

## 論文審査の結果の要旨

申請者氏名 小谷野 仁

生物群集は複雑な構造を持っており、サンプリング調査に基づく母集団特性の偏りのない推定が不可欠となる。この課題が必ずしも容易でないことは、たとえば腸内細菌や土壤微生物の群集の多様性を測定することを想定すれば、理解される。異なる環境、異なる時点での調査は規模も異なることは避けられない。予め母集団の構造に対して強い仮定を置くことなく、データの持つ情報をゆるやかにまとめ上げることにより、頑健な推定をすることが可能となる。しかし残念ながら、現在のところ、数理統計学と確率論の理論に裏打ちされた方法論はほとんど未整備の状態である。そこで本論文は、生態学・疫学の調査データを最大限に生かす統計的方法を提案する。第1章において問題の背景と先行研究を概観し、全体の構成を示したのち、生物多様性の推定(第2章、第3章)、小地域問題(第4章)、個体群密度の推定(第5章)という、生態学と疫学における3つの問題について、数理統計学と確率論による理論的整備を行う。

### 1. 生物群集の遺伝的多様性の推定

生物群集の持つメタゲノムからサンプリングされた相同な遺伝子配列に基づき、群集を構成する種の遺伝的多様性を偏りなく推定する尺度を提案する。集団遺伝学における遺伝的多様度は無作為交配を理論基盤に置いているため、種間の分散の尺度としてそのまま適用するための理論的根拠を提供しない。そこで第2章において、群集の遺伝的多様度をコンセンサス配列からの配列の平均距離として得られる配列間の分散で定義する。配列間の距離として Levenshtein 距離 (2つの文字列の違いを説明する置換と挿入・欠失の回数の最小値) を採用する。ただし、群集は一般に多くの部分群集からなり、階層構造を持つ。部分群集の複雑な構造を頑健にとらえるために、近接配列をまとめ上げることにより部分群集を定義する。群集の遺伝的多様度を、部分群集の配列とその部分群集のコンセンサス配列からの距離の期待値とコンセンサス配列とそのコンセンサス配列の間の距離の期待値の和で遺伝的多様度を表現する。

Yellowstone 国立公園にある8つの温泉からサンプリングされた1068の16SrRNA配列からなる標本について、大きさ200, 400, 600, 800の部分標本をとり、上で定義した遺伝的多様度を比較したところ、標本サイズによらず全標本に基づく推定値の周りに偏りなく分布していることが示された。分子系統樹の総枝長は標本サイズとともに増加し、平均枝長は標本サイズとともに減少したのと対照的である。野外の極限環境として温泉、アルカリ湖、塩湖、南極、深海溝、深海穴フィールド、温度をコントロールされた環境としてヒトの口腔、胃、腸、ウシの第一胃における微生物群集の16SrRNA標本を調査した結果、とりわけ極低温が遺伝的多様度を下げることが観測された。さらに異なる群集間で類似配列の割合により遺伝的共通性を求めたところ、類似の環境では遺伝的共通性が高く、異なる環境下では共通性は低いことが確かめられた。唯一例外は、

アルカリ湖が南極および深海溝と共通性が高かったことである。

第3章では、第2章で提案された遺伝的多様性の尺度の理論基盤を与える。まず、塩基配列をはじめとする文字列のなす集合に、距離と確率構造が与えられ、独立性が定義される。その上で、標本配列のコンセンサス配列および標本分散が標本サイズとともに母集団のコンセンサス配列、母分散に確率1で概収束すること、すなわち文字列の確率空間上の大数の法則が証明される。

## 2. 小地域の推定

第4章では、形質に地域差があることが予想される中で、広範な地域の状況を推定する問題を扱う。比較的多数の小地域をサンプリングするが、それぞれの地域からの標本は十分な数を採れない。しかも標本サイズは小地域ごとに不ぞろいである。地域ごとに個別に標本平均を求める方式では精度のよい推定は達成できない。そこで、地域間の違いを気候や人口などの測定可能な要因に還元できる固定効果とそれでは説明できないランダムな効果に分解する。ランダムな効果は、小地域の標本平均から固定効果を差し引いた残差として推定する。従って、標本平均をランダムな効果の予測値で補正することにより、推定精度を高めることができる。補正の強さは、小地域の標本サイズ、および地域間分散と地域内分散の相対比に依存する。本章では、この比の推定量として Efron-Morris(EM)型推定量を採用することにより、ミニマックス性(実際の比の如何に関らず精度がこの推定量よりも上回るような推定量を他に得ることはできないこと)が保証されることを証明し、シミュレーションにより特に EM2 型推定量が推奨されることを示した。

この結果を踏まえて、アフリカにおけるマラリア感染とそれによる死亡のリスク評価を行った。マラリアの生態には温度と降雨量に大きく左右されるため、赤道地域と乾燥地域に分け、殺虫剤処理を行った蚊帳の普及度、屋内残留噴霧の普及度、効マラリア薬アルテミシニンによる治療などの処置をリスクに影響を与える固定効果として解析に取り入れた。アフリカには53の国が存在するが、このうち13カ国のデータが入手可能である。解析の結果、国別の集計値に比べ、感染のリスク、それによる死亡のリスク、ともに1割ないし2割、低い推定結果が得られた。

## 3. 個体群密度の推定

第5章では、コストを抑えつつ個体群密度を精度良く推定するための、妥当な調査規模について考察する。ここでは調査領域を次第に拡大し、観測数がある値に達したときに調査を終了させる方式を考える。観測数はポアソン過程に従う。調査のコストは調査の規模とともに増大する一方、密度の推定量の精度は向上する。この問題は、調査コストと推定誤差を加えた損失関数の期待値を最小化させる最適停止則と密度推定法を求める逐次決定の問題として定式化される。簡単のため、均質な空間分布を仮定する。半径を拡大させていく円状あるいは球状の調査領域の場合について、期待損失を最小化する許容的な決定則を得る。

以上のとおり、本論文は、生物多様性の尺度と推定、広範な地域の調査における小地域問題、個体群密度の推定という、生態学と疫学における3つの問題について、数理統計学と確率論による理論的整備を行った。今後モデルを改良することにより、様々な場面での適用が可能となるであろう。これらの研究は学術上および応用上価値が高く、よって、審査委員一同は本論文が博士(農学)の学位論文として価値あるものと認めた。