



An empirical likelihood approach under cluster sampling with missing observations

Yves G. Berger¹

Received: 12 September 2017 / Revised: 22 June 2018 / Published online: 3 August 2018
© The Institute of Statistical Mathematics, Tokyo 2018

Abstract

The parameter of interest considered is the unique solution to a set of estimating equations, such as regression parameters of generalised linear models. We consider a design-based approach; that is, the sampling distribution is specified by stratification, cluster (multi-stage) sampling, unequal selection probabilities, side information and a response mechanism. The proposed empirical likelihood approach takes into account of these features. Empirical likelihood has been mostly developed under more restrictive settings, such as independent and identically distributed assumption, which is violated under a design-based framework. A proper empirical likelihood approach which deals with cluster sampling, missing data and multidimensional parameters is absent in the literature. This paper shows that a cluster-level empirical log-likelihood ratio statistic is pivotal. The main contribution of the paper is to provide the rigorous asymptotic theory and underlining regularity conditions which imply \sqrt{n} -consistency and the Wilks's theorem or self-normalisation property. Negligible and large sampling fractions are considered.

Keywords Design-based approach · Estimating equations · Stratification · Side information · Unequal probabilities

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10463-018-0681-x>) contains supplementary material, which is available to authorized users.

✉ Yves G. Berger
Y.G.Berger@soton.ac.uk

¹ Southampton Statistical Sciences Research Institute, University of Southampton, Southampton SO17 1BJ, United Kingdom