

論文の内容の要旨

論文題目 Development, test and application of a new intra-taxon
sampling method based on geographic information
(空間情報に基づく種内新規サンプリング手法の開発・検証・適用)

氏名 青木 聡志

本要旨は三部構成になっており、「A. サンプリング法の研究」、A の実験の妥当性の検証法を開発した「B. 効果量の研究」、開発したサンプリング手法を実用した「C. 適用研究」からなる。

A. サンプリング法

1. 研究背景

無作為抽出は統計学上の重要な仮定であるが、集団内あるいは分類群間のサンプリング手法はよく研究されているのに対し、野生分類群のサンプリング地点を決める分類群内のサンプリング手法の研究はほとんど存在しない。これまで植物分類学分野では、地図上の分布地点群からなるべくムラなく広い地域をカバーするように地点を選んで採取するサンプリング方法が一般的であり、客観的な方法がなかった。これは分類群の遺伝的多様度や形質を統計的に推定する上で仮定となるサンプリングが主観的に行われていることを意味し、データの正確性が損なわれている。そこで本研究では、野生分類群を対象とした客観的な標本抽出法を開発することを目的とする。

2. 定義・ソフトウェア・アルゴリズム

空間抽出という抽出方法を開発した。空間抽出は「分類群内で遺伝的に最も多様な集団を得るサンプリングが最適である」という生物多様性の解明を目指す分類学の考えに基づき、「同分類群個体は地理的に近いほど遺伝的にも近い」という集団遺伝学で立式されている仮定を利用する。この地理的隔離の仮定から、サンプリング候補地点間の距離から遺伝的相関を求める。この遺伝的相関の総和を比較することで、訪問可能な地点数内で遺伝的相関最小の地点群を最も多様な地点群として抽出する。さらに、空間抽出では一部地点でのサンプリングの成否を以後のサンプリング計画に反映させて採取地点を再計算する広義の adaptive sampling を行うことができる(図 1)。

空間抽出を行うために Samploc というソフトウェアを作成した。Samploc は n 個のサンプリング候補地点情報を取り込み、訪問可能な地点数 m と探索回数 r を入力することでサンプリング地点を算出する。このとき可能なサ

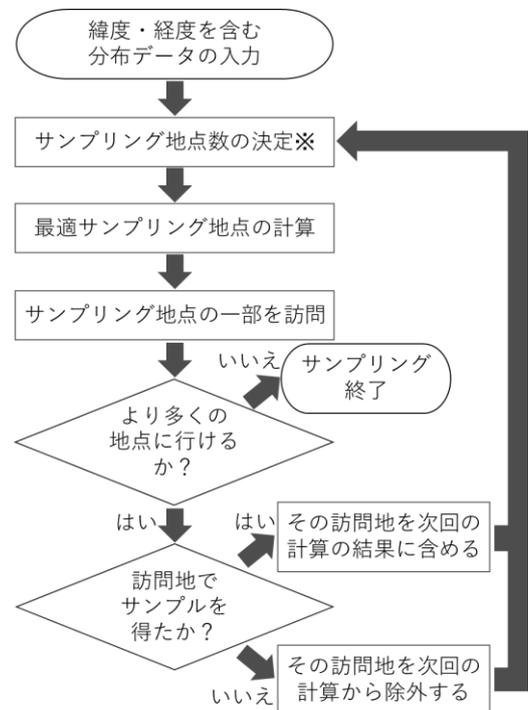


図 1.空間抽出の過程。※サンプリング地点数は研究の予算・時間で決定。

ンプリング地点の組み合わせは ${}_nC_m$ 通りあり、 ${}_nC_m \leq r$ かつ ${}_nC_m \leq 2147483647$ (32bit 符号つき整数の最大値)の場合、Samploc は総当たり法ですべての組み合わせの中から最適地点群を探索する。上記の条件に当たらない場合、Samploc はヒューリスティクス的一种である焼きなまし法で r 個の組み合わせを探索し、準最適地点群を与える。

3. 検証方法

既存の 20 分類群の分類・系統地理研究のデータを利用し、空間抽出の有効性を検証した。既存研究のサンプリング地点を母集団とみなし、母集団から 20%の地点数ごとに空間抽出と無作為抽出で地点群を再抽出してハプロタイプ多様度と塩基多様度を比較した。抽出率毎に、無作為抽出は 20 回反復、空間抽出は焼きなまし法を使っていれば 20 回反復、総当たり法なら 1 回探索した最適値で比較を行った。空間抽出の仮定が成立すれば、空間抽出の方がより高い多様度を示すと期待される。

4. 検証結果・考察

空間抽出の有効性は過半数の比較で示された。多様度が増加しなかった場合は、分布域に比して高い分散能力などの理由が考えられた。空間抽出は単純な仮定に基づいており、これに基づいた統計量は未開発であるが、野生分類群を扱う領域での統計学的基礎となるものと期待される。

B. 効果量

1. 研究の動機・背景

先の多様度比較実験はサンプルサイズを任意に大きくできるため、仮設検定を用いるのは不適切である。何故なら、サンプルサイズを大きくすれば、僅かな母平均の差でもそれが 0 でなければ有意差として検出できてしまうからである。この場合、サンプルサイズによらずに 2 群の違いを評価する効果量という統計量を利用できる。

ここで先の実験データの特性を考える。焼きなまし法を使う場合、空間抽出は準最適な地点群を抽出するため抽出を反復しても同じような地点を与え、データの分散は小さくなり、時に 0 になる。一方、無作為抽出にはこのような性質はない。よって、空間抽出と無作為抽出のデータを比較する際に等分散性を仮定するのは現実に即さない。焼きなまし法を使わない場合、空間抽出は最適地点群をただ 1 組与える。これを反復した無作為抽出のデータと比較する際は無作為抽出の 1 群と空間抽出 1 定数との比較とみなせる。ところが、既存の 2 群の平均値差の効果量は 2 群の等分散性を仮定しており、既存の 1 群の平均値と 1 定数との差の効果量は不偏統計量が導出されていなかった。

2. 考案した新規効果量・考察

等分散性を仮定しない 2 群の差の効果量とその不偏推定値 e を定義し、1 群と 1 定数との差の効果量の不偏推定量 c を示した。また、これらの効果量やその信頼区間を算出するための R のパッケージを作成、公開した。効果量 e は Welch の t 検定に対応するため、非常に広い適用範囲があると考えられる。これは近年 Student の t 検定よりも Welch の t 検定が推奨されているためである。効果量 c の適用範囲は比較的狭いが、対応のある 2 群の差(2 群の差と定数 0)や物理学における 1 群と物理定数との差、シミュレーション実験における 1 群と最適値との差などに適用できる。

C. 空間抽出の適用研究

1. 研究の動機・背景

開発した空間抽出に問題点や改良すべき点がないか調べるため、空間抽出を実用した小規模な分類研究を行った。イラクサ科 *Nanocnide* 属は東アジアから東南アジアの暖温帯から熱帯にかけて分布する草本のみからなる属で、3種または4種が認められている。*Nanocnide japonica* は属のタイプで、中国・台湾・韓国・日本に分布している。*N. zhejiangensis* は中国固有種である。琉球から記載された *N. lobata* と、中国で報告された *N. pilosa* に関しては分類見解に対立がある。Tateishi(2006)は *N. lobata* と *N. pilosa* の両者を認め、前者は南西諸島とインドシナに分布し、後者は中国および鹿児島県に分布するとしている。一方、Jiarui et al. (2003)は *N. pilosa* を *N. lobata* のシノニムとみなし、*N. lobata* が日本、中国、ベトナムに産するとしている。国内では立石(2006)の見解が採用されていて、*N. pilosa* は環境省指定絶滅危惧 IA 類として扱われている一方、国外では Jiarui et al. (2003)の見解が分類データベースなどで広く採用されている。この矛盾の解消が本研究の主目的である。既存の分子系統解析では、*N. lobata sensu* Tateishi と考えられるものは1個体しか使われておらず、*N. pilosa* と *N. lobata sensu* Tateishi の系統関係は明確になっていなかった。

2. 方法

サンプリング後に発表された *N. zhejiangensis* を除く *Nanocnide* 属各種のサンプルを、Samploc を使いながら国内および台湾・中国から集め、核 ITS 領域および MIG-seq 法による系統解析を行った。そして形態観察に基づく記載文とともに、系統解析に基づいた分類見解を示した。

3. 結果・考察

系統解析の結果、ITS・MIG-seq とともに *N. lobata sensu* Tateishi の単系統性を支持したものの、*N. pilosa* は側系統群となる(ITS)か、支持の低い単系統群となる(ブートストラップ値 68、MIG-seq)にとどまった。このため分岐学的種概念に基づき、*N. pilosa* を *N. lobata* の亜種とする分類見解を示した。Samploc を実用した結果、標本情報の信頼性による選別を行い、その結果をデータに示しておく必要性やサンプリング地点計算の新たな効率化方法、採取地点数の決め方の問題が明らかになった。