Asymptotic cumulants of some information criteria

Haruhiko Ogasawara

April 2015

Department of Information and Management Science
Otaru University of Commerce
Asymptotic cumulants of some information criteria

Haruhiko Ogasawara
Otaru University of Commerce

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.23500341. Author’s address: Department of Information and Management Science, Otaru University of Commerce, 3-5-21, Midori, Otaru 047-8501, Japan. Email: hogasa@res.otaru-uc.ac.jp
Asymptotic cumulants of some information criteria

Asymptotic cumulants of the Akaike and Takeuchi information criteria are given under possible model misspecification up to the fourth order with the higher-order asymptotic variances, where two versions of the latter information criterion are defined using observed and estimated expected information matrices. The asymptotic cumulants are provided before and after studentization using the parameter estimates by the weighted score method, which include the maximum likelihood and Bayes modal estimators as special cases. Higher-order bias corrections of the criteria are derived using log-likelihood derivatives, which yields simple results for cases under canonical parametrization in the exponential family. The results are illustrated by three examples.

Keywords: Akaike information criterion; Takeuchi information criterion; Kullback-Leibler distance; canonical parameters; higher-order bias correction.
1. Introduction

Typical information criteria are given by Akaike (1973) and Takeuchi (1976), which are called the Akaike information criterion (AIC) and Takeuchi information criterion (TIC), respectively. The criteria are used to assess the goodness of statistical models based on the Kullback-Leibler (1951) distance using the maximum likelihood estimators (MLEs) of associated parameters. In the AIC, it is assumed that a posited model holds or that a true model is a special case of the model employed. On the other hand in the TIC, possible model misspecification is considered. Stone (1977) derived the TIC in the context of cross validation. Linhart and Zucchini (1986, Proposition 2, Appendix A.2.1) also derived the TIC. For properties of the TIC, see Shibata (1989).

After the AIC and TIC were coined, information criteria with similar purposes have been introduced by e.g., Schwarz (1978; the Bayesian information criterion, BIC); Kishino and Hasegawa (1989), Ishiguro, Sakamoto and Kitagawa (1997; the extended information criterion, EIC), Shimodaira and Hasegawa (1999) for the methods using the bootstrap; Shibata (1989; the regularization information criterion, RIC) and, Konishi and Kitagawa (1996; the generalized information criterion, GIC; see also Konishi & Kitagawa, 2003; 2008, Chapters 5 to 8). In the RIC and GIC, the exclusive usage of the MLEs by the AIC and TIC was relaxed to cover e.g., robust and ridge-type estimators. For other information criteria, see Konishi and Kitagawa (2008) and Burnham and Anderson (2010).

The above information criteria are seen as point estimators of a corresponding population quantity with bias correction under correct model specification for the AIC and under possible model misspecification for the TIC, RIC and GIC. The population quantity is the so-called mean expected log-likelihood (Sakamoto, Ishiguro & Kitagawa, 1986, Equation (4.9)) associated with the Kullback-Leibler distance, where independent two-fold expectation is used one for data in the future for prediction and the other for current data for estimation with the same sample size denoted by \( n \). When \( n \) increases, the population value increases proportionately in an asymptotic sense. On the other hand, the terms of bias correction are of order \( O(1) \) for the AIC and \( O_p(1) \) for the TIC, RIC and GIC. For tractability, divide the information criteria by \( n \) yielding quantities per observation as \( n^{-1} \text{AIC} \) and \( n^{-1} \text{TIC} \). Then, the population value mentioned above is written symbolically...
as $O(1) + O(n^{-1})$ depending on $n$. The situation is somewhat different from that of typical parameter estimators as MLEs, where the population parameters usually do not depend on $n$. When $n$ becomes infinitely large, the population value $O(1) + O(n^{-1})$ for e.g., $n^{-1}\text{AIC}$ becomes $O(1)$, which is the expected log-likelihood averaged over observations, where the parameters are evaluated by their population values followed by expectation. The last population value of order $O(1)$ is also of interest as well as that of $O(1) + O(n^{-1})$.

The bias correction of the TIC was extended to the higher-order version by Konishi and Kitagawa (2003), which gives a refined point estimator of the population counterpart. On the other hand, statistical testing of the difference of the information criteria for different models have been developed by Steiger, Shapiro and Browne (1985) and Shimodaira (1997) under local alternatives and by Linhart (1988), and Kishino and Hasegawa (1989) under fixed alternatives. Interval estimation of the corresponding population quantities can also be done in similar manners. While the above methods of testing and estimation is for general models, the results for special models are available for the higher-order bias correction by Sugiura (1978) and Yanagihara, Sekiguchi and Fujikoshi (2003) and the asymptotic cumulants for standardized estimators by Yanagihara and Ohmoto (2005).

One of the purposes of this study is to derive general expressions of the higher-order bias corrections of $n^{-1}\text{AIC}$ and $n^{-1}\text{TIC}$ based on the parameter estimators by the weighted score method under possible model misspecification, where the expression is different from that of Konishi and Kitagawa (2003). The expression is given by the log-likelihood derivatives, which yields some transparent results for e.g., the cases of the natural exponential family. Note that Konishi and Kitagawa (2003) used the von Mises calculus (von Mises, 1947; Withers, 1983).

The second purpose is to give general formulas for the asymptotic cumulants of $n^{-1}\text{AIC}$ and $n^{-1}\text{TIC}$ up to the fourth order and the higher-order asymptotic variances before and after studentization for testing and interval estimation of the population quantities of interest. Three examples using basic distributions in statistics are shown. The first two examples of the exponential and normal distributions use MLEs under model misspecification, while the third example of the Bernoulli distribution uses the parameter estimators by the weighted score under correct model specification.
2. The higher-order asymptotic biases

Let $\theta$ be a $q \times 1$ vector of parameters in a statistical model with a $p \times 1$ vector $x^*$ of observable variables. Then, the log-likelihood of $\theta$ based on $n$ i.i.d. observations is denoted by

$$l = l(\theta \mid X^*) \equiv \sum_{j=1}^{n} l_j \equiv \sum_{j=1}^{n} \log f(x_j^* \mid \theta) \equiv f(X^* \mid \theta),$$

(2.1)

where $X^*$ is a $n \times p$ matrix whose rows $(x_j^*, j = 1, \ldots, n)$ are independent copies of $x^*$ or their realizations for simplicity of notation, and $f(x_j^* \mid \theta)$ is the probability density/mass function for a posited statistical model. The log-likelihood averaged over observations is denoted by $T = n^{-1}l$. Define

$$\hat{T}_{\text{ML}} = \overline{T}(\hat{\theta}_{\text{ML}} \mid X^*) = \overline{T}(\theta_{\text{ML}}(X^*) \mid X^*),$$

(2.2)

where $\hat{\theta}_{\text{ML}}$ is the MLE of the corresponding population quantity $\theta_0$. Let $\hat{\theta}_w$ be the vector of the parameter estimators by the weighted score method (WSEs) or the solution of $\theta$ satisfying

$$\frac{\partial \overline{T}(\theta \mid X^*)}{\partial \theta} + n^{-1}q^* = 0,$$

(2.3)

where $q^* = q^*(\theta)$, a function of $\theta$, is a $q \times 1$ weight vector, which becomes the log-prior derivatives in the case of Bayesian estimation but can be other general weights. Define

$$\overline{T}_w = \overline{T}(\hat{\theta}_w \mid X^*) = \overline{T}(\theta_w(X^*) \mid X^*),$$

(2.4)

whose special case is $\hat{\theta}_{\text{ML}}$ in (2.2) when $q^* = 0$. Let $Z^*$ be an independent copy of $X^*$, where $Z^*$ is interpreted as an independent data set in the future with the same sample size as $n$ from the viewpoint of prediction. Define

$$\overline{T}_0 = E_{\theta_0} \{\overline{T}(\theta_0 \mid Z^*)\} = \int_{\mathcal{R}(Z)} \overline{T}(\theta_0 \mid Z)g(Z \mid \zeta_0) dz,$$

(2.5)

where $g(Z \mid \zeta_0)$ is the true density of $Z^*$ determined by the parameter vector $\zeta_0$ of an appropriate size, and is possibly different from $f(Z \mid \theta_0)$. Equation (2.5) is to be
interpreted as the corresponding summation when \( g(Z | \zeta_0) \) is a probability mass.

Similarly, define
\[
\overline{t}_0 = \overline{T}(\theta_0 | X^*) = O_p(1) \quad \text{with} \quad E_g(\overline{t}_0) = \overline{T}'
\]
and
\[
\hat{t}_w^* = \int_{R(Z)} \overline{T}(\hat{\theta}_w | Z) g(Z | \zeta_0) dZ = \int_{R(Z)} \overline{T}(\theta_w(X^*) | Z) g(Z | \zeta_0) dZ = O_p(1).
\]

It is assumed that
\[
-2E_g(\hat{t}_w^* - \overline{t}_0^*) = n^{-1}b_1 + n^{-2}b_2 + O(n^{-3})
\]
holds, where \( n^{-1}b_1 \) and \( n^{-2}b_2 \) are defined as the asymptotic biases up to order \( O(n^{-2}) \) of \( -2\hat{t}_w^* \) whose population counterpart is \(-2E_g(\hat{t}_w^*) = O(1) \) for the AIC and TIC with \( n^{-2}b_2 \) being the higher-order added asymptotic bias.

In the following, we obtain an expression of \( b_2 \) different from that of Konishi and Kitagawa (2003) with \( b_1 \) being well known. For the expression, we use the formula of the expansion of \( \hat{\theta}_w = \theta_w(X^*) \) given by Ogasawara (2013, Equation (2.1) (see also 2015 for correction); 2014, Equation (2.4)):
\[
\hat{\theta}_w - \theta_0 = -n^{-1}\Lambda^{-1}q_0^* + \sum_{j=1}^{3} \Lambda^{(j)}l_0^{(j)} - n^{-1}(\hat{L}_w q_w^* - \Lambda^{-1}q_0^*)_{O_p(n^{-1/2})} + O_p(n^{-2})
\]
\[
= -n^{-1}\Lambda^{-1}q_0^* + \sum_{j=1}^{3} \Lambda^{(j)}l_0^{(j)} + n^{-1}\left[ \Lambda^{-1} \Lambda \Lambda^{-1}q_0^* - \Lambda^{-1} \frac{\partial q^*}{\partial \theta_0'} \Lambda^{(1)(1)} + \Lambda^{-1} \left[ \Lambda^{-1} \Lambda \Lambda^{-1}q_0^* \right] 2\right] + O_p(n^{-2})
\]
\[
\equiv -n^{-1}\Lambda^{-1}q_0^* + \sum_{j=1}^{3} \Lambda^{(j)}l_0^{(j)} + n^{-1}(l_0^{(w)})_{O_p(n^{-1/2})} + O_p(n^{-2}),
\]
where \( \Lambda = E_g(\partial^2 \overline{T} / \partial \theta \partial \theta' | \theta_0 = \theta_0) = E_g(\partial^2 \overline{T} / \partial \theta_0 \partial \theta'_0') = O(1), \quad q_0^* = q^*(\theta_0), \)
\( \Lambda^{(j)} = O(1), \quad l_0^{(j)} = O_p(n^{-j/2}) \) \( (j = 1, 2, 3), \quad \hat{L}_w = \frac{\partial^2 \overline{T}}{\partial \theta \partial \theta'} |_{\theta = \hat{\theta}_w} \equiv \frac{\partial^2 \overline{T}}{\partial \theta_w \partial \theta'_w}, \)
\[
5
\]
\( \mathbf{q}_w^* = \mathbf{q}^* (\hat{\mathbf{\theta}}_w), \) \( \mathbf{M} = \frac{\partial^2 \mathbf{T}}{\partial \mathbf{\theta}_0 \partial \mathbf{\theta}_0'}, - \mathbf{\Lambda} = O_p(n^{-1/2}), \frac{\partial \mathbf{q}_w^*}{\partial \mathbf{\theta}_0'} = \frac{\partial \mathbf{q}_w^* (\mathbf{\theta})}{\partial \mathbf{\theta}' \bigg|_{\mathbf{\theta}_0 = \mathbf{\theta}_0},} \\

\( \mathbf{J}_0^{(3)} = \frac{\partial^3 \mathbf{T}}{\partial \mathbf{\theta}_0 (\partial \mathbf{\theta}_0')^{<2>}}, \mathbf{x}^{<k>} = \mathbf{x} \otimes \cdots \otimes \mathbf{x} \) (\( k \) times of \( \mathbf{x} \)), \( \otimes \) denotes the Kronecker product, and \( (\cdot)^{O_p(n^{-1/2})} \) indicates that \( (\cdot) \) is of order \( O_p(n^{-1/2}) \) with other similar expressions.

The term \( \sum_{j=1}^{3} \Lambda^{(j)} \mathbf{l}_0^{(j)} \) in (2.9) (Ogasawara, 2010, Equation (2.4)) is given from the following expansion:

\[
\hat{\mathbf{\theta}}_{ML} - \mathbf{\theta}_0 = \sum_{j=1}^{3} \Lambda^{(j)} \mathbf{l}_0^{(j)} + O_p(n^{-2}), \tag{2.10}
\]

\[
\Lambda^{(1)} \mathbf{l}_0^{(1)} = -\mathbf{\Lambda}^{-1} \mathbf{\Lambda} \frac{\partial \mathbf{I}}{\partial \mathbf{\theta}_0},
\]

\[
\Lambda^{(2)} \mathbf{l}_0^{(2)} = \mathbf{\Lambda}^{-1} \mathbf{M} \mathbf{\Lambda}^{-1} \frac{\partial \mathbf{I}}{\partial \mathbf{\theta}_0} - \frac{1}{2} \mathbf{\Lambda}^{-1} \mathbf{E}_g (\mathbf{J}_0^{(3)}) \left( \mathbf{\Lambda}^{-1} \frac{\partial \mathbf{I}}{\partial \mathbf{\theta}_0} \right)^{<2>}
\]

\[
\Lambda^{(3)} \mathbf{l}_0^{(3)} = -\mathbf{\Lambda}^{-1} \mathbf{M} \mathbf{\Lambda}^{-1} \mathbf{M} \mathbf{\Lambda}^{-1} \frac{\partial \mathbf{I}}{\partial \mathbf{\theta}_0} + \frac{1}{2} \mathbf{\Lambda}^{-1} \mathbf{M} \mathbf{\Lambda}^{-1} \mathbf{E}_g (\mathbf{J}_0^{(3)}) \left( \mathbf{\Lambda}^{-1} \frac{\partial \mathbf{I}}{\partial \mathbf{\theta}_0} \right)^{<2>}
\]

\[
+ \mathbf{\Lambda}^{-1} \mathbf{E}_g (\mathbf{J}_0^{(3)}) \left[ \left( \mathbf{\Lambda}^{-1} \mathbf{M} \mathbf{\Lambda}^{-1} \frac{\partial \mathbf{I}}{\partial \mathbf{\theta}_0} \right) \otimes \left( \mathbf{\Lambda}^{-1} \frac{\partial \mathbf{I}}{\partial \mathbf{\theta}_0} \right) \right] - \frac{1}{2} \mathbf{\Lambda}^{-1} (\mathbf{J}_0^{(3)} - \mathbf{E}_g (\mathbf{J}_0^{(3)})) \left( \mathbf{\Lambda}^{-1} \frac{\partial \mathbf{I}}{\partial \mathbf{\theta}_0} \right)^{<2>}
\]

\[
- \frac{1}{2} \mathbf{\Lambda}^{-1} \mathbf{E}_g (\mathbf{J}_0^{(3)}) \left[ \left( \mathbf{\Lambda}^{-1} \frac{\partial \mathbf{I}}{\partial \mathbf{\theta}_0} \right) \otimes \left( \mathbf{\Lambda}^{-1} \mathbf{E}_g (\mathbf{J}_0^{(3)}) \left( \mathbf{\Lambda}^{-1} \frac{\partial \mathbf{I}}{\partial \mathbf{\theta}_0} \right)^{<2>} \right) \right]
\]

\[
+ \frac{1}{6} \mathbf{\Lambda}^{-1} \mathbf{E}_g (\mathbf{J}_0^{(4)}) \left( \mathbf{\Lambda}^{-1} \frac{\partial \mathbf{I}}{\partial \mathbf{\theta}_0} \right)^{<3>},
\]
\[ J^{(4)}_0 \equiv \frac{\partial^4 \mathcal{T}}{\partial \theta_0 (\partial \theta_0)'^{<3>}} , \quad l^{(1)}_0 = \frac{\partial \mathcal{T}}{\partial \theta_0} , \]

\[ l^{(2)}_0 = \left\{ v'(M) \otimes \frac{\partial \mathcal{T}}{\partial \theta_0}' , \left( \frac{\partial \mathcal{T}}{\partial \theta_0}' \right)^{<2>} \right\}' , \quad \equiv (l^{(2-1)}_0 , l^{(2-2)}_0)' = O_p(n^{-1}) , \]

\[ l^{(3)}_0 = \left[ v'(M)^{<2>} \otimes \frac{\partial \mathcal{T}}{\partial \theta_0}' , v'(M) \otimes \left( \frac{\partial \mathcal{T}}{\partial \theta_0}' \right)^{<2>} , \text{vec}' \{ J^{(3)}_0 - E_g (J^{(3)}_0) \} \otimes \left( \frac{\partial \mathcal{T}}{\partial \theta_0}' \right)^{<2>} \right] \]

\[ \equiv (l^{(3-1)}_0 , l^{(3-2)}_0 , l^{(3-3)}_0 , l^{(3-4)}_0)' = O_p(n^{-3/2}) , \]

where \( l^{(2-j)}_0 = O(1) \) \((j = 1, 2)\) and \( l^{(3-j)}_0 = O(1) \) \((j = 1, \ldots , 4)\) are defined implicitly by

\[ \Lambda^{(2)} l^{(2)}_0 = \sum_{j=1}^{2} \Lambda^{(2-j)} l^{(2-j)}_0 \quad \text{and} \quad \Lambda^{(3)} l^{(3)}_0 = \sum_{j=1}^{4} \Lambda^{(3-j)} l^{(3-j)}_0 ; \quad v'(M)^{<2>} = \{ v'(M) \}^{<2>} ; \quad v(\cdot) \]

is the vectorizing operator taking the non-duplicated elements of a symmetric matrix in parentheses; and \( \text{vec}(\cdot) \) is the vectorizing operator stacking the columns of a matrix sequentially.

Expand \( -2 \hat{\mathcal{L}}_w \) and \( -2 \hat{\mathcal{L}}^*_w \) as

\[ -2 \hat{\mathcal{L}}_w = -2 (\bar{t}_0)_{O_p(1)} - 2 \sum_{j=1}^{4} \frac{1}{j!} \left\{ \left( \frac{\partial \mathcal{T}}{\partial \theta_0} \right)^{<j>} \right\}_{O_p(n^{-j/2})} + O_p(n^{-3/2}) \quad (2.11) \]

and \( -2 \hat{\mathcal{L}}^*_w = -2 (\bar{t}^*_0)_{O(1)} - 2 \sum_{j=1}^{4} \frac{1}{j!} E_g \left\{ \left( \frac{\partial \mathcal{T}}{\partial \theta_0} \right)^{<j>} \right\}_{O(1)} + O_p(n^{-3/2}) , \)

respectively. Then, recalling \( E_g (\bar{t}_0) = \bar{t}_0^* \), we have
\[-2E_g \left( \hat{I}_w - \hat{I}_w^* \right) \]

\[= -2E_g \left[ \sum_{j=1}^{3} \frac{1}{j!} \left( \frac{\partial^j \hat{T}}{(\partial \theta_0)^{<j>}} \right) - E_g \left\{ \frac{\partial^j \hat{T}}{(\partial \theta_0)^{<j>}} \right\} \right]_{O_p(n^{-2})} + O(n^{-3}) \]

\[= -2E_g \left\{ \frac{\partial \hat{T}}{\partial \theta_0} (\hat{\theta}_w - \theta_0) \right\} \rightarrow O(n^{-2}) - E_g \left\{ \text{vec}'(M)(\hat{\theta}_w - \theta_0)^{<2>} \right\} \rightarrow O(n^{-2}) \]

\[-\frac{1}{3} \frac{1}{2} \left. \left( \frac{\partial \hat{T}}{\partial \theta_0} \right) \right|_{O(n^{-2})} \frac{\partial \hat{T}}{\partial \theta_0} \left( \hat{\theta}_w - \theta_0 \right) \right\} \rightarrow O(n^{-2}) + O(n^{-3}), \]

where the term of \( j = 4 \) in the remainder term of order \( O(n^{-3}) \); and \( E_g (\cdot) \rightarrow O(n^{-2}) \) indicates that the expectation is taken up to order \( O(n^{-2}) \).

Let \( \Gamma = n E_g \left( \frac{\partial \hat{T}}{\partial \theta_0} \frac{\partial \hat{T}}{\partial \theta_0} \right) \). When the model is true, \( \Gamma = -\Lambda = I_0 \), where \( I_0 \) is the population Fisher information matrix per observation. Under possible model misspecification, the last three expectations in (2.12) are given as

\[-2E_g \left\{ \frac{\partial \hat{T}}{\partial \theta_0} (\hat{\theta}_w - \theta_0) \right\} \]

\[= -2E_g \left\{ \frac{\partial \hat{T}}{\partial \theta_0} \left( -n^{-1} \Lambda^{-1} q_0^* + \sum_{j=1}^{3} \Lambda^{(j)} I_0^{(j)} + n^{-1} I_0^{(W)} \right) \right\} \]

\[= \left\{ 2E_g \left( \frac{\partial \hat{T}}{\partial \theta_0}, \Lambda^{-1} \frac{\partial \hat{T}}{\partial \theta_0} \right) \right\}_{O(n^{-2})} - \left\{ 2E_g \left( \frac{\partial \hat{T}}{\partial \theta_0}, \Lambda^{(2)} I_0^{(2)} \right) \right\}_{O(n^{-2})} \]

\[= \left\{ 2E_g \left( \frac{\partial \hat{T}}{\partial \theta_0}, \Lambda^{(3)} I_0^{(3)} \right) \right\}_{O(n^{-2})} - \left\{ 2E_g \left( n^{-1} \frac{\partial \hat{T}}{\partial \theta_0}, I_0^{(W)} \right) \right\}_{O(n^{-2})} + O(n^{-3}) \]
\[= 2n^{-1}\text{tr}(\Lambda^{-1}\Gamma) - 2n^{-2}\left[\sum_{a \geq b}^{q} \sum_{c,d=1}^{n} (\Lambda^{(2-1)})_{(d; ab, c)} n^2 E_g \left( m_{ab} \frac{\partial T}{\partial \theta_{0c}} \frac{\partial T}{\partial \theta_{0d}} \right) \right.\]

\[+ \sum_{a,b,c=1}^{q} (\Lambda^{(2-2)})_{(c; a,b)} n^2 E_g \left( \frac{\partial T}{\partial \theta_{0a}} \frac{\partial T}{\partial \theta_{0b}} \right) + \sum_{a,b,c,d,e,f=1}^{q} (\Lambda^{(3-1)})_{(f; ab, cd, e)} n^2 E_g \left( m_{ab} \frac{\partial T}{\partial \theta_{0e}} \right) + \sum_{a,b,c,d,e,f=1}^{q} (\Lambda^{(3-1)})_{(f; ab, cd, e)} n^2 E_g \left( m_{cd} \frac{\partial T}{\partial \theta_{0f}} \right)\]

\[\times \left\{ n \text{cov}_g (m_{ab}, m_{cd}) \gamma_{ef} + 2n \text{cov}_g \left( m_{ab} \frac{\partial T}{\partial \theta_{0e}} \right) n \text{cov}_g \left( m_{cd} \frac{\partial T}{\partial \theta_{0f}} \right) \right\} \]

\[+ \sum_{a,b,c,d,e,f=1}^{q} (\Lambda^{(3-2)})_{(e; ab, cd, d)} \sum_{(e,d,e)}^{3} n \text{cov}_g \left( m_{ab} \frac{\partial T}{\partial \theta_{0e}} \right) \gamma_{de} + \sum_{a,b,c,d,e,f=1}^{q} (\Lambda^{(3-3)})_{(f; abc, d, e)} \sum_{(d,e,f)}^{3} n \text{cov}_g \left( (J^{(3)})_{(a,b,c)} \right) \gamma_{ef} + \sum_{a,b,c,d,e,f=1}^{q} (\Lambda^{(3-4)})_{(d,a,b,e,c)} \gamma_{ab} \gamma_{cd} + \gamma_{ac} \gamma_{bd} + \gamma_{ad} \gamma_{bc}\]

\[+ \sum_{a,b,c=1}^{q} \lambda^{ab} (\Lambda^{-1} q_0^*) \text{cov}_g \left( m_{ab} \frac{\partial T}{\partial \theta_{0d}} \right) + \left( \frac{\partial q^*}{\partial \theta_0} \right) \Lambda^{-1} \Gamma \Lambda^{-1} \right)\]

\[- \text{tr}[E_g (J^{(3)}) \left( (\Lambda^{-1} q_0^*) \otimes (\Lambda^{-1} \Gamma \Lambda^{-1}) \right)] + O(n^{-3})\]

\[\equiv n^{-1} b_1 + n^{-2} c_1 + O(n^{-3}) \quad (b_1 = 2\text{tr}(\Lambda^{-1} \Gamma), c_1 = -2 \frac{[ \cdot ]}{(A)} (A))\]

where \((\Lambda^{(2-1)})_{(d; ab, c)}\) indicates the element of the \(d\)-th row and the column corresponding to \((M)_{ab} = m_{ab}\) (the \((a, b)\)th element of \(M\)) and \(\frac{\partial T}{\partial (\theta_0)_{c}} = \frac{\partial T}{\partial \theta_{0c}}\) of \(\Lambda^{(2-1)}\) with \((\cdot)_{c}\) being the \(c\)-th element of a vector with other expressions defined similarly;

\[\sum_{a \geq b}^{q} \sum_{b=1}^{a} \sum_{a=1}^{q} (\cdot), \sum_{e,f=1}^{q} (\cdot) = \sum_{e=1}^{q} \sum_{f=1}^{q} (\cdot); \text{cov}_g (\cdot)\) is the covariance using the distribution \(g(X^* | \xi_0)\); \(\sum_{(c,d,e)}^{3} (\cdot)\) is the sum of three symmetric terms with respect to \(c, d\) and \(e\); and \([ \cdot ]\) is for ease of finding correspondence;
\[-E_g \{ \text{vec}'(M)(\hat{\theta}_w - \theta_0)^{<2>} \} \]

\[= -E_g \left[ \text{vec}'(M) \left\{ 2(-n^{-1} \Lambda^{-1} q_0^*) \otimes \left( -\Lambda^{-1} \frac{\partial I}{\partial \theta_0} + \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right)_2 \right\}^{<2>} \right. \]

\[+ 2 \left( -\Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right) \otimes \left( \Lambda^{(2)} I_0^{(2)} \right) \right] \]

\[= -n^{-2} \left[ \sum_{a,b,c=1}^{q} (\Lambda^{-1} q_0^*)_a^{bc} n \text{cov}_g \left( m_{ab}, \frac{\partial I}{\partial \theta_{0c}} \right) + \sum_{a,b,c,d=1}^{q} \lambda^{ac} \lambda^{bd} n^2 E_g \left( m_{ab}, \frac{\partial I}{\partial \theta_{0c}}, \frac{\partial I}{\partial \theta_{0d}} \right) \right. \]

\[+ 2 \sum_{a,b,c=1}^{q} \sum_{d=2}^{q} \left( \Lambda^{(2-1)} \right)_{(b,d,c)}^{ac} n \text{cov}_g \left( m_{ab}, \frac{\partial I}{\partial \theta_{0c}} \right) n \text{cov}_g \left( m_{de}, \frac{\partial I}{\partial \theta_{0f}} \right) \]

\[+ n \text{cov}_g (m_{ab}, m_{de}) \gamma_{ef} \right] \]

\[= n^{-2} c_2 + O(n^{-3}), \]

\[-\frac{1}{3} E_g \{ \text{vec}' \{ J_0^{(3)} - E_g (J_0^{(3)}) \} (\hat{\theta}_w - \theta_0)^{<3>} \} \]

\[= -\frac{1}{3} E_g \left[ \text{vec}' \{ J_0^{(3)} - E_g (J_0^{(3)}) \} \left( -\Lambda^{-1} \frac{\partial I}{\partial \theta_{0d}} \right) \right] + O(n^{-3}) \]

\[= n^{-2} \sum_{a,b,c,d,e,f=1}^{q} \lambda^{ad} \lambda^{be} \lambda^{cf} n \text{cov}_g \left( J_0^{(3)}, \frac{\partial I}{\partial \theta_{0d}} \right) \gamma_{ef} + O(n^{-3}) \]

\[= n^{-2} c_3 + O(n^{-3}), \]

where \( \lambda^{bc} = (\Lambda^{-1})^{bc}. \) Then, from (2.13) to (2.15),

**Theorem 1.** Under (2.8) with regularity conditions for (2.9) and (2.10), the asymptotic biases \( n^{-1}b_1 \) and \( n^{-2}b_2 \) of \( -2\hat{\theta}_w \) up to order \( O(n^{-3}) \), based on the WSE \( \hat{\theta}_w \) derived by the estimation equation of (2.3), are given by

\[-2E_g (\hat{\theta}_w - \hat{\theta}_w) \]

\[= n^{-1} 2 \text{tr}(\Lambda^{-1} \Gamma) + n^{-2} (c_1 + c_2 + c_3) + O(n^{-3}) = n^{-1} b_1 + n^{-2} b_2 + O(n^{-3}), \]
where \( c_1, c_2 \) and \( c_3 \) are obtained by (2.13) to (2.15), respectively.

From (2.13) to (2.15), we find that \( b_1 \) and \( c_3 \) do not depend on \( q_0^* \) and are common to the results by the MLE \( \hat{\theta}_{ML} \) and the WSE \( \hat{\theta}_W \) while \( c_1 \) and \( c_2 \) depend on \( q_0^* \). A considerably simplified result is obtained in the following case.

**Corollary 1.** When the vector of canonical parameters in the exponential family of distributions is used under possible model misspecification,

\[
-2 E_g (\hat{I}_w - \hat{I}_w^*) = n^{-1} b_1 + n^{-2} c_1 + O(n^{-3}) \quad \text{with} \quad b_2 = c_1 \quad \text{and} \quad c_2 = c_3 = 0, \quad (2.17)
\]

where \( c_1 \) is simplified as

\[
-2 E_g \left\{ \frac{\partial^T}{\partial \theta_0} (\hat{\theta}_w - \theta_0) \right\} = 2 n^{-1} \text{tr}(\Lambda^{-1} \Gamma)
\]

\[
-2 n^{-2} \left[ \sum_{a,b,c=1}^q (\Lambda^{(2-2)})_{(c:a,b)} n^2 E_g \left( \frac{\partial^T}{\partial \theta_{0a}} \frac{\partial^T}{\partial \theta_{0b}} \frac{\partial^T}{\partial \theta_{0c}} \right) + \sum_{a,b,c,d=1}^q (\Lambda^{(3-4)})_{(d,a,b,c)} (\gamma_{ab} \gamma_{cd} + \gamma_{ac} \gamma_{bd} + \gamma_{ad} \gamma_{bc}) \right]
\]

\[
\quad + \text{tr} \left[ \frac{\partial q_0^*}{\partial \theta_0} \Lambda^{-1} \Gamma \right] - \text{tr} \left[ E_g (J_0^{(3)}) \{ (\Lambda^{-1} q_0^*) \otimes (\Lambda^{-1} \Gamma \Lambda^{-1}) \} \right] + O(n^{-3})
\]

\[= n^{-1} b_1 + n^{-2} c_1 + O(n^{-3}). \quad (2.18)
\]

Proof. Under canonical parametrization in the exponential family, it is known that

\[
\frac{\partial^T}{\partial \theta_0} \left( \frac{\partial^T}{\partial \theta_0} \right)_{(j:j)} = E_g \left( \frac{\partial^T}{\partial \theta_0} \left( \frac{\partial^T}{\partial \theta_0} \right)_{(j:j)} \right) \quad (j = 2,3,\ldots), \quad \text{which gives} \quad c_1 \quad \text{of} \quad (2.18) \quad \text{from} \quad (2.13) \quad \text{with} \quad M = O
\]

and \( J_0^{(3)} - E_g (J_0^{(3)}) = O \). The results of \( c_2 = c_3 = 0 \) are derived similarly from (2.14) and (2.15) with \( M = O \) and \( J_0^{(3)} - E_g (J_0^{(3)}) = O \), respectively. Q.E.D.

In the case of the MLE, the two terms associated with \( q_0^* \) in (2.18) vanish and recalling (2.10) for \( \Lambda^{(2-2)} \) and \( \Lambda^{(3-4)} \) in \( c_1 \) of (2.18), we have
\[ c_1 = -2 \left\{ \begin{array}{c} \sum_{a,b,c=1}^q (\Lambda^{(2-2)}(c,a,b)^{\alpha}E_{g} \left( \frac{\partial T}{\partial \theta_{0a}} \frac{\partial T}{\partial \theta_{0b}} \frac{\partial T}{\partial \theta_{0c}} \right) \\
+ \sum_{a,b,c,d=1}^q (\Lambda^{(3-4)}(d,a,b,c) \{ \gamma_{ab} + \gamma_{ac} + \gamma_{ad} \}) \end{array} \right\} \\
= -2 \left\{ \begin{array}{c} \sum_{a,b,c=1}^q \left\{ -\frac{1}{2} \Lambda^{-1}J_0^{(3)}(\Lambda^{-1})^{<2>} \right\} E_{g} \left( \frac{\partial T}{\partial \theta_{0a}} \frac{\partial T}{\partial \theta_{0b}} \frac{\partial T}{\partial \theta_{0c}} \right) \\
- \sum_{a,b,c,d=1}^q \left\{ \frac{1}{2} \Lambda^{-1}J_0^{(3)}(\Lambda^{-1})_{a} \otimes \left[ \Lambda^{-1}J_0^{(4)}(\Lambda^{-1})_{b} \otimes (\Lambda^{-1})_{c} \right] \right\} \end{array} \right\} \\
= -\text{vec}(J_0^{(3)}) \otimes n^2 E_{g} \left\{ \left( -\Lambda^{-1} \frac{\partial T}{\partial \theta_{0a}} \right)^{<3>} \right\} + \text{vec}(\Lambda^{-1} \Gamma \Lambda^{-1}) \text{vec}(\Lambda^{-1} \Gamma \Lambda^{-1}) \\
+ 2\text{vec}(J_0^{(3)}) \otimes (\Lambda^{-1} \Gamma \Lambda^{-1})^{<2>} \text{vec}(J_0^{(3)}) - \text{vec}(J_0^{(4)}) \otimes \text{vec}(\Lambda^{-1} \Gamma \Lambda^{-1})^{<2>} \right\}, \quad (2.19) \]

where \((\cdot)_{d\square}\) is the \(d\)-th row of a matrix and \((\cdot)_{a\square}\) is the \(a\)-th column of a matrix.

Under correct model specification and canonical parametrization, since

\[ \frac{\partial l_j}{\partial \theta_0} = \mathbf{x}^* - \mathbf{E}_j(\mathbf{x}^*) \] and \(-\Lambda = \Gamma = \mathbf{I}_0\), (2.19) becomes

\[ c_1 = \kappa_{f3}(\mathbf{x}^*) \kappa_{f3} \left( \mathbf{I}_0^{-1} \frac{\partial l_j}{\partial \theta_0} \right) - \text{vec}(\mathbf{I}_0^{-1}) \text{vec}(\mathbf{I}_0^{-1}) \text{vec}(\mathbf{I}_0^{-1}) \\
- 2\text{vec}(J_0^{(3)}) \otimes \text{vec}(J_0^{(3)}) + \kappa_{f4}(\mathbf{x}^*) \text{vec}(\mathbf{I}_0^{-1}) \text{vec}(\mathbf{I}_0^{-1}) \\
= \kappa_{f3}(\mathbf{I}_0^{-1/2} \mathbf{x}^*) \kappa_{f3} \left( \mathbf{I}_0^{-1/2} \frac{\partial l_j}{\partial \theta_0} \right) - \kappa_{f3}(\mathbf{I}_0^{-1/2} \mathbf{x}^*) \left[ \mathbf{I}_q \otimes \{ \text{vec}(\mathbf{I}_q) \} \right] \kappa_{f3}(\mathbf{I}_0^{-1/2} \mathbf{x}^*) \quad (2.20) \]

where \(\mathbf{x}^*\) is the \(q \times 1\) vector of observable variables associated with the minimum sufficient statistics \((p = q)\); \(\mathbf{I}_0^{-1/2} \mathbf{x}^*\); \(\kappa_{f_j}(\cdot)\) is the \(q' \times 1\) vector of the \(j\)-th
multivariate cumulants of a $q \times 1$ random vector in parentheses using the distribution $f(x^* | \theta_0)$ for $x^*$; for $l_j (j = 1, ..., n)$ see (2.1); $I_0^{-1/2}$ is a non-negative definite symmetric matrix-square-root of $I_0$ with $I_0^{-1/2} = (I_0^{1/2})^{-1}$ under the assumption of its existence; and $I_{(q)}$ is the $q \times q$ identity matrix.

Under correct model specification, since $\text{cov}_f(x^*) = I_0$ due to canonical parametrization, $\hat{x}^*$ is the vector of standardized variables with

$$\text{cov}_f(\hat{x}^*) = \text{cov}_f \left( I_0^{-1/2} \frac{\partial l_j}{\partial \theta_0} \right) \equiv \text{cov}_f \left( \frac{\partial \tilde{l}_j}{\partial \theta_0} \right) = I_{(q)} , \quad \text{where } \text{cov}_f(\cdot) \text{ is the exact covariance matrix using } f(x^* | \theta_0) . \text{ Then, } \kappa_{f3}(\hat{x}^*) \text{ and } \kappa_{f3}(\partial \tilde{l}_j / \partial \theta_0) (= \kappa_{f3}(\hat{x}^*)) \text{ are seen as } q^3 \times 1 \text{ vectors of the multivariate skewnesses of } \hat{x}^* \text{ and } \partial \tilde{l}_j / \partial \theta_0 , \text{ respectively. Similarly, } \kappa_{f4}(\hat{x}^*) \text{ is seen as a } q^4 \times 1 \text{ vector of the multivariate kurtoses of } \hat{x}^* . \text{ In the univariate case, (2.20) becomes the sum of } -2 \text{ times the squared skewness and the excess kurtosis.}

Similarly, under correct model specification, $b_1$ in the asymptotic bias of order $O(n^{-1})$ in (2.18) is also written as

$$b_1 = 2 \text{tr}(\Lambda^{-1} \Gamma) = -2q = -2 \text{vec}'(I_0) \text{vec}(I_0^{-1}) = -2 \kappa_{f2}'(x^*) \kappa_{f3} \left( I_0^{-1/2} \frac{\partial l_j}{\partial \theta_0} \right)$$

$$= -2 \kappa_{f2}'(\hat{x}^*) \kappa_{f2}(\hat{x}^*) = -2 \kappa_{f2}'(\hat{x}^*) \kappa_{f2}(\hat{x}^*) \quad (2.21)$$

The above results give

**Corollary 2.** Under correct model specification and canonical parametrization in the exponential family, when the multivariate skewnesses and kurtoses of the associated observable variables are zero, the MLE gives

$$-2E_f(\tilde{L}_{\text{ML}} - \tilde{L}_{\text{ML}}) = -n^{-1} 2q + O(n^{-3}) \quad (b_1 = -2q, b_2 = c_1 = c_2 = c_3 = 0) \quad (2.22)$$

where $E_f(\cdot)$ is defined using $f(x^* | \theta_0)$ similarly to $E_g(\cdot)$.

This can happen, for example, when the covariance matrix in the multivariate normal distribution is known, where the vector of canonical parameters is the mean vector. Since
\[
J^{(j)}_0 = \frac{\partial^2 I}{\partial \theta_0 (\partial \theta_0 ')^{j-1}} = E_g \left\{ \frac{\partial^2 I}{\partial \theta_0 (\partial \theta_0 ')^{j-1}} \right\} \quad (j = 2, 3, \ldots) \quad \text{under canonical parametrization,}
\]

the asymptotic expansion using the MLE corresponding to (2.12) higher than (2.12) is given only by the first term \(-2E_g \left\{ \frac{\partial^2 I}{\partial \theta_0 } (\hat{\theta}_{ML} - \theta_0 ) \right\} \), which is also given only by

\[
-2E_g \left\{ \frac{\partial^2 I}{\partial \theta_0 } \left( -\Lambda^{-1} \frac{\partial I}{\partial \theta_0 } \right) \right\} \quad \text{and} \quad -2E_g \{h(J^{(3)}_0 , J^{(4)}_0 , \ldots)\}, \quad \text{where} \quad h(\cdot) \quad \text{is the sum of multiplicative functions of the powers of the arguments. Then, we have}
\]

**Corollary 3.** When the covariance matrix \( \Sigma \) of the q-variate normal distribution is known, the MLE (the usual sample mean vector \( \overline{x} \)) of the population mean vector \( \mu_0 \) under possible model misspecification gives

\[
-2E_g (\hat{I}_{ML} - \hat{I}^*_ML) = -n^{-1}2q
\]

(2.23)

**Proof.** In the only non-vanishing term \(-2E_g \left\{ \frac{\partial^2 I}{\partial \theta_0 } (\hat{\theta}_{ML} - \theta_0 ) \right\} \) for the expansion of the left-hand side of (2.23),

\[
-2E_g \left\{ \frac{\partial^2 I}{\partial \theta_0 } \left( -\Lambda^{-1} \frac{\partial I}{\partial \theta_0 } \right) \right\} = -2\text{tr} \left\{ \Sigma^{-1}E_g \left\{ \frac{\partial^2 I}{\partial \theta_0 } \frac{\partial I}{\partial \theta_0 } \right\} \right\}
\]

= \(-n^{-1}2\text{tr}(\Sigma^{-1}\Sigma) = -n^{-1}2q \quad \text{under arbitrary distributions as long as} \quad \Sigma \quad \text{and} \quad \Sigma^{-1} \quad \text{exist. The remaining terms} \quad -2E_g \{h(J^{(3)}_0 , J^{(4)}_0 , \ldots)\} \quad \text{vanish when we use the normal distribution even under non-normality since} \quad J^{(j)}_0 = O(j = 3, 4, \ldots) \quad \text{in this case. Q.E.D.}
\]

Note that there is no remainder term in (2.23). An alternative direct proof of Corollary 3 is given as follows. Let \( z_j (j = 1, \ldots, n) \) be independent copies of \( x^* \) and \( E_{Z^*}(\cdot) \) denote an expectation over the distribution of \( Z^* \) or \( z_j (j = 1, \ldots, n) \). Then, by definition,

\[
-2\hat{I}^*_ML = -2E_{Z^*} \left[ \frac{-n^{-1}}{2} \sum_{j=1}^{n} (z_j - \overline{x})' \Sigma^{-1} (z_j - \overline{x}) - \frac{1}{2} \log \{(2\pi)^q | \Sigma|\} \right]
\]

= \( \text{tr}(\Sigma^{-1}\Sigma) + (\mu_0 - \overline{x})' \Sigma^{-1} (\mu_0 - \overline{x}) + \log \{(2\pi)^q | \Sigma|\} \)

= \( q + (\mu_0 - \overline{x})' \Sigma^{-1} (\mu_0 - \overline{x}) + \log \{(2\pi)^q | \Sigma|\} \)

(2.24)
which gives \(-2E_g(\hat{I}_{ML}^*) = (1 + n^{-1})q + \log \{(2\pi)^q \mid \Sigma\}\). On the other hand,

\[
-2E_g(\hat{I}_{ML}^*) = -2E_g \left[ -\frac{n^{-1}}{2} \sum_{j=1}^{n} (x_j - \bar{x})' \Sigma^{-1} (x_j - \bar{x}) - \frac{1}{2} \log \{(2\pi)^q \mid \Sigma\} \right] 
= (1 - n^{-1}) \text{tr}(\Sigma^{-1} \Sigma) + \log \{(2\pi)^q \mid \Sigma\}
= (1 - n^{-1})q + \log \{(2\pi)^q \mid \Sigma\}.
\]

(2.25)

Consequently, (2.24) and (2.25) yield \(-2E_g(\hat{I}_{ML}^* - \hat{I}_{ML}^*) = -n^{-1}2q\).

3. Bias correction for the AIC and TIC

Define

\[
n^{-1}\text{AIC}_W = -2\hat{I}_W + n^{-1}2q,
\]

(3.1)

\[
n^{-1}\text{TIC}^{(1)}_W = -2\hat{I}_W + n^{-1}\text{tr}(\hat{L}_W^-\hat{\Gamma}_W^-)
\]

and \(n^{-1}\text{TIC}^{(2)}_W = -2\hat{I}_W + n^{-1}\text{tr}(\hat{I}_{W^{(\Lambda)}}^-\hat{I}_{W}^{(\Gamma)})\) with \(\hat{I}_{W}^{(\Lambda)} = (\hat{I}_{W^{(\Lambda)}})^{-1}\)

where

\[
\hat{L}_W = \frac{\partial^2\ell}{\partial\hat{\theta}_W \partial\hat{\theta}_W}, \quad \hat{\Gamma}_W = n^{-1} \sum_{j=1}^{g} \frac{\partial l_j}{\partial\hat{\theta}_W} \frac{\partial l_j}{\partial\hat{\theta}_W}, \quad \hat{I}_{W^{(\Lambda)}} = \left\{-E_g \left( \frac{\partial^2\ell}{\partial\theta' \partial\theta} \right) \right\}_{\hat{\theta} \to \theta_W}
\]

and

\[
\hat{I}_{W}^{(\Gamma)} = \left\{E_g \left( \frac{\partial l_j}{\partial\theta} \frac{\partial l_j}{\partial\theta'} \right) \right\}_{\hat{\theta} \to \theta_W}.
\]

(3.2)

When the MLE is used, the subscript \(W\) in (3.1) becomes ML with \(\text{AIC}_{ML} = \text{AIC}\) (the usual AIC), \(\text{TIC}_{ML}^{(j)} = \text{TIC}^{(j)}(j = 1,2)\). The original definition of the Takeuchi information criterion (Takeuchi, 1976, Equation (15)) denoted by \(\text{TIC}_{ML} = \text{TIC}\) seems to be \(\text{TIC}_{ML}^{(2)} = \text{TIC}^{(2)}\) in (3.1), while the definition of the TIC by Linhart and Zucchini (1986, p.245), Konishi and Kitagawa (2008, p.60) and Burnham and Anderson (2010, Subsection 7.3.1) is \(\text{TIC}_{ML}^{(1)} = \text{TIC}^{(1)}\) in (3.1). The two matrices \(-\hat{L}_W\) and \(\hat{\Gamma}_W\) are observed information matrices given by \(\hat{\theta}_W\) and \(X\cdot\), which are estimators of \(-\Lambda\) and \(\Gamma\), respectively, and become the estimators of \(I_0\) under correct model specification. The two matrices \(\hat{I}_{W^{(\Lambda)}}\) and \(\hat{I}_{W}^{(\Gamma)}\) are also estimators of \(-\Lambda\) and \(\Gamma\), respectively, and are the
expected information matrices followed by estimation using $\hat{\theta}_w$ without $X^*$ except in $\theta_w(X^*)$. Since it is often difficult to derive the expectation $E_g(\cdot)$ in (3.2) when $g(x^* | \zeta_0)$ is unknown, $n^{-1}\text{TIC}^{(1)}_w$ is of practical use though $n^{-1}\text{TIC}^{(1)}_w$ is more complicated than $n^{-1}\text{TIC}^{(2)}_w$. The remaining combinations $n^{-1}\text{tr}(-\hat{L}_w^{-1}\hat{I}^{(f)}_w)$ and $n^{-1}\text{tr}(\hat{I}_w^{(-A)^{-1}}\hat{\Gamma}_w)$ for the correction term are not dealt with in this paper.

The higher order bias correction of $n^{-1}\text{AIC}_w$ is meaningless under model misspecification since the term $n^{-1}2q$ for bias correction is incorrect and should be replaced by that of $n^{-1}\text{TIC}^{(3)}_w$ which stands generically for $\text{TIC}^{(j)}_w (j = 1, 2)$. Consequently, this reduces to the higher-order bias correction of $n^{-1}\text{TIC}^{(3)}_w$ and will be dealt with later.

**Theorem 2.** Assume that a statistical model holds. Then, under regularity conditions, define

$$n^{-1}\text{AIC}_{w\to O(n^{-2})} \equiv n^{-1}\text{AIC}_w - n^{-2}\hat{b}_2 = -2\hat{T}_w + n^{-1}2q - n^{-2}(\hat{c}_1 + \hat{c}_2 + \hat{c}_3).$$

Then, $E_f(n^{-1}\text{AIC}_{w\to O(n^{-2})} + 2\hat{T}_w) = O(n^{-3})$, where $\hat{c}_1$, $\hat{c}_2$ and $\hat{c}_3$ are consistent estimators of $c_1$, $c_2$ and $c_3$, respectively.

In some special cases, $n^{-1}\text{AIC}_{ML}(= n^{-1}\text{AIC})$ gives the same result as that of Theorem 2 i.e., $E_f(n^{-1}\text{AIC} + 2\hat{T}_{ML}) = O(n^{-3})$. When the multivariate skewnesses and kurtoses of the associated observable variables are zero, from Corollary 2 we have this result. Similarly, when the covariance matrix of the multivariate normal distribution is known, Corollary 3 using the MLE of the mean vector gives the exact result $E_g(n^{-1}\text{AIC} + 2\hat{T}_{ML}) = 0$ even under non-normality.

For $n^{-1}\text{TIC}^{(3)}_w$, under possible model misspecification, define stochastic $\text{tr}^{(f)}_{\lambda}$ and $\text{tr}^{(f)}_{\lambda\Delta}$ in the expansions of $n^{-1}\text{TIC}^{(j)}_w (j = 1, 2)$ as follows.

**Definition 1.**
\begin{align*}
n^{-1}\text{TIC}^{(1)}_{W} &= -2\hat{I}_{W} + n^{-1}2\text{tr}(-\hat{L}_{W}^{\dagger}\hat{I}_{W}) \\
&= -2\hat{I}_{W} + n^{-1}2\text{tr}(-\Lambda^{-1}\Gamma) + 2(n^{-1}\text{tr}^{(T1)}_{\Delta}\gamma)_{\theta_{p}}(n_{-3/2}) + 2(n^{-1}\text{tr}^{(T1)}_{\Delta\Delta}\gamma)_{\theta_{p}}(n_{-5/2}) + O_{p}(n^{-5/2}) \\
(3.4)
\end{align*}

and

\begin{align*}
n^{-1}\text{TIC}^{(2)}_{W} &= -2\hat{I}_{W} + n^{-1}2\text{tr}(\hat{I}_{W}^{(\Lambda^{-A})-1}\hat{I}_{W}^{(\Gamma)}) \\
&= -2\hat{I}_{W} + n^{-1}2\text{tr}(-\Lambda^{-1}\Gamma) + 2(n^{-1}\text{tr}^{(T2)}_{\Delta\Delta}\gamma)_{\theta_{p}}(n_{-5/2}) + O_{p}(n^{-5/2}). \\
(3.5)
\end{align*}

For (3.4), let \( L_{0} = \left(\frac{\partial^{2}I}{\partial\theta_{0}\partial\theta_{0}'}\right)_{\theta_{p}(1)} \), then define stochastic \(-\Lambda^{-1}_{M}\) and \(-\Lambda^{-1}_{M}\Delta\) as follows:

\begin{align*}
-\hat{L}_{W} &= -L_{0}^{-1} + \sum_{j=1}^{q}L_{0}^{-1}\frac{\partial L_{0}}{\partial\theta_{0}}L_{0}^{-1}(\hat{\theta}_{W} - \theta_{0})_{j} \\
&+ \sum_{j,k=1}^{q}\left\{-L_{0}^{-1}\frac{\partial L_{0}}{\partial\theta_{0}}L_{0}^{-1}\frac{\partial L_{0}}{\partial\theta_{0}}L_{0}^{-1} + \frac{1}{2}L_{0}^{-1}\frac{\partial^{2}L_{0}}{\partial\theta_{0}\partial\theta_{0}}L_{0}^{-1}\right\} \\
&\times(\hat{\theta}_{W} - \theta_{0})_{j}(\hat{\theta}_{W} - \theta_{0})_{k} + O_{p}(n^{-3/2}) \\
&= -\Lambda^{-1} + \Lambda^{-1}\Delta\Delta\Lambda^{-1} - \Lambda^{-1}\Lambda\Delta\Lambda^{-1}
\end{align*}

\begin{align*}
+(\Lambda^{-1} - \Lambda^{-1}\Delta\Delta\Lambda^{-1})\sum_{j=1}^{q}\left[\frac{\partial L_{0}}{\partial\theta_{0}}\right]_{j} + \left[\frac{\partial L_{0}}{\partial\theta_{0}} - E_{g}\left(\frac{\partial L_{0}}{\partial\theta_{0}}\right)\right]_{j}(\Lambda^{-1} - \Lambda^{-1}\Delta\Delta\Lambda^{-1}) \\
\times\left\{-\Lambda^{-1}\frac{\partial I}{\partial\theta_{0}} - n^{-1}\Lambda^{-1}q_{0}^{*} + \Lambda^{(2)}(\Lambda^{-2})\right\}_{j}
\end{align*}

\begin{align*}
+\sum_{j,k=1}^{q}\left\{-\Lambda^{-1}E_{g}\left(\frac{\partial L_{0}}{\partial\theta_{0}}\right)\Lambda^{-1}E_{g}\left(\frac{\partial L_{0}}{\partial\theta_{0}}\right)\Lambda^{-1} + \frac{1}{2}\Lambda^{-1}E_{g}\left(\frac{\partial^{2}L_{0}}{\partial\theta_{0}\partial\theta_{0}}\right)\Lambda^{-1}\right\} \\
\times\left(\Lambda^{-1}\frac{\partial I}{\partial\theta_{0}}\right)_{j}(\Lambda^{-1}\frac{\partial I}{\partial\theta_{0}})_{k} + O_{p}(n^{-3/2})
\end{align*}
\[ = -\Lambda^{-1} + \left[ \Lambda^{-1} M \Lambda^{-1} - \Lambda^{-1} E_g \left( J_0^{(3)} \right) \left\{ \Lambda^{-1} \otimes \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right) \right\} \right] \quad \text{at } O_p(n^{-1/2}) \]

\[ + \left. \left[ -\Lambda^{-1} M \Lambda^{-1} M \Lambda^{-1} + \Lambda^{-1} M \Lambda^{-1} E_g \left( J_0^{(3)} \right) \left\{ \Lambda^{-1} \otimes \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right) \right\} \right] \right|_{(A)} \]

\[ + \Lambda^{-1} E_g \left( J_0^{(3)} \right) \left\{ \left( \Lambda^{-1} M \right) \otimes \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right) \right\} - \Lambda^{-1} \left( J_0^{(3)} - E_g \left( J_0^{(3)} \right) \right) \left\{ \Lambda^{-1} \otimes \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right) \right\} \]

\[ + \Lambda^{-1} E_g \left( J_0^{(3)} \right) \left\{ \Lambda^{-1} \otimes \left( -n^{-1} \Lambda \frac{\partial q^*}{\partial \theta_0} + \Lambda \frac{\partial q_1^*}{\partial \theta_0} \right) \right\} - \Lambda^{-1} \left[ E_g \left( J_0^{(3)} \right) \left\{ \Lambda^{-1} \otimes \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right) \right\} \right] \]

\[ + \frac{1}{2} \Lambda^{-1} E_g \left( J_0^{(4)} \right) \left\{ \Lambda^{-1} \otimes \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right)^{<2>} \right\} \quad \text{at } O_p(n^{-1}) \]

\[ \equiv -\Lambda^{-1} + \left( -\Lambda_M^{-1(A)} \right)_{O_p(n^{-1/2})} + \left( -\Lambda_M^{-1(\Delta A)} \right)_{O_p(n^{-1})} + O_p(n^{-3/2}). \quad (3.6) \]

Similarly, define stochastic \( \Gamma^{(A)}_M \) and \( \Gamma^{(\Delta A)}_M \) as follows:

\[ G \equiv G(\theta) \equiv G(\theta, X^*) \equiv \left( n^{-1} \sum_{j=1}^{q} l_j \theta_j \right)_{O_p(n)}, \quad G_0 \equiv G(\theta_0) \equiv G(\theta_0, X^*), \]

\[ G_0 = \Gamma + \left( M_G \right)_{O_p(n^{-1/2})}, \quad E_g(\theta) = \Gamma, \quad G_{0(1)} = \partial G_0 / \partial \theta_j, \]

\[ G_{0(j,k)} = \partial^2 \theta_j \partial \theta_0, \quad G_{0(1,j)} = \partial^2 \theta_0 \partial \theta_j, \quad (j, k = 1, ..., q), \]

\[ \hat{\Gamma}_w = G(\hat{\theta}_w, X^*) = (G(\theta))_{O_p(n)} + (\hat{\Gamma}_w - G_0)_{O_p(n^{-1/2})} \quad (3.7) \]

\[ = \Gamma + M_G + \sum_{j=1}^{q} \sum_{j=1}^{q} G_{0(j,k)}(\hat{\theta}_w - \theta_0), \frac{1}{2} \sum_{j, k=1}^{q} G_{0(j,k)}(\hat{\theta}_w - \theta_0), \frac{1}{2} \sum_{j, k=1}^{q} G_{0(j,k)}(\hat{\theta}_w - \theta_0)_{O_p(n^{-3/2})} \]

\[ = \Gamma + M_G + \sum_{j=1}^{q} \sum_{j=1}^{q} G_{0(j,k)} \left( -n^{-1} \Lambda \frac{\partial q^*}{\partial \theta_0} + \Lambda \frac{\partial q_1^*}{\partial \theta_0} \right)_{O_p(n^{-3/2})} \]

+ \sum_{j, k=1}^{q} G_{0(j,k)}(\Lambda I_{0(j)})_{O_p(n^{-3/2})} \]

\[ + O_p(n^{-3/2}) \]
\begin{align*}
&= \Gamma + \left\{ \mathbf{M}_G - \sum_{j=1}^q \mathbb{E}_g \left( \mathbf{G}_{0(j)}^{(3)} \right) \left( \Lambda^{-1} \frac{\partial \tilde{I}}{\partial \mathbf{\Theta}_0} \right)_{j} \right\} O_p(n^{-1/2}) \\
&+ \left\{ - \sum_{j=1}^q \left\{ \mathbf{G}_{0(j)}^{(3)} - \mathbb{E}_g \left( \mathbf{G}_{0(j)}^{(3)} \right) \right\} \left( \Lambda^{-1} \frac{\partial \tilde{I}}{\partial \mathbf{\Theta}_0} \right)_{j} + \sum_{j=1}^q \mathbb{E}_g \left( \mathbf{G}_{0(j)}^{(3)} \right) \left( - n^{-1} \mathbf{q}_0^* + \Lambda^{(2)} I_0^{(2)} \right)_{j} \right\} O_p(n^{-3/2}) \\
&+ \frac{1}{2} \sum_{j,k=1}^q \mathbb{E}_g \left( \mathbf{G}_{0(j,k)}^{(4)} \right) \left( \Lambda^{-1} \frac{\partial \tilde{I}}{\partial \mathbf{\Theta}_0} \right)_{j} \left( \Lambda^{-1} \frac{\partial \tilde{I}}{\partial \mathbf{\Theta}_0} \right)_{k} \right\} + O_p(n^{-3/2}) \\
&\equiv \Gamma + \left( \Gamma_M^{(\Lambda)} \right)_{O_p(n^{-1/2})} + \left( \Gamma_M^{(\Delta\Delta)} \right)_{O_p(n^{-1})} + O_p(n^{-3/2}).
\end{align*}

From (3.6) and (3.7), we have

**Lemma 1.** The stochastic correction term in (3.4) of $n^{-1} \text{TIC}_W^{(1)}$ in Definition 1 is expanded as

$$n^{-1} 2 \text{tr}(-\hat{\mathbf{I}}_W^{-1} - \hat{\mathbf{I}}_W^{-1}) = n^{-1} 2 \text{tr}(-\Lambda^{-1} \Gamma) + 2(n^{-1} \text{tr}(\Lambda^{(1)}))_{O_p(n^{-1})} + 2(n^{-1} \text{tr}(\Lambda^{(2)}))_{O_p(n^{-1})} + O_p(n^{-5/2})$$

$$\equiv n^{-1} 2 \text{tr}(-\Lambda^{-1} \Gamma) + 2\left\{ n^{-1} \text{tr}(\Lambda_M^{(1)} \Gamma - \Lambda_M^{(2)} \Gamma) \right\}_{O_p(n^{-1})} + O_p(n^{-5/2}),$$

(3.8)

where the stochastic quantities are given by (3.6) and (3.7).

For (3.5) of $n^{-1} \text{TIC}_W^{(2)}$, define stochastic $-\Lambda_1^{(-1)} - \Lambda_1^{(-1)(\Delta\Delta)}$, $\Gamma_1^{(\Lambda)}$ and $\Gamma_1^{(\Delta\Delta)}$ by omitting terms with $\mathbf{M}$, $\mathbf{J}_0^{(3)} - \mathbb{E}_g \left( \mathbf{J}_0^{(3)} \right)$, $\mathbf{M}_G$ and $\mathbf{G}_{0(j)}^{(3)} - \mathbb{E}_g \left( \mathbf{G}_{0(j)}^{(3)} \right)$ in (3.6) and (3.7) as follows:

$$\hat{\mathbf{I}}_W^{(-\Lambda^{-1})} = -\Lambda^{-1} - \left\{ \Lambda^{-1} \mathbb{E}_g \left( \mathbf{J}_0^{(3)} \right) \left\{ \Lambda^{-1} \otimes \left( \Lambda^{-1} \frac{\partial \tilde{I}}{\partial \mathbf{\Theta}_0} \right) \right\} \right\}_{O_p(n^{-1/2})}$$

$$+ \left\{ \Lambda^{-1} \mathbb{E}_g \left( \mathbf{J}_0^{(3)} \right) \left\{ \Lambda^{-1} \otimes \left( - n^{-1} \mathbf{q}_0^* + \Lambda^{(2)} I_0^{(2)} \right) \right\} - \Lambda^{-1} \left\{ \mathbb{E}_g \left( \mathbf{J}_0^{(3)} \right) \left\{ \Lambda^{-1} \otimes \left( \Lambda^{-1} \frac{\partial \tilde{I}}{\partial \mathbf{\Theta}_0} \right) \right\} \right\} + O_p(n^{-3/2})$$

$$\equiv -\Lambda^{-1} - (\Lambda_1^{(-1)}))_{O_p(n^{-1/2})} + (\Lambda_1^{(-1)(\Delta\Delta)})_{O_p(n^{-1})} + O_p(n^{-3/2}),$$
From (3.9) and (3.10), we have

**Lemma 2.** The stochastic correction term in (3.5) of \( n^{-1} \text{TIC}_w^{(2)} \) in Definition 1 is expanded as

\[
\begin{align*}
\hat{i}_w^{(t)} &= \Gamma - \left\{ \sum_{j=1}^{g} E_g \left( G_{0(j)}^{(1)} \right) \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right)_{j \in O_p(n^{-1/2})} \right\} + \left[ \sum_{j=1}^{n} E_g \left( G_{0(j)}^{(3)} \right) \left( -n^{-1} \Lambda^{-1} \eta_0^* + \Lambda^{(2)} I_0^{(2)} \right)_{j \in O_p(n^{-1/2})} \right. \\
&+ \frac{1}{2} \left. \sum_{j,k=1}^{g} E_g \left( G_{0(j,k)}^{(4)} \right) \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right)_{j} \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right)_{k} \right] + O_p(n^{-3/2}) \\
&= \Gamma + \left( \Gamma_1^{(\Lambda)} \right)_{O_p(n^{-3/2})} + \left( \Gamma_1^{(\Delta\Delta)} \right)_{O_p(n^{-1})} + O_p(n^{-3/2}).
\end{align*}
\]

(3.10)

where the stochastic quantities are given by (3.9) and (3.10).

For the bias correction of \( n^{-1} \text{TIC}_w^{(1)} \) (see (3.4) of Definition 1), we derive the expectations of the two stochastic terms \( 2(\text{tr}_{\Delta}^{(T1)})_{O_p(n^{-1/2})} \) and \( 2(\text{tr}_{\Delta\Delta}^{(T1)})_{O_p(n^{-1})} \), where the former expectation becomes \( \mathbb{E}_g(\text{tr}_{\Lambda}^{(T1)}) = \mathbb{E}_g \{ \text{tr}(-\Lambda^{-1} \Gamma - \Lambda^{-1} \Gamma^{(\Delta)}) \} = 0 \) (see (3.8) of Lemma 1) since \( \mathbb{E}_g(-\Lambda^{-1} \Gamma) = \mathbb{E}_g(\Gamma^{(\Lambda)}) = 0 \) by construction. The latter expectation (see (3.8)) is given as follows:
\[ E_g \{ 2(\text{tr}^{(T)}) \} \]
\[ = 2E_g \{ \text{tr}( -\Lambda^{-1}^{(\alpha)} \Gamma^{(\alpha)} - \Lambda^{-1}^{(\Lambda\alpha)} \Gamma - \Lambda^{-1}^{(\Lambda\Lambda)} ) \} \]
\[ = n^{-1} 2E_g \text{tr} \left[ \sum_{(A)} n \left( \Lambda^{-1} M \Lambda^{-1} - \Lambda^{-1} E_g (J^{(3)}_0) \left\{ \Lambda^{-1} \otimes \left( \Lambda^{-1} \frac{\partial l}{\partial \theta_0} \right) \right\} \right) \right. \]
\[ \times \left\{ M_G - \sum_{j=1}^q \frac{E_g (G^{(3)}_{0(j)})}{\Lambda^{-1}} \left( \Lambda^{-1} \frac{\partial l}{\partial \theta_0} \right) \right\} \]
\[ + \left[ -\Lambda^{-1} M \Lambda^{-1} M \Lambda^{-1} + \Lambda^{-1} M \Lambda^{-1} E_g (J^{(3)}_0) \left\{ \Lambda^{-1} \otimes \left( \Lambda^{-1} \frac{\partial l}{\partial \theta_0} \right) \right\} \right] \]
\[ + \Lambda^{-1} E_g (J^{(3)}_0) \left\{ (\Lambda^{-1} M \Lambda^{-1}) \otimes \left( \Lambda^{-1} \frac{\partial l}{\partial \theta_0} \right) \right\} - \Lambda^{-1} (J^{(3)}_0 - E_g (J^{(3)}_0)) \left\{ \Lambda^{-1} \otimes \left( \Lambda^{-1} \frac{\partial l}{\partial \theta_0} \right) \right\} \]
\[ + \Lambda^{-1} E_g (J^{(3)}_0) \left\{ \Lambda^{-1} \otimes \left( -n^{-1} \Lambda^{-1} q_0 + \Lambda^{(2)} l^{(2)} \right) \right\} - \Lambda^{-1} \left[ E_g (J^{(3)}_0) \left\{ \Lambda^{-1} \otimes \left( \Lambda^{-1} \frac{\partial l}{\partial \theta_0} \right) \right\} \right]^2 \]
\[ + \frac{1}{2} \Lambda^{-1} E_g (J^{(4)}_0) \left\{ \Lambda^{-1} \otimes \left( \Lambda^{-1} \frac{\partial l}{\partial \theta_0} \right) \right\} \Gamma \]
\[ - n \Lambda^{-1} \left[ \sum_{j=1}^q \left\{ \frac{E_g (G^{(3)}_{0(j)})}{\Lambda^{-1}} \left( \Lambda^{-1} \frac{\partial l}{\partial \theta_0} \right) \right\} + \sum_{j=1}^q \frac{E_g (G^{(3)}_{0(j)})}{\Lambda^{-1}} (-n^{-1} \Lambda^{-1} q_0 + \Lambda^{(2)} l^{(2)}_j) \right] \]
\[ + \frac{1}{2} \sum_{j,k=1}^q \frac{E_g (G^{(4)}_{0(j,k)})}{\Lambda^{-1}} \left( \Lambda^{-1} \frac{\partial l}{\partial \theta_0} \right)_j \left( \Lambda^{-1} \frac{\partial l}{\partial \theta_0} \right)_k \right] \]
\[\begin{align*}
&+ \left[ -\text{vec}'(\Lambda^{-1}) n E_g(M^{<2>}) \text{vec}(\Lambda^{-1} \Gamma \Lambda^{-1}) \\
&+ 2 \sum_{a,b,c=1}^q (\Lambda^{-1})_{a} E_g(J_0^{(3)}) \{ (\Lambda^{-1} \Gamma \Lambda^{-1})_{ab} \otimes (\Lambda^{-1})_{bc} \} n E_g \left( m_{ab} \frac{\partial \tilde{T}}{\partial \theta_{0c}} \right) \right] \\
&- \sum_{a=1}^q \text{tr} \left[ n E_g \left\{ \{ J_0^{(3)} - E_g(J_0^{(3)}) \} \frac{\partial \tilde{T}}{\partial \theta_{0a}} \right\} \{ (\Lambda^{-1} \Gamma \Lambda^{-1}) \otimes (\Lambda^{-1})_{,a} \} \right] \\
&+ \text{tr}[E_g(J_0^{(3)}) \{(\Lambda^{-1} \Gamma \Lambda^{-1}) \otimes \alpha_{W1} \}] \\
&- \sum_{a,b=1}^q \text{tr}[\Gamma \Lambda^{-1} E_g(J_0^{(3)}) \{(\Lambda^{-1} \otimes (\Lambda^{-1})_{,a}) \} E_g(J_0^{(3)}) \{(\Lambda^{-1} \otimes (\Lambda^{-1})_{,b}) \}] \gamma_{ab} \\
&+ \frac{1}{2} \text{vec}' E_g(J_0^{(4)}) \{ \text{vec}(\Lambda^{-1} \Gamma \Lambda^{-1}) \}^{<2>} \right] \\
&- \left[ - \sum_{a,b=1}^q \text{tr} n E_g \left[ \Lambda^{-1} \{ G_{0(a)}^{(3)} - E_g(G_{0(a)}^{(3)}) \} \frac{\partial \tilde{T}}{\partial \theta_{0b}} \right] \right] \eta_{ab} \\
&+ \sum_{j=1}^q \text{tr} \{ \Lambda^{-1} E_g(G_{0(j)}^{(3)})(\alpha_{W1} \}_j \} + \frac{1}{2} \sum_{j,k=1}^q \text{tr} \{ \Lambda^{-1} E_g(G_{0(j,k)}^{(4)})(\Lambda^{-1} \Gamma \Lambda^{-1})_{jk} \} \right] \\
&\equiv n^{-1} d^{(T1)},
\end{align*}\]

where \( n^{-1} \alpha_{W1} \equiv -n^{-1} \Lambda^{-1} \alpha_{0}^* + E_g(\Lambda^{(2)} \Gamma_0^{(2)}) \) is the vector of the asymptotic biases of \( \hat{\theta}_W \) up to order \( O(n^{-1}) \) under possible model misspecification.

For \( n^{-1} \text{TIC}_W^{(2)} \) (see (3.5) of Definition 1), similarly we have

\[2 E_g(\text{tr}^{(T2)}) = 2 E_g \{ \text{tr}(\Lambda^{-1}(\Lambda \Gamma - \Lambda^{-1} \Gamma)) \} = 0 \] by construction and
\[ E_g \{2(\text{tr}^{(T2)})\} = 2E_g \{\text{tr}(-\Lambda_{11}^{-1}(\Lambda_1) \Gamma_1^{-1} - \Lambda_{12}^{-1} \Delta_1 - \Lambda_1^{-1} \Delta_1^{-1})\} \]

\[= n^{-1} 2E_g \left[ \sum_{a,b,c,d=1}^g (\Lambda_{a,b}^{-1})_{a,b} E_g(J_{0}^{(3)}) \{(\Lambda_{a,b}^{-1})_{a,b} \otimes (\Lambda_{a,b}^{-1})_{a,b}\} \sum_{j=1}^g E_g(G_{0(j)}^{(3)}) \left\{ \Lambda_{a,b}^{-1} \frac{\partial I}{\partial \theta_0} \right\} \right] \]

\[\quad + n \left[ \Lambda_{a,b}^{-1} E_g(J_{0}^{(3)}) \left\{ \Lambda_{a,b}^{-1} \otimes \left( \Lambda_{a,b}^{-1} \frac{\partial I}{\partial \theta_0} \right) \right\} \right] \]

\[\quad + \frac{1}{2} \Lambda_{a,b}^{-1} E_g(J_{0}^{(4)}) \left\{ \Lambda_{a,b}^{-1} \otimes \left( \Lambda_{a,b}^{-1} \frac{\partial I}{\partial \theta_0} \right) \right\} \left( \begin{array}{c} \Gamma \\ \end{array} \right) - n^{-1} 2E_g \{\text{tr}(\Lambda_{a,b}^{-1} \Delta_1^{-1})\} \]

\[\equiv n^{-1} d^{(T2)}. \]

The higher-order bias corrections of \( n^{-1} \text{TIC}_w^{(j)} (j = 1, 2) \) are given as follows:

**Theorem 2.** Under possible model misspecification and some regularity conditions, define

\[ n^{-1} \text{TIC}_w^{(j)} \equiv n^{-1} \text{TIC}_w^{(j)} - n^{-2} (\hat{b}_2 + \hat{d}^{(Tj)}) \quad (j = 1, 2), \]

where \( \hat{b}_2 \) and \( \hat{d}^{(Tj)} \) are consistent estimators of \( b_2 = c_1 + c_2 + c_3 \) and \( d^{(Tj)} \). Then,
\[ E_g(n^{-1}\text{TIC}_{w}^{(j)} + 2\hat{T}_w^*) = O(n^{-3}). \]

4. Asymptotic cumulants

In Section 2, the bias of \(-2\hat{T}_w\) was defined as \(-2E_g(\hat{T}_w - \hat{T}_w^*)\) (see (2.8)) with the definitions of \(\hat{T}_w\) and \(\hat{T}_w^*\) by (2.4) and (2.7), respectively. In this section, the asymptotic cumulants of \(-2\hat{T}_w = -2\bar{T}(\hat{\theta}_w, X^*) = -2\bar{T}\{\theta_w(X^*), X^*\}\) using the density \(g(X^*|Z_0)\) are given, where the bias is defined as \(-2\{E_g(\hat{T}_w) - \bar{T}_0^*\}\) with \(\bar{T}_0^*\) being the population counterpart of \(\hat{T}_w\), which is the limiting value of \(\hat{T}_w\) when \(n\) is infinitely large. The value and the notation of \(\bar{T}_0^*\) are equal to those of (2.5) since

\[
\bar{T}_0^* = E_g\{\bar{T}(\theta_0|X^*)\} = \int_{R(X)} \bar{T}(\theta_0|X)g(X|Z_0)dX \\
= \int_{R(Z)} \bar{T}(\theta_0|Z)g(Z|Z_0)dZ = E_g\{\bar{T}(\theta_0|Z^*)\}.
\]

The asymptotic cumulants of \(n^{-1}\text{AIC}_w\) and \(n^{-1}\text{TIC}_{w}^{(j)} (j = 1, 2)\) are given before and after studentization up to the fourth order with the higher-order asymptotic variances. The studentization is for testing and interval estimation, where the population values of \(-2\hat{T}_w\) are defined in two ways as \(-2E_g(\hat{T}_w)\) and \(-2\bar{T}_0^*\). While these two values are of order \(O(1)\), the former depends on \(n\) in that the value is generally written as \(O(1) + O(n^{-1}) + O(n^{-2}) + \cdots\). When \(n\) is infinitely large, \(-2E_g(\hat{T}_w)\) becomes equal to \(-2\bar{T}_0^*\). So, \(-2\bar{T}_0^*\) is also of interest as well as \(-2E_g(\hat{T}_w)\). Note that asymptotically unbiased point estimators of the latter up to order \(O(n^{-1})\) are \(n^{-1}\text{AIC}_w\) under correct model specification and \(n^{-1}\text{TIC}_{w}^{(j)} (j = 1, 2)\) under possible model misspecification.

Under possible model misspecification, assume that the following hold with the definitions of the asymptotic cumulants whose factors of \(O(1)\) are \(\alpha^{(t)}_{w_k}\) for \(n^{-1}\text{AIC}_w\) and \(\alpha^{(t,j)}_{w_k}\) for \(n^{-1}\text{TIC}_{w}^{(j)} (j = 1, 2) (k = 1, \Delta 1, 2, \Delta 2, 3, 4)\):

24
\[ \kappa_{g1}(n^{-1}\text{AIC}_W) = -2(\overline{T}_0^{*})_{O(1)} + n^{-1}\alpha^{(A)}_{W1} + n^{-2}\alpha^{(A)}_{W1}\alpha + O(n^{-3}), \]
\[ \kappa_{g2}(n^{-1}\text{AIC}_W) = n^{-1}\alpha^{(A)}_{W2} + n^{-2}\alpha^{(A)}_{W1}\alpha + O(n^{-3}), \]
\[ \kappa_{g3}(n^{-1}\text{AIC}_W) = n^{-2}\alpha^{(A)}_{W3} + O(n^{-3}), \]
\[ \kappa_{g4}(n^{-1}\text{AIC}_W) = n^{-3}\alpha^{(A)}_{W4} + O(n^{-4}), \]
\[ \kappa_{g1}(n^{-1}\text{TIC}^{(j)}_W) = -2(\overline{T}_0^{*})_{O(1)} + n^{-1}\alpha^{(T,j)}_{W1} + n^{-2}\alpha^{(T,j)}_{W1}\alpha + O(n^{-3}), \] (4.2)
\[ \kappa_{g2}(n^{-1}\text{TIC}^{(j)}_W) = n^{-1}\alpha^{(T,j)}_{W2} + n^{-2}\alpha^{(T,j)}_{W1}\alpha + O(n^{-3}), \]
\[ \kappa_{g3}(n^{-1}\text{TIC}^{(j)}_W) = n^{-2}\alpha^{(T,j)}_{W3} + O(n^{-3}), \]
\[ \kappa_{g4}(n^{-1}\text{TIC}^{(j)}_W) = n^{-3}\alpha^{(T,j)}_{W4} + O(n^{-4}) \ (j = 1, 2). \]

From the asymptotic properties of \( n^{-1}\text{AIC}_W \) and \( n^{-1}\text{TIC}^{(j)}_W (j = 1, 2) \) given earlier we have, \( \kappa_{g1}(n^{-1}\text{AIC}_W + 2\hat{\ell}_W^{*}) = O(n^{-1}) \) under model misspecification and
\( \kappa_{g1}(n^{-1}\text{AIC}_W + 2\hat{\ell}_W^{*}) = O(n^{-2}) \) under correct model specification \( \alpha^{(A)*}_{W1} = 0 \) while
\( \kappa_{g1}(n^{-1}\text{TIC}^{(j)}_W + 2\hat{\ell}_W^{*}) = O(n^{-2}) \) with \( \alpha^{(T,j)*}_{W1} = 0 \ (j = 1, 2) \) under possible model misspecification. Other asymptotic cumulants for \( n^{-1}\text{AIC}_W + 2\hat{\ell}_W^{*} \) and \( n^{-1}\text{TIC}^{(j)}_W + 2\hat{\ell}_W^{*} \) using the notations \( \alpha^{(A)*}_{Wk} \) and \( \alpha^{(T,j)*}_{Wk} (k = \Delta 1, 2, \Delta 2, 3, 4) \), respectively, are defined similarly to (4.2).

Recall that \( n^{-1}\text{AIC}_W = -2\hat{\ell}_W + n^{-1}2q \) (see (3.1)) with the corresponding symbolic expressions of the asymptotic expansions of \( n^{-1}\text{TIC}^{(j)}_W (j = 1, 2) \) given by (3.4) and (3.5).

Then, for the asymptotic cumulants of (4.2), we expand the main term \(-2\hat{\ell}_W\) common to \( n^{-1}\text{AIC}_W \) and \( n^{-1}\text{TIC}^{(j)}_W (j = 1, 2) \):
\[ -2\hat{\ell}_W = -2(\overline{T}_0)_{O(1)} - 2\sum_{j=1}^{4} \frac{1}{j!} \left( \frac{\partial^{T}}{\partial \theta_0} \right)^{<j>} (\hat{\theta}_0 - \theta_0)^{<j>} + O_p(n^{-5/2}) \]
\[ = -2(\overline{T}_0)_{O(1)} - 2\sum_{j=1}^{4} \frac{1}{j!} \left( \frac{\partial^{T}}{\partial \theta_0} \right)^{<j>} \left\{ -n^{-1} \Lambda q_0^* + \sum_{k=1}^{3} \Lambda^{(k)} I_0^{(k)} + n^{-1}(I_0^{(W)})_{O_p(n^{-1/2})} \right\}^{<j>} \] (4.3)
\[ + O_p(n^{-5/2}) \]
\[
-2(\overline{t}_0)_{O_p(1)} - 2 \left( \frac{\partial \overline{t}}{\partial \theta_0} \right)_{O_p(n^{-1/2})} \left\{ -n^{-1} \Lambda^{-1} \mathbf{q}_0^* + \sum_{k=1}^{3} \Lambda^{(k)} \mathbf{l}_0^{(k)} + n^{-1} \mathbf{l}_0^{(W)} \right\}_{O_p(n^{-1/2})} \\
- \left\{ \frac{\partial^2 \overline{t}}{(\partial \theta_0 \gamma)^{<2>}} \right\}_{O_p(1)} \left\{ -n^{-1} \Lambda^{-1} \mathbf{q}_0^* + \sum_{k=1}^{3} \Lambda^{(k)} \mathbf{l}_0^{(k)} + n^{-1} \mathbf{l}_0^{(W)} \right\}_{O_p(n^{-1/2})}^{<2>}
\]

\[
- \frac{1}{3} \left\{ \frac{\partial^3 \overline{t}}{(\partial \theta_0^n)^{<3>}} \right\}_{O_p(1)} \left\{ -n^{-1} \Lambda^{-1} \mathbf{q}_0^* + \sum_{k=1}^{2} \Lambda^{(k)} \mathbf{l}_0^{(k)} \right\}_{O_p(n^{-1/2})}^{<3>}
\]

\[
- \frac{1}{12} E_g \left\{ \frac{\partial^4 \overline{t}}{(\partial \theta_0^n)^{<4>}} \right\} \left( \Lambda^{(1)} \mathbf{l}_0^{(1)} \right)_{O_p(n^{-5/2})}
\]

The five terms up to order \( O_p(n^{-1}) \) in the last expression of (4.3) are further expanded one by one as follows:

(i) 
\[
-2(\overline{t}_0)_{O_p(1)} = -2E_g (\overline{t}_0) - 2\overline{t}_0 - E_g (\overline{t}_0) \\
\equiv -2(\overline{t}_0)_{O(1)} - 2(\overline{t}_0 - \overline{t}_0)_{O_p(n^{-1/2})},
\]

(ii) 
\[
-2 \left( \frac{\partial \overline{t}}{\partial \theta_0} \right)_{O_p(n^{-3/2})} \left\{ -n^{-1} \Lambda^{-1} \mathbf{q}_0^* + \sum_{k=1}^{3} \Lambda^{(k)} \mathbf{l}_0^{(k)} + n^{-1} \mathbf{l}_0^{(W)} \right\}_{O_p(n^{-3/2})} \\
= 2 \left( n^{-1} \left( \frac{\partial \overline{t}}{\partial \theta_0} , \Lambda^{-1} \mathbf{q}_0^* \right) \right)_{O_p(n^{-3/2})} + 2 \left( \frac{\partial \overline{t}}{\partial \theta_0} , \Lambda^{-1} \frac{\partial \overline{t}}{\partial \theta_0} \right)_{O_p(n^{-1})} - 2 \left( \frac{\partial \overline{t}}{\partial \theta_0} , \Lambda^{(2)} \mathbf{l}_0^{(2)} \right)_{O_p(n^{-3/2})} \\
= \left( n^{-1} \left( \frac{\partial \overline{t}}{\partial \theta_0} , \Lambda^{(3)} \mathbf{l}_0^{(3)} \right) \right)_{O_p(n^{-3/2})} - 2 \left( n^{-1} \frac{\partial \overline{t}}{\partial \theta_0} , \mathbf{l}_0^{(W)} \right)_{O_p(n^{-3/2})}
\]

(iii) 
\[
- \left\{ \frac{\partial^2 \overline{t}}{(\partial \theta_0 \gamma)^{<2>}} \right\}_{O_p(1)} \left\{ -n^{-1} \Lambda^{-1} \mathbf{q}_0^* + \sum_{k=1}^{3} \Lambda^{(k)} \mathbf{l}_0^{(k)} + n^{-1} \mathbf{l}_0^{(W)} \right\}_{O_p(n^{-1/2})}^{<2>}
\]

\[
= -\text{vec}'(\Lambda + (M))_{O_p(n^{-1/2})} \left\{ -n^{-1} \Lambda^{-1} \mathbf{q}_0^* + \sum_{k=1}^{3} \Lambda^{(k)} \mathbf{l}_0^{(k)} + n^{-1} \mathbf{l}_0^{(W)} \right\}_{O_p(n^{-1/2})}^{<2>}
\]

\[
= \{-n^{-2} \text{vec}'(\Lambda)(\Lambda^{-1} \mathbf{q}_0^*)_{O(n^{-2})}\right\}_{O_p(n^{-3/2})} - 2\left[n^{-2} \text{vec}'(\Lambda) \left\{ (\Lambda^{-1} \mathbf{q}_0^*) \otimes \left( \Lambda^{-1} \frac{\partial \overline{t}}{\partial \theta_0} \right) \right\}_{O_p(n^{-3/2})}
\]

26
\[+2[n^{-1}\text{vec}'(A)\{(\Lambda^{-1}q_0^* \otimes (\Lambda^{(2)}I_0^{(2)})\}]_{O_p(n^{-2})} - \left\{\text{vec}'(A)\left(\Lambda^{-1}\frac{\partial I}{\partial \theta_0}\right)^{<2\gamma}>\right\}_{O_p(n^{-1})}\]

\[+2\left[\text{vec}'(A)\left(\Lambda^{-1}\frac{\partial I}{\partial \theta_0} \otimes (\Lambda^{(2)}I_0^{(2)})\right)\right]_{O_p(n^{-3/2})}\]

\[+2\left[\text{vec}'(A)\left(\Lambda^{-1}\frac{\partial I}{\partial \theta_0} \otimes (\Lambda^{(3)}I_0^{(3)})\right)\right]_{O_p(n^{-2})}\]

\[-\text{vec}'(A)(\Lambda^{(2)}I_0^{(2)})^{<2\gamma}>_{O_p(n^{-2})} + 2\left[n^{-1}\text{vec}'(A)\left(\Lambda^{-1}\frac{\partial I}{\partial \theta_0} \otimes I_0^{(W)}\right)\right]_{O_p(n^{-2})}\]

\[-2\left[n^{-1}\text{vec}'(M)\left(\Lambda^{-1}q_0^* \otimes \Lambda^{-1}\frac{\partial I}{\partial \theta_0}\right)\right]_{O_p(n^{-2})}\]

\[-\text{vec}'(M)\left(\Lambda^{-1}\frac{\partial l}{\partial \theta_0}\right)^{<2\gamma>}_{O_p(n^{-3/2})} + 2\left[\text{vec}'(M)\left(\Lambda^{-1}\frac{\partial l}{\partial \theta_0} \otimes (\Lambda^{(2)}I_0^{(2)})\right)\right]_{O_p(n^{-2})},\]

(iv)

\[-\frac{1}{3}\left(\partial \theta_0\right)^{<3\gamma>}\left(-n^{-1}\Lambda^{-1}q_0^* + \sum_{k=1}^{2} \Lambda^{(k)}I_0^{(k)}\right)^{<3\gamma>}\]

\[= -\frac{1}{3}\text{vec}'\left\{E_g(J_0^{(3)}) + \{J_0^{(3)} - E_g(J_0^{(3)})\}\right\}_{O_p(n^{-3/2})} \left(-n^{-1}\Lambda^{-1}q_0^* + \sum_{k=1}^{2} \Lambda^{(k)}I_0^{(k)}\right)^{<3\gamma>}\]

\[= \left[n^{-1}\text{vec}'\left\{E_g(J_0^{(3)})\right\}\left(\Lambda^{-1}q_0^* \otimes \Lambda^{-1}\frac{\partial l}{\partial \theta_0}\right)^{<2\gamma>}\right]_{O_p(n^{-2})}\]

\[+\frac{1}{3}\left[\text{vec}'\left\{E_g(J_0^{(3)})\right\}\left(\Lambda^{-1}\frac{\partial l}{\partial \theta_0}\right)^{<3\gamma>}\right]_{O_p(n^{-3/2})}\]

\[-\text{vec}'\left\{E_g(J_0^{(3)})\right\}\left(\Lambda^{-1}\frac{\partial l}{\partial \theta_0} \otimes (\Lambda^{(2)}I_0^{(2)})\right)\right]_{O_p(n^{-2})}\]

\[+\frac{1}{3}\text{vec}'\left\{J_0^{(3)} - E_g(J_0^{(3)})\right\}\left(\Lambda^{-1}\frac{\partial l}{\partial \theta_0}\right)^{<3\gamma>}_{O_p(n^{-2})},\]
\(- \frac{1}{12} E_g \left\{ \frac{\delta^4 I}{(\partial \theta_0)^4} \right\} \langle \Lambda^{(1)} I_0^{(1)} \rangle^{<4>} = - \frac{1}{12} \left[ \text{vec}' \{ E_g (J_0^{(4)}) \} \left( \Lambda^{-1} \frac{\delta I}{\partial \theta_0} \right) \right]_{\epsilon, (n^2)} \). \\

Using \( \frac{\partial I}{\partial \theta_0} \Lambda^{-1} \frac{\partial I}{\partial \theta_0} = \text{vec}'(\Lambda) \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right)^{<2>} \) and similar results in (4.4), (4.3) becomes

\[-2 \hat{I}_W = -2(\bar{T}_a - \bar{T}_0)_{\epsilon, (1)} - 2(\bar{T}_a - \bar{T}_0)_{\epsilon, (n^{-1/2})} + \left( \frac{\partial I}{\partial \theta_0} \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right)_{\epsilon, (n^{-1})} + \left( 2n^{-1} \frac{\partial I}{\partial \theta_0} \Lambda^{-1} q_0^* - 2 \frac{\partial I}{\partial \theta_0} \Lambda^{(2)} \right)_{\epsilon, (n^{-1})} - 2n^{-1} \text{vec}'(\Lambda) \left( \Lambda^{-1} q_0^* \right) \times \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right) \]

\[+ 2 \text{vec}'(\Lambda) \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right) \times \left( \Lambda^{(2)} \right) - \text{vec}'(M) \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right)^{<2>} \]

\[+ \frac{1}{3} \text{vec}' \{ E_g (J_0^{(3)}) \} \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right)^{<3>} \right\} \]

\[-\{ \text{vec}'(\Lambda) \left( \Lambda^{-1} q_0^* \right)^{<2>} \}_{\epsilon, (n^{-2})} \]

\[+ \left[ -2 \frac{\partial I}{\partial \theta_0} \Lambda^{(3)} I_0^{(3)} - 2n^{-1} \frac{\partial I}{\partial \theta_0} I_0^{(W)} + 2n^{-1} \text{vec}'(\Lambda) \left( \Lambda^{-1} q_0^* \right) \times \left( \Lambda^{(2)} I_0^{(2)} \right) \]

\[+ 2 \text{vec}'(\Lambda) \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right) \times \left( \Lambda^{(3)} I_0^{(3)} \right) - \text{vec}'(\Lambda) \left( \Lambda^{(2)} I_0^{(2)} \right)^{<2>} \]

\[+ 2n^{-1} \text{vec}'(\Lambda) \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right) \times \left( \Lambda^{(3)} I_0^{(3)} \right) - 2n^{-1} \text{vec}'(M) \left( \Lambda^{-1} q_0^* \right) \times \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right) \]

\[+ 2 \text{vec}'(\Lambda) \left( \Lambda^{-1} \frac{\partial I}{\partial \theta_0} \right) \times \left( \Lambda^{(3)} I_0^{(3)} \right) - \text{vec}'(\Lambda) \left( \Lambda^{(2)} I_0^{(2)} \right)^{<2>} \]
The last parenthetical results \( \overline{T}_{W}^{(j)} = \overline{T}_{M}^{(j)} \) (\( j = 1, \ldots, 4 \)) indicate that \( -2 \hat{T}_{W}^{(j)} \) is equal to \( -2 \hat{T}_{M}^{(j)} \) up to order \( O_{p}(n^{-3/2}) \). The remaining two terms of order \( O(n^{-2}) \) and \( O_{p}(n^{-2}) \) are relevant only to \( \alpha_{W1}^{(A)} \) and \( \alpha_{W1}^{(T)} \) (\( j = 1, 2 \)) in (4.2). In the last result of (4.5), the first term for \( \overline{T}_{W}^{(4)} \) can also be written as
vec'(Λ)(Λ^{(2)}I_0^{(2)})^{<2>}

= vec'(M) \left\{ \left( \Lambda^{-1} \frac{\partial \bar{l}}{\partial \theta_0} \right) \otimes \left( \Lambda^{-1} M \Lambda^{-1} \frac{\partial \bar{l}}{\partial \theta_0} \right) \right\}

- vec'[E_g(J_0^{(3)})] \left\{ \left( \Lambda^{-1} \frac{\partial \bar{l}}{\partial \theta_0} \right)^{<2>} \otimes \left( \Lambda^{-1} M \Lambda^{-1} \frac{\partial \bar{l}}{\partial \theta_0} \right) \right\}

+ \frac{1}{4} vec'[E_g(J_0^{(3)})] \left[ \left( \Lambda^{-1} \frac{\partial \bar{l}}{\partial \theta_0} \right)^{<2>} \otimes \left( \Lambda^{-1} E_g(J_0^{(3)}) \right) \left( \Lambda^{-1} \frac{\partial \bar{l}}{\partial \theta_0} \right)^{<2>} \right]

(recall (2.10)).

Noting that $n^{-1} \text{AIC}_W = -2\hat{l}_W + n^{-1}2q$, (4.5) gives

**Theorem 4.** Under possible model misspecification and regularity conditions for (4.2), the asymptotic cumulants of $n^{-1} \text{AIC}_W$ up to the fourth order with the higher-order asymptotic bias and variance are given as follows:

$$
\kappa_{g_1}(n^{-1} \text{AIC}_W) = -2(\hat{l}_0^{*})_{O(1)} + n^{-1}\{nE_g(\bar{t}_\text{ML}^{(2)}) + 2q\}_{O(1)} + n^{-2}\{n^2E_g(\bar{t}_\text{ML}^{(3)} + \bar{t}_\text{ML}^{(4)}) - q_0^{*} \Lambda^{-1} q_0^{*}\}_{O(1)}
$$

$$+ O(n^{-3}) = -2\bar{t}^{*}_0 + n^{-1}\{\text{tr}(\Lambda^{-1} \Gamma) + 2q\} + n^{-2}\{n^2E_g(\bar{t}_\text{ML}^{(3)} + \bar{t}_\text{ML}^{(4)}) - q_0^{*} \Lambda^{-1} q_0^{*}\} + O(n^{-3})
$$

$$= -2\bar{t}^{*}_0 + n^{-1}\alpha_{\text{ML1}}^{(A)} + n^{-2}\alpha_{\text{WAT1}}^{(A)} + O(n^{-3})
$$

$$= (\alpha_{W1}^{(A)} = \alpha_{\text{ML1}}^{(A)} = \text{tr}(\Lambda^{-1} \Gamma) + 2q),
$$

$$\kappa_{g_2}(n^{-1} \text{AIC}_W) = n^{-1}\{nE_g(\bar{t}_\text{ML}^{(1)})\}_{O(1)} + n^{-2}\{2n^2E_g(\bar{t}_\text{ML}^{(1)}\bar{t}_\text{ML}^{(2)}) + 2n^2E_g(\bar{t}_\text{ML}^{(1)}\bar{t}_\text{ML}^{(3)})
$$

$$+ n^2E_g(\bar{t}_\text{ML}^{(2)})^2 -(\alpha_{\text{ML1}}^{(A)} - 2q)^2\} + O(n^{-3})
$$

$$\equiv n^{-1}\alpha_{\text{ML2}}^{(A)} + n^{-2}\alpha_{\text{MLA2}}^{(A)} + O(n^{-3})
$$

$$= (nE_g(\bar{t}_\text{ML}^{(2)}) = \alpha_{\text{ML1}}^{(A)} - 2q = \text{tr}(\Lambda^{-1} \Gamma))
$$

$$= \alpha_{\text{ML2}}^{(A)} = nE_g(\bar{t}_\text{ML}^{(2)})^2 = 4E_g[(l_j - \bar{t}_0^{*})^2] = 4\text{var}_g(l_j), \alpha_{\text{WAT2}}^{(A)} = \alpha_{\text{MLA2}}^{(A)},
$$

$$\kappa_{g_3}(n^{-1} \text{AIC}_W) = n^{-2}\{n^2E_g(\bar{t}_\text{ML}^{(1)})^3 + 3n^2E_g(\bar{t}_\text{ML}^{(1)}\bar{t}_\text{ML}^{(2)}) - 3nE_g(\bar{t}_\text{ML}^{(2)})\alpha_{\text{ML2}}^{(A)}\} + O(n^{-3})
$$

$$\equiv n^{-2}\alpha_{\text{ML3}}^{(A)} + O(n^{-3}) \quad (\alpha_{W3}^{(A)} = \alpha_{\text{ML3}}^{(A)}),
$$
In the case of the canonical parameters under correct model specification as in Corollary 1, using (4.5) (see also (2.20)), the asymptotic biases become as follows:

\[
\kappa_{*} = E_g \{ n^{-1} AIC_w - E_g (n^{-1} AIC_w) \}^4 \cdot 3 \{ n^{-1} \alpha_{ML_2} + n^{-2} \alpha_{ML_2}^2 \} + O(n^{-4})
\]

\[
= E_g \{ n^{-1} AIC_w + 2 \overline{T_0^*} \}^4 \cdot n^{-3} [ -4 (\alpha_{ML_1} - 2q) \{ \alpha_{ML_3} + 3 (\alpha_{ML_1} - 2q) \alpha_{ML_2} \} + 6 (\alpha_{ML_1} - 2q)^2 \alpha_{ML_2} ] - 3n^2 (\alpha_{ML_2}^2) - 6n^{-3} \alpha_{ML_2} \alpha_{ML_2}^2 + O(n^{-4})
\]

\[
= E_g \{ n^{-1} AIC_w + 2 \overline{T_0^*} \}^4 \cdot 3n^{-2} (\alpha_{ML_2})^2
\]

\[
- n^{-3} (4 (\alpha_{ML_1} - 2q) \alpha_{ML_3} + 6 (\alpha_{ML_2} \alpha_{ML_2} + 6 \alpha_{ML_2} \alpha_{ML_1} - 2q)^2 \} + O(n^{-4})
\]

\[
= n^{-3} \left[ \left\{ \kappa_{*} (\overline{T_0^{(1)}}) \right\}_{(n^{-1})} + 4n^2 \left\{ \overline{T_0^{(1)}} \right\}_{(n^{-1})} + 6n^2 \left\{ \overline{T_0^{(1)}} \right\}_{(n^{-1})} \right] + O(n^{-4})
\]

\[
+ 4n^2 \left\{ \overline{T_0^{(1)}} \right\}_{(n^{-1})} \left\{ \overline{T_0^{(1)}} \right\}_{(n^{-1})} - 4 (\alpha_{ML_1} - 2q) \alpha_{ML_3} - 4 (\alpha_{ML_2} \alpha_{ML_2} - 6 \alpha_{ML_2} (\alpha_{ML_1} - 2q)^2 \} + O(n^{-4})
\]

\[
\equiv n^{-3} \alpha_{ML_4} + O(n^{-4}) \left( \alpha_{*} = \alpha_{ML_2} \right)
\]

The results for \( n^{-1} TIC_w(j = 1, 2) \) corresponding to Theorem 4 are given from (3.4) and (3.5) of Definition 1.

**Theorem 5.** Under possible model misspecification and regularity conditions for (4.2), the asymptotic cumulants of \( n^{-1} TIC_w(j = 1, 2) \) up to the fourth order with the higher-order asymptotic bias and variance are given as follows:

\[
\kappa_{*} (n^{-1} TIC_w)
\]

\[
=-2 \overline{T_0^*} + n^{-1} \{ \text{tr}(\Lambda^{-1} \Gamma) + 2q \} + n^{-2} \{ n^2 E_g (\overline{T_0^{(1)}} + \overline{T_0^{(1)}}) - q^* \Lambda^{-1} q^* \} + O(n^{-3})
\]

\[
= 2 \overline{T_0^*} + n^{-1} \alpha_{ML_1} + n^{-2} \alpha_{ML_1}^2 + O(n^{-3})
\]

\[
= 2 \overline{T_0^*} + n^{-1} \alpha_{ML_1} + n^{-2} \alpha_{ML_1}^2 + O(n^{-3})
\]

\[
\equiv 2 \overline{T_0^*} + n^{-1} \alpha_{ML_1} + n^{-2} \alpha_{ML_1}^2 + O(n^{-3})
\]

\[
(\alpha_{ML_1} = \alpha_{ML_1}^2 - 2q + 2 \text{tr}(\Lambda^{-1} \Gamma) = \text{tr}(\Lambda^{-1} \Gamma))
\]

31
\( \kappa_{g2}(n^{-1}TIC_{W}^{(j)}) = n^{-1}\alpha_{ML2} + n^{-2}\{\alpha_{MLA2} + 4nE_g(\tilde{I}_{ML}^{(1)}tr_{\lambda}^{(T,j)})\} + O(n^{-3}) \)
\[ = n^{-1}\alpha_{ML2} + n^{-2}\alpha_{WA2} + O(n^{-3}) \]
\( \alpha_{W2}^{(T)} = \alpha_{ML2} = \alpha_{A}^{(A)} = \alpha_{ML2}^{(A)} \),
\( \kappa_{g3}(n^{-1}TIC_{W}^{(j)}) = \kappa_{g3}(n^{-1}AIC_{ML}) + O(n^{-3}) \)
\[ = n^{-2}\alpha_{ML3} + O(n^{-3}) \]
\( \alpha_{W3}^{(T)} = \alpha_{ML3} = \alpha_{A}^{(A)} = \alpha_{ML3}^{(A)} \),
\( \kappa_{g4}(n^{-1}TIC_{W}^{(j)}) = \kappa_{g4}(n^{-1}AIC_{ML}) + O(n^{-3}) \)
\[ = n^{-3}\alpha_{ML4} + O(n^{-4}) \]
\( \alpha_{W4}^{(T)} = \alpha_{ML4} = \alpha_{A}^{(A)} = \alpha_{ML4}^{(A)} \) \( (j = 1, 2) \),

where the superscript \( (T-) \) indicates a result common to \( n^{-1}TIC_{W}^{(j)} (j = 1, 2) \).

In (4.8), \( \alpha_{W3}^{(T)} = \alpha_{ML3}^{(T)} = \alpha_{W3}^{(A)} = \alpha_{ML3}^{(A)} \) stems from the property that the third asymptotic cumulants of \( n^{-1}TIC_{W}^{(j)} (j = 1, 2) \) are given only by \( \tilde{I}_{W}^{(1)}(= \tilde{I}_{ML}^{(1)}) \) and \( \tilde{I}_{W}^{(2)}(= \tilde{I}_{ML}^{(2)}) \) of \(-\hat{\Delta}_{W}^{1}\) in (3.4) and (3.5) of Definition 1 (see the last parenthetical result of (4.5)) with the fixed term \( tr(\lambda^{-1}\Gamma) \) in (3.4) and (3.5) being irrelevant to the cumulants except that of the first order. The additional stochastic terms \( 2n^{-1}(tr_{\lambda}^{(T,j)})_{O_{p}(n^{-1/2})} \) for \( n^{-1}TIC_{W}^{(j)} (j = 1, 2) \) in (3.4) and (3.5) with \( \tilde{I}_{W}^{(j)}(= \tilde{I}_{ML}^{(j)}) \) \( (j = 1, 2, 3) \) in the expansion of \(-2\hat{\Delta}_{W} \) common to \( n^{-1}AIC_{W} \) and \( n^{-1}TIC_{W}^{(j)} (j = 1, 2) \) contribute to the higher-order added asymptotic variance \( n^{-2}\alpha_{WA2}^{(T)} (j = 1, 2) \) in (4.8). However, the contributions by \( 2n^{-1}tr_{\lambda}^{(T,j)} (j = 1, 2) \) are canceled when we derive the (asymptotic) fourth cumulants, giving \( \alpha_{W4}^{(T)} = \alpha_{ML4}^{(T)} = \alpha_{W4}^{(A)} = \alpha_{ML4}^{(A)} \) in (4.8).

For interval estimation of the population quantity \(-2\hat{T}_{0}^{*}\) as well as \(-2E_{g}(\hat{\Delta}_{W})\) by \( n^{-1}AIC_{W} \) and \( n^{-1}TIC_{W}^{(j)} (j = 1, 2) \), the following studentized estimators are defined:
\[ t_{W}^{(A)} = n^{1/2}(n^{-1}AIC_{W} + 2\hat{T}_{0}^{*}) \]
\( (\hat{\Sigma}_{W}^{(A)})^{1/2} \), \( t_{W}^{(T,j)} = n^{1/2}(n^{-1}TIC_{W}^{(j)} + 2\hat{T}_{0}^{*}) \)
\( (\hat{\Sigma}_{W}^{(A)})^{1/2} \) \( (j = 1, 2) \),
\[ t_{W}^{(A)} = n^{1/2}\{n^{-1}AIC_{W} + 2E_{g}(\hat{\Delta}_{W})\} \]
\( (\hat{\Sigma}_{W}^{(A)})^{1/2} \), \( t_{W}^{(T,j)} = n^{1/2}\{n^{-1}TIC_{W}^{(j)} + 2E_{g}(\hat{\Delta}_{W})\} \)
\( (\hat{\Sigma}_{W}^{(A)})^{1/2} \) \( (j = 1, 2) \),

where \( t_{W}^{(A)} \) and \( t_{W}^{(T,j)} \) \( (j = 1, 2) \) are for estimation of \(-2\hat{T}_{0}^{*}\) while \( t_{W}^{(A)} \) and
\( t_{W}^{(T,j)} \) \( (j = 1, 2) \) are for \(-2E_{g}(\hat{\Delta}_{W})\) under possible model misspecification; \( n^{-1}\hat{\Sigma}_{W}^{(A)} \) is the
robust estimator of the asymptotic variance $n^{-1} \alpha_{\text{ML2}}^{(A)}$ common to $n^{-1} \text{AIC}_w$ and $n^{-1} \text{TIC}^{(j)}_w (j = 1, 2)$:

$$
\hat{v}_w^{(A)} = 4(n - 1)^{-1} \sum_{j=1}^{n} (\hat{l}_{wj} - \hat{l}_w) = O_{p(1)}
$$

(4.10)

with $\hat{l}_{wj} \equiv l_j |_{\theta = \hat{\theta}_w}$ ($j = 1, ..., n$) and $\hat{l}_w = n^{-1} \sum_{j=1}^{n} \hat{l}_{wj}$ (for $l_j$ see (2.1)).

Under correct model specification, in many cases $\alpha_{\text{ML2}}^{(A)}$ may be explicitly obtained as a function of $\theta_0$. However, since this result depends on a model employed, the four versions of robust studentization in (4.9) are considered in this section. Define the stochastic quantity using $\theta_0$ in place of $\hat{\theta}_w$ in (4.10):

$$
\hat{v}_0^{(A)} = 4(n - 1)^{-1} \sum_{j=1}^{n} (l_{0j} - \bar{I}_0) = O_{p(1)}
$$

(4.11)

with $l_{0j} \equiv l_j |_{\theta = \theta_0}$ ($j = 1, ..., n$) and $\bar{I}_0 = n^{-1} \sum_{j=1}^{n} l_{0j}$. Then, $\hat{v}_0^{(A)}$ is an exactly unbiased robust estimator of $\alpha_{\text{ML2}}^{(A)}$ with $E_g(\hat{v}_0^{(A)}) = \alpha_{\text{ML2}}^{(A)}$ though $\hat{v}_0^{(A)}$ usually includes the unknown $\theta_0$. Generally, the estimator $\hat{v}_w^{(A)}$ is not an unbiased one but is a consistent estimator of $\alpha_{\text{ML2}}^{(A)}$.

Under possible model misspecification, assume that the following hold with the asymptotic cumulants, whose factors of order $O(1)$ are $\alpha_{(i)Wk}^{(A)} (k = 1, 2, \Delta 2, 3, 4)$ for $t_w^{(A)}$:

$$
\begin{align*}
\kappa_{g1}(t_w^{(A)}) &= n^{-1/2} \alpha_{(i)W1}^{(A)} + O(n^{-3/2}), \\
\kappa_{g2}(t_w^{(A)}) &= 1 + n^{-1} \alpha_{(i)W2}^{(A)} + O(n^{-2}) \quad (\alpha_{(i)W2}^{(A)} = 1), \\
\kappa_{g3}(t_w^{(A)}) &= n^{-1/2} \alpha_{(i)W3}^{(A)} + O(n^{-3/2}), \\
\kappa_{g4}(t_w^{(A)}) &= n^{-1} \alpha_{(i)W4}^{(A)} + O(n^{-2}).
\end{align*}
$$

(4.12)

Similarly, $\alpha_{(i)Wk}^{(T,j)}$ for $t_w^{(T,j)}$, $\alpha_{(i)Wk}^{(A)*}$ for $t_w^{(A)*}$ and $\alpha_{(i)Wk}^{(T,j)*}$ for $t_w^{(T,j)*} (j = 1, 2), (k = 1, 2, \Delta 2, 3, 4)$ are defined. These asymptotic cumulants are obtained.
However, since their derivations and results are relatively involved, they are shown in the first supplement to this paper (Ogasawara, 2016a).

Insert Tables 1 and 2 about here

5. Examples

Three examples are given in this section. Each of Examples 1 and 2 uses the MLE of a canonical parameter in the exponential family under model misspecification while Example 3 deals with the WSE of a canonical parameter in the exponential family under correct model specification. The asymptotic cumulants, obtained in Section 4, for the examples are shown in Tables 1 and 2, whose expository derivations are given in the supplements to this paper (Ogasawara, 2016a, 2016b).

Example 1: The MLE of the parameter in the exponential distribution is used when the gamma distribution with the shape parameter $\alpha$ being unequal to 1 holds. That is, the density

$$f(x^* = x | \lambda_0) = \lambda_0 \exp(-\lambda_0 x) \ (x > 0)$$

is used with $\theta_0 = \lambda_0$ when the true distribution is

$$g(x^* = x | \lambda_1, \alpha) = x^{\alpha-1} \lambda_1^\alpha \exp(-\lambda_1 x) / \Gamma(\alpha) \ (x > 0, \ \alpha \neq 1)$$

with $\xi_0 = (\lambda_1, \alpha)'$ and $\Gamma(\cdot)$ being the gamma function. By assumption $\alpha = 1$ is excluded. However, when $\alpha = 1$ in (5.2), this reduces to (5.1). The MLE of $\lambda_0$ is $1/\bar{x}$, where $\bar{x}$ is the sample mean of the observable variable. This gives the population $\lambda_0$ under model misspecification as

$$\lambda_0 = 1 / E_{\xi} (\bar{x}) = \lambda_1 / \alpha .$$

Example 2: The MLE of the mean in the univariate normal distribution with known variance $\sigma^2$ is used when the true distribution is non-normal with known variance $\sigma^2$. That is,

$$f(x^* = x | \mu_0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{ -\frac{(x - \mu_0)^2}{2\sigma^2} \right\}$$

with $\hat{\theta}_{ML} = \hat{\mu}_{ML} = \bar{x}$. In this example,
\[ \bar{t}_0^* = E_g(l_{0j}) = E_f(l_{0j}) = -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2}, \]

\[ E_g(\hat{\theta}_{ML}) = E_f(\hat{\theta}_{ML}) = \mu_0, \ n \text{var}_g(\hat{\theta}_{ML}) = n \text{var}_f(\hat{\theta}_{ML}) = \sigma^2. \]  

However, \( \text{var}_g(l_{0j}) = \frac{1}{4} \left\{ \kappa_{g4} \left( \frac{x^* - \mu_0}{\sigma} \right) + 2 \right\} \) under non-normality with \( \kappa_{g4}(\cdot) \neq 0 \) is not equal to \( \text{var}_f(l_{0j}) = 1/2 \) under normality.

**Example 3**: The WSE of the logit in the Bernoulli distribution is used under correct model specification. That is,

\[ \text{Pr}(x^* = x | \theta_0) = \pi_0^x (1 - \pi_0)^{1-x} (x = 0, 1), \quad \pi_0 = \frac{1}{1 + \exp(-\theta_0)}. \]  

While \( \hat{\theta}_{ML} = \log \frac{\bar{x}}{1 - \bar{x}} (\bar{x} \neq 0, 1) \), where \( \bar{x} \) is the usual sample proportion, \( \hat{\theta}_w \) in Example 3 is defined as the solution of \( \theta \) which maximizes

\[ \left\{ \prod_{j=1}^{n} \pi_{x_j} (1 - \pi)^{1-x_j} \right\} \{\pi(1-\pi)\}^{a/2} \text{ with } \pi = \frac{1}{1 + \exp(-\theta)}, \]  

where \( a \) is the sum of equal pseudocounts for two categories. The solution is given when

\[ \theta = \hat{\theta}_w = \log \frac{\bar{x} + n^{-1}0.5a}{1 - \bar{x} + n^{-1}0.5a}. \]

In the footnotes of the tables, general results associated with the tables are given (for derivation, see also Ogasawara, 2016a, b). In Examples 1 and 2, the results do not depend on scales since \( l \) (log-likelihood) is scale-free in these examples. Although \( \alpha \neq 1 \) is assumed in Example 1, \( \alpha = 1 \) gives the corresponding results under correct model specification. Note that in the latter case with \( \alpha = 1 \), all the results in Example 1 are given by fixed values. Under correct model specification, the bias-corrected \( n^{-1}\text{AIC}_{ML} \) up to order \( O(n^{-2}) \), denoted by \( n^{-1}\text{AIC}_{ML \rightarrow O(n^{-2})} \), is given by as simple as

\[ n^{-1}\text{AIC}_{ML \rightarrow O(n^{-2})} = -2\hat{\lambda}_{ML} + n^{-1}2 + n^{-2}2. \]  

Similarly, under normality, the results for Example 2 in the tables are given only by fixed values, where \( \kappa_j \)’s \( (\equiv \kappa_{g4} [(x^* - \mu_0) / \sigma] \)’s \( (j \neq 2) \) vanish. Note also that
In Example 2, \( n^{-1} \text{AIC}_{ML} (= n^{-1} \text{TIC}^{(j)}_{ML}, j = 1,2) \) is exactly unbiased even under non-normality (see (5.5) and Corollary 3). In Example 3, the results when \( \hat{\theta}_{ML} \) is used, are given by \( a = 0 \).

In Example 3, from Table 1 we have

**Corollary 4.** Under the assumption that the Bernoulli distribution holds, \( n^{-1} \text{AIC}_W \) for estimation of \(-2E_f (\hat{I}_W^*)\) using \( \hat{\theta}_W \) as the weighted score estimator of the logit with the total number \( a \) of equal pseudocounts for two categories gives no asymptotic bias up to order \( O(n^{-2}) \) when \( a = 1 \).

For the derivation of the higher-order asymptotic bias, see Ogasawara (2016b, Subsection S6.1). It is of interest to see that when \( a = 1 \), \( \hat{\theta}_W \) is also unbiased up to order \( O(n^{-1}) \) (see e.g., Ogasawara, 2013, Section 6). On the other hand, for estimation of \(-2\hat{I}_0^*\) the corresponding bias of \( n^{-1} \text{AIC}_W \) up to order \( O(n^{-2}) \) is

\[
n^{-1} + n^{-2} \{(1/6)(1-\bar{\pi}_0^{-1}) + (a^2/4)(1-2\pi_0)^2\bar{\pi}_0^{-1}\},
\]

which is minimized when \( a = 0 \) and \( \pi_0 \neq 1/2 \) while \( a \) is irrelevant to the asymptotic bias when \( \pi_0 = 1/2 \).

**References**


Biometrika, 83, 875-890.


Table 1. Asymptotic cumulants of \( n^{-1}\text{AIC}_{\text{ML}(W)} \) and \( n^{-1}\text{TIC}_{\text{ML}(W)}^{(j)} \) \((j = 1, 2)\) before studentization

<table>
<thead>
<tr>
<th></th>
<th>Example 1</th>
<th>Example 2</th>
<th>Example 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model distribution</strong></td>
<td>Exponential</td>
<td>normal with known (\sigma^2)</td>
<td>Bernoulli</td>
</tr>
<tr>
<td><strong>True distribution</strong></td>
<td>gamma, (\alpha \neq 1)</td>
<td>non-normal</td>
<td>Bernoulli</td>
</tr>
<tr>
<td><strong>Parameter</strong></td>
<td>canonical (the reciprocal of the scale)</td>
<td>canonical (mean)</td>
<td>canonical (logit)</td>
</tr>
<tr>
<td>AIC</td>
<td>(n^{-1}\text{AIC}_{\text{ML}})</td>
<td>(n^{-1}\text{AIC}<em>{\text{ML}}(= n^{-1}\text{TIC}</em>{\text{ML}}^{(3)}))</td>
<td>(n^{-1}\text{AIC}<em>{W}(= n^{-1}\text{TIC}</em>{W}^{(3)}))</td>
</tr>
<tr>
<td>(a_{\text{ML}(W)1}^{(A)})</td>
<td>(2 - \alpha^{-1})</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(a_{\text{ML}(W)A1}^{(A)})</td>
<td>(-(1/6)\alpha^{-2})</td>
<td>0</td>
<td>((1/6)(1 - t_0^{-1}) + (a^2/4)(1 - 2\pi_0)^2t_0^{-1})</td>
</tr>
<tr>
<td>(a_{\text{ML}(W)1}^{(A)*})</td>
<td>(2 - 2\alpha^{-1})</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(a_{\text{ML}(W)A1}^{(A)*})</td>
<td>(-2\alpha^{-2})</td>
<td>0</td>
<td>((a-1)((1 - 2\pi_0)^2t_0^{-1} + 2))</td>
</tr>
<tr>
<td>(a_{\text{ML}(W)2}^{(A)})</td>
<td>(4\alpha^{-1})</td>
<td>(\kappa_4 + 2)</td>
<td>(4\theta_0^2t_0)</td>
</tr>
<tr>
<td>(a_{\text{ML}(W)A2}^{(A)})</td>
<td>(2\alpha^{-2})</td>
<td>(-2(\kappa_4 + 1))</td>
<td>2</td>
</tr>
<tr>
<td>(a_{\text{ML}(W)3}^{(A)})</td>
<td>(-8\alpha^{-2})</td>
<td>(\kappa_8 + 12\kappa_4 + 4\kappa_4^2 + 8)</td>
<td>(-8\theta_0^3(1 - 2\pi_0) + 24\theta_0^2\kappa_0)</td>
</tr>
<tr>
<td>(a_{\text{ML}(W)4}^{(A)})</td>
<td>(32\alpha^{-3})</td>
<td>(\kappa_8 + 24\kappa_6 + 32\kappa_4\kappa_3)</td>
<td>(16\theta_0^6(1 - 6\pi_0 + 6\pi_0^2) + 128\theta_0^3(1 - 2\pi_0) + 192\theta_0^2t_0)</td>
</tr>
<tr>
<td>(\alpha_{\text{ML}(W)5}^{(A)})</td>
<td>(-64\alpha^{-4})</td>
<td>(\kappa_8 + 192\kappa_6 + 768\kappa_4\kappa_3 + 192\kappa_4^3 + 768\kappa_6^2 + 768\kappa_4^2\kappa_2 + 768\kappa_4\kappa_2^2 + 768\kappa_2^4 + 768\kappa_2^2\kappa_4^2 + 768\kappa_2^4)</td>
<td>(256\theta_0^{10}(1 - 10\pi_0 + 30\pi_0^2 - 10\pi_0^3 + \pi_0^4) + 1024\theta_0^7(1 - 4\pi_0) + 2304\theta_0^5(1 - 2\pi_0) + 4608\theta_0^3t_0)</td>
</tr>
<tr>
<td><strong>Higher-order bias correction</strong></td>
<td>(see the case of (n^{-1}\text{TIC}_{\text{ML}}^{(5)}) below)</td>
<td>(n^{-1}\text{AIC}<em>{\text{ML}}(= n^{-1}\text{TIC}</em>{\text{ML}}^{(3)}))</td>
<td>(n^{-1}\text{AIC}<em>{W}(= n^{-1}\text{TIC}</em>{W}^{(3)}))</td>
</tr>
<tr>
<td></td>
<td>is unbiased</td>
<td>(= n^{-1}\text{TIC}_{W\to O(n^{-2})}^{(3)})</td>
<td>(= -2\hat{t}_0 + n^{-2})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+ n^{-2}(1 - a)[(1 - 2x)^2 x(1 - x)]^{-1} + 2)</td>
<td></td>
</tr>
</tbody>
</table>

(to be continued)
Table 1. (continued)

| Example 1 \((\alpha_{\text{ML2}}^{(T3)} = \alpha_{\text{ML2}}^{(A)} \text{ in this example})\) |
|-------------------------------|---------------------------------------------------------|
| TIC                          | \(n^{-1}\text{TIC}_{\text{ML}}^{(3)}\)                 |
| \(\alpha_{\text{ML1}}^{(T3)}\) | \(\alpha^{-1}\)                                        |
| \(\alpha_{\text{MLI}}^{(T3)}\) | \(-\frac{1}{6\alpha^2} + \frac{2\alpha\psi''(\alpha) + \psi'(\alpha)}{\alpha\{\alpha\psi'(\alpha) - 1\}^2}\) |
| \(\alpha_{\text{ML1}}^{(T3)*}\) | \(0\)                                                  |
| \(\alpha_{\text{MLI}}^{(T3)*}\) | \(-\frac{2}{\alpha^2} + \frac{2\alpha\psi''(\alpha) + \psi'(\alpha)}{\alpha\{\alpha\psi'(\alpha) - 1\}^2}\) |

Higher-order bias correction

\[ n^{-1}\text{TIC}_{\text{ML}}^{(3)} \rightarrow O\left(\sigma^{-2}\right) \]

\[ = -2\hat{\alpha}_{\text{ML}} + n^{-1} \frac{2}{\alpha} + n^{-2} \left[ \frac{2}{\alpha^2} - \hat{\alpha}\psi''(\hat{\alpha}) + \psi'(\hat{\alpha}) \right] \]

Note. \(\alpha_{\text{MLj}}^{(A)}\) are for \(\kappa_{jg}(n^{-1}\text{AIC}_{\text{ML}} + 2\hat{t}_{0}^{*})\) while \(\alpha_{\text{MLj}}^{(A)*}\) are for \(\kappa_{jg}{\left\{n^{-1}\text{AIC}_{\text{ML}} + 2\hat{E}_g(\hat{t}_{\text{ML}}^{*})\right\}}\) \((j = 1, \Delta1, 2, \Delta2, 3, 4)\) in Examples 1 and 2. Similarly, \(\alpha_{\text{Wj}}^{(A)}\) and \(\alpha_{\text{Wj}}^{(A)*}\) are defined in Example 3. \(\psi'()\) and \(\psi''()\) are the first and second derivatives of the digamma function \(\psi()\), respectively.

Generally, \(\alpha_{\text{MLj}}^{(A)} = \alpha_{\text{MLj}}^{(A)*}\), \(\alpha_{\text{Wj}}^{(A)} = \alpha_{\text{Wj}}^{(A)*}\) \((j = 2, \Delta2, 3, 4)\), \(\alpha_{\text{Wj}}^{(A)} = \alpha_{\text{MLj}}^{(A)}\) \((j = 1, 2, \Delta2, 3, 4)\), \(\alpha_{\text{Wj}}^{(T3)} = \alpha_{\text{MLj}}^{(T3)} = \alpha_{\text{Wj}}^{(A)} = \alpha_{\text{MLj}}^{(A)}\) \((j = 2, 3, 4)\) .

In Example 1, \(\kappa_{jg} \equiv \kappa_{jg} \{(x^* - \mu_0) / \sigma\}\) and in Example 3 \(\hat{t}_{0} = \pi_0 (1 - \pi_0)\) is the population Fisher information per observation.
Table 2. Asymptotic cumulants of $n^{-3}AIC_{ML(W)}$ and $n^{-3}TIC_{ML(W)}^{(j)}$ ($j=1, 2$) after studentization

<table>
<thead>
<tr>
<th>Example 1</th>
<th>Example 2</th>
<th>Example 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AIC</strong></td>
<td>$n^{-3}AIC_{ML}$</td>
<td>$n^{-3}AIC_{ML}(=n^{-1}TIC_{ML}^{(3)})$</td>
</tr>
<tr>
<td>$\alpha_{(j)ML(W)}^{(A)}$</td>
<td>$\alpha^{1/2}-(1/2)\alpha^{-1/2}$</td>
<td>$(\kappa_4 + 2)^{-1}-(1/2)(\kappa_4 + 2)^{3/2}$</td>
</tr>
<tr>
<td></td>
<td>$\times(\kappa_6 + 12\kappa_4 + 6\kappa_3^2 + 8)$</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{(j)ML(W)}^{(A)*}$</td>
<td>$\alpha^{1/2}-\alpha^{-1/2}$</td>
<td>$-(1/2)(\kappa_4 + 2)^{-3/2}$</td>
</tr>
<tr>
<td></td>
<td>$\times(\kappa_6 + 12\kappa_4 + 6\kappa_3^2 + 8)$</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{(j)ML(W)2}$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\alpha_{(j)ML(W)3/2}$</td>
<td>$7/2\alpha^{-1} + 2$</td>
<td>$2-2(\kappa_4 + 2)^{-1}+(\kappa_4 + 2)^{-2}(-\kappa_6$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$+8\kappa_4\kappa_3 + 2\kappa_4^2 - 4\kappa_4 + 50\kappa_3^2)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$+(\kappa_4 + 2)^{-3}(7/4)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\times(\kappa_6 + 12\kappa_4 + 6\kappa_3^2 + 8)^2$</td>
</tr>
<tr>
<td>$\alpha_{(j)ML(W)3}$</td>
<td>$7/2\alpha^{-1} + 2$</td>
<td>$\alpha_{(j)MLA2}^{(A)}(\kappa_4 + 2)^{-2}$</td>
</tr>
<tr>
<td></td>
<td>$=$ $\alpha_{(j)MLA2}^{(A)*}$ in Example 1</td>
<td>$\times(\kappa_6 + 12\kappa_4 + 6\kappa_3^2 + 8)$</td>
</tr>
<tr>
<td>$\alpha_{(j)ML(W)4}$</td>
<td>$8\alpha^{-1} + 6$</td>
<td>$12-18(\kappa_4 + 2)^{-1}+(\kappa_4 + 2)^{-2}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\times(-2\kappa_6 - 48\kappa_4 + 64\kappa_4\kappa_3 - 70\kappa_4^2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$-294\kappa_4^2 - 144\kappa_3^2 - 84)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$+(\kappa_4 + 2)^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\times{12(\kappa_6 + 12\kappa_4 + 6\kappa_3^2 + 8)^2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$+12(\kappa_6 + 12\kappa_4 + 6\kappa_3^2 + 8)\kappa_3^2}$</td>
</tr>
</tbody>
</table>

Note. Generally, $\alpha_{(j)W}^{(A)} = \alpha_{(j)ML}^{(A)}$ $(j=1, 2, 3, 4)$, $\alpha_{(j)W}^{(A)*} = \alpha_{(j)ML}^{(A)*} = \alpha_{(j)W}^{(A)} = \alpha_{(j)ML}^{(A)} = 1$ and $\alpha_{(j)W}^{(A)*} = \alpha_{(j)ML}^{(A)*} = \alpha_{(j)W}^{(A)} = \alpha_{(j)ML}^{(A)} = \alpha_{(j)W}^{(A)} = \alpha_{(j)ML}^{(A)}$ $(j=3, 4)$. Generally, $\alpha_{(j)W}^{(A)} = \alpha_{(j)ML}^{(A)} = \alpha_{(j)W}^{(A)} = \alpha_{(j)ML}^{(A)} = 1$ and $\alpha_{(j)W}^{(A)*} = \alpha_{(j)ML}^{(A)*} = \alpha_{(j)W}^{(A)} = \alpha_{(j)ML}^{(A)} = \alpha_{(j)W}^{(A)} = \alpha_{(j)ML}^{(A)}$ $(j=3, 4)$. |
<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Author/s</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ホーキンズ＝サイモンの条件に関する諸説の統合について</td>
<td>ダスグプタ, ディパンカー</td>
<td>Jul.1992</td>
</tr>
<tr>
<td>2</td>
<td>Motivation and Causal Inferences in the Budgetary Control</td>
<td>Yoshihiro Naka</td>
<td>Aug.1992</td>
</tr>
<tr>
<td>3</td>
<td>プロセスによる労働者の能力に影響する対策の診断</td>
<td>マイサイモリィ, シュルキ</td>
<td>Nov.1992</td>
</tr>
<tr>
<td>5</td>
<td>Business Cycles with Asset Price Bubbles and the Role of Monetary Policy</td>
<td>Hiroshi Shibuya</td>
<td>Jun.1993</td>
</tr>
<tr>
<td>6</td>
<td>Continuous Double-Sided Auctions in Foreign Exchange Markets</td>
<td>Ryosuke Wada</td>
<td>Aug.1993</td>
</tr>
<tr>
<td>7</td>
<td>The Existence of Ramsey Equilibrium with Consumption Externality</td>
<td>Sadao Kanaya, Tomoichi Shinotsuka</td>
<td>Sep.1993</td>
</tr>
<tr>
<td>8</td>
<td>Money, Neutrality of Consumption Taxes, and Growth in Inter-temporal Optimizing Models</td>
<td>Jun-ichi Ito</td>
<td>Nov.1993</td>
</tr>
<tr>
<td>12</td>
<td>Sit-Down to Split: Flint GM Workers in 1937-1939</td>
<td>Satosshi Takata</td>
<td>Dec.1994</td>
</tr>
<tr>
<td>13</td>
<td>The Complementarity between Endogenous Protection and Direct Foreign Investment</td>
<td>Laixun Zhao</td>
<td>Feb.1995</td>
</tr>
<tr>
<td>14</td>
<td>Consumption Taxation and Tax Prepayment approach in Dynamic General equilibrium Models with Consumer Durables</td>
<td>Jun-ichi Itaya</td>
<td>Mar.1995</td>
</tr>
<tr>
<td>15</td>
<td>Regulatory System and Supervision of the Financial Institutions in Japan</td>
<td>Osamu Ito</td>
<td>May 1995</td>
</tr>
<tr>
<td>17</td>
<td>The Legacy of the Bubble Economy in Japan: Declining cross Shareholding and Capital Formation</td>
<td>Hiroo Hojo</td>
<td>May 1995</td>
</tr>
<tr>
<td>18</td>
<td>Stockownership in the U.S.: Capital Formation and Regulation</td>
<td>Marshall E. Blume</td>
<td>May 1995</td>
</tr>
<tr>
<td>19</td>
<td>International Joint Ventures and Endogenous Protection - A Political-Economy Approach</td>
<td>Laixun Zhao</td>
<td>Nov.1995</td>
</tr>
<tr>
<td>20</td>
<td>GM社をめぐるアフィリート労働研究: フィアントアイディアの現場像の吟味</td>
<td>高田聡</td>
<td>Feb.1996</td>
</tr>
<tr>
<td>23</td>
<td>A Note on the Impacts of Price Shocks on Wage in Unionized Economies</td>
<td>Laixun Zhao</td>
<td>May 1996</td>
</tr>
<tr>
<td>27</td>
<td>Characteristics and Reforms of Public Health Insurance System in Japan</td>
<td>Takashi Nakahama</td>
<td>Sep.1996</td>
</tr>
<tr>
<td>28</td>
<td>The Role of Local Governments in Urban Development Policy</td>
<td>Yoshinori Akiyama</td>
<td>Sep.1996</td>
</tr>
<tr>
<td>30</td>
<td>Comparison of Agricultural Policy in the U.S. and the Japan</td>
<td>Toshikazu Tateiwa</td>
<td>Oct.1996</td>
</tr>
<tr>
<td>31</td>
<td>US Health Insurance: Types, Patterns of Coverage and Constraints to Reform</td>
<td>Dwayne A. Banks</td>
<td>Oct.1996</td>
</tr>
<tr>
<td>36</td>
<td>A History of PBGC and Its Roles</td>
<td>C. David Gustafson</td>
<td>Nov.1996</td>
</tr>
<tr>
<td>ページ</td>
<td>タイトル</td>
<td>ディスカッションペーパーシリーズ</td>
<td>オーガニゼーション</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>---------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>44.</td>
<td>サハリン石油・ガス開発プロジェクトと北海道経済の活性化 第1号</td>
<td>北東アジア-サハリン研究会</td>
<td>May 1998</td>
</tr>
<tr>
<td>45.</td>
<td>企業間の競争と単権の不正課税行動について</td>
<td>伊藤 一</td>
<td>Jun.1998</td>
</tr>
<tr>
<td>46.</td>
<td>The Transition of Postwar Asia-Pacific Trade Relations</td>
<td>Masahiro Endoh</td>
<td>Jul.1998</td>
</tr>
<tr>
<td>53.</td>
<td>Tariff induced dumping in the intermediate-good market</td>
<td>Chisato Shibayama</td>
<td>Apr.2000</td>
</tr>
<tr>
<td>55.</td>
<td>サハリン石油・ガス開発プロジェクトと北海道経済の活性化 第2号</td>
<td>北東アジア-サハリン研究会</td>
<td>May 1999</td>
</tr>
<tr>
<td>58.</td>
<td>サハリン石油・ガス開発プロジェクトと北海道経済の活性化 第3号</td>
<td>北東アジア-サハリン研究会</td>
<td>May 2000</td>
</tr>
<tr>
<td>60.</td>
<td>Foreign Exchange Market Maker’s Optimal Spread with Heterogeneous Expectations</td>
<td>Kyosuke Wada</td>
<td>Jun.2000</td>
</tr>
<tr>
<td>63.</td>
<td>業態開発におけるイノベーションと競争 ~ ビブレのケースと</td>
<td>近藤公彦</td>
<td>Jan.2001</td>
</tr>
<tr>
<td>64.</td>
<td>Budget Distribution Problem</td>
<td>Tomoichi Shinotsuka</td>
<td>Feb.2001</td>
</tr>
<tr>
<td>67.</td>
<td>サハリン石油・ガス開発プロジェクトと北海道経済の活性化 第4号</td>
<td>北東アジア-サハリン研究会</td>
<td>Mar.2001</td>
</tr>
<tr>
<td>69.</td>
<td>環境保全型再生計画と環境保全に係る計画技術の研究</td>
<td>地域環境問題研究会</td>
<td>Oct.2001</td>
</tr>
<tr>
<td>70.</td>
<td>Additivity, Bounds, and Continuity in Budget Distribution Problem</td>
<td>Tomoichi Shinotsuka</td>
<td>Dec.2001</td>
</tr>
<tr>
<td>番号</td>
<td>タイトル</td>
<td>著者</td>
<td>発行年月</td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>-------</td>
<td>----------</td>
</tr>
<tr>
<td>78.</td>
<td>Optimal Multiobject Auctions with Correlated Types</td>
<td>Tomoichi Shinotsuka &amp; Simon Wilkie</td>
<td>Feb. 2002</td>
</tr>
<tr>
<td>79.</td>
<td>産業振興のサービス構成要素に関する重要度調査法に関する考察 -北海道への台湾人観光客の事例を中心に</td>
<td>稲葉由之 &amp; 沈潔如</td>
<td>Feb. 2003</td>
</tr>
<tr>
<td>81.</td>
<td>サハリン石油・ガス開発プロジェクトと北海道経済の活性化 第5号</td>
<td>北東アジア-サハリン研究会</td>
<td>Mar. 2002</td>
</tr>
<tr>
<td>82.</td>
<td>The Comparison Study on Retail Buyer Behaviour between UK, Australia and Japan</td>
<td>Hajime Itoh</td>
<td>Mar. 2003</td>
</tr>
<tr>
<td>83.</td>
<td>宿泊業のサービスの品質構成要素に関する重要度調査法に関する考察 -「買い手視角」での小売購買行動研究に向けて-</td>
<td>坂川裕司</td>
<td>May 2003</td>
</tr>
<tr>
<td>84.</td>
<td>Interdependent Utility Functions in an Intergenerational Context</td>
<td>Tomoichi Shinotsuka</td>
<td>May 2003</td>
</tr>
<tr>
<td>86.</td>
<td>マーケティング行動からみた小売業による需要創造 -明治期長崎の経営行動を考察対象として-</td>
<td>松尾 睦</td>
<td>Sep. 2003</td>
</tr>
<tr>
<td>87.</td>
<td>An Excerpt of Round-Robin Tournament by Excel's Macro</td>
<td>Masaru Uzawa</td>
<td>Apr. 2004</td>
</tr>
<tr>
<td>89.</td>
<td>Competition between Matching Markets</td>
<td>Koji Yokota</td>
<td>May 2005</td>
</tr>
<tr>
<td>90.</td>
<td>A Note on Jones' Model of Growth</td>
<td>Mutsuhiro Kato</td>
<td>Jun. 2006</td>
</tr>
<tr>
<td>94.</td>
<td>A Note on Optimal Taxation in the Presence of Externalities</td>
<td>Tomoichi Shinotsuka &amp; Ko Sumino</td>
<td>Feb. 2005</td>
</tr>
<tr>
<td>95.</td>
<td>A Note on the existence of a penalty for a knapsack problem</td>
<td>Iida Hiroshi</td>
<td>Mar. 2007</td>
</tr>
<tr>
<td>96.</td>
<td>Comments on knapsack problems with a penalty</td>
<td>Iida Hiroshi</td>
<td>Jul. 2007</td>
</tr>
<tr>
<td>100.</td>
<td>A Note on the existence of a penalty for a knapsack problem</td>
<td>Iida Hiroshi</td>
<td>Mar. 2007</td>
</tr>
<tr>
<td>103.</td>
<td>A Note on Optimal Taxation in the Presence of Externalities</td>
<td>Tomoichi Shinotsuka &amp; Ko Sumino</td>
<td>Feb. 2005</td>
</tr>
<tr>
<td>104.</td>
<td>A Note on Jones' Model of Growth</td>
<td>Mutsuhiro Kato</td>
<td>Mar. 2005</td>
</tr>
<tr>
<td>107.</td>
<td>A Note on the existence of a penalty for a knapsack problem</td>
<td>Iida Hiroshi</td>
<td>Mar. 2007</td>
</tr>
<tr>
<td>108.</td>
<td>Comments on knapsack problems with a penalty</td>
<td>Iida Hiroshi</td>
<td>Jul. 2007</td>
</tr>
<tr>
<td>113.</td>
<td>A Note on the existence of a penalty for a knapsack problem</td>
<td>Iida Hiroshi</td>
<td>Mar. 2007</td>
</tr>
<tr>
<td>114.</td>
<td>A Note on the existence of a penalty for a knapsack problem</td>
<td>Iida Hiroshi</td>
<td>Mar. 2007</td>
</tr>
<tr>
<td>115.</td>
<td>A Note on the existence of a penalty for a knapsack problem</td>
<td>Iida Hiroshi</td>
<td>Mar. 2007</td>
</tr>
<tr>
<td>116.</td>
<td>A Note on the existence of a penalty for a knapsack problem</td>
<td>Iida Hiroshi</td>
<td>Mar. 2007</td>
</tr>
<tr>
<td>Topic</td>
<td>Authors</td>
<td>Date</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>----------------------------------------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>162. Perverse effects of a ban on child labour in an overlapping</td>
<td>Kouki Sugawara &amp; Atsue Mizushima &amp; Koichi</td>
<td>Oct. 2013</td>
<td></td>
</tr>
<tr>
<td>generations model</td>
<td>Futagami</td>
<td></td>
<td></td>
</tr>
<tr>
<td>164. 18・19世紀前半北海沿岸農村社会の指導的地域役職者・領邦地方官吏</td>
<td>平井進</td>
<td>Mar. 2014</td>
<td></td>
</tr>
<tr>
<td>と土地区画：Landschaft Norderdithmarschen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165. ビジネスシステムの形成から見る6次産業化—バイオニアジャパングループの事例分析—</td>
<td>笹本香菜 &amp; 加藤敬太</td>
<td>Mar. 2014</td>
<td></td>
</tr>
<tr>
<td>166. ナップサック問題への2近似算法について雑感</td>
<td>飯田浩志</td>
<td>Jul. 2014</td>
<td></td>
</tr>
<tr>
<td>167. A further addendum to &quot;Some thoughts on the 2-approximation</td>
<td>Hiroshi Iida</td>
<td>Nov. 2014</td>
<td></td>
</tr>
<tr>
<td>algorithm for knapsack problems: A survey</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>168. ワイン産業における企業家コミュニティ北海道と山梨県の比較事例研究—</td>
<td>長村知幸</td>
<td>Jan. 2015</td>
<td></td>
</tr>
<tr>
<td>169. サハリン住宅建設企業の動向—2013年9月調査に基づいて—</td>
<td>小田福男</td>
<td>Mar. 2015</td>
<td></td>
</tr>
<tr>
<td>170. Asymptotic cumulants of some information criteria</td>
<td>Haruhiko Ogawara</td>
<td>Apr. 2015</td>
<td></td>
</tr>
</tbody>
</table>