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).

© 2015 Australian Mathematical Publishing Association Inc. 0004-9727/2015 $16.00

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


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