249

zoid: A mixture model for analyzing proportional data with zeros, ones, and overdispersion

Many ecological datasets are proportional, representing mixtures of constituent elements such as species, habitat types, or strains. Analyses of proportional data are challenged by categories with zero observations (zeros), cases where all observations fall into a single category (ones), and overdispersion. Conventional approaches to handling these challenges include aggregating data, removing observations, or manually adjusting values. An increasingly common approach avoids such ad-hoc adjustments by treating observations as arising from a mixture of statistical distributions. We describe and evaluate a mixture model (zoid) capable of handling observed data consisting of three possible categories: zeros, proportions, or ones. Instead of fitting data from single observational units (e.g., an individual organism), the model uses summed proportional contributions across units to estimate mixture proportions. Data arising as zeros or ones are modeled with binomial distributions and proportional observations are modeled using generalized beta distributions. Optional estimation of overdispersion and covariate influences expand model applications. We evaluate model performance, as implemented in Stan, using simulations and two ecological case studies. We show that zoid successfully estimates mixture proportions using simulated data with varying sample sizes and is robust to overdispersion and covariate structure. In the first case study, we model genetic information to estimate composition of a mixed-stock salmon fishery. In the second, we analyze stomach contents of Atlantic cod to illustrate how the model accommodates covariates and overdispersion. The zoid model represents a flexible tool for modeling proportional data in which zeros, ones, or overdispersion are expected a priori. Our implementation of the model as an R package, with corresponding descriptions, data examples, and user-specified model parameters, should facilitate its application to varied ecological datasets composed of proportional observations.