

Stratification is often used in multiple-recapture estimation for both producing estimates at the sub-population level, and for reducing the effects of heterogeneity of capture probabilities. When stratification variables include records with missing labels, researchers often either exclude the missing rows or replace the missing labels with a made up category. These approaches risk the introduction of important biases in estimates. A more principled approach is to use missing data techniques, like multiple imputation or joint modeling. These approaches can leverage the capture patterns themselves and, if available, auxiliary multivariate individual-level covariates, to estimate the missing stratification distribution. Here I present a fully Bayesian joint modeling approach that can directly fit multiple-recapture data with incomplete multivariate stratification data and incomplete multivariate auxiliary information. The approach combines Dirichlet process mixture models for each stratum's capture patterns, together with another Dirichlet-process mixture for the joint vector of strata and covariates, forming a full joint model. Estimation is performed by MCMC sampling, using both sample- and data-augmentation approaches that leverage special structures, resulting in a highly scalable algorithm. I apply this method to the problem of estimating the number of fatal victims of the Peruvian internal armed conflict between 1980 and 2000. This dataset documents about 35,000 observations obtained through record-linkage of 7 lists. Records include two stratification variables and 5 extra auxiliary discrete covariates, all of them with missing entries.