

東京理科大学 創域情報学部
情報理工学科 嘱託助教
もも ざき とも たか
桃崎 智隆

東京理科大学
理工学部 情報科学科 (2018年度入学) (現 創域情報学部 情報理工学科)
理工学研究科 情報科学科専攻 (2020年度修士課程修了) (現 創域理工学研究科)
理工学研究科 情報科学科専攻 (2023年度博士課程修了) (現 創域理工学研究科)

数字で守る海の命： 統計家が導く持続可能な海洋生態系

1. はじめに

このたびは、第19回学術奨励賞という荣誉ある賞を賜り、また本誌への執筆機会をいただき、心より御礼申し上げます。本稿では、私がノルウェー王立海洋研究所と進めてきた国際共同研究について、統計学になじみのない読者の方々にもできるだけ親しみを持っていただけるよう、噛み砕いてご紹介いたします。

統計学は、世の中にあふれるさまざまなデータから新しい科学的知見を見出すための分析手法を創り出す学問です。データを扱う実務家にとって解釈しやすい結果を返すこと、誰でも使える解析手法として広く実装されること、そして数学に裏打ちされた信頼性を備えること。この三つを両立する手法を生み出すことが、私たち統計家の使命だと考えています。

私の研究テーマの1つは、クジラの個体数と、それを取り巻く環境因子（天気・風力・視界・餌の資源量など）との関係性を分析するための統計手法の開発でした。日本と同じくノルウェーでも商業捕鯨が行われており、国際捕鯨委員会が定める捕獲上限を遵守しつつ、資源を持続可能に管理していくためには、さまざまな要因がクジラの観測数にどう関係するのかを常に把握する必要があります。そこで、ノルウェー王立海洋研究所が長年蓄積してきた目視調査データを統計的に分析する手法を、共同研究として開発しました。

2. 海洋調査データに潜む統計的課題

海洋調査データには、統計的に扱いにくい二つの特徴があります。一つ目は、性質の異なる変数が混在していることです。クジラの観測数のような離散データ、

餌の資源量のような連続データ、天気のようなカテゴリカルデータが同時に記録されています【表1】。連続データ同士であれば「相関係数」という馴染みのある指標で関係性を測れますが、種類の異なるデータの間ではそれが素直には使えません。

二つ目は、空間相関の存在です。海の上で観測されたデータは、近い場所どうしほど似た値になりやすいというのは、直感的にも自然な現象です。しかし統計的に見ると、この「位置的な近さによる類似性」は、本来分析したい「要因間の本質的な関係性」を歪めてしまう厄介な性質でもあります。

実際、空間相関を無視して従来の手法で分析を行うと、要因間の実際の関係性を大きく外してしまうことが、私たちの数値実験からも確認されました。海洋資源管理のように、誤った判断が漁業や生態系に直接影響する場面では、空間相関を適切に処理することが極めて重要となります。

3. 空間相関を考慮した順位による新しい分析手法

そこで本研究では、異なる種類のデータを同じ土台で扱い、かつ空間相関を考慮したうえで要因間の関係

【表1】 クジラ目視調査データの例。連続データ（緯度・経度・餌の資源量）、離散データ（観測数・風力）、カテゴリデータ（天気）が同一の調査記録に混在することを示す（数値はすべて架空例）。

ID	緯度	経度	餌の資源量	鯨の観測数	風力	天気
1	71.2	18.7	1.82	14	5	晴
2	71.2	19.0	1.06	12	4	晴
3	70.8	20.1	0.21	0	2	雨
⋮	⋮	⋮	⋮	⋮	⋮	⋮

性を分析できる新しい手法を開発しました¹⁾。その発想は、大きく三つのステップから成ります。

ステップ1：多変量正規分布の活用。 観測されたデータの裏側には、扱いやすい数学的性質を持つ「多変量正規分布」に従う潜在的な数値があると仮定します。この分布の「相関行列」と呼ばれる量が、要因間の関係性を素直に表してくれます。

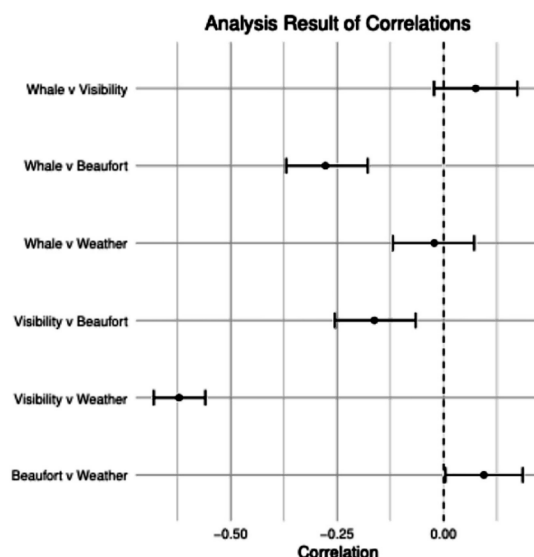
ステップ2：順位の活用。 ところが、ステップ1で仮定した多変量正規分布の数値は**潜在的なもので**、実際には直接観測できません。そこで橋渡し役となるのが**順位**です。観測データの順位は、その裏側にある潜在的な多変量正規分布の順位と等しいという性質があり、これを使えば観測データの順位から相関行列（要因間の関係性）を求められるのです。さらに順位を用いることで、連続データと離散データを同じ枠組みで扱えるという利点もあります。たとえば四つの観測値(1.82, 1.06, 0.21, 0.43)は順位に直すと(1, 2, 4, 3)となり、整数か実数かといった種類の違いが消え、共通の物差しで比較できます。

ステップ3：空間相関の取り込み。 ここまでのステップ1とステップ2では、観測地点どうしの位置関係は扱えていません。そこで、ステップ1の多変量正規分布のなかに、観測地点どうしの位置関係を反映する空間相関の構造を新たに導入しました。これは「近い地点ほど似た値になりやすい」という事実をモデルが自然に取り込む仕組みで、その効果の範囲も一緒に求められる設計になっています。こうして、**空間相関を考慮した順位による新しい分析手法**として、種類の異なるデータが混在し、かつ空間相関を持つ調査データに対しても、要因間の本質的な関係性を取り出せるようになりました。

4. バレンツ海のミンククジラ調査への応用

開発した手法を、ノルウェー王立海洋研究所が提供するバレンツ海（スヴァールバル諸島の南に広がる海域）におけるミンククジラを目視調査データに適用しました。データには、ミンククジラの観測数、視界良好度、風力階級、天気各要因が含まれます。分析の結果、次のような関係性が明らかになりました【図1】。

- ・**風力とクジラの観測数：負の相関。** 風が弱いときほどクジラはよく観測されます。
- ・**天気とクジラの観測数：相関はほぼなし。** 天気の良し悪しは観測数に大きく影響しません。
- ・**視界とクジラの観測数：やや正の相関。** 視界が良好なほどクジラを見つけやすい傾向があります。



【図1】 提案手法（spBGC）によるバレンツ海のミンククジラ調査データの解析結果。Whale=クジラの観測数、Visibility=視界良好度、Beaufort=風力階級、Weather=天気。各行は2変数間の相関係数（横軸）を示し、エラーバーは推定の幅、点線は相関なし（0）を表す。

風力や視界は海面の状態と関連が深く、観測のしやすさそのものに影響するため、こうした結果は生物学的にも調査運用上も妥当な解釈と整合します。本手法は、生態調査の指針づくりや資源管理にあたり、観測条件の影響と本質的な生態学的関係を切り分けて議論するための土台を提供してくれます。

5. おわりに

本研究で開発した手法は、海洋生態学に限らず、医療データや遺伝子データなど「種類の異なる変数が、空間的（あるいは時間的）に近い対象どうしで似た値を取りやすい」データであれば、幅広く応用が可能です。今後は計算アルゴリズムのさらなる高速化と、より複雑な従属構造に対する手法の開発に取り組んでいきたいと考えています。

最後に、本研究を導いてくださった共同研究者の中川智之先生、菅澤翔之助先生、そして貴重なデータと知見をご提供いただいたノルウェー王立海洋研究所の Solvang Kato Hiroko 先生に、心より感謝申し上げます。また、東京理科大学の関係各位、ならびに本稿執筆の機会を賜りました理窓博士会・科学フォーラム編集委員会の皆様にも深く御礼申し上げます。

参考文献

- 1) Momozaki, T., Nakagawa, T., Sugawara, S., and Solvang, H. K., Semi-parametric Copula Estimation for Spatially Correlated Multivariate Mixed Outcomes: Analyzing Visual Sightings of Fin Whales from a Line Transect Survey, *arXiv: 2312.12710*, 2023.