論文内容の要旨

論文題目

On an approximate formula for the distribution of 2-locus 2-allele model with mutual mutations

(相互突然変異がある場合の 2 座位 2 対立遺伝子モデルの分布に対する近似公式について)

氏名　　三浦　千明

理論集団遺伝学では、あるメンデル集団中に生じた突然変異(遺伝子)が、自然選択圧や浮動、集団の大きさの増減や分集団間での移住、近親婚や同類婚などの条件下で、集団中にどのように広まっていくのかを数理的に研究する事を目的の一つとする。この場合重要な指標の一つは突然変異遺伝子の集団内での頻度(分布)であり、頻度(分布の)変化のモデル化とその解析が研究の中心となる。またヘテロザイゴシティーのような集団遺伝学的統計量の多くがしばしば頻度分布から構成されることからも、その重要性が理解できる。

1 座位のみの場合、定常分布については、その数学的な扱い易さから十分豊富な結果が導かれている(Wright 1931, 1937)。また推移分布に関しても、Kimura(1955)や Crow and Kimura(1970)らの研究から、多くのの場合についてその解析的な解やその性質が求められている。2 座位間の連鎖がある場合については、これまでは主に 2 座位モデルに特異な問題である連鎖不平衡(LD)に焦点が当てられてきた。例えば浮動のみの場合の期待値の時間に沿った LD の推移(Ohta and Kimura 1969a)や定常状態での性質(Hill and Robertson 1968)、また再起突然変異がある場合の定常状態における偏差値(Ohta and Kimura 1969b)など、部分的な情報は得られていた。しかし各座位のアレル頻度や連鎖不平衡それ自体の頻度、またはハプロタイプ頻度について直接その頻度分布を求める試みは、その数学的な困難さからあまり研究されてこなかった。

一方確率論の分野においては、Watanabe(1987)やIkeda and Watanabe(1989)らによって、添え字付けられた伊藤過程によって表される確率微分方程のクラスにおいて、ウィーナー空間上でテイラー展開可能な条件が研究されてきた。更にYoshida(1992)によって数理統計への応用が研究され、確率過程の分布の具体的な漸近展開が得られるようになった。Takahashi and Kunitomo(2003)はこの理論を用いて、経済学に於けるデリバティブのプライシングの問題を考察し、漸近的な解とその有効性について研究した。これらの研究から得られた理論は小分散理論とよばれる。本論文では今まで導かれていなかった、それぞれの座位で相互に突然変異のある2座位間のモデルについて、小分散理論を応用することで、頻度分布の推移的状態について明示的な近似表現を導いた。

【モデル】

2つの座位A、BにそれぞれA$_1$とA$_2$、B$_1$とB$_2$のアレルがあり、再起突然変異率と各アレル頻度を以下のように決める。

$$\underset{p}{A} \underset{v_1}{\overset{u_1}{\rightleftharpoons}} \underset{1-p}{a} \qquad \underset{q}{B} \underset{v_2}{\overset{u_2}{\rightleftharpoons}} \underset{1-q}{b}$$

ハプロタイプA$_1$B$_1$、A$_1$B$_2$、A$_2$B$_1$、A$_2$B$_2$の頻度をそれぞれx_1、x_2、x_3、x_4とする時、連鎖不平衡の度合いを$D \equiv x_1x_4 - x_2x_3$で表す。この時Dも確率過程となる。また組み換え価をcとする。モデルは$p_t = p$、$q_t = q$、$D_t = D$を元とする3次元の確率微分方程式で表現する事が可能である。即ち

$$x_t = x_0 + \int_0^t \mu(x_s)ds + \int_0^t \sigma(x_s)dB_s$$

ここで

$$x_t = \begin{pmatrix} p_t \\ q_t \\ D_t \end{pmatrix} \qquad x_0 = \begin{pmatrix} p_0 \\ q_0 \\ D_0 \end{pmatrix} \qquad \mu(x_t) = \begin{pmatrix} v_1 - (u_1 + v_1)p_t \\ v_2 - (u_2 + v_2)q_t \\ -(1 + u_1 + v_1 + u_2 + v_2 + c)D_t \end{pmatrix}$$

$$\sigma\sigma^t = \begin{pmatrix} p_t(1-p_t) & D_t & D_t(1-2p_t) \\ D_t & q_t(1-q_t) & D_t(1-2q_t) \\ D_t(1-2p_t) & D_t(1-2q_t) & p_tq_t(1-p_t)(1-q_t) + D_t(1-2p_t)(1-2q_t) - D^2{}_t \end{pmatrix}$$

と書ける。p_0、q_0、D_0はt=0での初期値で、B_tは3次元のブラウン運動。

【漸近展開と近似公式】

上で述べた通り、形式的に$\varepsilon(0 \le \varepsilon \le 1)$によって添え字づけられた、頻度と連鎖不平衡の確率過程$x_t{}^{(\varepsilon)} = x_0{}^{(\varepsilon)} + \int_0^t \mu(x_s{}^{(\varepsilon)})ds + \varepsilon\int_0^t \sigma(x_s{}^{(\varepsilon)})dB_s$を、$\varepsilon$に沿って展開する事が保障される。つまり

$$x_t^{(\varepsilon)} = x_t^{(0)} + \varepsilon g_{1t} + \varepsilon^2 g_{2t} + \varepsilon^3 g_{3t} + \cdots \quad (0 < t < {}^{\exists}\tau)$$

となる。ここで各 g_{it} は非確率的な伊藤積分であり、従って平均 0 のブラウン運動に沿う。今 $y_t^{(\varepsilon)} \equiv (x_t^{(\varepsilon)} - x_t^{(0)}) / \varepsilon$ と置くと、$\lim_{\varepsilon \to \infty} y_t^{(\varepsilon)} = g_{1t}$ 。つまり $y_t^{(\varepsilon)}$ は各時点 t 、$\varepsilon \to 0$ で、平均 0、共分散行列 $Cov(g_{1t})$ の 3 次元正規分布に従う。よって、$x_t^{(\varepsilon)}$ は各時点 t 、$\varepsilon \to 0$ で、平均 $x_t^{(0)}$ 、共分散行列 $Cov(g_{1t})$ の 3 次元正規分布に従う。

以上から近似公式 $x_t^{(\varepsilon)} \sim Norm(x_t^{(0)}; Cov(g_{1t}))$ が導かれた。

【シミュレーションと例】

近似公式は x_t の真の分布を $\varepsilon \to 0$ という極限で、単峰型の分布である正規分布で近似したという事である。勿論 ε の与え方から、$x_t = x_t^{(1)}$ なので、この近似が非常に不正確になる条件がある。図 1 は時間があまり経過していない時、真の分布のシミュレーションと近似公式から得た分布を重ね合わせたものである。特に分布の裾が境界に初めて到達する前までの頻度分布は単峰型を保つので、近似公式はうまく機能する。しかし時間が十分に経った場合、もし突然変異率が低いならば遺伝的浮動の効果の方が比較的強く働き、定常状態に近づくにしたがって頻度分布は境界に分布が偏り、結果として単峰型ではなくなり、近似公式では上手く表現できない(図 2)。またこの場合 LD は単峰型にもかかわらず、公式が頻度分布の境界近くでの分布を表せないために連鎖平衡状態になる確率を小さく見積もってしまい、結果的に真の LD の分布を表現する事ができない。他方で、突然変異率が十分高い $(u_1, v_1, u_2, v_2 > \frac{1}{2})$ なら、頻度分布は境界から脱出する事無く推移時間を通して単峰型を保つので、公式を用いて上手く近似できる(図 3)。

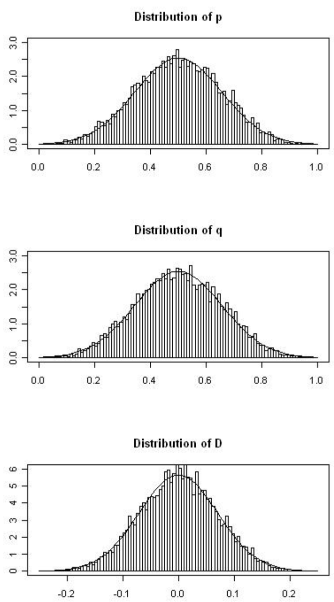

図 1

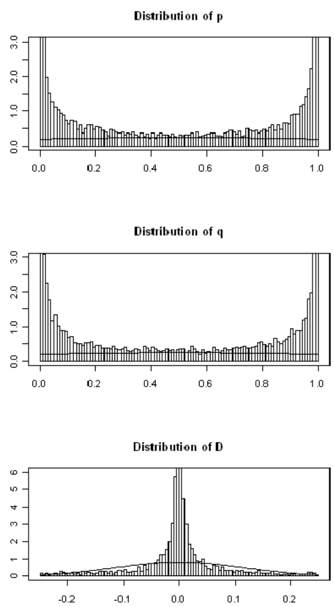

図 2

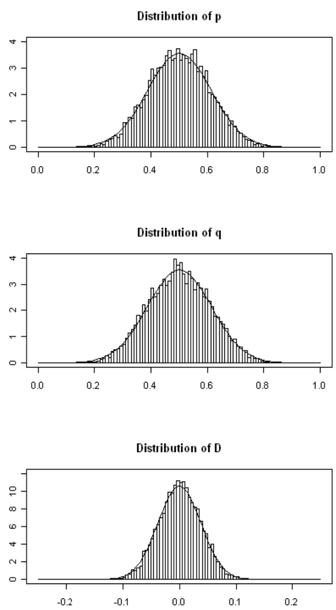

図 3

【定常状態と standard linkage deviation (SLD)】

小分散理論では任意の有限時間までにおける確率過程のテイラー展開を保証しているが、時間が無限大の極限においては理論的に展開が保障できない。しかしながら形式的には近似公式をもちいて時間が無限大の定常状態について考える事が可能である。そこでその形式的な近似公式から、Ohta and Kimura(1969b)で定義され目下のモデルについて正確な値が導かれた standard linkagdeviation(SLD)を構成し、LD に関してどの位の正確さを持つか考察する。まず自乗 SLD は以下で定義される。

$$\sigma^2 \equiv \frac{E[D^2]}{E[pq(1-p)(1-q)]}$$

この定義から、Ohta and Kimura(1969b)で求められた正確な値をσ^2_{exact}と表し、形式的な近似公式から求めた値をσ^2_{approx}と表す。図 4 は突然変異率を$\mu = u_1 = v_1 = u_2 = v_2$と置いて増加させた時に、近似から正確な値を引いた関数がどのように変化するかを、組み換え率を色々と変えて見たものである。破線はc=0.01、点線はc=1、実線はc=100 の場合を表す。図から分かるように、近似から求めた SLD は常に正確な物より値が大きくなるが、組み換え率が高いほどその差小さくなる。また組み換え率がどのような大きさであっても、突然変異率が高くなれば、定常状態でも近似は非常に正確になる事がわかる。

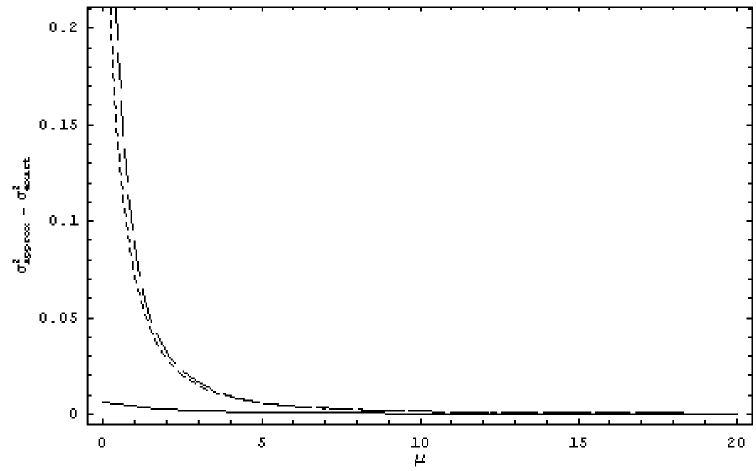

図 4

【議論】

小分散理論は非常に広範な応用を持つ理論である。本論文では 2 座位 2 対立遺伝子モデルについて考察し、近似解を求める事に成功した。この結果から(幾らか計算量が増えるものの)すぐに 2 座位多対立遺伝子モデルの近似解を求める事も可能である。また適切なパラメータセッティングに於いて、ancestral recombination graph にも応用が可能である。しかしながら真の分布が正規分布から遠ざかる時は、近似公式は境界をうまく表現する事ができない。この様な場合はまた、他の方法を用いて対処する必要がある。