$\lim_{n\to\infty} P_{00}^{(n)} = \lim_{n\to\infty} P_{10}^{(n)} = \frac{\beta}{\alpha+\beta}$$

$$\lim_{n\to\infty} P_{11}^{(n)} = \lim_{n\to\infty} P_{01}^{(n)} = \frac{\alpha}{\alpha+\beta}$$

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

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 μ 视为状态空间 E 上的行向量，那么 X_1 的分布 μ_1 为 μP. 自然地，X_n 的分布为 $\mu P^{(n)} = \mu P^n$.

Definition 4.8 (不变测度) 设$\pi = \{\pi_i, i \in E\}$为E上的测度，若π满足不变方程

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

则称π为P的不变测度*(invariant measure)*。进一步，若$\sum_{i \in E} \pi_i = 1$，则称π为不变分布。也称π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π为不变分布，如果$X_0 \sim \pi$，那么

$$(X_0, X_1, \cdots, X_n) \stackrel{d}{=} (X_m, X_{m+1}, \cdots, X_{m+n}), \quad \forall n, m \geqslant 1. \tag{4.6}$$

满足(4.6)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π也被称为平稳分布(Stationary Distribution).

注3. 若E是有限的，把π视为行向量，不变分布π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)；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

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

有无穷多个。

Example 4.9 (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π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

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

回顾转移概率的极限：

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

和不变分布之间有什么关系呢？