Statistics and Probability Letters 129 (2017) 216–221

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Asymptotic normality of one-step M-estimators based on non-identically distributed observations

Yuliana Yu. Linke

Sobolev Institute of Mathematics, Novosibirsk State University, Novosibirsk, Russia

ARTICLE INFO

Article history:
Received 14 December 2016
Received in revised form 29 April 2017
Accepted 29 May 2017
Available online 9 June 2017

Keywords:
One-step M-estimator
Initial estimator
Nonlinear regression

ABSTRACT

We find general conditions for asymptotic normality of two types of one-step M-estimators based on independent not necessarily identically distributed observations. As an application, we consider some examples of one-step approximation of quasi-likelihood estimators in nonlinear regression.

1. Introduction

Let $X_1, X_2, \ldots$ be independent observations that need not be identically distributed. We assume that their distributions depend on an unknown parameter $\theta \in \Theta \subset \mathbb{R}^m$. As one of the general methods to estimate parameter θ is usually considered M-estimation. By an M-estimator we mean a statistic $\widetilde{\theta}_n$ such that $\widetilde{\theta}_n$ is a solution to the equation

$$\sum_{i=1}^{n} M_i(t, X_i) = 0 \tag{1}$$

with probability tending to 1 as $n \to \infty$, where $M_i(t, x)$, $i = 1, \ldots, n$, are some known vector functions, with $\mathbb{E}M_i(\theta, X_i) = 0$ for all i. By a vector we always mean a column vector of height m. In what follows, the true parameter will be denoted by the symbol θ. Dependence on θ of the expectation $\mathbb{E}$ and probability $\mathbb{P}$ will not be indicated.

As is well known (see, e.g., Bai and Wu, 1997), $\widetilde{\theta}_n$ is asymptotically normal, i.e.,

$$A_{n,\theta}\left(\widetilde{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}_m(0, I) \tag{2}$$

under suitable regularity conditions, where $A_{n,\theta}$ denotes some $(m \times m)$-matrix which will be defined later; by $\mathcal{N}_m(0, I)$ we denote a random vector having the m-dimensional standard normal law. Unless explicitly stated otherwise, we always mean the limits as $n \to \infty$.

On the other hand, it may be difficult to find a consistent solution $\widetilde{\theta}_n$ to Eq. (1) and even its approximation with the help of iterative methods, especially if there is a few solutions to (1) (see, e.g., Small and Yang, 1999; Small and Wang, 2003). The situation becomes much easier if we know an initial estimator θ_n^* of the parameter θ which is consistent with a suitable rate of convergence. Put

$$\theta_n^{**} = \theta_n^* - \left(\sum_{i=1}^{n} D_i(\theta_n^*, X_i)\right)^{-1} \sum_{i=1}^{n} M_i(\theta_n^*, X_i), \tag{3}$$

E-mail address: linke@math.nsc.ru.

http://dx.doi.org/10.1016/j.spl.2017.05.020
0167-7152/

where $D_i(t, X_i)$ is the nondegenerate Jacobi matrix of the vector $M_i(t, X_i)$. The so-called *one-step M-estimator* θ_n^{**} is a one-step approximation of a solution to (1) obtained by Newton's iteration method if $t = \theta_n^*$ is the initial point. We prove that, under a wide spectrum of constraints on the exactness of θ_n^*, this estimator satisfies the limit relation

$$A_{n,\theta}\left(\theta_n^{**} - \theta\right) \xrightarrow{d} \mathcal{N}_m(0, \mathrm{I}). \tag{4}$$

So, the explicit estimator θ_n^{**} has the same asymptotic exactness as the M-estimator $\widetilde{\theta}_n$ satisfying (2). Moreover, the limit relation (4) may occur while (2) is not (see Remark 3).

Seemingly, the idea of one-step estimation goes back to R. Fisher. He considered an approximation of consistent maximum likelihood estimators based on a homogeneous sample. As regards one-step M-estimators of the form (3), they were particularly studied in the case of independent identically distributed observations and univariate parameters (see, e.g., the references in Linke, 2016). Our interest in one-step estimators (3) is mainly connected with the problem of approximation of least squares, maximum likelihood, and quasi-likelihood estimators for nonlinear regression problems. Notice that existence of several roots of Eq. (1) is rather typical for nonlinear regression models (see, e.g., Small and Yang, 1999; Small and Wang, 2003; see also Fig. 1). One of the basic elements of the one-step estimation methodology is the existence of sufficiently precise initial estimators. In Linke and Borisov (2017), an approach was suggested to construct such estimators for nonlinear regression models.

The idea of one-step estimation is widespread in the case when estimation is connected with finding the roots of some equations arising in various special statistical problems (for example, see Bickel, 1975; Simpson et al., 1992; Müller, 1994; Field and Wiens, 1994; Fan and Chen, 1999; Fan and Jiang, 2000; Welsh and Ronchetti, 2002; Bergesio and Yohai, 2011; Jurečková, 2012; Jurečková et al., 2012; Fan et al., 2014; and the references there).

The author thanks the referee for helpful comments.

2. Main results

We assume below that a matrix norm $\|\cdot\|$ is coordinated with a vector norm $|\cdot|$ and it is semimultiplicative, i.e., $|Ax| \le \|A\||x|$ and $\|AB\| \le \|A\|\|B\|$ for any matrices A, B, and every vector x. By convergence for vectors and matrices we mean the coordinate-wise convergence or, which is equivalent, convergence with respect to the norms chosen. By the expectation of a vector or a matrix, we mean the corresponding coordinate-wise expectation.

We will need the following conditions.

($\mathbf{A}_1$) Let $X_1, X_2, \dots, X_n$ be independent observations with values in a measurable space $\mathcal{X}$. The distributions $\mathcal{L}_{1,\theta}$, $\mathcal{L}_{2,\theta} \dots, \mathcal{L}_{n,\theta}$ of these observations depend on an unknown parameter $\theta \in \Theta \subset \mathbb{R}^m$, where Θ is an open set. In general, these distributions may depend on n and a secondary parameter ν of an arbitrary nature (we will not indicate dependence on ν).

($\mathbf{A}_2$) For every i, we define m-dimensional vectors $M_i(t, X_i)$ and $(m \times m)$-matrices $D_i(t, X_i)$ on Θ (which may depend on n) such that, for each interval $(t_1, t_2) \subset \Theta$, the equality

$$M_i(t_2, X_i) - M_i(t_1, X_i) = (t_2 - t_1)\int_0^1 D_i(t_1 + v(t_2 - t_1), X_i)dv$$

holds with probability 1, the mathematical expectation of each component of $D_i(\theta, X_i)$ is finite, and the relations $\mathbb{E}M_i(\theta, X_i) = 0$ and $\mathbb{E}|M_i(\theta, X_i)|^2 < \infty$ are valid.

($\mathbf{A}_3$) For all sufficiently large n, the matrix $J_{n,\theta} := \sum_{i=1}^n \mathbb{E}D_i(\theta, X_i)$ is nondegenerate and the matrix $I_{n,\theta} := \sum_{i=1}^n \mathbb{E}M_i(\theta, X_i) M_i^\top(\theta, X_i)$ is positive definite. Moreover, we have $\|I_{n,\theta}^{-1/2}\|\|J_{n,\theta}\| \to \infty$, $\sup_n \|I_{n,\theta}^{-1/2}\|\|I_{n,\theta}^{1/2}\| < \infty$, and

$$\sum_{i=1}^n D_i(\theta, X_i) J_{n,\theta}^{-1} \xrightarrow{p} I, \qquad I_{n,\theta}^{-1/2}\sum_{i=1}^n M_i(\theta, X_i) \xrightarrow{d} \mathcal{N}_m(0, I). \tag{5}$$

Put $\mathcal{E}_{n,\theta}(|\delta|) = \sum_{i=1}^n \mathbb{E}\omega_{i,\theta}(|\delta|, X_i)$, where

$$\omega_{i,\theta}(|\delta|, X_i) = \begin{cases} \sup\limits_{t:\,|t-\theta|\le|\delta|} \left\|\left(D_i(t, X_i) - D_i(\theta, X_i)\right)\right\| \|J_{n,\theta}^{-1}\| & \text{if } [\theta - \delta, \theta + \delta] \subset \Theta, \\ \infty \quad \text{otherwise.} \end{cases} \tag{6}$$

($\mathbf{A}_4$) $\limsup \mathcal{E}_{n,\theta}(|\delta|) \to 0$ as $\delta \to 0$.

($\mathbf{A}_5$) There is an estimator θ_n^* such that $\|I_{n,\theta}^{-1/2}\|\|J_{n,\theta}\||\theta_n^* - \theta|\mathcal{E}_{n,\theta}(|\theta_n^* - \theta|) \xrightarrow{p} 0$.

Theorem 1. *If the conditions* (A_1)–(A_5) *hold then the estimator* θ_n^{**} *in* (3) *is defined with probability tending to* 1 *and the relation* (4) *holds with* $A_{n,\theta} = I_{n,\theta}^{-1/2} J_{n,\theta}$.

Remark 1. The relations (5) are versions of the law of large numbers and the central limit theorem for triangular arrays of independent random variables. Sufficient conditions for fulfillment of such limit theorems are well known. Condition (A_5) is a universal constraint connecting the smoothness of the functions $M_i(\cdot, X_i)$ with the rate of convergence