Bull. Aust. Math. Soc. **93** (2016), 167–168
doi:10.1017/S0004972715000945

MIXING IT UP: NEW METHODS FOR FINITE MIXTURE MODELLING OF MULTI-SPECIES DATA IN ECOLOGY

FRANCIS K. C. HUI

(Received 15 June 2015; first published online 12 November 2015)

2010 *Mathematics subject classification*: primary 62P12; secondary 62F07, 62H86.

Keywords and phrases: statistics, parametric inference, multivariate analysis.

In this thesis, new methods are proposed for using finite mixture models to analyse multi-species data in ecology. Developments range from theoretical results to empirical studies, offering contributions to the literatures of finite mixture models, species distribution models, variable selection, cluster analysis and ordination.

To begin, a comparison on several real datasets demonstrates that mixture models offer better predictions of how species communities respond to the environment, compared to modelling species separately. This is achieved by borrowing strength across species—organisms with similar environmental responses are clustered together, forming a small number of *archetypal responses*. These results have appeared in [5].

A major challenge in applying mixture models generally is model selection, and two important contributions are made on this front. On how to choose the number of mixture components, complete likelihood information criteria (despite being a popular approach) are shown to potentially underfit the true number of components. As an alternative, a new observed likelihood information criterion is proposed, which is proven to be order consistent (see [3]). On how to choose the variables to enter into each component, two new penalties are proposed that exploit the grouped structure of covariates in mixture of regression models, leading to desirable asymptotic and finite sample properties (see [4]).

The performance of all penalised likelihood methods depends critically on the choice of tuning parameter. In the case of adaptive LASSO regression, a new information criterion is proposed for tuning parameter selection that, unlike previous criteria, explicitly accounts for the effect of penalisation on the bias–variance trade-off. The proposed criterion is shown to outperform many currently used criteria in selecting the tuning parameter (see [2]).

Thesis submitted to the University of New South Wales in December 2014; degree approved on 23 February 2015; supervisors David Warton (UNSW) and Scott Foster (CSIRO).

Apart from prediction, multi-species data are commonly analysed using cluster analysis and unconstrained ordination, to study how species composition varies spatially. To this end, a model-based approach to unconstrained ordination is proposed using latent variable models in [1]. This approach is then integrated with finite mixture models to produce a unified framework for simultaneous clustering and ordination. Examples and simulation demonstrate the advantages of model-based approaches over distance-based methods.

The thesis concludes by discussing several extensions to the methods proposed, with further applications to multi-species data.

References

[1] F. K. C. Hui, S. Taskinen, S. Pledger, S. D. Foster and D. I. Warton, 'Model-based approaches to unconstrained ordination', *Methods Ecol. Evol.* **6** (2015), 399–411.

[2] F. K. C. Hui, D. I. Warton and S. D. Foster, 'Tuning parameter selection for the adaptive lasso using ERIC', *J. Amer. Statist. Assoc.* **110** (2015), 262–269.

[3] F. K. C. Hui, D. I. Warton and S. D. Foster, 'Order selection in finite mixture models: complete or observed likelihood information criteria?', *Biometrika* (2015), doi:10.1093/biomet/asv027.

[4] F. K. C. Hui, D. I. Warton and S. D. Foster, 'Multi-species distribution modeling using penalized mixture of regressions', *Ann. Appl. Stat.* **9**(2) (2015), 866–882.

[5] F. K. C. Hui, D. I. Warton, S. D. Foster and P. K. Dunstan, 'To mix or not to mix: comparing the predictive performance of mixture models versus separate species distribution models', *Ecology* **94** (2013), 1913–1919.

FRANCIS K. C. HUI, School of Mathematics and Statistics,
Faculty of Science, CSIRO Digital Productivity Flagship,
University of New South Wales,
Sydney, NSW 2052, Australia
e-mail: fhui28@gmail.com

https://doi.org/10.1017/S0004972715000945 Published online by Cambridge University Press