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

M. R. Schofield\textsuperscript{a,b} and S. J. Bonner\textsuperscript{a,c}

\textsuperscript{a}Department of Statistics
University of Kentucky
Lexington, KY, USA
\textsuperscript{b}matthew.schofield@uky.edu
\textsuperscript{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 $A$ and latent (correctly identified) data $x$ with corresponding model $f(x|\theta)$. Fitting the model via MCMC is challenging since $x$ 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 $A$, and (ii) using these moves one-at-a-time to go between legitimate values of $x$. 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 $x_1$ and $x_2$ 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 $A$ that ensure irreducibility of the Markov chain when using one-at-a-time to update $x$. We illustrate the use of a Markov basis for the examples considered earlier.

**References**