

7. 時間スケーリングの分離と有限マルコフ連鎖の収束

可換モデルと呼ばれる広いクラスの繁殖モデルにおいて、時間スケールを適切に取ることにより、集団サイズ N を無限に大きくした極限で合祖過程(Coalescent process)と呼ばれる連続時間有限マルコフ連鎖が導かれることを第3章で示した。これは合祖過程が頑健性のあるモデルであることを示している。これまでのモデルでは離散時間で1世代当りに合祖が起こる確率は $1/N$ のオーダーであるが、他方分集団間の速い移住、植物でみられる自殖など毎世代一定の割合あるいは $1/N$ より大きなオーダーで生じる現象も存在する。このように様々な時間スケールで生じる現象を含むマルコフ連鎖モデルを連続時間近似する Separation of time scale と呼ばれる問題が近年研究されてきた。最初の7.1節で有限マルコフ連鎖の収束定理に関する Möhle の定理(1989)について紹介する。7.2節以下では収束定理の応用として順に Nordborg and Krone(2002), Sampson(2006), Möhle(1998), Kaj, Krone and Lascoux(2001)の結果を紹介する。最後の7.6節では Möhle の定理を拡張した形で Möhle and Notohara(2016)に従ってマルコフ連鎖の収束定理の証明を与える。

7.1 マルコフ連鎖の収束定理

N 個の半数体生物集団と毎世代可換モデルによる繁殖を仮定する。この集団から n 個の遺伝子をランダムに取り出したとき、1世代前に j 個の親に由来する確率 $p_{n,j}$ は第3章の結果より

$$p_{n,j} = \begin{cases} 1 - \frac{n(n-1)}{2} \frac{1}{N} + O(N^{-2}) & \text{if } j = n \\ \frac{n(n-1)}{2} \frac{1}{N} + O(N^{-2}) & \text{if } j = n-1 \\ O(N^{-2}) & \text{その他} \end{cases},$$

$$\text{また } Q = (q_{k,j}) \quad q_{k,j} = \begin{cases} -\frac{k(k-1)}{2} & \text{if } k = j \\ \frac{k(k-1)}{2} & \text{if } j = k-1 \\ 0 & \text{その他} \end{cases} \quad (7.1)$$

1世代の推移確率を $P_N = (p_{k,j})_{1 \leq k,j \leq n}$ とすると、 I を単位行列として $P_N = I + \frac{1}{N}Q + O(N^{-2})$ と表される。 N 世代を単位時間と取り $N \rightarrow \infty$ の極限を取ると良く知られた行列の性質により $\lim_{N \rightarrow \infty} P_N^{[M]} = \lim_{N \rightarrow \infty} \left(I + \frac{Q}{N} + O(N^{-2}) \right)^{[M]} = \exp(tQ)$ となり、 Q を生成作用素とする連続時間マルコフ連鎖が得られる。この収束定理を少し一般化した次の補題が Möhle(1998)によって示

された。 $d \times d$ 行列 $A = (a_{k,j})$ について、ノルムを $\|A\| = \max_k \sum_j |a_{k,j}|$ と定義する。 $\|AB\| \leq \|A\| \|B\|$ が成り立ち、 A が確率行列のとき $\|A\| = 1$ が成り立つ。

補題 7. 1 (Möhle(1998))

実数 $t \geq 0, K \geq 0$ とし、正の数 $(c_N)_{N \in \mathbb{N}}$ (\mathbb{N} は自然数の集合) を $\lim_{N \rightarrow \infty} c_N = 0$ とする。 $d \times d$ 行

列 A, B について極限 $\lim_{m \rightarrow \infty} A^m = P$ が存在し、 $\|A\| = 1$ のとき、

$$\lim_{N \rightarrow \infty} \sup_{\|B\| \leq K} \|(A + c_N B)^{\lfloor t/c_N \rfloor} - (P + c_N B)^{\lfloor t/c_N \rfloor}\| = 0. \quad (7.2)$$

さらに行列の列 $(B_N)_{N \in \mathbb{N}}$ について、極限 $G = \lim_{N \rightarrow \infty} P B_N P$ が存在するならば

$$\lim_{N \rightarrow \infty} (A + c_N B_N)^{\lfloor t/c_N \rfloor} = P - I + e^{tG} = p e^{tG} = e^{tG} P \quad (t \geq 0). \quad (7.3)$$

この補題より次の定理が成り立つ

定理 7. 2

$X_N = (X_N(r))_{r \in \mathbb{N}_0}$ (\mathbb{N}_0 は 0 及び自然数の集合) を有限の状態空間 S 上の斉時有限マルコフ

連鎖、その推移確率を Π_N とし、以下の条件を満たすとす。

(1) $A = \lim_{N \rightarrow \infty} \Pi_N$ が存在する。また十分大きな全ての N に対して $\Pi_N \neq A$ 。

(2) $P = \lim_{m \rightarrow \infty} A^m$ が存在する。 $P = (p_{ij})_{i,j \in S}$ とする。

(3) $G = \lim_{N \rightarrow \infty} P B_N P$ が存在する。ただし、全ての自然数 N に対して $B_N = \frac{\Pi_N - A}{c_N}$, $c_N = \|\Pi_N - A\|$ 。

初期分布 $P_{X_N(0)}$ がある確率分布 μ に分布収束するならば、マルコフ連鎖 $X_N(\lfloor t/c_N \rfloor)_{t \geq 0}$ の有

限次元分布は初期分布が $X_0 = \mu$ 、推移確率が $\Pi(t) = P - I + e^{tG} = P e^{tG}$ の連続時間マルコフ連鎖 $(X_t)_{t \geq 0}$ に有限次元分布収束する。

補題 7. 1、定理 7. 2 の証明は 7. 6 節でより拡張した形で与える。

補題 7. 1 より $\lim_{N \rightarrow \infty} \Pi_N^{\lfloor t/c_N \rfloor} = \lim_{N \rightarrow \infty} (A + c_N B_N)^{\lfloor t/c_N \rfloor} = P - I + e^{tG} = \Pi(t), t > 0$ 。

$(X_t)_{t \geq 0}$ の生成作用素は

$$\lim_{t \downarrow 0} \frac{\Pi(t) - \Pi(0+)}{t} = \lim_{t \downarrow 0} \frac{P - I + e^{tG} - P}{t} = \lim_{t \downarrow 0} \frac{e^{tG} - I}{t} = G. \quad (7.4)$$

マルコフ連鎖 $(X_t)_{t \geq 0}$ の推移確率は $\Pi(t) = Pe^{tG}$ より $P_t(i, j) = \sum_{k \in S} p_{ik}(e^{tG})_{kj}$ で与えられるので $(X_t)_{t \geq 0}$ は時刻 $t=0+$ に状態 $i \in S$ から $k \in S$ へ確率 p_{ik} で瞬時にジャンプし、その後は行列 G で生成されたマルコフ連鎖として推移する。

これにより、マルコフ性を用いて有限次元分布収束が導かれる。任意の $0 < t_1 < \dots < t_n$ に対して

$$\begin{aligned} \lim_{N \rightarrow \infty} P(X_N([t_1/c_N]), X_N([t_2/c_N]), \dots, X_N([t_n/c_N])) \\ = P(X(t_1) = x_1, X(t_2) = x_2, \dots, X(t_n) = x_n), \quad x_i \in S \quad (i = 1, 2, \dots, n) \end{aligned}$$

マルコフ連鎖 $X_N([t/c_N])$, $X(t)$ のサンプルパスはポアソン過程のサンプルパスのように、右連続、左極限を持つ関数空間 $D_S[0, \infty)$ 上に表現される。

$$D_S[0, \infty) = \{\omega: [0, \infty) \rightarrow S, \omega(t) \text{ は右連続、左極限を持つ, } t \geq 0\}$$

有限次元分布収束より詳細な収束の概念が $D_S[0, \infty)$ 上でマルコフ過程 $X_N([t/c_N])$ が $X(t)$ に分布収束する、すなわち弱収束という概念である。有限次元分布収束しても、 $D_S[0, \infty)$ 上で分布収束（弱収束）しない例も存在する。例えば $X_N(t) = \delta(t - N)$ とすると、 $X_N(t)$ の

$$\text{分布関数は } F_N(t) = \begin{cases} 1 & (t \geq N) \\ 0 & (t < N) \end{cases} \text{ であり、} \lim_{N \rightarrow \infty} X_N(t) = X(t) \equiv 0 \text{ で、有限次元分布収束は成}$$

立するが、分布関数は $\lim_{N \rightarrow \infty} F_N(t) = F(t) \equiv 0$ となり、 $F(t)$ は分布関数ではない。これは無

限のかなたに確率測度が逃げてしまうことによる。弱収束を示すための条件として、このようなことが起こらないことを保証する緊密性(tight)という条件を満たす必要がある。詳しくは Billingsley(1968)および Ethier and Kurtz(1986)を参照されたい。この章で扱う有限状態のマルコフ過程の弱収束については Ethier and Kurtz(1986)の Theorem 2.12 及び Kaj et al(2001)を参照されたい。

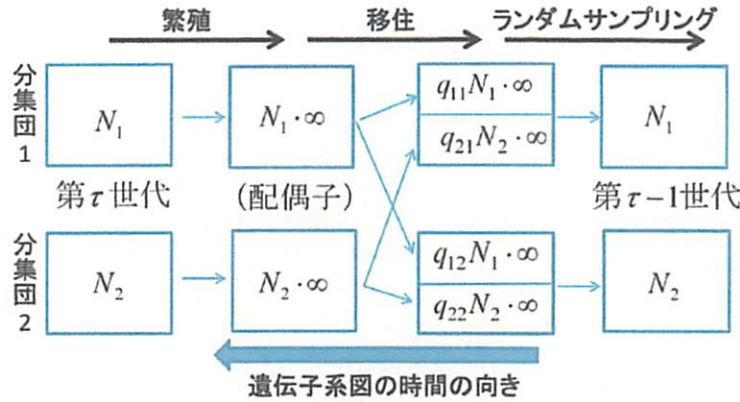
7. 2 速い移住率を持つ Structured coalescent モデル

第6章で紹介した地理的構造を含む Structured coalescent モデルにおいて、有限個の分集団に分かれ移住率が大きな速い移住(Fast migration)と呼ばれる場合に収束定理を応用する。 M 個の分集団からなる半数体生物集団を考え、分集団の集合を $\{1, 2, \dots, M\}$ 、分集団 k のサイズを $N_k = a_k N$ とする。

7. 2. 1 モデル

(1) 前向き過程：毎世代以下の3ステップで次世代の集団をつくる。

- ① 毎世代、各集団の個体が無限個の配偶子（種子）をつくる。
- ② 配偶子（種子）は M 個の分集団間を互いに独立に移住する。分集団 k から j ($k \neq j$) へ移住する確率（移住率）を $q_{k,j}$ とする。
- ③ 移住後、Wright-Fisher モデル、すなわちランダムサンプリングによって次世代の元の分集団サイズの集団が作られる。



この操作によって次世代の集団が作られるが、この集団から n 個の遺伝子をサンプルしたときその遺伝子系図を求めるために、後ろ向き過程を導く。

(2) 後ろ向き過程：

後ろ向き移住率 m_{jk} を分集団 j から任意に選んだ個体が 1 世代前に分集団 k から移住してきた確率とすると、

$$m_{jk} = \frac{a_k q_{kj}}{\sum_{i=1}^M a_i q_{ij}}$$

また状態空間を $E_i = \left\{ r = (r_1, r_2, \dots, r_M) \in \mathbb{N}^M : 1 \leq |r| = \sum_{j=1}^M r_j = i \right\}$ 、 $E = \bigcup_{i=1}^n E_i$ 、(\mathbb{N} は自然数の集合) とする。

全集団から計 n 個の遺伝子をサンプルし、その τ 世代前の祖先遺伝子の空間配置を $(Y_N(\tau))_{\tau \in \mathbb{N}_0} = (Y_{N,1}(\tau), \dots, Y_{N,M}(\tau)) \in E$ で表す。 $(Y_N(\tau))_{\tau \in \mathbb{N}_0}$ は状態空間 E 上のマルコフ連鎖で $Y_{N,i}(\tau)$ は τ 世代前に分集団 i で見出されるサンプルの祖先遺伝子の数である。

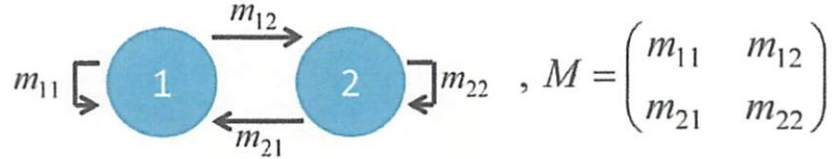
状態空間 E_i は $|E_i| = {}_M H_i = \binom{M+i-1}{i}$ 個の要素、 E は $|E| = \binom{n+M}{n} - 1$ 個の要素からなる。

7. 2. 2 サンプル数 $n=2$ 、分集団数 $M=2$ の場合

状態空間は $E_1 = \{(1,0), (0,1)\}$ 、 $E_2 = \{(2,0), (1,1), (0,2)\}$ 、 $E = E_1 \cup E_2$ となる。

サイズが $a_1 N$ と $a_2 N$ の二つの分集団間で下図のような後ろ向き移住があり、その定常分布

を $\gamma = (\gamma_1, \gamma_2)$ とする。方程式 $\gamma M = \gamma$ より $\gamma_1 = \frac{m_{12}}{m_{12} + m_{21}}, \gamma_2 = \frac{m_{21}}{m_{12} + m_{21}}$ 。



状態空間 E の中に順序を $(1,0), (0,1), (2,0), (1,1), (0,2)$ と付ける。この順序で移住による推移確率行列 A を表示し、その極限 $P = \lim_{m \rightarrow \infty} A^m$ を求めると

$$A = \begin{pmatrix} m_{11} & m_{12} & 0 & 0 & 0 \\ m_{21} & m_{22} & 0 & 0 & 0 \\ 0 & 0 & m_{11}^2 & 2m_{11}m_{12} & m_{12}^2 \\ 0 & 0 & m_{11}m_{21} & m_{12}m_{21} + m_{11}m_{22} & m_{22}m_{12} \\ 0 & 0 & m_{21}^2 & 2m_{21}m_{22} & m_{22}^2 \end{pmatrix}, \quad P = \lim_{m \rightarrow \infty} A^m = \begin{pmatrix} \gamma_1 & \gamma_2 & 0 & 0 & 0 \\ \gamma_1 & \gamma_2 & 0 & 0 & 0 \\ 0 & 0 & \gamma_1^2 & 2\gamma_1\gamma_2 & \gamma_2^2 \\ 0 & 0 & \gamma_1^2 & 2\gamma_1\gamma_2 & \gamma_2^2 \\ 0 & 0 & \gamma_1^2 & 2\gamma_1\gamma_2 & \gamma_2^2 \end{pmatrix}.$$

合祖(coalesce)による推移確率行列 B を表示すると

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1/(a_1N) & 0 & 1-1/(a_1N) & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1/(a_2N) & 0 & 0 & 1-1/(a_2N) \end{pmatrix} = I + \frac{1}{N}C, \quad C = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1/a_1 & 0 & -1/a_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1/a_2 & 0 & 0 & -1/a_2 \end{pmatrix}$$

となる。1世代の推移確率行列を Π_N とすると $\Pi_N = AB = A(I + C/N) = A + \frac{1}{N}AC$

補題 7. 1 を適用すると $c_N = 1/N$, $PA = P$ に注意して

$$G = P(AC)P = PCP = \left(\frac{\gamma_1^2}{a_1} + \frac{\gamma_2^2}{a_2} \right) \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \gamma_1 & \gamma_2 & -\gamma_1^2 & -2\gamma_1\gamma_2 & -\gamma_2^2 \\ \gamma_1 & \gamma_2 & -\gamma_1^2 & -2\gamma_1\gamma_2 & -\gamma_2^2 \\ \gamma_1 & \gamma_2 & -\gamma_1^2 & -2\gamma_1\gamma_2 & -\gamma_2^2 \end{pmatrix}$$

定理 7. 2 より、 $N \rightarrow \infty$ において次の推移確率 P_t の連続時間マルコフ連鎖に収束する。

$$P_t = P \exp(tG) = \begin{pmatrix} \gamma_1 & \gamma_2 & 0 & 0 & 0 \\ \gamma_1 & \gamma_2 & 0 & 0 & 0 \\ \gamma_1(1-e^{-\lambda t}) & \gamma_2(1-e^{-\lambda t}) & \gamma_1^2 e^{-\lambda t} & 2\gamma_1\gamma_2 e^{-\lambda t} & \gamma_2^2 e^{-\lambda t} \\ \gamma_1(1-e^{-\lambda t}) & \gamma_2(1-e^{-\lambda t}) & \gamma_1^2 e^{-\lambda t} & 2\gamma_1\gamma_2 e^{-\lambda t} & \gamma_2^2 e^{-\lambda t} \\ \gamma_1(1-e^{-\lambda t}) & \gamma_2(1-e^{-\lambda t}) & \gamma_1^2 e^{-\lambda t} & 2\gamma_1\gamma_2 e^{-\lambda t} & \gamma_2^2 e^{-\lambda t} \end{pmatrix}, \quad \lambda = \frac{\gamma_1^2}{a_1} + \frac{\gamma_2^2}{a_2}.$$

この推移確率より、サンプル遺伝子が1個の場合は瞬時に移住の定常分布 γ に到達し、サンプル遺伝子が2個の場合はまず瞬時に定常分布 $(\gamma_1^2, 2\gamma_1\gamma_2, \gamma_2^2)$ に達し、その後は初期分布に関係なく、率 λ で合祖が生じる Kingman の合祖過程として振る舞う。第6章で扱った地理的構造を持つモデルでは移住率を $O(N^{-1})$ で与え、連続時間近似の後に、移住率を限りなく大きくするという形で強移住極限(strong-migration limit)を考えた。この節のモデルは合祖が N^{-1} のオーダーの確率で起こるのに対して、毎世代 $O(1)$ で起こる速い移住率を持つモデルを考えた。このとき合祖が起こる前に多くの移住が起こり遺伝子の空間配置について移住による定常状態に到達する。その後合祖が起こり、最初の定常分布へのジャンプの後

は全体として1集団のように振る舞うことが示された。 $\lambda^{-1} = \left(\frac{\gamma_1^2}{a_1} + \frac{\gamma_2^2}{a_2} \right)^{-1}$ は合祖有効集団サイズ (coalescent effective size) と呼ばれる。第6章、定理6.5の N_e に対応するもので

ある。同様の議論 (調和平均 \leq 相加平均) により、 $\lambda^{-1} = \left(\frac{\gamma_1^2}{a_1} + \frac{\gamma_2^2}{a_2} \right)^{-1} \leq a_1 + a_2$ が成り立つ。

一般の M 個の分集団の場合は、後ろ向き移住率 $M = (m_{k,j})_{1 \leq k, j \leq M}$ の定常分布を

$\gamma = (\gamma_1, \dots, \gamma_M)$ とする、すなわち $\gamma M = \gamma$ 。遺伝子サンプルを $\alpha = (\alpha_1, \dots, \alpha_M)$, $|\alpha| = n$ とすると n 個の遺伝子の移住の下での定常分布は多項分布 $Mult(n; \gamma)$ に従う。すなわち、 $\beta = (\beta_i) \in E_n$

に対して $P(\beta) = \frac{n!}{\prod_{i=1}^M \beta_i!} \prod_{j=1}^M \gamma_j^{\beta_j}$ で与えられる。瞬時に n 個の遺伝子は移住による定常分布

$Mult(n; \gamma)$ に到達し、その後は祖先遺伝子数が k 個である状態の滞在時間を τ_n としたとき、

指数分布 $P(\tau_n > t) = \exp(-\lambda_n t)$, $\lambda_n = \frac{n(n-1)}{2\lambda^{-1}}$ に従う標準合祖過程として振る舞う。

強移住極限について、より広いクラスの研究として Hoesjjer(2011)の論文もある。次の節では、分集団サイズの変化的変化を伴うモデルへの応用を紹介する。

7. 3 分集団サイズの変動的を含む Structured coalescent モデル

有限個の分集団からなる地理的構造を含む遺伝子系図モデルでさらに分集団サイズの変動的を含むモデルについて離散マルコフ連鎖から連続時間マルコフ連鎖への収束を Sampson(2006)の研究に従って紹介する。ただし、移住率は N^{-1} のオーダーで起こる遅い移住率を仮定する。前節と同様に M 個の分集団からなる半数体生物集団を考える。各分集団のサイズを $N_k = a_k N$, ($1 \leq k \leq M$) とする。第 τ 世代の分集団サイズを $N(\tau) = (N_1(\tau), \dots, N_M(\tau))$ 、 $N(\tau)$ は状態空間 S_N の上を毎世代確率的に変動する定常な非周期既約マルコフ連鎖と仮定する。

$$|S_N| = s, S_N = \{\bar{a}_i N = (a_{i1}, a_{i2}, \dots, a_{iM}) N; i \in \{1, 2, \dots, s\}\},$$

状態 i での分集団 k のサイズを $a_{ik} N$ とする。分集団サイズのマルコフ連鎖 $N(\tau)$ の定常分布

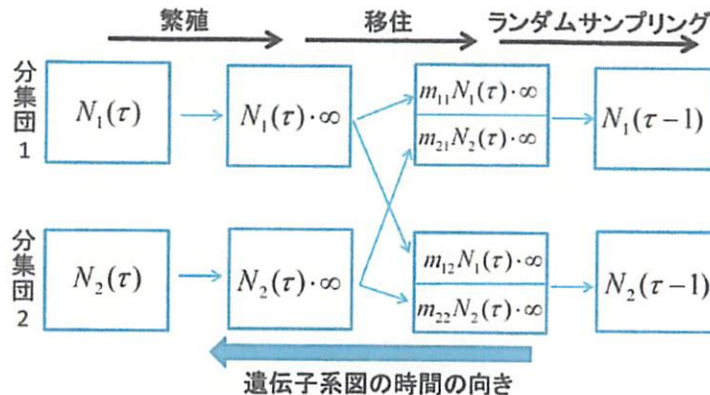
$$\text{を } \gamma = (\gamma_1, \dots, \gamma_s), \text{ 推移確率を } \tilde{T} = (\tilde{t}_{i,j}) \text{ とすると } \gamma \tilde{T} = \gamma.$$

(1) 前向き過程：毎世代以下の三つのステップで次世代の集団をつくる。

- ① 毎世代、各集団の個体が無限個の配偶子（種子）をつくる。
- ② 配偶子（種子）は M 個の分集団間を互いに独立に移住する。分集団 k から j ($k \neq j$) へ移住する確率（移住率）を $m_{k,j}$ とする。

$$m_{k,j} = \frac{\mu_{k,j}}{N} + o\left(\frac{1}{N}\right), \quad m_{k,k} = 1 - \sum_{j(\neq k)} m_{k,j}$$

- ③ 移住後、Wright-Fisher モデル、すなわちランダムサンプリングによって次世代の集団が作られる。



この操作によって次世代の集団が作られるが、この集団から n 個の遺伝子をサンプルしたときその遺伝子系図を求めるために、後ろ向きの過程を導く。

(2) 後ろ向きの過程

時間を現在から過去に遡る方向を正の向きにとり。τ 世代前の分集団の個体数を

$N(\tau) = (N_1(\tau), \dots, N_M(\tau))$ とする。推移確率を $T = (t_{i,j})$ 、ただし

$t_{i,j} = P(N(\tau+1) = \bar{a}_j N | N(\tau) = \bar{a}_i N)$ とすると、分集団サイズは定常なマルコフ過程なので

$t_{ij} = \frac{\gamma_j \tilde{f}_{j,i}}{\gamma_i}$ で与えられる。また $E = \left\{ r = (r_1, r_2, \dots, r_M) \in \mathbb{N}^M : 1 \leq |r| = \sum_{i=1}^M r_i \leq n \right\}$ 、(N は自然数の集合) とする。全集団から計 n 個の遺伝子をサンプルし、その τ 世代前の祖先遺伝子の空間配置を $(Y_N(\tau))_{\tau \in \mathbb{N}_0} = (Y_{N,1}(\tau), \dots, Y_{N,M}(\tau)) \in E$ で表す。Y_{N,i}(τ) は τ 世代前に分集団 i で見出されるサンプルの祖先遺伝子の数である。状態空間 E は $|E| = \binom{n+M}{n} - 1$ 個の要素からなる。

後ろ向き移住率 f_{klj} を分集団サイズが $\bar{a}_j N$ という条件下で分集団 k から任意に選んだ個体が 1 世代前に分集団 l から移住してきた確率とする。

$$f_{klj} = \frac{m_{lk} a_{jl} N}{\sum_{i=1}^M m_{li} a_{ji} N} = \frac{m_{lk} a_{jl}}{\sum_{i=1}^M m_{li} a_{ji}} \quad \text{より}$$

$$k \neq l \text{ のとき } f_{klj} = \frac{a_{jl} \mu_{lk} / N + o(N^{-1})}{\sum_{i(\neq k)} a_{ji} \mu_{ik} / N + a_{jk} \left(1 - \sum_{i(\neq k)} \mu_{ki} / N \right) + o(N^{-1})} = \mu_{lk} \frac{a_{jl}}{a_{jk}} \frac{1}{N} + o(N^{-1})$$

$$k = l \text{ のとき } f_{kkj} = 1 - \sum_{i(\neq k)} \mu_{ik} \frac{a_{jl}}{a_{ki}} \frac{1}{N} + o(N^{-1})$$

(3) 連続時間マルコフ連鎖への収束

以上のモデルの下で、全集団からサンプルした遺伝子の系図過程を求めるために次の様な集団サイズと祖先遺伝子の空間配置からなる二変量マルコフ連鎖を定義する。

$X_N(\tau) = (N(\tau), Y_N(\tau))$ 、時間軸は過去を正の向きにとり。X_N(τ) は E_N = S_N × E 上のマルコフ連鎖であり、その推移確率を Π_N とする。

$$\Pi_N = (\pi_{(i,\alpha),(j,\beta)})_{(i,\alpha),(j,\beta) \in E_N} = (P(X_N(\tau+1) = (\bar{a}_j N, \beta) | X_N(\tau) = (\bar{a}_i N, \alpha)))$$

ここで $\bar{a}_i N, \bar{a}_j N \in S_N, \alpha, \beta \in E$

$\alpha = (\alpha_i; 1 \leq i \leq M) \in E$ のレベルを $|\alpha| = \sum_{i=1}^M \alpha_i$ とする。 E の要素をレベル 1 から n まで並べ同じレベル内では適当に順序を付ける。 S_N の要素も s 個なので、直積の状態空間 $E_N = S_N \times E$ にも前節と同様に順序を定義することができる。この順序に従って E_N の要素を並べる。この順序に従って成分表示を行い推移確率 Π_N の行列表示をすることができる。このとき、次の定理が成り立つ。

定理 7. 3

推移確率行列 Π_N は次の形に表すことができる。

$$\Pi_N = A + \frac{B}{N} + o(N^{-1}), \quad A = (a_{(i,\alpha),(j,\beta)}) = (t_{ij} \delta_{\alpha,\beta})_{(i,\alpha),(j,\beta) \in E_N}, \quad B = (b_{(i,\alpha),(j,\beta)})_{(i,\alpha),(j,\beta) \in E_N}$$

$$b_{(i,\alpha),(j,\beta)} = \begin{cases} -t_{ij} \sum_{k=1}^M \left(\frac{\alpha_k(\alpha_k-1)}{2a_{jk}} + \alpha_k \sum_{\ell(\neq k)} \frac{\mu_{\ell k} a_{j\ell}}{a_{jk}} \right) & \beta = \alpha \text{ のとき} \\ t_{ij} \alpha_k \frac{\mu_{\ell k} a_{j\ell}}{a_{jk}} & \beta = \alpha - \varepsilon^k + \varepsilon^\ell, (k \neq \ell) \text{ のとき} \\ t_{ij} \frac{\alpha_k(\alpha_k-1)}{2a_{jk}} & \beta = \alpha - \varepsilon^k \text{ のとき} \\ 0 & \text{その他} \end{cases}$$

(証明)

マルコフ連鎖 $X_N(\tau)$ の状態変化は集団サイズの変化及び合祖または移住によるものなので、

(1) 合祖のみによる推移：1回の合祖のみによって状態 (i, α) から $(j, \alpha - \varepsilon^k)$ へ推移する確率を求める。集団サイズ過程が状態 i から j へ確率 t_{ij} で変化し、サンプル α の遺伝子が

全く移住をしない確率が $\prod_{\ell=1}^M f_{\ell j}^{\alpha_\ell}$ 、そのときサンプル α は全てサイズ $a_{j\ell} N$

($\ell=1, 2, \dots, M$) の1世代前の集団に行き着く。その中で分集団 k から来た α_k 個の遺伝子の中

の2個が共通な親をもつ確率が $\binom{\alpha_k}{2} \frac{1}{a_{jk} N} \left(\prod_{\ell=1}^{\alpha_k-1} \left(1 - \frac{\ell-1}{a_{jk} N} \right) \right)$ となるので、求める確率は

$$\pi_{(i,\alpha),(j,\alpha-\varepsilon^k)} = t_{ij} \left(\prod_{\ell=1}^M f_{\ell j}^{\alpha_\ell} \right) \binom{\alpha_k}{2} \frac{1}{a_{jk} N} \left(\prod_{\ell=1}^{\alpha_k-1} \left(1 - \frac{\ell-1}{a_{jk} N} \right) \right)$$

$$\text{ここで } \prod_{\ell=1}^M f_{\ell j}^{\alpha_\ell} = \prod_{\ell=1}^M \left(1 - \frac{1}{N} \sum_{z(\neq \ell)} \frac{\mu_{z\ell} a_{jz}}{a_{j\ell}} + o(N^{-1}) \right)^{\alpha_\ell} = 1 - \frac{1}{N} \sum_{\ell=1}^M \alpha_\ell \mu_{z\ell} \frac{a_{jz}}{a_{j\ell}} + o(N^{-1})$$

$$\prod_{\ell=1}^{\alpha_k-1} \left(1 - \frac{\ell-1}{a_{jk} N} \right) = 1 - \binom{\alpha_k-1}{2} \frac{1}{a_{jk} N} + o(N^{-1})$$

以上より

$$\pi_{(i,\alpha),(j,\alpha-\varepsilon^t)} = t_y \binom{\alpha_k}{2} \frac{1}{a_{jk} N} + o(N^{-1}) \quad (7.10)$$

(2) 移住のみによる推移：サンプル中の1遺伝子のみが後ろ向き移住をする確率は

$$k \neq \ell \text{ のとき, } N\pi_{(i,\alpha),(j,\alpha-\varepsilon^t+\varepsilon^t)} = t_y \left(\prod_{z(\neq k)} f_{zz|j}^{\alpha_z} \right) \alpha_k f_{kk|j}^{\alpha_k-1} f_{k\ell|j} = t_y \alpha_k \mu_{\ell k} \left(\frac{a_{j\ell}}{a_{jk}} \right) \frac{1}{N} + o(N^{-1})$$

複数回の合祖あるいは移住が起こる確率は $o(N^{-1})$ 、また $\sum_{\beta \in E} \pi_{(i,\alpha),(j,\beta)} = t_y$ なので、

$$\pi_{(i,\alpha),(j,\alpha)} = t_y \left\{ 1 - \frac{1}{N} \sum_{k=1}^M \left(\binom{\alpha_k}{2} \frac{1}{a_{jk}} + \alpha_k \sum_{\ell(\neq k)} \mu_{\ell k} \left(\frac{a_{j\ell}}{a_{jk}} \right) \right) \right\} + o(N^{-1}) \quad (7.11)$$

証明終わり

Möhle の定理を用いて、 $N \rightarrow \infty$ の極限において次の定理が成り立つ。

補題 7. 4

$$\lim_{N \rightarrow \infty} \Pi_N^{[M]} = P - I + e^{G}, \quad \text{ここで } P = (p_{(i,\alpha),(j,\beta)}) = \lim_{m \rightarrow \infty} A^m = (\gamma_j \delta_{\alpha,\beta})$$

$$G = (g_{(i,\alpha),(j,\beta)}) = PBP = (\gamma_j q_{\alpha,\beta}), \quad \gamma = (\gamma_1, \dots, \gamma_s) \text{ は定常分布, } q_{\alpha,\beta} = \sum_i \gamma_i \sum_j b_{(i,\alpha),(j,\beta)}$$

(証明)

$$A = (a_{(i,\alpha),(j,\beta)}) = (t_y \delta_{\alpha,\beta})_{(i,\alpha),(j,\beta) \in E_N} \text{ より,}$$

$$\begin{aligned} (A^2)_{(i,\alpha),(j,\beta)} &= \left(\sum_{(\ell,\varepsilon)} a_{(i,\alpha),(\ell,\varepsilon)} a_{(\ell,\varepsilon),(j,\beta)} \right) = \left(\sum_{(\ell,\varepsilon)} t_{i\ell} \delta_{\alpha,\varepsilon} t_{\ell j} \delta_{\varepsilon,\beta} \right) = \left(\left(\sum_{\ell} t_{i\ell} t_{\ell j} \right) \delta_{\alpha,\beta} \right) \\ &= \left((T^2)_{i,j} \delta_{\alpha,\beta} \right) \end{aligned}$$

$$\text{これより } P = \lim_{m \rightarrow \infty} A^m = (\gamma_j \delta_{\alpha,\beta}). \quad \text{また } G = (g_{(i,\alpha),(j,\beta)}) = PBP \text{ より}$$

$$\begin{aligned} g_{(i,\alpha),(j,\beta)} &= \sum_{(s,\zeta)} \sum_{(\ell,\varepsilon)} p_{(i,\alpha),(\ell,\varepsilon)} b_{(\ell,\varepsilon),(s,\zeta)} p_{(s,\zeta),(j,\beta)} = \sum_{(s,\zeta)} \sum_{(\ell,\varepsilon)} \gamma_{\ell} \delta_{\alpha,\varepsilon} b_{(\ell,\varepsilon),(s,\zeta)} \gamma_j \delta_{\zeta,\beta} \\ &= \gamma_j \sum_{\ell} \gamma_{\ell} \left(\sum_s b_{(\ell,\alpha),(s,\beta)} \right) \end{aligned} \quad \text{証明終}$$

状態空間が有限なので、有限マルコフ連鎖の弱収束について Ethier and Kurtz(1986) Theorem2.12 及び Kaj et al (2001) (7. 5 節参照) の手法により、次の定理を得る。

定理 7. 5

$N \rightarrow \infty$ のとき、離散時間マルコフ連鎖 $(Y_N((Nt)))_{t \geq 0}$ の初期分布 $Y_N(0)$ が μ に分布収束するならば、 $(Y_N((Nt)))_{t \geq 0}$ は初期分布 $Y(0) \stackrel{d}{=} \mu$ 、生成作用素 $Q = (q_{\alpha, \beta})_{\alpha, \beta \in E}$ の連続時間マルコフ連鎖 $Y = (Y(t))$ に空間 $D_E[0, \infty)$ 上で弱収束 (分布収束) する。

ここで

$$q_{\alpha, \beta} = \begin{cases} -\sum_{k=1}^M \left(\frac{\alpha_k(\alpha_k - 1)}{2c_k} + \alpha_k \sum_{l \neq k} \frac{\beta_{kl}}{2} \right), & \beta = \alpha \text{ のとき} \\ \alpha_k \beta_{kl} / 2, & \beta = \alpha - \varepsilon^k + \varepsilon^l \text{ のとき} \\ \alpha_k(\alpha_k - 1) / (2c_k), & \beta = \alpha - \varepsilon^k \text{ のとき} \\ 0 & \text{その他} \end{cases} \quad (7.12)$$

$$\text{ただし } \alpha_k = \sum_{i=1}^s \frac{\gamma_i}{a_{ik}}, \quad \beta_{kl} = 2\mu_{kl} \sum_{i=1}^s \gamma_i \frac{a_{il}}{a_{ik}}.$$

(証明)

$\eta_N : E_N \rightarrow E$, $\eta_N(\bar{a}_i N, \alpha) = \alpha$ とするとき、 η_N は可測。

$B(E_N)$ は E_N 上のバナッハ空間で、ノルムを $\|f\| = \sup_{x \in E_N} |f(x)|$ で定義する。 $B(E)$ は E_N を E

で置き換えたバナッハ空間である。 $f \in B(E_N)$, $x \in E_N$ に対して $\mathfrak{I}_N f(x) = \sum_{y \in E_N} f(y) \pi_{x,y}$

と定義する。このとき

$$\mathfrak{I}_N^{[M]}(f \circ \eta_N)(\bar{a}_i N, \alpha) = \sum_{(\bar{a}_j N, \beta) \in E_N} (f \circ \eta_N)(\bar{a}_j N, \beta) (\Pi_N^{[M]})_{(i, \alpha), (j, \beta)} = \sum_{\beta \in E} \sum_{j=1}^s f(\beta) (\Pi_N^{[M]})_{(i, \alpha), (j, \beta)}$$

他方 $\mathfrak{I}(t)f(\alpha) = \sum_{\beta \in E} f(\beta) (e^{tQ})_{\alpha, \beta}$ とする。Ethier&Kurtz(1986)の Theorem2.12 及び Kaj et

al(2001)の方法 (7. 5 節参照) より $N \rightarrow \infty$ において以下の等式、極限が成り立つことを示せばよい。

$$\left| \mathfrak{I}_N^{[M]}(f \circ \eta_N)(\bar{a}_i N, \alpha) - \mathfrak{I}(t)f(\alpha) \right| = \left| \sum_{\beta \in E} f(\beta) \left\{ \sum_{j=1}^s (\Pi_N^{[M]})_{(i, \alpha), (j, \beta)} - (e^{tQ})_{\alpha, \beta} \right\} \right| \rightarrow 0 \quad (N \rightarrow \infty)$$

結局 E は有限集合なので、各項 $\left\{ \sum_{j=1}^s (\Pi_N^{[M]})_{(i,\alpha),(j,\beta)} - (e^{Q^0})_{\alpha,\beta} \right\}$ について 0 に収束することを示せばよい。

補題 7. 4 より $\lim_{N \rightarrow \infty} \Pi_N^{[M]} = P - I + e^G$, ここで $P = (p_{(i,\alpha),(j,\beta)}) = \lim_{m \rightarrow \infty} A^m = (\gamma_j \delta_{\alpha,\beta})$

なので、 $\sum_{j=1}^s (\Pi_N^{[M]})_{(i,\alpha),(j,\beta)} = \sum_{j=1}^s (P - I)_{(i,\alpha),(j,\beta)} + \sum_{j=1}^s (e^{tG})_{(i,\alpha),(j,\beta)}$

ここで $\sum_{j=1}^s (P - I)_{(i,\alpha),(j,\beta)} = \sum_{j=1}^s \gamma_j \delta_{\alpha,\beta} - \sum_{j=1}^s \delta_{(i,\alpha),(j,\beta)} = \delta_{\alpha,\beta} - \delta_{\alpha,\beta} = 0$ 。これより、

$$\lim_{N \rightarrow \infty} \sum_{j=1}^s (\Pi_N^{[M]})_{(i,\alpha),(j,\beta)} = \sum_{j=1}^s (e^{tG})_{(i,\alpha),(j,\beta)} = \sum_{j=1}^s \left(\sum_{k=0}^{\infty} \frac{(tG)^k}{k!} \right)_{(i,\alpha),(j,\beta)} = \sum_{k=0}^{\infty} \frac{t^k}{k!} \left(\sum_{j=1}^s (G^k)_{(i,\alpha),(j,\beta)} \right)$$

また、 $(e^{Q^0})_{\alpha,\beta} = \sum_{k=0}^{\infty} \frac{t^k}{k!} (Q^k)_{\alpha,\beta}$ なので、 $\sum_{j=1}^s (G^k)_{(i,\alpha),(j,\beta)} = (Q^k)_{\alpha,\beta}$ を示せばよい。

帰納法で示す。 $k=1$ のとき、 $\sum_{j=1}^s (G)_{(i,\alpha),(j,\beta)} = \left(\sum_{j=1}^s \gamma_j q_{\alpha,\beta} \right)_{(i,\alpha),(j,\beta)} = q_{\alpha,\beta}$

k まで成り立つと仮定して $k+1$ のとき、

$$\begin{aligned} \sum_{j=1}^s (G^{k+1})_{(i,\alpha),(j,\beta)} &= \sum_{j=1}^s \sum_{(\ell,\varepsilon) \in E_N} g_{(i,\alpha),(\ell,\varepsilon)} (G^k)_{(\ell,\varepsilon),(j,\beta)} \\ &= \sum_{(\ell,\varepsilon) \in E_N} g_{(i,\alpha),(\ell,\varepsilon)} \left(\sum_{j=1}^s (G^k)_{(\ell,\varepsilon),(j,\beta)} \right) = \sum_{(\ell,\varepsilon) \in E_N} \gamma_\ell q_{\alpha,\varepsilon} (Q^k)_{\varepsilon,\beta} \\ &= \sum_{\ell=1}^s \gamma_\ell \left(\sum_{\varepsilon \in E} q_{\alpha,\varepsilon} (Q^k)_{\varepsilon,\beta} \right) = \sum_{\ell=1}^s \gamma_\ell (Q^{k+1})_{\alpha,\beta} = (Q^{k+1})_{\alpha,\beta} \end{aligned}$$

$k+1$ のときも成り立つ。以上より $(Y_N((Nt)))_{t \geq 0}$ は生成作用素 $Q = (q_{\alpha,\beta})_{\alpha,\beta \in E}$ の連続時間マルコフ連鎖 $Y = (Y(t))$ に空間 $D_E[0, \infty)$ 上で弱収束 (分布収束) することが示された。

分集団サイズの変化は毎世代起こる速い変化なので N 世代を単位時間とする時間スケールを取り、 $N \rightarrow \infty$ の極限を取ると、各分集団サイズが時間平均化された標準的な **Structured coalescent** 過程に収束することが示された。

7. 4 自殖を含む合祖モデル

二倍体生物集団の繁殖様式の中には、植物で見られる自家受粉すなわち自殖(selfing)という現象がある。この場合子供の二つの遺伝子は単一の親に由来している。ある一定の確率で自殖がある生物集団の遺伝子系図について多くの研究があるがこの節ではMöhle(1998)の結果を紹介する。N個の二倍体の雌雄同体個体からなる生物集団を仮定する。子供世代から任意に1個体の子供を取り出したとき、この子供は確率sで自殖によって、確率1-sで他殖(outcrossing)によって生まれると仮定する。自殖の場合は子供の2個の遺伝子は確率1/2で親の一つの遺伝子に由来し(coalesce)、確率1/2で親の異なる二つの遺伝子に由来する。この生物集団は確率的に自殖を行う個体と他殖を行う個体の2グループに分割されることになる。異なる個体からサンプルした二つの遺伝子の親遺伝子は全集団の2N個の遺伝子集団から任意交配によって選ばれたものとなる。よって、二つの遺伝子が合祖する確率は自殖、他殖の場合分けにより $s^2 \times \frac{1}{2N} + 2s(1-s) \times \frac{1}{2N} + (1-s)^2 \times \frac{1}{2N} = \frac{1}{2N}$ となる。

この集団からn個の遺伝子をランダムに取り出したとき、その遺伝子系図を考察する。

(1) サンプル数n=2の遺伝子系図

二つの遺伝子の系図についてr世代前の状態を D_r とする。 $D_r = 1, 2$ または3の三つの状態を考える。

$D_r = 1 \Leftrightarrow$ 二つの遺伝子が同一である。共通祖先に到達した状態。

$D_r = 2 \Leftrightarrow$ 二つの遺伝子が異なる個体に属す。

$D_r = 3 \Leftrightarrow$ 二つの遺伝子が同じ個体に属す。

ランダムに二つの遺伝子を集団からサンプルしたとき、初期分布は

$P(D_0=1)=0, P(D_0=2)=1-\frac{1}{2N-1}=\frac{2N-2}{2N-1}, P(D_0=3)=\frac{1}{2N-1}$ となる。マルコフ連鎖 D_r の

1世代当たりの推移確率 Π_N は $\Pi_N = \begin{pmatrix} 1 & 0 & 0 \\ 1/2N & 1-1/N & 1/2N \\ s/2 & 1-s & s/2 \end{pmatrix}$ で与えられる。

すなわち、状態2のとき、二つの遺伝子は異なる個体に属すので、任意交配で得られるので状態2からの推移確率は $p_{21}=1/2N, p_{22}=1-1/N, p_{23}=1/2N$ となる。

他方状態3から状態1の推移確率は自殖かつ同じ遺伝子に由来する場合なのでs/2、状態2への推移確率は他殖の場合なので1-s、最後に状態3への推移確率はs/2となる。

Π_N は $\Pi_N = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ s/2 & 1-s & s/2 \end{pmatrix} + \frac{1}{N} \begin{pmatrix} 0 & 0 & 0 \\ 1/2 & -1 & 1/2 \\ 0 & 0 & 0 \end{pmatrix} = A + \frac{1}{N}B$ と分解できる。

$$\text{これより、 } P = \lim_{m \rightarrow \infty} A^m = \lim_{m \rightarrow \infty} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{s(1-(s/2)^m)}{2} & (1-s)\frac{(1-(s/2)^m)}{1-s/2} & (s/2)^m \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ p & q & 0 \end{pmatrix}, \quad (7.16)$$

$$\text{ただし、 } p = \frac{s}{2-s}, \quad q = 1-p = \frac{2(1-s)}{2-s}.$$

$$(2-s)PBP = \begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ q & -q & 0 \end{pmatrix} \text{ 及び } ((2-s)PBP)^k = \begin{cases} I & (k=0 \text{ のとき}) \\ (-1)^k (2-s)PBP & (k \geq 1 \text{ のとき}) \end{cases} \text{ に注意して}$$

$(2-s)N$ 世代を単位時間とする時間スケールを取り、 $N \rightarrow \infty$ の極限を取ると 7. 1 節の補題 7. 1 より

$$\lim_{N \rightarrow \infty} \Pi_N^{[(2-s)Nt]} = \lim_{N \rightarrow \infty} \left(A + \frac{B}{N} \right)^{[(2-s)Nt]} = P - I + \exp((2-s)tPBP) = \Pi(t),$$

$$\Pi(t) = \begin{pmatrix} 1 & 0 & 0 \\ 1-e^{-t} & e^{-t} & 0 \\ 1-qe^{-t} & qe^{-t} & 0 \end{pmatrix} \text{ であり、またその生成作用素は}$$

$$G = \lim_{t \downarrow 0} \frac{\Pi(t) - \Pi(0+)}{t} = (2-s)PBP = \begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ q & -q & 0 \end{pmatrix} \text{ となる。}$$

離散時間マルコフ連鎖 $(D_{[(2-s)Nt]})_{t \geq 0}$ の有限次元分布は推移確率 $\Pi(t)$ の連続時間マルコフ連鎖 $(D_t)_{t \geq 0}$ に収束する。マルコフ連鎖 $(D_t)_{t \geq 0}$ において状態 3 は瞬時的状態であり、時刻 0+ 以降は通常の合祖過程となる。

(2) サンプル数 n の場合

集団から n 個の遺伝子をランダムにサンプルし、その遺伝子系図を考える。このとき現集団、および祖先集団の N 個の二倍体生物の個々についてその状態を $k \in \{0, 1, 2\}$ で表す。ある個体についてその遺伝子のうち k 個の遺伝子が n 個のサンプルあるいはその祖先遺伝子に含まれるとき、タイプ k と呼ぶ。 N 個のサンプルの離散時間の遺伝子系図を

$$(D_r)_{r \in \mathbb{N}_0} = (\mathfrak{R}_r, \mathfrak{N}_r)_{r \in \mathbb{N}_0} \text{ で表す。ここで } \mathfrak{R}_r \text{ はサンプルの } r \text{ 世代前の祖先の遺伝子数、} \mathfrak{N}_r \text{ は}$$

タイプ 2 の祖先個体数を表す。 (D_r) の状態空間を S_n とすると、

$$S_n = \{i = (b, x) : b \in \{1, 2, \dots, n\}, x \in \{0, 1, \dots, [b/2]\}, b \text{ は祖先遺伝子数、} x \text{ はタイプ 2 の個体数を表す。また } |S_n| = \sum_{b=1}^n ([b/2] + 1) = [n(n+4)/4]. (D_r) \text{ の初期分布を求めるために、確率}$$

$P(N, n, x) = P(D_0 = (n, x)) = P(\mathfrak{N}_0 = x)$ を考える。すなわち、集団から n 個の遺伝子をランダムに取り出したとき、タイプ 2 の個体を x 個体含んでいる確率である。明らかに $P(N, 1, x) = \delta_{0,x}$ 、

次の漸化式が成り立つ。

$$P(N, n+1, x) = \frac{n-2x+2}{2N-n} P(N, n, x-1) + \frac{2(N-n+x)}{2N-n} P(N, n, x) \quad n \geq 1 \quad (7.17)$$

右辺第1項は n 個のサンプル中に $x-1$ 個のタイプ 2 が含まれているとき、 $2N-n$ 個の遺伝子からサンプル中のタイプ 1 の個体 $((n-2(x-1))$ 個体) から残りの 1 個の遺伝子をサンプルする場合であり、第2項は n 個のサンプル中に x 個のタイプ 2 が含まれているとき、サンプルに含まれていない個体から遺伝子を 1 個サンプルすれば良いので、サンプルした個体数が $n-x$ であることに注意すると、その確率は $\frac{2(N-(n-x))}{2N-n}$ となる。

$$\begin{aligned} \text{この漸化式の解は } P(N, n, x) &= \frac{(n)_{2x}}{2^x x!} \times \frac{2^{n-x} (N)_{n-x}}{(2N)_n} \quad \text{ただし } (n)_k = n(n-1)\cdots(n-k+1) \\ &= \begin{cases} 1 - q_n / (2N) + O(N^{-2}) & \text{if } x=0 \\ q_n / (2N) + O(N^{-2}) & \text{if } x=1 \\ O(N^{-2}) & \text{その他} \end{cases} \quad \text{ただし } q_n = \frac{n(n-1)}{2}. \end{aligned} \quad (7.18)$$

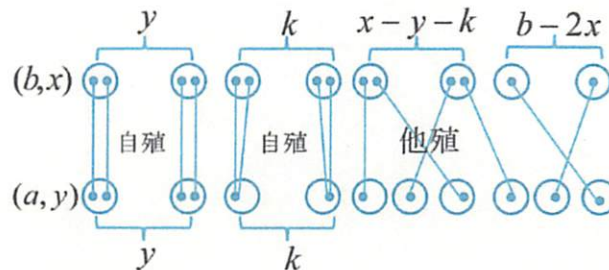
マルコフ連鎖 (D_r) の 1 世代の推移確率を求めよう。

定理 7. 6

$\Pi_N = (\pi_{i,j})_{i,j \in S_n} = P(D_r = j)$, ただし $i = (b, x)$, $j = (a, y) \in S_n$, $b \geq a$ とすると

$$\pi_{i,j} = \binom{x}{b-a} \binom{x-(b-a)}{y} (1-s)^{x-y-(b-a)} \left(\frac{s}{2}\right)^{y+b-a} + O(N^{-1}) \quad (7.19)$$

(証明) 他殖で合祖が起こる確率は $O(N^{-1})$ なので、 $N = \infty$ と仮定し、他殖による合祖を除外して自殖で合祖あるいは変化しない確率を求める。



上図において $N = \infty$ なので、一つの個体の相同な二つの遺伝子が任意交配によって共通な個体を親にもつことはありえないので、子供世代の y 個の個体は (自殖、非合祖)、 k 個の個体は (自殖、合祖)、 $x-y-k$ 個体は任意交配によって親から遺伝子を受け継いでいると仮定する。図より $y+k+\{b-2(y+k)\}=a$ 、よって $k=b-a$ 。子供世代の x 個のタイプ 2 遺伝子について y 個は (自殖、非合祖)、 $b-a$ 個は (自殖、合祖)、 $x-y-(b-a)$ 個は他殖となる

ので、求める確率は $\frac{x!}{(x-y-(b-a))!(b-a)!y!} (1-s)^{x-y-(b-a)} \left(\frac{s}{2}\right)^{b-a} \left(\frac{s}{2}\right)^y$ となり、整理すると $\pi_{i,j}$ を得る。

推移確率 $\Pi_N = (\pi_{i,j})$ の完全な形は困難であるが、 $x=0$ の場合は求めることができる。

すなわち、 $i=(b,0), j=(a,y)$ とすると $\pi_{i,j} = S(b,a) \frac{(2N)_a}{(2N)^b} P(N,a,y)$ となる。

なぜなら、 b 個の遺伝子は全て異なる個体に属するので、任意交配によって生まれた個体である。よって可能な親遺伝子の選び方は $(2N)^b$ 通り、このとき異なる a 個の親遺伝子を集団から選ぶ方法が $(2N)_a = 2N(2N-1)\cdots(2N-a+1)$ 通り、この a 個の部屋に b 個の遺伝子を割り当てる方法が $S(b,a)$ 通り、各々について (a,y) の状態、すなわちタイプ 2 が y 個ふくまれるようになる確率が $P(N,a,y)$ より得られる。また、このとき N のオーダーで整理すると

$$\pi_{i,j} = S(b,a) \frac{(2N)_a}{(2N)^b} P(N,a,y) = \begin{cases} 1 - q_b/N + O(N^{-2}) & \text{if } j=i=(b,0) \\ q_b/2N + O(N^{-2}) & \text{if } j=(b-1,0) \text{ または } j=(b,1) \\ O(N^{-2}) & \text{その他} \end{cases} \quad (7.20)$$

遺伝子系図過程 $(D_r)_{r \in N_0} = (\mathfrak{R}_r, \mathfrak{N}_r)_{r \in N_0}$ の推移確率 $\Pi_N = (\pi_{i,j})_{i,j \in S_n}$ は $\Pi_N = A + \frac{1}{N} B_N$ と分解できる。 $A = (a_{i,j}) = \lim_{N \rightarrow \infty} \Pi_N$, $B_N = N(\Pi_N - A)$ 。 $A = (a_{i,j})$, $i=(b,x), j=(a,y) \in S_n$ は

定理 7. 6 より $a_{i,j} = \frac{x!}{(x-y-(b-a))!(b-a)!y!} (1-s)^{x-y-(b-a)} \left(\frac{s}{2}\right)^{b-a} \left(\frac{s}{2}\right)^y$ 。これは、

三項展開 $\left((1-s) + \frac{s}{2} + \frac{s}{2} \right)^x$ の中の $(x-y-(b-a), b-a, y)$ の項に対応する。すなわち $(a_{i,j})_{i,j \in S_n}$ は三項分布 $M_n(x; 1-s, s/2, s/2)$ に従う。 $i=(b,x)$ を与えたとき、 $j=(a,y)$ が満たす条件は $x \geq x-y-(b-a) \geq 0$, $x \geq b-a \geq 0$, $x \geq y \geq 0$ 、すなわち $a-y \geq b-x \geq 0$, $b \geq a$, $x \geq y \geq 0$ 。

よって可能な範囲で $j=(a,y)$ を動かすと $\sum_{j \in S_n} a_{i,j} = 1$ より行列 A は確率行列である。 S_n の全要素に以下の順序 $(\tilde{\zeta})$ を定義する。 $i=(b,x) \stackrel{\text{定義}}{\tilde{\zeta}} j=(a,y) \Leftrightarrow b > a$ または $(a=b, x \geq y)$ 。 $a_{i,j} > 0$ ならば $i \tilde{\zeta} j$ であることが分かる。すなわち、 S_n の要素を順序 $\tilde{\zeta}$ に従って並べ、行列 A の成分をこの順序で配置すると、行列 A は下三角行列になる。行列 $B_N = (b_{i,j}^{(N)})$ については

$i=(b,x), x=0$ のときは前の結果より

$$b_{i,j} = \lim_{N \rightarrow \infty} b_{i,j}^{(N)} = \begin{cases} -q_b & \text{if } j = i \\ q_b/2 & \text{if } j = (b-1,0) \text{ または } (b,1) \\ 0 & \text{その他} \end{cases} \quad (7.21)$$

行列Aは確率行列なので、Aを推移確率とする S_n 上のマルコフ連鎖、すなわち自殖のみによって推移する遺伝子系図のプロセスを $(Z_m)_{m \in \mathbb{N}_0}$ とする。 $i = (b,0)$ のとき $a_{i,i} = 1$ なので状態

$i = (b,0)$ は吸収状態である。 $\lim_{m \rightarrow \infty} A^m = P = (p_{i,j})$ とすると、 $P = AP$ より

$$p_{i,j} = \sum_{k \in S_n} a_{i,k} p_{k,j} = \sum_{j \leq k \leq i} a_{i,k} p_{k,j} = a_{ii} p_{i,j} + \sum_{j \leq k < i} a_{i,k} p_{k,j} \text{ なるので}$$

$$p_{i,j} = \begin{cases} \delta_{i,j} & \text{if } i = (b,0) \text{ 吸収状態} \\ \left(\sum_{i \leq k \leq j} a_{i,k} p_{k,j} \right) / (1 - a_{i,i}) & \text{if } a_{i,i} < 1 \end{cases}$$

$$\text{この漸化式より } p_{i,j} = \begin{cases} \binom{x}{b-a} p^{b-a} q^{x-(b-a)} & \text{if } y=0 \\ 0 & \text{その他} \end{cases} \quad (7.22)$$

この結果は次の様に考えることができる。前節で見たように、サンプル数 $n=2$ のとき同一個体中の二つの遺伝子が自殖のみによって合祖する確率は $p = \frac{s}{2-s}$ 、最終的に合祖しない確率は $q = 1-p$ となる。

サンプル数 n で $i = (b,x)$ のとき x 個のタイプ 2 の個体について各々 2 個ずつの遺伝子について合祖するか否かは独立なので、最終的に状態 $j = (a,0)$ へ吸収されるためには $b-a$ 回の自殖による合祖と $x-(b-a)$ 回の非合祖があれば良いので二項分布

$$p_{i,j} = \binom{x}{b-a} p^{b-a} q^{x-(b-a)} \text{ となる。}$$

また以下の計算により $\sum_{k \in S_n} a_{i,k} p_{k,j} = p_{i,j}$ が成り立つことが分かる。 $i = (b,x), j = (a,y)$ とする。

$y > 0$ のとき、 $p_{k,j} = 0 (k \in S_n)$ なので明らか。よって $y=0, j = (a,0)$ とする。

$$\sum_{k=(c,z) \in S_n} a_{i,k} p_{k,j} = \sum_{c=a}^b \sum_{z=c-a}^{x-(b-c)} \binom{x}{b-c} \binom{x-(b-c)}{z} (1-s)^{x-(b-c)-z} \left(\frac{s}{2}\right)^{z+b-c} \times \binom{z}{c-a} p^{c-a} q^{z-(c-a)}$$

$$\text{ここで } \binom{x}{b-c} \binom{x-(b-c)}{z} \binom{z}{c-a} = \binom{x}{b-a} \binom{b-a}{c-a} \binom{x-(b-a)}{z-(c-a)}, \text{ また } p = \frac{s}{2-s}, q = 1-p \text{ より}$$

$$1-s = \frac{1-p}{1+p} = \frac{q}{1+p}, \frac{s}{2} = \frac{p}{1+p}, \text{ よって } (1-s)^{x-(b-c)-z} \left(\frac{s}{2}\right)^{z+b-c} p^{c-a} q^{z-(c-a)} = \frac{p^{z+b-a} q^{x-(b-a)}}{(1+p)^x} \text{ と変}$$

$$\begin{aligned}
h_{i,j} &= \sum_{k=(c,0) \in S_n} P_{i,k} \left\{ \frac{1}{2} q_c P_{(c-1,0),j} - q_c P_{(c,0),j} + \frac{1}{2} q_c P_{(c,1),j} \right\} \quad i=(b,x), j=(a,0) \text{ なるので } c=a, a+1 \\
&= P_{i,j} \left(\frac{1}{2} q_a P_{(a-1,0),j} - q_a P_{j,j} + \frac{1}{2} q_a P_{(a,1),j} \right) + P_{i,(a+1,0)} \left(\frac{1}{2} q_{a+1} P_{(a,0),j} - q_{a+1} P_{(a+1,0),j} + \frac{1}{2} q_{a+1} P_{(a+1,1),j} \right) \\
&= P_{i,j} \left(0 - q_a + \frac{1}{2} q_a q \right) + P_{i,(a+1,0)} \left(\frac{1}{2} q_{a+1} - 0 + \frac{1}{2} q_{a+1} p \right) = P_{i,j} q_a \left(\frac{q}{2} - 1 \right) + \frac{1}{2} (p+1) q_{a+1} P_{i,(a+1,0)} \\
&= \frac{1}{2-s} (q_{a+1} P_{i,(a+1,0)} - q_a P_{i,j}) = \frac{1}{2-s} (q_{a+1} B(x, p, b-a-1) - q_a B(x, p, b-a)) \quad (7.25)
\end{aligned}$$

$$B(x, p, k) = \binom{x}{k} p^k (1-p)^{x-k} \text{ は二項分布。}$$

補題 7. 1 より $\Pi(t) = \lim_{N \rightarrow \infty} \Pi_N^{(2-s)Nt} = P - I + e^{tG} = P e^{tG}$ 、ただし $G = (2-s)H$ 。

まとめると次の定理が成り立つ。

定理 7. 7

自殖を含む遺伝子系図のマルコフ連鎖 $(D_{[(2-s)Nt]})_{t \geq 0}$ の有限次元分布は推移確率

$\Pi(t) = P - I + e^{tG} = P e^{tG}$ の連続時間マルコフ連鎖 $(D_t)_{t \geq 0}$ のそれに収束する。行列 P

は式(7.22)で与えられ、行列 $G = (g_{i,j})$ は $i = (b, x), j = (a, y) \in S_n, q_a = \frac{a(a-1)}{2}, p = \frac{s}{2-s}$ とし

て $g_{i,j} = q_{a+1} B(x, p, b-a-1) - q_a B(x, p, b-a)$ である。

(補足)

$$(1) i = (b, 0) \text{ のとき、 } g_{i,j} = q_{a+1} \delta_{b,a+1} - q_a \delta_{a,b} = \begin{cases} -q_b & \text{if } j = (b, 0) \\ q_b & \text{if } j = (b-1, 0) \\ 0 & \text{その他} \end{cases} \text{。これは初期状態 } i = (b, 0)$$

から出発すると Kingman の合祖過程となることを示している。

(2) 一般の初期状態 $i = (b, x)$ から出発すると $\Pi(0+) = \lim_{t \downarrow 0} \Pi(t) = P$ より瞬時に状態 $j = (a, 0)$ へ

確率 $p_{i,j} = \binom{x}{b-a} p^{b-a} q^{x-(b-a)}$ で移り、その後は初期サンプル数 a の Kingman 合祖過程とし

て振る舞う。

(3) 同値類を考慮に入れた自殖を含む n -合祖過程

前節では自殖を含むモデルにおいて n 個の遺伝子サンプルの系図過程を考察してきた。状態空間 S_n はサンプル遺伝子の個数とサンプルのタイプで表示される空間であったが、第 3 章で紹介した Kingman の合祖過程と同様に、サンプル遺伝子のどの遺伝子が祖先を共有するか、サンプル遺伝子の集合に同値類を導入して考える。半数体生物集団から取り出し

た n 個の遺伝子 $\{I_1, I_2, \dots, I_n\}$ について、その同値類の集合を E_n とする。同じ同値類に属す遺伝子を (I_1, I_2, I_3) など (\cdot) で表示する。 $|E_n| = \sum_{i=1}^n S_n^{(i)}$ 、 $S_n^{(i)}$ は第 2 種スターリング数で n 個の遺伝子を i 個の同値クラスに分割する方法の数である。他方二倍体生物の場合は各遺伝子がどの個体に属しているかを指定しなくてはならない。例えば I_1 と I_2 が同じ個体に属す相同な遺伝子のとき $\langle I_1, I_2 \rangle$ と表示することにする。 n 個の遺伝子について保有個体も含めて表示すると $\xi = \langle \langle I_1, I_2 \rangle, \dots, \langle I_{2x-1}, I_{2x} \rangle, I_{2x+1}, \dots, I_n \rangle$ となる。前節と同様にサンプル中に含まれる遺伝子数に応じて各個体をタイプ k ($k=0,1,2$) と呼ぶことにする。 n 個のサンプル遺伝子について、タイプ 2 の個体が x 個あるとき、状態 ξ の様な n 個の遺伝子の配置の方法はタイプ 2 の個体の遺伝子を順次取り出せば良いので、並べ方は無視して

$$\frac{1}{x!} \prod_{k=0}^{x-1} \frac{(n-2k)(n-2k-1)}{2} = \frac{(n)_{2x}}{2^x x!} \text{ 通りとなる。二倍体生物の場合の } n \text{ 個の遺伝子について}$$

祖先の共有による同値類の全集合を S_n とすると、

$$|S_n| = \sum_{\xi \in S_n} 1 = \sum_{b=1}^n \sum_{\substack{\xi \in E_n \\ |\xi|=b}} \sum_{x=0}^{\lfloor b/2 \rfloor} \frac{(b)_{2x}}{2^x x!} = \sum_{b=1}^n S_n^{(b)} \sum_{x=0}^{\lfloor b/2 \rfloor} \frac{(b)_{2x}}{2^x x!} \quad (7.26)$$

$n=3$ の場合、サンプル遺伝子を I_1, I_2, I_3 とすると

$$E_3 = (\langle (I_1), (I_2), (I_3) \rangle, \langle (I_1, I_2), (I_3) \rangle, \langle (I_1, I_3), (I_2) \rangle, \langle (I_2, I_3), (I_1) \rangle, \langle (I_1, I_2, I_3) \rangle), |E_3| = 5$$

$$S_3 = \left(\begin{array}{l} \langle \langle (I_1), (I_2) \rangle, (I_3) \rangle, \langle \langle (I_1), (I_3) \rangle, (I_2) \rangle, \langle \langle (I_2), (I_3) \rangle, (I_1) \rangle, \langle (I_1), (I_2), (I_3) \rangle, \langle \langle (I_1, I_2), (I_3) \rangle \rangle \\ \langle \langle (I_1, I_2), (I_3) \rangle, \langle \langle (I_1, I_3), (I_2) \rangle, \langle \langle (I_1, I_3), (I_2) \rangle, \langle \langle (I_2, I_3), (I_1) \rangle \rangle \\ \langle \langle (I_2, I_3), (I_1) \rangle, \langle (I_1, I_2, I_3) \rangle \end{array} \right)$$

$\langle I_k, I_j \rangle$ は遺伝子 I_k, I_j が同一個体の遺伝子であることを表す記号であり、 $|S_3| = 11$ である。

E_3 は S_3 の部分集合となる。 n 個の遺伝子について半数体生物として b 個の同値クラスに分ける方法は $S_n^{(b)}$ 通りあり、二倍体生物として考える場合はその各々の同値類について、

$$\sum_{x=0}^{\lfloor b/2 \rfloor} \frac{(b)_{2x}}{2^x x!} \text{ 通りの配置があることより (7.26) 式を得る。}$$

二倍体生物における同値類を考慮に入れた遺伝子系図の合祖過程を $\{C_r\}_{r \in \mathbb{N}_0}$ とする。

初期状態を $\xi \in S_n$ 、含まれる遺伝子の数を $|\xi| = n$ 、タイプ 2 の個体数を $\|\xi\| = x \in \{0, \dots, \lfloor b/2 \rfloor\}$ とする。 $(\|\xi\|, \|\xi\|)$ を $\xi \in S_n$ の指標と呼び $I(\xi) = (\|\xi\|, \|\xi\|)$ で表すことにする。このとき、

$$\begin{aligned}
P(C_0 = \xi) &= \frac{P(N, n, x)}{(n)_{2x} / (2^x x!)} = \frac{2^{n-x} (N)_{n-x}}{(2N)_n} = \prod_{k=0}^{n-x-1} \frac{2N-2k}{2N-k} \prod_{k=n-x}^{n-1} \frac{1}{2N-k} \quad (7.27) \\
&= \begin{cases} 1 - q_n / 2N + O(N^{-2}) & (x=0 \text{ のとき}) \\ 1/2N & (x=1 \text{ のとき}) \\ 0 & (x \geq 2 \text{ のとき}) \end{cases}, \quad q_n = \frac{n(n-1)}{2}
\end{aligned}$$

マルコフ連鎖 $\{C_r\}_{r \in \mathbb{N}_0}$ の推移確率を $\Pi_N = (\pi_{\xi\eta})_{\xi, \eta \in \mathcal{S}_n} = (P(C_r = \eta | C_{r-1} = \xi))$ とする。

$\xi = \{(C_1, C_2), \dots, (C_{2x-1}, C_{2x}), C_{2x+1}, \dots, C_b\}$ $I(\xi) = (b, x)$ 及び
 $\eta = \{(D_1, D_2), \dots, (D_{2y-1}, C_{2y}), C_{2y+1}, \dots, C_a\}$ $I(\eta) = (a, y)$ とすると

定理 7. 10 の証明で述べたように、その指標から x 個のタイプ 2 遺伝子の中の y 個は (自殖、非合祖)、 $b-a$ 個は (自殖、合祖)、 $x-y-(b-a)$ 個は他殖とならなければならない。また、 ξ から η へ推移するためには ξ と η の間に以下の関係が成立しなければならない。

$$\begin{aligned}
D_i &= C_i \quad (i=1, \dots, 2y); \quad D_i = C_{2i-2y-1} \cup C_{2i-2y} \quad (i=2y+1, \dots, 2y+(b-a)); \\
D_i &= C_{i+b-a} \quad (i=2y+(b-a)+1, \dots, a)
\end{aligned}$$

この関係が成り立つとき $\xi \triangleleft \eta$ で表す。 (b, x) が与えられた時、指標 (a, y) を持つ η の個数は、

$\frac{x!}{(x-y-(b-a))!(b-a)!y!}$ 通りあるので推移確率 $\pi_{\xi\eta}$ は

$$\xi \triangleleft \eta \text{ のとき } \pi_{\xi, \eta} = (1-s)^{x-y-(b-a)} \left(\frac{s}{2}\right)^{y+b-a} + O(N^{-1}), \text{ その他は } \pi_{\xi, \eta} = O(N^{-1}) \text{ となる。} \quad (7.28)$$

$$\text{特に } \|\xi\| = 0 \text{ のとき、 } \pi_{\xi\eta} = \frac{(2N)_a}{(2N)^b} P(N, a, y) \frac{2^y y!}{(a)_{2y}} = \frac{2^{a-y} (N)_{a-y}}{(2N)^b} \quad (7.29)$$

$$= \begin{cases} 1 - b(b-1)/(2N) + O(N^{-2}) & \text{if } \xi = \eta \\ 1/(2N) + O(N^{-2}) & \text{if } \xi \triangleleft \eta \text{ または } \xi \rightarrow \eta \\ O(N^{-2}) & \text{その他} \end{cases}$$

ここで $\xi \triangleleft \eta \Leftrightarrow \xi \subseteq \eta, |\xi| = |\eta| + 1, \|\xi\| = \|\eta\|$ 、 $\xi \rightarrow \eta \Leftrightarrow \xi \subseteq \eta, |\xi| = |\eta|, \|\xi\| = \|\eta\| - 1$ 。

(7.28) より推移確率 Π_N は $\Pi_N = A + \frac{B_N}{N}$ 、ただし $A = (a_{\xi, \eta})$

$$a_{\xi, \eta} = \begin{cases} (1-s)^{x-y-(b-a)} \left(\frac{s}{2}\right)^{y+b-a} & \text{if } \xi \triangleleft \eta \\ 0 & \text{その他} \end{cases} \quad (7.30)$$

また同じ指標の η の個数に注意して $\sum_{\eta \in \mathcal{S}_n} a_{\xi, \eta} = 1$ であることも同様に確かめられる。

行列 B_N については $\|\xi\| = 0$ のとき(7.29)より

$$b_{\xi\eta} = \lim_{N \rightarrow \infty} b_{\xi\eta}^{(N)} = \begin{cases} -b(b-1)/2 & \text{if } \xi = \eta \\ 1/2 & \text{if } \xi < \eta \text{ または } \xi \rightarrow \eta \\ 0 & \text{その他} \end{cases} \quad (7.31)$$

また、定常分布 $\lim_{m \rightarrow \infty} A^m = P = (p_{\xi\eta})$ は方程式 $P = AP$ の解であるが、次式で与えられる。

$$p_{\xi\eta} = \begin{cases} p^{b-a} q^{x-(b-a)} & \text{if } \|\eta\| = 0, \xi \subseteq \eta \\ 0 & \text{その他} \end{cases} \quad (7.32)$$

計算は少々煩雑になるが $\sum_{\nu \in S_n} a_{\xi\nu} p_{\nu\eta} = p_{\xi\eta}$ が成り立つことが確かめられる。

行列 $H = \lim_{N \rightarrow \infty} PB_N P = (h_{\xi\eta})$ については

$$h_{\xi\eta} = \begin{cases} \frac{1}{2-s} \left(\sum_{\substack{\nu \in S_n \\ \nu < \eta}} p_{\xi\nu} - q_a p_{\xi\eta} \right) & \text{if } \|\eta\| = 0, \xi \subseteq \eta \\ 0 & \text{その他} \end{cases} \quad \text{で与えられる。}$$

以上の結果より、 $(2-s)N$ 世代を単位時間 $t=1$ に取り、Möhle(1998)の補題 7-1 を応用すると次の定理を得る。

定理 7. 8

マルコフ連鎖 $(C_{[(2-s)N]})_{t>0}$ は推移確率行列 $\Pi(t) = P - I + e^{Gt} = Pe^{Gt}$ の連続時間マルコフ連鎖 $(C(t))_{t>0}$ に有限次元分布収束する。ここで、行列 P は(7.32)式で与えられ、生成行列 G の成分は次式で与えられる。

$$g_{\xi,\eta} = \sum_{\substack{\nu \in S_n \\ \nu < \eta}} p_{\xi,\nu} - \frac{|\eta|(|\eta|-1)}{2} p_{\xi,\eta} \quad \text{任意の } \xi, \eta \in S_n$$

7. 5 種子の年齢を考慮にいた合祖モデル

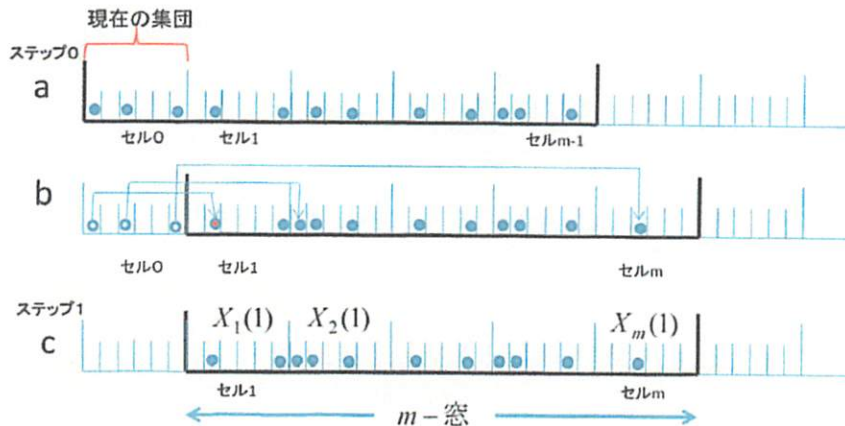
これまで、親から子へと1世代を単位とした世代交代のモデルを扱ってきたが、植物など何世代か前の種子が発芽し成長した個体が集団を構成するものも存在する。この様に幾世代かに渡って作られた配偶子によって次世代の集団が形成されるモデルとして Seed bank モデルが Kaj, Krone and Lascoux(2001)によって研究された。このモデルの遺伝子

系図についても収束定理（定理7. 2）とマルコフ連鎖の弱収束定理を適用することにより、標準的な合祖過程へ収束することが示される。

(1) モデル

集団は N 個の個体からなる半数体生物集団とする。各個体は数世代前に作られた種子から成長したものと仮定する。この種子が i 世代前の集団で作られた種子から任意にえらばれる確率を b_i とし、 i は $1 \leq i \leq m$ の範囲の自然数であり、 $\sum_{i=1}^m b_i = 1$ とする。 N 個の個体について i 世代前の種子によるものを X_i 個 ($1 \leq i \leq m$) とすると $X = (X_i; 1 \leq i \leq m)$

は多項分布 $P(X_i = x_i; 1 \leq i \leq m) = \frac{N!}{\prod_{i=1}^m x_i!} \prod_{i=1}^m b_i^{x_i}$ に従う。 $m = 1$ のときは通常の Wright-Fisher モデルとなる。下図に従ってモデルを説明する。



各セルは N 個のスロットから成り、各スロットは 1 個体を表す。セル 0 は現在の集団を表し、セル i ($1 \leq i \leq m-1$) は i 世代前の集団を表す。連続した m 個のセルを m -窓と呼ぶことにする。 $i = 0, 1, \dots, m-1$ の m 個のセルから合計 n 個の遺伝子サンプルを取り出す。セル i から取り出した個体数を $X_{i+1}(0)$ とする。 $\sum_{i=1}^m X_i(0) = n$ であり、現集団からの $X_1(0)$ 個のサンプルは成体（植物体）のサンプルであり、 $X_i(0)$ ($2 \leq i \leq m$) 個のサンプルは i 世代前の種子サンプルと考えることもできる。現在の集団だけから n 個のサンプルを取り出す場合は $X_1(0) = n$, $X_i(0) = 0$ ($2 \leq i \leq m$) となる。図のボールはサンプル遺伝子を表すが、初期時刻 (a) にはセル $m, m+1, m+2, \dots$ にはボールは全く含まれていない。図 (b) ではボールはそのまま m -窓を右に 1 セルだけずらす操作を行う。次にセル 0 にある各ボールを独立に確率 b_1, \dots, b_m でセル $1, 2, \dots, m$ に移す。セル内のどのスロットを選ぶかは等確率 ($1/N$) である。これは現集団のサンプル植物体が元の種子の状態に戻ることを意味する。移したボールが既にボールが占めているスロットに入るときは、この 2 個のボールは合祖 (coalesce) して、総ボール数は一つ減少する。セル 0 の $X_1(0)$ 個のボールを全て移した後の m -窓の各セルのボ

ールの数を $(X_1(1), X_2(1), \dots, X_m(1))$ とする。 $A_N(1) = \sum_{i=1}^m X_i(1) \leq n$ である。 $n - \sum_{i=1}^m X_i(1)$ は合祖の回数を表し、 $A_N(1)$ は m -窓に分配された祖先の数を表している。この操作を繰り返し、 k ステップ後の m -窓に分配された祖先の数を $X(k) = (X_1(k), X_2(k), \dots, X_m(k))$ とする。 $X_i(k)$ はセル $i+k-1$ にあるボールの数である。次のステップ $k+1$ は m -窓を右に 1 セルだけずらし、最も左のセル内の $X_1(k)$ のボールを右の m 個のセルに同様に移す。

次のステップ $k+1$ で m -窓の i 番目のセルに移されたボールの数を $R_i(k+1)$ とし、 $R(k+1) = (R_1(k+1), \dots, R_m(k+1))$ と置く。

$X_1(k)$ が与えられたとき、 $R(k+1)$ は多項分布 $Mult(X_1(k): b_1, \dots, b_m)$ に従う、すなわち

$$P(R(k+1) = (x_1, \dots, x_m)) = \frac{X_1(k)!}{\prod_{i=1}^m x_i!} \prod_{j=1}^m b_j^{x_j}。 R_i(k+1) の周辺分布は二項分布 $Bin(X_1(k): b_i)$ に従$$

う。 $X(k) = (X_1(k), \dots, X_m(k))$ は離散時間マルコフ連鎖となる。マルコフ連鎖 $X(k)$ は第 k 世代の m 個のセル内の遺伝子の数を表示しているので形状過程(Configuration process)と呼ぶことにする。

ここで、簡単な $n = 2, m = 2$ の例で推移確率を具体的に見てみよう。状態空間は $S_1 = \{(1,0), (0,1)\}, S_2 = \{(2,0), (1,1), (0,2)\}, S = S_1 \cup S_2$ となる。2 個の遺伝子をサンプルしたときの 1 世代の推移確率は以下ようになる。 S の要素をこの順に並べて行列表示する。

$$\begin{aligned} \Pi_N &= \begin{pmatrix} b_1 & b_2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ b_1^2/N & b_2^2/N & b_1^2(1-1/N) & 2b_1b_2 & b_2^2(1-1/N) \\ b_1/N & 0 & b_1(1-1/N) & b_2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{matrix} (1,0) \\ (0,1) \\ (2,0) \\ (1,1) \\ (0,2) \end{matrix} \\ & \quad \begin{matrix} (1,0) & (0,1) & (2,0) & (1,1) & (0,2) \end{matrix} \\ &= \begin{pmatrix} b_1 & b_2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & b_1^2 & 2b_1b_2 & b_2^2 \\ 0 & 0 & b_1 & b_2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} + \frac{1}{N} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ b_1^2 & b_2^2 & -b_1^2 & 0 & -b_2^2 \\ b_1 & 0 & -b_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (7.40) \end{aligned}$$

例えば形状過程 $X_N(k)$ が $(2,0)$ から $(2,0)$ へ推移するためには $m = 2$ なので、初期状態として左のセルにある 2 個の遺伝子の祖先が 1 世代前の同じセルの種子に由来し、かつ合祖が起こらなければよいので、 $P((2,0) \rightarrow (2,0)) = b_1^2(1-1/N)$ となる。 $(2,0)$ から $(1,1)$ への推移は 2 個の遺伝子が 1 個は一つ右のセルの種子に、もう 1 個が二つ右のセルの種子に由来すればよいので、 $P((2,0) \rightarrow (1,1)) = 2b_1b_2$ となる。 Π_N の他の成分についても同様に求めるこ

とができる。

(2) 主定理と合祖確率

以上のモデルの下で次の定理が成り立つ。

定理 7. 9

$E = \{1, 2, \dots, n\}$, $\beta_1 = \left(\sum_{i=1}^m ib_i \right)^{-1}$ とする。遺伝子系図過程 $A_N([Nt])$ は $N \rightarrow \infty$ において、

空間 $D_E[0, \infty)$ 内で生成作用素 $Q = (q_{i,j})_{i,j \in E}$ の連続時間マルコフ連鎖 $A(t)$ に弱収束する。

$$q_{i,j} = \begin{cases} -\beta_1^2 \frac{i(i-1)}{2} & \text{if } j = i \\ \beta_1^2 \frac{i(i-1)}{2} & \text{if } j = i-1 \\ 0 & \text{その他} \end{cases} \quad (7.41)$$

ℓ 個のボールが含まれるセルに r 個のボールが移入してきたとき、全く合祖(coalesce)が起こらない確率は合祖の回数を C とすると

$$\begin{aligned} P(C=0) &= \left(1 - \frac{\ell}{N}\right) \left(1 - \frac{\ell+1}{N}\right) \cdots \left(1 - \frac{\ell+r-1}{N}\right) = 1 - \frac{1}{N}(\ell + (\ell+1) + \cdots + (\ell+r-1)) + O(N^{-2}) \\ &= 1 - \frac{1}{N} \left(\ell r + \binom{r}{2} \right) + O(N^{-2}) \end{aligned}$$

同様にして

$$P(C \geq 1) = \frac{1}{N} \left(\ell r + \binom{r}{2} \right) + O(N^{-2}), \quad P(C \geq 2) = O(N^{-2})$$

$$\text{これより } P(C=1) = \frac{1}{N} \left(\ell r + \binom{r}{2} \right) + O(N^{-2}).$$

$X(k), R(k+1)$ が与えられたとき、 $C \geq 2$ の事象は確率 $O(N^{-2})$ なので、無視すると $X(k+1)$ が次式で与えられる。

$$X_i(k+1) = \begin{cases} X_{i+1}(k) + R_i(k+1) & \text{if } C=0 \\ X_{i+1}(k) + R_i(k+1) - 1 & \text{if } C=1 \end{cases}$$

まとめると、 $X(k) = (X_1(k), \dots, X_m(k))$ として

$$X(k+1) = \begin{cases} \alpha X(k) + R(k+1) & \text{with prob} = 1 - N^{-1} \left(\sum_{i=1}^m a_i(k) \right) + O(N^{-2}) \\ \alpha X(k) + R(k+1) - e_i & \text{with prob} = N^{-1} a_i(k) + O(N^{-2}) \\ \text{その他} & \text{with prob} = O(N^{-2}) \end{cases}$$

ただし、 $a_i(k) = X_{i+1}(k)R_i(k+1) + \binom{R_i(k+1)}{2}$, $\alpha X(k) = (X_2(k), \dots, X_m(k), 0)$ 。

$\Delta A_N(k) = A_N(k+1) - A_N(k)$ とすると、 $-\Delta A_N(k)$ はステップ k での合祖の数を表すが上記のことより、

$$P(\Delta A_N(k) = -1 | X(k), R(k+1)) = N^{-1} \left(\sum_{i=1}^{m-1} X_{i+1}(k)R_i(k+1) + \sum_{i=1}^m \binom{R_i(k+1)}{2} \right) + O(N^{-2}) \quad (7.42)$$

$R_i(k+1)$ は二項分布 $\text{Bin}(X_1(k); b_i)$ に従うことより、平均 $E[R_i(k+1)] = X_1(k)b_i$,

分散 $V[R_i(k+1)] = X_1(k)b_i(1-b_i)$ 、これより $E\left[\binom{R_i(k+1)}{2}\right] = \binom{X_1(k)}{2} b_i^2$ 。

故に(7.42)式について、 $X(k)$ を与えて $R(k+1)$ について期待値を取ると

$$P(\Delta A_N(k) = -1 | X(k)) = N^{-1} \left(X_1(k) \sum_{i=1}^{m-1} X_{i+1}(k)b_i + \binom{X_1(k)}{2} \sum_{i=1}^m b_i^2 \right) + O(N^{-2})$$

これより、状態 $X = (x_1, \dots, x_m) \in S_r$ のとき、次のステップで合祖が起こる確率は

$$P(C = 1 | X) = N^{-1} \left(x_1 \sum_{i=1}^{m-1} x_{i+1} b_i + \binom{x_1}{2} \sum_{i=1}^m b_i^2 \right) + O(N^{-2}) \quad (7.43)$$

N を大きくして行くと、合祖の確率は $O(N^{-1})$ で小さくなるので、二つの合祖の間の時間間隔は大きくなる。その間に形状(Configuration)過程は平衡状態に到達する。

レベル r の形状過程 $\{X^{(r)}(k)\}_{k=1,2,\dots}$ は以下のように定義される。

$$\begin{cases} X_j^{(r)}(k+1) = X_{j+1}^{(r)}(k) + R_j^{(r)}(k+1) & ; j = 1, 2, \dots, m \\ X_{m+1}^{(r)}(k) = 0 \end{cases} \quad (7.44)$$

$X_1^{(r)}(k)$ が与えられたとき、 $R^{(r)}(k+1) = (R_1^{(r)}(k+1), \dots, R_m^{(r)}(k+1))$ は多項分布

$\text{Mult}(X_1^{(r)}(k); b_1, \dots, b_m)$ に従う。分布 $P(B = i) = b_i$ を持つ確率変数を B とする。

$$\beta_j = \frac{P(B \geq j)}{E[B]} = \frac{\sum_{i=j}^m b_i}{\sum_{i=1}^m i b_i}, \quad (j=1,2,\dots,m) \text{ とすると、 } \beta_1 = \frac{1}{E[B]}, \quad \sum_{j=1}^m \beta_j = \frac{E[B]}{E[B]} = 1.$$

(β_j) は次の方程式を満たす。

$$j=1,\dots,m-1 \text{ のとき } \beta_j = \beta_{j+1} + \beta_1 b_j, \quad \beta_m = \beta_1 b_m \quad (7.45)$$

補題 7. 10

$r \in \{1,2,\dots,n\}$ に対してレベル r の形状過程 $\{X^{(r)}(k)\}_{k=1,2,\dots}$ の定常分布は

$X^{(r)}(\infty) = (X_1^{(r)}, \dots, X_m^{(r)}) \sim \text{Mult}(r; \beta_1, \dots, \beta_m)$ の多項分布で与えられる。

(証明) $X^{(r)}(k) \sim \text{Mult}(r; \beta_1, \dots, \beta_m)$ とする。その確率母関数は

$$\begin{aligned} E[u_1^{X_1^{(r)}(k)} \dots u_m^{X_m^{(r)}(k)}] &= \sum_{(y_1, \dots, y_m)} u_1^{y_1} \dots u_m^{y_m} P(X_i^{(r)}(k) = y_i, 1 \leq i \leq m) \\ &= \sum_{(y_1, \dots, y_m)} \frac{r!}{\prod_{i=1}^m y_i!} \prod_{i=1}^m \beta_i^{y_i} \prod_{i=1}^m u_i^{y_i} = (u_1 \beta_1 + \dots + u_m \beta_m)^r \end{aligned}$$

次に $X^{(r)}(k+1)$ の確率母関数を求める。(7.44)より

$$E[u_1^{X_1^{(r)}(k+1)} \dots u_m^{X_m^{(r)}(k+1)}] = E\left[\prod_{i=1}^m u_i^{X_i^{(r)}(k) + R_i^{(r)}(k+1)}\right] = E[u_1^{X_1^{(r)}(k)} \dots u_{m-1}^{X_{m-1}^{(r)}(k)} u_1^{R_1^{(r)}(k+1)} \dots u_m^{R_m^{(r)}(k+1)}]$$

ただし $X_{m+1}^{(r)}(k) = 0$ 。 $X_1^{(r)}(k)$ を与えたとき、 $R^{(r)}(k+1) = (R_1^{(r)}(k+1), \dots, R_m^{(r)}(k+1))$ は

多項分布 $\text{Mult}(X_1^{(r)}(k); b_1, \dots, b_m)$ に従うので

$$E[u_1^{R_1^{(r)}(k+1)} \dots u_m^{R_m^{(r)}(k+1)} | X_1^{(r)}(k)] = (b_1 u_1 + \dots + b_m u_m)^{X_1^{(r)}(k)}. \text{ また } X_1^{(r)}(k) \text{ を与えたとき、}$$

$(X_2^{(r)}(k), \dots, X_m^{(r)}(k))$ と $(R_1^{(r)}(k+1), \dots, R_m^{(r)}(k+1))$ は独立である。よって

$$E[u_1^{X_1^{(r)}(k+1)} \dots u_m^{X_m^{(r)}(k+1)}] = E\left[E[u_1^{X_2^{(r)}(k)} \dots u_{m-1}^{X_{m-1}^{(r)}(k)} | X_1^{(r)}(k)] E[u_1^{R_1^{(r)}(k+1)} \dots u_m^{R_m^{(r)}(k+1)} | X_1^{(r)}(k)]\right]$$

$$\begin{aligned}
&= E\left[(b_1 u_1 + \dots + b_m u_m)^{X_1^{(r)}(k)} u_1^{X_2^{(r)}(k)} \dots u_{m-1}^{X_m^{(r)}(k)}\right] = \{\beta_1(b_1 u_1 + \dots + b_m u_m) + \beta_2 u_1 + \dots + \beta_m u_m\}^r \\
&= \{(\beta_1 b_1 + \beta_2)u_1 + (\beta_1 b_2 + \beta_3)u_2 + \dots + (\beta_1 b_{m-1} + \beta_m)u_{m-1} + \beta_1 b_m u_m\}^r \quad (7.45) \text{より} \\
&= (\beta_1 u_1 + \beta_2 b_2 + \dots + \beta_m u_m)^r
\end{aligned}$$

これは $Mult(r; \beta_1, \dots, \beta_m)$ の確率母関数である。既約で非周期的マルコフ連鎖なので、多項分布 $Mult(r; \beta_1, \dots, \beta_m)$ は唯一の定常分布である。

合祖が起こる確率は $1/N$ のオーダーなので、合祖が起こるまでの平均待ち時間は N のオーダーである。従って次の合祖が起こるまでの時間に形状過程 $\{X^{(r)}(k)\}_{k=1,2,\dots}$ は定常分布に到達する。これは 7. 2 節で紹介した速い移住率を持つ Structured coalescent モデルにおいて、次の合祖が起こる前に速い移住率によって空間配置に関して定常分布に到達するという現象と同類の現象である。

$Y = (Y_1, \dots, Y_m)$ を多項分布 $Mult(r; \beta_1, \dots, \beta_m)$ に従う確率変数とすると、 X の確率母関数は $|X| = r$ のとき、 $U = (u_1 \beta_1 + \dots + u_m \beta_m)^r$ なので

$$\begin{cases} E[Y_i Y_{i+1}] = \frac{\partial^2 U}{\partial u_i \partial u_{i+1}} \Big|_{\substack{u_j=1 \\ (1 \leq j \leq m)}} = 2 \binom{r}{2} \beta_i \beta_{i+1} \quad (i=1, 2, \dots, m-1) \\ E\left[\binom{Y_i}{2}\right] = \frac{1}{2} \frac{\partial^2 U}{\partial u_i^2} \Big|_{\substack{u_j=1 \\ (1 \leq j \leq m)}} = \binom{r}{2} \beta_i^2 \end{cases} \quad (7.46)$$

この遺伝子系図過程 $X(k) = (X_1(k), \dots, X_m(k))$ はその形状の推移に多くの時間を使い形状が $x = (x_1, \dots, x_m)$ のとき、合祖率は(7.43)式より

$$\rho(x) = x_1 \sum_{i=1}^{m-1} x_{i+1} b_i + \binom{x_1}{2} \sum_{i=1}^m b_i^2 \quad (7.47)$$

x がレベル $r(|x| = r)$ のとき、これを形状過程の定常分布で平均すると r 個の祖先の状態のときの合祖率を求めることができる。(7.46)より

$$\begin{aligned}
\rho_r &= E[\rho(Y)] = \sum_{y \in S_r} \rho(y) \frac{r!}{\prod_{j=1}^m y_j!} \prod_{j=1}^m \beta_j^{y_j} = E\left[Y_1 \sum_{i=1}^{m-1} Y_{i+1} b_i + \binom{Y_1}{2} \sum_{i=1}^m b_i^2 \right] \\
&= 2 \sum_{i=1}^{m-1} \beta_i \beta_{i+1} \binom{r}{2} b_i + \beta_1^2 \binom{r}{2} \sum_{i=1}^m b_i^2 = \beta_1 \binom{r}{2} \left\{ 2 \sum_{i=1}^{m-1} \beta_{i+1} b_i + \beta_1 \sum_{i=1}^m b_i^2 \right\} \\
&= \beta_1^2 \binom{r}{2} \quad (7.48)
\end{aligned}$$

特に $m=1$ のときは、Wight-Fisher モデルになるが、このとき $\beta_1 = 1$ となり、通常の合祖

過程の合祖率 $\binom{r}{2}$ となる。

(3) 定理 7. 9 の証明

まず以下の記号を導入する。

1. $E = \{1, 2, \dots, n\}$: 祖先の数を表す空間
2. $E_N = S = S_1 \cup S_2 \cup \dots \cup S_n$, $S_r = \{x = (x_1, \dots, x_m; |x| = r)\}$
3. 関数 $\eta_N : S \rightarrow E$ 、 $x \in S$ に対して $\eta_N(x) = |x| = x_1 + \dots + x_m \in E$
4. $X_N(k), k = 0, 1, \dots$: 集団サイズ N のときの離散時間形状過程
5. $A_N(\lfloor Nt \rfloor) = \eta_N(X_N(\lfloor Nt \rfloor)) = |X_N(\lfloor Nt \rfloor)|$: N 世代を単位時間 ($t = 1$) としたときの祖先遺伝子数の過程。

$n = 2, m = 2$ の簡単な場合を例にとって定理 7. 12 を見てみよう。その推移確率は(7.40)

で与えられる。(7.40)式は $\Pi_N = A + \frac{1}{N}B$ と書けるが、定理 7. 12 に従って定常分布

$P = \lim_{k \rightarrow \infty} A^k$ を求める。

$$A = \begin{pmatrix} b_1 & b_2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & b_1^2 & 2b_1b_2 & b_2^2 \\ 0 & 0 & b_1 & b_2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} A_{11} & O \\ O & A_{22} \end{pmatrix}, \quad A_{11} = \begin{pmatrix} b_1 & b_2 \\ 1 & 0 \end{pmatrix}, \quad A_{22} = \begin{pmatrix} b_1^2 & 2b_1b_2 & b_2^2 \\ b_1 & b_2 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

これより、

$$P = \lim_{k \rightarrow \infty} A^k = \begin{pmatrix} \lim_{k \rightarrow \infty} A_{11}^k & O \\ O & \lim_{k \rightarrow \infty} A_{22}^k \end{pmatrix} = \begin{pmatrix} P_{11} & O \\ O & P_{22} \end{pmatrix}, \quad (7.49)$$

$$\text{ただし } P_{11} = \begin{pmatrix} \beta_1 & \beta_2 \\ \beta_1 & \beta_2 \end{pmatrix}, \quad P_{22} = \begin{pmatrix} \beta_1^2 & 2\beta_1\beta_2 & \beta_2^2 \\ \beta_1^2 & 2\beta_1\beta_2 & \beta_2^2 \\ \beta_1^2 & 2\beta_1\beta_2 & \beta_2^2 \end{pmatrix}$$

$$B = \begin{pmatrix} O & O \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ b_1^2 & b_2^2 & -b_1^2 & 0 & -b_2^2 \\ b_1 & 0 & -b_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad B_{21} = \begin{pmatrix} b_1^2 & b_2^2 \\ b_1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B_{22} = \begin{pmatrix} -b_1^2 & 0 & -b_2^2 \\ -b_1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$G = PBP = \begin{pmatrix} P_{11} & O \\ O & P_{22} \end{pmatrix} \begin{pmatrix} O & O \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} P_{11} & O \\ O & P_{22} \end{pmatrix} = \begin{pmatrix} O & O \\ G_{21} & G_{22} \end{pmatrix}, \quad (7.50)$$

ただし、 $G_{21} = P_{22}B_{21}P_{11} = \lambda \begin{pmatrix} \beta_1 & \beta_2 \\ \beta_1 & \beta_2 \\ \beta_1 & \beta_2 \end{pmatrix} = \lambda \tilde{P}_{11}$ とする、また $G_{22} = P_{22}B_{22}P_{22} = -\lambda P_{22}$

ここで、(7.45)より $\lambda = b_1^2\beta_1^2 + 2b_1\beta_1\beta_2 + b_2^2\beta_1^2 = \beta_1^2(b_1 + b_2)^2 + 2b_1\beta_1(\beta_2 - b_2\beta_1)^2 = \beta_1^2$ 。
定理 7. 2 よりマルコフ連鎖 $X_N(\lfloor t/N \rfloor)_{t \geq 0}$ の有限次元分布は $N \rightarrow \infty$ で、推移確率

が $\Pi(t) = P - I + e^{tG} = Pe^{tG}$ の連続時間マルコフ連鎖 $(X_t)_{t \geq 0}$ に有限次元分布収束する。

さらに祖先遺伝子の個数過程を $A_N(\lfloor Nt \rfloor) = \eta_N(X_N(\lfloor Nt \rfloor)) = |X_N(\lfloor Nt \rfloor)|$, $A(t) = |X(t)|$ とする。

$$\begin{aligned} G^k &= \begin{pmatrix} O & O \\ G_{21} & G_{22} \end{pmatrix}^k = \begin{pmatrix} O & O \\ G_{22}^{k-1}G_{21} & G_{22}^k \end{pmatrix} = \begin{pmatrix} O & O \\ (-1)^{k-1}\lambda^k P_{22}\tilde{P}_{11} & (-\lambda)^k P_{22} \end{pmatrix} \\ &= (-\lambda)^k \begin{pmatrix} O & O \\ -\tilde{P}_{11} & P_{22} \end{pmatrix} = (-\lambda)^k \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -\beta_1 & -\beta_2 & \beta_1^2 & 2\beta_1\beta_2 & \beta_2^2 \\ -\beta_1 & -\beta_2 & \beta_1^2 & 2\beta_1\beta_2 & \beta_2^2 \\ -\beta_1 & -\beta_2 & \beta_1^2 & 2\beta_1\beta_2 & \beta_2^2 \end{pmatrix} \end{aligned} \quad (7.51)$$

より、 $k \geq 1$ のとき $G^k = (-\lambda)^k \tilde{G}$ と表すことにする。 $G^0 = I$ (単位行列) である。

$$\begin{aligned} \Pi(t) &= P - I + e^{tG} = P - I + \sum_{k=0}^{\infty} \frac{(-\lambda t)^k}{k!} G^k = P - I + \left(I + \left(\sum_{k=1}^{\infty} \frac{(-\lambda t)^k}{k!} \right) \tilde{G} \right) \\ &= P + (e^{-\lambda t} - 1) \tilde{G} = \begin{pmatrix} \beta_1 & \beta_2 & 0 & 0 & 0 \\ \beta_1 & \beta_2 & 0 & 0 & 0 \\ \beta_1(1 - e^{-\lambda t}) & \beta_2(1 - e^{-\lambda t}) & \beta_1^2 e^{-\lambda t} & 2\beta_1\beta_2 e^{-\lambda t} & \beta_2^2 e^{-\lambda t} \\ \beta_1(1 - e^{-\lambda t}) & \beta_2(1 - e^{-\lambda t}) & \beta_1^2 e^{-\lambda t} & 2\beta_1\beta_2 e^{-\lambda t} & \beta_2^2 e^{-\lambda t} \\ \beta_1(1 - e^{-\lambda t}) & \beta_2(1 - e^{-\lambda t}) & \beta_1^2 e^{-\lambda t} & 2\beta_1\beta_2 e^{-\lambda t} & \beta_2^2 e^{-\lambda t} \end{pmatrix} \end{aligned}$$

$$\Pi(t) \text{ の第 1,2 列を加えると } \begin{pmatrix} 1 \\ 1 \\ 1-e^{-\lambda} \\ 1-e^{-\lambda} \end{pmatrix}, \text{ 第 3,4,5 列を加えると } \begin{pmatrix} 0 \\ 0 \\ e^{-\lambda} \\ e^{-\lambda} \\ e^{-\lambda} \end{pmatrix} \text{ となるが、}$$

これは $A(t)$ の推移確率が $P(A(t)=1|A(0)=1)=1$, $P(A(t)=1|A(0)=2)=1-e^{-\lambda}$,

$P(A(t)=2|A(0)=2)=e^{-\lambda}$ より、標準的な合祖過程に一致することを示している。

一般のサンプル数の場合については、少々計算が煩雑になるので概要だけを紹介する。まず離散時間モデルで 1 世代の推移確率 Π_N は次の形にかける。

$$\Pi_N = A + \frac{1}{N}B + O\left(\frac{1}{N^2}\right), \text{ ここで行列 } A \text{ 及び } B \text{ は次の形になる。}$$

$$A = \begin{pmatrix} A_{11} & O & O & \cdots & O & O \\ O & A_{22} & O & \cdots & O & O \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \vdots & \vdots \\ O & O & O & \cdots & A_{n-1,n-1} & O \\ O & O & O & \cdots & O & A_{n,n} \end{pmatrix},$$

$$B = \begin{pmatrix} -B_{11} & O & O & \cdots & O & O & O \\ B_{21} & -B_{22} & O & \cdots & O & O & O \\ O & \cdots & \cdots & & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ O & O & O & \cdots & B_{n-1,n-1} & -B_{n-1,n-1} & O \\ O & O & O & \cdots & O & B_{n,n-1} & -B_{n,n} \end{pmatrix}$$

小行列 $A_{i,i}$ は合祖を無視したレベル i の形状過程の推移確率、 $B_{i,j}$ は単一の合祖による推移

に関する行列である。一般に行列 C に対して $C(x,y)$ で第 (x,y) 成分を、 $C_{i,j}(x,y)$ で小行

列 $C_{i,j}$ の第 (x,y) 成分を表すことにする。(7.47)より $\sum_{y \in S_{i-1}} B_{i,i-1}(x,y) = \sum_{y \in S_i} B_{i,i}(x,y) = \rho(x)$ と

なる。

レベル i の状態の数を $d_i = |S_i|$ とすると、 $d_i = {}_m H_i = \binom{i+m-1}{m-1}$ 。

$A_{i,i}$ および $B_{i,i}$ は $d_i \times d_i$ 行列、 $B_{i,i-1}$ は $d_i \times d_{i-1}$ 行列である。推移確率行列 Π_N は $d \times d$ 行列

(ただし、 $d = |S| = \sum_{i=1}^n \binom{i+m-1}{m-1} = \binom{m+n}{m} - 1$) である。補題 7. 1 より、

$$\lim_{N \rightarrow \infty} \Pi_N^{[N]} = \lim_{N \rightarrow \infty} \left(A + \frac{1}{N} B + O\left(\frac{1}{N^2}\right) \right)^{[N]} = P - I + e^G$$

$$\text{ここで } P = \lim_{k \rightarrow \infty} A^k = \begin{pmatrix} \lim_{k \rightarrow \infty} A_{11}^k & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & \lim_{k \rightarrow \infty} A_{mm}^k \end{pmatrix} = \begin{pmatrix} P_{11} & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & P_m \end{pmatrix}$$

$P_{ii} = \lim_{k \rightarrow \infty} A_{ii}^k$ は推移確率 A_{ii} の定常分布であり、その行ベクトルは S_i の要素を適当な順に

並べて $y^{(1)}, \dots, y^{(d_i)}$ とすると、 $M_i = (M_i(y^{(1)}), \dots, M_i(y^{(d_i)}))$, (ただし $M_i(y)$ は多項分布

$Mult(i, \beta_1, \dots, \beta_m)$) となる。(7.48)より、 $\sum_{y \in S_i} M_i(y) \rho(y) = \beta_1^2 \binom{i}{2}$ が成り立つ。

$$\text{さらに } G = PBP = \begin{pmatrix} -G_{11} & O & O & \cdots & O & O \\ G_{21} & -G_{22} & O & \cdots & O & O \\ O & \cdots & \cdots & \cdots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & \cdots & G_{n-1,n-2} & -G_{n-1,n-1} & O \\ O & O & \cdots & O & G_{n,n-1} & -G_{n,n} \end{pmatrix}$$

ただし $G_{i,i} = P_{i,i} B_{i,i} P_{i,i}$, $G_{i,i-1} = P_{i,i} B_{i,i-1} P_{i-1,i-1}$ 。

また小行列 $G_{i,i}$ の (x, y) 成分は

$$G_{i,i}(x,y) = \sum_{u \in \mathcal{S}_i} \sum_{v \in \mathcal{S}_i} P_{i,i}(x,u) B_{i,i}(u,v) P_{i,i}(v,y) = \left(\sum_{u \in \mathcal{S}_i} P_{i,i}(x,u) \left(\sum_{v \in \mathcal{S}_i} B_{i,i}(u,v) \right) \right) M_i(y)$$

$$= \left(\sum_{u \in \mathcal{S}_i} P_{i,i}(x,u) \rho(u) \right) M_i(y) = \beta_1^2 \binom{i}{2} M_i(y)$$

よって $G_{i,i} = \beta_1^2 \binom{i}{2} P_{i,i}$ となる。

同様に $G_{i,i-1} = \beta_1^2 \binom{i}{2} \tilde{P}_{i,i-1}$ と書ける、ただし $\tilde{P}_{i,i-1}$ は行列 $P_{i-1,i-1}$ の同一な行ベクトルを i

個並べた $d_i \times d_{i-1}$ 行列である。 $G_{i,i}, G_{i,i-1}$ はいずれも x に依存せず、 $i = |x|$ のみに依存することに注意しよう。

(7.41) で定義される合祖過程の生成作用素を Q とすると、 $|x| = i$ のとき

$$\sum_{y \in \mathcal{S}_j} (A(t))(x,y) = \left(\lim_{N \rightarrow \infty} A_N([Nt]) \right) (x,y) = (e^{Qt})(i,j) \text{ が成り立ち、 } i\text{-合祖過程に収束する。}$$

(詳細は Kaj et al(2001) を参照)

また、第 7. 1 節の最後に述べたように、Ethier and Kurtz(1986) の p173(Theorem 2.12) により $E = \{1, 2, \dots, n\}$ 上のマルコフ連鎖 $A_N([Nt])$ は生成作用素 Q の連続時間マルコフ連鎖 $A(t)$ に空間 $D_E[0, \infty)$ 上で弱収束 (分布収束) する。

7. 6 収束定理の証明

Möhle の補題 7. 1、定理 7. 2 は次のように拡張できる。この節では拡張された収束定理とその証明を与える。

定理 7. 1 1 (Möhle and Notohara(2016))

(1) $(c_N)_{N \in \mathbb{N}}, (d_N)_{N \in \mathbb{N}}$ を $\lim_{N \rightarrow \infty} c_N = \lim_{N \rightarrow \infty} d_N = 0, \lim_{N \rightarrow \infty} \frac{c_N}{d_N} = 0$ を満たす二つの実数列とする。

$Q = (q_{i,j})$ を $d \times d$ の生成行列で $P = \lim_{m \rightarrow \infty} (I + d_N Q)^m$ および $(B_N)_{N \in \mathbb{N}}$ を $d \times d$ 行列で

$\|B_N\| \leq 1$ かつ $G = \lim_{m \rightarrow \infty} P B_N P$ が存在すると仮定する。このとき、すべての $t > 0$ に対して

$$\lim_{N \rightarrow \infty} (I + d_N Q + c_N B_N)^{\lfloor t/c_N \rfloor} = P - I + e^{tG} = P e^{tG} = e^{tG} P。$$

(2) $X_N = (X_N(r))_{r \in \mathbb{N}_0}$ を 1 ステップ推移確率が $\Pi_N = I + d_N Q + c_N B_N$ の有限の状態空間 S

上の斉時有限マルコフ連鎖とする。(1)の条件を満たすとき、初期分布 $P_{X_N(0)}$ がある

確率分布 μ に分布収束するならば、マルコフ連鎖 $X_N(\lfloor t/c_N \rfloor)_{t \geq 0}$ の有限次元分布は初期分布が $X_0^d = \mu$ 、推移確率が $\Pi(t) = P - I + e^{tG} = Pe^{tG}$ の連続時間マルコフ連鎖 $(X_t)_{t \geq 0}$ に有限次元分布収束する。

この定理は補題 7. 1 (Möhle(1998))の行列 A を行列 $A_N = I + d_N Q$ に置き換えたものとなっている。Möhle(1998)の証明を少し修正することにより得られる。

幾つかの補題を準備する。

補題 7. 1 2

$d \times d$ 行列 A について、極限 $\lim_{m \rightarrow \infty} A^m = P$ が存在するならば、 $\lim_{t \rightarrow \infty} e^{t(A-I)} = P$ 。

(証明) 任意の $\varepsilon > 0$ に対して、ある自然数 m_0 が存在して、 $m > m_0$ ならば $\|A^m - P\| < \varepsilon$ 。

$$\begin{aligned} \|e^{t(A-I)} - P\| &= \left\| e^{-t} \sum_{m=0}^{\infty} \frac{t^m}{m!} A^m - P \right\| = e^{-t} \left\| \sum_{m=0}^{\infty} \frac{t^m}{m!} (A^m - P) \right\| \\ &\leq e^{-t} \sum_{m=0}^{m_0} \frac{t^m}{m!} \|A^m - P\| + e^{-t} \sum_{m=m_0+1}^{\infty} \frac{t^m}{m!} \|A^m - P\| \\ &\leq \left(\sup_{0 \leq m \leq m_0} \|A^m - P\| \right) e^{-t} \left(\sum_{m=0}^{m_0} \frac{t^m}{m!} \right) + \varepsilon \rightarrow \varepsilon \quad (t \rightarrow \infty \text{ のとき}) \end{aligned}$$

ε は任意なので $\lim_{t \rightarrow \infty} e^{t(A-I)} = P$ 。

補題 7. 1 3

(1) 実数 $\delta > 0$ 、 $d \times d$ 行列 Q について $P = \lim_{m \rightarrow \infty} (I + \delta Q)^m$ が存在するならば $\lim_{t \rightarrow \infty} e^{tQ} = P$ 。

(2) 実数 $\delta_1, \delta_2 > 0$ 、 $d \times d$ 行列 Q について $P_1 = \lim_{m \rightarrow \infty} (I + \delta_1 Q)^m$ 、 $P_2 = \lim_{m \rightarrow \infty} (I + \delta_2 Q)^m$ が存在するならば、 $P_1 = P_2$ 。

(証明)

(1) 補題 7. 1 2 において、 $A = I + \delta Q$ と置くと、 $\lim_{t \rightarrow \infty} e^{tQ} = \lim_{s=t\delta \rightarrow \infty} e^{sQ} = P$ 。

(2)(1)より $P_1 = \lim_{t \rightarrow \infty} e^{tQ} = P_2$ 。

補題 7. 1 4

定理 7. 1 1 と同じ条件下で、ある自然数列 $(M_N)_{N \in \mathbb{N}}$ が存在して、 $\lim_{N \rightarrow \infty} M_N c_N = 0$,

$\lim_{N \rightarrow \infty} \sup_{m \geq M_N} \|(I + d_N Q)^m - P\| = 0$ が成り立つ。

(証明)

正の整数数列 $(M_N)_{N \in \mathbb{N}}$ を $\lim_{N \rightarrow \infty} M_N d_N = \infty$, $\lim_{N \rightarrow \infty} M_N c_N = 0$, $\lim_{N \rightarrow \infty} M_N d_N^2 = 0$ を満たすように

選ぶ。例えば $M_N = 1 + \left\lfloor \min(1/\sqrt{c_N d_N}, 1/\sqrt{d_N^3}) \right\rfloor$ と選べば条件を満たす。 $A_N = I + d_N Q$ と

すると、 Q は生成作用素で、 $\lim_{N \rightarrow \infty} d_N = 0$ なので、ある自然数 N_0 より大のすべての自然数

$N > N_0$ に対して、 $\|A_N\| = 1$ 。 $N > N_0$ とすると、自然数 m に対して

$\|A_N^{m+1} - P\| = \|(A_N^m - P)A_N\| \leq \|A_N^m - P\| \|A_N\| \leq \|A_N^m - P\|$ 、すなわち m とともに単調減少する。

従って $\lim_{N \rightarrow \infty} \|A_N^{M_N} - P\| = 0$ を示せば十分である。

$\|A_N^{M_N} - P\| \leq \|A_N^{M_N} - \exp(M_N d_N Q)\| + \|\exp(M_N d_N Q) - P\|$ なので $\lim_{N \rightarrow \infty} M_N d_N = \infty$ と補題 7.

1 3 より右辺第 2 項 $\lim_{N \rightarrow \infty} \|\exp(M_N d_N Q) - P\| = 0$ 。また、 $\|A_N\| = 1$, $\|\exp(d_N Q)\| = 1$ より右辺

第 1 項は

$$\|A_N^{M_N} - (e^{d_N Q})^{M_N}\| \leq M_N \|A_N - e^{d_N Q}\| = M_N \left\| \sum_{m=2}^{\infty} (d_N Q)^m / m! \right\| = M_N O(d_N^2) \rightarrow 0 \quad (N \rightarrow \infty)$$

(証明終わり)

[定理 7. 1 1 の証明]:

$A_N = I + d_N Q$ 、補題 7. 1 7 より $\lim_{m \rightarrow \infty} A_N^m = P$ とする。また、 $n = \lfloor t/c_N \rfloor$ とする。

$$\|(A_N + c_N B_N)^n - (P + c_N B_N)^n\| \leq \sum_{k=0}^n c_N^k \sum_{\substack{m_1, \dots, m_{k+1} \in \mathbb{N}_0 \\ m_1 + \dots + m_{k+1} = n-k}} \left\| A_N^{m_1} \prod_{j=2}^{k+1} B_N A_N^{m_j} - P^{m_1} \prod_{j=2}^{k+1} B_N P^{m_j} \right\|$$

補題 7. 14 より、任意の $\varepsilon > 0$ に対して、ある自然数列 $(M_N)_{N \in \mathbb{N}}$ が存在して、

$$\lim_{N \rightarrow \infty} M_N c_N = 0, \quad \|A_N^m - P\| < \varepsilon \text{ を満たす } (M_N)_{N \in \mathbb{N}} \text{ を取ることができる。}$$

全ての $j \in \{1, 2, \dots, k+1\}$ について $m_j \geq M_N$ ならば

$$\begin{aligned} & \left\| A_N^{m_1} B_N A_N^{m_2} B_N \cdots B_N A_N^{m_{k+1}} - P^{m_1} B_N P^{m_2} B_N \cdots B_N P^{m_{k+1}} \right\| \\ &= \left\| \left(A_N^{m_1} B_N A_N^{m_2} B_N \cdots B_N A_N^{m_{k+1}} - P^{m_1} B_N A_N^{m_2} B_N \cdots B_N A_N^{m_{k+1}} \right) + \right. \\ & \quad \left(P^{m_1} B_N A_N^{m_2} B_N \cdots B_N A_N^{m_{k+1}} - P^{m_1} B_N P^{m_2} B_N \cdots B_N A_N^{m_{k+1}} \right) + \\ & \quad \left. \cdots + \left(P^{m_1} B_N P^{m_2} B_N \cdots B_N A_N^{m_{k+1}} - P^{m_1} B_N P^{m_2} B_N \cdots B_N P^{m_{k+1}} \right) \right\| \\ &\leq \left\| (A_N^{m_1} - P^{m_1}) B_N A_N^{m_2} \cdots B_N A_N^{m_{k+1}} \right\| + \left\| P^{m_1} B_N (A_N^{m_2} - P^{m_2}) B_N \cdots B_N A_N^{m_{k+1}} \right\| + \cdots + \left\| P^{m_1} B_N \cdots (A_N^{m_{k+1}} - P^{m_{k+1}}) \right\| \\ &\leq \sum_{j=1}^{m_{k+1}} \|A_N^{m_j} - P^{m_j}\| = \sum_{j=1}^{k+1} \|A_N^{m_j} - P\| < (k+1)\varepsilon \end{aligned}$$

その他の項については

$$\begin{aligned} & \left\| A_N^{m_1} B_N A_N^{m_2} B_N \cdots B_N A_N^{m_{k+1}} - P^{m_1} B_N P^{m_2} B_N \cdots B_N P^{m_{k+1}} \right\| \\ &\leq \left\| A_N^{m_1} B_N A_N^{m_2} B_N \cdots B_N A_N^{m_{k+1}} \right\| + \left\| P^{m_1} B_N P^{m_2} B_N \cdots B_N P^{m_{k+1}} \right\| \leq 1 + 1 = 2 \end{aligned}$$

$$\text{これより、} \left\| (A_N + c_N B_N)^n - (P + c_N B_N)^n \right\| \leq \|A_N^n - P\| + S_1 + S_2$$

$$\begin{aligned} S_1 &= \sum_{k=1}^n c_N^k \sum_{\substack{m_1, \dots, m_{k+1} \geq M_N \\ m_1 + \dots + m_{k+1} = n-k}} (k+1)\varepsilon \leq \varepsilon \sum_{k=0}^n (k+1) c_N^k \sum_{\substack{m_1, \dots, m_{k+1} \in \mathbb{N}_0 \\ m_1 + \dots + m_{k+1} = n-k}} 1 = \varepsilon \sum_{k=0}^n (k+1) \binom{n}{k} c_N^k \\ &= \varepsilon \sum_{k=0}^n k \binom{n}{k} c_N^k + \varepsilon \sum_{k=0}^n \binom{n}{k} c_N^k = \varepsilon n c_N \sum_{k=1}^n \binom{n-1}{k-1} c_N^{k-1} + \varepsilon \sum_{k=0}^n \binom{n}{k} c_N^k \\ &= \varepsilon \{ n c_N (1 + c_N)^{n-1} + (1 + c_N)^n \} \sim \varepsilon e^t (t+1) \end{aligned}$$

$$S_2 = \sum_{k=1}^n c_N^k \sum_{\substack{m_1, \dots, m_{k+1} \in \mathbb{N}_0 \\ m_1 + \dots + m_{k+1} = n-k \\ \exists j \text{ with } m_j < M_N}} 2 \leq \sum_{k=1}^n c_N^k (k+1) \sum_{\substack{m_1, \dots, m_{k+1} \in \mathbb{N}_0 \\ m_1 + \dots + m_{k+1} = n-k \\ m_{k+1} < M_N}} 2 \leq \sum_{k=1}^n c_N^k (k+1) \sum_{\substack{m_{k+1}=0 \\ m_1, \dots, m_{k+1} \in \mathbb{N}_0 \\ m_1 + \dots + m_{k+1} = n-m_{k+1}-k}}^{M-1} 2$$

異なる k 種のものから重複を許して n 個取り出す重複組み合わせを ${}_k H_n$ とすると

$$\sum_{\substack{m_1, \dots, m_{k+1} \in \mathbb{N}_0 \\ m_1 + \dots + m_{k+1} = n-m_{k+1}-k}} 2 = 2 {}_k H_{n-m_{k+1}-k} = 2 \binom{n-m_{k+1}-1}{n-m_{k+1}-k} = 2 \binom{n-m_{k+1}-1}{k-1} \text{ より上式は}$$

$$\begin{aligned}
\text{予式} &= \sum_{k=1}^n c_N^k (k+1) \sum_{m_{k+1}=0}^{M-1} 2 \binom{n-m_{k+1}-1}{k-1} \leq \sum_{k=1}^n c_N^k (k+1) \sum_{m_{k+1}=0}^{M-1} 2 \binom{n-1}{k} \\
&\leq 2M_N \sum_{k=1}^n (k+1) \binom{n-1}{k-1} c_N^k = 2M_N c_N \sum_{k=0}^{n-1} (k+2) \binom{n-1}{k} c_N^k \\
&= 2M_N c_N \left\{ (n-1)c_N(1+c_N)^{n-2} + 2(1+c_N)^{n-1} \right\} \sim 2M_N c_N e^t (t+2) = O(c_N)
\end{aligned}$$

以上より $\lim_{N \rightarrow \infty} \left\| (A_N + c_N B_N)^{\lfloor t/c_N \rfloor} - (P + c_N B_N)^{\lfloor t/c_N \rfloor} \right\| = 0$ が示された。

次に $\lim_{N \rightarrow \infty} (P + c_N B_N)^{\lfloor t/c_N \rfloor} = P - I + e^G$ を示す。

$$\begin{aligned}
&(P + c_N B_N)^n - P + I - e^G \\
&= P^n + \sum_{k=1}^n c_N^k \left(\sum_{\substack{m_1, \dots, m_{k+1} \in \mathbb{N}_0 \\ m_1 + \dots + m_{k+1} = n-k}} P^{m_1} \prod_{j=2}^{k+1} B_N P^{m_j} \right) - P + I - \sum_{k=0}^{\infty} \frac{t^k}{k!} G^k \\
&= \sum_{k=1}^n \left(c_N^k \sum_{\substack{m_1, \dots, m_{k+1} \in \mathbb{N}_0 \\ m_1 + \dots + m_{k+1} = n-k}} P^{m_1} \prod_{j=2}^{k+1} B_N P^{m_j} - \frac{t^k}{k!} G^k \right) - \sum_{k=n+1}^{\infty} \frac{t^k}{k!} G^k \\
&= \sum_{k=1}^n \left(c_N^k \sum_{\substack{m_1, \dots, m_{k+1} \in \mathbb{N}_0 \\ m_1 + \dots + m_{k+1} = n-k}} \left(P^{m_1} \prod_{j=2}^{k+1} B_N P^{m_j} - (PB_N P)^k \right) \right) \\
&\quad + \sum_{k=1}^n \left(c_N^k \sum_{\substack{m_1, \dots, m_{k+1} \in \mathbb{N}_0 \\ m_1 + \dots + m_{k+1} = n-k}} (PB_N P)^k - \frac{t^k}{k!} G^k \right) - \sum_{k=n+1}^{\infty} \frac{t^k}{k!} G^k
\end{aligned}$$

$P^2 = P$ なので、全ての $j \in \{1, \dots, k+1\}$ について $m_j \geq 1$ ならば

$P^{m_1} B_N P^{m_2} \dots B_N P^{m_{k+1}} = (PB_N P)^k$ と書けるので

$$\begin{aligned}
(P + c_N B_N)^n - P + I - e^G &= \sum_{k=1}^n \left(c_N^k \sum_{\substack{m_1, \dots, m_{k+1} \in \mathbb{N}_0 \\ m_1 + \dots + m_{k+1} = n-k \\ \exists j \text{ with } m_j = 0}} \left(P^{m_1} B_N P^{m_2} \dots B_N P^{m_{k+1}} - (PB_N P)^k \right) \right) \\
&\quad + \sum_{k=1}^n \left(c_N^k \binom{n}{k} (PB_N P)^k - \frac{t^k}{k!} G^k \right) - \sum_{k=n+1}^{\infty} \frac{t^k}{k!} G^k
\end{aligned}$$

最初の項のノルムは次の式で抑えられる

$$\begin{aligned} & \left\| \sum_{k=1}^n c_N^k \sum_{\substack{m_1, \dots, m_{k+1} \in \mathbb{N}_0 \\ m_1 + \dots + m_{k+1} = n-k \\ \exists j \text{ with } m_j = 0}} (P^{m_1} B_N P^{m_2} \dots B_N P^{m_{k+1}} - (PB_N P)^k) \right\| \leq \sum_{k=1}^n c_N^k \sum_{\substack{m_1, \dots, m_{k+1} \in \mathbb{N}_0 \\ m_1 + \dots + m_{k+1} = n-k \\ \exists j \text{ with } m_j = 0}} 2 \\ & \leq \sum_{k=1}^n c_N^k (k+1) \sum_{\substack{m_1, \dots, m_{k+1} \in \mathbb{N}_0 \\ m_1 + \dots + m_{k+1} = n-k}} 2 = 2 \sum_{k=1}^n (k+1) \binom{n-1}{k-1} c_N^k = 2c_N \sum_{k=0}^{n-1} (k+2) \binom{n-1}{k} c_N^k \\ & = 2c_N \left((n-1)c_N (1+c_N)^{n-2} + 2(1+c_N)^{n-1} \right) \sim 2c_N e^t (t+2) = O(c_N) \end{aligned}$$

最後の項 $\sum_{k=n+1}^{\infty} \frac{t^k}{k!} G^k$ は e^G のテーラー展開の末尾部分なので $N \rightarrow \infty$ のとき 0 に収束する。

最後に $\lim_{N \rightarrow \infty} \sum_{k=1}^n \left(c_N^k \binom{n}{k} (PB_N P)^k - \frac{t^k}{k!} G^k \right) = 0$ を示す。

明らかに $c_N^k (n)_k \leq (c_N n)^k \leq t^k$ 、また $t - c_N = c_N(t/c_N - 1) < c_N n$ より

$$\begin{aligned} c_N^k (n)_k &= \prod_{i=0}^{k-1} (c_N(n-i)) > \prod_{i=0}^{k-1} (t - c_N(i+1)) = \prod_{i=1}^k (t - c_N i) \geq t^k - t^{k-1} c_N (1 + \dots + k) \\ &= t^k - t^{k-1} c_N k(k+1)/2 = t^k - (c_N/2) k(k+1) t^{k-1} \end{aligned}$$

$t^k - \frac{c_N}{2} k(k+1) t^{k-1} < c_N^k (n)_k < t^k$ なので、 $-\frac{c_N}{2} \frac{k+1}{(k-1)!} t^{k-1} < c_N^k \binom{n}{k} - \frac{t^k}{k!} < 0$ が成り立つ。

以上より $\left| c_N^k \binom{n}{k} - \frac{t^k}{k!} \right| < \frac{c_N}{2} \frac{k+1}{(k-1)!} t^{k-1}$ が成り立つ。

$$\begin{aligned} & \left\| \sum_{k=1}^n \left(c_N^k \binom{n}{k} (PB_N P)^k - \frac{t^k}{k!} G^k \right) \right\| \leq \sum_{k=1}^n \left| c_N^k \binom{n}{k} - \frac{t^k}{k!} \right| \|(PB_N P)^k\| + \sum_{k=1}^n \frac{t^k}{k!} \|(PB_N P)^k - G^k\| \\ & \leq \sum_{k=1}^n \left| c_N^k \binom{n}{k} - \frac{t^k}{k!} \right| + \sum_{k=1}^n \frac{t^k}{k!} k \|PB_N P - G\| \\ & \leq \sum_{k=1}^n \frac{c_N}{2} (k+1) \frac{t^{k-1}}{(k-1)!} + \|PB_N P - G\| \sum_{k=1}^n \frac{t^k}{(k-1)!} \\ & \leq \frac{c_N}{2} \sum_{k=0}^{n-1} (k+2) \frac{t^k}{k!} + \|PB_N P - G\| \sum_{k=0}^{n-1} \frac{t^{k+1}}{k!} \\ & \leq \frac{c_N}{2} (t+2) e^t + \|PB_N P - G\| t e^t \end{aligned}$$

上式は $N \rightarrow \infty$ で 0 に収束する。

(証明終わり)

これまでの議論は有限の状態空間上のマルコフ連鎖であったが、これを可算無限次元まで拡張することが可能である。状態空間を自然数全体 $N = \{1, 2, \dots\}$ とし、バナッハ空間

$\ell^\infty = \left\{ x = (x_i)_{i \in N} \in R^N : \|x\| = \sup_{i \in N} |x_i| < \infty \right\}$ に対して、 ℓ^∞ から ℓ^∞ への線形作用素を

$A = (a_{ij})_{i, j \in N}$ とする。すなわち $x = (x_i)_{i \in N} \in \ell^\infty$ に対して $(Ax)_i = \sum_{j \in N} a_{ij} x_j$ 、

ノルム $\|A\| = \sup_{i \in N} \sum_{j \in N} |a_{ij}| < \infty$ を仮定する。このような線形作用素の集合を

$L = \left\{ A = (a_{ij})_{i, j \in N} : \|A\| < \infty \right\}$ とする。空間 L は完備な距離空間であり、 $A, B \in L$ に対して

$\|AB\| \leq \|A\| \|B\|$ を満たす。 $A \in L$ に対して、指数関数 $e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}$ を有限次るときと同様に定

義できる。従って定理 7-1-1 の有限次行列は全て可算無限次の作用素 L の行列に置き換えても成り立つ。

7. 2 節から 7. 5 節で紹介した例は Möhle(1998)の定理を応用したものであり、毎世代 N に依存しない一定の確率で推移を起こす現象と世代当たり c_N の率で起こる現象を含む離散時間モデルの収束を扱っている。拡張された定理 7. 1 6 では毎世代 c_N よりも大きな率 d_N で起こる現象という形で一般化されている。7. 2 節から 7. 5 節で紹介した例はいずれも定理 7. 1 1 により拡張することができる。例えば 7. 2 速い移住率を持つ Structured coalescent モデルは 1 世代当たりの移住率を $1/N^\alpha$ (ただし $0 \leq \alpha < 1$) のオーダーとすると、このモデルは Nordborg and Krone(2001)で扱われたモデルとなる。彼らは収束については厳密に議論していないが、このモデルの連続時間マルコフ連鎖への収束が定理 7. 1 5 で示されたことになる。