Multiple Systems Estimation for sparse data, especially for quantifying the prevalence of Modern Slavery

Bernard W. Silverman

Multiple Systems Estimation is a key estimation approach for hidden populations such as the number of victims of Modern Slavery. The UK Government estimate (which I produced during the time I was Chief Scientific Adviser to the UK Home Office) of 10,000 to 13,000 victims was obtained by a multiple systems estimate based on six lists. A stepwise method was used to choose the terms in the model.
Further investigation shows that a small proportion of models give rather different answers, and that other model fitting approaches may choose one of these. Three data sets collected in the Modern Slavery context, together with a data set about the death toll in the Kosovo conflict, are used to investigate the stability and robustness of various Multiple Systems Estimate approaches. The crucial aspect is the way that interactions between lists are modelled, because these can substantially affect the results. A distinctive feature of data collected in this context that it is typical for the data to be sparse in that that many of the possible combinations of lists will have zero observed count.
I have investigated two approaches.   A Markov Chain Monte Carlo Bayesian approach is gives robust and stable results at least for the examples considered. Looking a bit deeper, this problem is an example of sparse contingency tables and, as has been noted by Rinaldo and Fienberg, for example, standard generalised linear model packages do not properly check for the existence and uniqueness of maximum likelihood estimators. Aspects considered include a correct way of finding maximum likelihood estimates if they exist; a stepwise approach to parameter fitting that does not rely on inappropriate information theoretic approaches; and implementation of a criterion for the existence of maximum likelihood estimates using insights that allow for very large numbers of models to be checked.

References (all available from www.bernardsilverman.co.uk)

Lax Chan, Bernard W. Silverman & Kyle Vincent. Multiple Systems Estimation for Sparse Capture Data: Inferential Challenges when there are Non-Overlapping Lists.

Lax Chan, Bernard W. Silverman & Kyle Vincent. SparseMSE: Multiple systems estimation for sparse capture data. R package.

Bernard W. Silverman.   Model fitting in Multiple Systems Analysis for the quantification of Modern Slavery: Classical and Bayesian approaches.