Journal of Multivariate Analysis 147 (2016) 20–37

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Asymptotic expansions for the estimators of Lagrange multipliers and associated parameters by the maximum likelihood and weighted score methods☆

Haruhiko Ogasawara

Department of Information and Management Science, Otaru University of Commerce, 3-5-21, Midori, Otaru 047-8501, Japan

ARTICLE INFO

Article history:
Received 18 September 2014
Available online 11 January 2016

AMS 2010 subject classifications:
62E20
62F12
62F15

Keywords:
One-sided confidence intervals
Asymptotic cumulants
Lagrange multipliers
Weighted score

ABSTRACT

In this paper, inverse expansions of parameter estimators are given in terms of their true values, where the estimators are obtained by the maximum likelihood and weighted score methods with constraints placed on the parameters using Lagrange multipliers. The corresponding expansions for estimated Lagrange multipliers are also given. These expansions are derived before and after studentization. The results with studentization give one-sided confidence intervals for the parameters up to third-order accuracy. As an application of the weighted score method, a modified Jeffreys prior to remove the asymptotic biases of the Lagrange multipliers as well as the parameter estimators is obtained under canonical parametrization in the exponential family.

1. Introduction

The Lagrange multiplier is an additional parameter that is introduced when a restriction is imposed on parameters to be estimated. When several restrictions are used, we have a vector of Lagrange multipliers. The estimators of the Lagrange multipliers can be used to test the restrictions by using asymptotic chi-square distribution, which is called the Lagrange multiplier test [3,4,35,36, Chapter 7]. It is known that the Lagrange multiplier test is equivalent to Rao's [32] score test (for developments of the score test and associated methods, see [11,13,10,33,34]). The higher-order distributions of the statistic of the score test are also available (e.g., [17,12] with a Bartlett correction; for a review see [19] and its references).

The purpose of this paper is to derive the asymptotic cumulants of non-studentized and studentized Lagrange multipliers and similar results of the associated parameter estimators with restrictions when both the weighted score (WS) and maximum likelihood (ML) are used for estimation. Using the results with studentization, one-sided confidence intervals (CIs) for the parameters up to third-order accuracy are given.

Note that the chi-square or quadratic form of the test statistic mentioned above corresponds to the two-sided confidence interval in the context of interval estimation and its applications in testing. Added terms in the asymptotic expansions of the distribution of the chi-square statistic higher than the usual asymptotic chi-square distribution are of order $O_p(n^{-k})$ ($k = 1, 2, \ldots$), where n is the sample size. The order $O_p(n^{-1})$ in the first additional term corresponds to that of the two-sided CI

☆ The author is indebted to an anonymous reviewer whose comments led to the improvement of an earlier version of this paper. This work was partially supported by a Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology (JSPS KAKENHI, Grant No. 26330031).

E-mail address: hogasa@res.otaru-uc.ac.jp.

http://dx.doi.org/10.1016/j.jmva.2015.12.015

which cancels the common term of order $O_p(n^{-1/2})$ in the upper and lower endpoints of the CI. When the one-sided CI is desired, the first additional term higher than that of the usual normal approximation is of order $O_p(n^{-1/2})$ (the CI with the additional term of order $O_p(n^{-1/2})$ is said to be second-order accurate while the CI based on the usual normal approximation has first-order accuracy). Since the additional term of order $O_p(n^{-1/2})$ is much simpler than that of $O_p(n^{-1})$, the one-sided CI with second-order accuracy has practical advantages.

2. Asymptotic expansions of the estimators with restrictions

Let $\boldsymbol{\theta}$ be the $q \times 1$ vector of parameters in a statistical model. Define the likelihood of $\boldsymbol{\theta}$ as $L = \prod_{i=1}^{n} f_i(\mathbf{x}_i | \boldsymbol{\theta})$, where $\mathbf{x}_i$ is the $p \times 1$ vector of the ith observation ($i = 1, \ldots, n$) and $f_i(\cdot)$ is the density or probability function of $\mathbf{x}_i$ given $\boldsymbol{\theta}$. Assume that the restrictions are given by an $r \times 1$ vector $\mathbf{h} = \mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$ with the corresponding vector $\boldsymbol{\eta}$ of Lagrange multipliers. It is also assumed that the model is identified without the restrictions. Let $\mathbf{q}^* = \mathbf{q}^*(\boldsymbol{\theta})$ be the $q \times 1$ weight vector in the WS method, which becomes the vector of log-prior derivatives in the case of Bayesian estimation, but can take other forms. Let $\bar{l} \equiv n^{-1} \log L$, then the weighted score estimators $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ and $\hat{\boldsymbol{\eta}}_{\mathrm{W}}$ for the population counterparts $\boldsymbol{\theta}_0$ and $\boldsymbol{\eta}_0$, respectively, are defined as the solution of the following equation (for the weighted score or modified score method, see [14], who deals with the cases without the restrictions and Lagrange multipliers):

$$\begin{pmatrix} \left.\dfrac{\partial \bar{l}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{W}}} + n^{-1} \left.\dfrac{\partial\, \mathbf{h}(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{W}}} \hat{\boldsymbol{\eta}}_{\mathrm{W}} + n^{-1}\mathbf{q}^*(\hat{\boldsymbol{\theta}}_{\mathrm{W}}) \\ \mathbf{h}(\hat{\boldsymbol{\theta}}_{\mathrm{W}}) \end{pmatrix} = \mathbf{0}. \tag{2.1}$$

Note that the maximum likelihood estimators (MLEs) denoted by $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ and $\hat{\boldsymbol{\eta}}_{\mathrm{ML}}$ are special cases of $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ and $\hat{\boldsymbol{\eta}}_{\mathrm{W}}$, respectively when $\mathbf{q}^* = \mathbf{0}$. Then, $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ and $n^{-1}\hat{\boldsymbol{\eta}}_{\mathrm{W}}$ are expanded as follows.

Lemma 1. *Under regularity conditions, the expansions of $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ and $n^{-1}\hat{\boldsymbol{\eta}}_{\mathrm{W}}$ are given by*

$$\begin{pmatrix} \hat{\boldsymbol{\theta}}_{\mathrm{W}} - \boldsymbol{\theta}_0 \\ n^{-1}\hat{\boldsymbol{\eta}}_{\mathrm{W}} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\theta}}_{\mathrm{ML}} - \boldsymbol{\theta}_0 \\ n^{-1}\hat{\boldsymbol{\eta}}_{\mathrm{W}} \end{pmatrix} + n^{-1}\mathbf{g}_0^{(\mathrm{W})} + n^{-1}\boldsymbol{\gamma}_0^{(\mathrm{W})} + O_p(n^{-2}) \tag{2.2}$$

with

$$\begin{pmatrix} \hat{\boldsymbol{\theta}}_{\mathrm{ML}} - \boldsymbol{\theta}_0 \\ n^{-1}\hat{\boldsymbol{\eta}}_{\mathrm{W}} \end{pmatrix} = \sum_{i=1}^{3} \boldsymbol{\Lambda}_{\mathrm{ML}}^{(i)}\mathbf{l}_0^{(i)} + O_p(n^{-2}),$$

where $n^{-1}\mathbf{g}_0^{(\mathrm{W})} = O_p(n^{-3/2})$ and $n^{-1}\boldsymbol{\gamma}_0^{(\mathrm{W})} = O(n^{-1})$ are additional terms by WS than those by ML in the expansion up to order $O_p(n^{-3/2})$, $\boldsymbol{\Lambda}_{\mathrm{ML}}^{(i)}\mathbf{l}_0^{(i)} = O_p(n^{-i/2})$ with $\boldsymbol{\Lambda}_{\mathrm{ML}}^{(i)} = O(1)$ and $\mathbf{l}_0^{(i)} = O_p(n^{-i/2})$ $(i = 1,\ 2,\ 3)$ are the terms in the expansions of the MLEs up to order $O_p(n^{-3/2})$.

The proof of Lemma 1 including the actual expressions of $\mathbf{g}_0^{(\mathrm{W})}$, $\boldsymbol{\gamma}_0^{(\mathrm{W})}$, $\boldsymbol{\Lambda}_{\mathrm{ML}}^{(i)}$ and $\mathbf{l}_0^{(i)}$ $(i = 1,\ 2,\ 3)$ is given in the Appendix. The lemma can be used to obtain the asymptotic properties of $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ and $n^{-1}\hat{\boldsymbol{\eta}}_{\mathrm{W}}$. For instance, we see that $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ is the same as $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ up to order $O_p(n^{-1/2})$. In the next section, the asymptotic cumulants of $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ and $n^{-1}\hat{\boldsymbol{\eta}}_{\mathrm{W}}$ before and after studentization are derived.

3. Asymptotic cumulants of the estimators of parameters and Lagrange multipliers

Define

$$\mathbf{L}_0 \equiv \left.\frac{\partial^2 \bar{l}}{\partial\, \boldsymbol{\theta} \partial\, \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \equiv \frac{\partial^2 \bar{l}}{\partial\, \boldsymbol{\theta}_0 \partial\, \boldsymbol{\theta}_0'}, \qquad \mathrm{E_T}(\mathbf{L}_0) \equiv \boldsymbol{\Lambda}_0 = O(1), \qquad \mathrm{E}_\theta(\mathbf{L}_0) \equiv -\mathbf{I}_0 = O(1), \tag{3.1}$$

where $\mathrm{E_T}(\cdot)$ is the expectation under an alternative true distribution, $\mathrm{E}_\theta(\cdot)$ is the expectation under correct model specification and $\mathbf{I}_0$ is the Fisher information matrix per observation. Define also

$$\mathbf{L}_0^* \equiv \begin{pmatrix} \mathbf{L}_0 & \dfrac{\partial \mathbf{h}_0'}{\partial\, \boldsymbol{\theta}_0} \\ \dfrac{\partial \mathbf{h}_0}{\partial\, \boldsymbol{\theta}_0'} & \mathbf{O} \end{pmatrix} \equiv \begin{pmatrix} \mathbf{L}_0 & \mathbf{H}_0 \\ \mathbf{H}_0' & \mathbf{O} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Lambda}_0 & \mathbf{H}_0 \\ \mathbf{H}_0' & \mathbf{O} \end{pmatrix} + \begin{pmatrix} \mathbf{L}_0 - \boldsymbol{\Lambda}_0 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \equiv \boldsymbol{\Lambda}_0^* + \begin{pmatrix} \mathbf{M}_0 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix},$$

$$\boldsymbol{\Lambda}_0^{*-1} \equiv \begin{pmatrix} \boldsymbol{\Lambda}_0^{(11)} & \boldsymbol{\Lambda}_0^{(12)} \\ \boldsymbol{\Lambda}_0^{(21)} & \boldsymbol{\Lambda}_0^{(22)} \end{pmatrix}, \qquad -\mathrm{E}_\theta(\mathbf{L}_0^*) = \begin{pmatrix} \mathbf{I}_0 & -\mathbf{H}_0 \\ -\mathbf{H}_0' & \mathbf{O} \end{pmatrix} \equiv \mathbf{I}_0^*, \qquad \mathbf{I}_0^{*-1} \equiv \begin{pmatrix} \mathbf{I}_0^{(11)} & \mathbf{I}_0^{(12)} \\ \mathbf{I}_0^{(21)} & \mathbf{I}_0^{(22)} \end{pmatrix}. \tag{3.2}$$