Nonparametric Estimation on Regression Coefficient and Population Size for Incomplete or Skewed Data

Huazhen Lin[*]

Department of Biostatistics, University of Washington, USA

[*]**Corresponding author:** Huazhen Lin, Department of Biostatistics, University of Washington, USA, Tel: 206-543-1044; E-mail: linhz@swufe.edu.cn

Citation: Huazhen Lin (2016) Nonparametric Estimation on Regression Coefficient and Population Size for Incomplete or Skewed Data. J Biom Biostat 7: 303. doi: 10.4172/2155-6180.1000303

Opinion

Long-tail and skewed data are frequently encountered in both economics and health-care fields. Since the data on the distribution tail are scarce yet very important, inappropriately handling of such data can lead to unstable and biased estimation results. We have developed a series of methods to analyze such data. Particularly, we developed a method to estimate the transformation function and error distribution function so to avoid the difficulty of specifying these functions in modeling. In addition, invoking jointly the smoothing technology, penalty and rank correlation, we have developed a new dimension-free calculation method to quickly select the important risk factors from a large number of potential risk factors. Furthermore, we proposed a semi-parametric latent transformation model to combine multiple skewed and long-tail outcomes in a data-driven way. The analysis of real data showed that our methods are more efficient and robust than the existing methods to identify influential risk factors.

Research area of primary interest is semi-parametric efficient nonparametric estimation. Based on penalized local linear method, we have developed a series of nonparametric methods to identify and estimate the significant varying-coefficient and component function in the survival analysis models and the generalized linear regression modes, respectively. Our estimation methods have been shown to be semi-parametric efficient in the sense of Bickel (1993).

In the area of population size estimation, we have done the following work. Capture-recapture experiment is one of the most commonly used methods for the population size study. Incomplete and inaccurate covariates often present analytic challenges in the capture-recapture experiment. Considering all unobserved data as missing data, we developed a new method of estimating population size for the capture-recapture experiment. Our estimators have a closed form to enjoy computational simplicity. Simulation results showed that our approach is more efficient than the existing methods.

Using local linear smoothing in a log-linear model, we developed a data-based estimation of population size for multi-list data. Compared with the existing estimators, ours has much smaller variance, hence is more efficient.

Finally, delayed report often occurs in the collection of capture-recapture observations. Since the distributions of the delayed report and event numbers are difficult to be specified appropriately, we developed a method to estimate the population size when both the distributions of the delayed report and event numbers are unknown. Our estimator is computationally feasible, consistent and asymptotically normal. Real data examples and simulation results showed that our estimator is more robust and efficient than the existing ones.