Journal of Statistical Planning and Inference 148 (2014) 101–110

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

A further study of the multiply robust estimator in missing data analysis

Peisong Han

Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, ON, Canada N2L 3G1

ARTICLE INFO

Article history:
Received 20 April 2013
Received in revised form
18 September 2013
Accepted 10 December 2013
Available online 17 December 2013

Keywords:
Augmented inverse probability weighting (AIPW)
Double robustness
Empirical likelihood
Extreme weights
Missing at random (MAR)
Propensity score

ABSTRACT

In estimating the population mean of a response variable that is missing at random, the estimator proposed by Han and Wang (2013) possesses the multiple robustness property, in the sense that it is consistent if any one of the multiple models for both the missingness probability and the conditional expectation of the response variable given the covariates is correctly specified. This estimator is a significant improvement over the existing doubly robust estimators in the literature. However, the calculation of this estimator is difficult, as it requires solving equations that may have multiple roots, and only when the appropriate root is used is the final estimator multiply robust. In this paper, we propose a new way to define and calculate this estimator. The appropriate root is singled out through a convex minimization, which guarantees the uniqueness. The new estimator possesses other desirable properties in addition to multiple robustness. In particular, it always falls into the parameter space, and is insensitive to extreme values of the estimated missingness probability.

1. Introduction

For estimation in the presence of missing data, double robustness is a desirable property due to its double protection on estimation consistency against model misspecification. An estimator is doubly robust if it is consistent when either the missingness probability or certain characteristic of the full data distribution, usually the conditional expectation of certain functions of the full data given the observed data, is correctly modeled. In the class of augmented inverse probability weighted (AIPW) estimators proposed and investigated by Robins et al. (1994), the one that is identified to be locally efficient has been found to be also doubly robust (Scharfstein et al., 1999), and has been advocated for routine use (Bang and Robins, 2005). Since then, a variety of doubly robust estimators have been proposed, including Tan (2006, 2008, 2010), Kang and Schafer (2007) and its discussion, Qin and Zhang (2007), Qin et al. (2008), Rubin and van der Laan (2008), Cao et al. (2009), van der Lann (2010), Tsiatis et al. (2011), Han (2012), and Rotnitzky et al. (2012). Most of these developments are based on estimation efficiency concerns. van der Laan and Gruber (2010) established the "collaborative" double robustness property of the estimators that solve the efficient influence curve estimating equation. Their results indicate that, consistency can still be achieved under certain combinations of dual misspecifications of the models.

Double robustness, however, still does not provide sufficient protection on estimation consistency in many practical studies, as it allows only a single model for the missingness probability and a single model for the conditional expectation. With an unknown data generating process, it is often risky to assume that one of these two models is correctly specified. To increase the likelihood of correct specification, multiple models may be postulated and fitted. The question then is how to

E-mail address: peisonghan@uwaterloo.ca

0378-3758/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2013.12.006

combine them into the estimation. See Robins et al. (2007) for some relevant discussion. In the setting of estimating the population mean of a response variable that is subject to ignorable missingness (Little and Rubin, 2002), Han and Wang (2013) proposed a weighted estimator that is consistent if any one of those multiple models is correctly specified, either for the missingness probability or for the conditional expectation. This multiply robust estimator apparently provides more protection on consistency than the doubly robust estimators.

One big challenge of Han and Wang's (2013) estimation procedure is that their calculation of the weight involves solving a set of equations that may have multiple roots, yet only when the appropriate root is used is the final estimator multiply robust. Although some ad hoc suggestions on the numerical implementation have been given, it is still very difficult in practice to single out the root that satisfies the technical assumptions made to guarantee the multiple robustness. In this paper, we provide a solution to the multiple-root problem by re-defining the estimator and proposing a new way of calculation. We define an objective function that is convex on a convex domain, so that a unique minimum exists for this objective function. We show that the unique minimizer is actually the appropriate root that should be used to calculate the weight in order to achieve multiple robustness. A Newton–Raphson algorithm searching for this minimizer is presented.

Our proposed estimator possesses other desirable properties in addition to multiple robustness. First, it has the "sample boundedness" property (Robins et al., 2007; Tan, 2010; Rotnitzky et al., 2012), in the sense that it always lies within the range of observed responses, and thus within the parameter space of the parameter of interest. Second, the performance of our estimator is not dramatically influenced by extreme values of the estimated missingness probability. It is well known in the literature that the sampling distribution of the AIPW estimators can be skewed and highly variable when the "inverse probability" weight itself is highly variable (Robins et al., 1995; Scharfstein et al., 1999, Robins and Wang, 2000). This occurs if, for some subjects whose response is observed, the estimated probability of observing the response is close to zero, causing those subjects to receive extremely large "inverse probability" weights compared to others. In such a case, through simulation studies, Kang and Schafer (2007) demonstrated that the AIPW estimators can have severe bias even if the models for both the missingness probability and the conditional expectation are only mildly misspecified. In recent literature, there have been many efforts trying to overcome this difficulty. See Kang and Schafer (2007) and its discussion, Cao et al. (2009), Tan (2010), and Rotnitzky et al. (2012). Because our proposed estimator is based on weight different from the "inverse probability", the impact of extreme values of the estimated missingness probability is mitigated.

2. A multiply robust estimator

To describe the multiply robust estimator in Han and Wang (2013), let Y denote the response variable, $\boldsymbol{X}$ the vector of covariates, and n the sample size. Let $R=1$ if Y is observed and $R=0$ if Y is missing. Assume that $P(R=1|Y=y, \boldsymbol{X}=\boldsymbol{x})=P(R=1|\boldsymbol{X}=\boldsymbol{x})$, the so-called missing at random mechanism (Little and Rubin, 2002), and denote this probability by $\pi(\boldsymbol{x})$. The observed data are independent and identically distributed triplets $(R_iY_i, \boldsymbol{X}_i, R_i)$ $(i=1,\ldots n)$. Let $m=\sum_{i=1}^n R_i$ be the number of subjects who have their response observed, and index these subjects by $i=1,\ldots,m$ without loss of generality. The goal is to estimate $\mu_0=E(Y)$.

The doubly robust AIPW estimator allows a single model for $\pi(\boldsymbol{x})$ and a single model for $a(\boldsymbol{x})=E(Y|\boldsymbol{X}=\boldsymbol{x})$, and has the form

$$\hat{\mu}_{\text{aipw}}=\frac{1}{n}\sum_{i=1}^n\left\{\frac{R_i}{\hat{\pi}(\boldsymbol{X}_i)}Y_i-\frac{R_i-\hat{\pi}(\boldsymbol{X}_i)}{\hat{\pi}(\boldsymbol{X}_i)}\hat{a}(\boldsymbol{X}_i)\right\},$$

where $\hat{\pi}(\boldsymbol{x})$ and $\hat{a}(\boldsymbol{x})$ denote the estimated values of $\pi(\boldsymbol{x})$ and $a(\boldsymbol{x})$, respectively. The consistency of $\hat{\mu}_{\text{aipw}}$ is achieved if either $\pi(\boldsymbol{x})$ or $a(\boldsymbol{x})$ is correctly modeled. The double robustness of $\hat{\mu}_{\text{aipw}}$ provides an extra protection on consistency compared to the inverse probability weighted (IPW) estimator (Horvitz and Thompson, 1952)

$$\hat{\mu}_{\text{ipw}}=\frac{1}{n}\sum_{i=1}^n\frac{R_i}{\hat{\pi}(\boldsymbol{X}_i)}Y_i,$$

which is consistent only if $\pi(\boldsymbol{x})$ is correctly modeled. As expected, $\hat{\mu}_{\text{aipw}}$ loses its consistency when neither $\pi(\boldsymbol{x})$ nor $a(\boldsymbol{x})$ is correctly modeled. The same is true for most existing doubly robust estimators, with an exception for the one in van der Laan and Gruber (2010), which may still be consistent under certain combinations of dual misspecifications of the models for $\pi(\boldsymbol{x})$ and $a(\boldsymbol{x})$.

To improve on double robustness, suppose now that multiple models $\mathcal{P}=\{\pi^j(\boldsymbol{\alpha}^j;\boldsymbol{x}):\ j=1,\cdots,J\}$ for $\pi(\boldsymbol{x})$ and multiple models $\mathcal{A}=\{a^k(\boldsymbol{\gamma}^k;\boldsymbol{x}):\ k=1,\ldots,K\}$ for $a(\boldsymbol{x})$ are postulated. Here $\boldsymbol{\alpha}^j$ and $\boldsymbol{\gamma}^k$ are the corresponding parameters, and we let $\hat{\boldsymbol{\alpha}}^j$ and $\hat{\boldsymbol{\gamma}}^k$ denote their estimators. Usually, $\hat{\boldsymbol{\alpha}}^j$ is taken to be the maximizer of the binomial likelihood

$$\prod_{i=1}^n\{\pi^j(\boldsymbol{\alpha}^j;\boldsymbol{X}_i)\}^{R_i}\{1-\pi^j(\boldsymbol{\alpha}^j;\boldsymbol{X}_i)\}^{1-R_i},$$

and $\hat{\boldsymbol{\gamma}}^k$ to be the regression coefficients of a generalized linear model for $a(\boldsymbol{x})$ based on the complete-case analysis.

Define

$$\hat{\theta}^j=\frac{1}{n}\sum_{i=1}^n\pi^j\left(\hat{\boldsymbol{\alpha}}^j;\boldsymbol{X}_i\right),\quad \hat{\eta}^k=\frac{1}{n}\sum_{i=1}^n a^k\left(\hat{\boldsymbol{\gamma}}^k;\boldsymbol{X}_i\right).$$