马氏链

应用随机过程补充讲义之二

1 定义及其基本性质

在这一章我们总假定状态空间 S 是有限或可数的. 假设 $\{X_n, n=0,1,2,\cdots\}$ 是取值于 S 的随机过程. 其有限维分布

$$P(X_1 \in A_1, X_2 \in A_2, \cdots X_n \in A_n), \quad A_i \subset S,$$

可以分解为

$$\sum_{x_i \in A_i, 1 \le i \le n} P(X_1 = x_1, X_2 = x_2, \cdots X_n = x_n);$$

而和式中每一项可以进一步改写为乘积,

$$\begin{aligned} & P(X_0 = x_0, X_1 = x_1, X_2 = x_2, \cdots X_n = x_n) \\ = \;& P(X_0 = x_0) \prod_{k=0}^{n-1} P(X_{k+1} = x_{k+1} | X_0 = x_0, X_1 = x_1, \cdots X_k = x_k). \end{aligned}$$

如果

$$P(X_{k+1} = x_{k+1} | X_0 = x_0, X_1 = x_1, \cdots X_k = x_k) = P(X_{k+1} = x_{k+1} | X_k = x_k), \tag{1}$$

则称随机过程 $\{X_n, n=0,1,2,\cdots\}$ 为 马尔可夫链, 简称 马氏链. 如果进一步,

$$P(X_{k+1} = y | X_k = x) = P(X_1 = y | X_0 = x)$$

对任意 $k \ge 1$ 恒成立，则称随机过程 $\{X_n\}$ 为 时齐马尔可夫链. 由于本章只讨论时齐马尔可夫链，时齐马尔可夫链也简称 马氏链. 我们称 $P(X_1 = j | X_0 = i)$ 为 (从 i 到 j 的一步)转移概率, 记为 $p(i,j)$; 称 $\{P(X_0 = x), x \in S\}$ 为 初始分布, 有时记为 $\mu(x)$. 反复利用 (1) 式得

$$P(X_0 = x_0, X_1 = x_1, X_2 = x_2, \cdots X_n = x_n) = \mu(x_0) \prod_{k=1}^{n} p(x_{k-1}, x_k). \tag{2}$$

可见有限维分布由初始分布和转移概率所决定. 讨论马氏链，只需给出初始分布和转移概率就可以了.

例 1. 随机游动是马氏链，设 $S_n = S_0 + \xi_1 + \xi_2 + \cdots + \xi_n$. 转移概率 $p(x,y) = P(\xi = y - x)$, 初始分布就是 S_0 的分布.

例 2. 状态空间 $S=\{0,1\}$, 转移概率 $p(0,1)=1-p(0,0)=p$, $p(1,0)=1-p(1,1)=q$. 这样决定的马氏链可以说是最简单的马氏链了. 当年马尔可夫研究普希金诗歌里元音字母和辅音字母交替出现的规律, 如果把元音字母记为 0, 辅音字母记为 1, 马尔可夫提出的模型正是这样.

例 3. Ehrenfest *模型*. 状态空间 $S=\{0,1,2,N\}$, $p(i,i+1)=1-i/N$, $p(i,i-1)=i/N$, 其他 $p(i,j)=0$ (习惯上人们只提及非零项). 这可以在物理上解释不可逆现象.

例 4. *带吸收壁的随机游动*. 状态空间 S 为整数格点, $p(0,0)=1$, 对任何 $x\neq 0$, $p(x,x+1)=p(x,x-1)=1/2$. 在到达原点之前, 该马氏链 X 就在做随机游动, 一旦到了原点就停止不动了. 这已不是严格意义上的随机游动, 但人们习惯称之为 *带吸收壁的随机游动*.

例 5. *带反射壁的随机游动*. 状态空间 $S=\{0,1,2,\cdots\}$, $p(0,1)=1$, 对任何 $x\geq 0$, $p(x,x+1)=p(x,x-1)=1/2$. 在到达原点之前, 该马氏链就在做随机游动, 一旦 (从状态 1) 到了原点, 下一步就一定返回状态 1.

例 6. *生灭链*. 状态空间 $S=\{0,1,2,\cdots\}$, $p(i,i+1)=p_i$, $p(i,i-1)=q_i$. 假设 $p_i+q_i=1$. Ehrenfest 模型, 带吸收壁的随机游动和带反射壁的随机游动都是生灭过程的特例.

例 7. *图上的简单随机游动*. 记图 $G=(V,E)$ 的顶点集为 V, 边集为 E, 对任意 $x,y\in V$, 以 $x\sim y$ 表示 x 和 y 之间有边相连. 进一步假设图 G 是简单图, 即两个顶点之间最多只有一条边相连, 没有从 x 到 x 的边. 以 d_x 表示顶点 x 的度数, 即与 x 相连的边的数目. 若 x 和 y 之间有边相连, 定义转移概率 $p(x,y)=1/d_x$, 否则为零. 对应的马氏链称为 *图 G 上的简单随机游动*. 当 G 为 d 维格点或规则树, 这与前一章的定义是一致的. 这正是这一名称的由来.

转移矩阵: 转移概率 $p(i,j)$ 有时也写成 p_{ij}. 若 S 为有限状态空间 $\{1,2,\cdots,n\}$, 称矩阵 $P=(p_{ij})$ 为 (一步)*转移 (概率) 矩阵*; 若 S 为可列状态空间, 我们仍习惯称转移概率全体 $\{p_{ij},i,j\in S\}$ 为转移阵, 记为 P.

称 $P(X_n=j|X_0=i)$ 为 (从 i 到 j 的)*n 步转移概率*, 记为 $p_{ij}(n)$, 有时也写成 $p_n(i,j)$. 补充规定 $p_{xy}(0)=\delta_{xy}$. 对任意 $x,y\in S$, $n,m\geq 0$,

$$\sum_{z\in S}p_n(x,z)p_m(z,y)=p_{n+m}(x,y).$$

此等式称为 Chapman-Kolmogorov 等式.

称矩阵 $(p_{ij}(n))$ 为 *n 步转移概率矩阵*, 暂记为 $\mathbf{P}(n)$. 因为 Chapman-Kolmogorov 等式, 当状态空间为有限时, $P(2)=P^2$, 归纳可证 n 步转移概率矩阵 $P(n)=P^n$. 而且形式上 $P(0)=P^0$ 也成立. 当状态空间为可列时, 我们并没有定义无限矩阵的乘积, 但可以依照 Chapman-Kolmogorov 等式, 认为 P^n 是 n 步转移概率全体 $P(n)$, 而且形式上 $P^nP^m=P^{n+m}$.

马尔可夫性质. 等式 (1) 称为马尔可夫性质, 有许多等价的命题 (习题 12), 其概率意义都是: 在已知随机过程现在所处的状态 $\{X_n=x\}$ 的条件下, 过去和未来的是相互独立的. 尤以下述命题为最.

命题 1 对任何 $r, s \geq 1$, $n_1 < n_2 < \cdots < n_r < n < m_1 < m_2 < \cdots < m_s$, 对任何 $x_1, \cdots, x_r$, $x, y_1, y_2, \cdots, y_s \in S$,

$$\begin{aligned} & P(X_{m_1} = y_1, \cdots, X_{m_s} = y_s; X_{n_1} = x_1, \cdots, X_{n_r} = x_r | X_n = x) \\ = \ & P(X_{n_1} = x_1, \cdots, X_{n_r} = x_r | X_n = x) \times P(X_{m_1} = y_1, \cdots, X_{m_s} = y_s | X_n = x). \end{aligned}$$

2 平稳分布

设马氏链的转移矩阵为 $\mathbf{P}$, 初始分布为 $\mu = \{\mu(x), x \in S\}$. 记 X_n 的分布为 μ_n, 则

$$\mu_n(y) = \sum_x \mu(x) p_n(x, y). \tag{3}$$

若 $\mu_n = \mu$ 对所有 n 都成立，则称 μ 为 *不变分布*.

若马氏链 $\{X_n\}$ 的初始分布是不变分布，则 $X_{n+1}, X_{n+2}, \cdots, X_{n+k}$ 的联合分布与 n 无关，即马氏链 $\{X_n\}$ 是一 *严平稳过程*, 故又称 μ 为 *平稳分布*. 在我们课程里 *不变分布* 和 *平稳分布* 是同义语.

马氏链可以没有不变分布，例如一维简单随机游动；也可以有多个不变分布，平凡的例子可以取转移阵为单位阵. 但对于有限状态空间马氏链，事情要简单的多，不变分布总存在. 为方便叙述，简化证明，本节讨论只限于有限状态空间. 许多结论可以推广到可列状态空间.

若状态空间 S 是有限的，总可以与 $\{1, 2, \cdots, n\}$ 做一一对应，因此不妨设 $S = \{1, 2, \cdots, n\}$. 把 μ 和 μ_n 视为行向量，则 (3) 式可以写为更简约的形式

$$\mu_n = \mu \mathbf{P}^n.$$

π 是不变分布当且仅当 $\pi = \pi \mathbf{P}$, 即 π 就是矩阵 $\mathbf{P}$ 的特征值为 1 的特征向量. 根据非负矩阵的 Frobenius 定理，不变分布总是存在的. 可以通过求解特征向量来确定不变分布.

以 $\mathcal{M}$ 记 S 上概率分布全体，或等价的，

$$\mathcal{M} = \{(a_1, a_2, \cdots, a_n); a_i \geq 0, \sum_{i=1}^n a_i = 1\}.$$

转移阵 $\mathbf{P}$ 诱导一个从 $\mathcal{M}$ 到 $\mathcal{M}$ 的映照： $\nu \to \nu \mathbf{P}$. 不变分布 π 是这一映照下的不动点. 现来考察 $\nu \mathbf{P}^n$ 与 π 的接近程度. 为此引入 $\mathcal{M}$ 中两点 μ 和 ν 之间的距离：

$$||\mu - \nu|| = \sum_{x \in S} |\mu(x) - \nu(x)| = 2 \max_{A \subset S} |\mu(A) - \nu(A)|.$$

容易验证

$$||\pi - \nu \mathbf{P}|| \leq ||\pi - \nu||.$$

但等号可以成立.

例 8.

$$\mathbf{P}=\begin{pmatrix}0 & 1\\ \\ 1 & 0\end{pmatrix},\quad \pi=(\frac{1}{2},\frac{1}{2}),\quad \nu=(1,0).$$

定义. (周期). 设 $x\in S$. $\{n\geq 1; p_{xx}(n)>0\}$ 的最大公约数称为状态 x 的 周期. 若所有状态的周期是一样, 称为马氏链的周期. 如果马氏链的周期为 1, 称该马氏链是 非周期的.

在前面的例 8 中, 马氏链的周期为 2.

命题 2. 假设转移阵 $\mathbf{P}=(p_{ij})$ 满足 $\min_{ij}p_{ij}>0$, π 是其不变分布, 则存在常数 $\alpha<1$, 对任何 $\mu,\nu\in\mathcal{M}$,

$$||\mu\mathbf{P}-\nu\mathbf{P}||\leq\alpha||\mu-\nu||;\quad ||\pi-\nu\mathbf{P}^k||\leq 2\alpha^k.$$

证明. 记 $\min_{ij}p_{ij}=\delta$. 由假设 $\sum_i\mu_ip_{ij}=\delta+\sum_i\mu_i(p_{ij}-\delta)$, 而 $(p_{ij}-\delta)\geq 0$,

$$\begin{aligned}||\mu\mathbf{P}-\nu\mathbf{P}|| &= \sum_j|\sum_i\mu_ip_{ij}-\sum_i\nu_ip_{ij}|\\ &= \sum_j|\sum_i\mu_i(p_{ij}-\delta)-\sum_i\nu_i(p_{ij}-\delta)|\\ &= \sum_j|\sum_i(\mu_i-\nu_i)(p_{ij}-\delta)|\\ &\leq \sum_j\sum_i|\mu_i-\nu_i|(p_{ij}-\delta)\\ &= \sum_i|\mu_i-\nu_i|\sum_j(p_{ij}-\delta)=(1-n\delta)||\mu-\nu||.\end{aligned}$$

取 $\alpha=1-n\delta$ 即得第一条不等式, 重复利用第一条不等式 k 次, 以及 $||\pi-\nu||\leq 2$, 即得第二条不等式. 证毕.

命题 2 的条件要求转移阵里的元素都不能为 0, 显然这样的条件是难以满足的. 一种减弱的办法是以 P^n 代替 P. 下面的命题就可以此思路证明.

定义 (不可约). 设 P 是状态空间 S 上的转移概率阵. 假设 C 是 S 的子集, 若对任何 $x\in C$, $\sum_{y\in C}p(x,y)=1$, 则称 C 是 P- 闭集. 显然 S 自身是 P- 闭集. 若 S 的任何真子集都不是 P- 闭集, 则称由 P 所决定的马氏链是 不可约 的.

命题 3 不可约非周期马氏链具有唯一不变分布, 对任何 $\nu\in\mathcal{M}$, 存在常数 $\alpha>0$, $||\pi-\nu\mathbf{P}^k||\leq 2e^{-\alpha n}$.

假如转移阵 $\mathbf{P}$ 的特征值为 $1, \lambda_2, \lambda_3, \cdots, \lambda_n$, 相应的特征向量为 $\pi, \pi_2, \pi_3, \cdots, \pi_n$, 成为 n 维欧氏空间的一组正交基. 任何 $\nu \in \mathcal{M}$ 可以表示为 $\pi, \pi_2, \pi_3, \cdots, \pi_n$ 的线性组合.

$$\nu = C_1\pi + C_2\pi_2 + C_3\pi_3 + \cdots + C_n\pi_n.$$

于是

$$\begin{aligned} \nu\mathbf{P}^k &= C_1\pi\mathbf{P}^k + C_2\pi_2\mathbf{P}^k + C_3\pi_3\mathbf{P}^k + \cdots + C_n\pi_n\mathbf{P}^k \\ &= C_1\pi + C_2\lambda_2^k\pi_2 + C_3\lambda_3^k\pi_3 + \cdots + C_n\lambda_n^k\pi_n. \end{aligned}$$

已知 $\max\{|\lambda_j|; 2 \le j \le n\} \le 1$, 只要 $\max\{|\lambda_j|; 2 \le j \le n\} = \beta < 1$, 上面等式的右端除第一项以外，其余各项均指数衰减，顺便推出 $C_1 = 1$.

$$||\pi - \nu P^k|| \le \sum_{j=2}^{n} |C_j||\lambda_j|^k||\pi_j|| \le C(1-\beta)^k.$$

我们称 $1 - \max\{|\lambda_j|; 2 \le j \le n\}$ 为 *谱隙*. 表示最大特征值 1 与其余特征值的最大模之间的距离. 近年的很多研究是关于保证谱隙为正的充分条件. 前面的例 8 中，另一个特征值为 -1, 所以谱隙是 0.

对于有限状态马氏链, 只要谱隙是正的, 任何初始分布都将收敛到不变分布, 而且收敛速度是指数阶的. 这几乎是人们能期望的最好结论. 在很长一段时间，人们热衷于讨论指数阶的大小, 尤其是随着状态数目的增长而增长的规律. 然而上个世纪八十年代 Persi Diaconis 发现 $||\pi - \nu\mathbf{P}^k||$ 的典型曲线有门阈现象 (threshold), 与指数曲线很不一样. 他用通俗的洗牌例子来说明. 一副牌有 54 张，摞成一叠可设想为 1 至 54 的排列，所以取状态空间为 1 至 54 的所有排列，洗牌就是从一种排列变为另一种排列，不变分布就是所有 1 至 54 的排列的均匀分布. 在此情形 Diaconis 发现只要洗 7 次效果就很好了. 再多洗效果改进不大. 这在实践中很有价值，尤其在随机算法的收敛方面.

3　常返与非常返

如果矩阵 $A = (a_{ij})$ 满足以下两条性质:

1. $a_{ij} \ge 0$,
2. $\sum_j a_{ij} = 1$,

则称 A 为 *随机矩阵*. 显然转移矩阵 $\mathbf{P} = (p_{ij})$ 是随机矩阵. 我们的问题是，任何随机矩阵也都是某个马氏链的转移矩阵吗？答案是肯定的. 如同随机游动一样，我们可以取 S^∞ 为样本空间 Ω. 每个样本点 ω 是一序列. 再选取 S 上的合适概率分布 μ 为初始分布，用 μ 和 A 来确定一族相容的有限维分布. 根据 Kolmogorov 的构造理论，这个随机过程就唯一确定了.

定义. 如果随机变量 τ 只取非负整数或 ∞, 而且 $\tau(\omega) \le n$ 是否成立可由 $X_0(\omega), X_1(\omega), X_2(\omega), \cdots, X_n(\omega)$ 的取值来判断，则称 τ 为 (关于马氏链 $\{X_n\}$ 的)停时. 有时为了强调

$P(\tau<\infty)=1$, 我们称 τ 为有限停时；如果存在 $M<\infty$, $P(\tau<M)=1$, 则称 τ 为有界停时；

例如第一章的随机游动首中时 τ_A 就是停时. 常数也是停时.

定理 (强马尔可夫性质). 设 τ 是关于马氏链 $\{X_n\}$ 的停时, 对任何 $x_1,\cdots,x_r,\ x,y\in S$,

$$P(X_{\tau+1}=y|X_\tau=x,X_{\tau-1}=x_1,\cdots X_{\tau-r}=x_r)=P(X_{\tau+1}=y|X_\tau=x).$$

证明 记 $B_k=\{\tau=k,X_k=x,X_{k-1}=x_1,\cdots X_{k-r}=x_r\}$. 当 $k<r$ 时， B_k 为空集. 当 $k\geq r$ 时，利用马氏性， $P(X_{k+1}=y|B_k)=p(x,y)$.

$$\begin{aligned}
& P(X_{\tau+1}=y|X_\tau=x,X_{\tau-1}=x_1,\cdots X_{\tau-r}=x_r)\\
= & \frac{P(X_{\tau+1}=y,X_\tau=x,X_{\tau-1}=x_1,\cdots X_{\tau-r}=x_r)}{P(X_\tau=x,X_{\tau-1}=x_1,\cdots X_{\tau-r}=x_r)}\\
= & \frac{\sum_{k=0}^{\infty}P(\tau=k,X_{\tau+1}=y,X_\tau=x,X_{\tau-1}=x_1,\cdots X_{\tau-r}=x_r)}{\sum_{k=0}^{\infty}P(\tau=k,X_\tau=x,X_{\tau-1}=x_1,\cdots X_{\tau-r}=x_r)}\\
= & \frac{\sum_{k=0}^{\infty}P(X_{k+1}=y,B_k)}{\sum_{k=0}^{\infty}P(B_k)}\\
= & \frac{\sum_{k=r}^{\infty}P(X_{k+1}=y|B_k)P(B_k)}{\sum_{k=r}^{\infty}P(B_k)}\\
= & \frac{\sum_{k=r}^{\infty}p(x,y)P(B_k)}{\sum_{k=r}^{\infty}P(B_k)}=p(x,y).
\end{aligned}$$

同理可证 $P(X_{\tau+1}=y|X_\tau=x)=p(x,y)$. 证毕.

从字面上我们也能得出 强马氏性 蕴含 马氏性. 在一般场合，反命题是不成立的. 但对于可数状态空间、离散时间参数的时齐马氏链，两者是一致的.

定义. 称 $G(x,y)=\sum_{n=0}^{\infty}p_{xy}(n)$ 为 格林函数. 若 $G(x,x)=\infty$, 则称状态 x 为 常返态; 否则就称为 暂态, 也叫做 非常返态. 若所有的状态同为常返或同为非常返，我们直接称马氏链是 常返的 或 非常返的.

注意定义只与转移阵 $\mathbf{P}$ 有关，与初始分布无关. 因此在许多场合我们常常忽略初始分布. 如需强调，则在下标中注明，如 P_μ, $P_x(\ \cdot\ )=P(\ \cdot\ |X_0=x)$. 令

$$\begin{aligned}
\sigma_x &= \min\{n\geq 1;X_n=x\};\\
\rho_{xy} &= P(\sigma_y<\infty|X_0=x).
\end{aligned}$$

注意 σ_x 与前面定义的首中时稍有不同，但也是停时. ρ_{xy} 表示从状态 x 出发的马氏链能在有限时间到达状态 y 的概率.

命题 4. $G(x,x)=1/(1-\rho_{xx})$. 等式两边同时收敛或同时发散.

证明 定义 $\sigma_1=\min\{n\ge 1;X_n=x\}$, 即前面的 σ_x. 归纳定义 $\sigma_k=\min\{n>\sigma_{k-1};X_n=x\}$, 其概率意义是马氏链第 k 次到达状态 x 的时刻.

$$
\begin{aligned}
&P_x(\sigma_2<\infty)=\sum_{n=2}^{\infty}P_x(\sigma_2=n)=\sum_{n=2}^{\infty}\sum_{m=1}^{n-1}P_x(\sigma_1=m,\sigma_2=n)\\
=\;&\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}P_x(\sigma_1=m,\sigma_2=m+n)\\
=\;&\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{P(X_0=X_m=X_{m+n}=x,X_j\ne x,1\le j<m \text{ 或 } m<j<n+m)}{P(X_0=x)}\\
=\;&\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{P(X_0=X_m=x,X_i\ne x,1\le i<m)}{P(X_0=x)}\\
&\times P(X_{m+n}=x,X_j\ne x,m<j<n+m|X_0=X_m=x,X_i\ne x,1\le i<m)\\
=\;&\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}P_x(X_m=x,X_i\ne x,1\le i<m)\\
&\times P(X_{m+n}=x,X_j\ne x,m<j<n+m|X_m=x)\\
=\;&\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}P_x(X_m=x,X_i\ne x,1\le i<m)P(X_n=x,X_j\ne x,0<j<n|X_0=x)\\
=\;&\left[\sum_{m=1}^{\infty}P_x(X_m=x,X_i\ne x,1\le i\le m-1)\right]^2=\rho_{xx}^2.
\end{aligned}
$$

归纳可证 $P_x(\sigma_k<\infty)=\rho_{xx}^k$.

再令 $N(\omega)=\sum_{k=1}^{\infty}1_{\{X_k(\omega)=x\}}$, 表示沿着序列 $\{X_n(\omega)\}$(也即样本点) 到过状态 x 的总次数. $N(\omega)=n$ 当且仅当 $\sigma_n<\infty$ 并且 $\sigma_{n+1}=\infty$. 于是 $P_x(N\ge k)=\rho_{xx}^k$.

$$
\sum_{k=1}^{\infty}\rho_{xx}^k=\sum_{k=1}^{\infty}P_x(N\ge k)=E_xN(\omega)=\sum_{k=1}^{\infty}P_x(X_k(\omega)=x)=\sum_{k=1}^{\infty}p_{xx}(n).
$$

两边加 1 即得结论.

推论 1. 若 $\rho_{xx}=1$, 则 $P_x(\sigma_k<\infty)=1$, 而且几乎处处 $N(\omega)=\infty$.

推论 2. 一维或二维的简单随机游动是常返的，而三维或更高维简单随机游动是非常返的.

据说上个世纪二十年代 Polya 在苏黎世大学周围街区随意漫步时总是一而再、再而三地遇到一对恋人. 为了说明自己并不在盯梢人家， Polya 把自己的随意漫步看成是二维格点上的简单随机游动，由于二维的简单随机游动是常返的，两个独立的简单随机游动一定会相遇无限次！日本学者 Kakutani 在 UCLA 演讲时给推论 2 一个绝妙的注解： A drunk man can find way home, a drunk bird can not.

4 可逆过程

定义. 如果存在 S 上的非负函数 $a: S \to R$, 对任何 $x, y \in S$,

$$a(x)p(x,y) = a(y)p(y,x)$$

均成立, 则称 a 为 配称分布, P 是 可配称的. 如果 a 是 S 上的概率分布, 则称 π 为 可逆分布, 称马氏链 P 为 可逆的 或 可逆过程 .

显然可逆分布是配称分布, 反之不然. 例如, d- 维简单随机游动是可配称的, 但不是可逆的. 但如果 $\sum_{x\in S} a(x) < \infty$, 经过归一化, $a(x)/\sum_{y\in S} a(y)$ 就成了可逆分布. 特别是有限状态空间上的可配称马氏链式可逆的. 如果可逆马氏链式不可约, 则对任何 $x \in S$, $\pi(x) > 0$. 其时确定可逆分布是比较容易的. 可先选定一个顶点 $x \in S$, 设 $\pi_x = a$, $C_x = 1$. 反复利用 (4) 式, 逐次确定 $\pi_y = C_y a$. 最后利用归一性确定 $a = (\sum_{y\in S} C_y)^{-1}$.

命题 5. 可逆分布是不变分布.

例 3(**续**). Ehrenfest 模型. 状态空间 $S = \{-N, -N+1, \cdots, 0, 1, 2, N\}$,

$$p(i, i+1) = \frac{1}{2} - \frac{i}{2N}, \quad p(i, i-1) = \frac{1}{2} + \frac{i}{2N}.$$

由 $\pi_i(N-i) = \pi_{i+1}(N+i+1)$ 可得 $\pi_i = C_{2N}^{N+i}/2^{2N}$. 当 N 很大时,

$$\sum_{|i|\geq \epsilon N} \pi_i \leq 2\int_{\sqrt{\epsilon N}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

这很小, 可以认为小概率事件在短时间内不会发生, 在物理上解释不可逆现象.

假设 $\mathbf{P}$ 是转移阵. 取状态空间 S 的子集 S_1, 对 $x, y \in S_1$, $x \neq y$, 定义 $p_1(x,y) = p(x,y)$, $p_1(x,x) = p(x,x) + \sum_{y\notin S_1} p(x,y)$, 则 $\mathbf{P}_1 = \{p_1(x,y); x, y \in S_1\}$ 是 S_1 上的转移概率阵, 若 $\mathbf{P}$ 是可配称的, 则 P_1 也是. 若 $\mathbf{P}$ 是可逆的, 则 P_1 也是可逆, 此时可取 $\pi_1(y) = \pi(y)/\sum_{z\in S_1} \pi(z)$.

设 $G = (V, E)$ 是一个无向图, 其中 V 是顶点集, E 是边的集合. 视 E 为 $V \times V$ 的子集, 以边的两个端点 x, y 来表示边. 以 $x \sim y$ 表示 $(x,y) \in E$; 由于 G 是无向图, $(x,y) = (y,x)$.

对于 $(x,y) \in E$, 赋以一个正数 $w(x,y)$, 称为边 (x,y) 的权重. 还可以把上述概念稍微拓广一点. 如果 $(x,y) \notin E$, 定义 $w(x,y) = 0$. 下一步, 定义顶点 x 的权重 $w(x) = \sum_{y\sim x} w(x,y)$. 定义

$$p(x,y) = \frac{w(x,y)}{w(x)}.$$

这样 $\{p(x,y); x, y \in S\}$ 构成一个马氏链的转移阵, 而且是可配称的, $w(x)$ 就是配称分布. 如果

$$\sum_{x\in V} w(x) < \infty,$$

则相应的马氏链是可逆的. 特别是对于有限图上述条件总成立.

反之，对任何可逆马氏链，把每个状态看成是一个顶点，如果 $p(x,y)>0$, 我们认为顶点 x 与 y 之间有边相连, 并赋予其权重 $w(x,y)=\pi(x)p(x,y)$. 显然 $w(x,y)=w(y,x)$. 这样我们得到一个带权的 (无向) 图. 所以，可逆马氏链和带权的图之间存在一一对应.

对于一般图 $G=(V,E)$, 如无特别声明，我们可以认为每条边的权重为 1. 相应的马氏链称为图 G 上的 *简单随机游动*. 在 d- 维格点上，这与前一章里定义的简单随机游动时一致的. 正因如此这里才如此命名，但一般而言，这里定义的简单随机游动并不符合第一章中随机游动的定义，而是随机游动的合理推广.

Dirichlet Principle. 假设 $x\in A\subset S$, 定义

$$\sigma_A=\min\{n;X_n\notin A\};\quad \sigma_x=\min\{n\geq 1;X_n=x\}.$$

定义 $\rho(x)=0,\rho(z)=1,z\notin A$; 对任何 $y\in A,y\neq x$, 定义

$$\rho(y)=P(\sigma_A<\sigma_x|X_0=y).$$

显然 $\rho\in\mathcal{F}$ 而 $\mathcal{F}=\{h:S\to[0,1];h(x)=0,h(u)=1,u\in A\}$.

命题 6 (Dirichlet Principle). 假设 $x\in A\subset S$, π 是 P 的配称分布，则

$$2\pi(x)P(\tau_A<\sigma_x|X_0=x)=\Phi(\rho)=\inf_{f\in\mathcal{F}}\Phi(f), \tag{4}$$

其中

$$\Phi(f)=\sum_{u\in A}\sum_{v\in S}\pi(u)p(u,v)(f(u)-f(v))^2.$$

证明. 首先设 A 是有限集，这时 $\Phi(f)$ 只有有限项，欲使其达到极小， f 应满足：对任何 $y\in A$, $y\neq x$,

$$f(y)=\sum_u p(y,u)f(u). \tag{5}$$

另一方面，利用全概公式和马氏性，不难验证 $\rho(\cdot)$ 满足 (6) 式. 因此由解的唯一性， (5) 式中第二个等号成立. 又

$$\begin{aligned}
\Phi(\rho) &= \sum_{u\in A}\sum_{v\in S}\pi(u)p(u,v)([1-\rho(u)]-[1-\rho(v)])^2\\
&= \sum_{u\in A}\sum_{v\neq u}\pi(u)p(u,v)\left([1-\rho(u)]^2+[1-\rho(v)]^2-2[1-\rho(u)][1-\rho(v)]\right)\\
&= 2\sum_{u\in A}\pi(u)[1-\rho(u)]\left([1-\rho(u)]-\sum_{v\neq u}p(u,v)[1-\rho(v)]\right)\\
&= 2\pi(x)\left(1-\sum_{v\neq x}p(x,v)[1-\rho(v)]\right)
\end{aligned}$$

$$
\begin{aligned}
&= 2\pi(x)\sum_{v\neq x} p(x,v)\rho(v) \\
&= 2\pi(x)\sum_{v\neq x} p(x,v)P(\tau_A<\sigma_x|X_0=v) \\
&= 2\pi(x)P(\tau_A<\sigma_x|X_0=x)
\end{aligned}
$$

当 A 是无限时,可用一列单调上升的有限子集序列 $\{A_n\}$ 来逼近, $A_n\subset A_{n+1}, \cup_n A_n=A$. 已知 (5) 式对 A_n 成立，取极限过渡到一般情形. 详见 Liggett, Interacting Particle Systems 证毕.

例 9: 二维整数格点 Z^2 上的简单随机游动是可配称的，可取配称分布 $\pi(y)=1$. Z^2 中的顶点 y 可用两个分量 (y_1,y_2) 来表示， $|y|=\max(|y_1|,|y_2|)$ 表示最大模. 应用 Ditrichilet Principle, 取 x 为原点， $A=\{y;|y|\le n\}$, 取 $f(y)=a_{|y|}$, 其中 $a_0=0$, a_n= 1, $\{a_k;k=1,2,\cdots,n-1\}$ 是待定常数. 则

$$
2P(\tau_A<\sigma_x|X_0=x)\le\Phi(f)=\sum_{k=1}^n 4(2k-1)(a_k-a_{k-1})^2.
$$

取最佳常数 a_k 可得

$$
P(\tau_A<\sigma_x|X_0=x)\le\frac{2}{\sum_{k=1}^n 1/(2k-1)}.
$$

当 n 趋向 ∞ 时右端趋向 0. 回忆第一节中 ρ_{xx} 的定义，不难推出,

$$
\rho_{xx}=\lim_n P(\tau_A>\sigma_x|X_0=x)=1-\lim_n P(\tau_A<\sigma_x|X_0=x)=1.
$$

这提供了证明常返性的一个新方法. 这一想法已被深入挖掘. 如果把带权的图看成是电网络，每条边都是一根导线，边 (x,y) 的电阻为 $1/w(x,y)$, 即电导为 $w(x,y)$. 再设 x 点的电压为 0, 所有 A 中顶点的电压为 1, 记顶点 y 的电压为 $v(y)$, 则

$$
v(y)=\sum_u p(y,u)v(u).
$$

由线性方程组 (5) 的解的唯一性， $\rho(y)=v(y)$. 这样就把问题转化为电网络的问题. 进而可以借用一些物理方面的知识或直观. 例如，电网络里有电阻、电流和功的概念. 在物理中我们可以定义两点之间的有效电阻 $R(x,y)$, 或者 x 到 A 之间的有效电阻 $R(x,A)$. 当 S 为无穷时，取 S 的一列单调上升的有限子集序列 $\{B_n\}$, $B_n\subset B_{n+1}$, $\cup_n B_n=S$, 设 $x\in B_1$, 则 $R(x,B_n^c)$ 单调上升，其极限 $\lim_n R(x,B_n^c)$ 称为从 x 到无穷远的有效总电阻，记为 $R(x,\infty)$.

命题 7. 假设 S 是可列无穷，则 $\rho_{xx}=1$ 当且仅当 $R(x,\infty)=\infty$.

5 分支过程

设 $\{\xi_{ij}, i, j \geq 1\}$ 独立同分布，取值非负整数，设 $X_0 = 1$,

$$X_{n+1} = \sum_{j=1}^{X_n} \xi_{nj}.$$

则 $\{X_n; n \geq 0\}$ 构成一马氏链，称为分支过程或分枝过程. 其转移概率可由 ξ 的共同分布所确定

$$P(\xi = k) = p_k, \quad k = 0, 1, 2, \cdots.$$

后者的概率意义更明显，也更常用，在一般场合只需给出 ξ 的分布即可.

历史上，1874 年英国人 Francis Galton 和 Reverend H.W. Watson 用此模型来描述、研究某个家族中男丁的数目，故又称 Galton-Watson 过程，或 GW 过程. 这时 ξ 表示某个父亲的儿子总数. 当然这模型也可用于描述某动物种群的个体数量变化，或核反应中裂变的原子个数. 因为这一原因，第二次世界大战以后美苏等国都非常重视.

如果 $X_n = 0$, 则 $X_{n+k} = 0$ ， $k \geq 1$. 即，一旦过程处于状态 0 ，就停止不动了. 状态 0 称为吸收态. 令 $\tau = \min\{n, X_n = 0\}$. 人们首先关心 $P(\tau < \infty)$ 是否为 1. 在这一节我们记

$$\rho = P(\tau < \infty), \quad f(s) = \sum_k p_k s^k, \quad m = \sum k p_k.$$

命题 8. (1)$EX_n = m^n$.

(2) ρ 是方程 $s = f(s)$ 的最小非零解. $\rho < 1$ 当且仅当 $m > 1$.

证明. (1) $EX_1 = E\xi = m$. 由归纳法,

$$EX_{n+1} = E\sum_{j=1}^{X_n} \xi_{nj} = E\sum_{l=1}^{\infty}\sum_{j=1}^{l} 1_{\{X_n = l\}}\xi_{nj}$$

$$= E\sum_{l=1}^{\infty} 1_{\{X_n = l\}}\sum_{j=1}^{l}\xi_{nj} = \sum_l P(X_n = l)ml = mEX_n = m^{n+1}.$$

(2) 利用全概公式

$$\rho = \sum P(\tau < \infty, X_1 = k) = \sum p_k P(\tau < \infty | X_0 = k) = \sum p_k \rho^k.$$

显然 1 是方程 $s = f(s)$ 的非零解. 注意到 $f(s)$ 在 [0,1] 区间内关于 s 单调上升，下凸，$f'(1) = m$. 当 $m < 1$ 时，曲线 $y = f(x)$ 与对角线 $y = x$ 在 [0,1] 区间内只在 $x = 1$ 相交. 当 $m = 1$ 时，曲线 $y = f(x)$ 与对角线 $y = x$ 在 [0,1] 区间内只在 $x = 1$ 相切. 当 $m > 1$, 曲线 $y = f(x)$ 与对角线 $y = x$ 在 [0,1) 区间内另有相交点. 证毕.

根据 $m<1$, $m=1$, $m>1$, 分别称分支过程是次 临界的, 临界的, 和 超临界的.

分支过程的每个样本对应一颗树. 设 Ω 为带根点的树全体. 每一个分支过程诱导了 Ω 上的概率分布, 称为 GW 分布, 记为 P_{GW}; 因为分支过程由一个非负整数上的概率分布 $\{p_k\}$ 所决定, 所以我们也说非负整数上的概率分布 $\{p_k\}$ 诱导了 Ω 上的概率分布. 以 GW 分布随机选取的树, 称为 GW 树.

设 Ω_0 为有限树全体, Ω_∞ 为无限树全体, 则 $\Omega=\Omega_0\cup\Omega_\infty$. 次临界和临界的分支过程诱导的概率分布全集中在 Ω_0, $P_{GW}(\Omega_0)$=1, $P_{GW}(\Omega_\infty)$=0. 在超临界情形, $P_{GW}(\Omega_0)=\rho$, $P_{GW}(\Omega_\infty)=1-\rho$. 对于 $A\subset\Omega_\infty$, 定义

$$P_{GW}^\infty(A)=\frac{P_{GW}(A)}{P_{GW}(\Omega_\infty)}.$$

即 P_{GW}^∞ 为 P_{GW} 限制在 Ω_∞ 的条件概率, 在许多场合我们更关心 P_{GW}^∞.

给定树 T 并选定一顶点 o 为根点, 对其他顶点 x, 定义 $|x|$ 为 x 与 o 的距离. 对于边 $e=(e_0,e_1)$, 定义 $|e|=\min\{|e_0|,|e_1|\}$. 取边 e 的权重 $w(e)=\lambda^{-|e|}$. 这样得到一个带权的图. 如前所述, 我们也可认为这是一个电网络, 边 e 的电阻 $R(e)=\lambda^{|e|}$. 则 $R(o,\infty)$ 是 λ 的单调函数. 定义

$$br(T)=\sup\{\lambda, R(o,\infty)<\infty\}.$$

称 $br(T)$ 为树 T 的 branching number.

命题 9. $P_{GW}^\infty(br(T)=m)=1$.

证明. 仅证容易的部分, $br(T)\le m$. 根据权重 $w(e)$ 定义树上的马氏链. 若 $\lambda\ge m$, 则相应的马氏链式常返的. 由此推出 $R(o,\infty)$=∞, 我们的策略是引用 Dirichlet Principle. 取

$$a_0=0;\quad a_l=1,\quad l\ge n;\quad a_k=\frac{\sum_{s=1}^k(\frac{\lambda}{m})^s}{\sum_{s=1}^n(\frac{\lambda}{m})^s},\quad 1\le k\le n-1.$$

在树 T 上第 k 层的顶点数为 X_k, 取 $f(y)=a_{|y|}$, 则

$$\begin{aligned}
&E_{GW}\sum_u\sum_{v\in S}\pi(u)p(u,v)(f(u)-f(v))^2\\
=&\ 2E_{GW}\sum_{k=1}^n\lambda^{-k+1}X_k(a_k-a_{k-1})^2\\
=&\ 2\sum_{k=1}^n\lambda^{-k+1}m^k(a_k-a_{k-1})^2=\frac{2\lambda}{\sum_{k=1}^n(\frac{\lambda}{m})^k}.
\end{aligned}$$

当 $n\to\infty$ 上式右端趋于 0, 于是 $P_{GW}(\rho_{oo}=0)=1$. 也即几乎所有的 GW 树的电阻是无穷大. 证毕.

树的分解. 选定一点为根点，涂为绿色，以其为端点添加 ξ 条边，其中 ξ 为随机变量，$P(\xi=k)=p_k(1-\rho^k)/(1-\rho)$. 把这 ξ 条边的另一端点随机地涂为绿色或红色，涂为绿色的概率为 $1-\rho$, 涂为红色的概率为 ρ. 如果所有端点全被涂为红色，则此次涂色无效，重新来过. 在每个绿色端点，重复上面的做法. 对于红色端点，添加 η 条边，其中 η 为随机变量， $P(\eta=k)=p_k\rho^{k-1}$, 所添新边的另一端全为红色. 按此规则进行下去，得一红绿相间的无穷树，其中绿色顶点组成的子树，称为主干，是按下列分布产生的 GW 树.

$$\hat{p}_0=0, \quad \hat{p}_k=\sum_{r=0}^{\infty} p_{k+r}\rho^r(1-\rho)^{k-1}C_{k+r}^r, \quad \forall k\geq 1.$$

红色顶点组成许多有限树，称为 bush, 连在主干上. 这红绿相间的树，如果不分颜色，恰好就是按 $\{p_k\}$ 产生的无穷 GW 树，即按 P_{GW}^{∞} 随机选取的一颗树.

阅读文献

Olle Haggstrom, *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press, Cambridge, 2002.

Peter G. Doyle & J. Laurie Snell, *Random Walks and Electrical Networks* (Carus Mathematical Monographs, No 22) Mathematical Association of America, 1984. http://www-ee.technion.ac.il/ adam/FUN/RWEN.pdf

John G. Kemeny & J. Laurie Snell, *Finite Markov Chains*, Springer, New York, 1983

参考文献

Theodore E. Harris, *The Theory of Branching Processes*, Springer, Berlin, 1963 & Dover Publications, Inc., New York, 1989

习题：

1. 对于 $x,y\in S$, 若存在 $x_0,x_1,\cdots,x_r\in S$, $x_0=x$, $x_r=y$, 且对任意 $1\leq k\leq r$, $p(x_{k-1},x_k)>0$, 则记为 $x\to y$. 若 $x\to y$ 且 $y\to x$, 则记为 $x\leftrightarrow y$. 补充规定 $x\leftrightarrow x$. 称 $\leftrightarrow$ 为互通关系. 证明互通关系是等价关系.

2. 不可约马氏链只有一个互通等价类.

3. 若 $x\leftrightarrow y$, 则 x 和 y 同为常返或同为非常返，且周期相同.

4. 试求比不可约更弱一些的充分条件，保证有限状态马氏链的不变分布唯一.

5. 证明：在二维整数格点上从原点到无穷远的电阻为 ∞; 而在三维整数格点上，从原点到无穷远的电阻是有限的. 这里使用前面的约定，即每条边的权重为 1, 也即电阻为 1.

6. 每个顶点都有 d 条边的树称为 d- 规则树. 假定 $d\geq 3$. 试证： (1) d- 规则树的 branching number 为 d. (2) $d-$ 规则树上的简单随机游动是非常返的.

7. 对于给定的马氏链，证明其所有不变分布全体构成一个凸集.

8. 考虑 $\{0,1,2,3,\cdots\}$ 上的马氏链 $\{X_n\}$, 其转移概率为 $p(0,1)=1$;

$$p(x,x+1)=\frac{x^2+2x+1}{2x^2+2x+1}, \quad p(x,x-1)=\frac{x^2}{2x^2+2x+1} \quad x\geq 1;$$

若 $|x-y|\geq 2$, 则 $p(x,y)=0$. 证明该马氏链是非常返的并计算 $\rho_x=P(\tau<\infty|X_0=x)$, 其中 $\tau=\min\{n\geq 1, X_n=0\}$. 提示： $\sum_{k=1}^{\infty}1/k^2=\pi^2/6$.

9. 证明：任何一列取整数值的独立随机变量序列是马氏链. 追加什么条件可使之成为时齐马氏链？

10. 某数据通信系统由 n 个中继站组成，从上一站向下一站传送信号 0 或 1 时，接收正确率为 p. 如用 X_0 表示初始站发出的数字. X_k 表示经过 k 次传送后接收到的数字，$\{X_k\}$ 构成马氏链. 试证

$$P(X_0 = 1|X_n = 1) = \frac{\alpha + \alpha(p-q)^n}{1 + (\alpha - \beta)(p-q)^n},$$

其中 $\alpha = P(X_0 = 1)$, $\beta = 1 - \alpha$, $q = 1 - p$. 请说明上述条件概率的实际意义.

11. 设状态 y 是非常返的, 对任意 x, 恒有 $\sum_{n=0}^{\infty} p_{xy}(n) \leq \sum_{n=0}^{\infty} p_{yy}(n)$. 试证明之.

12. 证明以下任一命题与马氏性等价.

1) 对任何 $m, n \geq 1$;对任何 $x_1, \cdots, x_n, y \in S$,
$P(X_{n+m} = y|X_n = x_n, X_1 = x_1, \cdots X_{n-1} = x_{n-1}) = P(X_{m+n} = y|X_n = x_n)$.

2) 对任何 $n_1 < n_2 < \cdots < n_r$;对任何 $x_1, \cdots, x_r \in S$,
$$P(X_{n_1} = x_1, \cdots X_{n_r} = x_r) = P(X_1 = x_1)\prod_{i=2}^{r} P(X_{n_i} = x_i|X_{n_{i-1}} = x_{i-1}).$$

3) 对任何 $r, s \geq 1, n_1 < n_2 < \cdots < n_r < n < m_1 < m_2 < \cdots < m_s$,
对任何 $x_1, \cdots, x_r,\ x, y_1, y_2, \cdots, y_s \in S$,
$P(X_{m_1} = y_1, \cdots, X_{m_s} = y_s|X_{n_1} = x_1, \cdots X_{n_r} = x_r, X_n = x)$
$= P(X_{n_1} = x_1, \cdots X_{n_r} = x_r|X_n = x)$.

13. 证明： $\rho_{xy} > 0$ 当且仅当 $x \to y$, 也等价于 $G(x, y) > 0$.

14. 有限状态空间 S 上的马氏链总是常返的.