Statistics and Probability Letters 140 (2018) 1–6

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A note on the equivalence of two semiparametric estimation methods for nonignorable nonresponse

Kosuke Morikawa [a,*], Jae Kwang Kim [b,c]

[a] Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan
[b] Department of Statistics, Iowa State University, Ames, IA 50011, USA
[c] Department of Mathematical Sciences, KAIST, Daejeon, 34141, Republic of Korea

ARTICLE INFO

Article history:
Received 30 October 2017
Received in revised form 20 March 2018
Accepted 31 March 2018
Available online 24 April 2018

MSC:
62D99
62F12

Keywords:
Empirical likelihood
Incomplete data
Nonignorable
Nonresponse
Not missing at random (NMAR)
Semiparametric estimation

ABSTRACT

We consider semiparametric estimation with nonignorable nonresponse data where only a parametric response model is assumed. We clarify the relationship of existing estimators and propose a new estimator which attains the semiparametric efficiency bound and is robust to model misspecification.

1. Introduction

Missing data problems are ubiquitous in many research areas, including econometrics, epidemiology, clinical study, and psychometrics. If analysts do not properly deal with missing data, then the results may be biased, which can lead to incorrect conclusions. Thus, a proper method for analyzing missing data needs to be developed. Also, it is preferable that the required assumptions in the proposed method be as weak as possible.

The required assumptions are strongly related to the outcome model or the response mechanism. In this paper, we focus on estimation with nonresponse data in which study variable is subject to missingness. Let y be the study variable, $\boldsymbol{x}$ be a fully observed d-dimensional covariate vector, and r be a response indicator of y, i.e., r takes the value 1 if y is observed, and takes the value 0 if y is missing. Thus, letting $\boldsymbol{z} = (\boldsymbol{x}^\top, y)^\top$, we observe $(\boldsymbol{x}, y)$ when $r = 1$, and observe only $\boldsymbol{x}$ when $r = 0$. The response mechanism is defined as the conditional probability $\pi(\boldsymbol{z}) = \Pr(R = 1 \mid \boldsymbol{z})$. If the mechanism does not depend on the study variable y, then it is called missing at random (MAR) and otherwise is called not missing at random (NMAR) (Little and Rubin, 2002; Kim and Shao, 2013). In the analysis of nonresponse data, MAR (NMAR) is also referred to as ignorable (nonignorable) missingness.

In this paper, we assume a parametric model on the response mechanism. Let $\pi(\boldsymbol{z}; \boldsymbol{\phi})$ be the parametric response model, where $\boldsymbol{\phi}$ is a q-dimensional parameter. In classical approaches for NMAR data, an outcome model $f(y \mid \boldsymbol{x})$, which is the

[*] Corresponding author.
E-mail addresses: morikawa@sigmath.es.osaka-u.ac.jp (K. Morikawa), jkim@iastate.edu (J.K. Kim).

https://doi.org/10.1016/j.spl.2018.03.020

conditional distribution of y given $\boldsymbol{x}$, is assumed in addition to the response model (Greenlees et al., 1982). This estimator has been criticized because of its sensitivity to model assumptions. Recently, some semiparametric methods, which do not require any outcome model, have been proposed.

Semiparametric estimation is mainly divided into two approaches: (i) the empirical likelihood (EL) approach; and (ii) the moment-based approach. Qin et al. (2002) derived a consistent and asymptotic normal estimator for $\boldsymbol{\phi}$ by using a technique of EL without using any outcome model. Kott and Chang (2010) proposed a moment-based estimator for $\boldsymbol{\phi}$, also without using any outcome model. Recently, Morikawa and Kim (2016) proposed two moment-based semiparametric adaptive estimators.

In this paper, we clarify the relationship between the EL estimator and the moment-base estimator, and show that there exists a specific case for which these two estimators are exactly the same. Also, we propose an estimation method that is robust to model misspecification. All technical details are given in the Supplementary Material (see Appendix A).

2. Previous semiparametric estimators

Let $\boldsymbol{z}_i = (\boldsymbol{x}_i, y_i)^\top$ $(i = 1, \dots, n)$ be independently and identically distributed realizations from unknown distribution $F(\boldsymbol{z})$, and r_i $(i = 1, \dots, n)$ be independently distributed taking binary values, either 0 or 1, with probability $\Pr(R_i = 1 \mid \boldsymbol{z}_i) = \pi(\boldsymbol{z}_i)$ for $i = 1, \dots, n$. Also, without loss of generality, assume that the first m elements are observed and that the remaining $(n-m)$ elements are missing in y_i, i.e., $r_i = 1$ for $i = 1, \dots, m$ and $r_i = 0$ for $i = m+1, \dots, n$. Qin et al. (2002) constructed the likelihood without using the data when $r = 0$ by

$$\prod_{i=1}^{m} \pi(\boldsymbol{\phi}; \boldsymbol{z}_i) dF(\boldsymbol{z}_i) \prod_{i=m+1}^{n} \int \{1 - \pi(\boldsymbol{\phi}; \boldsymbol{z})\} dF(\boldsymbol{z}) \tag{1}$$

and discretized the distribution F by w_i $(i = 1, \dots, m)$. The discretized distribution w_i can be estimated by maximizing $\prod_{i=1}^{m} w_i$ under the following constraints:

$$w_i \geq 0, \quad \sum_{i=1}^{m} w_i = 1, \quad \sum_{i=1}^{m} w_i \{\pi(\boldsymbol{\phi}; \boldsymbol{z_i}) - W\} = 0,$$

$W = \Pr(R = 1) = \int \pi(\boldsymbol{z}; \boldsymbol{\phi}_0) dF(\boldsymbol{z})$, and $\sum_{i=1}^{m} w_i \{\boldsymbol{h}(\boldsymbol{x}_i) - \bar{\boldsymbol{h}}_n\} = 0$, where $\boldsymbol{h} : \mathbb{R}^d \to \mathbb{R}^{p_1}$ $(p_1 \geq q-1)$ is an arbitrary function of $\boldsymbol{x}$, and $\bar{\boldsymbol{h}}_n = n^{-1} \sum_{i=1}^{n} \boldsymbol{h}(\boldsymbol{x}_i)$. The $\boldsymbol{h}(\boldsymbol{x})$ function helps to improve the efficiency. By introducing Lagrange multipliers, the solution to the above optimization problem is $\hat{w}_i^{-1} = m[1 + \boldsymbol{\lambda}_1^\top \{\boldsymbol{h}(\boldsymbol{x}_i) - \bar{\boldsymbol{h}}_n\} + \lambda_2 \{\pi(\boldsymbol{\phi}; \boldsymbol{z}_i) - W\}]$. By profiling out the unknown F with the estimates $\hat{w}_i$ $(i = 1, \dots, m)$ in (1) and taking the logarithm, we obtain the profile pseudo-loglikelihood:

$$\begin{aligned} &\ell(\boldsymbol{\phi}, W, \boldsymbol{\lambda}_1) \\ &= \sum_{i=1}^{m} \log \pi(\boldsymbol{\phi}; \boldsymbol{z}_i) - \sum_{i=1}^{m} \log[1 + \boldsymbol{\lambda}_1^\top \{\boldsymbol{h}(\boldsymbol{x}_i) - \bar{\boldsymbol{h}}_n\} + \lambda_2 \{\pi(\boldsymbol{\phi}; \boldsymbol{z}_i) - W\}] \\ &\quad + (n-m) \log(1 - W), \end{aligned} \tag{2}$$

where $\lambda_2 = (n/m - 1)/(1 - W)$. Qin et al. (2002) proposed a semiparametric estimator for $\boldsymbol{\phi}$ that maximizes the profile pseudo-loglikelihood. In the optimization procedure, some computational techniques are needed (see Chen et al., 2002) because the maximizer of (2) must satisfy $\hat{w}_i \geq 0$.

On the other hand, under the same assumptions, Kott and Chang (2010) proposed another semiparametric estimator that solves the following estimating equation:

$$\sum_{i=1}^{n} \left\{ \frac{r_i}{\pi(\boldsymbol{\phi}; \boldsymbol{z}_i)} - 1 \right\} \boldsymbol{g}(\boldsymbol{x}_i) = 0, \tag{3}$$

where $\boldsymbol{g} : \mathbb{R}^d \to \mathbb{R}^q$ is an arbitrary function of $\boldsymbol{x}$. This equation is called "calibration" equation in the literature of survey sampling. A typical choice for $\boldsymbol{g}$ when $d = 1$ is $\boldsymbol{g}(x) = (1, x, \dots, x^{q-1})^\top$. It is hard to decide the control variables in the calibration condition when $d > 1$. Also, when the dimension of $\boldsymbol{g}(\boldsymbol{x})$ is larger than q, say p_2, the model is over-identified and the generalized method of moments (GMM) method (Hansen, 1982) can be used to estimate $\boldsymbol{\phi}$. The GMM estimator can be constructed by

$$\hat{\boldsymbol{\phi}} := \arg\min_{\boldsymbol{\phi}} \sum_{i=1}^{n} \left\{ \frac{r_i}{\pi(\boldsymbol{\phi}; \boldsymbol{z}_i)} - 1 \right\}^2 \boldsymbol{g}(\boldsymbol{x}_i)^\top \hat{V}^{-1}(\boldsymbol{\phi}) \boldsymbol{g}(\boldsymbol{x}_i), \tag{4}$$

where $\hat{V}(\boldsymbol{\phi}) = n^{-1} \sum_{i=1}^{n} \{r_i / \pi(\boldsymbol{\phi}; \boldsymbol{z}_i) - 1\}^2 g(\boldsymbol{x}_i)^{\otimes 2}$ and $B^{\otimes 2} = BB^\top$ for any matrix B. The optimizations in (3) and (4) are much simpler than that of Qin et al. (2002) since there is no constraint in the optimization.