Misidentification in mark-recapture: have you got the moves?

M. R. Schofield[a,b] and **S. J. Bonner**[a,c]

[a]Department of Statistics
University of Kentucky
Lexington, KY, USA

[b]matthew.schofield@uky.edu
[c]simon.bonner@uky.edu

Keywords: Misidentification; Null basis; Markov basis

Abstract: Misidentification in mark-recapture studies can lead to biased estimation and inaccurate decision making. Link et al. (2010) proposed a solution when the observed data can be expressed as a linear function of known configuration matrix $\boldsymbol{A}$ and latent (correctly identified) data $\boldsymbol{x}$ with corresponding model $f(\boldsymbol{x}|\theta)$. Fitting the model via MCMC is challenging since $\boldsymbol{x}$ must satisfy the linear constraint. Link et al. overcame this difficulty by (i) finding a set of vectors (called moves) that form a basis for the null space of $\boldsymbol{A}$, and (ii) using these moves one-at-a-time to go between legitimate values of $\boldsymbol{x}$. However, we give examples that show that the approach of Link et al. may not be sufficient to produce an irreducible Markov chain; there may be (at least) two data vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ that we cannot transition between when applying the moves one-at-a-time, yet both satisfy the linear constraint. To solve this problem, we consider the notion of a Markov basis; a larger set of vectors (moves) that form a spanning set for the null space of $\boldsymbol{A}$ that ensure irreducibility of the Markov chain when using one-at-a-time to update $\boldsymbol{x}$. We illustrate the use of a Markov basis for the examples considered earlier.

References

Link, W. A., Yoshizaki, J., Bailey, L. L., and Pollock, K. H. (2010), "Uncovering a latent multinomial: analysis of mark-recapture data with misidentification," *Biometrics*, 66:178-185.