# 第 21 讲 离散时间分支过程

# 定义

## 灭绝概率

## 马尔可夫决策过程

考虑由同类生物 (或粒子) 构成的群体，其中的每个生物在寿终时以概率 $P(\xi = j) = p_j$ 分裂成 $j$ 个后代，且与其他生物的分裂情况独立. 其后代也按照相同的方式各自独立分裂自己的后代. 用 $X_n$ 表示第 $n$ 代生物的总数，称随机序列 $\{X_n\}$ 为 离散时间分支过程，也称为 Galton-Watson 分支过程.

现在用 $\xi_{nk}$ 表示第 $n$ 代的第 $k$ 个个体寿终时分裂成的后代数，则 $\{\xi_{nk}\}$ 是来自总体 $\xi$ 的随机变量. 对 $m = 1, 2, \cdots$，用 $X_0 = m$ 表示第 $0$ 代有 $m$ 个个体. 在条件 $X_0 = 1$ 下，有

$$
\begin{aligned}
X_1 &= \xi_{01}, \\
X_2 &= \sum_{k=1}^{X_1} \xi_{1k}, \\
&\cdots\cdots \\
X_n &= \sum_{k=1}^{X_{n-1}} \xi_{n-1,k}.
\end{aligned}
$$

用 $\mu = E\xi$ 表示总体 $\xi$ 的数学期望，这是一个个体分裂成的平均后代数. 因为 $X_{n-1}$ 是第 $n-1$ 代生物的个数，由 $\{\xi_{jk}|j \le n-2, k \ge 1\}$ 决定，所以与 $\{\xi_{n-1,k}|k \ge 1\}$ 独立，由瓦尔德定理我们有

$$\begin{aligned} EX_n &= E\xi_{n-1,k} EX_{n-1} = \mu EX_{n-1} = \mu^2 EX_{n-2} \\ &= \cdots = \mu^{n-1} EX_1 = \mu^n. \end{aligned}$$

当 $X_0 = 1$ 时，还可以计算出

$$\mathsf{Var}(X_n) = \begin{cases} \sigma^2 \mu^{n-1} \frac{\mu^n - 1}{\mu - 1}, & \mu \neq 1, \\ n\sigma^2, & \mu = 1. \end{cases}$$

因为已知 $X_{n-1} = i$ 后，$X_n$ 和 $(X_0, X_1, \cdots, X_{n-2})$ 独立，并且

$$P(X_n = j | X_{n-1} = i) = P(X_1 = j | X_0 = i), \ i, j \geq 0,$$

所以离散时间分支过程是马氏链.

## 定义

## 灭绝概率

## 马尔可夫决策过程

对于分支过程我们关心的是当 $X_0 = 1$ 时群体灭绝的概率 $p_0$. 由于群体灭绝的概率 $p_0$. 由于群体灭绝当且第一代每个个体的后代灭绝，所以有

$$P(\text{群体灭绝}|X_1 = j) = \rho_0^j,\ j = 0, 1, \cdots .$$

于是在条件 $X_0 = 1$ 下，容易计算

$$\begin{aligned} \rho_0 &= P(\text{群体灭绝}) \\ &= \sum_{j=0}^{\infty} P(\text{群体灭绝}|X_1 = j)p_j = \sum_{j=0}^{\infty} \rho_0^j p_j. \end{aligned} \tag{2.1}$$

定义

$$g(s) = \sum_{j=0}^{\infty} s^j p_j - s,$$

(2.1) 说明灭绝概率 $\rho_0$ 是 $g(s) = 0$ 的解.

**定理 2.1**

设 $p_0 > 0$，$p_0 + p_1 < 1$. 若 $X_0 = 1$，则

(1) $\rho_0$ 是 $g(s) = 0$ $(s \in [0,1])$ 的最小解;

(2) $\rho_0 = 1$ 的充分必要条件是 $\mu = E\xi \leq 1$.

## 证明.

(1) 我们只需证明如果 $\rho > 0$ 使得 $g(\rho) = 0$，则 $\rho \geq \rho_0$.
我们首先由归纳法证明对于 $n \geq 1$，$\rho \geq P(X_n = 0)$. 首先我们有

$$\rho = \sum_{j=0}^{\infty} p_j \rho^j = p_0 \rho^0 = P(X_1 = 0).$$

于是结论对 $n = 1$ 成立. 设 $\rho \geq P(X_{n-1} = 0)$，注意到 $X_0 = 1$，利用 $P(X_n = 0|X_1 = j) = [P(X_{n-1} = 0)]^j$，我们有

$$\rho = \sum_{j=0}^{\infty} p_j \rho^j \geq \sum_{j=0}^{\infty} p_j [P(X_{n-1} = 0)]^j = \sum_{j=0}^{\infty} P(X_1 = j) P(X_n = 0|X_1 = j) = P(X_n = 0).$$

这就得到 $\rho \geq P(X_n = 0)$. 对于 $n \geq 1$，$A_n = \{X_n = 0\}$ 是单调增加的，所以有

$$\rho_0 = P(\bigcup_{n=1}^{\infty} A_n) = \lim_{n \to \infty} P(A_n) \leq \rho.$$

这说明 $\rho_0$ 是最小解. 因为 $g(0) = p_0 > 0$，所以 $\rho_0 \in (0, 1]$.

## 证明.

(2) 函数 $g(s)$ 在 $(0,1]$ 中连续，且

$$g'(s)=\sum_{j=0}^{\infty} j s^{j-1} p_j - 1,\ g'(1)=\mu-1.$$

如果 $\mu \le 1$，对于 $s \in [0,1)$，我们有

$$g'(s) < g'(1) = \mu - 1 \le 0,$$

所以 $g(s)$ 是 $[0,1)$ 中的严格单调减函数. 由 $g(1)=0$ 知道 $\rho_0=1$ 是 $g(s)$ 在 $(0,1]$ 中的唯一零点，所以 $\rho_0=1$. 当 $\mu>1$ 时，我们证明 $\rho_0<1$. $g'(1)=\mu-1>0$ 说明 $g(s)$ 在 1 附近严格单调升. 又从

$$g''(s)=\sum_{j=2}^{\infty} j(j-1)s^{j-2}p_j > 0$$

知道 $g(s)$ 是严格的凸函数，于是再从 $g(0)=p_0>0$，$g(1)=0$ 知道 $g(s)=0$ 在开区间 $(0,1)$ 中有唯一解 $p_0$. □

**命题 2.2**

设 $\rho_0$ 是 $X_0 = 1$ 时分支过程 $\{X_n\}$ 的灭绝概率. 当 $p_0 > 0$，$p_0 + p_1 < 1$ 时，有

$$P(\lim_{n\to\infty} X_n = 0) = \rho_0,\ P(\lim_{n\to\infty} X_n = \infty) = 1 - \rho_0.$$

## 证明.

第一个等式就是灭绝概率的定义，自然成立. $0$ 是吸引状态. 对于 $i \geq 1$，由

$$p_{i0} = (P(\xi = 0))^i = p_0^i > 0,$$

我们知质点从 $i$ 出发以正概率不回到 $i$，说明 $i$ 不是常返的. 于是 $C = \{1, 2, \cdots, m\}$ 中的状态都是非常返的. 这说明马氏链最多访问 $C_m$ 有限次，最终离开 $C_m$. 定义 $W = \liminf_{n\to\infty} X_n$, 就有

$$P(W = 0) = \rho_0,$$

$$P(W \in C_m) = 0,$$

$$P(W \geq m) = 1 - P(W \in C_m) - P(W = 0) = 1 - \rho_0.$$

事件列 $B_m = \{W \geq m\}$ 单调减，用概率的连续性得到

$$P(W = \infty) = P(\bigcap_{m=1}^{\infty} \{W \geq m\}) = \lim_{m\to\infty} P(W \geq m) = 1 - \rho_0. \qquad \square$$

当生物的最初群体数 $X_0 = m$ 时，由于个体独立地分裂自己的后代，所以群体的平均增长速度为 $E[X_n|X_0 = m] = m\mu^n$，方差为

$$\mathsf{Var}(X_n|X_0 = m) = \begin{cases} m\sigma^2\mu^{n-1}\frac{\mu^n-1}{\mu-1}, & \mu \neq 1, \\ mn\sigma^2, & \mu = 1. \end{cases}$$

群体最终灭绝的概率为 $\rho_0^m$，群体数走向无穷的概率为 $1 - \rho_0^m$. 群体的初始数 $m$ 越大，最终灭绝的概率越小. 当 $\mu > 1$，方差 $\mathsf{Var}(X_n)$ 的指数增加说明每个个体的后代数发展得很快.

**例 2.3**

若 $p_0 = 1/2$, $p_1 = 1/4$, $p_2 = 1/4$, 确定 $\rho_0$.

证明.

因为 $\mu = E\xi \leq 1$, 所以 $\rho_0 = 1$. □

**例 2.4**

若 $p_0 = 1/4$, $p_1 = 1/4$, $p_2 = 1/2$, 确定 $\rho_0$.

证明.

我们知, $\rho_0$ 满足

$$\rho_0 = \frac{1}{4} + \frac{1}{4}\rho_0 + \frac{1}{2}\rho_0^2,$$

从而

$$2\rho_0^2 - 3\rho_0 + 1 = 0.$$

这个二次方程的最小正解为 $\pi_0 = 1/2$. □

考察一个过程，它在离散时间点上的观测是标号为 $1, \cdots, M$ 的 $M$ 个可能状态中的任意一个．在观测到过程的状态后必须选取一个动作，我们令 $A$ 表示所有可能动作的集合，并且假定它是有限集．
如果过程在时间 $n$ 处于状态 $i$，并且选取了动作 $a$，那么系统的下一个状态由转移概率 $P_{ij}(a)$ 确定．如果我们令 $X_n$ 表示过程在时间 $n$ 的状态，令 $a_n$ 表示在时间 $n$ 选取的动作，那么上面的描述就等价于

$$P\{X_{n+1} = j \mid X_0, a_0, X_1, a_1, \cdots, X_n = i, a_n = a\} = P_{ij}(a).$$

因此，转移概率只是当前状态和随后的动作的函数.

将选取动作的规则称为策略．我们将局限于这样一种策略：在任意时间策略规定的动作只依赖于当时过程的状态（而不依赖于任何以前的状态和动作）．然而，我们允许一个策略是“随机化”的，即可以按概率分布选取动作．换句话说，策略 $\beta$ 是一组数字的集合 $\beta = \{\beta_i(a), a \in A, i = 1, \cdots, M\}$，其含义是如果过程处于状态 $i$，则以概率 $\beta_i(a)$ 选取动作 $a$．当然，我们需要假定

$$0 \leqslant \beta_i(a) \leqslant 1, \quad \text{对于一切} i, a,$$
$$\sum_a \beta_i(a) = 1, \quad \text{对于一切} i.$$

在任意给定的策略 $\beta$ 下，状态序列 $\{X_n, n = 0, 1, \cdots\}$ 构成一个马尔可夫链，其转移概率 $P_{ij}(\beta)$ 给定为

$$P_{ij}(\beta) = P_\beta\{X_{n+1} = j \mid X_n = i\} = \sum_a P_{ij}(a)\beta_i(a),$$

其中最后的等式是通过以状态 $i$ 时所选取的动作为条件得到的．我们假设对于策略 $\beta$ 的每一个选取，得到的马尔可夫链 $\{X_n, n = 0, 1, \cdots\}$ 都是遍历的.

对于任意一个策略 $\beta$，令 $\pi_{ia}$ 表示在使用策略 $\beta$ 时，过程处于状态 $i$ 并且选取动作 $a$ 的极限（或稳态）概率，即

$$\pi_{ia} = \lim_{n\to\infty} P_\beta\{X_n = i, a_n = a\}.$$

向量 $\boldsymbol{\pi} = (\pi_{ia})$ 必须满足

$$\begin{aligned}
&\text{(i) 对于一切} i, a,\ \pi_{ia} \geqslant 0;\\
&\text{(ii)}\ \sum_i \sum_a \pi_{ia} = 1;\\
&\text{(iii) 对于一切} j,\ \sum_a \pi_{ja} = \sum_i \sum_a \pi_{ia} P_{ij}(a).
\end{aligned} \tag{4.33}$$

(i) 和 (ii) 是显然的，而 (iii) 是因为左边是处于状态 $j$ 的稳态概率，而右边是以前一步的状态与选取的动作为条件算得的同一个概率.

于是对于任意一个策略 $\beta$，存在一个满足 (i) $\sim$ (iii) 的向量 $\boldsymbol{\pi} = (\pi_{ia})$，$\pi_{ia}$ 等于使用策略 $\beta$ 时过程处于状态 $i$ 并且选取动作 $a$ 的稳态概率．反之亦然，也就是说，对于任意满足 (i) $\sim$ (iii) 的向量 $\boldsymbol{\pi} = (\pi_{ia})$，存在策略 $\beta$，使得如果使用了策略 $\beta$，那么过程处于状态 $i$ 并且选取动作 $a$ 的稳态概率等于 $\pi_{ia}$．为了验证最后的这个说法，假设 $\boldsymbol{\pi} = (\pi_{ia})$ 是一个满足 (i) $\sim$ (iii) 的向量．然后令策略 $\beta = \{\beta_i(a)\}$ 为

$$\beta_i(a) = P\{\text{策略}\beta \text{ 选取}a \mid \text{状态为}i\} = \frac{\pi_{ia}}{\sum_a \pi_{ia}}.$$

现在令 $P_{ia}$ 表示在使用策略 $\beta$ 时，过程处于状态 $i$ 并且选取动作 $a$ 的极限概率，我们需要证明 $P_{ia} = \pi_{ia}$．对此，首先注意 $\{P_{ia}, i = 1, \cdots, M, a \in A\}$ 是二维马尔可夫链 $\{(X_n, a_n), n \geqslant 0\}$ 的极限概率．因此它们是

$$
\begin{aligned}
&(\mathrm{i}') \ P_{ia} \geqslant 0, \\
&(\mathrm{ii}') \ \sum_i \sum_a P_{ia} = 1, \\
&(\mathrm{iii}') \ P_{ja} = \sum_i \sum_{a'} P_{ia'} P_{ij}(a') \beta_j(a)
\end{aligned}
$$

的唯一解，其中 (iii′) 成立是因为

$$
P\{X_{n+1} = j, a_{n+1} = a \mid X_n = i, a_n = a'\} = P_{ij}(a') \beta_j(a).
$$

由于

$$
\beta_j(a) = \frac{\pi_{ja}}{\sum_a \pi_{ja}},
$$

因此 $\{P_{ia}\}$ 是

$$P_{ia} \geqslant 0, \quad \sum_i \sum_a P_{ia} = 1, \quad P_{ja} = \sum_i \sum_{a'} P_{ia'} P_{ij}(a') \frac{\pi_{ja}}{\sum_a \pi_{ja}}$$

的唯一解．因此，为了证明 $P_{ia} = \pi_{ia}$，我们需要证明

$$\pi_{ia} \geqslant 0, \quad \sum_i \sum_a \pi_{ia} = 1, \quad \pi_{ja} = \sum_i \sum_{a'} \pi_{ia'} P_{ij}(a') \frac{\pi_{ja}}{\sum_a \pi_{ja}}.$$

前面的两个式子得自式 (4.33) 的 (i) 和 (ii)，而第三个式子等价于

$$\sum_a \pi_{ja} = \sum_i \sum_{a'} \pi_{ia'} P_{ij}(a'),$$

它得自式 (4.33) 的 (iii)．

于是我们已经证明了向量 $\boldsymbol{\pi} = (\pi_{ia})$ 满足式 (4.33) 的 (i) $\sim$ (iii)，当且仅当存在策略 $\beta$，使得在使用策略 $\beta$ 时，$\pi_{ia}$ 等于过程处于状态 $i$ 并且选取动作 $a$ 的稳态概率．事实上，这里的策略 $\beta$ 定义为 $\beta_i(a) = \pi_{ia} / \sum_a \pi_{ia}$．

上面的事实在确定最佳策略时十分重要．例如，假设只要处于状态 $i$ 并且选取动作 $a$ 就赚得某个报酬 $R(i, a)$．由于 $R(X_i, a_i)$ 表示在时间 $i$ 赚得的报酬，因此在策略 $\beta$ 下单位时间的平均报酬的期望可以表示为

$$\beta \text{ 下平均报酬的期望} = \lim_{n\to\infty} E_\beta \left[ \frac{\sum_{i=1}^{n} R(X_i, a_i)}{n} \right].$$

现在，如果令 $\pi_{ia}$ 表示处于状态 $i$ 并且选取动作 $a$ 的稳态概率，那么在时间 $n$ 的报酬的期望的极限等于

$$\lim_{n\to\infty} E[R(X_n, a_n)] = \sum_i \sum_a \pi_{ia} R(i,a),$$

由此推出

$$\beta \text{ 下平均报酬的期望} = \sum_i \sum_a \pi_{ia} R(i,a).$$

因此，确定最大化平均报酬的期望的策略问题就是求

$$\begin{aligned}
&\text{最大化}_{\boldsymbol{\pi}=(\pi_{ia})} \sum_i \sum_a \pi_{ia} R(i,a), \\
&\text{其中} \pi_{ia} \geqslant 0, \text{ 对于一切} i, a, \\
&\sum_i \sum_a \pi_{ia} = 1, \\
&\sum_a \pi_{ja} = \sum_i \sum_a \pi_{ia} P_{ij}(a), \quad \text{对于一切} j.
\end{aligned} \tag{3.1}$$

然而，上面的最大化问题是线性规划的一个特例，可以用称为单纯形法的标准线性规划算法求解．如果 $\boldsymbol{\pi}^* = (\pi_{ia}^*)$ 最大化了上述问题，那么最佳策略就是 $\beta^*$，其中

$$\beta_i^*(a) = \frac{\pi_{ia}^*}{\sum_a \pi_{ia}^*}.$$