

集団遺伝学における確率過程の問題
— 離散時間モデルの連続時間近似 —

九大理 松田博嗣, 石井一成

集団遺伝学の確率論的モデルにおいては、通常一世代当りの遺伝子頻度の変化に対する確率法則が仮定され、それより遺伝子頻度の平衡分布等を導くことが要請される¹⁾。世代当りの頻度変化が微小な場合、連続時間を用い、適当な確率微分方程式によってモデルを近似し、これを解くことによつて上記要請に近似的に応えるのが有用である。ここに於いて、確率性の起源が、集団の個体数が有限のためのサイズ効果、環境変化のための確率的自然選択かによつて、平衡分布にどのような特徴的差異が生ずるか等の問題があるが、これと相俟つて、上に述べた近似の正当づけが問われている。

さて、モデルは簡単な場合には、 n 世代における遺伝子頻度を x_n (n は整数)、 $\{\sigma_n\}$ を 2 状態 $\sigma_n = \pm 1$ をとる定常マルコフ鎖、 $F(x, \sigma)$, $G(x)$ を $x \in [0, 1]$ の連続実関数として、
$$\delta x_n \equiv x_{n+1} - x_n = \varepsilon F(x_n, \sigma_n) + \varepsilon^2 G(x) \quad (0 < \varepsilon \ll 1) \quad [1]$$

$$P(\sigma_{n+1} = -\sigma \mid \sigma_n = \sigma) = \delta^{(\varepsilon)} (> 0) \quad (\sigma = \pm 1) \quad [2]$$
によつて与えられる。²⁾ 推移確率 $\delta^{(\varepsilon)}$ は一般に ε に依存するとする。このモデルを連続時間上で定義された適当な確

率過程と対応 $n = n^{(\varepsilon)}(t)$, $x_n = x(t)$, $\sigma_n = \sigma(t)$ により近似した

1). γ を $O(1)$ の正定数として, (I) $\gamma^{(\varepsilon)} = \varepsilon\gamma$ のときは,

$$\frac{dx}{dt} = F(x(t), \sigma(t)), \quad P(\sigma(t+\Delta t) = -\sigma | \sigma(t) = \sigma) \approx \gamma \Delta t \quad (\Delta t \rightarrow 0) \quad [3]$$

なるマルコフ過程 $\{x(t), \sigma(t)\}$ に, $n^{(\varepsilon)} \equiv [t/\varepsilon]$; (II) $\gamma^{(\varepsilon)} = \gamma$ で,

$F(x, \sigma) = \sigma f(x)$ と書けるときは, Itô 方程式

$$dx = \left\{ \frac{G(x)}{V} + \frac{1}{2} f(x) \frac{df(x)}{dx} \left(1 - \frac{1}{V}\right) \right\} dt + f(x) dW \quad [4]$$

を充す拡散過程 $\{x(t)\}$ に, $n^{(\varepsilon)} \equiv [t/\varepsilon V]$, たゞし $V \equiv \sum_{n=-\infty}^{\infty} \langle \sigma_n \sigma_n \rangle$, が適当と考えられる。この理由として,

モデルマルコフ鎖 $\{x_n, \sigma_n\}$ の $\varepsilon \rightarrow 0$ における極限的性質をそれぞれの場合について述べる。³⁾

[3], [4] に対する平衡分布は適当な条件下で正確に求められる。確率変数 σ_n がサイズ効果を代表するときは, $\gamma^{(\varepsilon)} = \frac{1}{2}$, 従って場合(II)で, 近似的に $f(x) \propto \sqrt{x(1-x)}$; 確率的自然選択のときは, $\gamma^{(\varepsilon)} \leq \frac{1}{2}$, 従って場合(I), (II)が考えられ, $f(x) \propto x(1-x)$ である。突然変異の効果は $G(x)$ に含まれる。²⁾ 具体的に得られた平衡分布を解析すると, 確率性の起源がサイズ効果か, 確率的自然選択か, また時相関が強¹⁾ ($\gamma^{(\varepsilon)} \ll 1$) か, 弱¹⁾ ($\gamma^{(\varepsilon)} \approx \frac{1}{2}$) かによつて, モデルパラメタに対する分布の依存性の強弱を含む種々顕著な差異が見られることが判った。⁴⁾ また, 蛋白多型の実測データは, 確率的自然選択の効果がサイズ効果に優先すると考えた方が, 多くの場合理解し易いことが判った。⁵⁾

註

1) 集団遺伝学の確率論的モデルの入門書としては,

(1) Li, W, ed.: 'Stochastic models in population genetics,' (Dowden, Hutchinson & Ross, Inc. Stroudsburg, Pennsylvania, 1977)

(2) Maruyama, T.: 'Stochastic problems in population genetics.' (Springer-Verlag, Berlin, Heidelberg, New York, 1977)

(3) 福島正俊・石井一成 = '自然現象と確率過程' 入門 | 現代の数学 [10] (日本評論社, 1980)

などがある。(1)は編者の序文およびコメントに従って、1975年頃までの主要論文が集められており、この分野の研究全般の歴史的な流れを学ぶことが出来る。(2)はコルモゴロフの後退方程式を中心に据え、集団遺伝学における確率論的モデル研究の手法を生物系初学者に紹介した講義録である。地理的構造の扱りに特色があり、最近の結果まで含まれている。ただし、確率論的自然選択については余り述べられていない。(3)は中I部で確率過程の一般的考え方や手法を解説し、中II部で、その集団遺伝学への応用が紹介されている。総説的では無いが、揺動淘汰や生物進化モデルなどの新しい試みを含めて、ほぼ自己充足的に書か

れているのが特徴である。

集団遺伝学理論全般に渡り入門書としては、

木村資生(編): 'ヒト遺伝の基礎', 岩波講座, 現代生物
科学 6 (岩波書店, 1975) の2章, およびそこに紹
介されている参考書がある。また、

Roughgarden, J.: 'Theory of population gene-
tics and evolutionary ecology: an int-
roduction.' (Macmillan Pub. Co., Inc., New
York, 1979) もすすめられる。

2) モデル[1], [2]の由来:

(i) 確率的自然選択の場合: 遺伝子を2つのタイプに分類
し, 注目するタイプの遺伝子の個数を n 世代において N_n ,
他方のタイプのものを N_n^c とする。自己複製により次世代
にそれぞれ同じタイプの遺伝子も, 各遺伝子/個当り平均
 $w(\sigma_n)$ および $w^c(\sigma_n)$ 個作るとする。ただし, $\sigma_n = \pm 1$
は n 世代における環境を表わすパラメタである。 N_n ,
 N_n^c が十分に大きくて, 平均よりのずれの効果(サイズ効果)が
無視せられるとすると,

$$N_{n+1} = w(\sigma_n) N_n, \quad N_{n+1}^c = w^c(\sigma_n) N_n^c. \quad [A1]$$

$$\text{従って, } z_n \equiv \log \frac{x_n}{1-x_n}, \quad x_n \equiv \frac{N_n}{N_n + N_n^c}, \quad [A2]$$

$$\delta z_n \equiv z_{n+1} - z_n \quad \text{とおくと,} \quad [A3]$$

$$\delta Z_n = \log \left(\frac{w(\sigma_n)}{w^c(\sigma_n)} \right) \quad \text{となる.} \quad [A4]$$

注目するタイプの淘汰有利度 $\lambda(\sigma) \equiv \log(w(\sigma)/w^c(\sigma))$ と

ε のオ-ダの微小量とすると,

$$\begin{aligned} \delta x_n &= x_n(1-x_n) \delta Z_n + x_n(1-x_n) \left(\frac{1}{2} - x_n \right) (\delta Z_n)^2 + O(\varepsilon^3) \\ &= x_n(1-x_n) \lambda(\sigma_n) + x_n(1-x_n) \left(\frac{1}{2} - x_n \right) \{ \lambda(\sigma_n) \}^2 + O(\varepsilon^3) \end{aligned} \quad [A5]$$

である. ここで, 一般性を失うことなく,

$$\lambda(\sigma) = \bar{\lambda} + \varepsilon \sigma \quad (\bar{\lambda} \text{ は定数}) \quad [A6]$$

とおき, $\bar{\lambda} = O(\varepsilon^2)$ と仮定すると, [A5] より本文 [1]

に対応して,

$$F(x, \sigma) = x(1-x) \sigma \quad [A7]$$

$$G(x) = x(1-x) (\bar{\lambda} / \varepsilon^2) + x(1-x) \left(\frac{1}{2} - x \right) + (1-2x) (\mu / \varepsilon^2) \quad [A8]$$

が得られる. ただし, [A8] では, 突然変異率 μ

を $O(\varepsilon^2)$ とし, これによる x_n の変化が [A5] に重ね合わせられている. 環境 $\{\sigma_n\}$ はマルコフ鎖をなすと仮定し, 同一環境が持続し易いことより $\rho^{(t)} \leq \frac{1}{2}$ である.

(ii) サイズ効果は, 各遺伝子が次世代に作る遺伝子の数が正の分散 v をもつため, 平均数は一樣であっても, 有限集団では遺伝子頻度にランダムな変化を生ずる効果である.

これによる変化量の平均を $\langle \dots \rangle$ で表わすと,

$$\langle \delta N_n \rangle = \langle \delta N_n^c \rangle = 0 \quad [A9]$$

$$\langle \delta N_n \delta N_n^c \rangle = \langle \delta N_n \rangle \langle \delta N_n^c \rangle = 0$$

$$\langle (\delta N_n)^2 \rangle = \nu N_n, \quad \langle (\delta N_n^c)^2 \rangle = \nu N_n^c \quad [A10]$$

$$\text{なる仮定より, } N \equiv N_n + N_n^c \quad [A11]$$

として,

$$\langle \delta x_n \rangle = O\left(\frac{1}{N^2}\right) \quad [A12]$$

$$\langle (\delta x_n)^2 \rangle = \frac{\nu}{N} x_n(1-x_n) + O\left(\frac{1}{N^2}\right) \quad [A13]$$

である。従って、確率変数 σ_n でサイズ効果のランダム性

（近似的に代表することにより、時相関はゼロとする）

$\gamma^{(\varepsilon)} = \frac{1}{2}$, δx_n の平均と分散が $1/N$ によって最低次

で [A12], [A13] と一致するように,

$$\varepsilon F(x, \sigma) = \sqrt{\nu/N} \sqrt{x(1-x)} \sigma, \quad [A14]$$

突然変異の効果と前同様 $G(x)$ に含ませて,

$$\varepsilon^2 G(x) = \mu(1-2x) \quad [A15]$$

と取ればよい。

3) $\varepsilon \rightarrow 0$ の極限的性質: 先ず時刻 t と微小時間

Δt を固定して考え、もし

$$\sigma_{[t'/\varepsilon]} = \sigma_{[t/\varepsilon]} \quad (t \leq t' \leq t + \Delta t) \quad [A16]$$

ならば、 $\varepsilon \rightarrow 0$ で

$$\frac{dx}{dt} = F(x(t), \sigma(t)) \quad [A17]$$

となるが、 $\gamma^{(\varepsilon)} = \varepsilon \gamma$ のときは、 $\Delta t \rightarrow 0$ で [A16]

が成立つ確率は 1 に収束する。よって任意の t で [3]

が成立つ確率は 1 に収束する。

一方, $y^{(2)} = y$ のとき, 仮りに, 極限過程が拡散過程に
 収束すると仮定すると, その拡散方程式が [4] であるべき
 ことは,

$$\Delta x \equiv x(t+\Delta t) - x(t) \quad [A18]$$

$$\text{とあり, } \lim_{\Delta t \rightarrow 0} \lim_{\epsilon \rightarrow 0} \{ \langle \Delta x \rangle / \Delta t \} \quad \text{および} \quad \lim_{\Delta t \rightarrow 0} \lim_{\epsilon \rightarrow 0} \frac{\langle (\Delta x)^2 \rangle}{\Delta t}$$

をモデル [1][2] より計算することによって確かめられ
 る。これについては次の文献 4) の Appendix 2
 を参照のこと。

4) Matsuda, H. and Ishii, K. : Stationary
 gene frequency distribution in the
 environment fluctuating between two
 distinct states. J. Math. Biol. to be
 published.

5) Matsuda, H. and Gojobori, T. : Protein
 polymorphism and fluctuation of
 environments. Adv. in Biophys. 12,
 53 (1979).