Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Multivariate Analysis 96 (2005) 190–217

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

Robust semiparametric M-estimation and the weighted bootstrap

Shuangge Ma[a], Michael R. Kosorok[a,b,*]

[a] *Department of Statistics, University of Wisconsin-Madison, 1210 W. Dayton, Madison, WI 53706, USA*
[b] *Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, 1210 W. Dayton, Madison, WI 53706, USA*

Received 24 December 2003
Available online 11 November 2004

Abstract

M-estimation is a widely used technique for statistical inference. In this paper, we study properties of ordinary and weighted M-estimators for semiparametric models, especially when there exist parameters that cannot be estimated at the $\sqrt{n}$ convergence rate. Results on consistency, rates of convergence for all parameters, and $\sqrt{n}$ consistency and asymptotic normality for the Euclidean parameters are provided. These results, together with a generic paradigm for studying semiparametric M-estimators, provide a valuable extension to previous related research on semiparametric maximum-likelihood estimators (MLEs). Although penalized M-estimation does not in general fit in the framework we discuss here, it is shown for a great variety of models that many of the forgoing results still hold, including the $\sqrt{n}$ consistency and asymptotic normality of the Euclidean parameters. For semiparametric M-estimators that are not likelihood based, general inference procedures for the Euclidean parameters have not previously been developed. We demonstrate that our paradigm leads naturally to verification of the validity of the weighted bootstrap in this setting. For illustration, several examples are investigated in detail. The new M-estimation framework and accompanying weighted bootstrap technique shed light on a universal way of investigating semiparametric models.

AMS 1991 subject classification: 62F40; 62G20; 62G05

Keywords: M-estimation; Semiparametric models; Weighted bootstrap; Empirical processes

* Corresponding author. Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, 1210 W. Dayton, Madison, WI 53706, USA. Fax: +1 608 265 5579.

E-mail addresses: shuangge@biostat.wisc.edu (S. Ma), kosorok@biostat.wisc.edu (M.R. Kosorok).

0047-259X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2004.09.008

1. Introduction

The term "M-estimation" refers to a general method of estimation, where the estimators are obtained by maximizing (or minimizing) certain criterion functions. Suppose n independent, identically distributed (i.i.d.) observations $X_1 \dots X_n$ are drawn from P_γ, where γ is the unknown parameter of interest and may be infinite dimensional. A common type of criterion function takes the form $\mathbb{P}_n m_\gamma = n^{-1} \sum_{i=1}^n m_\gamma(X_i)$, where m_γ is a deterministic function and $\mathbb{P}_n$ is the empirical measure. The most widely used M-estimators include maximum-likelihood (MLE), ordinary least-squares (OLS), and least absolute deviation estimators. Semiparametric models are statistical models where at least one parameter of interest is not Euclidean. In this paper, we study asymptotic properties of ordinary and weighted M-estimators for semiparametric models that can be parametrized as $\gamma = (\theta, \eta) \mapsto P_{\theta,\eta}$, where θ is a Euclidean parameter and η belongs to an infinite-dimensional set.

Semiparametric maximum-likelihood estimation and M-estimation for parametric and nonparametric models has been studied extensively. For excellent references on the subject, see van der Vaart and Wellner [18] (hereafter abbreviated VW), [17, Chapter 25], [14] Important results include consistency and rate of convergence theorems for general M-estimators in [14,18] and efficiency results for the MLE of the Euclidean parameters in [2,7,17]. Moreover, a general theorem for investigating the asymptotic behavior of M-estimators for the Euclidean parameter in semiparametric models is given in [21].

An alternative approach to obtain the limit distribution of M-estimators is to derive a characterization as the solution of estimating equations of the form $\mathbb{P}_n \Psi(\hat{\gamma}_n) = 0$, as discussed in VW. Estimators obtained in this way are, in general, called Z-estimators. The Z-estimator theorem given in VW is especially useful when the estimator $\hat{\gamma}_n$ has convergence rate $\sqrt{n}$. Unfortunately, this theory does not apply when there exist parameters that cannot be estimated at the $\sqrt{n}$ rate.

Penalized M-estimation provides a flexible alternative to ordinary M-estimation. Examples of penalized M-estimators for semiparametric models include the penalized MLE for partial linear models of Mammen and van de Geer [10], the penalized MLE for the Cox model with interval censored data of Cai and Betensky [3], and the penalized MLE for transformation models with current status data of Ma and Kosorok [9].

Currently, there are no general inference procedures for the Euclidean component of a semiparametric M-estimator. It is interesting to note that although the empirical process bootstrap has been studied and used for a long time (see [1,4,11,20] for reference), no bootstrap results for semiparametric models are yet available for the setting where one of the parameters is not $\sqrt{n}$-consistent. The particular bootstrap we study in this paper is the weighted bootstrap which consists of i.i.d. positive random weights applied to each observation. This is in contrast to the nonparametric bootstrap where the random vector of observation weights is multinomial $(n, n^{-1}, \dots, n^{-1})$. In this case, the observation weights are exchangeable but not independent. We note, of course, that in both kinds of bootstraps the random weights are independent of the data. The reason we focus on the weighted bootstrap in this paper is that the i.i.d. behavior of the weights makes many of the proofs easier. While it is possible that our results also hold for the nonparametric bootstrap, such a determination is beyond the scope of this paper and appears to be quite difficult.