マルコフ連鎖の集団遺伝学への応用

システム情報数理学
伊藤　未希

2009 年 2 月 12 日

1　イントロダクション

集団遺伝学とは, 集団内の特定の遺伝子 (アリル) の分布や統計量を調べる学問である. ランダムに変化する遺伝子に対して, マルコフ過程を用いることで, モデルを簡単化できる. ここでは, 連続時間のマルコフ過程の例であるユール過程をアリルを対象としたモデルへと発展させた S_n-値 Moran モデルを構築する. S_n-値 Moran モデルでは状態に順列を用いることで, 集団の家族構成も読み取ることができる.

2　連続時間マルコフ過程の例

定義 2.1 正数の列 $\{\lambda_i\}$ を考える. 純粋出生過程とは次の条件をみたすマルコフ過程である.

$P_{i,i+1} = \lambda_i h + o(h)$

$P_{ii} = 1 - \lambda_i h + o(h)$

$P_{ij} = 0 \qquad (i > j)$

$\lambda_i > 0, \quad i = 0, 1, 2, \ldots$

定義 2.2 出生死亡過程の推移確率 $P_{ij}(t)$ は次の条件をみたす.

$P_{i,i+1}(h) = \lambda_i h + o(h) \quad i \geq 0$

$P_{i,i-1}(h) = \mu_i h + o(h) \quad i \geq 1$

$P_{ii}(h) = 1 - (\lambda_i + \mu_i)h + o(h) \quad i \geq 0$

$P_{ij}(0) = \delta_{ij}$

$\mu_0 = 0, \quad \lambda_0 > 0, \quad \mu_i, \lambda_i > 0, \quad i = 0, 1, 2, \ldots$

ユール過程

集団の個体間に相互作用がないと仮定する.

$\beta h + o(h)$:各個体が長さ h の時間内に新しい個体を 1 個生む確率

N:時刻 0 での個体数

とすると,

$$\begin{aligned} P_{i,i+1} &= \binom{i+N}{1}[\beta h + o(h)][1 - \beta h + o(h)]^{i+N-1} \\ &= (i+N)\beta h + o(h) \end{aligned}$$

$P(X(t) = i) = e^{-\beta t}(1 - e^{-\beta t}) \quad (i \geq 0)$

$X(t)$:時刻 t における集団の大きさ

移住を伴うユール過程

$P(X(t) = i) = \binom{i+\theta-1}{i} e^{-\theta t}(1 - e^{-t})^i$

θ:出生率 1 に対して移民が来る比率

3　S_n-値ユール過程

以下のルールに従って集団内の様子を順列で記述する.

- 発生順に個体を番号付け
- 移民が来たら新しいサイクルを右隣に作成
- 子供の番号は親の番号の左隣に挿入

例 $p = (5\ 3\ 1)(4\ 2)$

一番古い家族のメンバーは 1,3,5. 3 は 1 の子供, 5 は 3 の子供.

定義 3.1 サイズ n の順列の集合を S_n, すべてのサイズの順列の集合を $S = \cup_{n=0}^{\infty} S_n$ とおく. 移住を伴うユール過程は, 次のような推移率 $l_{pq}, (p \neq q)$ をもつ $S = \cup_{n=0}^{\infty} S_n$ 上のマルコフ連鎖 $\{\Pi(t), t \geq 0; \Pi(0) = (0)\}$ として表される.

$p = (0) \in S_0$ のとき

$$l_{pq} = \begin{cases} \theta & q = (1), \\ -\theta & q = p, \\ 0 & \text{その他.} \end{cases}$$

$p \in S_{n-1}, n > 1$ で p が形 $p = c_1 \ldots c_i \ldots c_k$ をもつとき,

$$l_{pq} = \begin{cases} \theta & q = c_1 \cdots c_i \cdots c_k\ (n), \\ 1 & q = c_1 \cdots c_i^j \cdots c_k, \\ -(n-1+\theta) & q = p, \\ 0 & \text{その他.} \end{cases}$$

$c_i = (n_1 \ldots n_{l(i)})$, $c_i^j = (n_1 \ldots n_{j-1}\ n\ n_j \ldots n_{l(i)})$

状態の推移だけに注目するとジャンプチェーンとなる.

定理 3.2 $\{\Pi(t), t \geq 0; \Pi(0) = 0\}$ のジャンプチェーンである $S = \cup_{n=0}^{\infty} S_n$ 上のマルコフ連鎖 $\{\tilde{\Pi}_n, n = 0, 1, 2, \ldots\}$ は

$$\begin{aligned} &\tilde{\Pi}_1 = I_1 \\ &\tilde{\Pi}_{n+1} = i_n(\tilde{\Pi}_n)Y_{n+1} \qquad n = 2, 3, \ldots \end{aligned}$$

を生成する.

$\{Y_n\}$:長さ 2 または長さ 1 のサイクルにのみ値をとる独立確率変数列

I_n:S_n 上の恒等写像
$p = c_1c_2\cdots c_k$ ならば, $i_n : S_n \to S_{n+1}$ は包含写像 $i_n(p) = c_1c_2\cdots c_k(n+1)$
このとき $Y_1 = I_1$ とし, $n > 1$ については Y_n は次のような分布をもつ.

$$P[Y_n = (n\ k)] = \frac{1}{n-1+\theta} \quad (1 \le k < n)$$
$$P[Y_n = I_n] = \frac{\theta}{n-1+\theta}$$

4 S_n-値 Moran モデル

Moran モデル（Moran, 1958 ; Kelly, 1979）

連続時間のアリルのモデル.

- 比率 μ で死ぬ個体がランダムに 1 つ選ばれる.
- 残った $n-1$ 個体の中から再びランダムに 1 つ選ばれ新しい個体が生まれる.
- 子供は確率 $1-u$ で親と同型, 確率 u で突然変異.

定義 4.1 S_n から S_n への写像 h_k を

$$h_k(p) = \hat{n}(s_r^{-1}ps_r) \qquad (r = \max\{k, p_k\})$$

によって定義する.
$s_r = (n\ n-1\ \ \ldots r+1\ r)$, $s_r \in S_n$
$\hat{n}(p) = (n\ p_n)p$, $p(k) = p_k$

定義 4.2

$$X_m = h_{Z_m}(X_{m-1})Y_m \qquad (m = 1, 2, \ldots)$$

によって定義される, 離散時間の確率過程 $\{X_m, m = 0, 1, \ldots\}$ を S_n-値 Moran モデルという.
$P[Z_m = k] = n^{-1}$, $(k = 1, \ldots, n)$, $\{Z_m\}$:$i.i.d$
$P[Y_m = (n\ k)] = \frac{1}{n-1+\theta} \quad (k = 1, \ldots, n-1)$
$P[Y_m = (n)] = \frac{\theta}{n-1+\theta}$

定理 4.3 S_n-値 Moran モデル $\{X_m\}$ の定常分布は,

$$r_q = \frac{\theta^{|q|}}{\theta_{(n)}}, \qquad q \in S_n$$

となる.
$\theta_{(n)} = \theta(\theta+1)\ldots(\theta+n-1)$, $|q| : q$ のサイクル数

連続時間への拡張

S_n-値 Moran モデルは事象を「1 回の死亡の後, 1 回の出生または突然変異が起こる」としていた. これを連続時間へ拡張すると, 出生, 死亡, 突然変異が独立に起こるモデルとなる. この連続時間のモデルの集団内のアリルの数だけに注目し, 一般的な移住を伴う出生死亡過程と等しくなれば, 家系の構成も含む移住を伴う出生死亡過程が構築できる.

定義 4.4 $p \in S_n$, $k = 1, \ldots, n$ とする. このとき連続時間の S_n-値 Moran モデルの事象を次のような写像 $f_k^{(n)} : S_n \to S_{n-1}$, $g_k^{(n)} : S_n \to S_{n+1}$, $h^{(n)} : S_n \to S_{n+1}$ で定義する.

死亡 $f_k^{(n)} : p \mapsto \pi_n(h_k(p)) = \pi_n(\hat{n}(S_r^{-1}pS_r))$
$(r = \max\{k, p_k\}, \pi_n : S_n \to S_{n-1})$
出生 $g_k^{(n)} : p \mapsto p(n+1)(n+1\ k)$
突然変異 $h^{(n)} : p \mapsto p(n+1)$

$j \in S_n$ のとき,

$$P_j(t+\Delta t) = \sum_{i \in S_{n-1} \cup S_n \cup S_{n+1}} P_{ij}P_i(t)$$

を求めたい. しかし, $j = f_k^{(n+1)}(i)$ をみたす道の総数のみが未知であり, 求める必要がある. 具体的に $S_3 \to S_2$, $S_2 \to S_1$ について見てみると次のような予想が立てられる.

予想

1 回の死亡により, ある $j \in S_n$ になるような $i \in S_{n+1}$ の総数は, $(n+1)^2$ となる.

根拠

現在の集団の状態がある順列 $i \in S_{n+1}$ で表されているとする. いま 1 回だけ死亡が起きたとすると, 得られる順列は, 1〜$n+1$ のそれぞれが死んだ場合の合計 $n+1$ 通り. $n+1$ 個体からなる順列の総数は $(n+1)!$ であるので, すべての $i \in S_{n+1}$ から 1 個体死亡する事象の総数は $(n+1)\cdot(n+1)!$. また $i \in S_{n+1}$ と死亡する個体が決められれば, 死亡の後に得られる順列 $j \in S_n$ は一つに定まる. 1 回の死亡により, $(n+1)\cdot(n+1)!$ 個の順列がそれぞれの $j \in S_n$ に平等に推移することがいえれば, n 個体からなる順列の総数は $n!$ であるので,

$$\frac{(n+1)\cdot(n+1)!}{n!} = (n+1)^2$$

となり, 予想が正しいといえる.

参考文献

主な参考文献.

[1] P. Joyce and S. Tavaré : Cycles, permutations and the structure of the Yule process with immigration, Stochastic Processes and their Applications 25 (1987), 309–314.

[2] P. Joyce and S. Tavaré : Random permutation and neutral evolution models, Stochastic Processes and their Applications 36 (1990), 245–261.

[3] S. カーリン (佐藤健一・佐藤由身子訳) : 確率過程講義, 産業図書, 1974.