

ミンククジラの身体データを例とした粗密がある空間データでの Fused Lasso による空間効果の推定

福井 敬祐 (大阪医科大・研究支援センター), 山村 麻理子 (広島大・教育),
柳原 宏和 (広島大・理), Solvang Hiroko; Nils Øien; & Tore Haug
(Institute of Marine Research, Marine Mammals Research Group, Norway)

空間データの解析において、分析対象とする空間の情報を満遍なく捉えられず、ある空間部分に偏って観測値が取られているデータがある。例として北東大西洋のミンククジラの身体データがあり、クジラが捕獲された位置 (緯度経度) や体の部位の長さや太さが記録されている。クジラの生態を観測する上では、回遊している北東大西洋全体での身体状態の把握が望ましい。しかし漁はほぼ決まった位置で行われるため、北東大西洋全体での空間データとしては粗密が生じている。先行研究である Solvang *et al.* [1] や Yamamura *et al.* [2] では、空間効果を多項式によりモデル化し、粗密の影響を大きく受けない推定を行っているが、多項式では柔軟な推定結果を望むことができない。スプラインなど、基底関数を配置するノンパラメトリック平滑化を用いることで、空間効果のより正確な推定を期待できるが、粗密のある空間データでは、標本数の少ない空間での正確な空間効果の推定結果が望めない。そこで分析の対象となる空間をいくつかに細分化し、Fused Lasso を用いて細分化した空間の空間効果を結合しつつ推定することで、データの粗密の影響を受けにくい空間効果の推定方法を提案する。

分析の対象となる空間を m 個に細分化し、空間 j ($j = 1, \dots, m$) における i 番目の目的変数と説明変数の関係を表すモデルを次のとおりおく。

$$y_{ij} = f(\mathbf{x}_{ij}) + \mu_j + \varepsilon_{ij}, \quad (i = 1, \dots, n_j; j = 1, \dots, m). \quad (1)$$

ただし、 \mathbf{x}_{ij} は $p \times 1$ の位置に無関係な説明変数ベクトル、 ε_{ij} は互いに独立な誤差変数を表しており、 $E[\varepsilon_{ij}] = 0$, $\text{Var}[\varepsilon_{ij}] = \sigma^2$ である。また、 μ_j は空間効果を示す未知パラメータであり、 f は空間以外の効果を示す項である。空間効果 $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)'$ と空間以外の効果 f は、後退当てはめ法を用いて推定する。 \hat{f} を f のノンパラメトリック推定値とすれば、空間効果 $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)'$ は、次の罰則付き残差誤差平方和の最小化により推定することができる。

$$\text{PRSS}(\boldsymbol{\mu}|\hat{f}) = \sum_{j=1}^m \sum_{i=1}^{n_j} \left\{ y_{ij} - \hat{f}(\mathbf{x}_{ij}) - \mu_j \right\}^2 + \lambda \sum_{j=1}^m \sum_{k \in D_j} |\mu_j - \mu_k|. \quad (2)$$

ただし、 λ は非負の正則化パラメータ、 D_j は空間 j に隣接する空間を示す集合である。北東大西洋のミンククジラの身体データにを用いた推定結果を発表当日に紹介する。

引用文献：

- [1] Solvang, H. K., Yanagihara, H., Øien, N. & Haug, T. (2017). Temporal and geographical variation in body condition of common minke whales (*Balaenoptera acutorostrata acutorostrata*) in the Northeast Atlantic. *Polar Biology*, **40**, 667–683.
- [2] Yamamura, M., Yanagihara, H., Solvang, H. K., Øien, N. & Haug, T. (2016). Canonical correlation analysis for geographical and chronological responses. *Procedia Computer Science*, **96**, 1351–1360.