When Ecological individual heterogeneity models and large data collide: An Importance Sampling approach

Blanca Sarzo[1,2], *Ruth King*[2], *Víctor Elvira*[2]

[1]Blanca.Sarzo@uv.es, Institut Cavanilles de Biodiversitat i Biologia Evolutiva, University of Valencia
[2]School of Mathematics and Maxwell Institute for Mathematical Sciences, University of Edinburgh

We consider the challenges that arise when fitting ecological individual heterogeneity models to large data sets. In particular, we focus on (continuous-valued) random effect models commonly used to describe individual heterogeneity present in ecological populations within the context of capture-recapture data, although the approach is more widely applicable to more general latent variable models. Within such models, the associated likelihood is expressible only as an analytically intractable integral. Common techniques for fitting such models to data include, for example, the use of numerical approximations for the integral, or a Bayesian data augmentation approach. However, as the size of the data set increases (i.e. the number of individuals increases), these computational tools may become computationally infeasible. We present an efficient Bayesian model-fitting approach, whereby we initially sample from the posterior distribution of a smaller subsample of the data, before correcting this sample to obtain estimates of the posterior distribution of the full dataset, using an importance sampling approach. We consider several practical issues, including the subsampling mechanism, computational efficiencies (including the ability to parallelise the algorithm) and combining subsampling estimates using multiple subsampled datasets. We initially demonstrate the feasibility (and accuracy) of the approach via simulated data before considering a challenging real dataset of approximately 30,000 guillemots (*Uria aalge*), and, using the proposed algorithm, obtain posterior estimates of the model parameters in substantially reduced computational time compared to the standard Bayesian model-fitting approach.

Keywords: Capture-recapture, intractable likelihood, random effects