

Aggregating Estimating Equations by Approximation

Ruibin Xi

Washington University in St. Louis, USA

Nan Lin

Washington University in St. Louis, USA

Data warehouses provide on-line analytical processing (OLAP) tools for interactive analysis of multidimensional data to facilitate effective data mining at multiple levels of abstraction. We consider the problem of achieving efficient computations and storage by saving low-dimensional compression measures in lower-level cells of the database such that the desired computation at a higher level can still be done accurately without accessing the raw data. While Chen et al. [*IEEE T. Knowl. Data En.* 18 (2006):1585–1599] found some compression measures such that the ordinary least squares estimates of linear models can be losslessly computed, such lossless calculations are usually not achievable in more general nonlinear models. In this article, we consider estimating equation estimation and propose some nearly lossless compression measures based on linear approximations to the estimating equation. We proved the strong consistency of the estimator from the aggregated linearized estimating equations and also provide some large deviation results. Simulation studies also suggests the computational advantage of our proposed method, which supports its more general application to massive data sets to gain computational efficiency by partitioning the raw data set into many smaller subsets followed by our aggregation method.

[Nan Lin, 1 Brookings Dr., St. Louis, MO 63130, USA; nlin@math.wustl.edu]