Statistica Sinica **33** (2023), 1-14
doi:https://doi.org/10.5705/ss.202019.0410

PENALIZED JACKKNIFE EMPIRICAL LIKELIHOOD IN HIGH DIMENSIONS

Zhouping Li, Jinfeng Xu, Na Zhao and Wang Zhou

Lanzhou University, City University of Hong Kong,
The University of Hong Kong and National University of Singapore

Abstract: The jackknife empirical likelihood (JEL) is an attractive approach for statistical inferences with nonlinear statistics, such as U-statistics. However, most contemporary problems involve high-dimensional model selection and, thus, the feasibility of this approach in theory and practice remains largely unexplored in situations in which the number of parameters diverges to infinity. In this paper, we propose a penalized JEL method that preserves the main advantages of the JEL and leads to reliable variable selection based on estimating equations with a U-statistic structure in high-dimensional settings. Under certain regularity conditions, we establish the asymptotic theory and oracle property for the JEL and its penalized version when the numbers of estimating equations and parameters increase with the sample size. Simulation studies and a real-data analysis are used to examine the performance of the proposed methods and illustrate its practical utility.

Key words and phrases: Estimating equations, high-dimensional data analysis, jackknife empirical likelihood, penalized likelihood, U-statistics, variable selection.

1. Introduction

Statistical inference based on estimating equations with a U-statistic structure (U-type estimating equations) is common in nonparametric and semiparametric situations, such as quantile and rank regressions (Jin et al. (2003)). Suppose that observations $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random vectors, and the unknown parameters $\theta = (\theta_1, \ldots, \theta_p)^T$ can be estimated by solving the following r ($r \geq p$) estimating equations:

$$U_n(\theta) = \binom{n}{k}^{-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} h(X_{i_1}, \ldots, X_{i_k}; \theta) = 0, \tag{1.1}$$

where $h(\cdot) = (h_1(\cdot), \ldots, h_r(\cdot))^T$ are symmetric in $X = (X_1, \ldots, X_k)^T$ and satisfy $Eh(X_1, \ldots, X_k; \theta_0) = 0$ with θ_0, which denotes the true value of θ. Here, Σ

Corresponding author: Zhouping Li, School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China. E-mail: lizhp@lzu.edu.cn.