学位の種類	博士（生体情報）
報告番号	甲第１５４６号
学位記番号	第１５号
氏名	小酒井 亮太
授与年月日	平成 28 年 3 月 25 日
学位論文の題名	地理的構造を持つ遺伝子系図モデルの研究
論文審査担当者	主査：　能登原 盛弘 副査：　鎌田 直子，渡邊 裕司，田嶋 文生

名古屋市立大学　博士学位論文

地理的構造を持つ遺伝子系図モデルの研究

2016 年

氏名　　小酒井亮太

名古屋市立大学システム自然科学研究科

目次

要旨

1 つの生物種集団の遺伝的組成の進化および遺伝的多様性の保持機構を解明することは集団遺伝学の重要なテーマである.1980 年代初頭より, 集団からサンプルした遺伝子の系図を考察する合祖理論 (Coalescent theory) が Kingman(1982) 及び Tajima(1983) によって導入された. 合祖理論はサンプルした複数の遺伝子の親遺伝子さらにその祖先遺伝子と遡るにつれ, 祖先の共有が生じてくるが, この現象を合祖 (Coalesce) と呼ぶ. さらに時間を遡ることにより最終的に共通な一つの祖先遺伝子に到達し, サンプル遺伝子の系図が完成する. さらにその系図上に生じる突然変異により, サンプルした遺伝子間の遺伝的な違い (DNA 塩基配列の差異) が生じる. 合祖理論は集団からサンプルした遺伝子の DNA 配列データと直結した解析が可能なモデルであり, モデルの導入以来, 理論, データ解析の両面から広範な研究がなされてきた. しかし, 生物集団は広い生息域を持ち, 地理的構造は生物集団の進化, 遺伝的多様性に大きな影響を与える. 生物集団の地理的構造の効果を合祖理論に取り入れることは自然な拡張である. 地理的な構造を考慮に入れた合祖モデルは Takahata(1988), Notohara(1990), Herbots(1994,1997) によって導入され Structured Coalescent Model(SCM) と呼ばれている. これは Kimura(1953) によって導入された飛び石モデル (Stepping stone model) を一般化したモデルと考えることもできる. すなわち, 生物集団が幾つもの小集団に分かれ毎世代, 各小集団内でランダムな交配が起こり, 同時に分集団間に個体の移住が生じるモデルである. このモデルで全集団からランダムにサンプルされた遺伝子の祖先遺伝子を遡ることにより遺伝子の系図が得られるが, 離散時間のモデルから分集団のサイズを一様に大きくする極限操作により, 数学的に扱いやすい SCM と呼ばれる連続時間マルコフ連鎖が得られる. 一般的な形は Notohara(1990) によって導入されたが, 数学的に厳密な証明は保存的移住及びライト・フィッシャータイプの繁殖という限定的な条件下で Herbots(1994,1997) によって与えられた. 本論文では自然集団に近い一般的な移住率及び可換モデルと呼ばれる一般的な繁殖様式において SCM が導かれることを示す. これにより, SCM がモデルの設定の細部によらない, 頑健なモデルであることが示されたことになる.

第 1 章（序論）では集団遺伝学, 分子進化の中立説, 遺伝子系図と遺伝的多様性など基本的事項について解説する.

第 2 章では Kingman(1982a,b,c) の合祖理論 (Coalescent theory) の基本事項およびサンプル遺伝子中に含まれる対立遺伝子の分布に関する Ewens のサンプリング公式について解説する.

第 3 章では本研究の主要結果, 即ち, 地理的構造を考慮に入れた一般的な離散時間モデルから極限操作により連続時間の SCM が導かれることを示す.

第 4 章では分集団間の遺伝的分化レベルの指標である固定指数 F について合祖理論の視点から研究した Herbots(1998) の結果, および有限個の分集団モデルではあるが一般的な移住率の下で SCM を導いた Sampson(2006) の結果など SCM について幾つかの結果を紹介する.

1　序論

1.1　集団遺伝学の成り立ち

19 世紀中頃,Darwin（1859）が名著「種の起源」を著して, 種が共通の祖先から進化する要因として自然選択（Natural Selection）の概念を導入した. Darwin は生物進化を‘変更を伴う遺伝（Decent With Modifications)’と呼んだ. この自然選択説に対する反応は大きく,Weismann が 1892 年に自然選択万能説を取り上げ, Darwin を非常に高く評価した. その姿勢はメンデルの法則と Darwin の法則を同時に指示する姿勢を見せた Mayr が彼を 19 世紀で二番目に重要な進化理論家であると提唱するほどであった. また彼自身も Darwin の自然選択説を実験的に検証しようとした最初の一人であった. このように Darwin 説が大きく騒がれる中, 20 世紀にはいると, 自然選択を除いた系統の変化が複数垣間見られた. de Vries が 1901 年にオオマツヨイグサに形態的に大きく異なる変異が頻発することを観察し, この現象に対して突然変異 (Mutation) という名称を与えるとともに突然変異説を唱えた. このことから, 1930 年までに遺伝の実態は染色体にあり, 染色体には遺伝の単位とみなすべきだと思われる多くの遺伝子があること, 遺伝子における‘変更’は染色体上に起こる突然変異に由来するものであることが次第に明らかになった. こうした遺伝学, 進化論の発展に基づいて生物の進化を数理科学的手法を用いて研究する試みが Haldane(進化の原因), Fisher(自然淘汰の遺伝学的理論), Wright(メンデル集団における進化) らによって行われた. それが集団遺伝学 (Population Genetics) の所以である. これらは現在の進化論の支配的な基盤となっている. 以後, 他にも様々な遺伝における‘変更’が明らかになった. 集団遺伝学においては, 環境に対して適応性を得た遺伝子はどのように集団中に拡がるか, 逆に適応性を得ることができなかった遺伝子はどのように集団から消えてゆくであろうか, 有性生殖と無性生殖の違いは何であるかなど生物進化に関する事例が多数挙げられ, 育種学, 優生学などの応用が生まれた. その他, 分子生物学の進化に伴って 1953 年に DNA の二重らせん構造が発見され, それに続いて 1972 年に米国のスタンフォード大学のポールバーグ教授が SV40DNA に大腸菌を導入するという異種の DNA 結合に成功した. この発見を基に遺伝子組み換え (Recombination) という言葉が登場することになる. このような発展に応じて, 分子の進化と生物の進化はどう関連しているのかを調べる研究が盛んになり, 分子集団遺伝学 (Molecular Population Genetics) と呼ばれる分野が拓かれてきた. その中で木村資生らは, 分子進化の定量的な観測値を得るためには自然選択とは無関係な中立突然変異が重要な役割をしているとする「分子進化の中立説」を唱えた. 理論集団遺伝学は 1950 年代半ばより国立遺伝学研究所の木村資生, 太田朋子, 丸山毅夫により日本が世界をリードする形で精力的な研究がなされ, この中で中立説が 1968 年に提唱された. この時期の研究は遺伝子頻度を対象とした拡散過程モデルが多く用いられ, 遺伝子頻度分布, 固定確率など多彩な研究がなされている.1980 年代に入ると Kingman(1982) 及び Tajima(1983) によって独立にサンプルされた中立な遺伝子系図を表現する,Coalescent モデル（合祖モデル）と呼ばれている確率モデルが導入された. その後 Coalescent モデルは遺伝子の組み換え, 自然選択, 集団の地理的構造など様々な要因を考慮にいれたモデルへ

と発展した. 遺伝子系図のモデルはサンプル遺伝子のデータ解析と直結し, 遺伝的多様性や集団の歴史を推定するための理論的枠組みを提供するモデルとして, 精力的に研究が続けられている.

1.2 分子進化の中立説

集団遺伝学 (Population Genetics) は生物集団の遺伝的組成の進化及び遺伝的多様性の保持機構を解明し, 生物集団の進化のメカニズムを解明することを重要なテーマとしている. 1920 年代頃より Fisher, Wright, Haldane などにより, 集団遺伝学の数学理論の基礎が確立されてきたが, ダーウインの自然選択説とメンデルの遺伝学の結合によって形成された総合進化説が主流であった. 1960 年代に電気泳動法により, 酵素多型（蛋白質多型）が高頻度で各種の生物集団内に存在することが次第に明らかになると, 分子レベルでは有利な突然変異によって進化が起こるとする自然選択説では説明できないと考えられるようになった. このような中で Kimura(1968) は「分子進化の中立説」を提唱した. 分子進化の中立説とは, 分子進化においては自然選択に対して有利でも不利でもない中立な突然変異と遺伝的浮動という生物集団の個体数有限性によって起こる偶然的な遺伝子頻度の変動によって分子進化は起こるとする説である. 突然変異は蛋白質のアミノ酸配列を変える非同義な変異とアミノ酸を変えない同義な変異があるが, 蛋白質の機能に影響を与えない同義な変異あるいは非同義でも機能にあまり影響を与えないものは中立な突然変異と考えることができる. 機能に重大な影響を与える突然変異は有害であり, 集団から次第に消えてゆく運命にある. 中立説では突然変異の多くは自然選択に対して中立または有害なものがほとんどであり, 分子進化に貢献する有利な突然変異は非常に少数と考える. 分子進化には大きく二つの特徴がある.「各タンパク質について年当たりの分子進化の一定性」と「変化様式の保守性」である. 木村資生の中立説によると「機能的に重要でない分子（または分子内の重要でない部分）ほど, そうでないものより進化の過程でアミノ酸や DNA 塩基の置換が急速に起こり, 置換率（進化速度）の最高は突然変異率できまる」とされる. 完全に中立であれば「分子進化速度＝突然変異率」が成り立ち, このような所見は中立説の理論的研究により示されており, 分子進化の知見を矛盾無く説明できる理論である.

1.3　遺伝子系図と祖先過程

遺伝子配列のデータから読み取れるものには統計量として多型的な（分離した）サイトの数（the number of segregating sites）がある．一般的に遺伝子配列はアデニン（A), シトシン（C), チミン（T), グアニン（G) の 4 つの塩基の配列によって表現される．

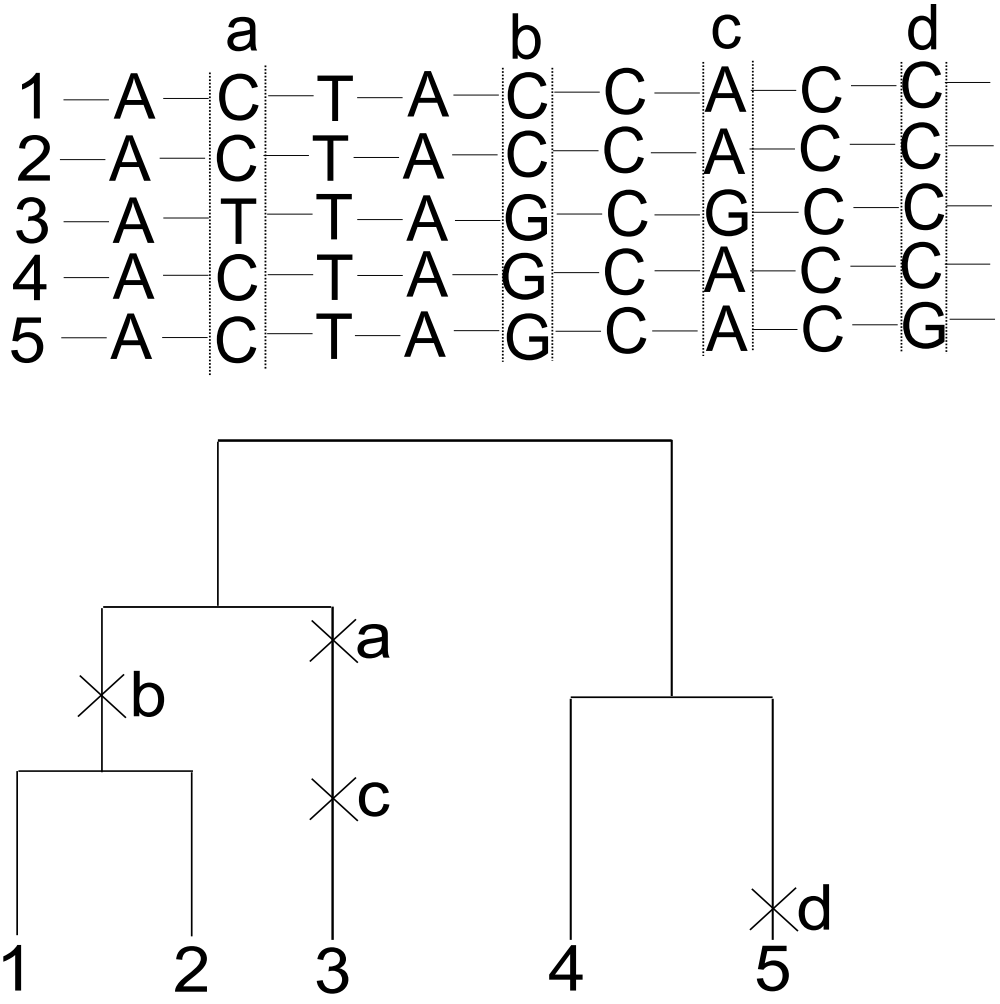

図 1　遺伝子系図と多型サイト (無限サイトモデルでの図. a,b,c,d のサイトは多型的なサイトであるが, これは系図上の示された位置で生じた突然変異に起因している.)

このような塩基置換を繰り返しながら, 時には表現型にも影響を及ぼし, 生物は進化を遂げてきた. Hammer(2001) 等によれば, 人の $Y-$ 染色体に関する次のようなデータの存在が知られている.

地域	サンプルサイズ	多型サイト数	平均塩基相違数
ヨーロッパ	355	16	2.48
北アフリカ	131	13	2.39
南アジア	133	14	2.56
南アフリカ	243	10	1.65

これらは $Y-$ 染色体に遺伝子組み換えがなく, 雄から 1 つ受け継ぐという, 非常に観察しやすいことに基づいて行われたものである.この結果はMichael Hammer等により, 世界各地から集計した 2000 人以上の男性に対して $Y-$ 染色体の検査を行い, その中から 4 つの地域に限定してデータを纏めたものである. $D_{i,j}$ を i 番目, j 番目の塩基配列の塩基相違数とすると, 平均塩基相違数
(塩基多様度)Πは

$$\Pi = \sum_{i<j} \frac{D_{i,j}}{\binom{n}{2}}$$

で表される.但し, n はサンプル数である.このデータからわかることは, 南アフリカの遺伝子の多様性が他の地域に比べて最も低いということである.しかし, 異なる集団から取り出した 2 つのサンプルについては$\Pi = 3.18$ で最も大きい(Hein et al.(2005)). これは, 全人類集団が任意交配集団ではないことを示唆している. すなわち, 任意交配集団ならば同じ集団からサンプルした場合と異なる集団からサンプルした場合の塩基多様度 に大きな差はないはずである. 異なる集団からのサンプルが高い塩基多様度を示すことは人類の各集団相互の地理的な隔たりが遺伝的多様性に影響を及ぼすことを表しており, 集団の地理的構造が遺伝的多様性および進化に与える影響を考察する必要性がある. ヒトに限らず, 生物集団の地理的構造を考慮に入れた遺伝子系図の視点から, 遺伝的多様性の問題を研究することは, 集団遺伝学の重要な研究テーマである.

1.4 研究の背景及び目的と意義

本研究と関連する古くから知られた地理的構造を考慮した集団遺伝学のモデルは Kimura(1953) による飛び石モデル (Stepping stone model) である. その後, このモデルに関する多くの研究があるが, 現在の視点から見るとサンプル数 2 個の場合の Structured coalescent モデル (SCM) と見なすことができる. Takahata(1988) 及び Tajima(1989) の 2 つの分集団モデルの先行研究があるが, 一般的な形での Structured coalescent モデルは Notohara(1990) によって導入された. その後多くの研究があるが, その全般的な解説については Wakeley(2009 第 5 章) を参照されたい. ただし Notohara(1990) においては数学的考察によるモデルの導出をしているが, 確率論的に厳密なものではなかった. Herbots(1994) はその学位論文で, 繁殖はライト・フィッシャーモデル, 移住は保存的移住という限定的な条件の下ではあるが厳密な証明を与えた. 本研究では可換モデルと呼ばれる一般的繁殖モデルと非保存的な場合も含む移住の下で, SCM が導かれることの厳密な証明を与えることができた. 可換モデルは次世代に残す子供の数の分布が親によって変わらないモデルであり, 自然選択が無い中立な遺伝子を対象としている. 一般的な条件下で SCM が導かれるということは SCM の頑健性, すなわち広いクラスの離散モデルから極限として同じ SCM が導かれることを意味し, 広いクラスのモデルに対して, 適切な条件の下で SCM がその近似遺伝子系図モデルとして利用できることを意味している. 自然集団では繁殖がライト・フィッシャーモデルに従い, 移住の前後で各分集団のサイズが不変ということはかなり不自然な仮定である. より広い繁殖様式である可換モデルでかつ非保存的移住という条件でも SCM が導かれるということは, 中立な遺伝子について, 自然集団からのサンプル遺伝子のデータ解析において適切な条件下で SCM が利用できることを保障していると言うこともできる. 第 2 章では集団遺伝学で, その扱いやすさから最も良く使われるライト・フィッシャーモデルとモランモデルについて, さらに Kingman(1982) に従い Coalescent モデル及び Ewens のサンプリング公式と呼ばれるサンプル中の各対立遺伝子の個数分布に関する公式が Coalescent 理論から導かれることを紹介する. 第 3 章が本研究の主要部分であるが, モデルの設定, 有限次元分布の収束及び弱収束の証明を与える. 本研究では無限個の分集団を含み, 上記の一般的条件の下での厳密な証明を与えた最初の研究結果である. 第 4 章では分集団間の分化レベルの指標である固定指数 F 統計量と SCM の関係について解説し, 簡単なモデル (サークル状飛び石モデルと島モデル) について具体的結果を示す. また有限個の分集団で各分集団サイズの確率的変動及び非保存的移住を含む移住モデルにおいて離散時間マルコフ連鎖から極限において SCM が導かれる Sampson(2006) の結果を紹介する. Sampson(2006) のモデルは有限個の分集団からなる有限マルコフ連鎖で表現されるモデルであり, 証明は Möhle(1998) の有限マルコフ連鎖の収束定理を利用したものであるが, 最近得られた収束定理の拡張 (Möhle and Notohara(2016)) により可算無限個の状態を含むモデルに拡張したときの有限次元分布の収束が導かれることを示す. 本研究では可算無限個の分集団を含むモデルを研究対象としており, この点が Sampson のモデルと異なる点である.

2　合祖モデル

合祖理論とは, 集団からサンプルした遺伝子について過去に遡ることによって祖先集団の中から共通祖先を見出す確率モデルの理論のことを言う. 合祖理論に関して有名な話はミトコンドリア・イブである. これはアメリカ在住の様々な人種の 134 人のミトコンドリア DNA の配列からその系図を探り, 下図に示されるように, 約 20 万年前に共通祖先に到達できるという発見であった.

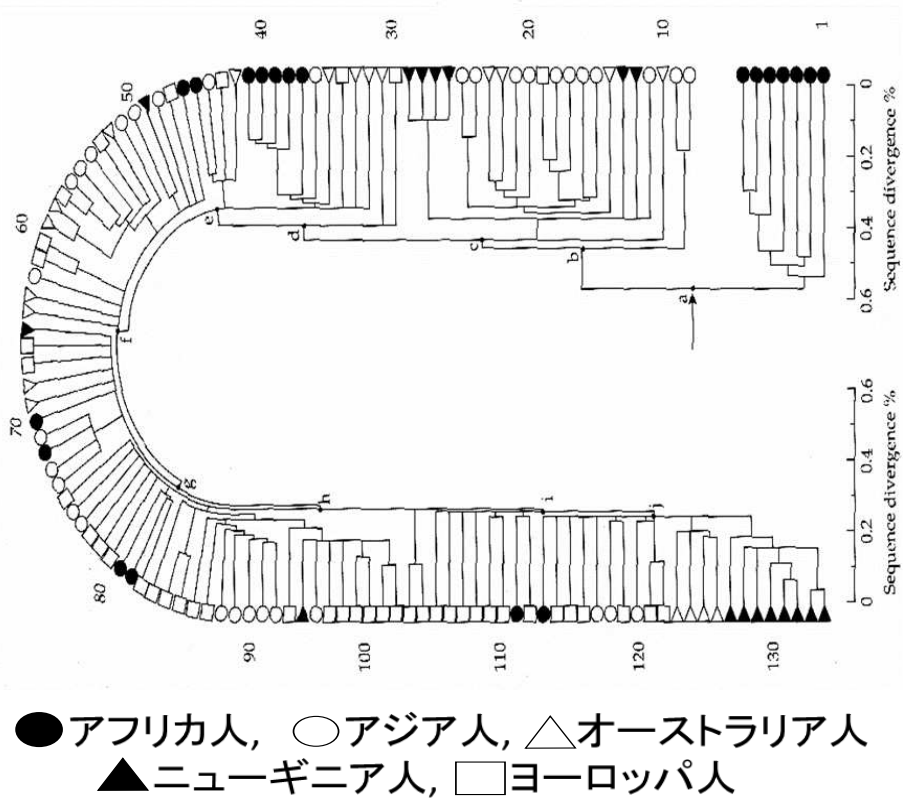

図 2　ミトコンドリアイブ (Cann et al.(1987))(アメリカ在住の様々な人種 134 人のミトコンドリア DNA 遺伝子をサンプルし, その系図を作成することにより, それらの共通祖先のミトコンドリア DNA 遺伝子が 20 万年前にあったことが判明した.)

この図から, サンプル遺伝子の祖先を辿ってゆくと, ただ一つの共通祖先 (MRCA;Most Recent Common Ancester) に到達することがわかる. 集団遺伝学で古くから最もよく用いられているライト・フィッシャーモデルやモランモデルを紹介する.

2.1 ライト・フィッシャーモデルとモランモデル

ライトフィッシャーモデルから紹介する. ν_i を i 番目の親が産む子供の数を表す確率変数としたとき, N 個の親個体が次世代に子供をそれぞれ $k_1, k_2, \cdots, k_N$ 個産む確率が多項分布

$$P[\nu_1 = k_1, \nu_2 = k_2, \cdots, \nu_N = k_N] = \frac{N!}{k_1!k_2!\cdots k_N!}(\frac{1}{N})^N, \quad \text{但し, } \sum_{l=1}^{N} k_l = N$$

となるモデルをライトフィッシャーモデルと言う. また, $\frac{1}{N}$ は N 個の親集団から 1 つの子供に対する, 1 つの親を見つけるときの確率である. これは, 可換モデルの一つで, 可換モデルとは次の条件を満たすものを言う：$(i_1, i_2, \cdots, i_N)$ を $(1, 2, \cdots, N)$ の任意の置換とするとき,

$$P[\nu_1 = k_1, \nu_2 = k_2, \cdots, \nu_N = k_N] = P[\nu_{i_1} = k_1, \nu_{i_2} = k_2, \cdots, \nu_{i_N} = k_N], \quad \text{但し, } \sum_{l=1}^{N} k_l = N$$

これは親によって産む子供の数の分布が変わらないことを意味する. 図 3 はライト・フィッシャーの繁殖のプロセスを表す図である.このライト・フィッシャーモデルに対して, 時間を遡ることを考える.今, 祖先過程 $\{{A_n}^N(t); t \geq 0\}$ を 0 世代でのサンプルが n である条件のもとで t 世代遡ったときの異なる祖先の数と定義すれば, 時刻 t で k 個の祖先遺伝子が 1 世代前に j 個の親を持つ確率は,

$$P({A_n}^N(t+1) = j | {A_n}^N(t) = k)$$

である.この確率を $g_{k,j}$とすれば, これら確率は次のように書ける:

$$g_{k,k} = \frac{N(N-1)\cdots(N-k+1)}{N^k} = 1 - \frac{k(k-1)}{2N} + O(\frac{1}{N^2}) \tag{1}$$

$$g_{k,k-1} = \binom{k}{2} \frac{N(N-1)\cdots(N-k+2)}{N^k} = \frac{k(k-1)}{2N} + O(\frac{1}{N^2}) \tag{2}$$

$j \leq k-2$ の時は

$$g_{k,j} = \frac{N(N-1)\cdots(N-j+1){S_k}^j}{N^k} = O(\frac{1}{N^2}) \tag{3}$$

但し, ${S_k}^j$ は k 個の子供が j 個のどの親に由来するか振り分けるときの組み合わせの数である（第２種スターリング数という）.これら確率 (1), (2), (3) を用いて遷移確率行列 $\mathbb{G} = (g_{k,j})$ を構成する. 実際, 単位行列 $\mathbb{I}$ ($\mathbb{I}_{k,l} = \delta_{k,l}$) と生成行列 $\mathbb{Q} = (\mathbb{Q}_{k,j})$(但し $\mathbb{Q}_{k,k} = -\frac{k(k-1)}{2}$, $\mathbb{Q}_{k,k-1} = \frac{k(k-1)}{2}$, また $j \neq k,\ k-1$ のとき, $\mathbb{Q}_{k,j} = 0$) を用いて $\mathbb{G}$ は以下のように表せる；

$$\mathbb{G} = \mathbb{I} + \frac{\mathbb{Q}}{N} + O(\frac{1}{N^2}) \tag{4}$$

N 世代を単位時間とするタイムスケールをとり, N に関して極限をとると, 生成作用素 $\mathbb{Q}$ に従う連続時間の斉時的マルコフ連鎖に収束する；

$$\lim_{N\to\infty} \mathbb{G}^{[Nt]} = \lim_{N\to\infty} \left(I + \frac{\mathbb{Q}}{N} + O(\frac{1}{N^2})\right)^{[Nt]} = e^{t\mathbb{Q}} \tag{5}$$

また,T_kを祖先の数が k である時の滞在時間とすると, この分布は平均$\dfrac{2}{k(k-1)}$に従う指数分布である；

$$P(T_k > t) = \lim_{N\to\infty} \left(1 - \frac{k(k-1)}{2N}\right)^{[Nt]} = e^{-\frac{k(k-1)}{2}t} \tag{6}$$

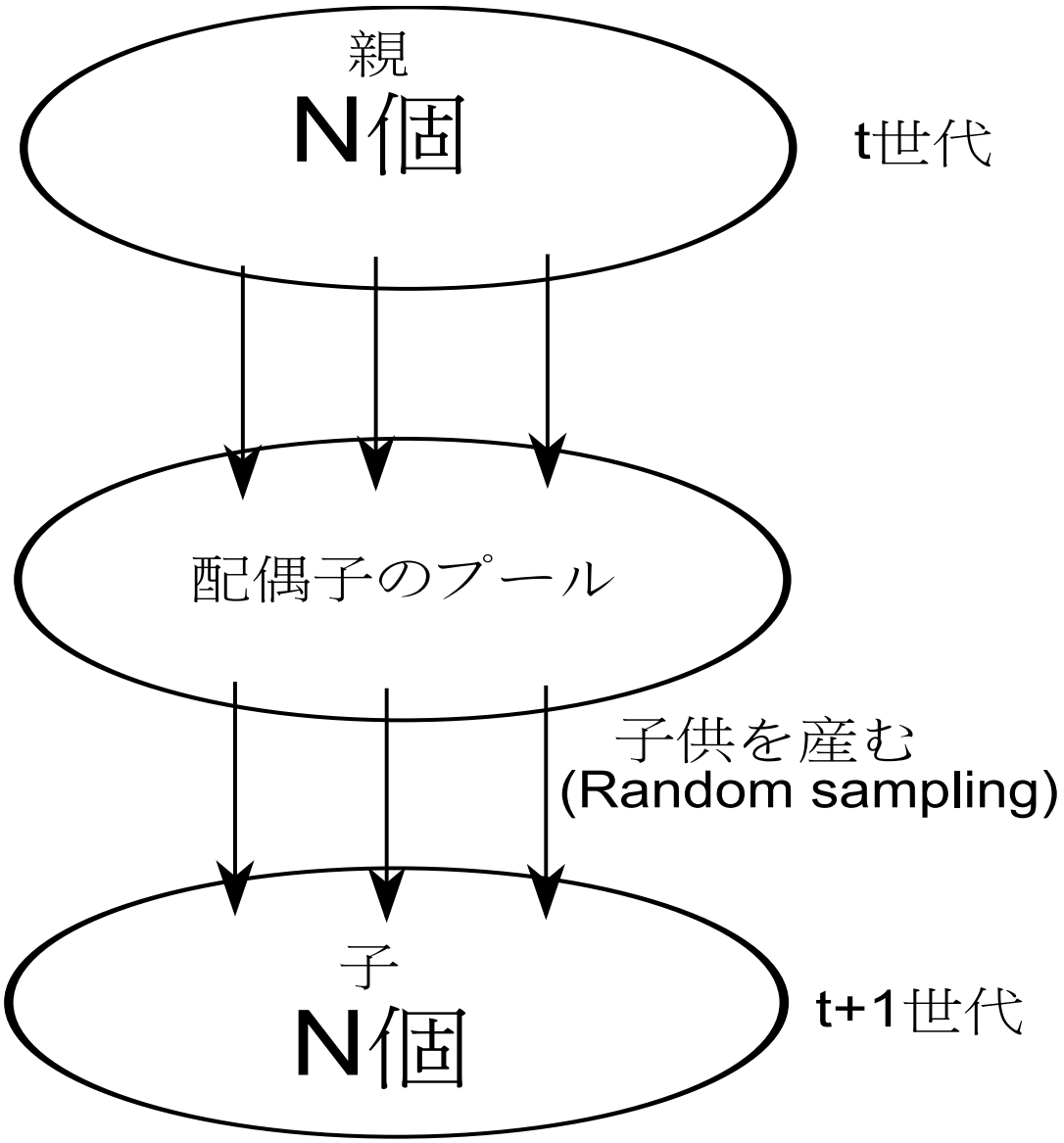

図 3　ライト・フィッシャーモデルの図 (N 個の半数体生物集団が配偶子のプールを作り, Random sampling によって抽出された個体により, 同じ集団サイズの次世代集団ができる.)

次にモランモデルを紹介する. モランモデルは毎世代 1 個体が死亡し, 死滅したところをその子供, もしくは他の親個体の子供により埋め合わせをし元の集団のサイズ (N) に戻る過程を表す. モランモデルに関しては, 以下のようにして連続時間のマルコフ連鎖の収束が導かれる：
モランモデルでは次のような遷移確率行列 $\hat{\mathbb{G}} = (\hat{g}_{k,j})$ に従う.

$$\hat{g}_{k,k} = 1 - \frac{k(k-1)}{N^2}$$

$$\hat{g}_{k,k-1} = \frac{k(k-1)}{N^2}$$

$$\hat{g}_{k,j} = 0, \quad j \leq k-2$$

$$\hat{\mathbb{G}} = \mathbb{I} + \frac{\hat{\mathbb{Q}}}{N^2}$$

但し, $\hat{\mathbb{Q}} = (\hat{\mathbb{Q}}_{k,j})$ で$\hat{\mathbb{Q}}_{k,k} = -k(k-1)$, $\hat{\mathbb{Q}}_{k,k-1} = k(k-1)$, $j \neq k$, $k-1$ のとき,$\hat{\mathbb{Q}}_{k,j} = 0$.
今,$\frac{N^2}{2}$世代を単位時間とするタイムスケールをとり,$N \to \infty$ の極限をとると, ライトフィッシャーモデルの場合と同じ生成作用素 $\mathbb{Q}$ に従う連続時間のマルコフ連鎖に収束する；

$$\lim_{N\to\infty} \hat{\mathbb{G}}^{[\frac{N^2}{2}t]} = \lim_{N\to\infty} \left(\mathbb{I} + \frac{\hat{\mathbb{Q}}}{N^2}\right)^{[\frac{N^2}{2}t]} = e^{t\hat{\mathbb{Q}}} \tag{7}$$

また,前と同様に,T_kを祖先の数が k である時の滞在時間とすると, この分布は平均$\dfrac{2}{k(k-1)}$に従う指数分布である；

$$P(T_k > t) = \lim_{N\to\infty} \left(1 - \frac{k(k-1)}{N^2}\right)^{[\frac{N^2}{2}t]} = e^{-\frac{k(k-1)}{2}t} \tag{8}$$

ライト・フィッシャーモデルもモランモデルもどちらも同じ滞在時間の分布に収束する. 図 4 はモランモデルの図である. ここでこれらのモデルにおける具体的な系図を見てみよう. 図 5 はサンプル数が $n = 6$ の場合の系図である. MRCA(Most Recent Common Ancester) へ到達するまでの時間の長さは,

$$T_{MRCA} = \sum_{i=2}^{n} T_i \quad (\text{但し、図 5 では } n = 6)$$

全体の枝の長さは,

$$T_{total} = \sum_{i=2}^{n} iT_i \quad (\text{但し、図 5 では } n = 6)$$

よって, それぞれの平均時間は (8) から,

$$E[T_{MRCA}] = E\Big[\sum_{i=2}^{n} T_i\Big] = \sum_{i=2}^{n} \frac{2}{i(i-1)} = 2\Big(1 - \frac{1}{n}\Big)$$

$$E[T_{total}] = E\Big[\sum_{i=2}^{n} iT_i\Big] = \sum_{i=2}^{n} \frac{2}{i-1} \cong 2log(n)$$

で表される. 今, 系図を書くと次のように書ける（図 5:サンプル数が 6 の場合）；

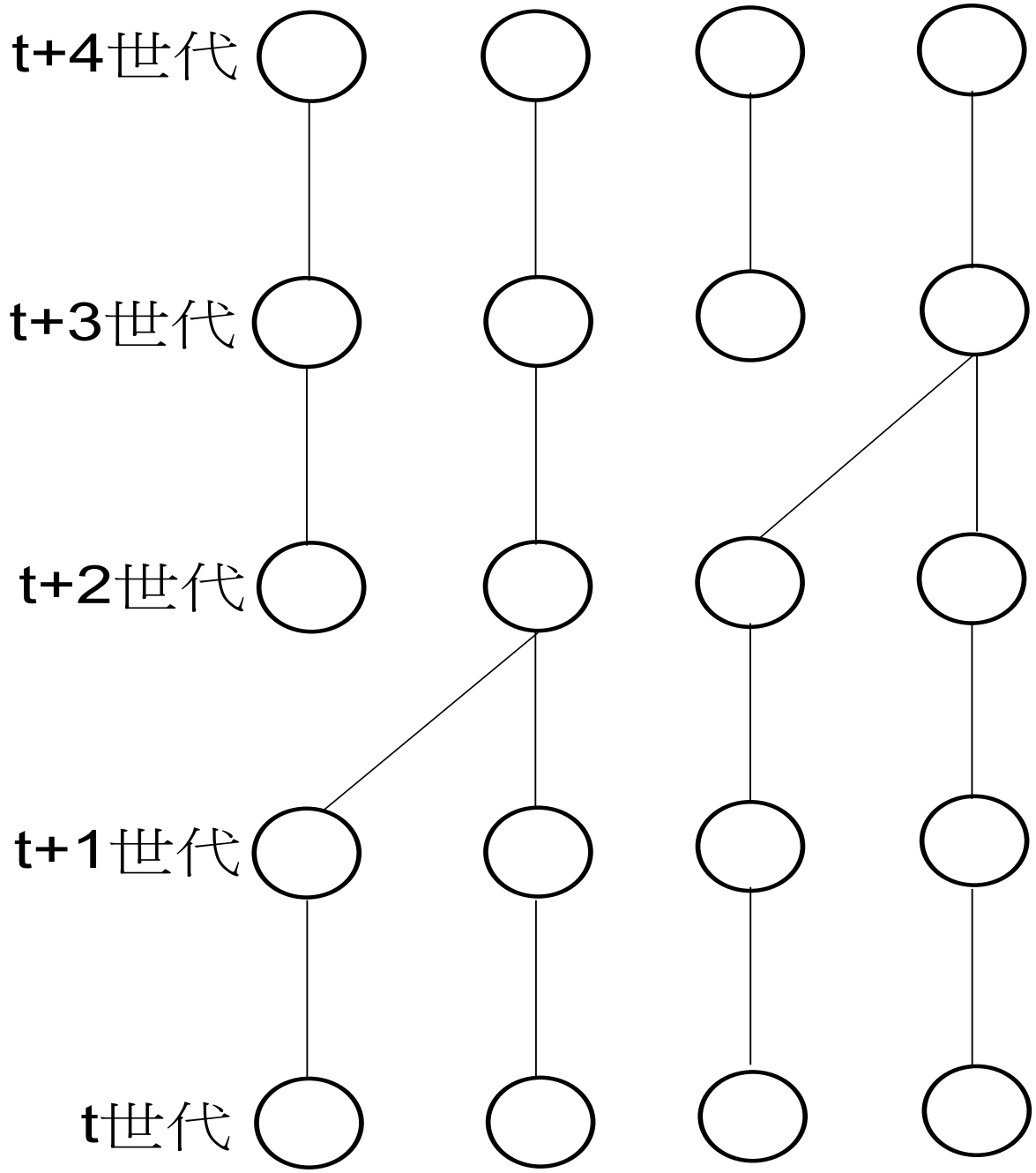

図 4　モランモデルの図 (N=4 の場合. 世代ごとに死んだ親個体を子供が埋め合わせをし, 集団サイズを同じくする.)

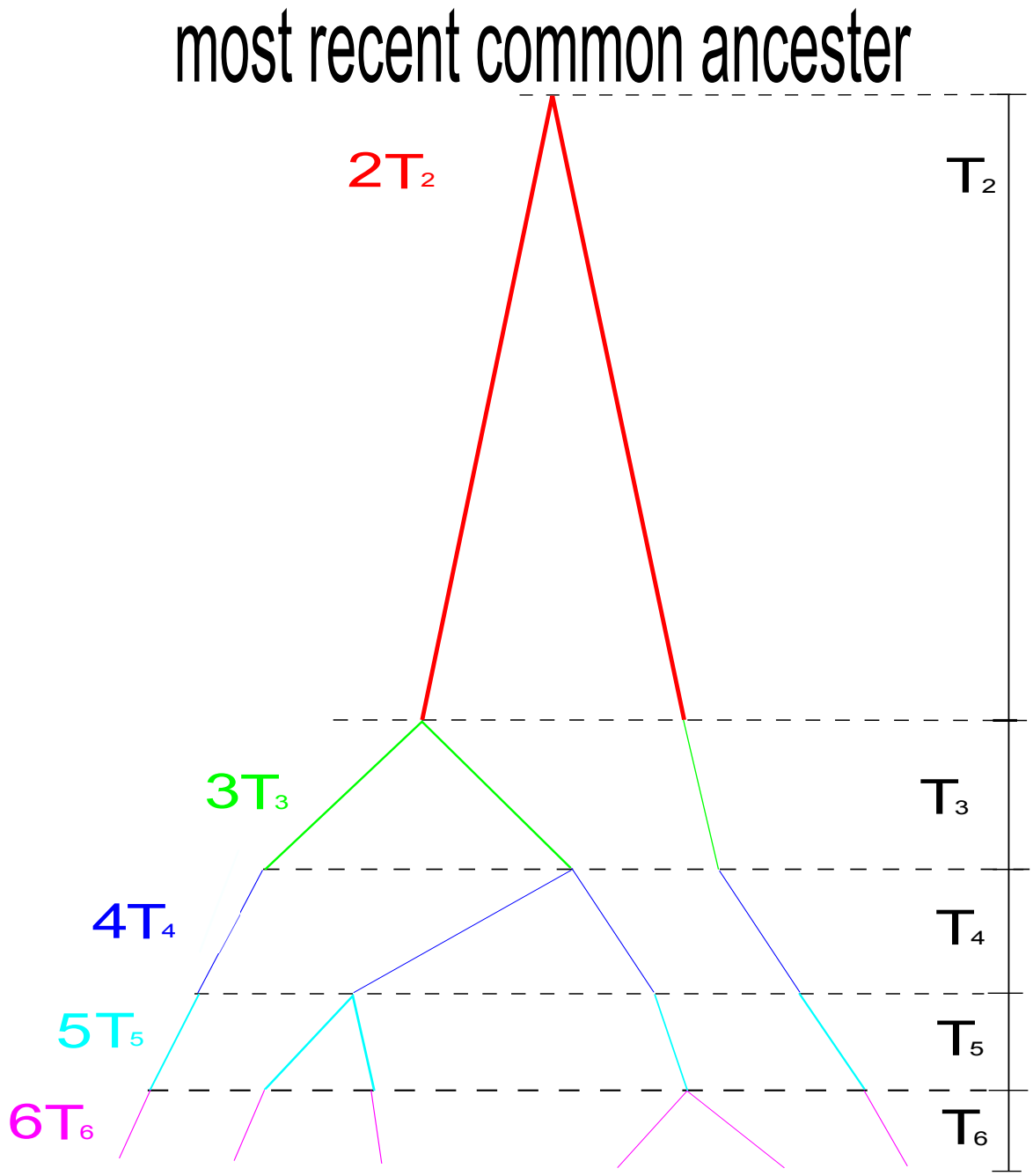

図 5　滞在時間を基にした遺伝子系統図 (サンプル数が 6 の場合. 滞在時間の分布が指数分布に従うため, サンプル数が減少するにつれ, 滞在時間が長くなる.)

次は 2-アレルタイプでのモランモデルとライトフィッシャーモデルを考える. 今, a, A の 2 つのアレルタイプが存在するとしよう. ライトフィッシャーモデルでは,アレルタイプ A の r 世代前の遺伝子の数を X_rとおくことにより,$p_{i,j}$を 1 世代で A タイプの遺伝子の数が i 個から j 個へ遷移する確率とすると,

$$p_{i,j} = P(X_{r+1} = j | X_r = i) = \frac{N!}{j!(N-j)!}\Big(\frac{i}{N}\Big)^j\Big(\frac{N-i}{N}\Big)^{N-j}, \quad 0 \leq i, j \leq N$$

平均を計算すれば, 二項分布の性質より,

$$E[X_{r+1}|X_r] = X_r \quad (\text{マルチンゲール})$$

更に平均をとれば,

$$E[X_{r+1}] = E[X_r]$$

であるから、この式から,

$$E[X_{r+1}] = E[X_0]$$

がわかる. これは, アレルタイプ A の遺伝子の平均個数が世代によって変わらないことを表す.次に分散を計算すれば,$V[X_r] = E[V[X_r|X_{r-1}]] + V[E[X_r|X_{r-1}]],\ V[X_r|X_{r-1}] = N * \frac{X_r}{N}\Big(1 - \frac{X_r}{N}\Big)$ であることを用いて,

$$V[X_r] = E[V[X_r|X_{r-1}]] + V[E[X_r|X_{r-1}]] = E\Big[N\frac{X_{r-1}}{N}\Big(1 - \frac{X_{r-1}}{N}\Big)\Big] + V[X_{r-1}]$$

$$= E[X_{r-1}] - \frac{E[(X_{r-1})^2]}{N} + V[X_{r-1}] = E[X_0] - \frac{V[X_{r-1}] - (E[X_{r-1}])^2}{N} + V[X_{r-1}]$$

よって,

$$= \Big(1 - \frac{1}{N}\Big)V[X_{r-1}] + \frac{E[X_0](N - E[X_0])}{N}$$

この漸化式を r について解けば, 分散について次の式が得られる.

$$V[X_r] = E[X_0](N - E[X_0])\Big(1 - \Big(1 - \frac{1}{N}\Big)^r\Big) + V[X_0]\Big(1 - \frac{1}{N}\Big)^r = E[X_0](N - E[X_0])\Big(1 - \Big(1 - \frac{1}{N}\Big)^r\Big)$$

今求めた分散の値を用いて, ヘテロ接合度と呼ばれる以下の式 ;

$$H(r) = E\Big[2\frac{X_r}{N}\Big(1 - \frac{X_r}{N}\Big)\Big] = 2\frac{E[X_r(N - X_r)]}{N^2}$$

の値を求めることで, このモデルの性質をより詳しく見てみよう. これは r 世代前において集団から 2 つの遺伝子を取り出した時, それらが異なるアレルタイプである確率を表す. X_rの平均, 分散より,

$$H(r) = E\Big[2\frac{X_r}{N}\Big(1 - \frac{X_r}{N}\Big)\Big] = \frac{2(NE[X_r] - E[(X_r)^2])}{N^2} = \frac{2(NE[X_0] - (V[X_r] + (E[X_r])^2))}{N^2}$$

$$= 2E\Big[\frac{X_0}{N} - \frac{(X_0)^2}{N^2}\Big]\Big(1 - \frac{1}{N}\Big)^r = H(0)\Big(1 - \frac{1}{N}\Big)^r$$

これより, $\lim_{r\to\infty} H(r) = 0$ となる.これは, 最終的に遺伝子の固定が起きることを表している.

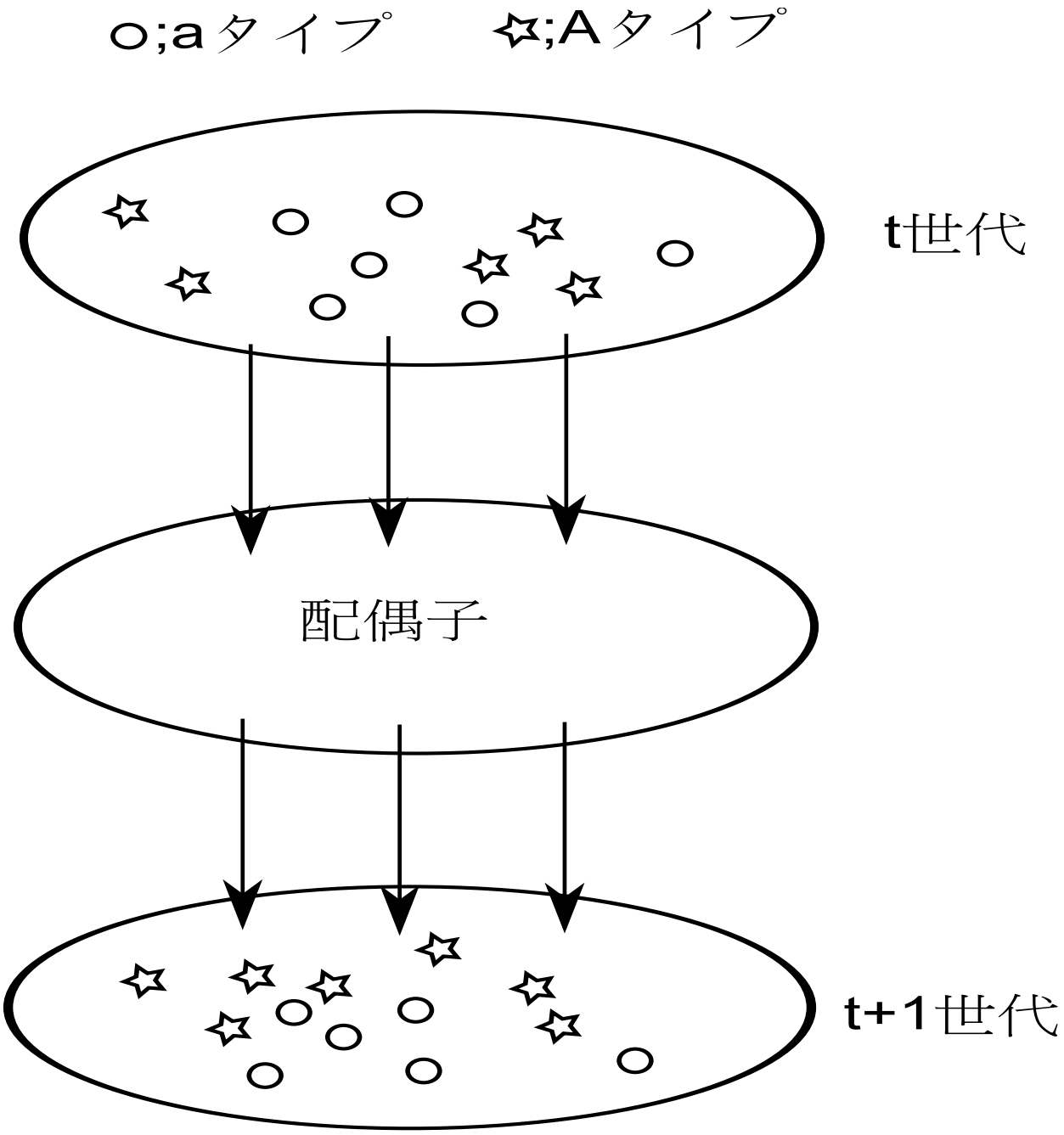

図 6　2 アレルタイプの場合のライト・フィッシャーモデルの図 (集団サイズが N である半数体生物集団を考える. 任意交配により, 2 つのアレルタイプが混ざり合っている.)

次にモランモデルの場合を考える. 同じくして X_rをアレルタイプ A の r 世代前の遺伝子の数とすると, 1 世代での遺伝子頻度における遷移確率は,

$$p_{i,j} = P(X_{r+1} = j | X_r = i) = \begin{cases} \dfrac{i(N-i)}{N^2} & (j = i+1 \ \text{又は} \ j = i-1) \\ \dfrac{i^2}{N^2} + \dfrac{(N-i)^2}{N^2} & (j = i \ \text{のとき}) \\ 0 & (\text{その他}) \end{cases} \tag{9}$$

この遷移確率を用いて平均を求めれば,

$$\begin{aligned} E[X_{r+1}|X_r] &= (X_r+1)P(X_{r+1} = X_r+1|X_r)+(X_r-1)P(X_{r+1} = X_r-1|X_r)+X_rP(X_{r+1} = X_r|X_r) \\ &= (X_r+1)\frac{X_r(N-X_r)}{N^2} + (X_r-1)\frac{X_r(N-X_r)}{N^2} + X_r\Big(\frac{(X_r)^2}{N^2} + \frac{(N-X_r)^2}{N^2}\Big) \\ &= 2X_r\frac{X_r(N-X_r)}{N^2} + X_r\Big(\frac{(X_r)^2}{N^2} + \frac{(N-X_r)^2}{N^2}\Big) = X_r \ \ (\text{マルチンゲール}) \end{aligned}$$

更に平均をとれば,

$$E[X_{r+1}] = E[X_r]$$

であるから, この式から,

$$E[X_{r+1}] = E[X_0]$$

であることがわかる. モランモデルに対してもライト・フィッシャーモデルと同様にアレルタイプ A の遺伝子の平均個数が世代によって変わらないことを表す. 分散は,

$$V[X_{r+1}|X_r] = E[(X_{r+1})^2|X_r] - E^2[X_{r+1}|X_r] = E[(X_{r+1})^2|X_r] - (X_r)^2$$

条件付平均の 2 乗モーメント $E[(X_{r+1})^2|X_r]$ を求めると,

$$\begin{aligned} E[(X_{r+1})^2|X_r] &= (X_r+1)^2P(X_{r+1} = X_r+1|X_r)+(X_r-1)^2P(X_{r+1} = X_r-1|X_r)+(X_r)^2P(X_{r+1} = X_r|X_r) \\ &= (X_r+1)^2\frac{X_r(N-X_r)}{N^2}+(X_r-1)^2\frac{X_r(N-X_r)}{N^2}+(X_r)^2\Big(\frac{(X_r)^2}{N^2}+\frac{(N-X_r)^2}{N^2}\Big) \\ &= 2((X_r)^2+1)\frac{X_r(N-X_r)}{N^2}+(X_r)^2\Big(\frac{(X_r)^2}{N^2}+\frac{(N-X_r)^2}{N^2}\Big) = (X_r)^2+2\frac{X_r}{N}-2\frac{(X_r)^2}{N^2} \end{aligned}$$

よって,

$$V[X_{r+1}|X_r] = E[(X_{r+1})^2|X_r] - (X_r)^2 = 2\frac{X_r}{N} - 2\frac{(X_r)^2}{N^2} = 2\frac{X_r}{N}\Big(1 - \frac{X_r}{N}\Big)$$

これらから,X_rの分散は,

$$V[X_r] = E[V[X_r|X_{r-1}]] + V[E[X_r|X_{r-1}]] = E\Big[2\frac{X_{r-1}}{N}\Big(1 - \frac{X_{r-1}}{N}\Big)\Big] + V[X_{r-1}]$$

$$= V[X_{r-1}]+2E[X_0]\frac{1}{N}-\frac{2}{N^2}(V[X_{r-1}]+E[(X_{r-1})^2]) = V[X_{r-1}]\Big(1-\frac{2}{N^2}\Big)+\frac{2E[X_0]}{N}\Big(1-\frac{E[X_0]}{N}\Big)$$

この漸化式を解けば,

$$V[X_r] = V[X_0]\Big(1-\frac{2}{N^2}\Big)^r + \frac{2E[X_0]}{N}\Big(1-\frac{E[X_0]}{N}\Big)\sum_{l=0}^{r-1}\Big(1-\frac{2}{N^2}\Big)^l$$

同様にヘテロ接合度を求めると,

$$E[H_{r-1}] = E\Big[2\frac{X_{r-1}}{N}\Big(1-\frac{X_{r-1}}{N}\Big)\Big] = E[V[X_r|X_{r-1}]] = V[X_r]-V[X_{r-1}]$$

$$= \Big(1-\frac{2}{N^2}\Big)^{r-1}\frac{2}{N^2}V[X_0] + \frac{2E[X_0]}{N}\Big(1-\frac{E[X_0]}{N}\Big)\Big(1-\frac{2}{N^2}\Big)^r$$

これより, $\lim_{r\to\infty} H(r) = 0$ となる.これは, ライト・フィッシャーモデルと同様に, 最終的にある遺伝子の固定が起きることを表している.

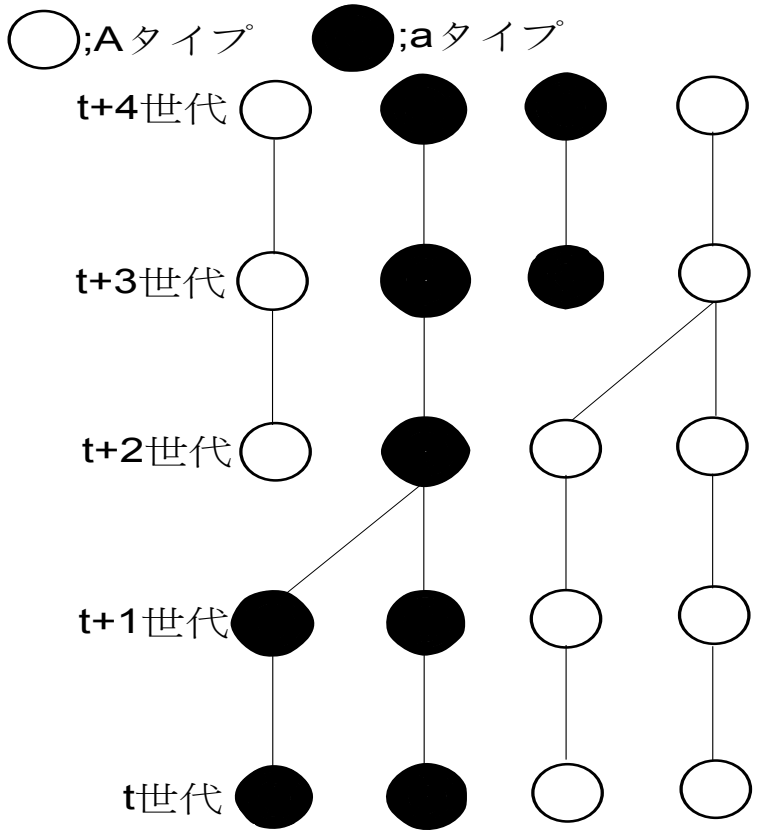

図 7　2 アレルタイプの場合のモランモデルの図 (N=4 の場合. 2 つのアレルタイプの個体が次世代の集団の大きさを自らの子孫によって同じくしている.)

2.2 Kingman の合祖理論 (Coalescent theory)

Kingman(1982a,b,c) はモランモデルやライト・フィッシャーモデルなどを含む可換モデルに対して, 遺伝子の系図を遡り, かつ集団サイズを無限に大きくすることで合祖過程と呼ばれる連続時間マルコフ連鎖 (死滅過程) に収束することを示した. N 個の半数体生物からなる集団を考え, この集団から n 個の個体 $I_1, I_2, \cdots, I_n$ をサンプルしたとする. その祖先によって同値関係 $R_N(t)$ を次のように定義する. 即ち, 2 つの個体 I_k, I_j が t 世代前に祖先を共有するとき $(I_k, I_j) \in R_N(t)$ と書き,この時刻で同じ同値類に属すということにする.(I_k, I_j) をクラスという. また, 初期状態はすべてが異なる状態でこれを $R_N(0) = \Delta$ で表すことにするとき,$\{R_N(t)\}_{t=0,1,2,\cdots}$ は同値関係を状態空間とするマルコフ連鎖である. 今, E_n を集合 $\{I_1, I_2, \cdots, I_n\}$ 上に定義される全ての同値関係の集合とするとき, 離散時間のマルコフ連鎖 $\{R_N(t)\}_{t=0,1,2,\cdots}$ の遷移確率を $\alpha, \beta \in E_n$ を用いて次のように構成しよう. αがβの細分である時$\alpha \subseteq \beta$で表す. 例えば $n = 5$ のとき,$\alpha = \{(I_1, I_2)(I_3)(I_4, I_5)\}, \beta = \{(I_1, I_2)(I_3, I_4, I_5)\}$ とすれば,αはβの細分である.これをマルコフ連鎖 $\{R_N(t)\}$ を用いて書き表せば,例えば,$R_N(0) = \{(I_1)(I_2)(I_3)(I_4)(I_5)\}, R_N(1) = \{(I_1, I_2)(I_3)(I_4, I_5)\}, R_N(2) = \{(I_1, I_2, I_3)(I_4, I_5)\}, R_N(3) = \{(I_1, I_2, I_3, I_4, I_5)\}$ とすれば,$R_N(0) \subseteq R_N(1) \subseteq R_N(2) \subseteq R_N(3)$ となり,またそれぞれクラスの数は $5, 3, 2, 1$ である.$|\alpha|$ は状態αにおけるクラスの数もしくは総サンプル数を表すとすれば,それぞれ $|R_N(0)| = 5, |R_N(1)| = 3, |R_N(2)| = 2, |R_N(3)| = 1$ である. t 世代前の祖先の状態が$\alpha \in E_n$ のとき, $t+1$ 世代前の祖先の状態が$\beta \in E_n$である推移確率 $P_{\alpha,\beta}$を求めると以下の様になる.
但し, $|\alpha| = |\beta| + 1$ でαがβの細分の時$\alpha \prec \beta$で表す.

定理 (Kingman(1982c)) v_iを i 番目の親が産む子供の個体数を表す確率変数で,$(v_1, v_2, \cdots, v_N)$ は可換モデル,かつ$\sum_{i=1}^{N} v_i = N$ を満たすとする.このとき,全ての i に対し極限値 $\lim_{N\to\infty} Var(v_i) = \sigma^2 > 0$ が存在,かつ,全ての $p, l \geq 1$ に対して$\sup_N E[{v_l}^p] < \infty$ が成り立つとき

$$P_{\alpha,\beta} = \begin{cases} 1 - \binom{|\alpha|}{2} \dfrac{\sigma^2}{N} + o(\dfrac{1}{N}) & (|\beta| = |\alpha| \text{ のとき}) \\ \dfrac{\sigma^2}{N} + o(\dfrac{1}{N}) & (|\beta| = |\alpha| - 1 \text{ のとき}) \\ o(\dfrac{1}{N}) & (\text{その他}) \end{cases} \tag{10}$$

$\dfrac{N}{\sigma^2}$世代を単位時間 $t = 1$ とする時間スケールを取り,$N \to \infty$ の極限をとると,サンプル数 k のときの滞在時間の分布はモランモデルとライトフィッシャーモデルの時と同様に,

$$P(T_k > t) = \lim_{N\to\infty} \Big(1 - \frac{k(k-1)\sigma^2}{2N}\Big)^{\left[\frac{Nt}{\sigma^2}\right]} = exp(-\frac{k(k-1)}{2}t) \quad (k \geq 2)$$

T_{MRCA}, T_{total}の平均時間はそれぞれ

$$E[T_{MRCA}] = 2(1 - \frac{1}{n})$$

$$E[T_{total}] = \sum_{i=2}^{n} \frac{2}{(i-1)} \cong 2log(n)$$

T_{MRCA}, T_{total}の密度関数はそれぞれ指数分布の畳み込みの式

$$f_{\sum_{i=1}^{n} T_i}(t) = \sum_{i=1}^{n} \lambda_i e^{-\lambda_i t} \prod_{j=1, j\neq i}^{n} \frac{\lambda_j}{\lambda_j - \lambda_i}$$

からすぐにわかる. 但し,

$$f_{T_i}(t) = \lambda_i e^{-\lambda_i t}$$

である. 証明は以下のように帰納法で示される. : $n = 3$ のとき

$$f_{T_2+T_3}(t) = \int_0^t f_{T_3}(s) f_{T_2}(t-s) ds = \int_0^t \lambda_3 e^{-\lambda_3 s} \lambda_2 e^{-\lambda_2 (t-s)} ds = \lambda_3 \lambda_2 e^{-\lambda_2 t} \int_0^t e^{-(\lambda_3 - \lambda_2)s} ds$$

$$= \frac{\lambda_3}{\lambda_3 - \lambda_2} \lambda_2 e^{-\lambda_2 t} (1 - e^{-(\lambda_3 - \lambda_2)t}) = \frac{\lambda_3}{\lambda_3 - \lambda_2} \lambda_2 e^{-\lambda_2 t} + \frac{\lambda_2}{\lambda_2 - \lambda_3} \lambda_3 e^{-\lambda_3 t}$$

$n = k - 1$ まで成立するとする.この時, $n = k$ で成立することを示す.

$$f_{\sum_{i=1}^{k} T_i}(t) = \int_0^t f_{\sum_{i=1}^{k-1} T_i}(s) f_{T_k}(t-s) ds = \int_0^t \sum_{i=1}^{k-1} \lambda_i e^{-\lambda_i s} \prod_{j=1, j\neq i}^{k-1} \frac{\lambda_j}{\lambda_j - \lambda_i} e^{-\lambda_k (t-s)} ds$$

$$= \sum_{i=1}^{k-1} \lambda_i (\frac{1}{\lambda_k - \lambda_i} e^{-\lambda_i t} + \frac{1}{\lambda_i - \lambda_k} e^{-\lambda_k t}) \lambda_k \prod_{j=1, j\neq i}^{k-1} \frac{\lambda_j}{\lambda_j - \lambda_i}$$

$$= \sum_{i=1}^{k-1} (\sum_{l=i,k} \lambda_l e^{-\lambda_l t} \prod_{m=k,i \ m\neq l} \frac{\lambda_m}{\lambda_m - \lambda_l}) \prod_{j=1, j\neq i}^{k-1} \frac{\lambda_j}{\lambda_j - \lambda_i}$$

$$= \sum_{i=1}^{k-1} \sum_{l=i,k} \lambda_l e^{-\lambda_l t} \prod_{m=k,i \ m\neq l} \frac{\lambda_m}{\lambda_m - \lambda_l} \prod_{j=1, j\neq i}^{k-1} \frac{\lambda_j}{\lambda_j - \lambda_i}$$

$$= \sum_{i=1}^{k-1} \sum_{l=i,k} \lambda_l e^{-\lambda_l t} \prod_{j=1, j\neq i}^{k} \frac{\lambda_j}{\lambda_j - \lambda_i} = \sum_{i=1}^{k} \lambda_i e^{-\lambda_i t} \prod_{j=1, j\neq i}^{k} \frac{\lambda_j}{\lambda_j - \lambda_i}$$

これで証明は完了した. 今, この式を用いて, $\lambda_i = \frac{i(i-1)}{2}$とすると,

$$f_{T_{MRCA}}(t) = \sum_{i=2}^{n} \frac{i(i-1)}{2} e^{-\binom{i}{2} t} \prod_{j=2, j\neq i}^{n} \frac{\binom{j}{2}}{\binom{j}{2} - \binom{i}{2}} = (-1)^n \prod_{j=2, j\neq i}^{n} \frac{j(j-1)}{(i-j)(i+j-1)}$$

$$= \frac{n!(n-1)!(-1)^n}{i(i-1)} \frac{1}{[(i-2)(i-3)\cdots 1][(-1)(-2)\cdots(-(n-i))]}$$
$$* \frac{1}{[(i+1)(i+2)\cdots(2i-2)][(2i)(2i+1)\cdots(i+n-1)]}$$
$$= \frac{(-1)^n n!(n-1)!(2i-1)i!}{i!(-1)^{n-i}(n-i)!(i+n-1)!} = \frac{(-1)^i(2i-1)n(n-1)\cdots(n-i+1)}{n(n+1)\cdots(n+i-1)}$$

以上のことから,

$$f_{T_{MRCA}}(t) = \sum_{i=2}^{n}(-1)^i \frac{(2i-1)n(n-1)\cdots(n-i+1)}{n(n+1)\cdots(n+i-1)} \frac{i(i-1)}{2} e^{-\binom{i}{2}t}$$

T_{total}に関しては,

$$f_{iT_i}(t) = \frac{i-1}{2} e^{-\frac{i-1}{2}t}$$

同様にして,

$$f_{T_{total}}(t) = \sum_{i=2}^{n} \frac{i-1}{2} e^{-\frac{i-1}{2}t} \prod_{j=2,j\neq i} \frac{\frac{j-1}{2}}{\frac{j-1}{2} - \frac{i-1}{2}} = \sum_{i=2}^{n} \frac{i-1}{2} e^{-\frac{i-1}{2}t} \prod_{j=2,j\neq i} \frac{j-1}{j-i}$$
$$= \sum_{i=2}^{n} \frac{i-1}{2} e^{-\frac{i-1}{2}t} \frac{1*2*\cdots*(i-2)*i*\cdots*(n-1)}{((2-i)(3-i)\cdots(-1))(n-i)!}$$
$$= \sum_{i=2}^{n} \frac{1}{2} e^{-\frac{i-1}{2}t} \frac{(n-1)!}{(-1)^{i-2}(i-2)!(n-i)!} = \sum_{i=2}^{n} \frac{n-1}{2} \frac{(n-2)!(-1)^{i-2}}{(i-2)!(n-i)!} e^{-\frac{i-2}{2}t} e^{-\frac{t}{2}}$$
$$= \frac{n-1}{2} e^{-\frac{t}{2}} \{\sum_{i=2}^{n} \binom{n-2}{i-2} (-1)^{i-2} (e^{-\frac{t}{2}})^{i-2}\}$$

$i-2=k$ とすると,

$$= \frac{n-1}{2} e^{-\frac{t}{2}} \{\sum_{k=0}^{n-2} \binom{n-2}{k} (-1)^k (e^{-\frac{t}{2}})^k\} = \frac{n-1}{2} e^{-\frac{t}{2}} (1-e^{-\frac{t}{2}})^{n-2}$$

よって,

$$f_{T_{total}}(t) = \frac{n-1}{2} e^{-\frac{t}{2}} \sum_{j=0}^{n-2} \binom{n-2}{j} 1^{n-2-j} (-e^{-\frac{t}{2}})^j = \frac{n-1}{2} e^{-\frac{t}{2}} (1-e^{-\frac{t}{2}})^{n-2}$$

また, 次のような式変形も成り立つ.

$$\prod_{j=2,j\neq i} \frac{\frac{j-1}{2}}{\frac{j-1}{2} - \frac{i-1}{2}} = \prod_{2\leq j\leq i-1} \frac{\frac{j-1}{2}}{\frac{j-1}{2} - \frac{i-1}{2}} * \prod_{i+1\leq j\leq n} \frac{\frac{j-1}{2}}{\frac{j-1}{2} - \frac{i-1}{2}}$$

$$\prod_{2\leq j\leq i-1} \frac{\frac{j-1}{2}}{\frac{j-1}{2} - \frac{i-1}{2}} = \frac{1*2\cdots*(i-2)}{(2-i)(3-i)\cdots(-2)(-1)} = (-1)^{i-2} = (-1)^i$$

$$\prod_{i+1\leq j\leq n}\frac{\frac{j-1}{2}}{\frac{j-1}{2}-\frac{i-1}{2}}=\frac{i(i+1)\cdots(n-1)}{1*2*\cdots(n-i-1)(n-i)}=\frac{(n-1)!}{(i-1)!(n-i)!}=\binom{n-1}{i-1}$$

よって, 次の様な表現式も成り立つ.

$$f_{T_{total}}(t)=\sum_{i=2}^{n}(-1)^i\binom{n-1}{i-1}\frac{i-1}{2}e^{-\frac{i-1}{2}t}$$

2.3 有限次元分布の収束 (Kingman(1982c))

これまで, ライトフィッシャーモデルではタイムスケールが N, モランモデルではタイムスケールを$\frac{N^2}{2}$とした離散時刻マルコフ連鎖に対し, N に関して極限をとることにより連続時間の同じ生成作用素に従うマルコフ連鎖に収束することを示した. 今度はもっと一般的なタイムスケールにおいて, 同じく生成作用素 $\mathbb{Q}$ に従う連続時間のマルコフ連鎖に収束することを示すことで定式化する. $\mathbb{A}=(a_{\epsilon,\eta})$ を $a_{\epsilon,\eta}\geq 0$, $\sum_{\eta}a_{\epsilon,\eta}=1$ を満たす確率行列とする.今, ノルムを $||\mathbb{A}||=\max_{\epsilon}\sum_{\eta}|a_{\epsilon,\eta}|$ とすると, $\mathbb{A}$ は縮小作用素 $(||\mathbb{A}||\leq 1)$ である. 2 つの縮小作用素 $\mathbb{A}_i,\mathbb{B}_i(i=1,2,\cdots,r)$ を用いれば $||\mathbb{A}_1\mathbb{A}_2\cdots\mathbb{A}_r-\mathbb{B}_1\mathbb{B}_2\cdots\mathbb{B}_r||\leq\sum_{i=1}^{r}||\mathbb{A}_i-\mathbb{B}_i||$.まず, $r=2$ の時,

$$\begin{aligned}||\mathbb{A}_1\mathbb{A}_2-\mathbb{B}_1\mathbb{B}_2||&\leq||\mathbb{A}_1\mathbb{A}_2-\mathbb{A}_1\mathbb{B}_2||+||\mathbb{A}_1\mathbb{B}_2-\mathbb{B}_1\mathbb{B}_2||\\&\leq||\mathbb{A}_1||\cdot||\mathbb{A}_2-\mathbb{B}_2||+||\mathbb{A}_1-\mathbb{B}_1||\cdot||\mathbb{B}_2||\\&\leq\sum_{i=1,2}||\mathbb{A}_i-\mathbb{B}_i||\end{aligned}$$

同様にして一般の r の場合を証明する.

$$\begin{aligned}||\mathbb{A}_1\mathbb{A}_2\cdots\mathbb{A}_r-\mathbb{B}_1\mathbb{B}_2\cdots\mathbb{B}_r||&\leq||\mathbb{A}_1\mathbb{A}_2\cdots\mathbb{A}_{r-1}\mathbb{A}_r-\mathbb{A}_1\mathbb{A}_2\cdots\mathbb{A}_{r-1}\mathbb{B}_r||\\&+||\mathbb{A}_1\mathbb{A}_2\cdots\mathbb{A}_{r-1}\mathbb{B}_r-\mathbb{A}_1\mathbb{A}_2\cdots\mathbb{B}_{r-1}\mathbb{B}_r||+\cdots\\&\cdots+||\mathbb{A}_1\mathbb{B}_2\cdots\mathbb{B}_{r-1}\mathbb{B}_r-\mathbb{B}_1\mathbb{B}_2\cdots\mathbb{B}_r||\leq\sum_{i=1}^{r}||\mathbb{A}_i-\mathbb{B}_i||\end{aligned}$$

これより,

$$||{\mathbb{P}_N}^{[\frac{t}{c_N}]}-(\mathbb{I}+c_N\mathbb{Q})^{[\frac{t}{c_N}]}||\leq\frac{t}{c_N}||\mathbb{P}_N-(\mathbb{I}+c_N\mathbb{Q})||=t||\frac{\mathbb{P}_N-\mathbb{I}}{c_N}-\mathbb{Q}||$$

ゆえに,

$$\lim_{N\to\infty}{\mathbb{P}_N}^{[\frac{t}{c_N}]}=\lim_{N\to\infty}(\mathbb{I}+c_N\mathbb{Q})^{[\frac{t}{c_N}]}=e^{t\mathbb{Q}}$$

但し, $\mathbb{P}_N$は (α,β) 成分が $P_N(\alpha,\beta)$ で表される離散時刻の遷移確率行列で, c_Nは N に関する時間スケールである. これより, 連続時間の生成作用素 $\mathbb{Q}$ に従うマルコフ連鎖 $\{\mathrm{R(t)}\}$ への収束が導かれる. 後, 有限次元分布の収束を証明する. 離散時間マルコフ連鎖 $\{R_N(t)\}$ のタイムスケールを

$\frac{1}{c_N}$として取り直した確率過程 $R_N\Big([\frac{t}{c_N}]\Big)$ を$\hat{R}_N(t)$ とした時,任意の $n \in \mathbb{N}$($\mathbb{N}$ は自然数全体),時点 $t_1 < t_2 < \cdots < t_n,\ x_1, \cdots, x_n \in E_n$に対し,

$$\lim_{N\to\infty} P(\hat{R}_N(t_1) = x_1, \hat{R}_N(t_2) = x_2, \cdots, \hat{R}_N(t_n) = x_n) = P(R(t_1) = x_1, R(t_2) = x_2, \cdots, R(t_n) = x_n)$$

となることを示す. 証明は以下のようである.今,初期状態を x_0とすると,

$$P(\hat{R}_N(t_1) = x_1, \hat{R}_N(t_2) = x_2, \cdots, \hat{R}_N(t_n) = x_n) = (P_N(x_0, x_1))^{[\frac{1}{c_N}t_1]} \cdots (P_N(x_{n-1}, x_n))^{[\frac{1}{c_N}t_n]-[\frac{1}{c_N}t_{n-1}]}$$

N に関して極限をとれば,

$$\begin{aligned}
&\lim_{N\to\infty} P(\hat{R}_N(t_1) = x_1, \hat{R}_N(t_2) = x_2, \cdots, \hat{R}_N(t_n) = x_n) \\
&= (e^{t_1\mathbb{Q}})_{x_0,x_1}(e^{(t_2-t_1)\mathbb{Q}})_{x_1,x_2} \cdots (e^{(t_n-t_{n-1})\mathbb{Q}})_{x_{n-1},x_n} \\
&= P(R(t_1) = x_1, R(t_2) = x_2, \cdots, R(t_n) = x_n)
\end{aligned}$$

これより,有限次元分布の収束が証明された.Kingmanの合祖理論に関しては,$c_N = \frac{\sigma^2}{N}$として行えば,ライトフィッシャーモデルやモランモデルと同じ生成作用素;

$$\mathbb{Q}_{\alpha,\beta} = \begin{cases} -\binom{|\alpha|}{2} & (|\beta| = |\alpha| \text{ のとき}) \\ 1 & (|\beta| = |\alpha| - 1 \text{ のとき}) \\ 0 & (\text{その他}) \end{cases} \tag{11}$$

に従う連続時間のマルコフ連鎖に収束する.この結果からもわかるように,Kingmanの合祖理論においてもライトフィッシャーモデルやモランモデルのように同じ生成作用素に従う連続時間のマルコフ連鎖に収束することがわかった.また,有限次元分布の収束により,離散時刻から連続時間への経路（$path$）の収束が導かれた.

2.4 突然変異を含む合祖過程

この節では突然変異を考慮に入れたとき, サンプル遺伝子の DNA 配列中の分離サイトの数の分布およびサンプル中の異なるアレルタイプの数に関する Ewens のサンプリング公式を導く.

突然変異は常に異なる塩基サイトに起こるという無限サイトモデル(Infinite-site model)を仮定する. 1 遺伝子当たりの突然変異率を $u = \dfrac{\theta}{2N}$とする.

突然変異の数を K とすると, 時間 t の間に生じる突然変異の数は次の二項分布に従う.

$$P(K = k|t) = \binom{N}{k} \left(\frac{\theta t}{2N}\right)^k \left(1 - \frac{\theta t}{2N}\right)^{N-k}$$

よって, N に関して極限をとれば, ポアソン分布になる.

$$\lim_{N\to\infty} \binom{N}{k} \left(\frac{\theta t}{2N}\right)^k \left(1 - \frac{\theta t}{2N}\right)^{N-k} = \left(\frac{\theta t}{2}\right)^k \frac{1}{k!} e^{-\frac{\theta t}{2}}$$

となる. 今これを連続時間の突然変異数における確率とし,

$$P(K = k|t) = \left(\frac{\theta t}{2}\right)^k \frac{1}{k!} e^{-\frac{\theta t}{2}}$$

と書くことにする. サンプルした n 本の配列の中にある分離サイト (segregating sites) の数を S とする. 無限サイトモデルのもとでサンプル中の分離サイトの数は系図上の全長 (T_{total}) 上に生じた突然変異の数に等しいので, この確率は次のように表される：

$$P(S = k) = \int_0^\infty P(S = k|t) f_{T_{total}}(t) dt = \left(\frac{\theta}{2}\right)^k \sum_{i=2}^{n} (-1)^i \binom{n-1}{i-1} \frac{i-1}{2} \int_0^\infty \frac{t^k}{k!} e^{-\frac{\theta+i-1}{2} t} dt$$

部分積分法によって,

$$= \left(\frac{\theta}{2}\right)^k \sum_{i=2}^{n} (-1)^i \binom{n-1}{i-1} \frac{i-1}{2} \left(\frac{2}{\theta+i-1}\right)^{k+1} = \sum_{i=2}^{n} (-1)^i \binom{n-1}{i-1} \left(\frac{i-1}{\theta+i-1}\right) \left(\frac{\theta}{\theta+i-1}\right)^k$$

突然変異数の平均及び分散は Watterson(1975) より

$$E[S] = E[K]E[T_{total}] = \theta \sum_{i=1}^{n-1} \frac{1}{i} \cong \theta log(n)$$

$$V[S] = V[K]E[T_{total}] + E^2[K]V[T_{total}] = \left(\frac{\theta}{2}\right)\left(2\sum_{i=1}^{n-1} \frac{1}{i}\right) + \left(\frac{\theta}{2}\right)^2 \left(4\sum_{i=1}^{n-1} \frac{1}{i^2}\right) = \theta \sum_{i=1}^{n-1} \frac{1}{i} + \theta^2 \sum_{i=1}^{n-1} \frac{1}{i^2}$$

備考: X_iを各 i について独立同分布な確率変数とする. $Y = \sum_{i=1}^{K} X_i (K$ は確率変数) とした時, 平均, 分散に対して次の式が成り立つ.

$$E[Y] = E[K]E[X_i]$$

$$V[Y] = E[K]V[X_i] + V[K]E^2[X_i]$$

Ewens のサンプリング公式 (Ewens(1972)); ある集団から n 個のサンプルを取り出したとき, その中に含まれる各アレルタイプの遺伝子の個数についてEwensのサンプリング公式と呼ばれる式がある.アレルタイプの数を p とする.

a_k : サンプル中に k 個含まれているアレルタイプの数

$a=(a_1,a_2,\cdots,a_n)$ とすると, $\sum_{k=1}^{n} ka_k=n,\ \sum_{k=1}^{p} a_k=p$ である.この時,

$$P(a_1,a_2,\cdots,a_n)=\frac{n!}{(\theta)_n}\prod_{j=1}^{n}\frac{\theta^{a_j}}{j^{a_j}a_j!}$$

但し, $(\theta)_n=\theta(\theta+1)\cdots(\theta+n-1)$ である.この式の導出方法を説明しよう. 系図上で祖先の数が j の時, 生じる事象がその時点で k 個の祖先をもつアレルクラスで合祖が起こる確率は

$$\frac{k(k-1)}{2}\ /\ \frac{j\theta+j(j-1)}{2}$$

突然変異である確率は

$$\frac{\theta}{2}\ /\ \frac{j\theta+j(j-1)}{2}$$

であるから, よって, 1 から n までナンバーのついた遺伝子中に, $A_1,A_2,\cdots,A_p$の p 種のアレルをそれぞれ $k_1,k_2,\cdots,k_p$個ずつ含んでいる確率は

$$P(k_1,k_2,\cdots,k_p)=\prod_{j=1}^{n}\frac{1}{\frac{j\theta+j(j-1)}{2}}\prod_{i=1}^{p}\Big(\prod_{l=2}^{k_i}\frac{l(l-1)}{2}\Big)\Big(\frac{\theta}{2}\Big)^p\frac{n!}{\prod_{i=1}^{p}k_i!}=\frac{\theta^p}{(\theta)_n}\prod_{i=1}^{p}(k_i-1)!$$

$a=(a_1,a_2,\cdots,a_n)$ だから, i 個のサンプルを含むアレルタイプが a_i個あるので $\prod_{i=1}^{p} a_i!$通りある.そのそれぞれのアレルタイプ内で i 個の遺伝子について順番の付け方は $i!$通り.よって, n 個のサンプル遺伝子において, アレルタイプと, そのタイプ内での順番によって

$$\frac{n!}{\prod_{j=1}^{n}(j!)^{a_j}a_j!}$$

通りある. $k_i=j$ となるラベル i の個数は a_j個あることから,

$$\prod_{i=1}^{p}(k_i-1)!=\prod_{j=1}^{n}((j-1)!)^{a_j}$$

となることがわかるから, これを前式の $P(k_1,k_2,\cdots,k_p)$ に用いると, 求める式が出る. 具体的な説明を以下の図 8 で示す.

(例)

$n=5, p=2$ の場合, $k_1=3, k_2=2$ とすれば,

$$\frac{\frac{3*2}{2}}{\frac{5\theta+5*4}{2}} * \frac{\frac{2*1}{2}}{\frac{4\theta+4*3}{2}} * \frac{\frac{2*1}{2}}{\frac{3\theta+3*2}{2}} * \frac{\frac{\theta}{2}}{\frac{2\theta+2*1}{2}} * \frac{\frac{\theta}{2}}{\frac{\theta}{2}} = \prod_{j=1}^{5} \frac{1}{\frac{j\theta+j(j-1)}{2}} \prod_{i=1}^{2} \Big(\prod_{l=2}^{k_i} \frac{l(l-1)}{2} \Big) \Big(\frac{\theta}{2} \Big)^2$$ となる.

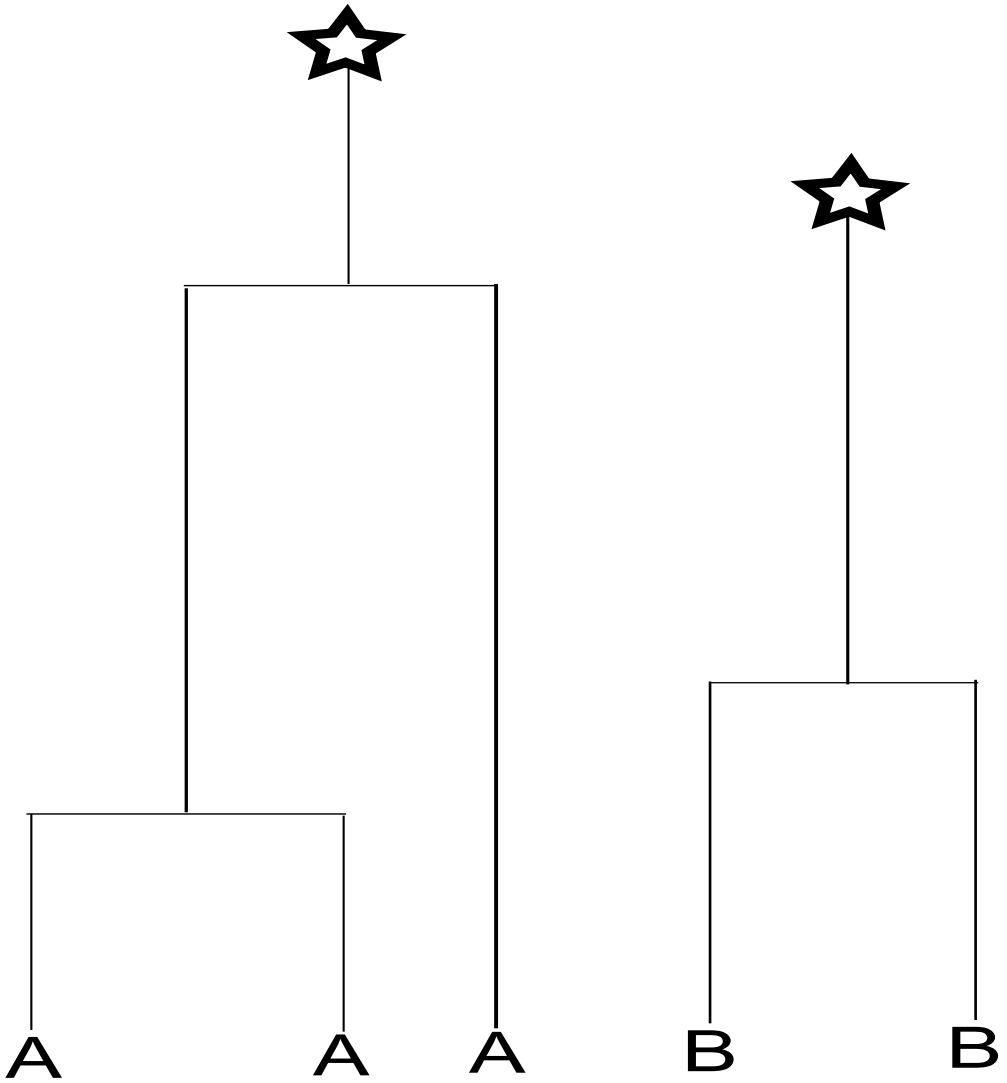

図 8　$n=5, p=2, k_1=3, k_2=2$ の場合の系図例

3　地理的構造を持つ合祖モデル (Structured Coalescent model)

第 2 章の合祖モデルは任意交配集団という条件で導かれるものであり, 現実の生物集団の生息域には様々な地理的制約があり, 個体間の交配は多かれ少なかれある程度の範囲に限られる. このような地理的構造を考慮に入れたモデルには歴史的には Malecot(1967), Kimura(1953), Kimura and Weiss(1964), Maruyama(1970) 等の多くの研究がある. これらの研究は離れた分集団間からサンプルした 2 つの遺伝子が Identical by Descent である確率（同じ祖先に由来する確率）を求めたものである.Kimura(1953) のモデルは飛び石モデルと呼ばれているが, 地理的構造を持つ Structured Coalescent Model(SCM) は飛び石モデルを任意個数のサンプルに拡張したモデルと考えることもできる. 一般的な形での SCM は Notohara(1990) によって導入されたが, 数学的に厳密な意味での SCM の導出の証明は Herbots(1994,1997) によって保存的移住率でかつ繁殖がライト・フィッシャーモデルという限定的な条件下で与えられた. 本論文では一般的な非保存的な移住率でかつ可換モデルという一般的な繁殖モデルで, 適正な条件の下で SCM が導かれることを示す. 本章では, 移住後の集団のサイズが変化する非保存的移住 (Non-conservative migration) とより一般的な繁殖 (Cannings' reproduction) の場合に対しタイムスケールを取り直した離散時刻の祖先過程が, 集団サイズ (N) を無限に大きくすることにより, 生成作用素；

$$\mathbb{Q}_{\alpha,\beta} = \begin{cases} -\sum_{i\in S}\left(\alpha_i\dfrac{M_i}{2} + \dfrac{\sigma^2\alpha_i(\alpha_i-1)}{2c_i}\right) & (\beta=\alpha\text{のとき}) \\ \alpha_i\dfrac{M_{i,j}}{2} & (\beta=\alpha-\epsilon^i+\epsilon^j\ (i\neq j)\text{ のとき}) \\ \dfrac{\sigma^2\alpha_i(\alpha_i-1)}{2c_i} & (\beta=\alpha-\epsilon^i\text{のとき}) \\ 0 & (\text{その他}) \end{cases}$$

に従う地理的構造を持つ合祖過程に収束することを示す. 但し, 後に詳しく述べるが, α_iは分集団 i に属する祖先の数, $S=\{k,j,\cdots\}$ で k,j は各分集団を表す.$M_{i,j}$は分集団 i から分集団 j に移ってきた個体数の割合. また, $M_i=\sum_{j\neq i}M_{i,j}$である. α,βは祖先の地理的配置を表す無限次元のベクトルで, c_iは分集団 i のサイズを決定する比例定数である. $\mathbb{Q}$ の成分に対しては, $\alpha_i\dfrac{M_{i,j}}{2}$は祖先$\alpha_i$の中の 1 つが分集団 i から分集団 j に移住する割合, $\dfrac{\sigma^2\alpha_i(\alpha_i-1)}{2c_i}$は分集団 i の祖先数がα_iである時の合祖率を表している. $N_i=2c_iN$ を繁殖後の分集団 i の個体数, $N_i{}^*$を 移住後の分集団 i の個体数とすれば, 1 世代当たりの集団サイズの推移は次の図 9 のように表される；

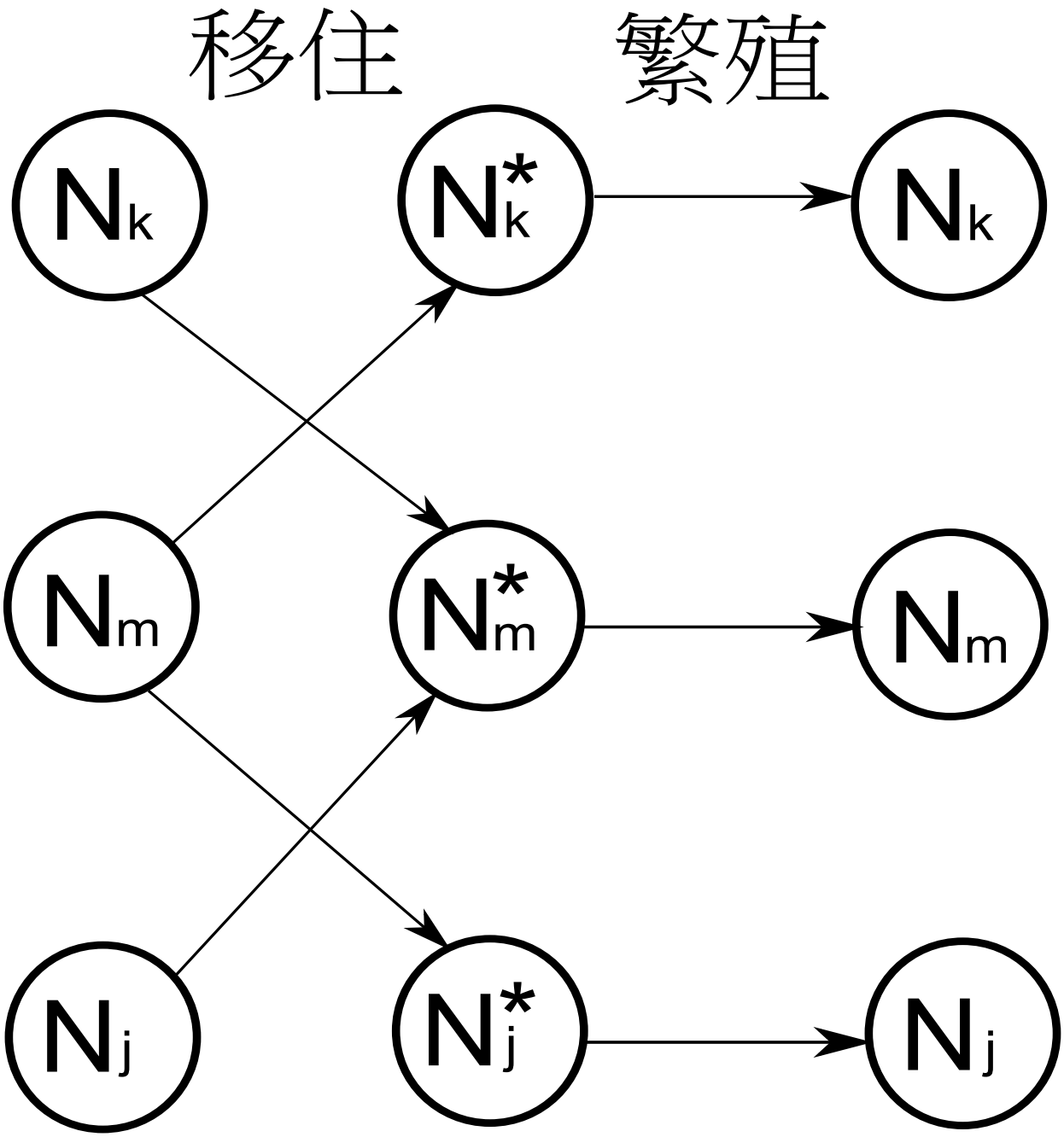

図 9　移住と繁殖のプロセス (分集団の数が 3 つの場合. 移住前の集団サイズを N_k, N_j, N_m とした時, 移住によって集団サイズが ${N_k}^*, {N_m}^*, {N_j}^*$ に変化し, 繁殖によって元の集団サイズ N_k, N_j, N_m に戻る.)

3.1 離散時間モデル

任意交配を行っている一つの生物集団からランダムに取り出した複数の中立な遺伝子サンプルの系図は, 1980 年代初めに Kingman 等によってある条件下で, Coalescent モデルと呼ばれる連続時間マルコフ連鎖モデルで表現されることが示された. しかし, 生物集団が任意交配を行うことは稀であり, 生物集団がもつ地理的な構造は遺伝的多様性及び進化に大きな影響を持つ. この研究では地理的構造を持つ生物集団からランダムに取り出した複数の遺伝子の系図を表現するモデルを離散時間モデルから出発し, 集団サイズを大きくするという極限操作により連続時間の Structured Coalescent モデルと呼ばれる連続時間マルコフ過程へ弱収束 (確率分布の収束) することを証明する. Herbots(1997) は保存的移住 (conservative migration) という条件の下で離散時間モデルから連続時間モデルへの収束を示したが, これをより一般の移住率の下で証明を与えることができた.

移住率について

$q_{i,j}$を $\sum_{j\neq k} q_{k,j} \leq 1$ を満たす, 分集団 i から j へ移住する割合とする.
今, 次のような (i)(ii) を仮定する ;
(i) c_i と K は定数で $1 \leq c_i < K$ を満たす.
(ii)

$$q_{i,j} = \frac{{q_{i,j}}^*}{4N}, \qquad i \neq j.$$

ここで,

$$\sup_{k\in S} \sum_{j\in S,\ j\neq k} {q_{k,j}}^* < \infty, \ \ \sup_{k\in S} \sum_{j\in S,\ j\neq k} {q_{j,k}}^* < \infty \tag{12}$$

が成り立つ. 後向き移住率 $m_{i,j}$は移住後の分集団 i での総個体数と分集団 j から分集団 i へ移住した個体数との割合であり, 次式で与えられる.

$$m_{i,j} = \frac{N_j q_{j,i}}{{N_i}^*}, \qquad i \neq j.$$

また, 分集団 i から他の分集団へ移住する割合 m_i は,

$$m_i = \sum_{j\neq i} m_{i,j} = \sum_{j\neq i} \frac{N_j q_{j,i}}{{N_i}^*}$$

で与えられる. 分集団 i における移住後の個体数 ${N_i}^*$ について, 次のように表すことができる.

$$\begin{aligned} {N_i}^* &= \sum_{k\neq i} N_k q_{k,i} + N_i\Big(1 - \sum_{j\neq i} q_{i,j}\Big) \\ &= \sum_{k\neq i} 2c_k N \frac{{q_{k,j}}^*}{4N} + 2c_i N\Big(1 - \sum_{j\neq i} \frac{{q_{i,j}}^*}{4N}\Big) \end{aligned}$$

$$= 2c_iN + \frac{1}{2}\Big(\sum_{k\neq i} c_k {q_{k,i}}^* - c_i \sum_{j\neq i} {q_{i,j}}^*\Big) = 2c_iN + \frac{Q_i}{2} \tag{13}$$

但し,

$$Q_i = \sum_{k\neq i} c_k {q_{k,i}}^* - c_i \sum_{j\neq i} {q_{i,j}}^* \tag{14}$$

で表され, 任意の $i \in S$ に対して, $Q_i = 0$ が成り立つとき, 保存的移住と呼ばれる. Q_i について, 仮定から,

$$|Q_i| = |\sum_{k\neq i} c_k {q_{k,i}}^* - c_i \sum_{j\neq i} {q_{i,j}}^*| \leq K|\sum_{k\neq i} {q_{k,i}}^* + \sum_{j\neq i} {q_{i,j}}^*|$$

$$\leq K\Big(\sup_k \sum_{j\neq k} {q_{k,j}}^* + \sup_k \sum_{j\neq k} {q_{j,k}}^*\Big) \leq C \tag{15}$$

但し, C は i に依存しないある正定数である.また $m_{i,j}$ についての評価から, 連続時間での移住率は,

$$m_{i,j} = \frac{N_j q_{j,i}}{{N_i}^*} = \frac{2c_jNq_{j,i}}{2c_iN + \frac{1}{2}Q_i} \leq \frac{K}{2}\frac{{q_{j,i}}^*}{2N - \frac{1}{2}C}$$

$$\lim_{N\to\infty} 4Nm_{i,j} = \frac{c_j}{c_i}{q_{j,i}}^*$$

$$\frac{c_j {q_{j,i}}^*}{c_i + \frac{Q_i}{4N}} \leq \frac{c_j {q_{j,i}}^*}{1 - \frac{C}{4N}} \leq \frac{K{q_{j,i}}^*}{1 - \frac{C}{4N}} < \infty$$

その他, ルベーグの収束定理を用いて,

$$\lim_{N\to\infty} 4Nm_i = \lim_{N\to\infty} \sum_{j\neq i} 4Nm_{i,j} = \lim_{N\to\infty} \sum_{j\neq i} \frac{c_j {q_{j,i}}^*}{c_i + \frac{Q_i}{4N}} = \sum_{j\neq i} \frac{c_j}{c_i} {q_{j,i}}^*$$

よって, 以下では, 連続時間の移住率をそれぞれ

$$M_i = \sum_{j\neq i} \frac{c_j}{c_i} {q_{j,i}}^*$$

$$M_{i,j} = \frac{c_j}{c_i} {q_{j,i}}^*$$

で定義する. 他, 十分 N が大きいときは,

$$4Nm_i \leq 2\sum_{j\neq i} \frac{c_j}{c_i - \frac{C}{4N}} {q_{j,i}}^* \leq 2K \sum_{j\neq i} {q_{j,i}}^* = 2K\Big(\sup_i \sum_{j\neq i} {q_{j,i}}^*\Big) = M$$

但し, M は正定数. 従って,

$$\sup_{i\in S} m_i \leq \frac{M}{4N}, \quad \sup_{i\in S} M_i \leq K \sup_{i\in S} \sum_{j\neq i} {q_{j,i}}^* \leq \frac{M}{2} \tag{16}$$

となる.

繁殖について（Cannings' Reproduction)(Cannings(1974))

非負整数全体を $\mathbb{Z}_+$で表す. 即ち, $\mathbb{Z}_+ = \{0, 1, 2, \cdots\}$ である. 現在から過去に向かって番号付けをした世代を $r \in \mathbb{Z}_+$で表す. $r = 0$ は現在の世代, $r = 1$ は 1 世代前の世代を表す.
${\nu_i}^{(l,r)}$ は分集団 l , 第 r 世代前における i 番目の個体の子供の数を表す.

この $\{{\nu_i}^{(l,r)}\}$ について, 次の仮定をおく.

仮定

(i) 任意の $l \in S$ に対して, $\displaystyle\sum_{i=1,2,\cdots,{N_l}^*} {\nu_i}^{(l,r)} = N_l$

(ii) $l(\in S)$ は固定, $({\nu_1}^{(l,r)}, {\nu_2}^{(l,r)}, \cdots, {\nu_{N_l}}^{(l,r)})$ は $r(\in \mathbb{Z}_+)$ に関して独立同分布に従う.

$(iii) l \in S, r \in \mathbb{Z}_+$を固定した時, ${\nu_1}^{(l,r)}, {\nu_2}^{(l,r)}, \cdots, {\nu_{{N_l}^*}}^{(l,r)}$は可換である.

即ち, $\{n_1, n_2, \cdots, n_d\}$ を $\{1, 2, \cdots, d\}(d = {N_l}^*)$ の任意の置換とするとき,

$$P\{{\nu_1}^{(l,r)} = s_1, {\nu_2}^{(l,r)} = s_2, \cdots, {\nu_d}^{(l,r)} = s_d\} = P\{{\nu_{n_1}}^{(l,r)} = s_1, {\nu_{n_2}}^{(l,r)} = s_2, \cdots, {\nu_{n_d}}^{(l,r)} = s_d\}$$

が成り立つ. ここで, $s_1, s_2, \cdots, s_d \in \mathbb{Z}_+$, かつ $s_1 + s_2 + \cdots + s_d = N_l$である.

(iv) 分散を$\sigma^2 (= E[({\nu_1}^{(l,r)})^2] - E^2[{\nu_1}^{(l,r)}]) > 0$ と置いたとき,

$$\lim_{N\to\infty} \sup_{l\in S} |E[\{{\nu_1}^{(l,r)}\}^2] - (\sigma^2 + 1)| = 0$$

(v) $k \geq 1$,

$$ {K_k}^* = \sup_{l\in S, N} E[\{{\nu_1}^{(l,r)}\}^k] < \infty \tag{17}$$

${c_N}^l$はある世代で分集団 l に含まれる選ばれた 2 個体が 1 世代遡った時に共通の親を持つ確率とすると,

$${c_N}^l = \frac{\sum_{i=1}^{{N_l}^*} E[{\nu_i}^{(l,r)}({\nu_i}^{(l,r)} - 1)]}{N_l(N_l - 1)} = \frac{{N_l}^* E[\{{\nu_i}^{(l,r)}\}^2]}{N_l(N_l - 1)} - \frac{1}{N_l - 1}$$

となる. ここで,

$$\lim_{N\to\infty} 2N{c_N}^l = \frac{\sigma^2}{c_l} \tag{18}$$

が成り立つことが容易にわかる.

後ろ向き移住行列

まずはじめにいくつか新しい仮定をおく.

$\alpha = (\alpha_1, \alpha_2, \cdots)$. 但し, $\alpha_i \in \mathbb{Z}_+$は分集団 i に属する祖先の数. 時刻の意味合いを付けて$\alpha = \alpha(t), t = 0, 1, 2, \cdots$とするとき, $|\alpha(0)| = n, |\alpha| = \sum_{k\in S} \alpha_k$は全ての分集団の祖先の総数を表す.

集合 $E = \{\alpha \in {\mathbb{Z}_+}^S : \sum_{i\in S} \alpha_i \leq n\}$ は全ての分集団における地理的配置を表すベクトルの全体集合である.

$\alpha = (\alpha_i; i \in S), \beta = (\beta_i; i \in S), \alpha \pm \beta \in E$ のとき, $\alpha \pm \beta = (\alpha_i \pm \beta_i; i \in S)$ と定義する.

また, $\epsilon^k \in E$ を単位ベクトル $(\epsilon^k)_i = \delta_{k,i}$で定義する.

先ず, 後ろ向き移住について考察する.

${R_N}^{(m)}(\alpha)$; αの中の 2 つ以上の祖先が移住者である確率とする.

$$
\begin{aligned}
{R_N}^{(m)}(\alpha) &\leq \sum_{k\in S} \frac{\binom{\alpha_k}{2}\binom{{N_k}^* - 2}{m_k {N_k}^* - 2}}{\binom{{N_k}^*}{m_k {N_k}^*}} + \sum_{k\in S} \frac{\binom{\alpha_k}{1}\binom{{N_k}^* - 1}{m_k {N_k}^* - 1}}{\binom{{N_k}^*}{m_k {N_k}^*}} \sum_{l\neq k} \frac{\binom{\alpha_l}{1}\binom{{N_l}^* - 1}{m_l {N_l}^* - 1}}{\binom{{N_l}^*}{m_l {N_l}^*}} \\
&\leq \sum_{k\in S} (\alpha_k m_k)^2 + \sum_{k\in S} \alpha_k m_k \sum_{l\neq k} \alpha_l m_l = (\sum_{k\in S} \alpha_k m_k)^2
\end{aligned}
$$

であるから, (14) により,

$$
{R_N}^{(m)}(\alpha) \leq \frac{M^2 |\alpha|^2}{16 N^2}
$$

となる. また, α の中の一つが分集団 i から分集団 j に移る確率は,

$$
\begin{aligned}
&\frac{\binom{\alpha_i}{1}\binom{{N_i}^* - \alpha_i}{m_{i,j} {N_i}^* - 1}}{\binom{{N_i}^*}{m_{i,j} {N_i}^*}} \frac{\binom{{N_i}^* - m_{i,j} {N_i}^* - \alpha_i + 1}{m_i {N_i}^* - m_{i,j} {N_i}^*}}{\binom{{N_i}^* - m_{i,j} {N_i}^*}{m_i {N_i}^* - m_{i,j} {N_i}^*}} \prod_{k\neq i} \frac{\binom{{N_k}^* - \alpha_k}{m_k {N_k}^*}}{\binom{{N_k}^*}{m_k {N_k}^*}} \\
&= \alpha_i m_{i,j} \frac{{N_i}^*}{{N_i}^* - m_i {N_i}^* - \alpha_i + 1} \prod_{k\in S} \prod_{a=0,\cdots,\alpha_k - 1} \frac{{N_k}^* - m_k {N_k}^* - a}{{N_k}^* - a}
\end{aligned}
$$

今, ${R_N}^{(m)}(\alpha, \beta)$ を後ろ向き移住において, α から β に変化し, かつ, α の中の 2 つ以上の個体が移住者である確率とすると, 後ろ向き移住における確率は次のようになる.

$P_N{}^{(m)}(\beta|\alpha) =$

$$\begin{cases} 1-\sum_{i\in S}\alpha_i m_i \dfrac{{N_i}^*}{{N_i}^*-m_i{N_i}^*-\alpha_i+1}\prod_{k\in S}\prod_{a=0,\cdots,\alpha_k-1}\dfrac{{N_k}^*-m_k{N_k}^*-a}{{N_k}^*-a}-\sum_{\gamma\neq\alpha}{R_N}^m(\alpha,\gamma) \\ \qquad\qquad if\ \beta=\alpha \\ \alpha_i m_{i,j}\dfrac{{N_i}^*}{{N_i}^*-m_i{N_i}^*-\alpha_i+1}\prod_{k\in S}\prod_{a=0,\cdots,\alpha_k-1}\dfrac{{N_k}^*-m_k{N_k}^*-a}{{N_k}^*-a}+{R_N}^m(\alpha,\alpha-\epsilon^i+\epsilon^j) \\ \qquad\qquad if\ \beta=\alpha-\epsilon^i+\epsilon^j\ (j\neq i) \\ {R_N}^{(m)}(\alpha,\beta)\qquad otherwise \end{cases}$$

また, 次式が成立する.

$$\sum_{\beta\neq\alpha}{R_N}^{(m)}(\alpha,\beta)\leq{R_N}^{(m)}(\alpha)\leq(\sum_{k\in S}\alpha_k m_k)^2$$

後ろ向き繁殖行列

${R_N}^{(r)}(\alpha)$ をα に含まれる 2 つ以上のペアが共通の親を持つ確率とすると,

$${R_N}^{(r)}(\alpha)=\sum_{v=2}^{\infty}{P_N}^{(r)}\big(\{exactly\ v\ pairs\ of\ individuals\ in\ \alpha\ each\ share\ a\ parent\}\big)\quad(19)$$

まず初めに ${R_N}^{(r)}(\alpha)$ の評価をする. 明らかに次の不等式が成立する.
但し, ここでは $a_l={N_l}^*, b_l=N_l$とした.

$$\frac{a_l}{b_l}=1+\frac{1}{2N}\frac{Q_l}{2c_l}\leq 2,\quad \frac{a_l-1}{b_l-1}\leq 2\frac{a_l}{b_l}\leq 4,\quad \frac{1}{b_l-1}\leq\frac{1}{b_l-2}\leq\frac{1}{b_l-3}\leq\frac{1}{N}$$

$$|{c_N}^l|\leq 2\frac{{K_2}^*}{N},\ for\ N\geq\max(3,\frac{C}{4})$$

そこで, αに属する 2 つ, もしくはそれ以上のペアが共通の親を持つ事象は次の 2 つに分けて考えられる：
$Case(1)$: 2 つのペアが違う分集団に属していて, 共通の親を持つ事象
$Case(2)$: 2 つのペアが同じ分集団に属していて, 共通の親を持つ事象
明らかに,

$${R_N}^{(r)}(\alpha)\leq P(Case(1))+P(Case(2))$$

まず, $P(Case(1))$ について評価する.

$l_1 \neq l_2$, $\sum_{i=1}^{a_{l_1}} \frac{{\nu_i}^{(r,l_1)}({\nu_i}^{(r,l_1)}-1)}{b_{l_1}(b_{l_1}-1)}$と$\sum_{i=1}^{a_{l_2}} \frac{{\nu_i}^{(r,l_2)}({\nu_i}^{(r,l_2)}-1)}{b_{l_2}(b_{l_2}-1)}$は互いに独立である.よって,

$$P(Case(1)) \leq \binom{V}{2} \frac{\left({2K_2}^*\right)^2}{N^2} \leq \frac{n^4({2K_2}^*)^2}{8N^2} \leq \frac{C_1}{N^2}$$

但し,$V = \sum_{i \in S} \binom{\alpha_i}{2}, C_1 = \frac{n^4({K_2}^*)^2}{2}$

次に $P(Case(2))$ について評価する.

$$P(Case(2)) \leq \sum_{l \in S} ({P_l}^{(1)} + {P_l}^{(2)})$$

但し,

${P_l}^{(1)} = \binom{\alpha_l}{3} \sum_{i=1}^{a_l} \frac{E[{\nu_i}^{(l,r)}({\nu_i}^{(l,r)}-1)({\nu_i}^{(l,r)}-2)]}{b_l(b_l-1)(b_l-2)}$

$${P_l}^{(2)} = \binom{\binom{\alpha_l}{2}}{2} \sum_{i \neq j, 1 \leq i,j \leq a_l} \frac{E[{\nu_i}^{(l,r)}({\nu_i}^{(l,r)}-1){\nu_j}^{(l,r)}({\nu_j}^{(l,r)}-1)]}{b_l(b_l-1)(b_l-2)(b_l-3)}$$

${P_l}^{(1)}$ をまず評価すると,

$${P_l}^{(1)} \leq \frac{{\alpha_l}^3}{6} \frac{a_l E[{\nu_1}^{(l,r)^3}]}{b_l(b_l-1)(b_l-2)} \leq \frac{{\alpha_l}^3}{6} \frac{2{K_3}^*}{N^2}$$

だから,

$$\sum_{l \in S} {P_l}^{(1)} \leq \frac{2n^3{K_k}^*}{6N^2} = \frac{n^3{K_3}^*}{3N^2}$$

次に ${P_l}^{(2)}$ を評価すると,

$${P_l}^{(2)} \leq \frac{a_l(a_l-1)}{b_l(b_l-1)} \binom{\frac{{\alpha_l}^2}{2}}{2} \frac{E[{\nu_1}^{(l,r)^2}{\nu_2}^{(l,r)^2}]}{(b_l-2)(b_l-3)} \leq \frac{{\alpha_l}^4{K_4}^*}{N^2}$$

より,

$$\sum_{l \in S} {P_l}^{(2)} \leq \frac{n^4{K_4}^*}{N^2}$$

これらのことから,

$$P(Case(2)) \leq \Big(\frac{n^3}{3}{K_3}^* + n^4{K_4}^*\Big)\frac{1}{N^2} = \frac{C_2}{N^2}$$

但し, $C_2 = \Big(\frac{n^3}{3}{K_3}^* + n^4{K_4}^*\Big)$ である. 従って,

$${R_N}^{(r)}(\alpha) \leq \frac{C_1 + C_2}{N^2}$$

がわかる. $\beta \notin \{\alpha\} \cup \{\alpha - \epsilon^i; i \in S\}$ のとき, ${R_N}^{(r)}(\alpha, \beta)$ は後ろ向き繁殖においてαからβへ移る確率とする.このとき,

$${R_N}^{(r)}(\alpha) = \sum_{\beta \notin \{\alpha\} \cup \{\alpha - \epsilon^i; i \in S\}} {R_N}^{(r)}(\alpha, \beta)$$

となる. ${R_N}^{(r)}(\alpha, \alpha - \epsilon^i)$ は, 次の式を満たす量として, 定義する.

$${P_N}^{(r)}(\alpha - \epsilon^i | \alpha) = \binom{\alpha_i}{2} {c_N}^i - {R_N}^{(r)}(\alpha, \alpha - \epsilon^i)$$

ここで, ${P_N}^{(r)}(\alpha - \epsilon^i|\alpha)$ は後ろ向き繁殖で α から $\alpha - \epsilon^i$ へ移る確率である.

$$\sum_{i \in S} {P_N}^{(r)}(\alpha - \epsilon^i|\alpha) = {P_N}^{(r)}\Big(\{exactly\ one\ pair\ of\ individual\ in\ \alpha\ each\ share\ a\ parent\}\Big) \tag{20}$$

であり, かつ,

$$\sum_{i \in S} \binom{\alpha_i}{2} {c_N}^i = \sum_{v=1}^{\infty} v{P_N}^{(r)}\Big(\{exactly\ v\ pairs\ of\ individuals\ in\ \alpha\ each\ share\ a\ parent\}\Big) \tag{21}$$

だから, (21) の両辺から（20）の両辺を取り去って,

$$\sum_{i \in S} {R_N}^{(r)}(\alpha, \alpha - \epsilon^i) = \sum_{v=2}^{\infty} v{P_N}^{(r)}\Big(\{exactly\ v\ pairs\ of\ individuals\ in\ \alpha\ each\ share\ a\ parent\}\Big) \tag{22}$$

が成り立つことがわかる. (19) と (22) により, 不等式,

$$2{R_N}^r(\alpha) \leq \sum_{i \in S} {R_N}^{(r)}(\alpha, \alpha - \epsilon^i) \leq \binom{n}{2} {R_N}^{(r)}(\alpha) \tag{23}$$

が成り立つことがわかる. この不等式を用いて,

$$\sum_{\beta \neq \alpha} {R_N}^{(r)}(\alpha, \beta) \leq \Big(\binom{n}{2} + 1\Big){R_N}^{(r)}(\alpha) \tag{24}$$

が成り立つことがわかる.

従って, 後ろ向き繁殖における推移確率 ${P_N}^{(r)}(\beta|\alpha)$ は,
${P_N}^{(r)}(\beta|\alpha) =$

$$\begin{cases} 1-\sum_{i\in S}\binom{\alpha_i}{2}{c_N}^i+\sum_{i\in S}{R_N}^{(r)}(\alpha,\alpha-\epsilon^i)-\sum_{\gamma\notin\{\alpha\}\cup\{\alpha-\epsilon^i:i\in S\}}{R_N}^{(r)}(\alpha,\gamma) & (\beta=\alpha\text{のとき}) \\ \binom{\alpha_i}{2}{c_N}^i-{R_N}^{(r)}(\alpha,\alpha-\epsilon^i) & (\beta=\alpha-\epsilon^i\text{のとき}) \\ {R_N}^{(r)}(\alpha,\beta) & (\text{その他}) \end{cases}$$

となる.

3.2 有限次元分布の収束

後向き移住と後ろ向き繁殖を組み合わせて, 後ろ向き遷移確率は次のようになる.

$$P_N(\beta|\alpha)=\sum_{\gamma}{P_N}^{(r)}(\gamma|\alpha){P_N}^{(m)}(\beta|\gamma)$$

$\mathbb{P}_N, {\mathbb{P}_N}^{(m)}, {\mathbb{P}_N}^{(r)}$を, それぞれ移住と繁殖, 移住, 繁殖に対する 1 ステップの遷移確率行列とし, $\{\alpha^{(N)}(\tau)\}_{\tau\in\mathbb{Z}_+}=\{({\alpha_i}^{(N)}(\tau))_{i\in S}\}_{\tau\in\mathbb{Z}_+}$は初期状態$\alpha^{(N)}(0)=\alpha$をもつ, 遷移確率 $\mathbb{P}_N$に従う E 上の離散時刻マルコフ連鎖とする.このとき, $\{{\alpha_i}^{(N)}([2Nt])\}$ の有限次元分布が, 初期状態$\alpha(0)=\alpha$である地理的構造を持つ合祖過程 $\{\alpha(t)\}_{t\geq 0}$の有限次元分布に収束することを示す. 1 ステップの後ろ向き遷移確率を行列の形で表せば,

$$\mathbb{P}_N={\mathbb{P}_N}^{(r)}{\mathbb{P}_N}^{(m)}$$

となる. $\mathbb{I}$ を

$$\mathbb{I}_{\alpha,\beta}=\delta_{\alpha,\beta}=\begin{cases}1 & (\beta=\alpha\text{のとき}) \\ 0 & (\text{その他})\end{cases}$$

とすると,

$${\mathbb{P}_N}^{(m)}=\mathbb{I}+\frac{{\mathbb{Q}_N}^{(m)}}{2N}+{\mathbb{R}_N}^{(m)}\qquad {\mathbb{P}_N}^{(r)}=\mathbb{I}+\frac{{\mathbb{Q}_N}^{(r)}}{2N}+{\mathbb{R}_N}^{(r)}$$

但し,

$$({\mathbb{R}_N}^{(m)})_{\alpha,\beta} = \begin{cases} -\sum_{\gamma\neq\alpha} {R_N}^{(m)}(\alpha,\gamma) & (\beta=\alpha\text{のとき}) \\ {R_N}^{(m)}(\alpha,\beta) & (\text{その他}) \end{cases}$$

$$({\mathbb{R}_N}^{(r)})_{\alpha,\beta} = \begin{cases} \sum_{i\in S} {R_N}^{(r)}(\alpha,\alpha-\epsilon^i) - \sum_{\beta\notin\{\alpha\}\cup\{\alpha-\epsilon^i\}} {R_N}^{(r)}(\alpha,\beta) & (\beta=\alpha\text{のとき}) \\ -{R_N}^{(r)}(\alpha,\alpha-\epsilon^i) & (\beta=\alpha-\epsilon^i\text{のとき}) \\ {R_N}^{(r)}(\alpha,\beta) & (\text{その他}) \end{cases}$$

$({\mathbb{Q}_N}^{(m)})_{\alpha,\beta} =$

$$\begin{cases} -\sum_{i\in S}\alpha_i(2Nm_i)\frac{{N_i}^*}{{N_i}^*-m_i{N_i}^*-\alpha_i+1}\prod_{k\in S}\prod_{a=0,\cdots,\alpha_k-1}\frac{{N_k}^*-m_k{N_k}^*-a}{{N_k}^*-a} & (\beta=\alpha\text{のとき}) \\ \alpha_i(2Nm_{i,j})\frac{{N_i}^*}{{N_i}^*-m_i{N_i}^*-\alpha_i+1}\prod_{k\in S}\prod_{a=0,\cdots,\alpha_k-1}\frac{{N_k}^*-m_k{N_k}^*-a}{{N_k}^*-a} & (\beta=\alpha-\epsilon^i+\epsilon^j\ (j\neq i)\text{のとき}) \\ 0 & (\text{その他}) \end{cases}$$

$$({\mathbb{Q}_N}^{(r)})_{\alpha,\beta} = \begin{cases} -\sum_{i\in S}\binom{\alpha_i}{2}\Big(\Big(1+\frac{Q_i}{4c_iN}\Big)\frac{E[\{{\nu_1}^{(i,r)}\}^2]}{c_i-\frac{1}{2N}}-\frac{1}{c_i-\frac{1}{2N}}\Big) & (\beta=\alpha\text{のとき}) \\ \binom{\alpha_i}{2}\Big(\Big(1+\frac{Q_i}{4c_iN}\Big)\frac{E[\{{\nu_1}^{(i,r)}\}^2]}{c_i-\frac{1}{2N}}-\frac{1}{c_i-\frac{1}{2N}}\Big) & (\beta=\alpha-\epsilon^i\text{のとき}) \\ 0 & (\text{その他}) \end{cases}$$

特に, ${\mathbb{Q}_N}^{(r)}$ については,

$${c_N}^l = \frac{\sum_{i=1}^{{N_l}^*}E[{\nu_i}^{(l,r)}({\nu_i}^{(l,r)}-1)]}{N_l(N_l-1)} = \frac{1}{2N}\Big(\Big(1+\frac{Q_l}{4c_lN}\Big)\frac{E[\{{\nu_1}^{(l,r)}\}^2]}{c_l-\frac{1}{2N}}-\frac{1}{c_l-\frac{1}{2N}}\Big)$$

であることを用いた.

今, 分集団の数は可算無限個存在するから, この行列に次のようなノルムを与える.

$$||\mathbb{A}|| = \sup_{\alpha} \sum_{\beta} |\mathbb{A}_{\alpha,\beta}|$$

${\mathbb{Q}_N}^{(m)}$を評価すれば,

$$\sum_{i\in S} \alpha_i m_i \frac{{N_i}^*}{{N_i}^* - m_i {N_i}^* - \alpha_i + 1} \prod_{k\in S} \prod_{a=0,\cdots,\alpha_k-1} \frac{{N_k}^* - m_k {N_k}^* - a}{{N_k}^* - a}$$

$$= \sum_{i\in S} \alpha_i m_i \prod_{l=0}^{\alpha_i-2} \frac{{N_i}^* - m_i {N_i}^* - l}{{N_i}^* - 1 - l} \prod_{k\in S, k\neq i} \prod_{\alpha=0}^{\alpha_k-1} \frac{{N_k}^* - m_k {N_k}^* - a}{{N_k}^* - a} \leq \frac{nM}{4N}$$

$$||{\mathbb{Q}_N}^{(m)}|| \leq nM \tag{25}$$

${\mathbb{Q}_N}^{(r)}, {\mathbb{R}_N}^{(m)}, {\mathbb{R}_N}^{(r)}$ を評価すれば,

$$||{\mathbb{Q}_N}^{(r)}|| \leq 2 \sum_{i\in S} \binom{\alpha_i}{2} \left(\left(1 + \frac{Q_i}{4c_i N}\right) \frac{E[\{{\nu_1}^{(i,r)}\}^2]}{c_i - \frac{1}{2N}} - \frac{1}{c_i - \frac{1}{2N}} \right)$$

$$\leq 2 \sum_{i\in S} \binom{\alpha_i}{2} \left(\left(1 + \frac{C}{4N}\right) \frac{E[\{{\nu_1}^{(i,r)}\}^2]}{1 - \frac{1}{2N}} \right) \leq 4 \binom{n}{2} (1 + C) {K_2}^* < \infty \tag{26}$$

$$||{\mathbb{R}_N}^{(m)}|| \leq \frac{M^2 n^2}{8N^2}, \quad ||{\mathbb{R}_N}^{(r)}|| \leq \frac{n^2 (C_1 + C_2)}{N^2} \tag{27}$$

但し, 最後の式に関しては次の 2 つの関係式を用いた.

$$\sum_{i\in S} {R_N}^{(r)}(\alpha, \alpha - \epsilon^i) \leq \binom{n}{2} {R_N}^{(r)}(\alpha) \leq \binom{n}{2} \frac{(C_1 + C_2)}{N^2}$$

$$\sum_{\beta\neq\alpha} {R_N}^{(r)}(\alpha, \beta) \leq \left(\binom{n}{2} + 1 \right) {R_N}^{(r)}(\alpha) \leq \left(\binom{n}{2} + 1 \right) \frac{(C_1 + C_2)}{N^2}$$

以上より,

$$\mathbb{P}_N = \mathbb{I} + \frac{\mathbb{Q}_N + \pi_N}{2N} \tag{28}$$

が成り立つ. 但し,

$$\mathbb{Q}_N = {\mathbb{Q}_N}^{(m)} + {\mathbb{Q}_N}^{(r)}$$

$$\pi_N = 2N \left({\mathbb{R}_N}^{(r)} + {\mathbb{R}_N}^{(m)} + {\mathbb{R}_N}^{(r)} {\mathbb{R}_N}^{(m)} + \frac{{\mathbb{Q}_N}^{(r)} {\mathbb{R}_N}^{(m)}}{2N} + \frac{{\mathbb{R}_N}^{(r)} {\mathbb{Q}}^{(m)}}{2N} + \frac{{\mathbb{Q}_N}^{(m)} {\mathbb{Q}}^{(r)}}{4N^2} \right) \tag{29}$$

であるから,

$$\lim_{N\to\infty} \mathbb{Q}_N = \mathbb{Q}$$

即ち,

$$\text{任意の}\alpha,\beta \in E \text{ に対して } \lim_{N\to\infty} (\mathbb{Q}_N)_{\alpha,\beta} = \mathbb{Q}_{\alpha,\beta}$$

が成り立つことがわかる. 但し,

$$\mathbb{Q}_{\alpha,\beta} = \begin{cases} -\sum_{i\in S}\left(\alpha_i \frac{M_i}{2} + \frac{\sigma^2\alpha_i(\alpha_i-1)}{2c_i}\right) & (\beta = \alpha\text{のとき}) \\ \alpha_i \frac{M_{i,j}}{2} & (\beta = \alpha - \epsilon^i + \epsilon^j \ (i\neq j)\ \text{のとき}) \\ \frac{\sigma^2\alpha_i(\alpha_i-1)}{2c_i} & (\beta = \alpha - \epsilon^i\text{のとき}) \\ 0 & (\text{その他}) \end{cases}$$

有限次元分布の収束を示すために次の式が成り立つことを証明する；

$$\lim_{N\to\infty} {\mathbb{P}_N}^{[2Nt]} = e^{t\mathbb{Q}}$$

まず, $||\mathbb{Q}|| \leq nM + 2\sigma^2 \binom{n}{2} < \infty$ より, $e^{t\mathbb{Q}} = \sum_{v=0}^{\infty} \frac{t^v\mathbb{Q}^v}{v!}$ が存在する. なぜなら成分ごとにも,

$$\text{任意の}\alpha,\beta \in E \text{ に対し}, |(e^{t\mathbb{Q}})_{\alpha,\beta}| \leq \sum_{v=0}^{\infty} \frac{t^v|(\mathbb{Q}^v)_{\alpha,\beta}|}{v!} \leq \sum_{v=0}^{\infty} \frac{t^v||\mathbb{Q}||^v}{v!} = e^{t||\mathbb{Q}||} < \infty$$

が成立するからである. よって,

$${\mathbb{P}_N}^{[2Nt]} = \left(\mathbb{I} + \frac{\mathbb{Q}_N + \pi_N}{2N}\right)^{[2Nt]} = \sum_{v=0}^{[2Nt]} \binom{[2Nt]}{v} (\tfrac{1}{2N})^v (\mathbb{Q}_N + \pi_N)^v$$

$$= \sum_{v=0}^{[2Nt]} \frac{[2Nt]([2Nt]-1)\cdots([2Nt]-v+1)}{(2N)^v} \frac{(\mathbb{Q}_N+\pi_N)^v}{v!}$$

この成分については,

$$({\mathbb{P}_N}^{[2Nt]})_{\alpha,\beta} = \sum_{v=0}^{\infty} a_{v,N} \tag{30}$$

但し,

$$a_{v,N} = I_{\{v\leq[2Nt]\}} \frac{[2Nt]([2Nt]-1)\cdots([2Nt]-v+1)}{(2N)^v} \frac{(\mathbb{Q}_N+\pi_N)^v_{\alpha,\beta}}{v!} \tag{31}$$

$$I_{\{v\leq[2Nt]\}} = \begin{cases} 1 & (v\leq[2Nt] \text{ のとき}) \\ 0 & (\text{その他}) \end{cases}$$

(25),(26) によって,

$$||\mathbb{Q}_N|| \leq ||{\mathbb{Q}_N}^{(m)}|| + ||{\mathbb{Q}_N}^{(r)}|| \leq C^* < \infty \tag{32}$$

但し,

$$C^* := nM + 4\binom{n}{2}(1+C){K_2}^* \tag{33}$$

(25),(26),(27),(29) によって,

$$||\pi_N|| \leq \frac{k_1}{N}, \quad ||\mathbb{A}_N|| \leq \frac{k_2}{N} \tag{34}$$

但し, この $\mathbb{A}_N$ とは, 次式を満たす行列のことである:

$$(\mathbb{Q}_N+\pi_N)^v = {\mathbb{Q}_N}^v + \mathbb{A}_N$$

また,k_1, k_2は N に依存しない定数である. 今, 行列 $\mathbb{V}$ を,

$$(\mathbb{V})_{\alpha,\beta} = \begin{cases} nK\sum_{j\neq i}{q_{i,j}}^* + 2\binom{\alpha_i}{2}\{(1+C){K_2}^*\} & (\beta=\alpha\text{のとき}) \\ nK{q_{i,j}}^* & (\beta=\alpha-\epsilon^i+\epsilon^j\ (i\neq j)\text{ のとき}) \\ 2\binom{\alpha_i}{2}\{(1+C){K_2}^*\} & (\beta=\alpha-\epsilon^i\text{のとき}) \\ 0 & (\text{その他}) \end{cases}$$

と定義すると,

$$||\mathbb{V}|| \leq 2nK\sup_{i\in S}\sum_{j\neq i}{q_{i,j}}^* + 4\binom{n}{2}(1+C){K_2}^* < \infty$$

が成立し, しかも, 成分ごとには,

$$|(\mathbb{Q}_N)_{\alpha,\beta}| \leq (\mathbb{V})_{\alpha,\beta}$$

が成立していることがわかる. よって, 有界収束定理から,

$$\lim_{N\to\infty}({\mathbb{Q}_N}^v)_{\alpha,\beta} = \lim_{N\to\infty}\sum_{\gamma_1,\gamma_2,\cdots,\gamma_{v-1}}(\mathbb{Q}_N)_{\alpha,\gamma_1}\cdot(\mathbb{Q}_N)_{\gamma_1,\gamma_2}\cdots(\mathbb{Q}_N)_{\gamma_{v-1},\beta}$$

$$= \sum_{\gamma_1,\gamma_2,\cdots,\gamma_{v-1}}(\mathbb{Q})_{\alpha,\gamma_1}\cdot(\mathbb{Q})_{\gamma_1,\gamma_2}\cdots(\mathbb{Q})_{\gamma_{v-1},\beta} = (\mathbb{Q}^v)_{\alpha,\beta}$$

以上のことから,

$$\lim_{N\to\infty}(\mathbb{Q}_N+\pi_N)^v = \mathbb{Q}^v$$

任意の $v\in\mathbb{Z}+$ に対して,

$$\lim_{N\to\infty}a_{v,N} = \frac{t^v(\mathbb{Q}^v)_{\alpha,\beta}}{v!}\quad \alpha,\beta\in S \tag{35}$$

十分 大きな全ての自然数 N に対して, (31), (32), (34) から,

$$|a_{v,N}| \le \frac{t^v||\mathbb{Q}_N+\pi_N||^v}{v!} \le \frac{t^v(C^*+1)^v}{v!} \qquad \alpha,\beta\in S \tag{36}$$

となる. このことから,

$$\sum_{v=0}^{\infty}\frac{t^v(C^*+1)^v}{v!} = e^{t(C^*+1)} < \infty$$

となることがわかる. よって (30), (31) と (35), (36) を用いて, 有界収束定理から,

$$\text{任意の}\alpha,\beta\in E\text{ に対して、}\lim_{N\to\infty}({\mathbb{P}_N}^{[2Nt]})_{\alpha,\beta} = \sum_{v=0}^{\infty}\frac{t^v(\mathbb{Q}^v)_{\alpha,\beta}}{v!} = (e^{t\mathbb{Q}})_{\alpha,\beta}$$

このことを用いて, 有限次元分布の収束を示す.

$$\begin{aligned}
&P\{\alpha^{(N)}([2Nt_1]) = x_1\ ,\ \cdots\ ,\ \alpha^{(N)}([2Nt_N]) = x_N\}\\
&= ({\mathbb{P}_N}^{[2Nt_1]})_{\alpha,x_1}({\mathbb{P}_N}^{[2Nt_2]-[2Nt_1]})_{x_1,x_2}\cdots({\mathbb{P}_N}^{[2Nt_n]-[2Nt_{n-1}]})_{x_{n-1},x_n}\\
&\to (e^{t_1\mathbb{Q}})_{\alpha,x_1}(e^{(t_2-t_1)\mathbb{Q}})_{x_1,x_2}\cdots(e^{(t_n-t_{n-1})\mathbb{Q}})_{x_{n-1},x_n},\quad N\to\infty
\end{aligned}$$

即ち,

$$\lim_{N\to\infty}P\{\alpha^{(N)}([2Nt_1]) = x_1\ ,\ \cdots\ ,\ \alpha^{(N)}([2Nt_N]) = x_N\} = P\{\alpha(t_1) = x_1\ ,\cdots\ ,\ \alpha(t_N) = x_N\}$$

が示された.よって,E が可算集合だから, このことから, $\{\alpha^{(N)}([2Nt]);t\ge 0\}$ の有限次元分布は合祖過程 $\{\alpha(t);t\ge 0\}$ の有限次元分布に収束する.

3.3 地理的構造を持つ合祖過程の弱収束

相対コンパクト性

E を $\mathbb{R}^S$の部分空間とみなすとき（但し, $\mathbb{R}$ は実数全体.), 次のノルムで距離空間 (E, d) を定義する.

$$||\mathbb{X}|| = \sup_{i \in S} |x_i| \quad (\mathbb{X} = (x_i)_{i \in S} \in \mathbb{R}^S) \tag{37}$$

このノルムの下で E は可分かつ完備である. Ethier and Kurtz(1986) の Chapter 3, Corollary 7,4 を用いて $\{\alpha^{(N)}([2Nt])\}$ の相対コンパクト性を示すために, 次の 2 条件を確認する：

(a) 任意の $\eta > 0$ と $t \geq 0$ に対して,

$$\liminf_{N \to \infty} P\{\alpha^{(N)}([2Nt]) \in \Gamma_{\eta,t}\} \geq 1 - \eta \tag{38}$$

となるようなコンパクト集合 $\Gamma_{\eta,t} \subset E$ が存在する.

(b) 任意の $\eta > 0$ と $T \geq 0$ に対して,

$$\limsup_{N \to \infty} P\{\omega'(\alpha^{(N)}([2Nt])\ ,\ \delta\ ,\ T) \geq \eta\} \leq \eta \tag{39}$$

となるような$\delta > 0$ が存在する.

但し, $\omega'(\alpha^{(N)}([2Nt])\ ,\ \delta\ ,\ T) = \inf_{\{t_i\}} \max_i \sup_{s,t \in [t_{i-1}, t_i)} ||\alpha^{(N)}([2Ns]) - \alpha^{(N)}([2Nt])||$ である.

また, 時刻の列 $\{t_i\}$ に対しては次なる条件を満たすようにとった：

(i) $0 = t_0 < t_1 < \cdots < t_{k-1} < T \leq t_k$

(ii) $\min_i (t_i - t_{i-1}) > \delta$

まず (a) を証明する. t を固定する. この時, $S_m \subset S, S_m$; 有限集合, $S_1 \subset S_2 \subset \cdots \subset S_m \subset S_{m+1} \subset \cdots, \cup_{m=1}^{\infty} S_m = S$ とし, $\Gamma_m = \{\alpha \in E; \alpha_i = 0\ if\ i \notin S_m\}$（有限集合）とおけば $\Gamma_1 \subset \Gamma_2 \subset \cdots, \cup_{m=1}^{\infty} \Gamma_m = E$ である.この時,

$$\lim_{n \to \infty} P\{\alpha(t) \in \Gamma_n\} = P\{\alpha(t) \in \cup_{n=1}^{\infty} \Gamma_n\} = P\{\alpha(t) \in E\} = 1$$

となる. また, 有限次元分布の収束により,

$$P\{\alpha(t) \in \Gamma_n\} = \lim_{N \to \infty} P\{\alpha^{(N)}([2Nt]) \in \Gamma_n\}$$

これらのことから, (a) は明らかである. (b) の証明に移る. p_N を

$$N \in \mathbb{Z}_+\ , \qquad p_N = \frac{C^* + 1}{2N} \tag{40}$$

とおく.ここで, N を十分大きくとると, $p_N < 1$, また, そのような各 N に対して, 離散時間マルコフ連鎖 $(Z_N\ ,\ \xi_N) = \{(Z_N(\tau)\ ,\ \xi_N(\tau))\ ;\ \tau = 0, 1, 2, \cdots\}$ は状態空間 $(\mathbb{Z}_+/\{0\}) \times E$ をもち, その

遷移確率：

$$P\{(Z_N(\tau+1)\ ,\ \xi_N(\tau+1))=(j,\beta)|(Z_N(\tau)\ ,\ \xi_N(\tau))=(i,\alpha)\}=$$

$$\begin{cases} 1-p_N & (j=i\ and\ \beta=\alpha\text{のとき}) \\ p_N-\sum_{\gamma\in E;\gamma\neq\alpha}P_N(\gamma|\alpha) & (j=i+1\ and\ \beta=\alpha\text{のとき}) \\ P_N(\beta|\alpha) & (j=i+1\ and\ \beta\neq\alpha\text{のとき}) \\ 0 & (\text{その他}) \end{cases}$$

但し, $P_N(\beta|\alpha)$ は祖先過程 $\{\alpha^{(N)}(\tau)\}_{\tau\in\mathbb{Z}_+}$ に対するαからβへの遷移確率である. (29), (32), (34), (40) から, 任意の$\alpha\in E$ に対して,

$$\sum_{\gamma\in E;\gamma\neq\alpha}P_N(\gamma|\alpha)=\frac{\sum_{\gamma\neq\alpha}(\mathbb{Q}_N+\pi_N)_{\alpha,\gamma}}{2N}\leq p_N$$

であることから, 上記の遷移確率行列 の要素は全て正である. 他, 明らかなことだが, 全ての $\eta\ ,\ T>0$ に対して,

$$P\{\omega'(\alpha_N([2N*])\ ,\ \delta\ ,\ T)\geq\eta\}=P\{\omega'(\xi_N([2N*])\ ,\ \delta\ ,\ T)\geq\eta\} \tag{41}$$

である. 次に過程 $(Z_N\ ,\ \xi_N)$ の 飛躍時間 (jump time)の列を $0=\rho_0<\rho_1<\cdots$ となるように構成し, $\tau_i=\rho_i-\rho_{i-1}\ ,\ i\in\mathbb{Z}_+$ (inter-jump times) とする. τ_i は各 i について互いに独立で, それぞれ平均 $\frac{1}{p_N}$ の幾何分布に従うものである (独立同分布). 過程 $(Z_N\ ,\ \xi_N)$ のジャンプする確率は各世代 p_N である. 今, 固定された$\eta>0$ と $T>0$ を仮定する.次の集合の大小関係を証明する ; $\exists J\in\mathbb{Z}_+$ と$\delta>0$ に対して,

$$\{\omega'(\xi_N([2N*])\ ,\ \delta\ ,\ T)<\eta\}\supset\{\rho_J\geq 2NT\quad\text{かつ}\quad\tau_i>2N\delta\ ,\ i=1,2,\cdots,J\} \tag{42}$$

証明するにあたって, まず右の集合を吟味しよう. $k_N=min\{i\ :\ \rho_i\geq 2NT\}$ とする. 但し $1\leq k_N\leq J$ で, 分割 $t_i=\frac{\rho_i}{2N}\quad(i=0,1,\cdots,k_N)$ は $0=t_0<t_1<\cdots<t_{k_{N-1}}<T\leq t_{k_N}$, かつ, $t_i-t_{i-1}>\delta\quad(i=1,\cdots,k_N)$ を満たす分割である.この時, 過程 $(Z_N(\tau)\ ,\ \xi_N(\tau))$ は時刻 $\rho_{i-1}\leq\tau\leq\rho_i$の間で定数の値をとる. 即ち,

$$\omega'(\xi_N([2N*])\ ,\ \delta\ ,\ T)=0$$

よって, 集合の包含関係 (42) が証明された. 直ちに,

$$P\{\omega'(\xi_N([2N*])\ ,\ \delta\ ,\ T)<\eta\}\geq P\{\rho_J\geq 2NT\quad\text{かつ}\quad\tau_i>2N\delta\ ,\ i=1,2,\cdots,J\}$$

がわかる. (b) を示すためには, 以下の式を示せばよい：

$$\liminf_{N\to\infty} P\{\rho_J \geq 2NT \quad \text{かつ} \quad \tau_i > 2N\delta\ ,\ i=1,2,\cdots,J\} \geq 1-\eta$$

実際,

$$\begin{aligned}
&P\{\rho_J \geq 2NT \quad \text{かつ} \quad \tau_i > 2N\delta\ ,\ i=1,2,\cdots,J\} \\
&= P\{\rho_J \geq 2NT | \tau_i > 2N\delta\ ,\ i=1,2,\cdots,J\} P\{\tau_i > 2N\delta\ ,\ i=1,2,\cdots,J\} \\
&= P\{\rho_J \geq 2NT | \tau_i > 2N\delta\ ,\ i=1,2,\cdots,J\} (P\{\tau_i > 2N\delta\})^J
\end{aligned}$$

$\rho_J = \sum_{i=1}^{J} \tau_i$ (J 回目の飛躍時刻) であるから, このことから,

$$P\{\rho_J \geq 2NT | \tau_i > 2N\delta\ ,\ i=1,2,\cdots,J\} \geq P\{\rho_J \geq 2NT\} \tag{43}$$

がわかる. 実際, 各 i について $P\{\tau_i = k\} = p_N(1-p_N)^{k-1}$, $k \geq 1$ だから,

$$\begin{aligned}
&P\{\rho_J \geq 2NT\} = P\{\tau_1 + \tau_2 + \cdots + \tau_J \geq 2NT\} \\
&= {p_N}^J \sum_{l_i \geq 1\ ,\ i=1,\cdots,J.\ and\ \sum_{m=1}^{J} l_m \geq 2NT} (1-p_N)^{(\sum_{m=1}^{J} l_m - J)}
\end{aligned} \tag{44}$$

これで右辺が変形できた. 今度は左辺を以下のように変形する.

$$\begin{aligned}
&P\{\rho_J \geq 2NT \quad \text{かつ} \quad \tau_i > 2N\delta\ ,\ i=1,2,\cdots,J\} \\
&= P\{\tau_1 + \tau_2 + \cdots + \tau_J \geq 2NT\ ,\ \tau_i > 2N\delta\ ,\ i=1,2,\cdots,J\} \\
&= {p_N}^J \sum_{l_i \geq 2N\delta\ ,\ i=1,\cdots,J.\ and\ \sum_{m=1}^{J} l_m \geq 2NT} (1-p_N)^{(\sum_{m=1}^{J} l_m - J)}
\end{aligned}$$

$$P\{\tau_i > 2N\delta\ ,\ i=1,2,\cdots,J\} = (P\{\tau_i > 2N\delta\})^J = (1-p_N)^{J(2N\delta-1)}$$

これらから左辺の条件付確率を計算すれば,

$$\begin{aligned}
&P\{\rho_J \geq 2NT | \tau_i > 2N\delta\ ,\ i=1,2,\cdots,J\} \\
&= {p_N}^J \sum_{l_i \geq 2N\delta\ ,\ i=1,\cdots,J.\ and\ \sum_{m=1}^{J} l_m \geq 2NT} (1-p_N)^{(\sum_{m=1}^{J} l_m - J)} / (1-p_N)^{J(2N\delta-1)} \\
&= {p_N}^J \sum_{l_i \geq 2N\delta\ ,\ i=1,\cdots,J.\ and\ \sum_{m=1}^{J} l_m \geq 2NT} (1-p_N)^{(\sum_{m=1}^{J} l_m - 2N\delta J)}
\end{aligned}$$

$m_i = l_i - 2N\delta + 1$, $i=1,2,\cdots,J$ とおくと,

$$= {p_N}^J \sum_{\sum_{k=1}^{J} m_k \geq 2NT - 2N\delta J + J\ and\ m_i \geq 1\ ,\ i=1,2,\cdots,J} (1-p_N)^{\sum_{k=1}^{J} m_k - J} \tag{45}$$

求める不等式 (43) が成立することがこれらの計算によって明らかとなった. 但し, ここでは,$2N\delta \leq 1$ としてよいから, $-2N\delta J + J \geq 0$ であることを用いた.

また,

$$P\{\rho_J \geq 2NT\} = P\{Z_N([2NT]) - Z_N(0) < J\}$$

であるから, 以上のことから,

$$P\{\omega'(\alpha_N([2N*])\ ,\ \delta\ ,\ T) \geq \eta\} \leq P\{Z_N([2NT]) - Z_N(0) < J\}(P\{\frac{\tau_i}{2N} > \delta\})^J \tag{46}$$

ここで, τ_iは幾何分布なので, N に関して極限をとると, $\frac{\tau_i}{2N}$は平均 $C^* + 1$ に従う指数分布となる (確率変数を X とする).また, $Z_N([2NT]) - Z_N(0)$ が二項分布 $B([2NT], p_N)$ に従うので, 極限は平均 $T(C^* + 1)$ に従うポアソン分布となる (確率変数を Z とする). (46) から, すぐに

$$\liminf_{N\to\infty} P\{\omega'(\alpha_N([2N*])\ ,\ \delta\ ,\ T) \geq \eta\} \geq P\{Z < J\}(P\{X > \delta\})^J \tag{47}$$

が導き出される. (47) の右辺について $J \to \infty$, $\delta \to 0$ とすると, 1 に収束する.よって (b) が証明された. 最後に次のことについて示す.

‘‘N に関して極限をとれば, 祖先過程 $\{\alpha^{(N)}([2Nt]) :\ t \geq 0\}$ は空間 $D_E[0, \infty)$ の中で生成作用素 $\mathbb{Q}$ に従う, 地理的構造を持つ合祖過程 $\{\alpha(t) :\ t \geq 0\}$ に弱収束する.’’

証明については, Ethier and Kurtz(1986) の Chapter 3 の Theorem 7.8 の (b) を用いる.
$\{\alpha^{(N)}([2Nt])\}_{N\in\mathbb{Z}^+}$は相対コンパクトであり, かつ有限次元分布が収束しているから $\alpha^{(N)}([2Nt])$ は$\alpha(t)$ に弱収束する.

4　地理的構造を持つ遺伝子系図に関する種々の結果

4.1　共通祖先に到達するまでの時間の分布

D：分集団の数, $N_1, N_2, N_3, \cdots, N_D$をそれぞれ $1, \cdots, D$ でラベルされた分集団のサイズとする. c_iを各分集団の集団サイズを決定する比例定数とし, $N_i = 2c_iN$ が成立するものとする.今祖先の数における地理的配置を表すベクトルを$\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_D)$ とするとき, 前章の結果から $\alpha^{(N)}([2Nt])$ は N に関して極限をとった時, 生成作用素,

$$\mathbb{Q}_{\alpha,\beta} = \begin{cases} -\sum_{i\in S}\left(\alpha_i\dfrac{M_i}{2} + \dfrac{\sigma^2\alpha_i(\alpha_i-1)}{2c_i}\right) & (\beta = \alpha\text{のとき}) \\ \alpha_i\dfrac{M_{i,j}}{2} & (\beta = \alpha - \epsilon^i + \epsilon^j\ (i \neq j)\ \text{のとき}) \\ \dfrac{\sigma^2\alpha_i(\alpha_i-1)}{2c_i} & (\beta = \alpha - \epsilon^i\text{のとき}) \\ 0 & (\text{その他}) \end{cases} \tag{48}$$

に従うマルコフ過程$\alpha(t)$ に弱収束することがわかった.この一般的な形の生成作用素に対して, 次の定理が成り立つ：

定理：T の母関数 (ラプラス変換) の方程式 (Notohara(2000))
サンプル遺伝子が 1 つの共通祖先に到達するまでの時間の長さを $T = \inf\{t; |\alpha(t)| = 1\}$, αでの滞在時間を$\tau(\alpha)$, $f(\alpha) = E[e^{-\lambda T}|\alpha(0) = \alpha]$ と置く.この時, 以下の式が成り立つ.

$$\sum_{\beta}\mathbb{Q}_{\alpha,\beta}f(\beta) = \lambda f(\alpha) \tag{49}$$

但し, 全ての k に対して, $f(\epsilon^k) = 1$ である.
証明：

$$\begin{aligned} f(\alpha) &= E[e^{-\lambda T}|\alpha(0) = \alpha] = E[e^{-\lambda(T-\tau(\alpha))-\lambda\tau(\alpha)}|\alpha] \\ &= E[E[e^{-\lambda\tau(\alpha)}e^{-\lambda(T-\tau(\alpha))}|F_{\tau(\alpha)}]|\alpha] = E[e^{-\lambda\tau(\alpha)}E[e^{-\lambda(T-\tau(\alpha))}|F_{\tau(\alpha)}]|\alpha] \\ &= E[e^{-\lambda\tau(\alpha)}E[e^{-\lambda(T-\tau(\alpha))}|\alpha(\tau(\alpha))]|\alpha](\text{強マルコフ性}) \\ &= E[e^{-\lambda\tau(\alpha)}|\alpha]\sum_{\beta\neq\alpha}\frac{\mathbb{Q}_{\alpha,\beta}}{|\mathbb{Q}_{\alpha,\alpha}|}f(\beta) = \frac{1}{\lambda - \mathbb{Q}_{\alpha,\alpha}}\sum_{\beta\neq\alpha}\mathbb{Q}_{\alpha,\beta}f(\beta) \end{aligned}$$

境界条件については直ちに導出できる.(証完)

これらの結果を用いて,サンプル数が 2 の場合に応用する. T を合祖するまでの時刻,τを状態の遷移を起こすまでの滞在時間とすると,$T = \tau + T(\theta_\tau w)$ が成り立つ. 但し,θ_tは時刻に関する遷移作用素である.今考えているのはたった 2 つのサンプルであるから,$\alpha = 2\epsilon^i$とするとき $E_\alpha(T)$ を $E({T^i}_w)$ と書き,$\alpha = \epsilon^i + \epsilon^j$とするとき,$E_\alpha(T)$ を $E({T^{i,j}}_b)$ と書くことにする. ${T^i}_w$を分集団 i における合祖するまでの時間. w は $within$ の意味である. ${T^{i,j}}_b$は分集団 i と分集団 j にある 2 つのサンプルが合祖するまでの時間を表す. b は $between$ の意味である. $|\alpha|$ を祖先の総数とするとき,T は $T = \inf\{t > 0; |\alpha| = 1\}$ と書くことができる.

$(i)\alpha = 2\epsilon^i$の時

$$\mathbb{Q}_{\alpha,\beta} = \begin{cases} -\left(M_i + \dfrac{\sigma^2}{c_i}\right) & (\beta = \alpha\text{のとき}) \\ M_{i,j} & (\beta = \alpha - \epsilon^i + \epsilon^j \ (i \neq j) \text{ のとき}) \\ \dfrac{\sigma^2}{c_i} & (\beta = \alpha - \epsilon^i\text{のとき}) \\ 0 & (\text{その他}) \end{cases} \tag{50}$$

$(ii)\alpha = \epsilon^i + \epsilon^j$の時

$$\mathbb{Q}_{\alpha,\beta} = \begin{cases} -\left(\dfrac{M_i}{2} + \dfrac{M_j}{2}\right) & (\beta = \alpha\text{のとき}) \\ \dfrac{M_{i,k}}{2} & (\beta = \alpha - \epsilon^i + \epsilon^k \ (i \neq k) \text{ のとき}) \\ \dfrac{M_{j,k}}{2} & (\beta = \alpha - \epsilon^j + \epsilon^k \ (k \neq j) \text{ のとき}) \\ 0 & (\text{その他}) \end{cases} \tag{51}$$

先程証明した，T の母関数の方程式を用いて, Bahlo and Griffiths(2000)のラプラス変換の式を導く. 但し,$\sigma^2 = 1$ とする. 生成作用素（(50), (51) から,

$$f(\epsilon^i + \epsilon^j) = \frac{1}{\lambda + \frac{M_i}{2} + \frac{M_j}{2}} (\mathbb{Q}_{\epsilon^i+\epsilon^j, 2\epsilon^i} f(2\epsilon^i) + \mathbb{Q}_{\epsilon^i+\epsilon^j, 2\epsilon^j} f(2\epsilon^j)$$

$$+ \sum_{k \neq j,i} \mathbb{Q}_{\epsilon^i+\epsilon^j, \epsilon^i+\epsilon^k} f(\epsilon^i + \epsilon^k) + \sum_{l \neq j,i} \mathbb{Q}_{\epsilon^i+\epsilon^j, \epsilon^l+\epsilon^j} f(\epsilon^l + \epsilon^j))$$

$$= \frac{1}{\lambda + \frac{M_i}{2} + \frac{M_j}{2}} (M_{j,i} f(2\epsilon^i) + M_{i,j} f(2\epsilon^j) + \sum_{k \neq j} M_{j,k} f(\epsilon^i + \epsilon^k) + \sum_{l \neq i} M_{i,l} f(\epsilon^l + \epsilon^j))$$

$$= \frac{1}{\lambda + \frac{M_i}{2} + \frac{M_j}{2}} (\sum_{k \neq j} M_{j,k} f(\epsilon^i + \epsilon^k) + \sum_{l \neq i} M_{i,l} f(\epsilon^l + \epsilon^j))$$

$$= \frac{\frac{M_i}{2} + \frac{M_j}{2}}{\lambda + \frac{M_i}{2} + \frac{M_j}{2}} \Big(\frac{1}{\frac{M_i}{2} + \frac{M_j}{2}} \sum_{k \neq j} M_{j,k} f(\epsilon^i + \epsilon^k) + \sum_{l \neq i} M_{i,l} f(\epsilon^l + \epsilon^j) \Big)$$

$$= \Big(1 + \frac{2\lambda}{M_i + M_j}\Big)^{-1} 2 \Big(\sum_{l \neq i} \frac{M_{i,l} f_{l,j}(s)}{M_i + M_j} + \sum_{l \neq j} \frac{M_{j,l} f_{l,i}(s)}{M_i + M_j} \Big) \tag{52}$$

$$f(2\epsilon^i) = \frac{1}{\lambda + M_i + \frac{1}{c_i}} \Big(2 \sum_{k \neq i} \mathbb{Q}_{2\epsilon^i, \epsilon^i + \epsilon^k} M_{i,k} f(\epsilon^i + \epsilon^k) + \mathbb{Q}_{2\epsilon^i, \epsilon^i} \Big)$$

$$= \frac{M_i + \frac{1}{c_i}}{\lambda + M_i + \frac{1}{c_i}} \Big(2 \sum_{k \neq i} M_{i,k} f(\epsilon^i + \epsilon^k) + \frac{1}{c_i M_i + 1} \Big)$$

$$= \Big(1 + \frac{\lambda}{M_i + \frac{1}{c_i}}\Big)^{-1} \Big(\frac{1}{c_i M_i + 1} + 2 \sum_{l \neq i} \frac{M_{i,l} f_{l,i}(s)}{M_i + \frac{1}{c_i}} \Big) \tag{53}$$

サンプル数を 2 つにした場合でも, 式は単純にはならない. 具体的な結果を導くためには, もう少し単純なモデルにしないと, 詳しい結果は得られそうにない. 次の節では, アイランドモデルと円形状配列である飛び石モデルについて説明し, 合祖と近親交配の関係を詳しく述べている. また, これら分集団間の分化の指標として F_{ST} を計算する.

4.2 固定指数 F（Herbots(1994)）

生物がある程度の集落や群集をもって生活していると同時にある程度の近い系統をもつものだと考えるのは当然のことであろう. 近親的なものの間では任意交配とは考えられず, 近親交配として取り扱われる. その場合, ハーディ・ワインベルグの法則のような任意交配を考えたようなモデルではなく, 近親交配係数 F を用いて表されることになる.ここで F は, ある個体の持つ 2 つの相同遺伝子が共通祖先の持つ同じ遺伝子に由来する確率を指す. 実際に 2 つのアレルタイプ A,a が存在し, それぞれの頻度を P_A, P_a とした場合で考えればそれぞれの式は,

ハーディ・ワインベルグの場合

$$P_{AA} = {P_A}^2, P_{Aa} = 2P_AP_a, P_{aa} = {P_a}^2$$

近親交配の場合

$$P_{AA} = {P_A}^2 + P_A(1-P_A)F, P_{Aa} = 2P_AP_a - 2P_AP_aF, P_{aa} = {P_a}^2 + P_a(1-P_a)F$$

である.この F の値は時には F_{ST}であり, F_{ST}を考える場合は集団に階層構造がある場合である.
もし $F = 1$ ならばこれは a, A のどちらかのアレルタイプで統一された分集団で構成されていることを表す. 即ち, それぞれの分集団ごとに各アレルタイプの系統が出来上がっているので, そこから任意に選び出された 2 つのサンプルが同じ祖先遺伝子で構成されている確率は 1 である.よって F_{ST}とは, 任意に選び出された 2 つのサンプルが同じ祖先遺伝子に由来する確率である.言い直せば, 家系同一性 (Identical-by-descent), もしくは合祖し, 突然変異がない場合の確率である.よって合祖理論を論じるときは祖先からの遺伝的距離が非常に近いものを考えていることになる. その他 F_{ST}は全分集団間での遺伝子の多様性を測る尺度ともなりうる.この確率を次のように定義する.

$$F_{ST} = \frac{H_T - H_S}{H_T} \qquad (\textbf{Wright(1950)})$$

それぞれの記号は次のような意味を持つ：H_Tは全分集団間でのヘテロ接合度, H_Sは各分集団間でのヘテロ接合度である.この式を少し書き換えれば, 以下のようになる.

$$F_{ST} = \frac{f_0 - \bar{f}}{1 - \bar{f}} \tag{54}$$

但し,

$f_0(= 1 - H_S)$; 任意に選ばれた 1 つの分集団からランダムに選び出された 2 つのサンプルが同じアレルである確率
$\bar{f}(= 1 - H_T)$; 全分集団からランダムに選び出された 2 つのサンプルが同じアレルである確率
とした.これら f_0と$\bar{f}$はもう少し具体的に書き表すことができて, 突然変異率を $u = \dfrac{\theta}{N}$とするとき

の離散時刻での時刻 t まで突然変異を起こさない確率は,

$$\frac{\theta}{N}\Big(1-\frac{\theta}{N}\Big)^{t-1}$$

よって連続時間での確率は,

$$\lim_{N\to\infty}\frac{\theta}{N}\Big(1-\frac{\theta}{N}\Big)^{[Nt]-1}=e^{-\theta t}$$

よって, この時刻 t を合祖するまでの滞在時間 T で置き換えれば, $E[e^{-\theta T}]$ は合祖するまで突然変異を起こさない場合の平均確率となる.今, T_0 を 1 つの分集団からランダムに選び出された 2 つのサンプルが合祖するまでの滞在時間, $\bar{T}$ を全分集団からランダムに選び出された 2 つのサンプルが合祖するまでの滞在時間とするとき, $f_0, \bar{f}$ は次式で与えられる.

$$f_0=E[e^{-\theta T_0}],\ \bar{f}=E[e^{-\theta \bar{T}}]$$

よって,

$$F_{ST}=\frac{E[e^{-\theta T_0}]-E[e^{-\theta \bar{T}}]}{1-E[e^{-\theta \bar{T}}]}\qquad(\textbf{Slatkin(1991)})$$

で与えられる. ここでは中立な突然変異における無限アレルモデルの下に考察を行っている.
以降, この F_{ST} の値を用いて様々な地理的構造の場合を見ていく.

d 次元トーラス状格子モデル

分集団が d 次元トーラス状の格子空間に配置されていて, p^d 個の分集団からなる.分集団のサイズは全て等しく (全ての i に対し, $c_i=c$), 移住率はどの分集団から移るに際しても同じく一定の移住率に従うものとする.今, $K=\{k=(k_1,k_2,k_3,\cdots,k_d);k_i=0,1,2,\cdots,p-1\}$ とおく. 移住率は空間的に一様で $k,j\in K$ に対して $k-j=i$ のみに依存することとする.この移住率を $m_{k,j}=m_{k-j}=m_i$ と書くことにする.実際にラプラス変換 $f(\alpha)$ を求めよう.式 (49) より,

$$\sum_j(m_j+m_{-j})f(k-j)+\frac{1}{c}(1-f(k))\delta_{k,0}=\lambda f(k)$$

但し, $f(\vec{0})$ は 2 つのサンプルが同じ分集団に滞在する時刻に対するラプラス変換 (確率) である. d 次元ベクトル $\theta=(\theta_1,\cdots,\theta_d)$ を各成分が $\theta_r=\frac{2\pi q}{p}(q=0,1,\cdots,p-1)$ の値をとるとする.母関数を

$$H(\theta)=\sum_k f(k)e^{-i\theta\cdot k}\qquad\text{但し, }\theta\cdot k=\sum_{r=1}^{d}\theta_r k_r$$

逆変換により,

$$H(\theta)=\frac{1}{p^d}\sum_k f(k)e^{-i\theta\cdot k}$$

先程と同様にして,

$$(M(\theta)+M(-\theta))H(\theta)+\frac{1}{c}(1-f(\vec{0}))=\lambda H(\theta)$$

但し, $M(\theta)=\sum_k m_k e^{-i\theta\cdot k}$. これらから,

$$H(\theta)=\frac{f(\vec{0})-1}{c(M(\theta)+M(-\theta)-\lambda)}$$

$$f(\vec{0})=\frac{1}{p^d}\sum_\theta H(\theta)=\frac{f(\vec{0})-1}{cp^d}\Big(\sum_\theta \frac{1}{M(\theta)+M(-\theta)-\lambda}\Big)$$

この式から,

$$f(\vec{0})=\frac{\frac{1}{cp^d}\sum_\theta \frac{1}{M(\theta)+M(-\theta)-\lambda}}{\frac{1}{cp^d}\sum_\theta \frac{1}{M(\theta)+M(-\theta)-\lambda}-1}=\frac{\sum_\theta \frac{1}{M(\theta)+M(-\theta)-\lambda}}{\sum_\theta \frac{1}{M(\theta)+M(-\theta)-\lambda}-cp^d}$$

よって,

$$f(k)=\frac{1}{p^d}\sum_{\theta'}H(\theta)e^{-i\theta'\cdot k}=\frac{1}{p^d}\sum_{\theta'}\frac{f(\vec{0})-1}{c(M(\theta')+M(-\theta')-\lambda)}e^{-i\theta'\cdot k}$$

$$=\frac{1}{p^d}\sum_{\theta'}\frac{\frac{\sum_\theta \frac{1}{M(\theta)+M(-\theta)-\lambda}}{\sum_\theta \frac{1}{M(\theta)+M(-\theta)-\lambda}-cp^d}-1}{c(M(\theta')+M(-\theta')-\lambda)}e^{-i\theta'\cdot k}$$

$$=\frac{1}{p^d}\sum_{\theta'}\frac{\frac{cp^d}{\frac{1}{M(\theta)+M(-\theta)-\lambda}-cp^d}}{c(M(\theta')+M(-\theta')-\lambda)}e^{-i\theta'\cdot k}=\frac{\sum_{\theta'}\frac{e^{-i\theta'\cdot k}}{M(\theta')+M(-\theta')-\lambda}}{\sum_\theta \frac{1}{M(\theta)+M(-\theta)-\lambda}-cp^d}$$

多次元では複雑な式となるので, 単純な場合で観測してみよう. この式を用いることによって, 1 次元のサークル状の飛び石モデルを具体的に考察してみる. Herbots(1994)の論文に従い $d=1, p=n, \lambda=s, c=1$ とする. 移住率を

$$M_{i,j}=\begin{cases}(1-a)\dfrac{M}{2} & (j=i-1\ \text{のとき})\\ a\dfrac{M}{2} & (j=i+1\ \text{のとき})\\ -\dfrac{M}{2} & (j=i\ \text{のとき})\\ 0 & (\text{その他})\end{cases}$$

とすると,

$$M(\theta)+M(-\theta)=M_{i,i-1}e^{i\theta*(-1)}+M_{i,i+1}e^{i\theta*1}+M_{i,i-1}e^{-i\theta*(-1)}+M_{i,i+1}e^{-i\theta*1}+2M_{i,i}$$

$$=(1-a)\frac{M}{2}e^{-i\theta}+a\frac{M}{2}e^{i\theta}+(1-a)\frac{M}{2}e^{i\theta}+a\frac{M}{2}e^{-i\theta}+2(-\frac{M}{2})$$

$$=(1-a)\frac{M}{2}(e^{-i\theta}+e^{i\theta})+a\frac{M}{2}(e^{-i\theta}+e^{i\theta})-M=-M(1-\cos(\theta))$$

これより次式を得る.

$$f(k) = \frac{\frac{1}{n}\sum_{l=0}^{n-1} \frac{\cos(\frac{2\pi lk}{n})}{s+M(1-\cos(\frac{2\pi l}{n}))}}{1+\frac{1}{n}\sum_{l=0}^{n-1} \frac{1}{s+M(1-\cos(\frac{2\pi l}{n}))}}$$

s の微分によって, 平均と分散の式を求めれば,

$$E[T_k] = n + \frac{k(n-k)}{M}$$

$$V[T_k] = n^2 + \frac{n(n^2-1)}{3M} + \frac{k(n-k)(n^2+1-2k(n-k))}{3M^2}$$

この 2 つの式からわかるように, $M \to \infty$ とすれば, 2 つのサンプルの距離に関係なく, 分集団の数だけで決まる. この結果は後ほど述べるアイランドモデルの時と同じである. アイランドモデルの場合とサークル状の飛び石モデルの場合の F_{ST} の値を比較する. まずアイランドモデルの場合を調べるため, (52) と (53) から, ラプラス変換 $f_{ii}(s) = f_0(s)$ と $f_{ij}(s) = f_1(s)$ の値を求める. 記号を合わせるために $D = n, M_{i,j} = \frac{M}{2(n-1)}, M_i = M$ とすれば, (52) 式は,

$$f_1(s) = \Big(1+\frac{2s}{2M}\Big)^{-1} 2\frac{1}{2M}\Big(2\frac{M}{2(n-1)}(n-2)f_1(s) + 2\frac{M}{2(n-1)}f_0(s)\Big)$$

$$f_1(s) = \Big(1+\frac{s}{M}\Big)^{-1} \frac{1}{M}\Big(\frac{M}{n-1}(n-2)f_1(s) + \frac{M}{n-1}f_0(s)\Big)$$

$$(M+s)f_1(s) = \frac{M}{n-1}(n-2)f_1(s) + \frac{M}{n-1}f_0(s)$$

$$Mf_1(s) + (n-1)sf_1(s) - Mf_0(s) = 0$$

同じく (53) 式は,

$$f_{ii}(s) = f_0(s) = \Big(1+\frac{s}{M+1}\Big)^{-1}\Big(\frac{1}{M+1} + \frac{Mf_1(s)}{M+1}\Big)$$

$$f_0(s)(1+M+s) = 1 + 2Mf_1(s), \text{よって}, (1+M+s)f_0(s) - Mf_1(s) = 1$$

この 2 つの式を解いて,

$$f_0(s) = \frac{M+(n-1)s}{M+(nM+n-1)s+(n-1)s^2}$$

$$f_1(s) = \frac{M}{M+(nM+n-1)s+(n-1)s^2}$$

それぞれ s について微分すれば, 平均, 分散は,

$$E_{2\epsilon^i}[T] = n \quad E_{\epsilon^i+\epsilon^j}(T) = 1 + \frac{M+1}{M}(n-1) \ \ (Wakeley(2009))$$

$$V_{\epsilon^i+\epsilon^j}[T] = n^2 + 2\frac{(n-1)^2}{M} = n^2\Big(1 + 2\frac{(n-1)^2}{Mn^2}\Big) \quad (Wakeley(2009))$$

$$V_{2\epsilon^i}[T] = n^2 + 2\frac{(n-1)^2}{M} + \frac{(n-1)^2}{M^2} = n^2\Big(1 + \frac{(n-1)^2}{n^2}\Big(\frac{2}{M} + \frac{1}{M^2}\Big)\Big) \quad (Wakeley(2009))$$

$M \to \infty$ とすれば, 1 次元のサークル状の飛び石モデルの場合と同様の結果であることが確認できる. さて F_{ST}の値を求めよう. 計算すれば,

$$\bar{f} = \frac{1}{n}f_0 + (1 - \frac{1}{n})f_1 = \frac{nM + (n-1)s}{M + (nM + n - 1)s + (n-1)s^2}$$

(54) 式に代入すれば,

$$F_{ST} = \frac{1}{1 + M\frac{n^2}{(n-1)^2} + \frac{\theta n}{n-1}} \quad n \to \infty \quad \to \frac{1}{1 + M + \theta}$$

この結果から, 分集団の数を無限に多くすると, F_{ST} の値は少し大きくなる.
次に 1 次元のサークル状の飛び石モデルについて考察する. この場合, $\bar{f} = \dfrac{f(0) + f(d)}{2}$となるから F_{ST}の値は,

$$F_{ST}(d) = \frac{f_0 - \bar{f}}{1 - \bar{f}} = \frac{f(0) - f(d)}{2 - (f(0) + f(d))} = \frac{\frac{1}{n}\sum_{k=0}^{n-1}\frac{1-\cos(\frac{2\pi kd}{n})}{\theta + M(1-\cos(\frac{2\pi k}{n}))}}{2 + \frac{1}{n}\sum_{k=0}^{n-1}\frac{1-\cos(\frac{2\pi kd}{n})}{\theta + M(1-\cos(\frac{2\pi k}{n}))}}$$

但し, $s = \theta$ と置いた.この $F_{ST}(d)$ の値はある分集団と d ステップ離れた分集団のみに観点を置いて考えられたものであるから, 実際には, $\bar{f}$を全ての分集団に対して考えなければならない. $\bar{f}$は算術平均を用いて次のように置き換えられる ;

$$\bar{f} = \begin{cases} \dfrac{1}{n}f(0) + \dfrac{2}{n}\displaystyle\sum_{d=1}^{\frac{n}{2}-1} f(d) + \dfrac{1}{n}f\Big(\dfrac{n}{2}\Big) & (n \text{ が偶数のとき}) \\ \dfrac{1}{n}f(0) + \dfrac{2}{n}\displaystyle\sum_{d=1}^{\frac{n-1}{2}} f(d) & (n \text{ が奇数のとき}) \end{cases}$$

この時, 三角関数に関する 2 つの公式 ;

$$\sum_{r=1}^{n}\cos(rx) = \cos\Big(\frac{(n+1)}{2}x\Big)\sin\Big(\frac{nx}{2}\Big)\frac{1}{\sin\Big(\frac{x}{2}\Big)}$$

$$\sum_{r=1}^{n}\sin(rx) = \sin\Big(\frac{(n+1)}{2}x\Big)\cos\Big(\frac{nx}{2}\Big)\frac{1}{\sin\Big(\frac{x}{2}\Big)}$$

を用いて $\bar{f}$ を変形するための式を導く. これら上式は $\sum_{k=0}^{n} e^{ikx} = \sum_{k=0}^{n} \cos(kx) + i\sum_{k=0}^{n} \sin(kx)$ の展開式から直ちに求められる. 即ち,

$$\sum_{k=0}^{n} e^{ikx} = \frac{1-e^{i(n+1)x}}{1-e^{xi}} = \frac{e^{-\frac{(n+1)x}{2}i} - e^{\frac{(n+1)x}{2}i}}{e^{-\frac{x}{2}i} - e^{\frac{x}{2}i}} \frac{e^{\frac{(n+1)x}{2}i}}{e^{\frac{x}{2}i}}$$

$$= \frac{-2\sin\left(\frac{(n+1)x}{2}\right)}{-2\sin\left(\frac{x}{2}\right)} e^{nxi} = \frac{\sin\left(\frac{(n+1)x}{2}\right)}{\sin\left(\frac{x}{2}\right)} \left(\cos\left(\frac{nx}{2}\right) + i\sin\left(\frac{nx}{2}\right)\right)$$

この公式から, $k = 1, 2, \cdots$ とすると,

$$\sum_{d=0}^{n-1} \cos\left(\frac{2\pi kd}{n}\right) = 1 + \sum_{d=1}^{n-1} \cos\left(\frac{2\pi kd}{n}\right) = 1 + \cos(k\pi)\sin\left(\frac{n-1}{n}\pi k\right) \frac{1}{\sin\left(\frac{\pi k}{n}\right)}$$

三角関数の公式より,

$$\sin\left(\frac{n-1}{n}\pi k\right) = \sin(\pi k)\cos\left(\frac{k\pi}{n}\right) - \cos(k\pi)\sin\left(\frac{k\pi}{n}\right) = -\cos(k\pi)\sin\left(\frac{k\pi}{n}\right)$$

よって,

$$\sum_{d=0}^{n-1} \cos\left(\frac{2\pi kd}{n}\right) = 1 - \cos^2(k\pi) = 0$$

求められる結果は,

$$\sum_{d=0}^{n-1} \cos\left(\frac{2\pi kd}{n}\right) = \begin{cases} 0 & (k \neq 0 \text{ のとき}) \\ n & (k = 0 \text{ のとき}) \end{cases}$$

これら三角関数の公式により, $\bar{f}$は次のように求められる.

$$\bar{f} = \frac{1}{\theta\{n + \sum_{k=1}^{n-1} \frac{1}{\theta + M(1-\cos(\frac{2\pi k}{n}))}\}}$$

(54) から, F_{ST}の値は,

$$F_{ST} = \frac{\sum_{k=1}^{n-1} \frac{1}{\theta + M(1-\cos(\frac{2\pi k}{n}))}}{n + \sum_{k=1}^{n-1} \frac{1}{\theta + M(1-\cos(\frac{2\pi k}{n}))}}$$

アイランドモデルの時と同様に分集団の数を無限にした極限を考える. そのため, 微分積分における平均の公式 ;

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} f\left(a + (b-a)\frac{k}{n}\right) = \frac{1}{b-a} \int_a^b f(x)dx$$

を用いる. 但し, $b > a$ で, かつ, $f(x)$ は $[a, b]$ で連続である.この式を用いれば,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n-1} \frac{1}{\theta + M(1-\cos(\frac{2\pi k}{n}))} = \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{\theta + M(1-\cos(x))} dx$$

ここで複素積分を用いて, 単位円周上で積分を行う. $z = e^{ix}$と置けば, $dx = \dfrac{dz}{iz}$であるから,

$$\frac{1}{2\pi}\int_0^{2\pi} \frac{1}{\theta + M(1-\cos(x))}dx = \frac{1}{2\pi}\int_{|z|=1} \frac{1}{\theta + M(1-\frac{z+z^{-1}}{2})}\frac{dz}{iz}$$

$$= \frac{-1}{2\pi i}\int_{|z|=1} \frac{2}{Mz^2 - 2(\theta+M)z + M}dz$$

分母に関して, z の解を求めれば,

$$z = \frac{(\theta+M) + \sqrt{(\theta+2M)\theta}}{M},\ \frac{(\theta+M) - \sqrt{(\theta+2M)\theta}}{M}$$

$$\Big|\frac{(\theta+M) + \sqrt{(\theta+2M)\theta}}{M}\Big| > 1, \Big|\Big(\frac{(\theta+M) + \sqrt{(\theta+2M)\theta}}{M}\Big)\Big(\frac{(\theta+M) - \sqrt{(\theta+2M)\theta}}{M}\Big)\Big| = 1$$

であるから, $\Big|\dfrac{(\theta+M) - \sqrt{(\theta+2M)\theta}}{M}\Big| < 1$.よって, 単位円周内にあるのは$\dfrac{(\theta+M) - \sqrt{(\theta+2M)\theta}}{M}$だけであるから, 留数 (Res) を求めると,

$$Res\Big(\frac{(\theta+M) - \sqrt{(\theta+2M)\theta}}{M}\Big) = \Big[\frac{2}{\frac{d}{dz}(Mz^2 - 2(\theta+M)z + M)}\Big]_{z=\frac{(\theta+M)-\sqrt{(\theta+2M)\theta}}{M}} = \frac{-1}{\sqrt{\theta(\theta+2M)}}$$

求められる式は,

$$\frac{1}{2\pi}\int_0^{2\pi} \frac{1}{\theta + M(1-\cos(x))}dx = \frac{-1}{2\pi i}(2\pi i)\frac{-1}{\sqrt{\theta(\theta+2M)}} = \frac{1}{\sqrt{\theta(\theta+2M)}}$$

よって,

$$\lim_{n\to\infty} F_{ST} = \frac{\frac{1}{\sqrt{\theta(\theta+2M)}}}{1 + \frac{1}{\sqrt{\theta(\theta+2M)}}} = \frac{1}{1+\sqrt{\theta(\theta+2M)}}\Big(> \frac{1}{1+M+\theta}\Big)$$

結果, 分集団の数を無限に多くした場合は F_{ST}の値がサークル状の飛び石モデルの場合よりアイランドモデルの場合の方が小さくなることがわかった. 同じ総移住率$\dfrac{M}{2}$という条件の下で, サークル状の飛び石モデルの方が分集団の分化が起こりやすいことを示している.アイランドモデルの方が分集団間の相互移住の関係が強く, 予想される結果である.

4.3　地理的構造を持つ非保存的移住の合祖モデル ((K.Sampson(2006))

この節では分集団の数が有限かつ各世代において集団サイズの変動がある場合を取り扱っている Sampson(2006) の結果を紹介する. この内容によれば, 各分集団サイズが定常分布に従い, 毎世代変化する. 収束に関しては有限状態マルコフ連鎖に関する次の Möhle(1998) の補題を利用する：

補題 (Möhle(1998))

固定された $t, K \geq 0$ と $\lim_{N\to\infty} c_N = 0$ となる正の実数列 c_Nを仮定する. 行列 $\mathbb{A} = (a_{ij})$ のノルムを $||\mathbb{A}|| := \max_i \sum_j |a_{ij}|$ で定義する.この時, $\mathbb{A}$ が $||\mathbb{A}|| = 1$, $\lim_{m\to\infty} \mathbb{A}^m =: \mathbb{P}$ を満たすならば,

$$\lim_{N\to\infty} \sup_{||\mathbb{B}||\geq K} ||(\mathbb{A} + c_N\mathbb{B})^{[\frac{t}{c_N}]} - (\mathbb{P} + c_N\mathbb{B})^{[\frac{t}{c_N}]}|| = 0$$

もし, $\mathbb{G} := \lim_{N\to\infty} \mathbb{P}\mathbb{B}_N\mathbb{P}$ を満たす行列の列 $(\mathbb{B}_N)_{N\in\mathbb{Z}_+/\{0\}}$が存在すれば,

$$\lim_{N\to\infty} (\mathbb{A} + c_N\mathbb{B}_N)^{[\frac{t}{c_N}]} = \mathbb{P} - \mathbb{I} + e^{t\mathbb{G}} \quad \forall t > 0$$

□

以下では Sampson(2006) のモデルを説明する.

(i) 前向きのプロセス

半数体生物集団 ($haploid\ model$) を考える. M 個のコロニーを考え, それぞれ集団のサイズを $N_k = a_k N (1 \leq k \leq M)$ とする.分集団のサイズは状態空間 $S_N = \{a_i N = (a_{i1}N, , \cdots, a_{iM}N); i \in \{1, 2, \cdots, s\}\}$ 上で定常分布$\gamma = (\gamma_1, \cdots, \gamma_s)$ に従う定常マルコフ連鎖である. i は集団サイズの変化量を与えたものであり, それが毎世代, s 通りの変化をもって変化していく. 前向き移住率は次の式で与えられる：

$$m_{k,j} = \frac{\mu_{k,j}}{N} + o\Big(\frac{1}{N}\Big) \quad for\ k \neq j,\ m_{k,k} = 1 - \sum_{j\neq k} m_{k,j}$$

非保存的移住であるから,

$$N_k \sum_{j\neq k} m_{k,j} \neq \sum_{j\neq k} N_j m_{j,k}$$

が成立する. 但し, 保存的移住の場合も含んでいる.

(ii) 後向きのプロセス

全集団からサイズ n のサンプルを取り出し, その祖先遺伝子のプロセス $Y_N(\tau)$ を

$$E = \{r = (r_1, r_2, \cdots, r_M) \in \{\mathbb{Z}_+/\{0\}\}^M : 1 \leq |r| = \sum_{i=1}^{M} r_i \leq n\}$$

の集合の上で変化するマルコフ過程とする.ここで, $Y_N(\tau)$ の第 i 成分はτ世代前に分集団 i に見いだされる祖先遺伝子の数である. $N_i(\tau)$ をτ世代前の分集団 i の集団のサイズとすると, 全てのコロニーの集団のサイズはベクトル $N(\tau) = (N_1(\tau), \cdots, N_M(\tau))$ によって与えられ, かつ, 推移確率を

$$t_{ij} = P(N(\tau + 1) = a_j N | N(\tau) = a_i N)$$

で与えれば, 定常分布γから, $\gamma T = \gamma$が成り立つ.但し, $(T)_{i,j} = t_{i,j}$である. また, 後ろ向き移住率 $f_{kl|j}$を集団サイズが $a_j N$ という条件の下で分集団 k から任意に選んだ 1 個体が 1 世代前に分集団 l から移住してきた確率とするとき, 次の式で与えられる :

$$f_{kl|j} = \frac{m_{lk} a_{jl} N}{\sum_{i=1}^{M} m_{ik} a_{ji} N} = \frac{m_{lk} a_{jl}}{\sum_{i=1}^{M} m_{ik} a_{ji}}$$

$k \neq l$ のとき $f_{kl|j} = \left(\frac{a_{jl}\mu_{lk}}{N} + o\left(\frac{1}{N}\right)\right) \frac{1}{\sum_{i \neq k} \frac{a_{ji}\mu_{ik}}{N} + a_{jk}(1 - \sum_{i \neq k} \frac{\mu_{ki}}{N}) + o\left(\frac{1}{N}\right)} = \mu_{lk} \frac{a_{jl}}{a_{jk} N} + o\left(\frac{1}{N}\right)$

$k = l$ のとき $f_{kk|j} = 1 - \sum_{l \neq k} \mu_{lk} \frac{a_{jl}}{a_{jk} N} + o\left(\frac{1}{N}\right)$

2 変量マルコフ連鎖

$X_N(\tau) = (N(\tau), Y_N(\tau))$ を集合 $E_N(= S_N \times E)$ の上での 2 変量マルコフ連鎖とする.毎世代, 分集団のサイズの変動と祖先の地理的配置を同時に考えたプロセスである. Coalescent rateは集団のサイズに依存するため, このようにして考えられた.推移確率を

$$\pi_{(i,\gamma),(j,\beta)} = P(X_N(\tau+1) = (a_j N, \beta) | X_N(\tau) = (a_i N,\ \gamma))$$

で与える.1 個体の移住あるいは 2 個体のCoalesceの確率は$\frac{1}{N}$のオーダーとなり,
それ以外の変化の確率は $o(\frac{1}{N})$(高次の無限小) となることにより,1 世代当たりの推移確率$\Pi_N = (\pi_{(i,\gamma),(j,\beta)})$ について次の命題が成り立つ;

命題（Sampson(2006))

$\Pi_N = \mathbb{A} + \dfrac{\mathbb{B}}{N} + o(\dfrac{1}{N})$; $\mathbb{A} = (a_{(i,\gamma),(j,\beta)}) = (t_{ij}\delta_{\gamma,\beta})$, $\mathbb{B} = (b_{(i,\gamma),(j,\beta)})$

$$b_{(i,\gamma),(j,\beta)} = \begin{cases} -t_{ij}\displaystyle\sum_{k=1}^{M}\Big(\frac{\gamma_k(\gamma_k-1)}{2a_{jk}} + \gamma_k\sum_{l\neq k}\frac{\mu_{lk}a_{jl}}{a_{jk}}\Big) & (\gamma=\beta\text{のとき}) \\ t_{ij}\gamma_k\dfrac{\mu_{lk}a_{jl}}{a_{jk}} & (\beta=\gamma-\epsilon^k+\epsilon^l,\ k\neq l\ \text{のとき}) \\ t_{ij}\dfrac{\gamma_k(\gamma_k-1)}{2a_{jk}} & (\beta=\gamma-\epsilon^k\text{のとき}) \\ 0 & (\text{その他}) \end{cases} \tag{55}$$

□

まず, Möhle(1998)の補題により,

$$\lim_{N\to\infty}\Pi_N{}^{[Nt]} = \mathbb{P} - \mathbb{I} + e^{t\mathbb{G}}$$

ここで,

$$\mathbb{P} = (p_{(i,\gamma),(j,\beta)}) = \lim_{m\to\infty}\mathbb{A}^m = (\gamma_j\delta_{\gamma,\beta})$$

$$\mathbb{G} = (g_{(i,\gamma),(j,\beta)}) = \mathbb{P}\mathbb{B}\mathbb{P} = (\gamma_j q_{\gamma,\beta}), \qquad q_{\gamma,\beta} = \sum_i \gamma_i \sum_j b_{(i,\gamma),(j,\beta)}$$

更に状態空間 E が有限であることにより, Ethier and Kurtz(1986)の定理およびKaji(2001), から次の定理が成り立つ.

定理（Sampson(2006)）

初期分布 $Y_N(0) \to w,\ N \to \infty$(法則収束) のとき, $Y_N([Nt])_{t\geq 0} \to Y = (T(t))_{t\geq 0},\ N \to \infty$(法則収束) かつ, $Y(0) \overset{d}{\cong} w$ が成立するとき, $Y = (Y(t))$ の生成作用素：$\mathbb{Q} = (q_{\gamma,\gamma'})_{\gamma,\gamma'\in E}$は,

$$q_{\gamma,\gamma'} = \begin{cases} -\sum_{k=1}^{M} \alpha_k \Big(\dfrac{\gamma_k(\gamma_k - 1)}{2} + \gamma_k \sum_{l\neq k} \dfrac{\beta_{kl}}{2} \Big) & (\gamma = \beta \text{のとき}) \\ \gamma_k \dfrac{\beta_{kl}}{2} & (\beta = \gamma - \epsilon^k + \epsilon^l,\ k \neq l \text{ のとき}) \\ \alpha_k \dfrac{\gamma_k(\gamma_k - 1)}{2} & (\beta = \gamma - \epsilon^k \text{のとき}) \\ 0 & (\text{その他}) \end{cases} \tag{56}$$

と表される. 但し, $\alpha_k = \sum_{i=1}^{s} \dfrac{\gamma_i}{\alpha_{ik}},\quad \beta_{kl} = 2\mu_{lk} \sum_{i=1}^{s} \gamma_i \dfrac{a_{il}}{a_{ik}}$

(Sampson の証明の概要)

$\eta_N : E_N \to E,\quad \eta_N(a_i N, \gamma) = \gamma,\ f \in \mathbb{B}(E)$ とするとき,

$$\Im_N{}^{[Nt]}(f \circ \eta_N)\,(a_i N, \gamma) = \sum_{(a_j N, \gamma') \in E_N} f \circ \eta_N(a_j N, \gamma')(\Pi_N{}^{[Nt]})_{(i,\gamma),(j,\gamma')}$$

他方, $\Im(t)f(\gamma) = \sum_{\gamma' \in E} f(\gamma')(e^{t\mathbb{Q}})_{\gamma,\gamma'}$とする. Ethier and Kurtz(1986)の定理およびKaji et al(2001)から,

$$|\Im_N{}^{[Nt]}(f\circ\eta_N)\,(a_i N, \gamma) - \Im(t)f(\gamma)| = |\sum_{\gamma' \in E} f(\gamma')(\sum_{j=1}^{s} (\Pi_N{}^{[Nt]})_{(i,\gamma),(j,\gamma')} - (e^{t\mathbb{Q}})_{\gamma,\gamma'})| \to 0,\quad N \to \infty$$

を示せばよい. E は有限集合なので, 各項について収束を示せば十分である. Möhle(1998)の補題と少々の計算により次式が成立する.

$$\lim_{N\to\infty} |\sum_{i=1}^{s} (\Pi_N{}^{[Nt]})_{(i,\gamma),(j,\gamma')} - (e^{t\mathbb{Q}})_{\gamma,\gamma'})| = |\sum_{j=1}^{s} (e^{t\mathbb{G}})_{\gamma,\gamma'} - (e^{t\mathbb{Q}})_{\gamma,\gamma'}| = 0$$

□

以上が Sampson(2006) の結果である. Möhle and Notohara(2016) によれば, この結果をもう少し拡張できて,

定理 (Möhle and Notohara(2016))

c_Nと d_Nを $\lim_{N\to\infty} c_N = 0, \lim_{N\to\infty} d_N = 0, \lim_{N\to\infty} \frac{c_N}{d_N} = 0$ を満たす正の実数列とする. $\mathbb{P} := \lim_{m\to\infty}(\mathbb{I} + d_N\mathbb{Q})^m$とする時, 生成行列 $\mathbb{Q}$ と $\mathbb{G} := \lim_{N\to\infty} \mathbb{P}\mathbb{B}_N\mathbb{P}$ を満たす行列の列 $(\mathbb{B}_N)_{N\in\mathbb{Z}_+/\{0\}}$が存在すれば, 全ての $t > 0$ に対し,

$$\lim_{N\to\infty}(\mathbb{I} + d_N\mathbb{Q} + c_N\mathbb{B}_N)^{[\frac{t}{c_N}]} = \mathbb{P} - \mathbb{I} + e^{t\mathbb{G}} = \mathbb{P}e^{t\mathbb{G}} = e^{t\mathbb{G}}\mathbb{P}$$

また,全ての $N \in \mathbb{Z}_+/\{0\}$ に関して, 各世代の遷移確率行列が $\mathbb{I} + d_N\mathbb{Q} + c_N\mathbb{B}_N$に従う, 可算有限な状態空間 S 上の離散時刻マルコフ連鎖を $(X_N(r))_{r\in\mathbb{Z}_+}$ とするとき, もし, 初期分布の列 $P_{X_N(0)}$がある確率測度μに弱収束するならば, マルコフ連鎖 $(X_N([\frac{t}{c_N}]))_{t\geq 0}$の有限次元分布は初期分布$\mu$, 遷移確率行列$\Pi(t) = \mathbb{P} - \mathbb{I} + e^{t\mathbb{G}} = \mathbb{P}e^{t\mathbb{G}} = e^{t\mathbb{G}}\mathbb{P}(t > 0)$ の連続時間のマルコフ連鎖 $(X_t)_{t\geq 0}$の有限次元分布に収束する.

さらに, この定理は無限次元の行列に対して以下のように拡張できる；

$l^\infty = \{x = (x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}; ||x|| = \sup_{i\in\mathbb{N}} |x_i| < \infty\}, \mathbb{A} = (a_{i,j})_{i,j\in\mathbb{N}}$ 但し, $||\mathbb{A}|| = \sup_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} |a_{i,j}| < \infty$ を満たすものとする.このとき $x \in l^\infty$に対して $(\mathbb{A}x)_i = \sum_{j\in\mathbb{N}} a_{i,j}x_j$と定義すると, 行列 $\mathbb{A}$ はバナッハ空間 l^∞から l^∞への線形作用素となる.このような l^∞から l^∞への線形作用素全体を L とすると, L は完備距離空間であり, $\mathbb{A}, \mathbb{B} \in L$ の時 $||\mathbb{A}\mathbb{B}|| \leq ||\mathbb{A}||||\mathbb{B}||$.また, 行列 $\mathbb{A}$ の指数関数 $e^{\mathbb{A}} = \sum_{n=0}^{\infty} \frac{\mathbb{A}^n}{n!}$も有限次元の場合と同様に定義できる.この時生成行列 $\mathbb{Q} = (q_{i,j})$ 及び行列の列 $(\mathbb{B}_N)_{N\in\mathbb{Z}_+/\{0\}}$が無限次元の L の行列である場合に対しても定理は成り立つ.

この定理により, Sampson の定理は以下のように拡張される；

1. 分集団サイズの変動の 1 世代当たりの推移確率を, l を $0 \leq l < 1$ の実数として,

$$P(N(\tau+1) = a_jN | N(\tau) = a_iN) = \begin{cases} \frac{t_{i,j}}{N^l} & if\ i \neq j \\ 1 - \frac{t_{i,j}}{N^l} & if\ i = j \end{cases}$$

と置くと, 1 世代当たりの推移確率は$\Pi_N = \mathbb{I} + \frac{\mathbb{A}}{N^l} + \frac{\mathbb{B}}{N} + o(\frac{1}{N})$ と書ける.ここで $\mathbb{A} = (t_{i,j}\delta_{\gamma,\beta})$, 但し $t_{i,i} = -\sum_{j\neq i} t_{i,j}$とする.このとき, $N \to \infty$ においてSampsonの結果がそのまま成り立つ.

2. 分集団のサイズの状態空間および分集団の個数を可算無限とし ($s = \infty$、$M = \infty$) 分集団のサイズ変動の 1 世代当たりの推移確率を 1.で定義されたものとする.このとき, $N \to \infty$ において離散時間マルコフ連鎖から, $s = \infty, M = \infty$ とした (56) を生成作用素とする連続時間マルコフ連鎖への有限次元分布の収束が成り立つ.

5　まとめ

本論文では中立な遺伝子で, 非保存的移住率と可換モデルと呼ばれる一般的な繁殖モデルの下で, 生物の地理的構造を考慮に入れた離散時間モデルから出発し, 各分集団のサイズが一様に無限に大きくする極限操作により, 集団からサンプルした遺伝子の祖先の遺伝子系図を表現するSCM が導かれることを厳密に証明した. これは可換モデルの下で Coalescent 過程を導出したKingman(1982) の定理の集団構造を持つ場合への自然な拡張ということができる. 4.1 節, 4.2 節で共通祖先に到達するまでの時間 (Coalescence time) の分布及び固定指数 F について, トーラス状格子モデルの場合に具体的な解を求めた. しかし, 一般の SCM において, このような具体的な解を求めることは困難である. SCM は移住と合祖によって推移する連続時間マルコフ連鎖であり, 状態 α での滞在時間はパラメーター $|\mathbb{Q}_{\alpha,\alpha}|$ の指数分布に従い, その滞在の後, 確率 $\dfrac{\mathbb{Q}_{\alpha,\beta}}{|\mathbb{Q}_{\alpha,\alpha}|}$（ただし$\beta \neq \alpha$）で状態$\alpha$から状態$\beta$ へ推移する. 梅田 (2005) はこの性質を利用してコンピューター・シミュレーションを行い, 共通祖先に到達するまでの時間の分布, サンプル遺伝子の分離サイトの数の分布を求めた (結果は Notohara and Umeda(2006) で発表). 第 4 章の 4.3 で紹介したように, 分集団の個数が有限の場合には Möhle の補題の応用として, SCM が導かれることが Sampson(2006) によって示された. さらに Sampson のモデルは各分集団のサイズが有限個の状態を確率的に変動する場合も含んでいる. 他方, 本論文では分集団の個数は可算無限個であり, 離散時間から連続時間マルコフ連鎖の弱収束の証明は, 3.3 節に示したように有限個数の場合に比べるとより精密な議論を必要とする. 分集団サイズが変動する Sampson の結果を可算無限個の分集団を含む場合に拡張することは, 興味ある問題であるが, Möhle and Notohara(2016) の定理の利用により有限次元分布の収束までは示すことができるが, 弱収束の証明は未解決の問題である. SCM は長年集団遺伝学で使われて来たモデルであり, SCM を利用したデータ解析ソフトも Genetree, Migrate-n など多く開発されている. 本研究によって, このようなデータ解析のための基本モデルである SCM の頑健性を証明したと言うことができる.

謝辞

本研究を遂行するにあたって, 名古屋市立大学大学院システム自然科学研究科教授, 能登原盛弘先生に御指導を頂きまして, 心から深く感謝の意を表します. また, ゼミに参加してくださった副指導教員である鈴木善幸教授, 村瀬香准教授ともに理論集団遺伝学を研究させて頂きました事, ありがとうございました. 特に, 客員教授である清水昭信先生には非常に熱心かつ丁寧に御指導頂きました. 誠にありがとうございました.

参考文献

[1] Bahlo,M. and Griffiths,R.C. (2000). Coalescence times for two genes from a subdivided population. *J.Math.Biol.* **43,** 397–410.

[2] Cann,R.L., Stoneking,M.and Wilson,A.C. (1987). Mitochondrial DNA and human evolution. *Nature***325,**31-36.

[3] Cannings, C. (1974). The latent roots of certain Markov chains arising in genetics: A new approach, I. Haploid models. *Adv. Appl. Prob.* **6,** 260–290.

[4] Ethier,S.N. and Kurtz,T.G. (1986) *Markov Processes:Characterization and Convergence* Wiley, New York.

[5] Ewens,W.J.(1972).The sampling theory of selectively neutral alleles. *Theor.Popul.Biol.***3**,87–112.

[6] Hammer,M.F.,Karafet,T.M.,Redd,A.J.,Jarjanazi,H.,Santachiara-Benerecetti,S.,Soodyall,H.,Zegura,S.L.(2001) Hierarchical patterns of global human Y-chromosome diversity.*Mol.Biol.Evol.***18**(7),1189-1203.

[7] Hein,J. Schierup,M. and Wiuf,C.(2005) *Gene genealogies, Variation and Evolution* Oxford, University Press.

[8] Herbots, H.M. (1994). Stochastic models in population genetics: genealogical and genetic differentiation in structured populations. PhD diss. University of London.

[9] Herbots, H.M. (1997). The structured coalescent. In: P. Donnelly and S. Tavare: Progress in population genetics and human evolution (IMA Volumes in Mathmatics and its Applications, vol. 87, pp. 231-255) New York:Springer–Verlag.

[10] Kaj,I. Krone,S.M. and Lascoux,M. (2001). Coalescent theory for seed bank models. *J.Appl.Prob.* **38,** 285–301.

[11] Kimura,M.(1953)."Stepping-Stone" Model of Population. Annual Report. *National Institute of Genetics* **3,** 62-63.

[12] Kimura, M.(1968). Evolutionary rate at the molecular level. *Nature* **217**, 624-626.

[13] Kimura,M. and Weiss, G.H.(1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49,** 561-576.

[14] Kingman,J.F.C. (1982a). On the genealogy of large populations. *J.Appl.Prob.* **19A,** 27–43.

[15] Kingman,J.F.C. (1982b). The coalescent. *Stochastic Process* **13,** 235–248.

[16] Kingman,J.F.C. (1982c). Exchangeability and the Evolution of Large Population. In ‘‘Exchangeability in Probability and Statistics’’ G.Koch and F.Spizzichino(North-Holland Pub. Comp.), 97-112.

[17] Malecot,G.(1967). Identical loci and Relationship. Proc. Fifth Berkeley Symp.

Math.Prob. **4,** 317-332.

[18] Maruyama, T.(1970). Stepping stone models of finite length. *Adv.Appl. Prob.* **2,** 229-258.

[19] Möhle,M. (1998).A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. *Adv.Appl.Prob.* **30,** 493–512.

[20] Möhle,M. and Notohara,M.(2016 年掲載予定). An extension of a convergence theorem for Markov chains arising in population genetics.

[21] Notohara,M. (1990).The coalescent and the genealogical process in geographically structured populations. *J.Math.Biol.* **36,** 188–200.

[22] Notohara,M.(2000).A perturbation method for the structured coalescent with strong migration. *J.Appl.Prob.* **37**, 148-167.

[23] Notohara,M. and Umeda,T.(2006). The coalescence time of sampled genes in the structured coalescent model. *Theor.Popul.Biol.* **70**, 289-299.

[24] Sampson,K.Y. (2006). Structured Coalescent With Nonconservative Migration. *J.Appl.Prob.* **43**, 351–362.

[25] Slatkin,M.(1991). Inbreeding coefficients and coalescence times. *Genet.Res.Camb.***58**,167-175.

[26] Tajima,F.(1983).Evolutionary relationship of DNA sequences in finite populations. *Genet.Res,Camb.* **52**, 213–222.

[27] Tajima,F.(1989).DNA Polymorphism in a Subdivided Population: The Expected Number of Segregating Sites in the Two-Subpopulation Model. *Genetics.* **123**, 229-240.

[28] Takahata,N.(1988).The coalescent in two partially isolated diffusion populations. *Genet.Res,Camb.* **52**, 213–222.

[29] 梅田高呂 (2005):A Study of Structured Coalescent by Monte Carlo Simulation(地理的構造を持つ合祖モデルのモンテカルロシミュレーション (修士論文)).

[30] Wakeley,J. (2009).Coalescent Theory. *An Introduction.* Roberts and Company Publishers.

[31] Watterson,G.A.(1975). On the number of segregating sites in genetical model without recombination. *Theor. Popul. Biol.***7**, 256-276.

[32] Wilkinson–Herbots,H.M.(1998).Genealogy and subpopulation differentiation under various models of population structure. *J.Math.Biol.* **37**, 535–585.

[33] Wright,S.(1950). Genetical structure of populations. *Nature*, **166**, 247-249.

発表論文

[学術雑誌論文]

1. Ryouta Kozakai, Akinobu Shimizu and Morihiro Notohara, "Convergence to the structured coalescent processes", Journal of Applied Probability, 2016 年掲載予定.

[口頭発表]

1. 小酒井　亮太,　清水昭信,　能登原盛弘,　”地理的構造を持つ遺伝子系図モデル”, 日本応用数理学会, 2014 年 9 月 5 日.
2. 小酒井　亮太,　”Structured Coalescent 過程への収束”,　統計数理研究所研究集会「科学における確率」, 2015 年 7 月 7 日.
3. 小酒井　亮太,　”地理的構造を持つ合祖過程”, 生命情報科学若手の会第 7 回研究会, 2015 年 10 月 2 日.

用語集

1. 分子進化の中立説：分子レベルでの進化は自然選択に対して有利でも不利でもない中立な突然変異と遺伝的浮動によって起こるとする説.1968 年に木村資生によって提唱された.
2. 遺伝的浮動：生物集団の個体数が有限であるために起こる遺伝子頻度の偶然的変動. 特に個体数が小さい集団ほど変動が大きい. 遺伝的浮動により遺伝的多様性は減少して行き, 最終的に遺伝子の固定, あるいは消失が起こる.
3. マルコフ性：確率論における確率過程の持つ特性の一種で, その過程の将来状態の条件付確率分布が, 現在の状態のみに依存し, 過去の如何なる状態にも依存しない特性を持つことを言う. マルコフ性のある確率過程をマルコフ過程と呼び, とりうる状態が離散的 (有限または可算) なものをマルコフ連鎖という. 合祖モデルもマルコフ連鎖である.
4. Coalescent theory（合祖理論) : Kingman(1982) 及び Tajima(1983) によって独立に提唱され, マルコフ連鎖を用いた理論. 集団から任意に取り出した遺伝子の祖先を遡ると, サンプル遺伝子の祖先は次第に共有され, 最終的に 1 つの共通祖先に到達する. 共通祖先に到達するまでの時間, 遺伝的多様性などを求めることができる.
5. 任意交配 (random mating) : 集団中の雄と雌の個体間で選り好みなく, ランダムに交配が行われること. 反対に表現型が似た者同士の交配様式は同類交配 (assortative mating) という.
6. 遺伝的多型：生物集団の多くには様々な遺伝的変異がみられること. 人の血液型はその例である.1960 年代に電気泳動法により, タンパク質の遺伝的多型の存在が明らかになった. 近年は遺伝子の DNA 塩基配列における多型として SNP(Single Nucleotide Polymorphism) などが知られている.
7. 分集団：生物集団がいくつかの小集団に分かれて繁殖を行っているとき, 各小集団のことで各分集団内では任意交配が仮定されることが多い.
8. 飛び石モデル：1953 年, 木村資生によって提案されたモデル. 生物集団が分集団に分かれ, 各分集団内での任意交配による繁殖と, 分集団間での個体の移住を考慮に入れたモデル.
9. 島モデル：Wright(1943) によって提案されたモデル. 個体数が同じいくつかの分集団からなり, 全ての分集団間で同じ率での移住を仮定したモデル.