59

Large Data and Complex Ecological Models: When Worlds Collide

We consider the challenges that arise when fitting complex ecological models to "large" data sets. In particular we consider capture-recapture data where random effect (or hierarchical) models are commonly used to represent individual heterogeneity, often present in ecological populations under study. In general these models lead to a likelihood that is expressible only as an analytically intractable integral. Common techniques for fitting such models to data include, for example, the use of direct 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 multiple subsampling estimates using multiple subsampled datasets. An empirical simulation study for an individual heterogeneity capture-recapture model demonstrates the approach provides reliable posterior estimates but at a substantially reduced computational cost compared to a standard Bayesian data augmentation approach. The method is subsequently applied to a population of approximately 30,000 ringed guillemots, where we specify an additive logistic model for the survival probabilities allowing for age, time and individual heterogeneity. Fitting the model using standard Bayesian data augmentation techniques are infeasible, yet our new approach provides consistent posterior estimates in the order of a couple of hours (or less).