3．遺伝子系図の合祖過程

集団遺伝の古典的理論として遺伝子頻度の変化を記述するマルコフ連鎖モデルおよび、拡散過程モデルについて紹介してきた。近年集団遺伝学においては遺伝子の系図を表現する合祖過程(Coalescent Process)と呼ばれる確率モデルが導入され広範に研究されている。この章ではこの遺伝子の系図に関するマルコフ過程モデルについて紹介する。

3．1　合祖モデル

近交係数(inbreeding coefficient)、同祖的 (identical by descent)など遺伝子系図に関連する概念は古くから集団遺伝学では用いられている。Felsenstein(1971)は遺伝子系図の視点からN個の半数体生物から成る有限集団の遺伝的変異の減衰速度を論じた。突然変異および自然選択は無いものと仮定しよう。集団からランダムに選んだr個の個体（遺伝子）がちょうどs個の異なる親（親遺伝子）を持つ確率をG_{rs}とする。当然$r<s$のとき$G_{rs}=0$である。第t世代にランダムに取り出したk個の遺伝子が異なるj種のアレルを含む確率を$P_t(k,j)$とすると、１世代遡ることにより漸化式$P_t(k,j)=\sum_{s=1}^{N}G_{ks}P_{t-1}(s,j)$を得る。

$N\times N$行列を$G=(G_{rs})$, $P(t)=(P_t(k,j))$とすると、$P(t)=GP(t-1)$と表される。これより、$P(t)=G^tP(0)$を得る。ここで行列の積G^tの(k,j)成分はk個のサンプル遺伝子がt世代遡った祖先集団でj個の異なる祖先に由来する確率を表す。すなわちサンプル遺伝子の系図を表すマルコフ連鎖の推移行列であることに注意しよう。行列Gの性質を調べることにより、Felsenstein は十分大きなtに対して漸近的に$P_t(N,k)\sim A_kG_{kk}{}^t$（$A_k$は定数）が成り立つことを示し、全集団から$k$種のアレルの遺伝的多様性が失われる率が漸近的に毎世代G_{kk}で表されることを証明した。集団の遺伝的多様性の消失速度については第３．３．２節で詳しく論じる。さらに、Gladstien(1978)は可換モデルについて詳しい研究をしているが、ここでは簡単に次の定理を紹介しておこう。

定理３．１

可換モデルにおいては上記の確率G_{rs}は次式で与えられる。

$$G_{rs}=\left[\binom{N}{s}\Big/\binom{N}{r}\right]\sum_{R\in A(r,s)}E\left[\prod_{i=1}^{s}\binom{\theta_i}{r_i}\right]$$

ただし、$A(r,s)=\{R=(r_1,r_2,...,r_s)\,;\,r_1,..,r_s>0,\ \ r_1+...+r_s=r\}$、

$\sum_{R\in A(r,s)}$ は$A(r,s)$に属す全ての$R=(r_1,..,r_s)$についての和を表す。

（証明） s個の親から生まれたr個のサンプル中でi番目の親から生まれた子供の数を$r_i\ (1\le i\le s)$で表す。G_{rs}はランダムに選んだr個の遺伝子がs個の親遺伝子を持つ確率なので、まず全集団からランダムにr個の遺伝子を選ぶ方法が$\binom{N}{r}$通りあり、前の世代からs個の親個体を選ぶ方法は$\binom{N}{s}$通り。さらに、これらの間に親子関係の結び方を考えると、i番目の親が生む子供の数はθ_i、その中からr_i個の子供を選ぶ方法が各々独立に$\binom{\theta_i}{r_i}$通りあるので、可換モデルに従って$(\theta_1,..,\theta_s)$に関する確率分布で期待値を取ると定理の結果を得る。

幾つかのモデルで具体例を計算してみよう。

例１　ライト・フィッシャーモデル

$$G_{kj}=\frac{N(N-1).....(N-j+1)}{N^k}S_k^{(j)} \tag{3.1}$$

$S_k^{(j)}$は第２種スターリング数で異なるk個のものをj個のブロックに分ける場合の数を表す（詳しくは付録 (C)参照)。

例２　モランモデル

$$G_{kk}=1-\frac{k(k-1)}{N^2}=1-G_{k,k-1} \tag{3.2}$$

Kingman(1982a,b,c)は可換モデルにおいてある条件の下で、集団のサイズNを無限大にする拡散近似と同じ時間スケールの極限操作により、上記の遺伝子系図のプロセスが合祖過程(Coalescent Process)と呼ばれている連続時間のマルコフ連鎖（死滅過程）に収束することを示した。同じ合祖過程モデルはほとんど同時期に田嶋(Tajima(1983))によっても発表されている。

まず、可換モデルの一つライト・フィッシャーモデルを例にとってみよう。

N個の半数体生物から成る集団を考えよう。この集団から２個の遺伝子をサンプルしたとき、異なる親を持つ確率$P(2,2)$を考えると、ある一つの遺伝子の親はN通りあるので$P(2,2)=\frac{N(N-1)}{N^2}=1-\frac{1}{N}$。さらに$\tau$世代遡っても異なる祖先に由来する確率を$P_\tau(2,2)$

とすると、$P_\tau(2,2)=\left(1-\frac{1}{N}\right)^\tau$ となる。N 世代を単位時間 $t=1$ とする時間スケールを取り

$[Nt]$ 世代遡っても異なる祖先に由来する確率を $P_t^N(2,2)$ とすると $P_t^N(2,2)=\left(1-\frac{1}{N}\right)^{[Nt]}$

さらに $N\to\infty$ とすると、$P_t(2,2)=\lim_{N\to\infty}\left(1-\frac{1}{N}\right)^{[Nt]}=\exp(-t)$、平均 1 の指数分布に従う。

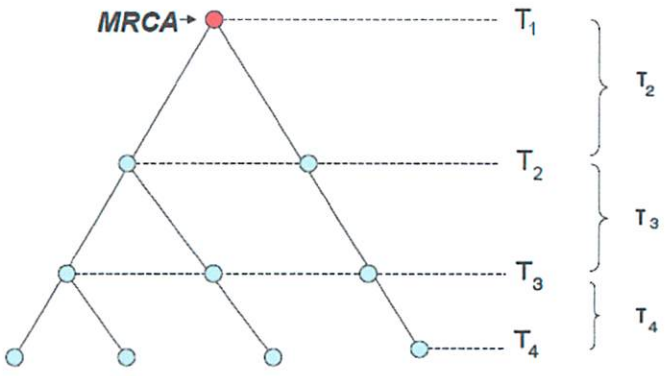
より一般に n 個のサンプル遺伝子が k 個の親に由来する確率は(3.1)より

$$P(n,k)=G_{n,k}=\frac{N(N-1)\cdots(N-k+1)}{N^n}S_n^{(k)}$$
$$=\begin{cases}\frac{n(n-1)}{2N}+O\left(\frac{1}{N^2}\right) & if\ k=n-1\\ 1-\frac{n(n-1)}{2N}+O\left(\frac{1}{N^2}\right) & if\ k=n\\ O\left(\frac{1}{N^2}\right) & if\ k\le n-2\end{cases} \qquad (3.3)$$

N 個のサンプル遺伝子が $[Nt]$ 世代遡っても異なる n 個の祖先をもつ確率 $P_t^N(n,n)$ は

$P_t^N(n,n)=\left(1-\frac{n(n-1)}{2N}+O\left(\frac{1}{N^2}\right)\right)^{[Nt]}$、よって $N\to\infty$ の極限をとると、

$P_t(n,n)=\lim_{N\to\infty}\left(1-\frac{n(n-1)}{2N}+O\left(\frac{1}{N^2}\right)\right)^{[Nt]}=\exp\left(-\frac{n(n-1)}{2}t\right)$、$P_t(n,n)$ は平均が $2/n(n-1)$ の指数分布に従う。状態 n への滞在時間は平均 $2/n(n-1)$ の指数分布に従い、その後状態 $n-1$ へ推移する。



サンプル数 4 の場合の図
状態 4,3,2 の滞在時間を
τ_4,τ_3,τ_2 とする。
$T_4=0,\quad T_1=\tau_4+\tau_3+\tau_2$

n 個のサンプル遺伝子は最終的に確率 1 で一つの共通祖先(Most Recent Common Ancestor)に到達する。その待ち時間を T_1 とすると平均 $E[T_1]=\sum_{k=2}^{n}\frac{2}{k(k-1)}=2\left(1-\frac{1}{n}\right)$ となる。この確率過程を合祖過程(Coalescent Process)と呼ぶ。

Kingman(1982)はより一般の可換モデルにおいて同様に合祖過程が導かれることを証明した。N個の半数体生物から成る集団から n 個の個体 $I_1,I_2,...I_n$ をサンプルしたとする。その祖先の共有関係によってサンプル遺伝子の集合に次のような同値関係 $R_r^{(N)}$ を定義する。すなわち、二つの個体 I_k,I_j が r 世代前に祖先を共有するとき $(I_k,I_j)\in R_r^{(N)}$ と書き、この時刻で同じ同値類に属すと言う。$R_r^{(N)}$ は r 世代前に祖先を共有するか否かによって $I_1,I_2,...I_n$ 上に定義された同値類である。初期状態は全てが異なるという状態でこれを $R_0^{(N)}=\Delta$ で表す、また r 世代前にこの n 個の個体が単一の共通祖先を持つならばこれを $R_r^{(N)}=\Theta$ で表すことにする。$R_r^{(N)}$ は $I_1,I_2,...I_n$ 上に定義された同値関係を状態空間とするマルコフ連鎖である。また $R_r^{(N)}$ に含まれるクラス数を $A_r^{(N)}=\left|R_r^{(N)}\right|$ で定義する。例えば n=5 のとき、

$R_0^{(N)}=\{(I_1)(I_2)(I_3)(I_4)(I_5)\}, R_1^{(N)}=\{(I_1I_2)(I_3I_4)(I_5)\},$

$R_2^{(N)}=\{(I_1,I_2)(I_1,I_2,I_5)\}, R_3^{(N)}=\{(I_1,I_2,I_3,I_4,I_5)\}=\Theta$ とすると

$A_0^{(N)}=5,\ A_1^{(N)}=3,\ A_2^{(N)}=2,\ A_3^{(N)}=1$ となる。

集合 $\{I_1,I_2,...I_n\}$ 上に定義される全ての同値関係の集合を E_n とする。$\alpha,\beta\in\mathrm{E}_n$ に対して、α が β の細分であるとき $\alpha\subseteq\beta$ で表す。例えば $\alpha=\{(I_1,I_2)(I_3,I_4)(I_5)\}$, $\beta=\{(I_1,I_2)(I_3,I_4I_5)\}$ のとき $\alpha\subseteq\beta$ である。また $\alpha=\{(I_1,I_2)(I_3,I_4)(I_5)\}$, $\beta=\{(I_1,I_3)(I_2,I_4I_5)\}$ のとき、α は β の細分ではない。α に含まれるクラス数を $|\alpha|$ とすると、$\alpha\subseteq\beta$ のとき $|\alpha|\geq|\beta|$ である。n 個のサンプル遺伝子が r 世代前の祖先によって $\alpha\in\mathrm{E}_n$ のとき、$r+1$ 世代前に状態 $\beta\in\mathrm{E}_n$ である推移確率 $P_{\alpha,\beta}$ を求める。同値関係による $\{I_1,I_2,...I_n\}$ の分割を

β : $\{C_1, C_2, ..., C_b\}$、$|\beta| = b$ とする。また、α は β の細分なので、

α : $\{C_{11}, ..., C_{1a_1}, C_{21}, ..., C_{2a_2}, ..., C_{b1}, .., C_{ba_b}\}$　　ただし $\sum_{i=1}^{b} a_i = |\alpha| = a$、

$\bigcup_{i=1}^{a_i} C_{ji} = C_j$ $(j = 1,2,..,b)$ である。

N個の個体から成る集団からランダムに a 個の異なる個体を取り出したとき、その親が異なる b 個体である確率なので、$(j_1, j_2, \cdots, j_b)$ を $\{1,2,\cdots,N\}$ の N 個の自然数から選んだ異なる b 個の自然数とするとき、各親個体 j_i が生む k_i 個の子供から a_i 個の子供を選べば良いので、可換モデルであることに注意すると次式が成り立つ。

$$P_{\alpha,\beta} = \frac{1}{N_{[a]}} E\left[\sum_{(j_1,j_2,...,j_b)} (\nu_{j_1})_{[a_1]} (\nu_{j_2})_{[a_2]} ..... (\nu_{j_b})_{a_b}\right]$$

$$= \frac{1}{N_{[a]}} \sum_{(j_1,j_2,...,j_b)} \sum_{(k_1,...,k_b)} (k_1)_{[a_1]} (k_2)_{[a_2]} .... (k_b)_{[a_b]} P(\nu_{j_1} = k_1, ..., \nu_{j_b} = k_b)$$

$$= \frac{N_{[b]}}{N_{[a]}} \sum_{(k_1,...,k_b)} (k_1)_{[a_1]} ..... (k_b)_{[a_b]} P(\nu_1 = k_1, ..., \nu_b = k_b) = \frac{N_{[b]}}{N_{[a]}} E[(\nu_1)_{[a_1]} ... (\nu_b)_{[a_b]}] \quad (3.4)$$

ただし　$x_{[k]} = x(x-1)(x-2)...(x-k+1)$

$\{P_{\alpha,\beta}; \alpha, \beta \in \mathrm{E}_n\}$ はマルコフ連鎖 $R_r^{(N)}$ の推移確率である。

定義：$\alpha, \beta \in \mathrm{E}_n$ に対して、$\alpha \subseteq \beta$, $|\alpha| = |\beta| + 1$ のとき、$\alpha \prec \beta$ で表す。すなわち

β は α の中の二つのクラスが合体(合祖)したものである。このとき次の定理を得る。

<u>定理３．２ (Kingman(1982c))</u>

極限値 $\lim_{N\to\infty} Var(\nu_1) = \sigma^2 > 0$ が存在し、すべての $p \geq 1$ に対して $\sup_N E[\nu_1^{\,p}] < \infty$ のとき

$$P_{\alpha,\beta} = \begin{cases} \dfrac{\sigma^2}{N} + o(\dfrac{1}{N}) & if \quad \alpha \prec \beta \\ 1 - \dfrac{a(a-1)}{2} \times \dfrac{\sigma^2}{N} + o(\dfrac{1}{N}) & if \quad \alpha = \beta \\ o(\dfrac{1}{N}) & otherwise \end{cases} \qquad \text{ただし} a = |\alpha|$$

(証明)

(ｉ) $\alpha \prec \beta$ のとき、$|\alpha| = |\beta| + 1$ より $a_1 = 2, a_2 = 1, ...a_b = 1$ の場合を示せば十分である。

$$P_{\alpha,\beta}=\frac{1}{N_{[a]}}E[\sum_{(j_1,\ldots,j_b)}\nu_{j_1}(\nu_{j_1}-1)\nu_{j_2}\ldots\nu_{j_b}]=\frac{1}{N_{[a]}}E\left[\sum_{j=1}^{N}\{\nu_j(\nu_j-1)(\sum_{\substack{(j_2,\ldots,j_b)\\ j_i\neq j(2\le i\le b)}}\nu_{j_2}\ldots\nu_{j_b})\}\right]$$

$$\le\frac{1}{N_{[a]}}E\left[\sum_{j=1}^{N}\nu_j(\nu_j-1)\{\nu_1+\nu_2+..+\nu_N\}^{b-1}\right]$$

$\sum_{i=1}^{N}\nu_i=N, b=a-1$ より

$$P_{\alpha,\beta}\le\frac{N^{a-2}}{N_{[a]}}E[\sum_{j=1}^{N}\nu_j(\nu_j-1)]=\frac{N^{a-1}}{N_{[a]}}E[\nu_1(\nu_1-1)]=\frac{\sigma^2}{N}+o(\frac{1}{N}) \qquad (3.5)$$

$S=\sum_{j=1}^{N}\left\{\nu_j(\nu_j-1)\left(\sum_{\substack{(j_2,\ldots,j_b)\\ j_i\neq j(2\le i\le b)}}\nu_{j_2}\ldots\nu_{j_b}\right)\right\}$ とすると、$\sum_{\substack{(j_2,\ldots,j_b)\\ j_i\neq j}}\nu_{j_2}\nu_{j_3}\ldots\nu_{j_b}$ は$(\nu_1+..+\nu_N-\nu_j)^{b-1}$

の展開式から、いずれかのiについてν_iの２次以上を含む項を全て引いたものに等しい

$\sum_{\substack{(j_2,\ldots,j_b)\\ j_i\neq j}}\nu_{j_2}\nu_{j_3}\ldots\nu_{j_b}\ge(N-\nu_j)^{b-1}-\sum_{l\neq j}\{(\nu_i+\sum_{k\neq l,j}\nu_k)^{b-1}$の中で$\nu_i$の２次以上のを含む項の和$\}$

{　} 内は$b-1=a-2$個の$\left(\nu_i+\sum_{k\neq l,j}\nu_j\right)$の積なので、その中から２個は$\nu_i$、他は$\nu_j$を除

く任意の変数を選んで掛けた項全体の和より小さく、$\sum_{k\neq j}\nu_k\le N$なので結局

$$\sum_{\substack{(j_2,\ldots,j_b)\\ j_i\neq j}}\nu_{j_2}\nu_{j_3}\ldots\nu_{j_b}\ge(N-\nu_j)^{b-1}-\binom{a-2}{2}\sum_{l\neq j}\nu_l^{\,2}N^{a-4}$$

$p\ge1, 0\le x\le1$のとき、$(1-x)^p\ge1-px$ より

$$(N-\nu_j)^{b-1}=(N-\nu_j)^{a-2}=N^{a-2}\left(1-\frac{\nu_j}{N}\right)^{a-2}\ge N^{a-2}\left\{1-\frac{(a-2)\nu_j}{N}\right\}$$　これより

$$S=\sum_{j=1}^{N}\left\{\nu_j(\nu_j-1)\left(\sum_{\substack{(j_2,\ldots,j_b)\\ j_i\neq j(2\le i\le b)}}\nu_{j_2}\ldots\nu_{j_b}\right)\right\}$$

$$\ge\sum_{j=1}^{N}\nu_j(\nu_j-1)[N^{a-2}\{1-\frac{(a-2)\nu_j}{N}\}-\binom{a-2}{2}\sum_{l\neq j}\nu_l^{\,2}N^{a-4}]$$

$$= N^{a-2}\sum_{j=1}^{N}\nu_j(\nu_j-1)-(a-2)N^{a-3}\sum_{j=1}^{N}\nu_j{}^2(\nu_j-1)-\binom{a-2}{2}N^{a-4}\sum_j\sum_{i\neq j}\nu_j(\nu_j-1)\nu_i{}^2$$

$$P_{\alpha,\beta}=\frac{E[S]}{N_{[a]}} \geq \frac{N^{a-1}}{N_{[a]}}\left\{E[\nu_1(\nu_1-1)]-\frac{(a-2)}{N}E[\nu_1{}^3]-\frac{\binom{a-2}{2}}{N}E[\nu_1{}^2\nu_2{}^2]\right\}$$

$$=\frac{\sigma^2}{N}+o(\frac{1}{N}) \quad \cdots\cdots \qquad (3.6)$$

(3.5)(3.6)より $P_{\alpha,\beta}=\frac{\sigma^2}{N}+o(\frac{1}{N})$。

（ii） $\alpha\subseteq\beta,\ |\alpha|\geq|\beta|+2$ すなわち $a\geq b+2$ のとき

$$P_{\alpha,\beta}=\frac{N_{[b]}}{N_{[a]}}E[(\nu_1)_{[a_1]}(\nu_2)_{[a_2]}\cdots(\nu_b)_{[a_b]}]\leq\frac{N_{[b]}}{N_{[a]}}E[\nu_1{}^{a_1}\nu_2{}^{a_2}\cdots\nu_b{}^{a_b}]$$

$$\leq\frac{N_{[b]}}{N_{[a]}}(E[\nu_1{}^a])^{\frac{a_1}{a}}(E[\nu_2{}^a])^{\frac{a_2}{a}}\cdots(E[\nu_b{}^a])^{\frac{a_b}{a}} \qquad \text{(ヘルダーの不等式(注))}$$

$$=\frac{N_{[b]}}{N_{[a]}}E[\nu_1{}^a]=o(\frac{1}{N})$$

（注） Appendix(E) ヘルダーの不等式で $f_i=\nu_i^{a_i}$, $\frac{1}{p_i}=\frac{a_i}{a}$ $(i=1,2,\ldots,b)$ と置く。

（iii） $\alpha=\beta$ のとき

$\alpha\not\subset\beta$ のとき、明らかに $P_{\alpha,\beta}=0$。$|\alpha|=a$ のとき $a=b+1$ となる組み合わせは $\binom{a}{2}$ 通り

よって（i）（ii）より $P_{\alpha,\alpha}=1-\frac{a(a-1)}{2}\times\frac{\sigma^2}{N}+o(\frac{1}{N})$ を得る。

（証明終わり）

$$Q=(Q_{\alpha,\beta}),\ \text{ただし}\ Q_{\alpha,\beta}=\begin{cases}1 & if\ \alpha\prec\beta\\ -\frac{a(a-1)}{2} & if\ \alpha=\beta,\ |\alpha|=a\\ 0 & \text{その他}\end{cases} \qquad (3.7)$$

とすると推移確率は $P=(P_{\alpha,\beta})=I+c_NQ+o(c_N),\ c_N=\frac{\sigma^2}{N}$ と表される。 (3.8)

集団から二つの遺伝子をランダムに取り出したとき、同じ親を持つ確率をc_Nとすると

$$c_N = \frac{1}{N(N-1)}\sum_{i=1}^{N} E[\nu_i(\nu_i - 1)] = \frac{E[\nu_1(\nu_1 - 1)]}{N-1} = \frac{Var(\nu_1)}{N-1}$$

（1） Wright-Fisher model

ν_1は二項分布$B(N,\frac{1}{N})$に従うので、ν_1の母関数は$g(x) = E[e^{x\nu_1}] = (\frac{e^x}{N} + 1 - \frac{1}{N})^N$

$$Var(\nu_1) = N \times \frac{1}{N} \times (1 - \frac{1}{N}) = 1 - \frac{1}{N} \quad 、\ \lim_{N\to\infty} Var(\nu_1) = 1 \quad c_N = \frac{Var(\nu_1)}{N-1} = \frac{1}{N}$$

$$\sup_N E[\nu_1{}^p] = \sup_N \lim_{x\to 0} \frac{\partial^p}{\partial x^p} g(x) < \infty$$

（2） Moran model

$$P(\nu_1 = 0) = \frac{N-1}{N^2},\ P(\nu_1 = 1) = \frac{1}{N^2} + (\frac{N-1}{N})^2 = \frac{N^2 - 2N + 2}{N^2},\ P(\nu_1 = 2) = \frac{N-1}{N^2}$$

$$E[\nu_1] = 1,\ Var(\nu_1) = \frac{2(N-1)}{N^2}, \quad c_N = \frac{2}{N^2}$$

定理３．２の結果よりＮ世代を単位時間とする時間のスケーリングを行い、Ｎを無限に大きくすると遺伝子の系図を表す合祖過程(Coalescent Process)と呼ばれる連続時間マルコフ連鎖が得られる。

定理３．３

一般に行列$A = (a_{\xi\eta})$は$a_{\xi\eta} \geq 0, \sum_{\eta} a_{\xi\eta} = 1$を満たすとき確率行列と呼ばれる。このとき

ノルム$\|A\| = \max_{\xi} \sum_{\eta} |a_{\xi\eta}|$に関して縮小作用素（$\|A\| \leq 1$）であるので行列$A_i, B_i (i = 1,2,..,r)$が

縮小作用素のとき、ｒに関する数学的帰納法により$\|A_1 A_2 ... A_r - B_1 B_2 ... B_r\| \leq \sum_{i=1}^{r} \|A_i - B_i\|$

が成り立つ。これより

$$\|P_N^{[t/c_N]} - (I + c_N Q)^{[t/c_N]}\| \leq \frac{t}{c_N} \|P_N - (I + c_N Q)\| = t \left\| \frac{P_N - I}{c_N} - Q \right\| \ 。$$

故に $\lim_{N\to\infty} P_N^{[t/c_N]} = \lim_{N\to\infty} (I + c_N Q)^{[t/c_N]} = e^{tQ}$。これより時間スケールされたプロセス

$(R^{(N)}_{[t/c_N]})_{t\geq 0}$の有限次元分布収束が導かれる。

サンプルαの定理３．３で得られた連続時間マルコフ連鎖を$\alpha_t = \lim_{N\to\infty} R^{(N)}_{[t/c_N]}$とし$n-$合祖過程と呼ぶ。$|\alpha_t|$は時刻$t$での祖先遺伝子数を表し、初期条件は$|\alpha_0| = n$で$|\alpha_t|$は時間と共に減少する死滅過程である。連続時間マルコフ連鎖α_tが状態$|\alpha_t| = k$に滞在する時間をτ_kとすると、

$$P(\tau_k > t) = \exp(-\frac{k(k-1)}{2}t), \qquad k = n, n-1, \ldots, 2 \tag{3.9}$$

分布密度　$\frac{d}{dt}P(\tau_k \le t) = \frac{k(k-1)}{2}\exp(-\frac{k(k-1)}{2}t)$,　　τ_n, $\tau_{n-1}, \ldots\ldots, \tau_2$は独立。

祖先遺伝子数$|\alpha_t|$の推移確率を$P_t(k,j) = P(|\alpha_t| = j \| \alpha_0| = k)$とすると次の方程式を満たす。

$$\frac{d}{dt}P_t(k,j) = \frac{k(k-1)}{2}\{P_t(k-1,j) - P_t(k,j)\} \qquad P_0(k,j) = \delta_{k,j} \tag{3.10}$$

この解については次の節で突然変異を含むより一般的な形で求める。

$n-$合祖過程α_tのジャンプ過程を$\{\mathfrak{R}_k ; k = n, n-1, \ldots, 1\}$とすると、$\beta \in \mathrm{E}_n$, $|\beta| = k$のとき

$P(\alpha_t = \beta \| \alpha_0| = n) = P(|\alpha_t| = k \| \alpha_0| = n)P(\mathfrak{R}_k = \beta | \mathfrak{R}_n = \Delta)$、すなわち二つの過程$|\alpha_t|$と$\mathfrak{R}_k$は独立である。

３．２　遺伝的多様性と遺伝子系図

生物集団内の遺伝的多様性はDNA塩基配列の多様性であり、それはDNAに生じる突然変異に起因する。この節ではまず突然変異を含む遺伝子系図について考えよう。
突然変異はそれまでに存在しない常に新しいタイプであるという無限対立遺伝子モデルを仮定する。このとき、系図上に現れる突然変異は常に新しいアレルタイプの起源と見なされる。サンプル遺伝子の系図を過去に遡って行くと、そのアレルタイプの起源となる祖先遺伝子に到達する。同一の突然変異を起源にもつ遺伝子を同祖的(identical by descent)と呼ぶことにすると、サンプル遺伝子の祖先遺伝子に含まれる家系数(Lines of Descent)に関するマルコフ過程を得る。突然変異は毎世代$\frac{\vartheta}{2N}$の率で起こり、生じた突然変異は常に新しいタイプとする(infinite-allele model)。k個のサンプル遺伝子について、時間をtだけ遡った集団において突然変異を経ることなしにj個の祖先に由来する確率を$P_t(k,j)$とすると、

１世代当たりの変化は１世代を$\Delta t = \frac{1}{N}$の時間スケールに取ると

$$P_{t+\Delta t}(k,k)=\left(1-\frac{\vartheta}{2N}\right)^{k}\left(1-\frac{k(k-1)}{2N}\right)P_t(k,k)+O\left(\frac{1}{N^2}\right)$$

$$P_{t+\Delta t}(k,j)=\left(1-\frac{\vartheta}{2N}\right)^{k}\left(1-\frac{k(k-1)}{2N}\right)P_t(k,j)+\left(\frac{k\vartheta}{2N}+\frac{k(k-1)}{2N}\right)P_t(k-1,j)+O\left(\frac{1}{N^2}\right)$$

これより次の微分方程式を得る。

定理３．４

$$\frac{d}{dt}P_t(k,j)=\frac{k(k+\vartheta-1)}{2}\{P_t(k-1,j)-P_t(k,j)\}\qquad(k>j)$$

$$\frac{d}{dt}P_t(j,j)=-\frac{j(j+\vartheta-1)}{2}P_t(j,j),\qquad P_0(k,j)=\delta_{k,j}$$

$\vartheta=0$ とすると定理３．３の合祖過程が得られる。

この方程式の解は Tavare(1984,p129)で解説されているが、ここではラプラス変換を用いた方法を紹介しよう。

$Q_\lambda(k,j)=\int_0^\infty e^{-\lambda t}P_t(k,j)dt$ としよう。上式の両辺をラプラス変換すると

$\lambda Q_\lambda(k,j)-\delta_{k,j}=\alpha_i\{Q_\lambda(k-1,j)-Q_\lambda(k,j)\}$　ただし $\alpha_i=\frac{i(i+\vartheta-1)}{2}$。

これより $k>j$ のとき　$Q_\lambda(k,j)=\frac{\alpha_k}{\lambda+\alpha_k}Q_\lambda(k-1.j)+\frac{\delta_{k,j}}{\lambda+\alpha_k}=\frac{\alpha_k}{\lambda+\alpha_k}Q_\lambda(k-1,j)$

これを繰り返すと $Q_\lambda(k,j)=\{\prod_{i=j+1}^{k}\frac{\alpha_i}{\lambda+\alpha_i}\}Q_\lambda(j,j)=\frac{1}{\alpha_j}\prod_{i=j}^{k}\frac{\alpha_i}{\lambda+\alpha_i}$、

ただし $P_t(j,j)=\exp\{-\frac{j(j+\vartheta-1)}{2}t\}$ より $Q_\lambda(j,j)=\frac{1}{\lambda+\alpha_j}$ である。

さらにこの式を部分分数に分解すると

$$Q_\lambda(k,j)=\frac{1}{\alpha_j}\sum_{i=j}^{k}\frac{A_i}{\lambda+\alpha_i}\ ,\quad \text{ただし}\ A_i=\left(\prod_{r=j}^{k}\alpha_r\right)\Big/\prod_{\substack{r=j\\ r\neq i}}^{k}(\alpha_r-\alpha_i),\ i\geq j$$

ラプラス逆変換により

$$P_t(k,j)=\sum_{i=j}^{k}\frac{A_i}{\alpha_j}\exp\left[-\frac{i(i+\vartheta-1)}{2}t\right]\qquad(3.11)$$

と表現される。ただし

$$\frac{A_i}{\alpha_j}=\left(\prod_{r=j+1}^{k}\alpha_r\right)\Bigg/\prod_{\substack{r=j\\ r\neq i}}^{k}(\alpha_r-\alpha_i)=\prod_{r=j+1}^{k}\{r(r+\vartheta-1)\}\Bigg/\prod_{\substack{r=j\\ r\neq i}}^{k}\{(r-i)(r+i+\vartheta-1)\}$$

$$=\frac{k!(2i+\vartheta-1)\prod_{r=j+1}^{k}(r+\vartheta-1)}{j!(-1)^{i-j}(k-i)!\prod_{r=j}^{k}(r+i+\vartheta-1)}=\frac{(-1)^{i-j}(2i+\vartheta-1)k_{[i]}(j+\vartheta)_{(i-1)}}{j!(i-j)!(k+\vartheta)_{(i)}}$$

ここで$x_{(j)}=x(x+1).....(x+j-1),\quad x_{[j]}=x(x-1)....(x-j+1)$

さらにサンプル数kを無限に大きくすると、

$$\lim_{k\to\infty}\frac{A_i}{\alpha_j}=\lim_{k\to\infty}\frac{(-1)^{i-j}(2i+\vartheta-1)k_{[i]}(j+\vartheta)_{(i-1)}}{j!(i-j)!(k+\vartheta)_{(i)}}=\frac{(-1)^{i-j}(2i+\vartheta-1)(j+\vartheta)_{(i-1)}}{j!(i-j)!}$$

$k\to\infty$とすると、$P_t(j)=\lim_{k\to\infty}P_t(k,j)$は全集団の遺伝子が$t$遡った祖先集団で突然変異を経ることなく$j$個の遺伝子と同祖的(lines of descent)につながっている確率となる。

$$P_t(j)=\lim_{k\to\infty}P_t(k,j)=\sum_{i=j}^{\infty}\frac{(-1)^{i-j}(2i+\vartheta-1)(j+\vartheta)_{(i-1)}}{j!(i-j)!}\exp\left[-\frac{i(i+\vartheta-1)}{2}t\right]\qquad(3.12)$$

特に$\vartheta=0$とすると、全集団の遺伝子がt遡ったj個の祖先遺伝子の子孫である確率は

$$P_t(j)=\sum_{i=j}^{\infty}\frac{(-1)^{i-j}(2i-1)(j)_{(i-1)}}{j!(i-j)!}\exp\left[-\frac{i(i-1)}{2}t\right]\qquad(3.13)$$

$j=1$とすると、t遡った祖先集団で一つの共通祖先に由来する確率が

$$P_t(1)=\sum_{i=1}^{\infty}\frac{(-1)^{i-1}(2i-1)(1)_{(i-1)}}{(i-1)!}\exp\left[-\frac{i(i-1)}{2}t\right]=1-\sum_{i=2}^{\infty}(-1)^{i}(2i-1)\exp\left[-\frac{i(i-1)}{2}t\right](3.14)$$

遺伝子はアデニン(A)、チミン(T)、シトシン(C)、グアニン(G)の４種の塩基の一次配列として表現される。集団から３本の遺伝子をサンプルし、次のようなＤＮＡ塩基配列が得られたと仮定しよう。

ATTGCCTAGGTCAACTGGACCTGA　（サンプル１）
ATTGTCTAGCTCAACTGGACCTGA　（サンプル２）
ATTGCCTAGCTCAACTGGACCTGA　（サンプル３）

全部で２４個の塩基から成るDNA領域で、ほとんど同じ配列であるが、例えば左から５番目の塩基サイトはサンプル１と３はシトシン（C）であるがサンプル２はチミン（T）であり、同様に１０番目もサンプルによって塩基の種類（G，C）が異なる。この様に、サ

ンプルによって異なる塩基を持つようなサイトを分離サイト(segregating site)という。一般に二つのサンプル遺伝子について、その共通祖先までの合祖時間が長いと、それだけ多くの突然変異を蓄積し分離サイトの数も増加する。従って、逆に分離サイトの数が多いと、共通祖先までの合祖時間も長いと予想される。そこで、サンプル中の分離サイトの情報から共通祖先までの総合祖時間を推定する問題を考えてみよう。サンプル遺伝子の系図過程で j 個の祖先遺伝子である期間の長さをτ_j（$2 \le j \le n$）,とする。$T = \sum_{j=2}^{n} \tau_j$ は合祖時間の合計、$L = \sum_{j=2}^{n} j\tau_j$ は系図の全ての枝(branch)の長さの合計である。

サンプル遺伝子数が２個のとき、田島(Tajima(1983))は分離サイトの数が$S = k$という条件の下で、同一祖先までの合祖時間の分布が次のガンマ分布で与えられることを示した。合祖時間の分布は指数分布$P(T = t) = \exp(-t)$に従い、$T = t$の条件下で分離サイトの数は平均$\frac{\vartheta}{2} \times 2t = \vartheta t$のポアソン分布$P(S = k|T = t) = \frac{(\vartheta t)^k}{k!}\exp(-\vartheta t)$に従うので

$$P_2(T = t|S = k) = \frac{P_2(T = t, S = k)}{P_2(S = k)} = \frac{P_2(T = t)P_2(S = k|T = t)}{P_2(S = k)}$$

$$= \frac{\exp(-t) \times \frac{(\vartheta t)^k}{k!}\exp(-\vartheta t)}{\frac{1}{(1+\vartheta)}(\frac{\vartheta}{1+\vartheta})^k} = \frac{(1+\vartheta)^{k+1}}{k!}t^k \exp[-(1+\vartheta)t] = Gamma(k+1, \frac{1}{1+\vartheta}) \quad (3.15)$$

ここで$\vartheta = 2Nu$, u は１世代１遺伝子当たりの突然変異率である。

ここでは一般に n 個の遺伝子をサンプルしたとき、その分離サイトから共通な祖先までの総合祖時間を推定する問題を考える。そのために、(T, S)の同時母関数を定義する。

$G(\lambda, z) = E[\exp(-\lambda T) \times z^S] = \sum_{k=0}^{\infty} z^k \{\int_0^{\infty} P(T = t, S = k)e^{-\lambda t}dt\}$ ただし $\lambda \ge 0, 0 \le z \le 1$。

このとき、次のようにして容易に母関数を求めることができる(Tavare et al(1997))。

補助定理３．５

$$G(\lambda, z) = \prod_{j=2}^{n} \left\{ \frac{j(j-1)}{j(j-1) + \vartheta j(1-z) + 2\lambda} \right\}$$

（証明）突然変異は系図の全長Ｌのとき、平均$\frac{\vartheta L}{2}$のポアソン分布で発生し、滞在時間τ_jは分布密度が指数分布$\frac{d}{dt}P(\tau_j \le t) = \frac{j(j-1)}{2}\exp\left(-\frac{j(j-1)}{2}t\right)$に従うことより、

$$G(\lambda,z)=E[E[\exp(-\lambda T)z^S|\tau_2,..,\tau_n]]=E[\exp(-\lambda T)E[z^S|\tau_2,..,\tau_n]]$$

$$=E[\exp(-\lambda T)\{\sum_{k=0}^{\infty}z^kP(S=k|L)\}]=E\left[\exp(-\lambda T)\left\{\sum_{k=0}^{\infty}z^k\times\frac{(\frac{\vartheta L}{2})^k}{k!}\exp(-\frac{\vartheta L}{2})\right\}\right]$$

$$=E[\exp(-\lambda T)\exp(\frac{\vartheta Lz}{2})\exp(-\frac{\vartheta L}{2})]=E[\exp(-\lambda T-\frac{\vartheta}{2}L(1-z))]$$

$$=E[\exp\{-\lambda\sum_{j=2}^{n}\tau_j-\frac{\vartheta}{2}(1-z)\sum_{j=2}^{n}j\tau_j\}]=\prod_{j=2}^{n}E[\exp\{-\lambda\tau_j-\frac{\vartheta}{2}(1-z)j\tau_j\}]$$

$$=\prod_{j=2}^{n}[\int_0^{\infty}\exp\{-\lambda t-\frac{\vartheta}{2}(1-z)jt\}\times\frac{j(j-1)}{2}\exp(-\frac{j(j-1)}{2})dt]$$

$$=\prod_{j=2}^{n}\left(\frac{j(j-1)}{j(j-1)+\vartheta j(1-z)+2\lambda}\right)$$

ここで$E[\bullet|\tau_2,\cdots,\tau_n]$は$\tau_2,\cdots\tau_n$を与えた下での条件付き期待値を表す（付録(H)参照)。

母関数$G(\lambda,z)$をzのベキ級数に展開すると

$$G(\lambda,z)=\sum_{k=0}^{\infty}z^k\left\{\sum_{|r|=k}\{\prod_{j=2}^{n}(\frac{\vartheta}{\vartheta+j-1})^{r_j}(\frac{j-1}{\vartheta+j-1})\}\{\prod_{j=2}^{n}(\frac{j(\vartheta+j-1)}{2\lambda+j(\vartheta+j-1)})^{r_j+1}\}\right\}$$

ここで$\sum_{|r|=k}$は$|r|=\sum_{i=2}^{n}r_i=k$を満たす全ての $n-1$次元ベクトル $r=(r_2,...,r_n)$についての和を表す。これより

$$\int_0^{\infty}P(T=t,S=k)e^{-\lambda t}dt=\sum_{|r|=k}\{\prod_{j=2}^{n}(\frac{\vartheta}{\vartheta+j-1})^{r_j}(\frac{j-1}{\vartheta+j-1})\}\{\prod_{j=2}^{n}(\frac{j(\vartheta+j-1)}{2\lambda+j(\vartheta+j-1)})^{r_j+1}\}.$$

最後にラプラス逆変換により次の定理を得る。

定理３．６

$$P(T=t,S=k)=\sum_{|r|=k}\left\{\prod_{j=2}^{n}\left(\frac{\vartheta}{\vartheta+j-1}\right)^{r_j}\left(\frac{j-1}{\vartheta+j-1}\right)\right\}\left\{\underset{j=2}{\overset{n}{*}}Gamma\left(r_j+1,\frac{2}{j(\vartheta+j-1)}\right)\right\}$$

$$\text{ここで}Gamma\left(r_j+1,\frac{2}{j(\vartheta+j-1)}\right)=\frac{\left(\frac{j(\vartheta+j-1)}{2}\right)^{r_j+1}}{r_j!}t^{r_j}\exp\left[-\frac{j(\vartheta+j-1)}{2}t\right]$$

また $\overset{n}{\underset{j=2}{*}} Gamma(,)$ はガンマ分布の畳み込みを意味する。さらに

$P(S=k)=\sum_{|r|=k}\prod_{j=2}^{n}\left\{\left(\frac{\vartheta}{\vartheta+j-1}\right)^{r_j}\left(\frac{j-1}{\vartheta+j-1}\right)\right\}$ より $S=k$ の条件下での T の条件付分布は

$P_n(T=t|S=k)=\frac{P(T=t,S=k)}{P(S=k)}$ で与えられる。

定理３．６の内容は第１章、補助定理１．１７に基づき次のように説明できる。

一つの遺伝子の系図上で突然変異率は $\frac{\vartheta}{2}$ である。故に k 個遺伝子の系図のいずれかで mutation が起こるまでの待ち時間を σ_k とすると指数分布

$P(\sigma_k>t)=\left\{\exp\left[-\frac{\vartheta}{2}t\right]\right\}^k=\exp\left[-\frac{\vartheta k}{2}t\right]$ で与えられる。また状態 k での滞在時間 τ_k は

$P(\tau_k>t)=\exp\left[-\frac{k(k-1)}{2}t\right]$。k 個のサンプル遺伝子について最初に突然変異または合祖が起こるまでの待ち時間 $Min(\sigma_k,\tau_k)$ の分布は

$$P(Min(\sigma_k,\tau_k)>t)=\exp\left[-\frac{k\vartheta}{2}t\right]\exp\left[-\frac{k(k-1)}{2}t\right]=\exp\left[-\left(\frac{k\vartheta+k(k-1)}{2}\right)t\right]$$

変化が起きたとき、それが突然変異、合祖である確率はそれぞれ

$$P(\text{突然変異})=P(Min(\sigma_k,\tau_k)=\sigma_k)=\frac{k\vartheta/2}{\{k\vartheta+k(k-1)\}/2}=\frac{\vartheta}{\vartheta+k-1}$$

$$P(\text{合祖})=P(Min(\sigma_k,\tau_k)=\tau_k)=\frac{k(k-1)/2}{\{k\vartheta+k(k-1)\}/2}=\frac{k-1}{\vartheta+k-1}$$

これより k 個のサンプル遺伝子の系図について r 回突然変異が生じ最後に coalesce が起こる確率は $\left(\frac{\vartheta}{\vartheta+k-1}\right)^r\left(\frac{k-1}{\vartheta+k-1}\right)$ 、r 回の突然変異と１回の合祖が起こるまでの待ち時間は指数分布 $Exp\left(\frac{2}{k(\vartheta+k-1)}\right)$ の $r+1$ 重畳み込みで与えられるので、ガンマ分布 $Gamma\left(r+1,\frac{2}{k(\vartheta+k-1)}\right)$ となる（ガンマ分布については付録 A 参照）。

$|r|=\sum_{j=2}^{n}r_j=k,$ を満たすベクトル $r=(r_2,\ldots,r_n)$ について $\left(\frac{\vartheta}{\vartheta+j-1}\right)^{r_j}\left(\frac{j-1}{\vartheta+j-1}\right)$ は j 個

の系統が $j-1$ 個の系統に合祖する前に r_j 回突然変異が起こる確率であり、

$\prod_{j=2}^{n}\left(\frac{\vartheta}{\vartheta+j-1}\right)^{r_j}\left(\frac{j-1}{\vartheta+j-1}\right)$ は共通な一つの祖先に到達するまでに、$r=(r_2,...,r_n)$ に対応した突然変異と合祖が起こる確率である。状態 j から $j-1$ へ合祖するまでに r_j 回の突然変異が起こるとき状態 j の滞在時間はガンマ分布 $\mathrm{Gamma}(r_j+1,\frac{2}{j(\vartheta+j-1)})$ に従う。$\eta^r{}_j(j=2,..,n)$ をガンマ分布 $\mathrm{Gamma}\left(r_j+1,\frac{2}{j(\vartheta+j-1)}\right)$ に従う独立な確率変数とすると、総計 $\eta^r=\sum_{j=2}^{n}\eta_j$ の分布はこれらのガンマ分布の畳み込みで表される。

$r=(r_2,...,r_n)$ に対して、$A(r)=\prod_{j=2}^{n}\left\{\left(\frac{\vartheta}{\vartheta+j-1}\right)^{r_j}\left(\frac{j-1}{\vartheta+j-1}\right)\right\}\Big/P(S=k)$ とすると、明らかに $\sum_{|r|=k}A(r)=1$。T の条件付分布は $P_n(T=t|S=k)=\sum_{|r|=k}^{r}A(r)P(\eta^r=t)$ となる。

ガンマ分布 $Gamma(\alpha,\beta)$ の平均、分散はそれぞれ $\alpha\beta$ および $\alpha\beta^2$ である。よって

$$E[\eta^r]=\sum_{j=2}^{n}\frac{2(r_j+1)}{j(\vartheta+j-1)}\ \text{ and }\ Var[\eta^r]=\sum_{j=2}^{n}\frac{4(r_j+1)}{j^2(\vartheta+j-1)^2}.$$

条件 $S=k$ の下での合祖時間 T の条件付平均および分散は次式のようになる。

$$\begin{cases}E\{T|S=k]=\sum_{|r|=k}A(r)E[\eta^r]\\ Var[T|S=k]=\sum_{|r|=k}A(r)Var[\eta^r]+\sum_{|r|=k}A(r)\{E[\eta^r]-E\}^2\ \ where\ \ E=E[T|S=k]\end{cases}\tag{3.16}$$

分離したサイト数 S について別の表現式を紹介しよう。系図の全枝の長さ $L=\sum_{j=2}^{n}j\tau_j$ が与えられたとき、突然変異の数はパラメーター $\frac{\vartheta}{2}L$ のポアソン分布に従うので

$P(S=k|L)=\frac{(\vartheta L/2)^k}{k!}\exp\left(-\frac{\vartheta L}{2}\right)$。$L$ の分布密度を $f_L(t)$ とすると

$P(S=k)=\int_0^\infty P(S=k|t)f_L(t)dt$ で与えられる。$L_j=j\tau_j\ (j=2,3,\cdots,n)$ の分布は

$P(L_j = j\tau_j > t) = P\left(\tau_j > \frac{t}{j}\right) = \exp\left(-\frac{j(j-1)}{2} \times \frac{t}{j}\right) = \exp\left(-\frac{(j-1)t}{2}\right)$ より

分布密度は $f_{L_j}(t) = \frac{j-1}{2}\exp\left(-\frac{(j-1)t}{2}\right)$ となる。$L_2, L_3, \cdots, L_n$ は互いに独立で

$L = \sum_{j=2}^{n} L_j$ なので、L の分布密度は $\{f_{L_j}(t); j = 2,3,\cdots,n\}$ の畳み込みで得られるので

最終的に、$f_L(t) = \sum_{j=2}^{n}(-1)^j \binom{n-1}{j-1} \frac{j-1}{2}\exp\left(-\frac{(j-1)}{2}t\right)$ となる。　　(3.17)

また、これは $f_L(t) = \frac{n-1}{2}e^{-t/2}\left(1-e^{-t/2}\right)^{n-2}$ という表示も可能である。

これより、分離サイト数 S の確率分布は

$$P(S=k) = \int_0^\infty \frac{(\vartheta t/2)^k}{k!}\exp\left(-\frac{\vartheta t}{2}\right) \times \left(\sum_{j=2}^{n}(-1)^j \binom{n-1}{j-1}\frac{j-1}{2}\exp\left(-\frac{(j-1)}{2}t\right)\right)dt$$

$$= \left(\frac{\vartheta}{2}\right)^k \sum_{j=2}^{n}(-1)^j \binom{n-1}{j-1}\frac{j-1}{2}\int_0^\infty \frac{t^k}{k!}\exp\left(-\frac{(\vartheta+j-1)}{2}t\right)dt$$

$$= \sum_{j=2}^{n}(-1)^j \binom{n-1}{j-1}\left(\frac{j-1}{\vartheta+j-1}\right)\left(\frac{\vartheta}{\vartheta+j-1}\right)^k. \qquad (3.18)$$

3．3　合祖過程の種々の性質

合祖過程を用いて集団内の遺伝子の系図や多様性に関する様々な性質を導くことができる。ここでは Kingman(1982),Tavare(1984)から幾つかの結果を紹介する。

この節では突然変異は仮定しない Kingman の合祖過程から導かれる性質を紹介する。

3．3．1合祖過程のジャンプ過程とファミリーサイズ過程

合祖過程 α_t のジャンプ過程 $\{\Re_k; k = n, n-1, \ldots, 1\}$ の推移確率を求める。$\{\Re_k\}$ の

1回の推移確率は $\xi, \eta \in \mathrm{E}_n$ に対し　$P(\Re_{k-1} = \eta | \Re_k = \xi) = \begin{cases} \frac{2}{k(k-1)} & \xi \succ \eta \text{のとき} \\ 0 & \text{その他} \end{cases}$。

これより次の定理を得る。

定理３．７

$n-$合祖過程 α_t において、祖先サイズ過程 $|\alpha_t|$ とジャンプ過程 $\{\Re_k\}$ は独立であり

$\alpha_t = \Re_{|\alpha_t|}$ と表される。また $\Re_k$ の分布は次式で与えられる。

$$P(\Re_k = \xi | \Re_n = \Delta) = \frac{(n-k)!k!(k-1)!}{n!(n-1)!}\lambda_1!\lambda_2!\ldots.\lambda_k! \qquad (3.19)$$

ここで$\lambda_1,\lambda_2,...,\lambda_k$は$\xi$に含まれる$k$個の同値類の各クラスサイズで、$\lambda_1+...+\lambda_k=n$を満たす。

（証明）$\alpha_t=\Re_{|\alpha_t|}$は３．１節で述べたので、式（3.19）を証明する。nから始め、逆向きの数学的帰納法で示す。

（ｉ）$k=n$のとき、$\lambda_1=...=\lambda_n=1$なので、(3.19) の右辺＝１が成り立つ。

（ii）kのとき成り立つと仮定して、$\xi\in \mathrm{E}_n,|\xi|=k$のとき$P_k(\xi)=P(\Re_k=\xi|\Re_n=\Delta)$

と書くことにする。

（iii）$k-1$のとき、$\eta\in \mathrm{E}_n,|\eta|=k-1$に対して

$$P_{k-1}(\eta)=\sum_{\xi}P(\Re_{k-1}=\eta|\Re_k=\xi)P_k(\xi)=\frac{2}{k(k-1)}\sum_{\xi\succ\eta}P_k(\xi)$$

ηのクラスサイズを$\lambda_1,\lambda_2,...,\lambda_{k-1}$（ただし$\lambda_1+...+\lambda_{k-1}=n$）とする。$\xi$に含まれる二つのクラスが１クラスに合併して$\eta$の$k-1$個のクラスが出来るので$\xi$のクラスサイズは$1\le i\le k-1,1\le \nu\le \lambda_i-1$に対して、$\lambda_1,....,\lambda_{i-1},\nu,\lambda_i-\nu,\lambda_{i+1},.....,\lambda_{k-1}$と$\eta$の$i$番目のクラスを二つに分割する形で与えられる。故に全ての可能なi,νについて加えればよいので、λ_i個の遺伝子をνと$\lambda_i-\nu$の二つのクラスに分割する方法は$\frac{1}{2}\binom{\lambda_i}{\nu}$通りに注意して

$$\begin{aligned}P_{k-1}(\eta)&=\sum_{i=1}^{k-1}\sum_{\nu=1}^{\lambda_i-1}\frac{2}{k(k-1)}\times\frac{(n-k)!k!(k-1)!}{n!(n-1)!}\lambda_1!....\lambda_{i-1}!\nu!(\lambda_i-\nu)!....\lambda_{k-1}!\times\frac{1}{2}\binom{\lambda_i}{\nu}\\&=\frac{(n-k)!(k-1)!(k-2)!}{n!(n-1)!}\sum_{i=1}^{k-1}\lambda_1!...\lambda_{i-1}!\lambda_{i+1}!....\lambda_{k-1}!\{\sum_{\nu=1}^{\lambda_i-1}\nu!(\lambda_i-\nu)!\binom{\lambda_i}{\nu}\}\\&=\frac{(n-k+1)!(k-1)!(k-2)!}{n!(n-1)!}\lambda_1!\lambda_2!....\lambda_{k-1}!\end{aligned}$$

よって$k-1$のときも成り立ち証明された。

$n-$合祖過程から派生するマルコフ過程としてファミリーサイズ過程$\{F_t;t\ge 0\}$がある。$\alpha\in \mathrm{E}_n$がサイズiの同値クラスをm_i個（$i=1,2,...,n$）含むとき$f(\alpha)=(m_1,m_2,..,m_n)$と書くことにする。$\sum_{i=1}^{n}m_i$は$\alpha$に含まれる同値クラスの数であり、$\sum_{i=1}^{n}im_i=n$である。このとき、$n-$合祖過程$\alpha_t$に対し、$F_t=f(\alpha_t)$によってファミリーサイズ過程$F_t$を定義する。$\alpha_t$はｎ個の異なるクラスに始まり、最終的に１つのクラスに合祖するのでファミリーサイズ過程は$F_0=(n,0,...,0)$に始まり$F_\infty=(0,..,0,1)$に終わる。ファミリーサイ

ズ過程の推移確率$P(F_t = \mathrm{M}|F_0)$（ただし$\mathrm{M} = (m_1, m_2, ..., m_n)$、$F_0 = (n,0,...,0)$）はクラス数を$\sum_{i=1}^{n} m_i = k$とすると

$$P(F_t = \mathrm{M}|F_0) = P(|\alpha_t| = k \| \alpha_0| = n)\left\{ \sum_{\xi, f(\xi)=M} P(\Re_k = \xi) \right\}$$

$$= P_t(n,k) \frac{(n-k)!k!(k-1)!}{n!(n-1)!} \lambda_1! \lambda_2! ... \lambda_k! \left\{ \sum_{\xi, f(\xi)=M} 1 \right\}$$

$f(\xi) = M$を満たすξの数は$\dfrac{n!}{(m_1!....m_n!)(\lambda_1!...\lambda_k!)}$なので

$$P(F_t = \mathrm{M}|F_0) = P_t(n,k) \times \frac{(n-k)!(k-1)!}{(n-1)!} \times \frac{k!}{m_1!...m_n!} = P_t(n,k) \times \left\{ \binom{k}{M} \Big/ \binom{n-1}{k-1} \right\} \quad (3.20)$$

ここで$\binom{k}{M} = \dfrac{k!}{m_1!....m_n!}$である。 (Tavare(1984)

3．3．2　集団からアレルが消失する速度

突然変異が存在しないとき、集団内の遺伝的多様性は時間とともに消失してゆく。この消失の速度を Felsenstein(1971)に従い遺伝子系図を利用して導いた Tavare(1984)の結果を紹介する。

時刻 t にサンプルした n 個の遺伝子が j 種のアレルタイプを含む確率を$Q_t(n, j)$とする。このサンプル遺伝子の系図を遡ると

$$Q_t(n, j) = \sum_{k=j}^{n} P_t(n,k) Q_0(k, j) \quad (3.21)$$

ここで$P_t(n,k)$は n の遺伝子がk個の祖先に由来する$n-$合祖過程の推移確率である。

$n \to \infty$として全集団がj種のアレルタイプを含む確率を$Q_t(j) = \lim_{n\to\infty} Q_t(n, j)$とすると

(3.10)を使って

$$Q_t(j) = \lim_{n\to\infty} Q_t(n, j) = \sum_{k=j}^{\infty} P_t(k) Q_0(k, j) \quad (3.22)$$

ここで$Q_0(k, j)$は時刻$t = 0$に取り出したk個のサンプルがちょうどj種のアレルタイプを持つ確率である。このとき次の補助定理が成り立つ。

補助定理３．８　(Tavare(1984),(7.4)式)

$t = 0$の初期集団に K 種のアレルタイプが存在したと仮定し、頻度を$p_1, p_2, ..., p_K$　（ただ

し$\sum_{i=1}^{K} p_i = 1$）とする。このとき

$$Q_0(k,j) = \sum_{s=1}^{j} (-1)^{j-s} \binom{K-s}{j-s} \sum_{S} (p_{i_1} + \ldots + p_{i_s})^k \quad , j = 1,2,\ldots, \min(k,K) \qquad (3.23)$$

ここで$\sum_{S}$ は与えられた各 s に対して$1 \le i_1 < i_2 < \ldots < i_s \le K$を満たすすべての$(i_1,..,i_s)$

についての和を表す。

（証明）アレルタイプを$\{1,2,\ldots,K\}$とし、この中からj種のアレル$1 \le i_1 < i_2 < \ldots < i_j \le K$

を固定して考える。このj種の中からr種の部分集合を取り出し$I = \{i_1, i_2,..,i_r\}, r \le j$とする。集団からランダムに$k$個の遺伝子をサンプルするとき、$A(I) = A(i_1,\ldots i_r)$をこの$k$個のサンプルに含まれるアレルタイプは$i_1,\ldots,i_r$の部分集合であるという事象、また
$B(I) = B(i_1,..,i_r)$を含まれるアレルタイプがちょうど$i_1,\ldots,i_r$であるという事象とする。
二つのアレルタイプの部分集合$J_1 = (k_1,..,k_s), J_2 = (j_1,..,j_u)$について$J_1 \ne J_2$, $J_1 \subseteq I$
$J_2 \subseteq I$ならば、$B(J_1) \bigcap B(J_2) = \phi$、すなわち$B(J_1)$と$B(J_2)$は排反事象である。よって
$A(I) = \sum_{J \subseteq I} B(J)$なので　$P(A(I)) = \sum_{J \subseteq I} P(B(J))$。このとき包含と排除の原理（付録 D 参照）により$P(B(I)) = \sum_{J \subseteq I} (-1)^{i-j} P(A(J))$と表せる。ここで$i = |I|$, $j = |J|$（各集合の要素数）。

これより$Q_0(k,j) = \sum_{J} P(B(i_1,i_2,..,i_j))$、ここで$\sum_{J}$ はK種の中の全ての j 種のアレルタイプ$1 \le i_1 < i_2 < \ldots < i_j \le K$についての和を表す。包含と排除の原理より

$$Q_0(k,j) = \sum_{J} P(B(i_1,i_2,..,i_j)) = \sum_{J} \sum_{L \subseteq J} (-1)^{j-s} P(A(L)) \qquad （ただし$s = |L| \le |J| = j$）$$

$$= \sum_{s=1}^{j} \sum_{L:|L|=s} P(A(L)) (-1)^{j-s} \left(\sum_{J(\supseteq L)} 1 \right)$$

ここで$\sum_{J(\supseteq L)} 1$は集合$L = (k_1,..,k_s)$を部分集合として含む集合 J（ただし$|J| = j$）の数である。
これは K 種のアレルタイプからＬに含まれる s 種のアレルを除いた$K - s$種のアレルから$j - s$種のアレルを選ぶ場合の数に等しいので$\sum_{J(\supseteq L)} 1 = \binom{K-s}{j-s}$。また$L = (k_1,..,k_s)$のとき

$P(A(L)) = (p_{k_1} + \ldots + p_{k_s})^k$なので(3.23)を得る。

$k < j$ のとき明らかに $Q_0(k,j) = 0$ であるが、このとき(3.23)の右辺＝0も成り立つ。
これに注意し、(3.23)を((3.22)に代入すると

$$Q_t(j) = \sum_{k=1}^{\infty} P_t(k) Q_0(k,j) = \sum_{k=1}^{\infty} P_t(k) \left\{ \sum_{s=1}^{j} (-1)^{j-s} \binom{K-s}{j-s} \sum_{S} (p_{i_1} + \ldots + p_{i_s})^k \right\}$$

$$= \sum_{s=1}^{j} (-1)^{j-s} \binom{K-s}{j-s} \left\{ \sum_{S} \sum_{k=1}^{\infty} P_t(k) (p_{i_1} + \ldots + p_{i_s})^k \right\} \qquad (3.24)$$

二つのアレル A,a を仮定しアレル A の初期頻度が p のとき、$f(p;t) = \sum_{k=1}^{\infty} P_t(k) p^k$ とすると $f(p;t)$ は時刻 t における集団がアレルAに固定している確率を表す。以上をまとめると

定理３．９ (Tavare(1984), (7.6)式)
時刻 t に集団に存在するアレルの種類数がちょうど j である確率は次式で与えられる。

$$Q_t(j) = \sum_{s=1}^{j} (-1)^{j-s} \binom{K-s}{j-s} \sum_{S} f(p_{i_1} + \ldots + p_{i_s}; t) \qquad (3.25)$$

これらの結果は Littler(1975), Griffiths(1979)でも得られている。
さらに集団内の平均のアレル数を求めると

$$\sum_{j=1}^{\infty} j Q_t(j) = \sum_{j=1}^{\infty} j \left\{ \sum_{k=1}^{\infty} P_t(k) Q_0(k,j) \right\} = \sum_{k=1}^{\infty} \left\{ \sum_{j=1}^{\infty} j Q_0(k,j) \right\} P_t(k)$$

アレルタイプを $A_1, A_2, \ldots, A_K$、初期頻度を $p_1, p_2, \ldots, p_K$ とすると
$\sum_{j=1}^{\infty} j Q_0(k,j) = \sum_{j=1}^{K} \{1 - (1-p_j)^k\} = K - \sum_{j=1}^{K} (1-p_j)^k$、よって平均アレル数は

$$\sum_{j=1}^{\infty} j Q_t(j) = \sum_{k=1}^{\infty} \left\{ K - \sum_{j=1}^{K} (1-p_j)^k \right\} P_t(k) = K - \sum_{j=1}^{K} f(1-p_j; t) \qquad (3.26)$$

と表される。

３．４　モランモデルとサブサンプルの遺伝子系図

Kingman の合祖過程は定理３．２、定理３．３で示した様に Cannings の可換モデルからある条件の下で導かれるが、この範疇に入らないモデルとしてモランモデルがある。まず Watterson(1984)によるモランモデルの遺伝子系図について紹介し、さらにサンプル遺伝子の中からさらに遺伝子をサンプルしたサブサンプルを考えたとき、サンプルとそのサブサンプルの入れ子構造の遺伝子系図を考察した Saunder et al.(1984)の結果を紹介する。

３．４．１　モランモデルの遺伝子系図

第２．１節で紹介したN個の半数体生物集団のモランモデルを仮定する。すなわち毎世代１個の親が選ばれ$1-u$の確率で親と同じタイプの子供を、確率uで親と異なる全く新しいタイプの子供を１個生む。突然変異は無限対立遺伝子モデルを仮定する。その後、N個の親の中からランダムに選ばれた１個体が集団から除かれ元の集団サイズに戻る。以上を１世代とするモランモデルで、時刻$t=0$に集団からランダムにｎ個の遺伝子をサンプルし、この遺伝子サンプルを$\{I_1,I_2,.....,I_n\}$とする。これらの遺伝子の中に次のような二種類の同値関係$\{D_t,F_t;t=1,2,3,...\}$を定義する。二つの遺伝子I_k,I_jが突然変異を経由する事無くｔ世代前に共通な祖先を持つとき、$(I_k,I_j)\in D_t$と書き、二つの遺伝子はD_t同値という。D_tは突然変異を経ずにｔ世代前に同じ祖先を共有するもので分類されたクラスの集合である。D_tに含まれるクラス数を$|D_t|$で、D_tに含まれるクラスξ_i $(i=1,2,..,|D_t|)$のサイズをλ_iとする。二つの遺伝子I_k,I_jが１世代前からｔ世代前までの何れかの時点で、ある突然変異遺伝子を共通祖先として持つとき$(I_k,I_j)\in F_t$と書き、二つの遺伝子はF_t同値という。同様にクラス数を$|F_t|$で、F_tに含まれるクラスη_j $(j=1,2,..,|F_t|)$のサイズをμ_jとする。ξ_iはｎ個のサンプル遺伝子のｔ世代前まで突然変異を経由しない１つの祖先を共有するクラスを表わし、η_jは突然変異を起こした１つの祖先を共有するクラスを表している。

二つの同値類によるｎ個の遺伝子$\{I_1,I_2,.....,I_n\}$の分割を

$R_t=\{\xi_i,i=1,2,..,|D_t|;\eta_j,j=1,2,..,|F_t|\}=\{\xi;\eta\}$とする。$\sum_{i=1}^{p}\lambda_i+\sum_{j=1}^{q}\mu_j=n$（ただし$p=|D_t|,\ q=|F_t|$）である。モランモデルの遺伝子系図過程$R_t$の推移確率を考えよう。

初期状態は$|D_0|=n,|F_0|=0$、$R_0=\{\xi_i=(I_i),i=1,2,..,n;\phi\}$である。

（ｉ）第t世代の状態を$R_t=\{\xi;\eta\}$、$R_{t+1}=\{\xi^*;\eta\}$とする。Kingman の合祖過程と同様にξ^*がξの中の二つのクラスを一つのクラスに融合することで得られるとき$\xi\prec\xi^*$と書くことにする。ξの中のクラスξ_iとξ_jが１世代前に共通な親を持ちξ^*でひとつのクラス$\xi_i\cup\xi_j$となったと仮定すると、これは親$\xi_i\cup\xi_j$が突然変異を起こさずに子供ξ_iを残した

場合と子供ξ_jを残した場合の二通りあるので、その確率は$\xi \prec \xi^*$のときに限り

$P(R_{t+1}=\{\xi^*;\eta\}|R_t=\{\xi;\eta\})=\frac{2(1-u)}{N^2}$となる。　（２．２節　例２参照）(3.27)

（ii）t世代前の祖先ξ_iが、その１世代前のある親個体の突然変異を起こした子供であるとき、第t+1世代ではξ_iはF同値類に属すクラスになる。これを$R_t=\{\xi;\eta\}$および$R_{t+1}=\{\xi-\xi_i;\eta+\xi_i\}$と表わすことにするとその推移確率は

$$P(R_{t+1}=\{\xi-\xi_i;\eta+\xi_i\}|R_t=\{\xi;\eta\})=\frac{u}{N}\ 。 \tag{3.28}$$

（iii）変化を起こさない確率は

$$P(R_{t+1}=R_t|R_t=\{\xi;\eta\})=1-\frac{p(p-1)(1-u)}{N^2}-\frac{pu}{N},\ \text{ただし}\ p=|D_t| \tag{3.29}$$

これは（i）の場合の数が$\frac{p(p-1)}{2}$通り、（ii）がp通りあることによる。

１世代当たりの推移確率が(3.27)(3.28)(3.29)で与えられる遺伝子系図のマルコフ連鎖$\{R_t;t=0,1,2,...\}$の推移はD同値類の中の二つのクラスの合祖かまたは突然変異による遷移なので、１世代の$\{R_t\}$の推移確率はD同値類の数$|D_t|$にしか依存しない。**Kingman**の合祖過程と同様に$\{R_t\}$のジャンプ過程を$\{\Re_k;k=n,n-1,...,0\}$とすると$R_t=\Re_{|D_t|}$が成り立つ。$R_t=\{\xi;\eta\}$、$p=|D_t|, q=|F_t|$とすると、

$$P(R_t=\{\xi;\eta\})=P(|D_t|=p)P(R_t=\{\xi;\eta\},|F_t|=q\ \big||D_t|=p)=P(|D_t|=p)P(\Re_p=\{\xi;\eta\})$$

が成り立つ。このとき次の定理が成り立つ。

定理３．１０

初期条件$|D_0|=n,|F_0|=0$、$R_0=\{\xi_i=(I_i),i=1,2,..,n;\phi\}$のとき

$$P(|F_t|=q,\ R_t=\{\xi;\eta\}\ \big||D_t|=k)=\frac{(n-k)!k!\vartheta^q}{n!(k+\vartheta)_{(n-k)}}\prod_{i=1}^{k}\lambda_i!\prod_{j=1}^{q}(\mu_j-1)! \tag{3.30}$$

ただし$k=0,1,..,n;\ q=0,1,..,n-k;\ \vartheta=\frac{Nu}{1-u},\ \sum_{i=1}^{p}\lambda_i+\sum_{j=1}^{q}\mu_j=n$。

（証明）

$P(\mathfrak{R}_p=\{\xi;\eta\})=P(|F_t|=q,R_t=\{\xi;\eta\}\,\big|\,|D_t|=p)$なので、ジャンプ過程の確率分布

$P(\mathfrak{R}_p=\{\xi;\eta\})$が(3.30)の右辺で与えられることを$p$に関する逆向きの帰納法で示す。

（ｉ） $p=n$のとき、$\mathfrak{R}_n=\{\xi_1,....,\xi_n;\phi\}$,$|\xi_i|=\lambda_i=1\,(1\le i\le n),q=0$なので(3.30)の右辺

$$=\frac{0!n!\vartheta^0}{n!(n+\vartheta)_{(0)}}=1$$が成り立つ。

（ⅱ） $p=k+1$のとき成り立つと仮定して、ジャンプ過程の状態をＲ，Ｗで表わすと

$$P(\mathfrak{R}_k=R)=\sum_W P(\mathfrak{R}_{k+1}=W)P(\mathfrak{R}_k=R|\mathfrak{R}_{k+1}=W) \tag{3.31}$$

ジャンプ過程$\{\mathfrak{R}_k\}$のＷからＲへの状態変化は合祖又は突然変異によるので、場合分けをして考える。

(a) 合祖による遷移： WのＤ同値類に属す二つのクラスξ_{i_1}とξ_{i_2}が合祖してＲの一つのクラス$\xi_l=\xi_{i_1}\cup\xi_{i_2}$になったと仮定しよう。$|\xi_l|=\lambda_l$,$|\xi_{i_1}|=\lambda_{i_1}$とすると$|\xi_{i_2}|=\lambda_l-\lambda_{i_1}$。

ξ_lを二つのクラスξ_{i_1} ,ξ_{i_2}に分割する方法は$\frac{1}{2}\binom{\lambda_l}{\lambda_{i_1}}$通り。$1\le\lambda_{i_1}\le\lambda_l-1$なので$W\to R$が合祖による遷移である確率は

$$P(\mathfrak{R}_k=R|\mathfrak{R}_{k+1}=W)=\frac{2(1-u)/N^2}{((k+1)k(1-u)/N^2)+(k+1)u/N}=\frac{2}{(k+1)(k+\vartheta)}$$

(3.26)で合祖による遷移の部分だけの和は帰納法の仮定より

$$\sum_W P(\mathfrak{R}_{k+1}=W)\frac{2}{(k+1)(k+\vartheta)}=\frac{1}{(k+1)(k+\vartheta)}\sum_{l=1}^{k}\left[\sum_{s=1}^{\lambda_l-1}\binom{\lambda_l}{s}\left\{\frac{(n-k-1)!(k+1)!\vartheta^q}{n!(k+1+\vartheta)_{(n-k-1)}}\right.\right.$$

$$\left.\left.\times\lambda_1!...s!(\lambda_l-s)!....\lambda_k!\left(\prod_{j=1}^{q}(\mu_j-1)!\right)\right\}\right]$$

$$=\frac{(n-k-1)!\vartheta^q k!}{n!(k+1+\vartheta)_{(n-k-1)}(k+\vartheta)}\left(\prod_{j=1}^{q}(\mu_j-1)!\right)\left\{\sum_{l=1}^{k}\sum_{s=1}^{\lambda_l-1}\binom{\lambda_l}{s}\lambda_1!....s!(\lambda_l-s)!...\lambda_k!\right\}$$

$$=\frac{(n-k-1)!\vartheta^q k!}{n!(k+\vartheta)_{(n-k)}}\left(\prod_{j=1}^{q}(\mu_j-1)!\right)\left(\prod_{l=1}^{k}\lambda_l!\right)\left(\sum_{l=1}^{k}(\lambda_l-1)\right) \tag{3.32}$$

(b) 突然変異による遷移

$W=\{\xi_1,...,\xi_k,\eta_j;\ \eta_1,...,\eta_{j-1},\eta_{j+1},..,\eta_q\}$, $R=\{\xi_1,..,\xi_k;\eta_1,..,\eta_q\}$ とする。すなわちRの中のF同値類の中の$\eta_j(1\le j\le q)$がWの中のD同値類に属していた一つのクラスの突然変異によって遷移した確率を考える。(a)と同様にして$P(\mathfrak{R}_k=R|\mathfrak{R}_{k+1}=W)=\dfrac{\vartheta}{(k+1)(k+\vartheta)}$、

突然変異による変化の部分の和を取ると

$$\sum_W P(\mathfrak{R}_{k+1}=W)\frac{\vartheta}{(k+1)(k+\vartheta)}=\frac{\vartheta}{(k+1)(k+\vartheta)}\sum_{j=1}^{q}\frac{(n-k-1)!(k+1)!\vartheta^q}{n!(k+1+\vartheta)!}\left\{\left(\prod_{i=1}^{k}\lambda_i!\right)\mu_j!\right\}$$
$$\times\prod_{\substack{s=1\\ s\neq j}}^{q}(\mu_s-1)!$$
$$=\frac{(n-k-1)!k!\vartheta^q}{n!(k+\vartheta)_{(n-k)}}\left(\prod_{i=1}^{k}\lambda_i!\right)\left(\prod_{j=1}^{q}(\mu_j-1)!\right)\left(\sum_{j=1}^{q}\mu_j\right)\qquad(3.33)$$

(3.27)(3.28)を加え合わせると$\sum_{i=1}^{k}(\lambda_i-1)+\sum_{j=1}^{q}\mu_j=n-k$なので(3.30)を得る。

よって、全ての$k=n,n-1,...0$について成り立つ。

さらに$|D_t|$の分布について次の定理が成り立つ。

定理３．１１

$$P(|D_t|=k)=\sum_{i=k}^{n}\left\{1-\frac{2d_i}{N(N+\vartheta)}\right\}^t(-1)^{i-k}(2i+\vartheta-1)\frac{(k+\vartheta)_{(i-1)}n_{[i]}}{k!(i-k)!(n+\vartheta)_{(i)}}\qquad(3.34)$$

ただし、$k=0,1,2,...,n;\ d_i=\dfrac{i(i+\vartheta-1)}{2}$,$x_{(n)}=x(x+1)...(x+n-1)$,

$x_{[n]}=x(x-1)...(x-n+1)$ 。$k=0$のときは、$i=0$の項は１とする。

（証明）数学的帰納法を用いる。

（ｉ）$k=n$のとき　(3.34)は$P(|D_t|=n)=\left(1-\dfrac{2d_n}{N(N+\vartheta)}\right)^t$、これは第ｔ世代まで全く変化がない確率であり明らかに正しい。

（ii）$t=0$のとき(3.34)は(3.11)の$t=0$の時の右辺と一致するので$P(|D_0|=k)=\delta_{n,k}$。

$(k,t)=(n,0)$ を出発点として k は $n,n-1,n-2,...,0$; t は 0,1,2,.... と二変数の数学的帰納法で証明する。全ての $\{(k,0);k=n,n-1,...,0\}$ および $\{(n,t);t=0,1,2,...\}$ については成り立っているので、$t-1$ のとき、$k=n,n-1,...,0$ について成り立ち、t のとき $k=n,(n-1),..,p$ まで成り立つと仮定する。

（iii）時刻 t および $k=p-1$ のとき

$$P(|D_t|=p-1)=P(|D_{t-1}|=p)P(|D_t|=p-1\,||D_{t-1}|=p) + P(|D_{t-1}|=p-1)P(|D_t|=p-1\,||D_{t-1}|=p-1)$$

$$=P(|D_{t-1}|=p)\frac{2d_p}{N(N+\vartheta)}+P(|D_{t-1}|=p-1)\left\{1-\frac{2d_{p-1}}{N(N+\vartheta)}\right\}$$

$P(|D_{t-1}|=p),P(|D_{t-1}|=p-1)$ は帰納法の仮定より(3.34)で与えられるので代入し整理すると(3.34)を得る。

(3.34)で世代を $[\frac{N^2}{2}t]$ と置き、$N^2/2$ 世代を単位時間とするタイムスケールを取り $N\to\infty$ の極限を取ると $\lim_{N\to\infty}\left\{1-\frac{2d_i}{N(N+\vartheta)}\right\}^{[N^2t/2]}=\exp(-d_it),\ d_i=\frac{i(i+\vartheta-1)}{2}$ なので

$$\lim_{N\to\infty}P(|D_{[N^2t/2]}|=k)=\sum_{i=k}^{n}(-1)^{i-k}\frac{(2i+\vartheta-1)(k+\vartheta)_{(i-1)}n_{[i]}}{k!(i-k)!(n+\vartheta)_{(i)}}\exp[-\frac{i(i+\vartheta-1)}{2}t] \qquad (3.35)$$

となり(3.11)と一致する。また、突然変異率 $u=0$ のとき、遺伝子系図過程 R_t は合祖過程となり、常に $|F_t|=0$。(3.30)において $q=0,\ \vartheta=0$ とすると

$$P(R_t=\{\xi;\phi\}\,||D_t|=k)=\frac{(n-k)!k!}{n!k_{(n-k)}}\lambda_1!...\lambda_k!=\frac{(n-k)!k!(k-1)!}{n!(n-1)!}\lambda_1!....\lambda_k! \qquad (3.36)$$

これは(3.19)と一致する。突然変異率 $u>0$ のとき、$t\to\infty$ において $|D_t|\to 0$ となり、全てのサンプルはある突然変異を祖先とする遺伝子となり遺伝子系図過程 R_t は次の定常分布に収束する。

$$P(|F_\infty|=q,\ R_\infty=\{\phi;\eta\}\,|D_\infty=\phi)=\frac{\vartheta^q}{\vartheta_{(n)}}(\mu_1-1)!(\mu_2-1)!....(\mu_q-1)! \qquad (3.37)$$

これは第4章で論じる Ewens のサンプリング公式と呼ばれる式(4.3)に対応し、定常状態の集団からランダムに取り出した遺伝子が q 種類のアレルタイプを含み、それぞれ μ_i 個

$(i=1,2,..,q)$のサンプルを含んでいる確率を表わす。

3．4．2　サブサンプルの遺伝子系図

N個の半数体生物から成る離散時間モランモデルを考える。集団からn個の遺伝子をサンプルし、さらにこのn個のサンプルからm個$(m \le n)$のサブサンプルを取り出す。このサンプルとサブサンプルの遺伝子系図に関するSaunder, Tavare and Watterson(1984)の結果を紹介する。突然変異はないものとする。$A_1^N(t)$をn個のサンプルのt世代前の祖先遺伝子の数、$A_2^N(t)$をm個のサブサンプルの祖先遺伝子の数とする。前節の記法では$A_1^N(t)=|D_t|$である。$(A_1^N(t),A_2^N(t))$の同時推移確率を

$$P_t((n,m),(k,j))=P(A_1^N(t)=k,A_2^N(t)=j \mid A_1^N(0)=n,\ A_2^N(0)=m) \tag{3.38}$$

$A_1^N(t)$の推移確率を$g_t(n,k)=P(A_1^N(t)=k \mid A_1^N(0)=n)$とすると、定理3．11より

$$g_t(n,k)=\sum_{i=k}^{n}\left\{1-\frac{i(i-1)}{N^2}\right\}^t \frac{(-1)^{i-k}(2i-1)k_{(i-1)}n_{[i]}}{k!(i-k)!n_{(i)}} \tag{3.39}$$

二つの遺伝子系図過程$A_1^N(t),A_2^N(t)$についてそれぞれ祖先の数がr個になる最初の時刻を$T_1(r)=\min\{t;A_1^N(t)=r\}$, $T_2(r)=\min\{t;A_2^N(t)=r\}$とし、$A_2^{N*}(r)=A_2^N(T_1(r))$とする。$A_2^{N*}(r)$はサンプルの系図過程がr個の祖先に成った時点でのサブサンプルの祖先遺伝子の数を表わす。このとき、次の補助定理が成り立つ。

補助定理3．12
$A_1^N(0)=n,A_2^N(0)=m$とする。このとき、$\{A_2^{N*}(n-i);\ i=0,1,2,..,n-1\}$は推移確率が次式で与えられる時間的に非一様なマルコフ連鎖である。

$$P(A_2^{N*}(r-1)=k-1 \mid A_2^{N*}(r)=k)=1-P(A_2^{N*}(r-1)=k \mid A_2^{N*}(r)=k)=\frac{k(k-1)}{r(r-1)} \tag{3.40}$$

（証明）
これは遺伝子系図過程でサブサンプルの祖先の数が減少するときは必ずサンプルの祖先の数が減少するので、(3.40)はサンプルの祖先数がrになった時点でサブサンプルの祖先がkのとき、サンプルの系図過程が合祖を起こし$r-1$になった時点でそれがサブサンプルの祖先の合祖でもある確率なので、明らかに場合の数の比より${}_kC_2/{}_rC_2=\frac{k(k-1)}{r(r-1)}$を得る。また
$A_2^{N*}(r)=k$のとき、明らかに$A_2^{N*}(r-1)=k$または$k-1$なので(3.40)を得る。
この結果はモランモデルに限らず全ての可換モデルで成り立つ。

これより $A_2^N(t)$ の条件付分布が得られる。

定理３．１３

$$P(A_2^N(t)=j \mid A_1^N(t)=k, A_1^N(0)=n,\ A_2^N(0)=m)$$
$$=\frac{(n-m)!(n-k)!m!(m-1)!k!(k-1)!}{(m-j)!(k-j)!n!(n-1)!j!(j-1)!}\times\frac{(n+j-1)!}{(k+m-1)!(n+j-k-m)!} \tag{3.41}$$

（証明）

$A_1^N(t)=k, A_2^N(t)=j, A_1^N(t+1)=k$ ならば $A_2^N(t+1)=j$ なので、左辺は

$P(A_2^{N*}(k)=j \mid A_1^N(0)=n, A_2^N(0)=m)$ に等しい。すなわち、時刻 $T_1(k)$ において(3.41)が成り立つことを示せば十分である。$P(A_2^{N*}(k)=j \mid A_1^N(0)=n, A_2^N(0)=m)=\phi(k,j)$ と書くと

補助定理３．１２より $k\le n-1,\ j\le m-1$ のとき

$$\phi(k,j)=\phi(k+1,j+1)\frac{j(j+1)}{k(k+1)}+\phi(k+1,j)\{1-\frac{j(j-1)}{k(k+1)}\} \tag{3.42}$$

境界条件は $\phi(n,m)=1,\ j>m$ のとき $\phi(n,j)=0$ である。まず $j=m$ のとき(3.41)が成り立つことを証明する。$k+1$ まで成り立つと仮定すると

$$\phi(k,m)=\phi(k+1,m)\{1-\frac{m(m-1)}{k(k+1)}\}$$
$$=\frac{(n-m)!(n-k-1)!m!(m-1)!(k+1)!k!}{0!(k+1-m)!n!(n-1)!m!(m-1)!}\times\frac{(n+m-1)!}{(k+m)!(n-k-1)!}\times\frac{(k+m)(k-m+1)}{k(k+1)}$$
$$=\frac{(n-m)!(n-k-1)!m!(m-1)!k!(k-1)!}{0!(k-m)!n!(n-1)!m!(m-1)!}\times\frac{(n+m-1)!}{(k+m-1)!(n-k-1)!}$$

これより $j=m$ のとき全ての k について(3.41)が成り立つことが示された。最後に (k,j) まで成り立つと仮定して $(k,j-1)$ のとき(3.42)より

$$\phi(k,j-1)=\phi(k+1,j)\frac{j(j-1)}{k(k+1)}+\phi(k+1,j-1)\{1-\frac{(j-1)(j-2)}{k(k+1)}\}$$

数学的帰納法により上式を整理すると証明される。

定理３．１３と(3.39)より $(A_1^N(t),A_2^N(t))$ の結合分布も得られる。

$$P_t((n,m),(k,j))=P(A_1^N(t)=k, A_2^N(t)=j \mid A_1^N(0)=n,\ A_2^N(0)=m)$$
$$=P(A_2^N(t)=j \mid A_1^N(t)=k, A_1^N(0)=n, A_2^N(0)=m)P(A_1^N(t)=k \mid A_1^N(0)=n)$$
$$=g_t(n,k)\phi(k,j) \tag{3.43}$$

$g_1(n,k)=\begin{cases} n(n-1)/N^2 & (k=n-1\text{のとき}) \\ 1-n(n-1)/N^2 & (k=n\text{のとき}) \end{cases}$ より１世代当たりの $(A_1^N(t),A_2^N(t))$ の推

移確率 $P((n,m),(k,j))$ は

$$P((n,m),(k,j))=\begin{cases} \dfrac{n(n-1)}{N^2}\times\dfrac{m(m-1)}{n(n-1)}=\dfrac{m(m-1)}{N^2} & (k=n-1, j=m-1) \\ \dfrac{n(n-1)}{N^2}\times\dfrac{(n-m)(n+m-1)}{n(n-1)}=\dfrac{(n-m)(n+m-1)}{N^2} & (k=n-1,\ j=m) \\ \{1-\dfrac{n(n-1)}{N^2}\}\times 1=1-\dfrac{n(n-1)}{N^2} & (k=n,\ j=m) \end{cases}$$

(3.44)

$A_1^N(T_2(j))$ はサブサンプルの系図過程が j 個の祖先に到達した最初の時刻におけるサンプル遺伝子の祖先の数である。故に $A_1(T_2(j))=\max\{k;A_2^*(k)=j\}$ である。定理３．１３より次の定理を得る。

定理３．１４

$$P(A_1^N(T_2(j)=k\,|\,A_1^N(0)=n,A_2^N(0)=m)=P(A_2^{N*}(k+1)=j+1,\ A_2^{N*}(k)=j)$$
$$=\frac{(n-m)!(n-k-1)!}{(m-j-1)!(k-j)!}\times\frac{m!(m-1)!k!(k-1)!}{n!(n-1)!j!(j-1)!}\times\frac{(n+j)!}{(k+m)!(n+j-k-m)!}$$

（証明）

$A_1^N(T_2(j))=k$ となるのは $A_1^N=k+1$ のとき $A_2^N=j+1$、かつ $A_1^N=k$ のとき $A_2^N=j$、すなわち合祖がサブサンプルの祖先内で起こるということなので補助定理３．１２より

$$P(A_1^N(T_2(j)=k\,|\,A_1^N(0)=n,A_2^N(0)=m)=P(A_1^{N*}(k+1)=j+1,\ A_2^{N*}(k)=j)$$
$$=\phi(k+1,j+1)\times\frac{j(j+1)}{k(k+1)}$$

これを整理して上の結果を得る。

特に $j=1$ とするとサブサンプルの遺伝子系図 $A_2(t)$ が共通な一つの祖先に到達したときのサンプル遺伝子の系図 $A_1(t)$ の条件付分布が得られる。

系３．１５

$$P(A_1^N(T_2(1))=k\,|\,A_1^N(0)=n,A_2^N(0)=m)=(m-1)(n+1)\frac{k!m!(n-k-1)!(n-m)!}{(m+k)!(n-1)!(n-k-m+1)!}$$

特に$k=1$とすると

$P(A_1^N(T_2(1))=1\big|A_1^N(0)=n,A_2^N(0)=m)=P(T_1(1)=T_2(1))=\dfrac{(m-1)(n+1)}{(m+1)(n-1)}$となり非常に高い確率でサンプルとサブサンプルが同時に共通な祖先に到達することが分かる。$n\to\infty$とすると、全集団の遺伝子とm個のサンプルが同時に共通な祖先に到達する確率は

$P(A_1^N(T_2(1))=1\big|A_1^N(0)=\infty,A_2^N(0)=m)=\dfrac{m-1}{m+1}$となる。遺伝子の系図過程は祖先遺伝子の数が1個づつ減少する死滅過程であり、サンプルとサブサンプルの共通祖先の集合は高い確率である世代で一致することを意味する。これをサンプルとサブサンプル遺伝子系図過程のカップリング(coupling)と呼ぶ。ある世代でカップリングが生じるとそれ以後は二つの遺伝子系図は同一の系図過程となる。カップリングが最初に起こった世代におけるサンプルおよびサブサンプルの祖先の数をLとすると、初期条件$A_1(0)=n,A_2(0)=m$のとき、その確率$P(L=j)$は$\phi(j,j)=P(A_2^{N*}(j)=j\big|A_1^N(0)=n,A_2^N(0)=m)$を用いて

$$P(L=j)=\phi(j,j)-\phi(j+1,j+1)=2j(n-m)\frac{m!(m-1)!(n-j-1)!(n+j+1)!}{n!(n-1)!(m+j)!(m-j)!}\qquad(3.45)$$

突然変異を考慮に入れたモランモデルを考える。サンプル、サブサンプルの遺伝子系図で3．4．1節で定義した突然変異を経由しない祖先遺伝子の同値クラス類D_tに着目したプロセスを考えよう。すなわち$A_1^N(t)=|D_t|$であり、$A_1^N(t),A_2^N(t)$はサンプル、サブサンプルの突然変異を経由していない祖先遺伝子の数を表わす。$A_1^N(t)$の1世代あたり推移確率を$h_{n,k}=P(A_1^N(1)=k\big|A_1^N(0)=n)$とすると、(3.27)(3.28)(3.29)より

$$h_{n,k}=\begin{cases}1-\dfrac{n(n+\vartheta-1)}{N(N+\vartheta)} & (k=n)\\ \dfrac{n(n+\vartheta-1)}{N(N+\vartheta)} & (k=n-1)\\ 0 & \text{その他}\end{cases}\qquad\text{ただし}\vartheta=\frac{Nu}{1-u}。\qquad(3.46)$$

$A_1^N(t)$の推移確率を$h_t(n,k)=P(A_1^N(t)=k\big|A_1^N(0)=n)$とすると$h_t(n,k)=P(|D_t|=k)$より(3.34)式で与えられる。$(A_1^N(t),A_2^N(t))$の同時分布を

$P_t((n,m),(k,j))=P(A_1^N(t)=k,A_2^N(t)=j\big|A_1^N(0)=n,A_2^N(t)=m)$とすると、1世代当たりの推移確率は

$$P_1((n,m),(k,j))=\begin{cases}1-\dfrac{n(n+\vartheta-1)}{N(N+\vartheta)} & (k=n, j=m)\\ \dfrac{(n-m)(n+m+\vartheta-1)}{N(N+\vartheta)} & (k=n-1,\ j=m)\\ \dfrac{m(m+\vartheta-1)}{N(N+\vartheta)} & (k=n-1,\ j=m-1)\\ 0 & \text{その他}\end{cases} \tag{3.47}$$

例えば$(k,j)=(n-1,m-1)$となるのはサンプル遺伝子の祖先D_tが合祖または突然変異によってnから$n-1$に減少しそれがサブサンプルの系図に含まれる場合なので、その確率は$\dfrac{n(n+\vartheta-1)}{N(N+\vartheta)}\times\dfrac{m(m+\vartheta-1)}{n(n+\vartheta-1)}=\dfrac{m(m+\vartheta-1)}{N(N+\vartheta)}$となる。

$(A_1^N(t),A_2^N(t))$の状態空間は$\Im=\{(k,j)\,|k=0,1,..,n;\ j=0,1,....,\min(m,k)\}$である。

定理３．１３の拡張として次の定理を得る。

定理３．１６

$$\begin{aligned}&P(A_2{}^N(t)=j\,|A_1{}^N(t)=k,A_1{}^N(0)=n,A_2{}^N(0)=m)\\&=\frac{(n-m)!m!(n-k)!\Gamma(m+\vartheta)\Gamma(n+j+\vartheta)\Gamma(k+\vartheta)}{n!(n-k-m+j)!j!(k-j)!(m-j)!\Gamma(n+\vartheta)\Gamma(j+\vartheta)\Gamma(m+k+\vartheta)}\end{aligned} \tag{3.48}$$

（証明）

補助定理３．１２と同様にして$\{A_2^{N*}(n-i);\ i=0,1,2,..,n\}$は推移確率が次式で与えられる非斉次のマルコフ連鎖である。

$$P(A_2^{N*}(k-1)=j-1\,|A_2^{N*}(k)=j)=1-P(A_2^{N*}(k-1)=j\,|A_2^{N*}(k)=j)=\frac{j(j+\vartheta-1)}{k(k+\vartheta-1)}$$

定理３．１３と同様にして(3.48)の右辺は$\phi(k,j)=P(A_2^{N*}(k)=j\,|A_1(0)=n,A_2(0)=m)$

に等しい。(3.42)と同様に次式を得る。

$$\phi(k,j)=\phi(k+1,j+1)\frac{(j+\vartheta)(j+1)}{(k+\vartheta)(k+1)}+\phi(k+1,j)\{\frac{(k-j+1)(k+j+\vartheta)}{(k+\vartheta)(k+1)}\} \tag{3.49}$$

境界条件$\phi(n,j)=\begin{cases}1 & (j=m)\\ 0 & (j>m)\end{cases}$の下で、これより(3.43)を得る。

$(A_1{}^N(t),A_2{}^N(t))$の同時分布$P_t((n,m),(k,j))$は(3.48)と$A_1(t)$の推移確率$h_t(n,k)$を用いて

$$P_t((n,m),(k,j))=h_t(n,k)\phi(k,j) \tag{3.50}$$

またサブサンプルの系図過程が$A_2{}^N(t)=j$となる最初の時刻でのサンプルの系図過程

$A_1^N(T_2(j))$について、定理３，１４と同様に次の定理が成り立つ

定理３．１７

$$P(A_1^N(T_2(j)=k \mid A_1^N(0)=n, A_2^N(0)=m) = P(A_2^{N*}(k+1)=j+1,\ A_2^{N*}(k)=j)$$
$$= \frac{(n-m)!(n-k-1)!m!k!\Gamma(m+\vartheta)\Gamma(n+j+1+\vartheta)\Gamma(k+\vartheta)}{(m-j-1)!(k-j)!j!(n-k+j-m)!n!\Gamma(n+\vartheta)\Gamma(j+\vartheta)\Gamma(m+k+1+\vartheta)}$$

ただし$0 \le j \le k \le n-m+j$。

３．４．３　サブサンプルの連続時間遺伝子系図過程

離散時間モランモデルを用いて考察してきたが、$N^2/2$世代を単位時間とし集団のサイズＮを無限に大きくすると連続時間の系図過程が得られる。まず突然変異がないとき、(3.34)より

$$g_t(n,k) = \lim_{N\to\infty} P(A_1^N(\frac{N^2 t}{2}) = k \mid A_1^N(0)=n) = \sum_{i=k}^{n} \exp[-\frac{i(i-1)}{2}t] \frac{(-1)^{i-k}(2i-1)k_{(i-1)}n_{[i]}}{k!(i-k)!n_{(i)}}$$

定理３．１３はサンプルの系図過程$A_1(\cdot)$がジャンプした時点でのサブサンプルの過程$A_2(\cdot)$の状態のみに依存しているので、連続時間モデルにおいてもそのジャンプ過程に着目すればそのまま成り立つ。定理３．１３で$n \to \infty$とすると

$$P(A_2(t)=j \mid A_1(t)=k, A_1(0)=\infty,\ A_2(0)=m)$$
$$= \lim_{n\to\infty} \frac{(n-m)!(n-k)!m!(m-1)!k!(k-1)!}{(m-j)!(k-j)!n!(n-1)!j!(j-1)!} \times \frac{(n+j-1)!}{(k+m-1)!(n+j-k-m)!}$$
$$= \left\{ \binom{k}{j}\binom{m-1}{j-1} \middle/ \binom{k+m-1}{k-1} \right. \qquad j=1,2,\cdots,\min(m,k) \qquad (3.51)$$

これは全集団が時間ｔ遡った祖先集団のｋ個の祖先に由来するという条件の下で集団からランダムに取り出されたｍ個の遺伝子が同じ祖先集団のｊ個の祖先に由来する確率を表わしている。(3.46)の分母はｊ個の個体をｋ個の祖先に割り当てる重複組み合わせの数$\binom{k+m-1}{k-1}$に等しく、分子はｋ個の祖先からｊ個の祖先を選ぶ組み合わせ$\binom{k}{j}$とｍ個の個体をｊ個の祖先に空部屋が無いように割り当てる方法$\binom{m-1}{j-1}$の積となっている。すなわちランダムな組み合わせから期待される確率であり、中立遺伝子ということから帰結される当然の結果でもある。

次に突然変異を仮定すると、離散モデルから$N \to \infty$として連続時間近似を行ない同様に

$\lim_{N\to\infty}(A_1^N([\frac{N^2t}{2}]), A_2^N([\frac{N^2t}{2}])) = (A_1(t), A_2(t))$ とすると $(A_1(t), A_2(t))$ は空間 $\mathfrak{I}$ の連続時間マルコフ連鎖で、その生成作用素は次式の様になる。

$$Q((n,m),(k,j)) = \begin{cases} -\dfrac{n(n+\vartheta-1)}{2} & (k=n,\ j=m) \\ \dfrac{(n-m)(n+m+\vartheta-1)}{2} & (k=n-1,\ j=m) \\ \dfrac{m(m+\vartheta-1)}{2} & (k=n-1,\ j=m-1) \end{cases} \tag{3.52}$$

3．5　集団サイズの変動を含む遺伝子系図

これまでの議論は全て集団の大きさが一定という前提で考えてきた。しかし、現実の集団は常に様々な環境変動等の要因による個体数の変動がある。遺伝子系図におけるその影響を考えてみよう。離散世代モデルで集団のサイズが世代 τ の関数として $\{N(\tau);\tau=0,1,2,...\}$、ただし $N(0)=N$ と与えられると仮定する。この集団から任意に二つの遺伝子を取り出したとき、共通な祖先に達するまでの時間（世代数）を $\tau_2(N)$ とすると $[Nt]$ 世代以上合祖が起きない確率は $P(\tau_2(N)>[Nt]) = \prod_{\tau=1}^{[Nt]}\left(1-\frac{1}{N(\tau)}\right)$ となる。両辺の対数を取ると　$\log P(\tau_2(N)>[Nt]) = \sum_{\tau=1}^{[Nt]}\log\left(1-\frac{1}{N(\tau)}\right)$

集団サイズの相対比を表す関数として $f_N(t) = \frac{N([Nt])}{N} = \frac{N(\tau)}{N}$　$(\frac{\tau-1}{N} < t \le \frac{\tau}{N}$ のとき）

とする。$f_N(t)$ は t の階段関数である。さらに $f(t) = \lim_{N\to\infty} f_N(t)$ とする。

$x>1$ のとき不等式 $x \le -\log(1-x) \le \frac{x}{1-x}$ より $x = \frac{1}{N(\tau)} = \frac{1}{Nf_N(\tau)}$ とすると、

$$\frac{1}{N}\sum_{\tau=1}^{[Nt]}\frac{1}{f_N(\tau)} \le -\sum_{\tau=1}^{[Nt]}\log\left(1-\frac{1}{N(\tau)}\right) \le \frac{1}{N}\sum_{\tau=1}^{[Nt]}\frac{1}{f_N(\tau)-(1/N)}$$

従って $N\to\infty$ の極限を取り、$\tau_2 = \lim_{N\to\infty} T_2(N)/N$ とすると、区分求積法により

$$\log P(\tau_2>t) = \lim_{N\to\infty}\log P(\tau_2(N)>[Nt]) = \lim_{N\to\infty}\sum_{\tau=1}^{[Nt]}\log\left(1-\frac{1}{N(\tau)}\right) = -\int_0^t\frac{1}{f(s)}ds$$

すなわち $P(\tau_2>t) = \exp[-\Lambda(t)]$,　ただし $\Lambda(t) = \int_0^t\frac{1}{f(s)}ds$。集団サイズが一定のときは $f(t)=1$ として $P(\tau_2>t) = \exp[-t]$ となる。十分時間を遡ると任意の二つの遺伝子は確率

1で共通祖先を持つので$\lim_{t\to\infty} P(\tau_2 > t) = 0$より$\lim_{t\to\infty} \Lambda(t) = \infty$を仮定する。$\tau_2$の分布密度は

$\frac{d}{dt}P(\tau_2 \le t) = \frac{1}{f(t)}\exp[-\Lambda(t)]$、平均は$E[\tau_2] = \int_0^\infty \exp[-\Lambda(t)]dt$である。

一般に遺伝子の数がn個のとき、最初の合祖までの待ち時間をτ_nとすると、同様の議論により$N \to \infty$の極限を取ると、$P(\tau_n > t) = \exp[-\frac{n(n-1)}{2}\Lambda(t)]$となる。集団サイズが変動する場合も3．1節と同様に$n-$合祖過程α_t^νを導くことができるが、これは時間的に非一様な純粋死滅過程であり、上添え字νは変動サイズモデルを表わす。集団サイズ変動モデルでの合祖過程α_tの祖先の数を$A^\nu(t) = |\alpha_t^\nu|$とすると、微小時間$h$での推移確率は

$$P(A^\nu(t+h) = j \,|\, A^\nu(t) = k) = \begin{cases} \frac{k(k-1)}{2f(t)}h + o(h) & (j = k-1) \\ 1 - \frac{k(k-1)}{2f(t)}h + o(h) & (j = k) \\ o(h) & (\text{その他}) \end{cases} \quad (3.53)$$

各状態$A^\nu(.) = n, n-1, ...,2$での滞在時間を$\tau_n, \tau_{n-1}, ..., \tau_2$とすると、その同時分布密度は

$$P(\tau_n = t_n, \tau_{n-1} = t_{n-1}, ..., \tau_2 = t_2) = \prod_{j=2}^{n}\left\{\frac{j(j-1)}{2f(s_j)}\exp[-\frac{j(j-1)}{2}\{\Lambda(s_j) - \Lambda(s_{j+1})\}]\right\} \quad (3.54)$$

ここで$0 \le t_n, ..., t_2 < \infty$, $s_{n+1} = 0$, $s_n = t_n$, $s_j = t_j + t_{j-1} + ... + t_n$ $(j = 2, ..., n-1)$。

$S_j = \tau_j + \tau_{j+1} + ... + \tau_n$とすると$S_2$は一つの共通祖先までの合祖時間であり、条件

$S_{j+1} = s$の下でτ_jの分布は　$P(\tau_j > t \,|\, S_{j+1} = s) = \exp[-\frac{j(j-1)}{2}\{\Lambda(t+s) - \Lambda(s)\}]$

祖先数の過程$A^\nu(t)$は時間的に非一様なマルコフ連鎖である。集団サイズ一定のKingmanのn-合祖過程の祖先数を$A(t) = |\alpha_t|$とすると、集団サイズ変動における過程$A^\nu(t)$は推移確率$P(A^\nu(t+h) = j \,|\, A^\nu(t) = k)$が(3.53)式で与えられることから

$A^\nu(t) = A(\Lambda(t))$, ($A^\nu(0) = A(0) = n$, $t \ge 0$) と表現される。すなわち、集団サイズ変動モ

デルの合祖過程 $A^{\nu}(t)$ は集団サイズ一定の n-合祖過程の関数 $\Lambda(t)$ による決定論的時間変更によって得られることが分かる。

Kingman の $n-$ 合祖過程の推移確率を $g_{n,k}(t)=P(A(t)=k|A(0)=n)$ とすると

過程 $A^{\nu}(t)$ の推移確率は次式で与えられる。

$$P(A^{\nu}(t)=k|\,A^{\nu}(0)=n)=g_{n,k}(\Lambda(t))$$
$$=\sum_{j=k}^{n}\frac{(2j-1)(-1)^{j-k}k_{(j-1)}n_{[j]}}{k!(j-k)!n_{(j)}}\exp[-\frac{j(j-1)}{2}\Lambda(t)] \qquad (3.55)$$

(Tavare(2004))

この章で紹介したモデルは全て半数体生物、すなわち雌雄の区別がない生物という条件のモデルであるが、ヒトをはじめとして、有性生殖を行う生物についての合祖過程の導出は厳密には証明が必要である。この問題については、Möhle(1998c)等を参照されたい。また、合祖過程が導かれるためには、定理３．２で述べた条件 $\lim_{N\to\infty}Var(\nu_1)=\sigma^2>0$、すべての $p\geq 1$ に対して $\sup_{N}E[\nu_1^p]<\infty$ を見たさなければならない。これは、個々の個体が次世代に残す子供の数の分散が $N\to\infty$ においても有限という条件であるが、例えば１個体の子供が全集団のある一定の割合（例えば１０％など）を占めるような確率が正であるような場合、あるいは親世代の各個体は繁殖能力に違いはないが、ランダムに選ばれた数個体のみが次世代に子供を残せると言う様な場合この条件は満たされない。このようなとき、サンプルした３個体以上の個体が同時に合祖するという事態が生じる。このように多数の個体が同時に合祖を起こすようなプロセスは $\Lambda-$*Coalescent* と呼ばれているが、この話題については、Sagitov(1999), Pitman(1999),Möhle and Sagitov(1999, 2003)などを参照されたい。