

## 集団遺伝学における拡散モデルのジャンプ型モデルへの収束

筑紫女学園短期大学・一般教育 飯塚 勝 (Masaru Iizuka)

### §1. 離散時間モデル, 拡散モデル, ジャンプ型モデル

集団遺伝学においては、モデルを記述する基本方程式が差分方程式で与えられることが多い(離散時間モデル)。確率的自然淘汰(環境のランダムな変動)や、集団の個体数有限によるランダム・サンプリング・ドリフトなど確率的要因が存在するときは、それは確率差分方程式となり、その解は離散時間確率過程となる。しかしながら、これらの確率過程は一般に非常に複雑なので、これらを連続時間確率過程、とくに、拡散過程で近似することが多い(拡散モデル)。

離散時間モデルから連続時間モデルへの収束の確率過程論的証明も多岐行なわれている(Iizuka and Matsuda (1982), Iizuka (1987), および、そこに引用された文献を参照)。

集団中に二つのタイプ(アレル),  $A_1, A_2$ , が存在する場合の拡散モデルを考える。 $A_1$ の時刻 $t$ における集団中の相対

的頻度 (遺伝子頻度) を  $x(t)$ ,  $t \geq 0$  とする。  $x(t)$  の時間変化の要因としては、突然変異, 自然淘汰, ランダム・サンプリング・ドリフトなどがある。突然変異率を  $\mu$  (簡単のために、突然変異は  $A_1$  と  $A_2$  の間で対称とする)、自然淘汰の強度を  $S$ , 集団の個体数を  $N$  とする。ここで、決定論的自然淘汰の場合、 $S$  は  $A_1$  の淘汰有利度  $s$  であり、確率的自然淘汰の場合、 $S$  は  $A_1$  の淘汰有利度の自相関  $V$  に対応する。拡散モデルの下で、Kolmogorov の後向き方程式

$$\frac{dE_x[f(x(t))]}{dt} = E_x[Lf(x(t))] \quad (1.1)$$

が成立つ。但し、 $E_x[f(x(t))]$  は  $f(x(t))$  の  $x(0) = x$  の下での期待値であり、生成作用素  $L$  は典型的な場合

$$\left. \begin{aligned} Lf(x) &= \left\{ a(\alpha, x) \frac{d^2}{dx^2} + b(\alpha, \theta, x) \frac{d}{dx} \right\} f(x) \\ a(\alpha, x) &= x(1-x) + c\alpha x^2(1-x)^2 \\ b(\alpha, \theta, x) &= \alpha x(1-x)g(x) + \theta(1-2x) \end{aligned} \right\} \quad (1.2)$$

と表わされる。  $\alpha$  は  $NS$ ,  $\theta$  は  $N\mu$  に対応する。決定論的自然淘汰の場合は  $c=0$  で、遺伝子淘汰モデルでは  $g(x)=1$ , 超優性モデルでは  $g(x)=1-2x$  である (Crew and Kimura (1970), Ewens (1979))。確率的自然淘汰の場合は  $c=1$ ,  $g(x)=\beta(1-2x)$ ,

$\beta \geq 1$  である。SAS-CFF モデル (Gillespie (1978)) では  $\beta > 1$  であり、TIM モデル (Takahata, Ishii and Matsuda (1975)) では  $\beta = 1$  となる。 $x(t)$  の確率密度関数  $\phi(t, x)$  が存在すれば

Kolmogorov の前向き方程式

$$\frac{\partial \phi(t, x)}{\partial t} = L^* \phi(t, x) \quad (1.3)$$

が成立つ。但し、 $L^*$  は  $L$  の共役作用素である。定常分布  $\phi(x)$  が存在すれば、

$$L^* \phi(x) = 0 \quad (1.4)$$

をみたす。

集団中に  $K$  種類のアレル,  $A_1, A_2, \dots, A_K$ , が存在するときは、 $A_i$  の時刻  $t$  における遺伝子頻度を  $x_i(t)$ ,  $i=1, 2, \dots, K$  とし、 $x(t) = (x_1(t), x_2(t), \dots, x_{K-1}(t))$ ,  $x = (x_1, x_2, \dots, x_{K-1})$  とすると、拡散モデルの下で、(1.1), (1.2), (1.4) が成立つ。但し、 $L$  は

$$L = \sum_{i,j=1}^{K-1} a_{ij}(\alpha, x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{K-1} b_i(\alpha, x) \frac{\partial}{\partial x_i} \quad (1.5)$$

である。 $K \geq 3$  に対して、たとえば、(1.4) を explicit に解くことは困難なことが多い(いわゆる、フラックワス・ゼロをみたす解が存在しないことが多い)。さらに、 $L, L^*$  は境界で退化した偏微分作用素であり、(1.1), (1.2), (1.4) の教

値解析も困難な場合が多い。

したがって、拡散モデルを、再び、より簡単なモデルで近似することが考えられる。もちろん、モデルのパラメータ、 $\alpha, \theta$  に関する条件が必要となるであろう。実際、Gillespie (1983a, 1983b) は次の予想を提起した。  $x^\varepsilon(t) = (x_1^\varepsilon(t), x_2^\varepsilon(t), \dots, x_{k-1}^\varepsilon(t))$  を

$$L^\varepsilon = \sum_{i,j=1}^{k-1} a(\alpha_\varepsilon, x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{k-1} b(\alpha_\varepsilon, \theta_\varepsilon, x) \frac{\partial}{\partial x_i} \quad (1.6)$$

を生成作用素とする拡散過程とする。 strong selection - weak mutation limit, すなわち  $\alpha_\varepsilon \rightarrow \infty, \theta_\varepsilon \rightarrow 0$  ( $\varepsilon \rightarrow 0$ ) のとき、  $x^\varepsilon(t)$  はジャンプ型マルコフ過程  $x(t)$  に収束する。但し、  $\alpha_\varepsilon \theta_\varepsilon = \text{const.}$  とする。さらに、  $x(t)$  の状態空間は、決定論的自然淘汰の場合は、  $\{x \mid b(1,0,x) = 0\}$  で有限個の点からなる。確率的自然淘汰の場合は、(1.6) で  $\alpha_\varepsilon = \alpha$  (正定数),  $\theta_\varepsilon = 0$  とおき、さらに、  $\alpha_\varepsilon, \theta_\varepsilon$  に依存しない項を 0 とおいた生成作用素に対応する拡散過程に定常分布が存在するとして (一般に複数個)、  $x(t)$  はこれら有限個の定常分布の値をとる measure-valued ジャンプ型マルコフ過程である。

まず、この予想に関するいくつかの疑向、問題点を述べる。

(1) ここでは収束の意味は明らかにされていないが、  $x^\varepsilon(t)$  に対応する関数空間上の確率測度の弱収束がありえないことは容易に分る。したがって、より弱い収束概念、たとえば、有

限次元分布の収束が、その可能性として考えられる。(2)ここでは、いかなる time-scaling の下で極限をとるかが明らかにされていない。(3)  $\alpha_2$  と  $\theta_2$  の関係は  $\alpha_2 \theta_2 = \text{const.}$  であるか。(4) Gillespie の予想は確率過程論的に新しいタイプ(連続からジャンプ型へ)の極限定理を主張している。

以下において、Gillespie の予想を肯定的に証明する。但し、以下の方法は一次元確率過程 ( $k=2$ ) にのみ適用できるものである。したがって、実際に応用上必要とされる  $k \geq 3$  の場合は未解決である。しかしながら、上にも述べたように、一次元の場合ですら不明な点が多いので、まず、この場合に明確な結果を得ておくことは重要であると思われる。なお、関連した文献としては、Freidlin and Wentzell (1984), Kipnis and Newman (1985) がある。前者は多次元の場合を一般的に扱っているが、Gillespie の予想に対応するような具体的な極限定理は考察されていない。後者は、物理学における metastable behavior (力学系に微小な確率的擾動が加わったときの系の漸近挙動) や量的形質の遺伝に対応して、一次元で生成作用素が

$$L^\varepsilon = \varepsilon \frac{d^2}{dx^2} - G'(x) \frac{d}{dx} \quad (1.7)$$

なる場合にジャンプ型マルコフ過程への有限次元分布の収束

を証明している。但し、 $q(x)$  は double-well 型ポテンシャルで、 $q'(x)$  はその導関数である。この証明においては (1.7) の右辺の1項が  $\alpha$  に依存しないことが本質的であり、したがって、Gillespie の予想に対しては有効でない。

## §2. bi-generalized diffusion processes と極限定理

一次元拡散過程  $x^\varepsilon(t)$  の生成作用素  $L^\varepsilon$  の

$$L^\varepsilon = a_\varepsilon(x) \frac{d^2}{dx^2} + b_\varepsilon(x) \frac{d}{dx} \quad (2.1)$$

なる表現を Kolmogorov の標準形という。通常の極限定理では、 $a_\varepsilon(x)$  と  $b_\varepsilon(x)$  の収束 ( $\varepsilon \rightarrow 0$ ) から有限次元分布の収束を導くことが多い (semigroup approximation theorem)。しかし、Gillespie の予想の場合は、 $\varepsilon \rightarrow 0$  のとき  $a_\varepsilon(x)$  と  $b_\varepsilon(x)$  が発散してしまうので、この方法は有効でない。一方、(2.1) は Feller の標準形

$$L^\varepsilon = \frac{d}{dm_\varepsilon(x)} \frac{d}{dS_\varepsilon(x)} \quad (2.2)$$

$$S_\varepsilon(x) = \int^x e^{-B_\varepsilon(y)} dy : \text{scale function} \quad (2.3)$$

$$m_\varepsilon(x) = \int^x \frac{1}{a_\varepsilon(y)} e^{B_\varepsilon(y)} dy : \text{speed measure} \quad (2.4)$$

と表現することができる。ここで、 $B_2(x) = \int^x \frac{b_2(y)}{a_2(y)} dy$  である。実は、単調増加で連続な関数  $S_2(x)$  と単調増加で右連続な関数  $m_2(x)$  を与えると、一次元拡散過程が一意的に決定される (Itô and McKean (1965))。

さらに、条件を弱めて、単調増加で連続な関数  $S_2(x)$  と非減少で右連続な関数  $m_2(x)$  を与えると、一意的に一次元確率過程が決定される。このクラスを *generalized diffusion processes* という。birth and death process はその例である (Kotani and Watanabe (1982))。birth and death process はジャンプ型マルコフ過程であるから、 $S_2(x)$  と  $m_2(x)$  の収束から有限次元分布の収束が言えれば、極限がジャンプ型マルコフ過程となりうる事が十分考えられる。

しかしながら、 $S_2(x)$  の極限を考えると、単調増加性と連続性が成立しない場合もある。したがって、 $S_2(x)$  と  $m_2(x)$  の収束から有限次元分布の収束を言おうとするとき、*generalized diffusion processes* をさらに一般しておく方が都合がよいと考えられる。実際、Ogura (1988) は、非減少関数  $S_2(x)$  と非減少かつ右連続な関数  $m_2(x)$  を与えると、一意的に一次元確率過程が決定することを明らかにし、このクラスを *bi-generalized diffusion processes* と名付けた。さらに、Ogura (1988) は次の極限定理 (詳細な条件は原論文を参照) を証明した。

### 極限定理 (Y. Ogura)

$S_2(x), s(x)$  を非減少関数とする。  $m_2(x), m(x)$  を非減少かつ右連続な関数とする。  $x^\varepsilon(t), x(t)$  をそれぞれ  $(S_\varepsilon(x), m_\varepsilon(x), c(x), m(x))$  に対応する bi-generalized diffusion process とする。このとき、 $s(x)$  の連続点で  $S_\varepsilon(x)$  が  $s(x)$  に、 $m(x)$  の連続点で  $m_\varepsilon(x)$  が  $m(x)$  に収束するならば、補足的条件の下で、 $t > 0$  に対して、 $x^\varepsilon(t)$  の有限次元分布は  $x(t)$  のそれに収束する。

この極限定理を用いて、次節で Gillespie の予想を証明する。なお、与えられた  $s(x), m(x)$  から対応する  $x(t)$  を同定する方法は Minami, Ogura and Tomisaki (1985) や Ogura (1988) を参照。

### §3. 結果

§1 で述べた拡散モデル (一次元) の列  $x^\varepsilon(t)$ ,  $\varepsilon > 0$  を考える。すなわち、 $x^\varepsilon(t)$  は、

$$\left. \begin{aligned} L^\varepsilon &= a_\varepsilon(x) \frac{d^2}{dx^2} + b_\varepsilon(x) \frac{d}{dx} \\ a_\varepsilon(x) &= x(1-x) + c\alpha_\varepsilon x^2(1-x)^2 \\ b_\varepsilon(x) &= \alpha_\varepsilon x(1-x)g(x) + \theta_\varepsilon(1-2x) \end{aligned} \right\} (3.1)$$



を生成作用素とする一次元拡散過程である。  $a, b$  ( $0 \leq a < b \leq 1$ )  
 に対して

$$I_-^\varepsilon(a, b) = \int_a^b e^{-B_\varepsilon(x)} dx \quad (3.2)$$

$$I_+^\varepsilon(a, b) = \int_a^b \frac{1}{q_\varepsilon(x)} e^{B_\varepsilon(x)} dx \quad (3.3)$$

とおく。但し、 $B_\varepsilon(x) = \int^x \frac{b_\varepsilon(y)}{a_\varepsilon(y)} dy$  である。さらに、 $\delta$   
 ( $0 < \delta \leq 1$ ) に対して、

$$\tau_\varepsilon = \tau_\varepsilon(\delta) = I_+^\varepsilon(0, \delta) I_-^\varepsilon(0, 1) \quad (3.4)$$

とおく。  $x^\varepsilon(t)$  に対して時間を  $\tau_\varepsilon$  で scale した確率過程を  
 $X_\varepsilon(t) = x^\varepsilon(\tau_\varepsilon t)$  とする。境界  $\{0\}$  または  $\{1\}$  が正則境界の場合  
 は、集田遺伝学で通常行なわれているように、反射壁条件  
 をおくことにする。このとき、 $X_\varepsilon(t)$  の生成作用素  $L_\varepsilon$  は  
 $L_\varepsilon = \tau_\varepsilon L^\varepsilon$  であり、scale function  $S_\varepsilon(x)$ , speed measure  $m_\varepsilon(x)$  は

$$S_\varepsilon(x) = \begin{cases} I_-^\varepsilon(0, x) / I_-^\varepsilon(0, 1) & (0 \leq x \leq 1) \\ x & (\text{otherwise}) \end{cases} \quad (3.5)$$

$$m_\varepsilon(x) = \begin{cases} 0 & (x \leq 0) \\ I_+^\varepsilon(0, x) / I_+^\varepsilon(0, \delta) & (0 \leq x \leq 1) \\ I_+^\varepsilon(0, 1) / I_+^\varepsilon(0, \delta) & (x \geq 1) \end{cases} \quad (3.6)$$

となる。前節の極限定理を用いて、以下の結果が証明され

る。なお、収束はすべて有限次元分布の意味である。

(1) 決定論的遺伝子淘汰モデル ( $c=0, g(x)=1$ )

$\alpha_\varepsilon \rightarrow \infty, \theta_\varepsilon \rightarrow 0$  とする。但し、 $\lim_{\varepsilon \rightarrow 0} \frac{\log \theta_\varepsilon}{\alpha_\varepsilon} \leq -1$  とする。  
任意の  $\delta$  ( $0 < \delta < 1$ ) を固定する。このとき、

$$S_\varepsilon(x) \rightarrow S(x) = \begin{cases} 1 & (0 < x \leq 1) \\ x & (\text{otherwise}) \end{cases} \quad (3.7)$$

$$m_\varepsilon(x) \rightarrow m(x) = \begin{cases} 0 & (x < 0) \\ 1 & (0 \leq x < 1) \\ \infty & (x \geq 1) \end{cases} \quad (3.8)$$

がそれぞれ  $S(x), m(x)$  の連続点で成立つ。したがって、 $t > 0$  に対して、 $x_\varepsilon(t)$  は次のジャンプ型マルコフ過程  $x(t)$  に収束する。

$$\left. \begin{aligned} P(x(t) = 0 \mid x(0) = 0) &= e^{-t} \\ P(x(t) = 1 \mid x(0) = 0) &= 1 - e^{-t} \end{aligned} \right\} \quad (3.9)$$

$$P(x(t) = 1 \mid 0 < x(0) \leq 1) = 1 \quad (3.10)$$

(2) 決定論的超優性モデル ( $c=0, g(x)=1-2x$ )

$\alpha_\varepsilon \rightarrow \infty, \theta_\varepsilon \rightarrow 0$  とする。但し、 $\lim_{\varepsilon \rightarrow 0} \frac{\log \alpha_\varepsilon}{\theta_\varepsilon} \leq -\frac{1}{4}, \frac{\theta_\varepsilon}{\sqrt{\alpha_\varepsilon}} e^{\alpha_\varepsilon/4} \rightarrow \infty$  とする。任意の  $\delta$  ( $0 < \delta < \frac{1}{2}$ ) を固定する。このとき

$$S_{\varepsilon}(x) \rightarrow S(x) = \begin{cases} \frac{1}{2} & (0 < x < 1) \\ x & (\text{otherwise}) \end{cases} \quad (3.11)$$

$$m_{\varepsilon}(x) \rightarrow m(x) = \begin{cases} 0 & (x < 0) \\ 1 & (0 \leq x < \frac{1}{2}) \\ \infty & (x \geq \frac{1}{2}) \end{cases} \quad (3.12)$$

がそれぞれ  $S(x), m(x)$  の連続点で成立つ。したがって、 $t > 0$  に対して、 $x_{\varepsilon}(t)$  は次のジャンプ型マルコフ過程  $x(t)$  に収束する。

$$\left. \begin{aligned} P(x(t) = 0 \mid x(0) = 0) &= e^{-2t} \\ P(x(t) \geq \frac{1}{2} \mid x(0) = 0) &= 1 - e^{-2t} \end{aligned} \right\} \quad (3.13)$$

$$P(x(t) \geq \frac{1}{2} \mid 0 < x(0) < \frac{1}{2}) = 1 \quad (3.14)$$

(3) SAS-CFF モデル ( $c=1, g(x) = \beta(1-2x), \beta > 1$ )

$\alpha_{\varepsilon} \rightarrow \infty, \theta_{\varepsilon} \rightarrow 0, S=1$  とする。但し、 $\frac{\log \theta_{\varepsilon}}{\log \alpha_{\varepsilon}} \rightarrow 0$  とする。

このとき、

$$S_{\varepsilon}(x) \rightarrow S(x) = \begin{cases} \frac{1}{2} & (0 < x < 1) \\ x & (\text{otherwise}) \end{cases} \quad (3.15)$$

$$m_{\varepsilon}(x) \rightarrow m(x) = \begin{cases} 0 & (x \leq 0) \\ \frac{\int_0^x [y(1-y)]^{\beta-2} dy}{\int_0^1 [y(1-y)]^{\beta-2} dy} & (0 \leq x \leq 1) \\ 1 & (x \geq 1) \end{cases} \quad (3.16)$$

がそれぞれ  $S(x), m(x)$  の連続点で成立つ。したがって、 $t > 0$

に対して、 $X_\varepsilon(t)$  は次の確率過程  $x(t)$  に収束する。  $B$  を  $[0, 1]$  上の任意の Borel 集合として、

$$P(x(t) \in B) = \int_B dm(x) \quad (3.17)$$

但し、 $dm(x)$  は (3.16) で定義された  $m(x)$  より導かれた確率測度、すなわち、突然変異とランダム・サンプリング・ドリフトが不在のときの SAS-CFF モデルの定常分布である。

(4) TIM モデル ( $c=1, g(x) = 1-2x$ )

このとき、技術的な条件がチェックできていないので証明されていないが、形式的に考察すると次の命題が成立つことが期待される。  $\alpha_\varepsilon \rightarrow \infty, \theta_\varepsilon \rightarrow 0, \delta = 1$  とする。このとき、 $t > 0$  に対して、 $X_\varepsilon(t)$  は次のジャンプ型マルコフ過程  $x(t)$  に収束する。  $0 \leq x \leq 1$  とする。

$$\left. \begin{aligned} P(x(t)=0 | x(0)=x) &= \frac{1+(1-2x)e^{-4t}}{2} \\ P(x(t)=1 | x(0)=x) &= \frac{1-(1-2x)e^{-4t}}{2} \end{aligned} \right\} \quad (3.18)$$

(3.18) は突然変異とランダム・サンプリング・ドリフトが不在の場合、TIM モデルでは擬固定 (quasi-fixation) が起こることに対応している。

以上では、 $\alpha_2 \rightarrow \infty$ ,  $\theta_2 \rightarrow 0$  の場合のみを考察したが、 $\alpha_2 \rightarrow \alpha$  ( $0 < \alpha < \infty$ ),  $\theta_2 \rightarrow 0$  (moderate selection - weak mutation limit),  $\alpha_2 \rightarrow 0$ ,  $\theta_2 \rightarrow 0$  (weak selection - weak mutation limit) の場合も含めて系統的な解析の結果を報告準備中である (Iizuka and Ogura (1989))。

#### 引用文献

1. Crow, J.F. and M. Kimura (1970) An Introduction to Population Genetics Theory. Harper and Row, New York.
2. Ewens, W.J. (1979) Mathematical Population Genetics. Springer, Berlin-Heidelberg - New York.
3. Freedlin, M.I. and A.D. Wentzell (1984) Random Perturbations of Dynamical Systems. Springer, Berlin-Heidelberg - New York.
4. Gillespie, J.H. (1978) A general model to account for enzyme variation in natural populations, V. The SAS-CFF model. *Theoret. Population Biology*, 14, 1-45.
5. Gillespie, J.H. (1983a) Some properties of finite populations experiencing strong selection and weak mutation. *The American Naturalist*, 121, 691-708.

6. Gillespie, J. H. (1983b) A simple stochastic gene substitution model. *Theoret. Population Biology*, 23, 202-215.
7. Iizuka, M. (1987) Weak convergence of a sequence of stochastic difference equations to a stochastic ordinary differential equation. *J. Math. Biol.*, 25, 643-652.
8. Iizuka, M. and H. Matsuda (1982) Weak convergence of discrete time non-Markovian processes related to selection models in population genetics. *J. Math. Biol.*, 15, 107-127.
9. Iizuka, M. and Y. Ogura (1989) Convergence of one-dimensional diffusion processes to a jump process related to population genetics. (in preparation)
10. Itô, K. and H. P. McKean (1965) *Diffusion Processes and Their Sample Paths*. Springer, Berlin-Heidelberg - New York.
11. Kipnis, K. and C. M. Newman (1985) The metastable behavior of infrequently observed random, one-dimensional diffusion processes. *SIAM J. Appl. Math.*, 45, 972-982.
12. Kotani, S. and S. Watanabe (1982) Krein's spectral theory of strings and generalized diffusion processes. *Functional Analysis and Markov Processes*, 235-254. ed. by M. Fukushima, *Lecture Notes in Math.* 923, Springer, Berlin-Heidelberg - New York.

13. Minami, N., Y. Ogura and M. Tomisaka (1985) Asymptotic behavior of elementary solutions of one-dimensional generalized diffusion equations. *Ann. Probability*, 13, 693-715.
14. Ogura, Y. (1988) One-dimensional bi-generalized diffusion processes. *J. Math. Soc. Japan*, (in press)
15. Takahata, N., K. Ishii and H. Matsuda (1975) Effect of temporal fluctuation of selection coefficient on gene frequency in a population. *Proc. Nat. Acad. Sci. USA*, 72, 4541-4545.