

## 2. 生物集団の繁殖モデル

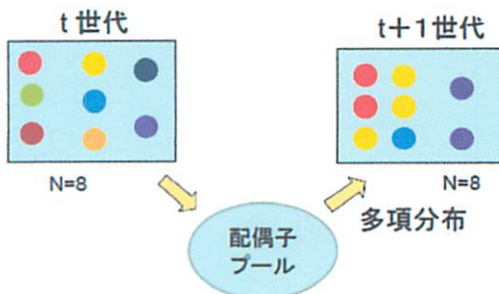
生物集団は有限個の個体より成り、各個体がそれぞれ子供を残すことにより、遺伝的形質を後の世代へと伝えている。その過程で起こる様々な作用によって、その生物集団の遺伝的組成に少しずつ変化が生じ、長い時間の後には進化的現象として現れてくる。このプロセスの基礎となる繁殖のモデルは多くの研究者によって考えられてきた。この章では集団遺伝学で古くから用いられて来た離散時間および連続時間のマルコフ連鎖モデルを紹介する。特に自然選択に対して良くも悪くもない遺伝子は中立な遺伝子と呼ばれ、分子進化において重要な役割を果たすものである。歴史的には、離散時間、離散状態のモデルを連続近似した拡散過程モデルがフィッシャー、ライト、木村、太田、丸山等によって研究され、集団遺伝学の数学的理論として確立された(参照: Crow and Kimura(1970))。さらに、近年はこのマルコフ連鎖から時間を現在から過去へと遡ることにより、遺伝子の系図を辿るプロセス、合祖過程が盛んに研究されている。その基礎となるモデルを紹介する。

### 2. 1 ライト・フィッシャーモデルとモランモデル

この二つのモデルは集団遺伝学で古くから最も良く用いられるモデルである。ライト・フィッシャーモデル(Fisher 1930, Wright 1931)をまず紹介しよう。N個の半数体生物から成る集団とある遺伝子座を考えよう。半数体なので、生物集団を単にN個の遺伝子の集合と見なすこともできる。このN個の個体はその配偶子を作りそれを集めた配偶子のプールを考える。このプールからN個の遺伝子をランダムに取り出し(ランダムサンプリング)、次世代の生物集団が作られる。個体数は常に一定とする。いずれの個体も配偶子を作る能力は等しく、配偶子プールから取り出した1個の配偶子がN個の親の何れのものであるかは等確率とする。このことは、各個体に自然選択に対して有利不利の違いが無く、中立であることを表している。非常に単純化されたモデルであるが、このモデルは有限個体数の生物集団の繁殖モデルとして、その特性を端的に表す優れたモデルとして、最も良く利用される。N個の個体が次世代に子供をそれぞれ $j_1, j_2, \dots, j_N$  (ただし $\sum_{i=1}^N j_i = N$ )となる確率は次の多項分布で与えられる、

$$P(j_1, j_2, \dots, j_N) = \frac{N!}{j_1! j_2! \dots j_N!} \left(\frac{1}{N}\right)^N \quad (2.1)$$

#### ライト・フィッシャー モデル



注目している遺伝子座にいくつかの対立遺伝子（アレル）を考える。例えば血液型の対立遺伝子 A,B,O などは良い例である。r個のアレル  $A_1, A_2, \dots, A_r$  を仮定しよう。ある世代 t にアレル  $A_i$  ( $1 \leq i \leq r$ ) の集団内での数が  $k_i$  であるとする。当然  $\sum_{i=1}^r k_i = N$  である。さらに突然変異の存在を仮定しよう。1世代当たりアレル  $A_i$  が  $A_j$  (ただし  $i \neq j$ ) へ突然変異する確率を  $u_{ij}$  とする。  $u_{ii} = 1 - \sum_{j \neq i} u_{ij}$  は突然変異を起こさない確率である。毎世代突然変異の後、ランダムサンプリングによって次世代の集団が決まると仮定する。次世代におけるアレル  $A_i$  の個数を  $j_i$  ( $1 \leq i \leq r$ ) とすると、  $k = (k_1, \dots, k_r)$  から  $j = (j_1, \dots, j_r)$  への推移確率  $P(j|k)$  は次の多項分布で与えられる。

$$P(j|k) = \frac{N!}{\prod_{i=1}^r j_i!} \prod_{i=1}^r x_i^{j_i} \quad (2.2)$$

ただし  $x_i = \sum_{j=1}^r \frac{k_j}{N} u_{ji}$  は配偶子プールにおける突然変異後のアレル  $A_i$  の頻度である。

$S = \{ k = (k_1, \dots, k_r) | k_i (1 \leq i \leq r) \text{ は } 0 \text{ 以上の整数, } \sum_{i=1}^r k_i = N \}$  とすると、第1章で述べたように、単位時間の推移確率が(2.2)で与えられる S 上の離散時間マルコフ連鎖が決まる。

次にモラン(Moran 1958)によって導入されたモデルを紹介しよう。N個の個体から成る半数体生物集団を仮定する。毎世代ランダムに1個体が死亡し、その個体はN個の個体のいずれかの子供で置き換えられる。どの個体の子供であるかは等しく確率  $\frac{1}{N}$  とする。二つの

アレル A, a を持つ自然選択に対して中立な遺伝子座を考える。第 t 世代において、アレル A の個数を  $X(t)$  とする。各世代集団の中の1個体が死亡しN個の個体のいずれかの子供で置き換わるが、その親がアレルタイプ A のとき確率  $1-u$  で同じアレルタイプの子供、確率  $u$  で a タイプの子供で置き換える。同様に a タイプの親については

$1-v$  と  $v$  で A および a タイプの子供を残す。従って、 $X(t)$  の1世代当たりの推移確率は以下の様になる。

$$\begin{aligned}
P(X(t+1) = j+1 | X(t) = j) &= (1 - \frac{j}{N}) \left\{ \frac{j}{N}(1-u) + (1 - \frac{j}{N})v \right\} = \lambda_j \\
P(X(t+1) = j-1 | X(t) = j) &= \frac{j}{N} \left\{ \frac{j}{N}u + (1 - \frac{j}{N})(1-v) \right\} = \mu_j \\
P(X(t+1) = j | X(t) = j) &= 1 - \lambda_j - \mu_j
\end{aligned} \tag{2.3}$$

毎時刻、各状態  $j$  から確率  $\lambda_j, \mu_j, 1 - \lambda_j - \mu_j$  でそれぞれ  $+1, -1, 0$  だけ移動する離散時間マルコフ連鎖として表現される。

この Moran モデルは連続時間になると、次のように表現される。時刻  $t$  でのアレル A の集団内個数を  $X(t)$  とする。  $X(t) = j$  のとき、微小時間  $\Delta t$  の間に変化が起こる確率は  $(\lambda_j + \mu_j)\Delta t + o(\Delta t)$ 、すなわち変化が起こるまでの待ち時間は指数分布

$\exp[-(\lambda_j + \mu_j)t]$  に従うものとする。そして、確率  $\frac{\lambda_j}{\lambda_j + \mu_j}$  で  $+1$ 、 $\frac{\mu_j}{\lambda_j + \mu_j}$  で

$-1$  だけ移動する出生および死滅が起こる。  $P_i(j, k) = P(X(t) = k | X(0) = j)$  を  $X(t)$  の推移確率とすると、次のコルモゴロフ後退方程式で表現される。

$$\frac{d}{dt} P_i(j, k) = \lambda_j P_i(j+1, k) + \mu_j P_i(j-1, k) - (\lambda_j + \mu_j) P_i(j, k), \quad P_0(j, k) = \delta_{k,j} \tag{2.4}$$

この出生死滅過程の推移確率は具体的に求めることができるが、詳しくは Moran(1958), Karlin & McGregor(1962) を参照されたい。

## 2.2 分枝過程モデルと可換モデル

Karlin & McGregor(1964) はライト・フィッシャーモデルなどを含むより一般的な直積分枝過程モデルを導入した。まず必要な分枝過程に関する基本事項から説明しよう。分枝過程とは 1.3 節で紹介した様に、幾つかの粒子が、それぞれ独立に同じ確率法則にしたがって毎世代子供を生み、その子供達もお互いに独立に子供を生むというプロセスである。1 個体が次の世代に  $k$  個の子を残す確率を  $p_k$  とする。  $\{p_k\}$  の母関数を

$f(s) = \sum_{k=0}^{\infty} p_k s^k$  ( $0 \leq s \leq 1$ ) とする。明らかに  $|f(s)| \leq \sum_{k=0}^{\infty} p_k = 1$  であり、1 個の個体が産

む平均の子供の数は  $\sum_{k=0}^{\infty} k p_k = \lim_{s \rightarrow 1} \frac{d}{ds} f(s)$  で与えられる。

$k$  個の個体があり、次世代に合計  $j$  個の子供となる確率を  $P(k, j)$  とする。各個体は独立に

子供を残すので、 $i$  番目の個体が生む子供の数を  $n_i$  とすると、 $P(k, j) = \sum_{n_1+\dots+n_k=j} p_{n_1} p_{n_2} \dots p_{n_k}$ 、

ここで  $\sum$  は  $n_1 + \dots + n_k = j$  となる全ての  $(n_1, \dots, n_k)$  に関する和を表す。その母関数を

$$P_k(s) = \sum_{j=0}^{\infty} P(k, j) s^j \text{ とすると、 } P_k(s) = \sum_{j=0}^{\infty} \left\{ \sum_{n_1+\dots+n_k=j} p_{n_1} \dots p_{n_k} \right\} s^j = \left\{ \sum_{i=0}^{\infty} p_i s^i \right\}^k = \{f(s)\}^k$$

母関数は 1 個体に関する母関数の積 (累乗) となる。Karlin&McGregor(1964)は次のような分枝過程モデルを導入した。二つのアレル A と a を仮定しよう。A タイプの個体は独立に母関数  $f(z) = \sum_{k=0}^{\infty} p_k z^k$  に従って子供を生み、a タイプは互いに独立に別の母関数

$g(w) = \sum_{k=0}^{\infty} q_k w^k$  に従って子供を生むとする。  $X_t, Y_t$  を第  $t$  世代における、A および a の個

体数とする。各個体は独立に、またそれぞれのタイプに従って子供を生むので、初期時刻  $(X_0, Y_0) = (i_1, i_2)$  が次世代に  $(X_1, Y_1) = (j_1, j_2)$  となる遷移確率は

$$P((X_1, Y_1) = (j_1, j_2) | (X_0, Y_0) = (i_1, i_2)) = P(X_1 = j_1 | X_0 = i_1) P(Y_1 = j_2 | Y_0 = i_2) \\ = \{f(z)\}^{i_1} \{g(w)\}^{i_2} \text{ における } z^{j_1} w^{j_2} \text{ の係数}$$

集団の個体数が一定  $N$  であるという条件を課することにより、ライト・フィッシャーモデルを含むより一般的なモデルを導くことができる。すなわち、 $X_0 = i, Y_0 = N - i$  とし、

$X_1 + Y_1 = N$  という条件付で  $X_1 = j$  となる確率  $P_{i,j}$  を求めると上記の結果より

$$P_{i,j} = P(X_1 = j | X_0 = i, Y_0 = N - i, X_1 + Y_1 = N) \\ = \frac{P(X_1 = j, Y_1 = N - j | X_0 = i, Y_0 = N - i)}{P(X_1 + Y_1 = N | X_0 = i, Y_0 = N - i)} \quad (2.5) \\ = \frac{\{f(z)\}^i \{g(w)\}^{N-i} \text{ における } z^j w^{N-j} \text{ の係数}}{\{f(s)\}^i \{g(s)\}^{N-i} \text{ における } s^N \text{ の係数}}$$

母関数  $f(z), g(w)$  を与えることにより種々のモデルが導かれる。特に、生まれてくる子供の数がポアソン分布  $p_k = \frac{\lambda^k}{k!} \exp(-\lambda)$  および  $q_k = \frac{\mu^k}{k!} \exp(-\mu)$  ( $k=0,1,2,\dots$ ) で与えられるときは、母関数は  $f(z) = \exp[\lambda(z-1)]$ ,  $g(w) = \exp[\mu(w-1)]$  となる。このとき(2.5)は、

$$P_{i,j} = {}_N C_j \left\{ \frac{i\lambda}{i\lambda + (N-i)\mu} \right\}^j \left\{ \frac{(N-i)\mu}{i\lambda + (N-i)\mu} \right\}^{N-j} \quad (2.6)$$

となる。 $i\lambda$  および  $(N-i)\mu$  は  $i$  個の A タイプ個体および  $N-i$  個の a タイプ個体がそれぞ

れ次世代に残す平均の子供の数であり、特に $\lambda = \mu$ のときは二つのアレルに対するライト・フィッシャーモデルとなる。 $\lambda \neq \mu$ のときは、平均の子供の数が異なり、自然選択の効果を表すと考えられる。さらに突然変異を考慮に入れたモデルを考えてみよう。

アレルタイプ A および a の個体は前と同様に。それぞれ確率母関数  $f(z), g(w)$  に従って親と同じタイプの子供を生むとする。次に突然変異によってタイプの転換が起こるが、A から a への突然変異率を  $u$ , a から A への突然変異率を  $v$  とする。そして、

$A(z, w) = (1-u)z + uw$ ,  $B(z, w) = (1-v)w + vz$  とすると、 $A(z, w), B(z, w)$  は A 型及び a 型の子供が突然変異でそれぞれの確率でタイプを転換する確率母関数になる。繁殖および突然変異の 2 つを合成すると A 型の親個体から生まれる子供の数の確率母関数  $F(z, w)$

$$\text{は } F(z, w) = f(A(z, w)) = \sum_{n=0}^{\infty} p_n \{(1-u)z + uw\}^n = \sum_{m,n=0}^{\infty} c(m, n) z^m w^n .$$

ただし  $c(m, n) = p_{m+n} \binom{m+n}{m} (1-u)^m u^n$  は A 型の子供が  $m$  個、a 型の子供が  $n$  個生まれる確率である。同様に a 型の親から生まれる子供の数の母関数は  $G(z, w) = g(B(z, w))$  となる。初期の A 及び a の個体数を  $X_0 = i, Y_0 = j$  ( $i + j = N$ ) とし、集団の大きさを一定  $N$  という条件の下で、この分枝課程かれ誘導される繁殖のマルコフ連鎖モデルの推移確率  $P_{ij}$  は

(2.6)と同様にして、

$$\begin{aligned} P_{ij} &= P(X_1 = j | X_0 = i, Y_0 = N - i, X_1 + Y_1 = N) \\ &= \frac{\{F(z, w)\}^i \{G(z, w)\}^{N-i} \text{における } z^j w^{N-j} \text{ の係数}}{\{F(t, t)\}^i \{G(t, t)\}^{N-i} \text{における } t^N \text{ の係数}} \end{aligned} \quad (2.7)$$

特にポアソン分布に従って子供が生まれ、その後突然変異によってタイプの転換があると  $F(z, w) = \exp[\lambda\{(1-u)z + uw - 1\}]$ ,  $G(z, w) = \exp[\mu\{vz + (1-v)w - 1\}]$  なので、

上式を計算すると  $P_{ij} = {}_N C_j \left(\frac{A}{A+B}\right)^j \left(\frac{B}{A+B}\right)^{N-j}$  が得られる。ただし、A, B は

$$A = \lambda i(1-u) + \mu(N-i)v, \quad B = \lambda i u + \mu(N-i)(1-v) \text{ で与えられる。}$$

さらに多タイプのアレルを含むモデル、上のように直積分枝過程から誘導されるマルコフ連鎖の推移確率行列の固有値、固有ベクトルなど求められているが、興味のある読者は Karlin & McGregor (1964), Karlin (1969) を参照されたい。

次に、Cannings (1974, 1975) による可換モデルを紹介する。ここでは突然変異は無視する。このモデルはライト・フィッシャーモデルを含み、中立な遺伝子の一般的な繁殖モデルである。中立な遺伝子では 1 個の個体から生まれる子供の数は、個体あるいは遺伝子型によらず同じ確率分布に従う。毎世代一定数  $N$  個の半数体生物から成る集団と 2 つの対立遺伝子 A, a をもつ遺伝子座を仮定する。第  $t$  世代におけるアレル A の個数を

$X_i$  ( $X_i = 0, 1, 2, \dots, N$ ) とする。  $(\xi_i, \eta_i)$   $i = 1, 2, \dots, N$  を 2次元の確率変数とする。ここで、 $\xi_i = 0$  または 1 であり、 $i$  番目の個体はその世代で死亡するときは  $\xi_i = 0$ 、次の世代まで生き延びるときは  $\xi_i = 1$  とする。また  $\eta_i$  は  $i$  番目の個体から生まれる子供の数を表す。個体数が一定  $N$  より  $\sum_{i=1}^N \xi_i + \sum_{i=1}^N \eta_i = N$ 、次世代のアレル A の数  $X_{i+1}$  は

$$X_{i+1} = \sum_{i=1}^{X_i} \xi_i + \sum_{i=1}^{X_i} \eta_i = \sum_{i=1}^{X_i} \theta_i \quad \text{ただし } \theta_i = \xi_i + \eta_i \quad \text{となる。個体数一定より } \sum_{i=1}^N \theta_i = N \text{ と}$$

なり、 $\xi_i$  と  $\eta_i$  を分ける必要はない。以上の条件の下で、確率  $P(\theta_{i_1} = j_1, \theta_{i_2} = j_2, \dots, \theta_{i_k} = j_k)$

(ただし、 $i_1, \dots, i_k$  は  $1, 2, \dots, N$  の異なる  $k$  個の数) が任意の  $k$  ( $1 \leq k \leq N$ ) に対して  $(i_1, i_2, \dots, i_k)$  に依存しないとき、可換モデル(Exchangeable model)と呼ばれる。

ライト・フィッシャーモデルは上記の確率が多項分布(2.1)で与えられる可換モデルである。後で見るように、モランモデル、自然選択を含まない直積分枝過程モデルなども全て可換モデルである。 $(\theta_1, \theta_2, \dots, \theta_N)$  の分布の可換性よりマルコフ連鎖  $X_i$  の推移確率は

$$P_{ij} = P(X_{i+1} = j | X_i = i) = P\left(\sum_{k=1}^i \theta_k = j\right) \text{ となる。このとき、次の定理が成り立つ。}$$

定理 2. 1

可換モデルの推移確率行列を  $P = (P_{ij})$  ( $i, j = 0, 1, 2, \dots, N$ ) とする。このとき行列  $P$

の固有値は  $\lambda_0 = 1$ ,  $\lambda_j = E[\theta_1 \theta_2 \dots \theta_j]$  (ただし  $j = 1, 2, \dots, N$ ) で与えられる。

(証明)

行列  $P$  が正則行列  $Z$  と上半三角行列  $A$  によって  $P = ZAZ^{-1}$  すなわち  $PZ = ZA$  と表される

とき、 $P$  の固有値と  $A$  の固有値は一致し、行列  $A$  の対角成分が固有値である。なぜなら、

$$\text{行列 } P \text{ の固有方程式は } |P - \lambda E| = |ZAZ^{-1} - \lambda E| = |Z(A - \lambda E)Z^{-1}| = |Z||A - \lambda E||Z^{-1}| = |A - \lambda E| = 0$$

より、行列  $A$  の固有方程式と一致する。よって、固有値も一致し上半三角行列の固有値は明らかに対角成分である。次のような  $(N+1) \times (N+1)$  の行列  $Z$  を考える。

$$Z = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1^2 & 1^3 & \dots & 1^N \\ 1 & 2 & 2^2 & 2^3 & \dots & 2^N \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & N & N^2 & N^3 & \dots & N^N \end{pmatrix} \quad (2.8)$$

行列  $Z$  の行列式はファン・デルモント行列式であり、0 ではないので、行列  $Z$  は正則行列である。故に  $Z^{-1}PZ = A$  となる行列  $A$  が存在するので、この行列  $A = (a_{kj})$  が上半三角行列、

すなわち  $j < k$  のとき  $a_{kj} = 0$  であることを示せばよい。  $Z^{-1}PZ = A$  は  $PZ = ZA$  と同値なので、  
 $j < i$  のとき行列  $PZ$  の  $(i, j)$  成分は  $\theta_1, \theta_2, \dots$  の互換性により

$$(PZ)_{ij} = \sum_{k=0}^N P_{ik} k^j = E[(X_{i+1})^j | X_i = i] = E\left[\left(\sum_{k=1}^i \theta_k\right)^j\right] = iE[\theta_1^j] + \dots + i_{[j]}E[\theta_1\theta_2 \dots \theta_j]$$

ただし、  $i_{[j]} = i(i-1)(i-2)\dots(i-j+1)$ 。

行列  $A$  が上半三角行列のとき、行列  $ZA$  の  $(i, j)$  成分は  $j < k$  のとき  $a_{kj} = 0$  より

$\sum_{k=0}^i a_{kj} i^k = a_{jj} i_{[j]} + \{i^{j-1}, i^{j-2}, \dots$  の項} となり  $i$  の  $j+1$  乗以上の項が現れない。また逆も成り立つ。行列  $PZ$  の  $(i, j)$  成分はこの形をしており、  $j < k$  のとき  $a_{kj} = 0$ 、すなわち行列  $A$  は上半三角行列である。また固有値は  $a_{jj} = E[\theta_1\theta_2 \dots \theta_j]$  となることが証明された。

この定理を用いて低次の固有値を求めてみよう。明らかに  $\lambda_1 = E[\theta_j] = 1$ 。次に

$\lambda_2 = E[\theta_1\theta_2]$  であるが、  $\sum_{j=1}^N \theta_j = N$  より、  $\theta_j$  の可換性を利用すると

$$N^2 = E\left[\left(\sum_{j=1}^N \theta_j\right)^2\right] = NE[\theta_1^2] + N(N-1)E[\theta_1\theta_2]。$$

これより、  $\lambda_2 = E[\theta_1\theta_2] = 1 - \frac{E[\theta_1^2] - 1}{N-1}$  となる。

また  $\theta_1$  の確率母関数を  $\phi(s) = \sum_{k=0}^N P(\theta_1 = k) s^k$  とすると

$$\lim_{s \rightarrow 1} \frac{d^2}{ds^2} \phi(s) = \phi''(1) = \sum_{k=0}^N k(k-1)P(\theta_1 = k) = E[\theta_1^2] - E[\theta_1] = E[\theta_1^2] - 1。$$

これより、  $\lambda_2 = 1 - \frac{\phi''(1)}{N-1}$  を得る。同様に  $N^3 = E\left[\left(\sum_{j=1}^N \theta_j\right)^3\right]$  と可換性を利用して

$$\lambda_3 = E[\theta_1\theta_2\theta_3] = \frac{N^2 - 3NE[\theta_1^2] + 2E[\theta_1^3]}{(N-1)(N-2)}$$
 となる。より高次の固有値は単一の  $\theta_1$  のみで

表すことができず、より複雑となる。

幾つかの具体的なモデルについて調べてみよう。

例 1. ライト・フィッシャーモデル

$\theta = (\theta_1, \dots, \theta_N)$  は多項分布に従う。  $j = (j_1, j_2, \dots, j_N)$ ,  $\sum_{i=1}^N j_i = N$  とすると、

$P(\theta = \mathbf{j}) = \frac{N!}{j_1! j_2! \dots j_N!} \left(\frac{1}{N}\right)^N$ 。これより固有値は次のように与えられる。

$$\begin{aligned} \lambda_i &= E[\theta_1 \theta_2 \dots \theta_i] = \sum_{j_1, j_2, \dots, j_N} j_1 j_2 \dots j_i \times \frac{N!}{j_1! j_2! \dots j_i! j_{i+1}! \dots j_N!} \left(\frac{1}{N}\right)^N \\ &\quad j_1 - 1 = k_1, \dots, j_i - 1 = k_i, \quad j_{i+1} = k_{i+1}, \dots, j_N = k_N \text{ とおくと} \\ &= \frac{N^{[i]}}{N^i} \sum_{\substack{k_1, k_2, \dots, k_N \\ k_1 + \dots + k_N = N-i}} \frac{(N-i)!}{k_1! \dots k_N!} \left(\frac{1}{N}\right)^{N-i} = \frac{N^{[i]}}{N^i} \end{aligned}$$

### 例 2. 離散時間 Moran モデル

個々の個体が次世代まで生き残る確率あるいは子供を残す確率を考えると、ある個体が死亡しその個体の子供が次世代に残るとき  $\theta = (1, 1, \dots, 1)$  となる。また、死亡する個体が次世代に子供を残せないとき  $\theta = (2, 0, 1, \dots, 1)$  となる。その確率は

$$P(\theta = (1, 1, \dots, 1)) = \left(\frac{1}{N}\right)^2 \times N = \frac{1}{N}$$

$$P(\theta = (2, 0, 1, \dots, 1)) = \left(1 - \frac{1}{N}\right) \frac{1}{N(N-1)} = \frac{1}{N^2}$$

これより固有値は

$$\lambda_j = E\left[\prod_{i=1}^j \theta_i\right] = 1 \times \frac{1}{N} + \frac{1}{N^2} \{(N-j) \times 2j + (N-j)(N-j-1)\} = 1 - \frac{j(j-1)}{N^2}$$

### 例 3. 木村モデル

Kimura ( ) はライト・フィッシャーモデルの無限大の配偶子プールを考える代わりに各個体が毎世代  $r$  個の配偶子を作るモデルを考えた。集団の個体数を  $N$  とすると  $Nr$  個の全配偶子のプールからランダムサンプリングによって  $N$  個の遺伝子を抽出し、次世代の集団が作られる。  $j$  番目の個体がつくる  $r$  個の配偶子の中で、次世代の集団に寄与する配偶子の数を  $\theta_j$  とすると、  $\theta = (\theta_1, \dots, \theta_N)$  の分布は次の一般超幾何分布で与えられる。

$$P(\theta = \mathbf{x}) = \prod_{j=1}^N \binom{r}{x_j} / \binom{Nr}{N}, \quad \mathbf{x} = (x_1, x_2, \dots, x_N) \quad \sum_{j=1}^N x_j = N. \text{ これより、固有値は}$$

$$\lambda_j = E\left[\prod_{i=1}^j \theta_i\right] = \sum_{\mathbf{x} | \sum x_i = N} x_1 x_2 \dots x_j \prod_{k=1}^N \binom{r}{x_k} / \binom{Nr}{N} = \frac{r^j N^{[j]}}{(rN)^{[j]}}.$$

## 2. 3 遺伝子頻度の拡散過程

生物集団の繁殖モデルとしてマルコフ連鎖モデルを紹介してきたが、一般に離散時間



のマルコフ連鎖は状態空間が複雑になるにつれその解析が困難になる。集団遺伝学ではこのマルコフ連鎖からある種の極限操作により導かれる拡散過程モデルが伝統的に多く用いられてきた。拡散方程式による表現によって、種々のモデルの詳しい解析が可能となる。この節では、これまで紹介したマルコフ連鎖モデルと関連する幾つかの拡散過程モデルについて、その結果だけ簡単に紹介する。

2. 2節(2.7)式で紹介した Karlin&MacGregor(1964)の突然変異を含む分枝過程モデルで、ただし二つのアレル A, a が生む子供の数が同じ母関数  $F(z, w) = f((1-u)z + uw)$  に従う中立な遺伝子とする。個体数  $N$  の集団で第  $t$  世代でのアレル A の数を  $X(t, N)$ 、A から a へ

の突然変異率を  $u = \frac{\mathcal{G}_1}{N}$  a から A を  $v = \frac{\mathcal{G}_2}{N}$ 、 $N$  世代を単位時間とする時間スケールを取つ

て遺伝子頻度を  $Y_N(t) = \frac{X([Nt], N)}{N}$  ( $[\cdot]$  はガウス記号) とする。  $N \rightarrow \infty$  の極限を取ると

$Y_N(t)$  は次のコルモゴロフ後退方程式で表現される拡散過程  $Y_t$  に収束することが示される。

$$\frac{\partial}{\partial t} P_t(x, y) = \frac{\sigma}{2} x(1-x) \frac{\partial^2}{\partial x^2} P_t(x, y) + \{\mathcal{G}_1 - (\mathcal{G}_1 + \mathcal{G}_2)x\} \frac{\partial}{\partial x} P_t(x, y) \quad (2.9)$$

ただし、  $P_t(x, y) = P(Y_t = y | Y_0 = x)$  (ただし  $P_0(x, y) = \delta_{x,y}$ ) は拡散過程  $Y_t$  の推移確率で

ある。この拡散方程式(2.9)の厳密解は Crow and Kimura(1956)および Goldberg(1950)によって次のように表現される。

$$P_t(x, y) = \sum_{k=0}^{\infty} X_k(y) \exp[-k\{(\mathcal{G}_1 + \mathcal{G}_2) + \frac{\sigma(k-1)}{2}\}t] \quad (2.10)$$

ただし

$$X_k(y) = y^{B-1}(1-y)^{(A-B)-1} F(A+k-1, -k, A-B, 1-y) \times F(A+k-1, -k, A-B, 1-x) \\ \times \frac{\Gamma(A-B+k)\Gamma(A+2k)\Gamma(A+k-1)}{k!\Gamma^2(A-B)\Gamma(B+k)\Gamma(A+2k-1)}$$

ここで  $F(, , ,)$  は超幾何関数 (Appendix (B)参照)、  $\Gamma(\cdot)$  はガンマ関数 (Appendix(A)参照)

$A = \frac{2(\mathcal{G}_1 + \mathcal{G}_2)}{\sigma}$ 、  $B = \frac{2\mathcal{G}_1}{\sigma}$  である。さらに  $t \rightarrow \infty$  とすると Wright によっても知られていた

$$\text{定常分布へ収束する } \lim_{t \rightarrow \infty} P_t(x, y) = \frac{\Gamma(A)}{\Gamma(B)\Gamma(A-B)} y^{B-1}(1-y)^{A-B-1} \quad (\text{ベータ分布}).$$

詳しくは Crow and Kimura(1970)を参照されたい。

さらに多くのアレルタイプがある場合には、  $A_1, A_2, \dots, A_d$  の  $d$  個のアレルを仮定し多タイプの分枝過程から、上記と同様の極限操作により、拡散過程に収束することが示される。

アレル  $A_k$  の頻度を  $x_k$  ( $1 \leq k \leq d$ )、ただし  $\sum_{k=1}^d x_k = 1$  とする。  $X = (x_1, x_2, \dots, x_{d-1})$  とすると

$d - 1$ 次元空間の単体領域  $D = \{X = (x_1, \dots, x_{d-1}); x_k \geq 0 (1 \leq k \leq d-1), \sum_{k=1}^{d-1} x_k \leq 1\}$  上の次のコルモゴロフ後退方程式で表現される拡散過程が得られる。

$$\frac{\partial}{\partial t} P_t(X, Y) = \frac{\sigma}{2} \sum_{p, q=1}^{d-1} x_p (\delta_{p,q} - x_q) \frac{\partial^2}{\partial x_p \partial x_q} P_t(X, Y) + \sum_{p=1}^{d-1} b_p(X) \frac{\partial}{\partial x_p} P_t(X, Y) \quad (2.11)$$

$$\text{ただし } b_p(X) = \sum_{k=1}^{d-1} x_k \vartheta_{kp} + (1 - \sum_{k=1}^{d-1} x_k) \vartheta_{dp} + (1 - x_p) m_p - x_p \left( \sum_{k=1}^d m_k - m_p \right)$$

ここで、 $\vartheta_{kj}$  はアレル  $A_k$  から  $A_j$  への突然変異率、 $m_p$  は移住を表す。さらに自然選択を含む分枝過程から拡散過程への収束については佐藤 (Satou(1978)) によって示されている。

モランモデルについても拡散過程への収束を論じることができるが、このときは  $N^2$  世代を単位時間とする時間スケールを取り、遺伝子頻度を  $Y_N(t) = \frac{X([N^2 t], N)}{N}$  とし、

$N \rightarrow \infty$  の極限を取ることにより得られる。(2.11) のような複数の対立遺伝子を含む多次元拡散過程の推移確率については一般的には求められていないが、中立遺伝子で突然変異に関するある条件の下では解が求められている。アレル  $A_k$  から  $A_j$  への突然変異率を  $u_{k,j}$  とし

$u_{k,j} = u_j$  すなわち  $j$  のみに依存するとき、(2.11) で移住は無く、 $\sigma = \frac{1}{2}$  となるように時間スケールを取ると

$$\frac{\partial}{\partial t} P_t(X, Y) = \frac{1}{4} \sum_{p, q=1}^{d-1} x_p (\delta_{p,q} - x_q) \frac{\partial^2}{\partial x_p \partial x_q} P_t(X, Y) + \sum_{p=1}^{d-1} \{u x_p - u_k\} \frac{\partial}{\partial x_p} P_t(X, Y) \quad (2.12)$$

定常分布は前進方程式

$$\frac{1}{4} \sum_{p, q=1}^{d-1} y_p (\delta_{p,q} - y_q) \frac{\partial^2}{\partial y_p \partial y_q} P(Y) + \sum_{p=1}^{d-1} \frac{\partial}{\partial y_p} \{(u_k - u y_k) P(Y)\} = 0, \quad (2.13)$$

を解いて次のディレクレ分布が得られる。

$$P(Y) = \frac{\Gamma(4u)}{\Gamma(4u_1) \dots \Gamma(4u_d)} \prod_{k=1}^d y_k^{4u_k - 1} \quad \text{ただし } u = \sum_{k=1}^d u_k \quad (2.14)$$

方程式(2.12)の解、すなわち多次元の推移確率  $P_t(X, Y)$  については Shimakura(1977), Griffiths(1978), Perlow(1979)を参照されたい。Ohta and Kimura(1972)は当時電気泳動法によって発見されたタンパク質多型を中立説の立場から説明するモデルとして1次元格子上の格子点を各アレルの状態と見なしたステップワイズ突然変異モデル(Stepwise mutation model)を提案した。各整数を1つのアレルと考え、遺伝子頻度を

$X = (\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots)$ 、ただし  $\sum_{k=-\infty}^{\infty} x_k = 1$  とする。このとき次のような無限次元の拡散方程式で表現される。

$$\frac{\partial}{\partial t} P_i(X, Y) = \frac{1}{2} \sum_{k,j} x_k (\delta_{k,j} - x_j) \frac{\partial^2}{\partial x_k \partial x_j} P_i(k, j) + \sum_{k=-\infty}^{\infty} \left( \sum_{j=-\infty}^{\infty} x_j u_{j,k} \right) x_k \frac{\partial}{\partial x_k} P_i(X, Y) \quad (2.15)$$

$u_{j,k}$  はアレル  $j$  から  $k$  への突然変異率である。さらにこのモデルはアレルの状態空間を一般化した測度値確率過程(Measure valued process)として研究されている。詳しくは Ethier and Kurtz(1993)、Feng(2010)などを参照されたい。

## 2. 4 自然選択による遺伝子頻度の変化

ダーウインの自然選択説では、適者生存すなわち環境に適応し、生存力、繁殖力が勝った個体の子供が集団内で多数を占めるようになる。もちろん子供とは繁殖年齢まで達する子供を意味している。この傾向が幾世代も遺伝して継続されると、その個体の子孫が集団の多数を占めることになる。ある個体が生きて生んだ子供のうち繁殖年齢まで成長できた子供の数を個体適応度という。自然選択が働くのは個体レベルであり、個々の遺伝子ではない。ある遺伝子型(対立遺伝子)に対して、周りの遺伝子について平均した適応度をその遺伝子型(対立遺伝子)の遺伝的適応度と呼ぶ。遺伝的適応度は相対適応度として、1を基準として表される。ある遺伝形質(遺伝子型あるいは対立遺伝子)の遺伝的適応度が1より大きいときは、その遺伝子は集団内で世代とともに増えてゆく傾向があり、1より小さいときは減少の傾向がある。

二倍体生物、 $A_1, A_2$  の二つの対立遺伝子を持つ遺伝子座を仮定して、自然選択の下での遺伝子頻度の変化を考える。遺伝子型は  $A_1A_1, A_1A_2, A_2A_2$  の3つのタイプがある。対立遺伝子  $A_1, A_2$  の頻度を  $p, q$  ( $p+q=1$ )、任意交配を仮定すると、それぞれの適応度、頻度などは以下ようになる。

表 2.1

遺伝子型	$A_1A_1$	$A_1A_2$	$A_2A_2$
自然選択前頻度	$p^2$	$2pq$	$q^2$
各遺伝子型の適応度	$w_{11}$	$w_{12}$	$w_{22}$
自然選択後頻度	$p^2 w_{11} / \bar{w}$	$2pq w_{12} / \bar{w}$	$q^2 w_{22} / \bar{w}$
平均適応度	$\bar{w} = p^2 w_{11} + 2pq w_{12} + q^2 w_{22}$		

次世代の対立遺伝子  $A_1$  の頻度を  $p'$  とすると

$$p' = \frac{p^2 w_{11} + pq w_{12}}{\bar{w}} = \frac{p(p w_{11} + q w_{12})}{\bar{w}} = \frac{p w_1}{\bar{w}}, \quad (2.16)$$

ただし  $w_1 = p w_{11} + q w_{12}$  ( $A_1$  の平均適応度)

これより

$$\Delta p = p' - p = \frac{p(w_1 - \bar{w})}{\bar{w}} = \frac{pq\{p(w_{11} - w_{12}) + q(w_{12} - w_{22})\}}{\bar{w}} \quad (2.17)$$

遺伝子頻度の変化は以下のようになる。

(1) 優劣がない場合(No dominance):  $w_{12} = (w_{11} + w_{22})/2$

$w_{11} > w_{22}$  とすると  $w_{11} > w_{12} > w_{22}$ 、このとき  $\Delta p > 0$  より  $p$  は増加。

$p \rightarrow 1$  ( $q \rightarrow 0$ )

(2)  $A_1$  の完全優性(dominance):  $w_{11} = w_{12}$

このとき  $\Delta p = \frac{pq^2(w_{12} - w_{22})}{\bar{w}}$ 。

$w_{11} = w_{12} > w_{22}$  のとき、 $\Delta p > 0$  より  $p \rightarrow 1$  ( $q \rightarrow 0$ )、

$w_{11} = w_{12} < w_{22}$  のとき、 $\Delta p < 0$  より  $p \rightarrow 0$  ( $q \rightarrow 1$ )。

(3) 超優性(Over-dominance):  $w_{11} < w_{12} > w_{22}$

$\Delta p = \frac{pq\{p(w_{11} - w_{12}) + q(w_{12} - w_{22})\}}{\bar{w}} = 0$ ,  $p > 0, q > 0$  とすると

$p(w_{11} - w_{12}) + q(w_{12} - w_{22}) = 0$ 。

条件  $w_{11} < w_{12} > w_{22}$  より  $0 < \tilde{p} = \frac{w_{12} - w_{22}}{(w_{12} - w_{11}) + (w_{12} - w_{22})} < 1$ 。

世代が進むにつれて、遺伝子頻度は平衡頻度  $\hat{p}$  に収束する。

超優勢では遺伝的多型(対立遺伝子  $A_1, A_2$  の共存)が可能であるが、(1)(2)の場合は最終的に遺伝的多型が消失してゆく。遺伝的多様性を保持するには、突然変異の存在が不可欠である。適応度について  $w_{11} \geq w_{12} \geq w_{22}$  のとき、対立遺伝子  $A_2$  は消失してゆくの、多型を保持するには  $A_1$  から  $A_2$  への突然変異が必要である。この突然変異率を  $u$  としよう。次世代の  $A_1$  の頻度は自然選択後の頻度にさらに突然変異を起こさない確率

$1-u$  を掛けて  $p' = \frac{p(pw_{11} + qw_{12})(1-u)}{\bar{w}}$  となる。

自然選択と突然変異の存在下で平衡頻度を求めてみよう。

相対的適応度を  $w_{11} = 1, w_{12} = 1 - hs, w_{22} = 1 - s$  とする。  $s$  を選択係数、  $h$  を優性の度合いと呼ぶ。  $0 < s, h < 1$  とする。このとき、平衡条件  $p' = p$  より

$p(pw_{11} + qw_{12})(1-u) = p\bar{w}$ ,  $0 < p$  とすると  $(pw_{11} + qw_{12})(1-u) = \bar{w}$ 。

$A_2$  の頻度  $q$  について方程式を表すと、

$\{(1-q) + (1-hs)q\}(1-u) = (1-q)^2 + 2(1-hs)q(1-q) + (1-s)q^2$

整理すると、  $u + (1-u)hsq = 2hsq(1-q) + sq^2$

(i)  $h = 0$  ( $A_2$  の完全劣性) のとき

$u = sq^2$ , これより平衡頻度は  $\hat{q} = \sqrt{u/s}$  となる。

(ii)  $h > 0$  かつ  $h \gg 2\sqrt{u/s}$  のとき、

$(1-2h)sq^2 + (1+u)shq - u = 0$ 、通常  $0 < u \ll 1$  なので、近似的に

$(1-2h)sq^2 + shq - u = 0$ 。

$$\text{解は } \hat{q} = \frac{-sh + \sqrt{s^2h^2 + 4u(1-2h)s}}{2(1-2h)s} = \frac{-sh + sh\sqrt{1 + 4u(1-2h)/(sh^2)}}{2(1-2h)s}$$

$\left| \frac{4(1-2h)u}{sh^2} \right| \ll 1$  のとき、すなわち  $h^2 \gg \left| \frac{4(1-2h)u}{s} \right|$ 、例えば  $h \gg 2\sqrt{\frac{u}{s}}$  のとき

$$\text{平衡頻度は } q \cong \frac{-sh + sh(1 + 2u(1-2h)/(sh^2))}{2(1-2h)s} = \frac{u}{sh} \quad \text{となる。}$$

離散世代での遺伝子頻度変化を考察したが、連続時間で考えてみよう。

適応度を  $w_{11} = 1 + s_{11}\Delta t$ ,  $w_{12} = 1 + s_{12}\Delta t$ ,  $w_{22} = 1 + s_{22}\Delta t$ 、突然変異率を  $A_1 \rightarrow A_2$  を  $u\Delta t$ ,  $A_2 \rightarrow A_1$  を  $v\Delta t$  とする。

平均適応度は  $\bar{w} = 1 + \bar{s}\Delta t$ , ただし  $\bar{s} = p^2s_{11} + 2pqs_{12} + q^2s_{22}$ 。このとき

$$p' = \frac{p\{p(1 + s_{11}\Delta t) + q(1 + s_{12}\Delta t)\}}{1 + \bar{s}\Delta t} (1 - u\Delta t) + \left\{ 1 - \frac{p\{p(1 + s_{11}\Delta t) + q(1 + s_{12}\Delta t)\}}{1 + \bar{s}\Delta t} \right\} v\Delta t$$

$$\Delta p = p' - p = \frac{pq\{(p(s_{11} - s_{12}) + q(s_{12} - s_{22})) - up + qv\}\Delta t + O((\Delta t)^2)}{1 + \bar{s}\Delta t},$$

$$\Delta t \rightarrow 0 \text{ とすると、} \frac{dp}{dt} = pq\{p(s_{11} - s_{12}) + q(s_{12} - s_{22})\} + vq - up \text{ をえる。} \quad (2.18)$$

さらに有限集団における遺伝子頻度の機会的浮動(random genetic drift)を考慮に入れ、推移確率を  $P_i(x, y)$  とすると、コルモゴロフ前向き方程式は

$$\frac{\partial}{\partial t} P_i(x, y) = \frac{1}{4} \frac{\partial^2}{\partial y^2} \{y(1-y)P_i(x, y)\} + \frac{\partial}{\partial y} \{(pq(ps_1 - qs_2) + vq - up)P_i(x, y)\} \quad (2.19)$$

ただし  $s_1 = s_{11} - s_{12}$ ,  $s_2 = s_{22} - s_{12}$

定常分布は  $\frac{1}{4} \frac{\partial^2}{\partial y^2} \{y(1-y)P(y)\} + \frac{\partial}{\partial y} \{(pq(ps_1 - qs_2) + vq - up)P(y)\} = 0$  を解き、

ライトの公式より  $P(y) = Ce^{2sx}x^{2v-1}(1-x)^{2ux-1}$  (Cは規格化定数) を得る。 (2.20)

詳しくは参考文献(Crow and Kimura(1970))を参照されたい。

環境の変動により自然選択圧が時間とともに変化する時も遺伝子頻度の確率的变化が生じ、集団遺伝学では、これも広い意味で「遺伝子頻度の機会的変動」に含まれる。自然選択の確率的变化に伴う遺伝子頻度の拡散過程モデルも重要なモデルであるが、Kimura(1954), Takahata, Ishii and Matsuda(1975)などを参照されたい。