

Multiple systems estimation in the presence of censored cells

Ruth King¹; Oscar Rodriguez de Rivera Ortega²; Rachel McCrea²

- ¹ School of Mathematics, University of Edinburgh, UK
- ² School of Mathematics, Statistics and Actuarial Science, University of Kent, UK

Abstract:

Multiple systems estimation focuses on the estimation of hidden population sizes. Applications range from the number of people who inject drugs in an area, civilian casualties within wars, webpages on a given topic or modern day slaves. Data are collected on the given population via a series of lists that partially observe individuals in the population. Assuming that individuals are uniquely identifiable, lists can be cross-classified and the corresponding data summarised via an incomplete contingency table, providing the number of unique individuals observed by each distinct combination of lists. The table is incomplete as the number of individuals not observed by any of the lists is not observable.

Within this presentation we consider multiple systems estimation for contingency table data, where the observed cell entries may be censored and only presented to lie within a given interval. We describe how we can analyse such data using log-linear models and obtain estimates of the corresponding total population size taking into account the interval censoring. We compare this approach to some form of bounded approach, by setting the censored cell entries to the lower and upper limits of the interval. For the dataset that we consider we demonstrate that even for relatively small specified intervals, the estimate of the total population size, and selected log-linear model, is sensitive to how these censored cells are dealt with in the analysis.

Keywords:

Incomplete contingency table; log-linear model; maximum likelihood estimate