

数理遺伝学入門

— 遺伝子系図の確率モデル —

はじめに

集団遺伝学は、R.A.Fisher, S.Wright, J.B.S. Halden, G.Malecot などによる数理統計あるいは確率過程を用いたおよそ1世紀にわたる理論的研究の歴史がある。日本においては、1950年代半ばより、「分子進化の中立説」の提唱者、木村資生博士をはじめ太田朋子、丸山毅夫の諸先生方によって主に拡散過程モデルを用いた理論集団遺伝学の指導的な研究がなされてきた。他方、1980年代初めより、J.F.C.Kingman、田嶋文生およびR.R.Hudsonによって導入された Coalescent モデル（合祖モデル）は新たな集団遺伝学モデルとして注目され、精力的に研究されてきた。これは、集団からサンプルした遺伝子の系図を表現する確率モデルであり、現実の遺伝子配列データと直結した基礎理論として実験、理論の両面から注目され広範な研究及び応用がなされている。特に近年ではマルコフ連鎖モンテカルロシミュレーションによるベイズ推定など多型データ解析への実際的な応用研究がされている。モデル自体はマルコフ過程の基礎的知識で十分理解できるものであり、多くの興味ある問題を提供し、今後の遺伝学あるいは生態学など種々の分野への応用も期待される。この分野は現在も発展途上にあり、より進んだ内容、最新の詳しい情報については巻末に上げた参考論文を直接御覧いただきたい。ここで紹介する内容は遺伝子系図の基本的かつ数学的側面に限定されたものであり、実際的なデータ解析への応用を伴うような集団遺伝学の多くの論文は含まれていないことを断っておかなければならない。数学のレベルとしては大学の理工系学部の知識で十分理解できる内容であり、それを超える事項については付録に簡単な解説を付けている。また生物学に関する専門的な知識も特に必要とはしない。

第1章では本書で必要となるマルコフ連鎖について紹介し、第2章は生物集団の種々の繁殖マルコフ連鎖モデルについて、第3章は合祖過程(Coalescent process)について、第4章は集団からサンプルした遺伝子の対立遺伝子構成に関する Ewens のサンプリング公式と関連する内容について、第5章は自然選択、組み換えを含む遺伝子系図過程について、第6章は地理的構造など構造を持つ合祖過程(Structured coalescent process)について、第7章は異なる時間スケールで起こる現象の連続時間マルコフ連鎖モデルについて、第8章は無限サイトモデルにおけるサンプルの諸量の分布と中立仮説検定法について紹介する。紹介する内容については解説あるいは証明を可能な限り丁寧に付けたつもりである。そのために、紹介できるテーマは限定されることになるが、数理遺伝学の基礎および手法については十分な知識を習得することができると思う。

2016年12月

能登原 盛弘

目次

1. マルコフ連鎖
 1. 1 マルコフ性と推移確率
 1. 2 到達確率と到達時間
 1. 3 離散時間マルコフ連鎖の例
 1. 4 指数分布と連続時間マルコフ連鎖
 1. 5 出生死滅過程とコルモゴロフ方程式
 1. 6 ポアソン過程および有限マルコフ過程
2. 生物集団の繁殖モデル
 2. 1 ライト・フィッシャーモデルとモランモデル
 2. 2 分枝過程モデルと可換モデル
 2. 3 遺伝子頻度の拡散過程
 2. 4 自然選択による遺伝子頻度の変化
3. 遺伝子系図の合祖過程
 3. 1 合祖モデル
 3. 2 遺伝的多様性と遺伝子系図
 3. 3 合祖過程の種々の性質
 3. 4 モランモデルとサブサンプルの遺伝子系図
 3. 5 集団サイズの変動を含む遺伝子系図
4. Ewens のサンプリング公式、ポアソン・ディリクレ分布とポリアの壺様モデル
 4. 1 Ewens のサンプリング公式
 4. 2 アレルの年齢
 4. 3 固有変異の多型 (Unique Event Polymorphism)
 4. 4 ディリクレ分布とポリアの壺モデル
 4. 5 ポアソン・ディリクレ分布
 4. 6 Ewens のサンプリング公式とポリアの壺様モデル
5. 自然選択、組み換えを含む遺伝子系図
 5. 1 自然選択を含む遺伝子系図 (Ancestral Selection Graph)
 5. 2 組み換えを含む遺伝子系図 (Ancestral Recombination Graph)
6. 構造を持つ合祖過程(Structured Coalescent)
 6. 1 地理的構造を持つ集団における遺伝子系図
 6. 2 合祖時間分布と分離サイトの数の分布
 6. 3 大きな移住率および小さな移住率における遺伝子系図モデル
 6. 4 分離サイト数からの合祖時間の推定
 6. 5 自然選択による構造を持つ遺伝子系図 (Hudson&Kaplan のモデル)

7. 時間スケーリングの分離と有限マルコフ連鎖の収束
 7. 1 マルコフ連鎖の収束定理
 7. 2 速い移住率を持つ Structured Coalescent モデル
 7. 3 分集団サイズの確率的変動を含む Structured Coalescent モデル
 7. 4 自殖を含む合祖モデル
 7. 5 種子の年齢を考慮に入れた合祖モデル
 7. 6 収束定理の証明
8. 無限サイトモデルの種々の結果と中立仮説の検定
 8. 1 サイト頻度と突然変異の年齢
 8. 2 分離サイト数と塩基多様度
 8. 3 Tajima's D テスト
 8. 4 その他の検定法

付 録

- (A) 指数分布・ポアソン分布・ガンマ分布
- (B) 超幾何関数
- (C) スターリング数
- (D) 包含と排除の原理
- (E) ヘルダーの不等式
- (F) 再帰性と非再帰性
- (G) ラプラス変換
- (H) 条件付き確率と条件付き期待値
- (I) マルコフ時間と強マルコフ性
- (J) ファインマン・カッツの公式