Lecture Notes of Probability Theory
Lesson 10

范舟
zhou.fan@sjtu.edu.cn
上海交通大学 ACM 班

目录

1 内容摘要

在 Lesson 10 中，吴老师所讲授的内容主要有：

- 介绍了 Box and Subboxes 问题，并详述其证明过程.
- 回顾了一道此前的习题，并引入高阶马尔可夫链的概念.
- 引入马尔可夫链中周期性的定义，并介绍了两个与周期性相关的定理及其证明.
- 介绍了与马尔可夫链历史信息有关的一个定理，并给出其证明的第一部分.

笔者所整理的 lecture notes 中如有存在疏漏之处，还请各位读者批评指正.

2 Box and Subboxes 问题

2.1 问题描述

Box and Subboxes 问题 集合 A 与 $B^{(i)}(i=1,2,\dots,n)$ 均是 m 个集合的笛卡尔积[1]

$$A = A_1 \times A_2 \times \cdots \times A_{m-1} \times A_m$$
$$B^{(i)} = B_1^{(i)} \times B_2^{(i)} \times \cdots \times B_{m-1}^{(i)} \times B_m^{(i)} \quad (i=1,2,\dots,n)$$

且有

$$A = \bigcup_{i=1}^{n} B^{(i)}$$

$$1_A \equiv \sum_{i=1}^{n} 1_{B^{(i)}} \pmod 2 \tag{1}$$

对于某个集合 S, 函数 1_S 定义[2]如下：

$$1_S(x) = \begin{cases} 1, & x \in S \\ 0, & x \notin S \end{cases}$$

求证 若 A 与 $B^{(i)}$ 满足

$$B_j^{(i)} \neq A_j \quad (i=1,2,\dots,n \quad j=1,2,\dots,m) \tag{2}$$

则有

$$n \geq 2^m$$

[1] Cartesian product

[2] 此处 1_S 函数的概念在 Lesson 4 中引入，为了方便读者，在这里重新叙述其定义.

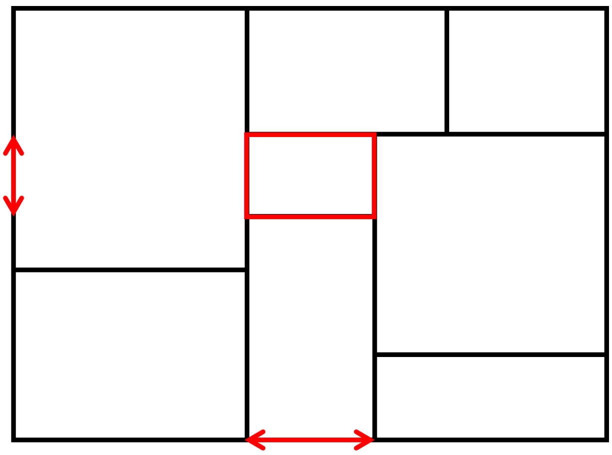

图 1: 二维情况下的一种可能划分

评注 读者也可以思考将条件 (1) 改为如下的条件 (3) 时的证明，条件 (3) 是条件 (1) 的一个特例，注意到其等价于：A 是 $B^{(i)}(i=1,2,\dots,n)$ 的不交并.

$$1_A = \sum_{i=1}^{n} 1_{B^{(i)}} \tag{3}$$

下文的“解答与证明”部分给出了在 (1) 式条件下的证明，因为 (1) 是更弱的条件，这一证明具有一般性，包含了条件 (3) 下的结论.

2.2 直观认识

可以将 A 看作一个 m 维笛卡尔积空间中的“盒子”(box)，而每一个 $B^{(i)}$ 是包含于 A 的一个“子盒子”(subbox)，$B^{(i)}$ 的全体的并即为 A. 将 subbox 投影到第 j 个维度上，$B_j^{(i)}$ 是 A_j 的一个子集.

对于二维的情况，图 1 展示了一种可能的 subbox 的划分，图中的 box 是各个 subbox 的不交并. 以图中标注为红色的 subbox 为例，其在两个维度上的投影均为一段子区间. 有必要指出，为了使展示效果直观，图中的 subbox 在每个维度上均是连续的，但这并不是必要的性质. 事实上，我们只讨论 A 在每个维度上的投影 A_j 均为有限集的情况.

2.3 证明与解答

证明 考虑 A 的一个大小为奇数的子集 C

$$C = C_1 \times C_2 \times \cdots \times C_{m-1} \times C_m$$

$$C_j \subseteq A_j \quad (j=1,2,\dots,m)$$

$$|C| \equiv 1 \pmod 2 \tag{4}$$

由 (4) 式可以立即得到

$$|C_j| \equiv 1 \pmod 2 \quad (j = 1, 2, \ldots, m)$$

令 $O(A)$ 为 A 的所有大小为奇数的子集的全集，即

$$O(A) = \{C \mid C \subseteq A, |C| \equiv 1 \pmod 2\}$$

进一步地，令

$$O_i(A) = \{C \mid C \in O(A), C \cap B^{(i)} \in O(A)\} \tag{5}$$

由定义 (5) 可以直接得到

$$O(A) \supseteq \bigcup_{i=1}^{n} O_i(A)$$

下面证明

$$O(A) = \bigcup_{i=1}^{n} O_i(A) \tag{6}$$

只需证明，对于任意的 $C \in O(A)$，一定存在 $i \in \{1, 2, \ldots, n\}$，使得 $C \in O_i(A)$. 使用反证法，假设存在 $C \in O(A)$，且对于任意的 $i \in \{1, 2, \ldots, n\}$ 都有 $C \notin O_i(A)$，即

$$|C \cap O_i(A)| \equiv 0 \pmod 2 \quad (i = 1, 2, \ldots, n)$$

根据 (1) 式，有

$$|C| \equiv \sum_{i=1}^{n} |C \cap O_i(A)| \pmod 2$$

即得 $|C| \equiv 1 \pmod 2$，这与 $C \in O(A)$ 矛盾，因此假设不成立，即证得 (6).

对于任意的 $i \in \{1, 2, \ldots, n\}$，考虑随机事件“在 $O(A)$ 中随机抽取一个元素 C，$C \in O_i(A)$”的概率. 由于 $C \in O(A)$，$C \in O_i(A)$ 等价于 $|C \cap B^{(i)}| \equiv 1 \pmod 2$，即

$$|C_j \cap B_j^{(i)}| \equiv 1 \pmod 2$$

由于各个维度的子事件“$|C_j \cap B_j^{(i)}| \equiv 1 \pmod 2$”是独立的，有

$$\begin{aligned}
&P\Big(C \in O_i(A) \,\Big|\, C \in O(A)\Big) \\
=&P\Big(|C \cap B^{(i)}| \equiv 1 \pmod 2 \,\Big|\, C \in O(A)\Big) \\
=&\prod_{j=1}^{m} P\Big(|C_j \cap B_j^{(i)}| \equiv 1 \pmod 2 \,\Big|\, C \in O(A)\Big) \\
=&\prod_{j=1}^{m} P\Big(|C_j \cap B_j^{(i)}| \equiv 1 \pmod 2 \,\Big|\, |C_j| \equiv 1 \pmod 2, C_j \subseteq A_j\Big)
\end{aligned} \tag{7}$$

下面证明[3]，对于任意的 $j \in \{1, 2, \ldots, m\}$，有

$$P\Big(|C_j \cap B_j^{(i)}| \equiv 1 \pmod 2 \,\Big|\, |C_j| \equiv 1 \pmod 2, C_j \subseteq A_j\Big) = \frac{1}{2} \tag{8}$$

[3]Lesson 10 课上所讲的这一部分证明过程原本存在一定疏漏，感谢刘天怡同学指出并提醒.

对于任意一个集合 $S \neq \varnothing$，其大小为奇数的子集与大小为偶数的子集个数相等，即

$$P\Big(|T| \equiv 1 \pmod 2 \,\Big|\, T \subseteq S\Big) = P\Big(|T| \equiv 0 \pmod 2 \,\Big|\, T \subseteq S\Big) = \frac{1}{2} \tag{9}$$

记 $D_1 = C_j \cap B_j^{(i)}, D_2 = C_j \cap (A_j \setminus B_j^{(i)})$，则 C_j 是 D_1 与 D_2 的不交并.
由条件 (2) 知，$A_j \setminus B_j^{(i)} \neq \varnothing$. 为了方便，我们忽略为空集的 $B^{(i)}$，这并不影响 $n \geq 2^m$ 的证明[4]. 又有 $D_1 \subseteq B_j^{(i)}, D_2 \subseteq (A_j \setminus B_j^{(i)})$. 因此，根据 (9) 有

$$P\Big(|D_1| \equiv 1 \pmod 2 \Big) = \frac{1}{2}$$
$$P\Big(|D_2| \equiv 0 \pmod 2 \Big) = \frac{1}{2}$$

根据条件概率的定义，有

$$\begin{aligned}
&P\Big(|C_j \cap B_j^{(i)}| \equiv 1 \pmod 2 \,\Big|\, |C_j| \equiv 1 \pmod 2, C_j \subseteq A_j\Big) \\
=&\frac{P\Big(|C_j \cap B_j^{(i)}| \equiv 1 \pmod 2, |C_j| \equiv 1 \pmod 2, C_j \subseteq A_j\Big)}{P\Big(|C_j| \equiv 1 \pmod 2, C_j \subseteq A_j\Big)} \\
=&\frac{P\Big(|D_1| \equiv 1 \pmod 2, |D_1| + |D_2| \equiv 1 \pmod 2\Big)}{P\Big(|D_1| + |D_2| \equiv 1 \pmod 2\Big)} \\
=&\frac{P\Big(|D_1| \equiv 1 \pmod 2, |D_2| \equiv 0 \pmod 2\Big)}{P\Big(|D_1| + |D_2| \equiv 1 \pmod 2\Big)} \\
=&\frac{\frac{1}{2} \times \frac{1}{2}}{\frac{1}{2}} = \frac{1}{2}
\end{aligned}$$

即证得 (8). 进一步地，根据 (7) 有

$$\begin{aligned}
&P\Big(C \in O_i(A) \,\Big|\, C \in O(A)\Big) \\
=&\prod_{j=1}^{m} P\Big(|C_j \cap B_j^{(i)}| \equiv 1 \pmod 2 \,\Big|\, |C_j| \equiv 1 \pmod 2, C_j \subseteq A_j\Big) \\
=&\prod_{j=1}^{m} \frac{1}{2} = \frac{1}{2^m}
\end{aligned}$$

又有

$$P\Big(C \in O_i(A) \,\Big|\, C \in O(A)\Big) = \frac{|O_i(A)|}{|O(A)|}$$

即得

$$\frac{|O_i(A)|}{|O(A)|} = \frac{1}{2^m}$$

结合 (6) 可得

$$n \geq 2^m$$

证毕.

[4]这是由于“非空的 $B^{(i)}$ 个数大于等于 2^m”本身蕴含了“$B^{(i)}$ 个数大于等于 2^m”.

2.4 拓展延伸

本题中存在这一条件

$$B_j^{(i)} \neq A_j, \quad (i = 1, 2, \ldots, n \quad j = 1, 2, \ldots, m)$$

可以将其直观理解为：在每一个维度，至少需要 2 个 subbox 才可将这一维度全部覆盖. 感兴趣的读者可以进一步思考，将这一条件改为下列条件之一时，类似的结论是否依然成立，并可以思考其证明（或证伪）的过程.

1. 在每一个维度，至少需要 3 个 subbox 才可将这一维度全部覆盖. 此时证明 $n \geq 3^m$.
2. 至少需要 n_j 个 subbox 才可将第 j 个维度全部覆盖. 此时证明 $n \geq \prod_{j=1}^m n_j$.

3 习题回顾：高阶马尔可夫链

3.1 问题描述

已知 $M_1 = (X_0, X_1, X_2, \ldots)$ 是有限状态空间 Ω 上的马尔可夫链[5]，f 是从 Ω 到另一个有限状态空间 Ω' 的映射

$$f : \Omega \to \Omega'$$

则 $M_2 = \Big(f(X_0), f(X_1), f(X_2), \ldots\Big)$ 是否仍是马尔可夫链？

评注 此道习题是在 Lesson 9 提出的，当时已经有同学举出反例证明 M_2 不一定仍是马尔可夫链. 读者可以思考：当 M_1 与 f 满足何种条件时，M_2 仍是马尔可夫链？在 Lesson 10 中，提出关于本题的进一步思考：M_2 是否一定是高阶马尔可夫链[6]？下面引入高阶马尔可夫链的定义.

高阶马尔可夫链 高阶马尔可夫链是满足广义马尔可夫性质的随机变量序列 $(X_0, X_1, X_2, \ldots)$，即下一个状态 X_{n+1} 的概率分布只依赖于此前的有限个状态 $X_n, X_{n-1}, \ldots, X_{n-k}$. 即存在 k 使得

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

[5] Markov chain

[6] higher-order Markov chain

4 马尔可夫链的周期性

4.1 定义

马尔可夫链状态的周期 对于有限状态集 $\{S_1, S_2, \dots, S_k\}$ 上的马尔可夫链 M，设其转移概率矩阵为 P，定义其某一状态 S_i 的周期如下

$$d(S_i) \triangleq \gcd\{n \mid n \geq 1, (P^n)_{ii} > 0\}$$

特别地，若 $d(S_i) = 1$，我们称状态 S_i 是非周期的[7].

评注 状态 S_i 的周期的定义即为从 S_i 出发经过若干步转移又返回 S_i 的所有环路长度的最大公因数.

马尔可夫链的周期性 若 M 是有限状态集 $\{S_1, S_2, \dots, S_k\}$ 上的马尔可夫链，则 M 是非周期的当且仅当对于任意的 $S_i (i = 1, 2, \dots, k)$ 都有 $d(S_i) = 1$.

4.2 习题

求证 马尔可夫链状态的周期 $d(S_i)$ 可以对于马尔可夫链的强联通分支定义，即对于马尔可夫链 M 的任意强联通分支 G，有

$$d(S_i) = d(G), \quad \forall S_i \in G \tag{10}$$

4.3 定理 1

4.3.1 定理描述

定理 1 若 M 是有限状态集 $\{S_1, S_2, \dots, S_k\}$ 上的马尔可夫链，其转移概率矩阵为 $P_{k\times k}$，则 M 是非周期的当且仅当存在 $N < \infty$ 使得

$$(P^n)_{ii} > 0, \quad \forall n > N, \forall i \in \{1, 2, \dots, k\}$$

4.3.2 定理 1 的证明

为了证明定理 1，我们首先引入下面的引理 1.

引理 1 令 A 为一个正整数的集合，使得 $\gcd A = 1$ 且 A 对于加法满足封闭性，即

$$x + y \in A, \quad \forall x, y \in A \tag{11}$$

那么有 $\mathbb{N} \setminus A$ 是有限集, 证毕.

[7] aperiodic

引理 1 的证明 由于 $\gcd A = 1$，因此 1 可以表示为 A 中元素的线性组合，即

$$1 = \sum_j x_j a_j, \quad a_j \in A, x_j \neq 0 \tag{12}$$

令 $c = \sum_j |x_j| a_j, N = c^2$，则对于任意的 $n \geq N$，由带余除法，可将 n 表示为

$$n = qc + r, \quad 0 \leq r < c, q \geq c$$

再根据 (12)，有

$$\begin{aligned} n &= q \sum_j |x_j| a_j + r \sum_j x_j a_j \\ &= \sum_j (q|x_j| + r x_j) a_j \end{aligned}$$

因为 $q > r$，有 $q|x_j| + r x_j \geq 0$，再根据 (11) 可得 $n \in A$，因此 $|\mathbb{N} \setminus A| \leq N$，即证得其为有限集.

定理 1 的证明 首先证明定理 1 中等价性的一个方向.
设 M 是有限状态集 $\{S_1, S_2, \ldots, S_k\}$ 上的非周期马尔可夫链，其转移概率矩阵为 $P_{k \times k}$.
对于任意的 $i \in \{1, 2, \ldots, k\}$，令

$$A_i = \{n \mid (P^n)_{ii} > 0\}$$

因为 M 是非周期的，有

$$\gcd A_i = 1 (i = 1, 2, \ldots, k)$$

令 $B_i = \mathbb{N} \setminus A_i$，由引理 1 知 B_i 为有限集. 令 $B = \bigcup_{i=1}^k B_i$，则 B 为有限集. 因此可以取到 B 的最大值 N，有

$$(P^n)_{ii} > 0 \quad \forall n > N, \forall i \in \{1, 2, \ldots, k\} \tag{13}$$

下面证明定理 1 中等价性的另一方向.
设已知 (13)，取质数 p_1, p_2 使得 $N < p_1 < p_2$，则对于任意 $i \in \{1, 2, \ldots, k\}$，有

$$(P^{p_1})_{ii} > 0, (P^{p_2})_{ii} > 0$$

则 $d(S_i) \mid gcd(p_1, p_2)$，即 $d(S_i) = 1$. 因此 S_i 是非周期的，进一步有 M 是非周期的, 证毕.

4.3.3 拓展延伸

习题 定理 1 要求 M 是有限状态集上的马尔可夫链，其转移概率矩阵 P 相应也是一个有限大小的矩阵，读者不妨思考，若 M 是无限状态集上的马尔可夫链，定理 1 中的等价性是否仍然成立？若 M 是高阶马尔可夫链，定理 1 是否成立？

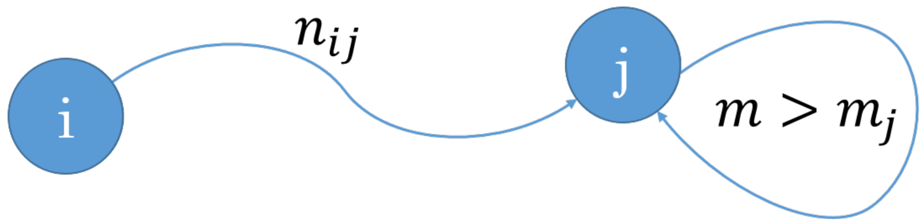

图 2: 定理 2 中状态转移的一条路径

评注 在对于无限状态集上的马尔可夫链或高阶马尔可夫链讨论此定理时，需要首先定义无限状态马尔可夫链或高阶马尔可夫链的周期性.

4.4 定理 2

4.4.1 定理描述

定理 2 令 $M = (X_0, X_1, \dots)$ 为有限状态集 $\{S_1, S_2, \dots, S_k\}$ 上的不可约[8]且非周期的马尔可夫链，设其转移概率矩阵为 P，那么存在 $N < \infty$，使得

$$P^n > 0, \quad \forall n \geq N$$

评注 定理 1 显示，对于非周期的马尔可夫链，当 n 足够大时有 P^n 的对角线元素恒大于 0. 定理 2 进一步显示了，若增加不可约的条件，当 n 足够大时 P^n 的所有元素均恒大于 0.

4.4.2 定理 2 的证明

证明 因为 M 是不可约的马尔可夫链,对于任意的 $i, j \in \{1, 2, \dots, k\}$,存在 $n_{ij} > 0$,使得 $(P^{n_{ij}})_{ij} > 0$. 因为 M 是非周期的马尔可夫链，根据定理 1，对于任意的 $i \in \{1, 2, \dots, k\}$，存在 m_i，使得

$$(P^m)_{ii} > 0, \quad \forall m \geq m_i$$

令

$$N = \max_{i,j}\{n_{ij} + m_j\}$$

则对于任意的 $i, j \in \{1, 2, \dots, k\}$，任意的 $n \geq N$，令 $m = n - n_{ij}$，则 $m \geq m_j$，有

$$(P^n)_{ij} \geq (P^{n_{ij}})_{ij}(P^m)_{jj} > 0$$

证毕.

[8]一个马尔可夫链被称为是不可约的 (irreducible)，当且仅当其状态转移图是强连通图.

直观理解 可以将定理 2 的证明过程理解为，对于任意的状态 S_i, S_j，可以按照如下的方式找到一条概率大于 0 的转移步数为 $n \geq N$ 的路径：如图 2，首先从 S_i 出发转移 n_{ij} 步转移到 S_j，再由 S_j 出发转移 $m = n - n_{ij}$ 步返回 S_j.

5 马尔可夫链的起点

5.1 定理 3

这里我们直接使用前面一道习题的结论 (10)，即可以定义马尔可夫链中一个强连通分量的周期.

定理 3 设有限状态马尔可夫链 M 的强连通分量为 $G_1, G_2, \ldots, G_k$，其中 $G_i(i = 1, 2, \ldots, k)$ 的周期为 d_i. 则如下的极限存在

$$\lim_{n \to \infty} P_{ij}^{nd} = \Pi_{ij} \tag{14}$$

定理 3 的证明留作习题.

5.2 定理 4

定理 4 有限状态的马尔可夫链必有一个起点[9]. 即对于有限状态空间 $\{S_1, S_2, \ldots, S_k\}$ 上的马尔可夫链 M，设其转移概率矩阵为 P，对于任一概率分布 $\alpha \in \Delta^{k-1}$，令 $B = \{\beta \in \Delta^{k-1} \mid \exists i \geq 0, \beta P^i = \alpha\}$，存在 N，使得对于任意的 $\beta \in B$，有

$$\exists i \leq N, \beta P^i = \alpha$$

其中，Δ^{k-1} 表示 $k-1$ 阶概率单纯型，即

$$\Delta^{k-1} = \{x \in R^k \mid x \geq 0, \sum_i x_i = 1\}$$

直观理解 可以将定理 4 直观理解为，从有限状态马尔可夫链的任意状态出发，只能向前回溯有限多步历史. 在这里对于历史路径中可能存在的环路中的转移步数不进行重复累计，即对于概率分布 α 的每个祖先分布 β，我们只关心 β 转移到 α 所需的最小步数.

$$\beta \to \beta P \to \beta P^2 \to \cdots \to \alpha$$

下面给出定理 4 的部分证明[10].

[9] beginning

[10] 在 Lesson 10 中，由于时间所限，所讲的定理 4 证明是不完整的，证明的最后一部分在 Lesson 11 中给出.

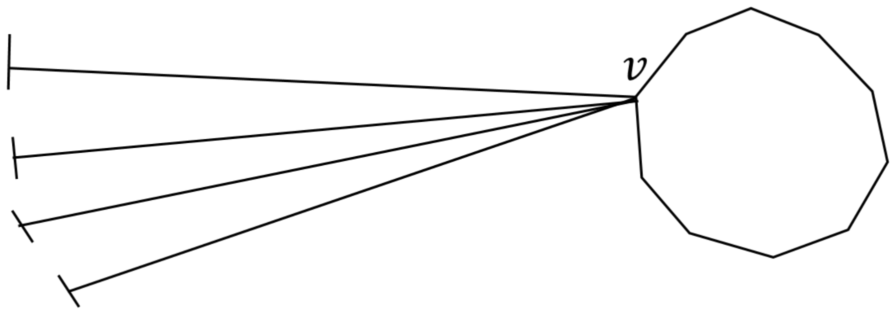

图 3: 定理 4 证明思路的直观展示

证明 设 $v \in \Delta^{k-1}$，且存在一个无限序列 $v^{(1)}, v^{(2)}, \ldots$ 满足

$$\lim_{n\to\infty} v^{(n)} P^{nd} = v$$

由因为 Δ^{k-1} 是一个紧集，存在聚点 u 与子列 $v^{(n_1)}, v^{(n_2)}, \ldots$，使得

$$\lim_{i\to\infty} v^{(n_i)} = u$$

同时根据 (14) 有

$$\lim_{i\to\infty} P^{n_i d} = \Pi$$

又有

$$\Pi P^d = \Pi$$

则

$$\begin{aligned} v &= u\Pi \\ &= u\Pi P^d \\ &= vP^d \end{aligned}$$

因此 v 处于一个周期为 d 的环上.

直观理解 上面的证明显示满足上述条件的 v 一定处于一个周期为 d 的环路上. 如图 3 所示，v 也不可能向前回溯出长度无限的“尾巴”，因为若假设存在这样无限长的回溯路径，那么在这条路径上连接 v 的节点 v' 一定也处于周期为 d 的环路上，这与假设矛盾. 这样我们证得了，v 向前回溯的“尾巴”一定是有限长度的. 有必要指出，这并不能证明定理 4 的结论，因为即使每条向前回溯路径的长度是有限的，若存在无限多条这样的路径，那么回溯路径的长度仍然可能没有上界. 定理 4 证明的下一部分即证明这样的“尾巴”只有有限多条，这一部分内容在 Lesson 11 中给出.