Journal of Multivariate Analysis 101 (2010) 2149–2167

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Asymptotic expansions for the pivots using log-likelihood derivatives with an application in item response theory[☆]

Haruhiko Ogasawara

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

A R T I C L E   I N F O

Article history:
Received 21 July 2009
Available online 24 April 2010

AMS subject classifications:
62E20
62H10

Keywords:
Pivots
Log-likelihood derivatives
Inverse expansion
Sandwich estimator
Item response theory

A B S T R A C T

Asymptotic expansions of the distributions of the pivotal statistics involving log-likelihood derivatives under possible model misspecification are derived using the asymptotic cumulants up to the fourth-order and the higher-order asymptotic variance. The pivots dealt with are the studentized ones by the estimated expected information, the negative Hessian matrix, the sum of products of gradient vectors, and the so-called sandwich estimator. It is shown that the first three asymptotic cumulants are the same over the pivots under correct model specification with a general condition of the equalities. An application is given in item response theory, where the observed information is usually used rather than the estimated expected one.

1. Introduction

For testing and estimation of population parameters using maximum likelihood estimators (MLEs), pivots or pivotal statistics are frequently used with their asymptotic normality. Let $\boldsymbol{\theta}$ be a $q \times 1$ vector of unknown parameters in a model with its MLE and population counterpart being $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, respectively. Denote a generic parameter by θ representing one of the elements of $\boldsymbol{\theta}$ with $\hat{\theta}$ and θ_0 defined similarly. Then, a pivot is generally given as

$$t_V = \frac{N^{1/2}(\hat{\theta} - \theta_0)}{\{(\hat{\mathbf{V}})_{\theta\theta}\}^{1/2}} \quad \text{with } \hat{\mathbf{V}} = \mathbf{V}(\hat{\mathbf{I}}, \hat{\mathbf{H}}, \hat{\mathbf{G}}), \tag{1.1}$$

where $\hat{\mathbf{V}}$ is a $q \times q$ matrix of the consistent estimator of N times the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ under correct model specification; $\mathbf{V}$ is assumed to be differentiable with respect to $\mathbf{I}$, $\mathbf{H}$ and $\mathbf{G}$ that are $q \times q$ matrices associated with Fisher information; N is the sample size; $(\cdot)_{\theta\theta}$ is a diagonal element of the matrix in parentheses corresponding to θ; and $\hat{\mathbf{I}} = \mathbf{I}(\hat{\boldsymbol{\theta}})$, $\hat{\mathbf{H}} = \mathbf{H}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{G}} = \mathbf{G}(\hat{\boldsymbol{\theta}})$ are matrix functions, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, defined in the following.

Let $\bar{l}$ be the mean log-likelihood of $\boldsymbol{\theta}$:

$$\bar{l} = \bar{l}(\boldsymbol{\theta}) = \bar{l}(\boldsymbol{\theta}|\mathbf{Y}) = N^{-1}\sum_{i=1}^{N} l_i = N^{-1}\sum_{i=1}^{N} l_i(\boldsymbol{\theta}) = N^{-1}\sum_{i=1}^{N} \ln f(\mathbf{y}_i|\boldsymbol{\theta}), \tag{1.2}$$

☆ This work was partially supported by a Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology, No. 20500249.

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

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.04.001

where $\mathbf{Y}$ is the $N \times p$ matrix of independent observations whose ith row is $\mathbf{y}_i'$, and $f(\mathbf{y}_i|\boldsymbol{\theta})$ is the joint density or probability of the p variables at $\mathbf{y}_i$ in a posited model. Then, $\mathbf{I}$, $\mathbf{H}$ and $\mathbf{G}$ are defined as

$$\begin{aligned}\mathbf{I} = \mathbf{I}(\boldsymbol{\theta}) &= \mathrm{E}_{\boldsymbol{\theta}}\left(\frac{\partial l_i}{\partial \boldsymbol{\theta}}\frac{\partial l_i}{\partial \boldsymbol{\theta}'}\right) = \mathrm{E}_{\boldsymbol{\theta}}\left(N\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}}\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}'}\right) = \mathrm{E}_{\boldsymbol{\theta}}\left(-\frac{\partial^2 \bar{l}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right) \quad (i = 1, \ldots, N), \\ \mathbf{H} = \mathbf{H}(\boldsymbol{\theta}) &= -\frac{\partial^2 \bar{l}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} \quad \text{and} \quad \mathbf{G} = \mathbf{G}(\boldsymbol{\theta}) = N^{-1}\sum_{i=1}^{N}\frac{\partial l_i}{\partial \boldsymbol{\theta}}\frac{\partial l_i}{\partial \boldsymbol{\theta}'},\end{aligned} \tag{1.3}$$

where $\mathrm{E}_{\boldsymbol{\theta}}(\cdot)$ denotes that the expectation is taken under the model. When the matrices are evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, we obtain $\mathbf{I}_0 = \mathbf{I}(\boldsymbol{\theta}_0) = O(1)$, which is a population information matrix per observation, while $\mathbf{H}_0 = \mathbf{H}(\boldsymbol{\theta}_0) = \mathbf{H}(\boldsymbol{\theta}_0, \mathbf{Y})$ and $\mathbf{G}_0 = \mathbf{G}(\boldsymbol{\theta}_0) = \mathbf{G}(\boldsymbol{\theta}_0, \mathbf{Y})$ as well as $\hat{\mathbf{I}} = \mathbf{I}(\hat{\boldsymbol{\theta}})$, $\hat{\mathbf{H}} = \mathbf{H}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{G}} = \mathbf{G}(\hat{\boldsymbol{\theta}})$ are stochastic quantities. Note that $\hat{\mathbf{I}}$ is the estimated expected information matrix and that $\hat{\mathbf{H}}$ and $\hat{\mathbf{G}}$ are observed ones with $\boldsymbol{\theta}$ evaluated at $\hat{\boldsymbol{\theta}}$. Though $\mathrm{E}_{\boldsymbol{\theta}}(\mathbf{H}_0) = \mathrm{E}_{\boldsymbol{\theta}}(\mathbf{G}_0) = \mathbf{I}_0$, generally $\mathrm{E}_{\boldsymbol{\theta}}(\hat{\mathbf{H}}) \neq \mathbf{I}_0$ and $\mathrm{E}_{\boldsymbol{\theta}}(\hat{\mathbf{G}}) \neq \mathbf{I}_0$. In this paper, we deal with the following pivots:

$$\begin{aligned}&z = \frac{N^{1/2}(\hat{\theta} - \theta_0)}{\{(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1\prime})_{\theta\theta}\}^{1/2}}, \qquad t = \frac{N^{1/2}(\hat{\theta} - \theta_0)}{\{(\hat{\mathbf{I}}^{-1})_{\theta\theta}\}^{1/2}}, \\ &t_H = \frac{N^{1/2}(\hat{\theta} - \theta_0)}{\{(\hat{\mathbf{H}}^{-1})_{\theta\theta}\}^{1/2}}, \qquad t_G = \frac{N^{1/2}(\hat{\theta} - \theta_0)}{\{(\hat{\mathbf{G}}^{-1})_{\theta\theta}\}^{1/2}} \quad \text{and} \quad t_R = \frac{N^{1/2}(\hat{\theta} - \theta_0)}{\{(\hat{\mathbf{H}}^{-1}\hat{\mathbf{G}}\hat{\mathbf{H}}^{-1})_{\theta\theta}\}^{1/2}},\end{aligned} \tag{1.4}$$

where z is given from (1.1) when $\hat{\mathbf{V}}$ is replaced by $\mathbf{V}(\boldsymbol{\Lambda}, \boldsymbol{\Gamma})$ with $\mathbf{I}$ being null, $\boldsymbol{\Lambda} = O(1)$ and $\boldsymbol{\Gamma} = O(1)$. Let

$$\mathbf{L} = \frac{\partial^2 \bar{l}}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'} \equiv \frac{\partial^2 \bar{l}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -\mathbf{H}_0. \tag{1.5}$$

Then, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ are defined as

$$\boldsymbol{\Lambda} = \mathrm{E}_{\mathrm{T}}(\mathbf{L}) \equiv \mathrm{E}_{\boldsymbol{\theta}_{\mathrm{T}}}(\mathbf{L}) \quad \text{and} \quad \boldsymbol{\Gamma} \equiv \mathrm{E}_{\mathrm{T}}(\mathbf{G}_0), \tag{1.6}$$

where $\boldsymbol{\theta}_{\mathrm{T}}$ is a $q_{\mathrm{T}} \times 1$ vector of parameters in a true distribution when the model with $\boldsymbol{\theta}$ is misspecified, and $\mathrm{E}_{\mathrm{T}}(\cdot)$ indicates that the expectation is taken over the true distribution rather than that of the misspecified one. Though $\boldsymbol{\theta}_{\mathrm{T}}$ is usually unknown in practice, z is used in this paper for comparison to the remaining pivots. Under correct model specification, $-\boldsymbol{\Lambda} = \boldsymbol{\Gamma} = \mathbf{I}_0$ and all the five pivots have unit asymptotic variances. However, when the model is misspecified, generally only z and t_R have this property.

In the remaining sections of this paper, general formulas of the asymptotic expansions of the distributions of the five pivots will be derived up to $O(N^{-1})$ with the required asymptotic cumulants up to the fourth-order and the higher-order asymptotic variance under possible model misspecification. The asymptotic cumulants will be derived using the expectations of (products of) log-likelihood derivatives based on the true distribution i.e., $\mathrm{E}_{\mathrm{T}}(\cdot)$. The method of the asymptotic expansions of the distributions of $\hat{\theta}$ using the log-likelihood derivatives is called inverse expansion and has been developed by Bartlett [4], Lawley [19], Hayakawa [17], Akahira and Takeuchi [2], McCullagh [21], Ferguson [14], Stafford [27] and Viraswami and Reid [29]. An advantage of the methods using log-likelihood derivatives is that their exact moments are usually available especially when the model is correctly specified (see e.g., [5, p. 12] and [28, Sections 19.17–19.22]).

For statistics similar to (1.4), Stafford [27] obtained the asymptotic cumulants, up to the fourth order, of the elements of the following vectors

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\hat{\mathbf{H}}^{-1/2} \quad \text{and} \quad (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'(\hat{\mathbf{H}}^{-1}\hat{\mathbf{G}}\hat{\mathbf{H}}^{-1})^{-1/2} \tag{1.7}$$

under correct model specification, where $\mathbf{A}^{-1/2}$ is the matrix square root of $\mathbf{A}^{-1}$ satisfying $\mathbf{A}^{-1/2}\mathbf{A}^{-1/2\prime} = \mathbf{A}^{-1}$. Though in the case of a single parameter i.e., $q = 1$, the expressions of (1.7) become equal to t_H and t_R, respectively, they are generally different when $q > 1$. Stafford [27] derived the asymptotic cumulants using computer algebra developed by him (see also [3]), but showed the results only up to the third asymptotic cumulants except some special cases since his general expression is "two hundred pages long in the multi-parameter case" [27, p. 15]. The general formula of the asymptotic expansion when $q > 1$ under possible model misspecification was given by Viraswami and Reid [29, Section 3.3, Appendix 2] partially using computer algebra with a modified method of Stafford [27] for restricted cases which are similar to (1.7) but more tractable in that $\hat{\mathbf{H}}$ and $\hat{\mathbf{G}}$ are replaced by $\mathbf{H}_0$ and $\mathbf{G}_0$, respectively.

In this paper, the corresponding full results of the five pivots of (1.4) will be given up to the fourth asymptotic cumulants under possible model misspecification. It will be shown that the first three asymptotic cumulants of the distributions of t, t_H, t_G and t_R are the same over the pivots under correct model specification. The general conditions for these equalities to hold will also be provided. An application will be shown in item response theory (IRT, for IRT see e.g., [8]), where $q = 10$ with and without model misspecification. In IRT, the information matrix based on the multinomial distribution with 2^n categories or response patterns can be derived, where n is the number of dichotomously scored (i.e., correct or incorrect responses)