

# 生態データのためのポアソン点過程モデル —準線形モデリング—

福井大学 小森 理, 横浜市立大学 三枝 祐輔, 統計数理研究所 江口 真透

生物の生息分布を精度よく推定することは、その生物と環境との関係を理解するうえでもまた生物保全の観点からも重要なことである。しかしながら往々にして生物の生息に関するデータを収集する場合、その生物を誤って見逃す可能性が常にあるため、正確な存在不在 (presence-absence) のデータを得ることは網羅的な調査をする場合を除きかなり難しいのが現状である。そこで在データのみ (presence-only-data) から生物の生息分布の推定を行うことが近年盛んになってきた。代表的なものにエントロピー最大化分布を使った Maxent[3] がある。また推定精度向上のため生息分布とは直接関係のない、観測地点へのアクセスのしやすさ等のバイアス項を考慮することも必要となる。そこで今回我々はポアソン点過程の枠組みから生息分布の強度に対し以下のようなモデルを考えた。

$$\lambda_\tau(s) = \exp \left[ \frac{1}{\tau} \log \{ \pi \exp(\tau \boldsymbol{\beta}^\top \mathbf{x}(s)) + (1 - \pi) \exp(\tau \boldsymbol{\delta}^\top \mathbf{z}(s)) \} \right]$$

ただし  $s$  は観測地点、 $\mathbf{x}(s)$  はその観測地点での環境変数、 $\mathbf{z}(s)$  はバイアス変数、 $\pi$  は混合比、 $\boldsymbol{\beta}$ ,  $\boldsymbol{\delta}$  はそれぞれの係数とする。またこの  $\tau \in \mathbb{R}$  は形状パラメータであり、 $\lim_{\tau \rightarrow 0} \lambda_\tau(s) = \exp\{\pi \boldsymbol{\beta}^\top \mathbf{x}(s) + (1 - \pi) \boldsymbol{\delta}^\top \mathbf{z}(s)\}$  は通常のポアソン点過程モデルに帰着し、 $\lambda_0(s) = \pi \exp(\boldsymbol{\beta}^\top \mathbf{x}(s)) + (1 - \pi) \exp(\boldsymbol{\delta}^\top \mathbf{z}(s))$  は superposed ポアソン点過程に帰着し、 $\lim_{\tau \rightarrow \infty} \lambda_\tau(s) = \exp(\max\{\boldsymbol{\beta}^\top \mathbf{x}(s), \boldsymbol{\delta}^\top \mathbf{z}(s)\})$ 、 $\lim_{\tau \rightarrow -\infty} \lambda_\tau(s) = \exp(\min\{\boldsymbol{\beta}^\top \mathbf{x}(s), \boldsymbol{\delta}^\top \mathbf{z}(s)\})$  となるため、提案モデルは従来のモデルの拡張ととらえることができる。このようなモデルは準線形モデルと呼ばれ、Kolmogorov-Nagumo 平均を基にしたモデリングであり、すでにロジスティック回帰モデルの枠組みでいくつか提案されている [2 1]。また上記のモデルに対するポアソン点過程の対数尤度は、 $s_i$  をデータ地点、 $y_i$  を presence のとき 1、それ以外るとき 0 とする変数とすると ( $i = 1, \dots, n$ )

$$\tilde{L}_\tau(\boldsymbol{\beta}, \boldsymbol{\delta}) = \sum_{i=1}^n w_i \left[ y_i \log(\lambda_\tau(s_i)) - \lambda_\tau(s_i) \right]$$

となり、これを最大化するようにパラメータ  $\boldsymbol{\beta}$ ,  $\boldsymbol{\delta}$ ,  $\tau$  を推定する。ただし  $w_i$  はあらかじめ定まっている重みである。このモデルは生息分布に直接関係のある環境変数と、擾乱項のバイアス変数を別々のクラスターとしてモデリングしている点が特徴であり、従来法では推定不能であったそれぞれの切片項をも同時に推定できる。いくつかのシミュレーションで提案法の有用性を示すとともに、実データ解析の結果も当日報告する。

## 参考文献

- [1] KOMORI, O., EGUCHI, S., IKEDA, S., OKAMURA, H., ICHINOKAWA, M. AND NAKAYAMA, S. (2016). An asymmetric logistic regression model for ecological data. *Methods in Ecology and Evolution* **7**, 249–260.
- [2] OMAE, K., KOMORI, O. AND EGUCHI, S. (2017). Quasi-linear score for capturing heterogeneous structure in biomarkers. *BMC Bioinformatics* **18**, 308.
- [3] PHILLIPS, S. J., ANDERSON, R. P. AND SCHAPIRE, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231–259.