



Approximating the full likelihood for marginal $2 \times J$ contingency tables and case-control data.

Markus Chagas Stein*

Department of Statistics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil; and
Department of Statistics, University of Auckland, Auckland, New Zealand - m.stein@auckland.ac.nz

Chris Wild

Department of Statistics, University of Auckland, Auckland, New Zealand - c.wild@auckland.ac.nz

Alastair Scott

Department of Statistics, University of Auckland, Auckland, New Zealand - a.scott@auckland.ac.nz

Abstract

In this work, we consider likelihood-based approaches for constrained two-way contingency tables. Very often only the marginal distributions of variables of interest are available and inferences for unit-record effects are subject to a range of biases. Examples of such problem appear in so-called "ecological inference" and disclosure limitation. In this case, the only reliable way of reducing bias and the impact of uncheckable assumptions is supplementing the marginal distributions by other, e.g. individual-level, information. Our interest is in case-control sampling as auxiliary information to the group-level distributions. Maximum likelihood estimation can be a very challenging task, because its calculations require the enumeration of all tables consistent with the observed data. However, good approximations to the full likelihood can be carried out by sampling possible tables. We compare inferences via true likelihood in $2 \times J$ tables and estimated likelihoods, through uniform sampling methods and an informative sampling scheme based on independent binomial distributions. Preliminary results show that large improvements in computational time can be obtained for minimal losses in efficiency of estimation.

Keywords: Ecological inference; Contingency tables; Case-control sampling; Maximum likelihood estimation.