Journal of Mathematical Research with Applications
Jan., 2013, Vol. 33, No. 1, pp. 1–10
DOI:10.3770/j.issn:2095-2651.2013.01.001
Http://jmre.dlut.edu.cn

Bias Correction for Alternating Iterative Maximum Likelihood Estimators

Gang YU[1,*], Wei GAO [2], Ningzhong SHI [2]

1. *School of Mathematics and Quantitative Economics, Dongbei University of Finance and Economics, Liaoning* 116025*, P. R. China;*
2. *Key Laboratory for Applied Statistics of MOE and School of Mathematics and Statistics, Northeast Normal University, Liaoning* 130024*, P. R. China*

Abstract In this paper, we give a definition of the alternating iterative maximum likelihood estimator (AIMLE) which is a biased estimator. Furthermore we adjust the AIMLE to result in asymptotically unbiased and consistent estimators by using a bootstrap iterative bias correction method as in Kuk (1995). Two examples and simulation results reported illustrate the performance of the bias correction for AIMLE.

Keywords maximum likelihood estimation (MLE); alternating iterative maximum likelihood estimator (AIMLE); asymptotic normality; bias correction.

MR(2010) Subject Classification 62F10; 62F12; 62F40

1. Introduction

The method of maximum likelihood is, by far, the most popular technique for deriving estimators, which can be expressed in the form

$$L(\widehat{\theta}(x);x)=\max_{\theta\in C}L(\theta;x), \tag{1}$$

where $L(\theta;x)$ is log-likelihood function, x is a given sample point, θ is a k dimensional vector of parameters, $\widehat{\theta}(x)$ is the MLE of θ and C is a convex subset of R^k. If the function $L(\theta;x)$ is concave in θ, then the problem (1) can be transformed into the convex programming problem, which has been solved by some approaches provided by Avirel [1], Peressini et al [9] and Stoer [12]. In many estimating problems, however, the log-likelihood function $L(\theta;x)$ may not be concave. Recently, Shi et al [10] studied the problem (1) when $L(\theta;x)$ is semi-concave by using an alternating iterative method (AIM). A semi-concave function is defined as follows:

Definition *Let C be a convex subset of R^{k_1} and D be a convex subset of R^{k_2}. A function $f(\theta,\varphi;x)$ is said to be a semi-concave function of (θ,φ) if*

Received June 26, 2011; Accepted March 27, 2012
Supported by the National Natural Science Foundation of China (Grant Nos. 71171035; 71173029; 10931002; 11071035), the Program for New Century Excellent Talents (Grant No. NCET-10-315) and Excellent Talents Program of Liaoning Educational Committee (Grant No. 2008RC15).
* Corresponding author
E-mail address: yugang543@163.com (Gang YU); gaow@nenu.edu.cn (Wei GAO); shinz@nenu.edu.cn (Ningzhong SHI)

i) $f(\theta,\varphi;x)$ *is defined on* $C\times D$;

ii) For any given $\theta\in C, f(\theta,\cdot;x)$ *is strictly concave on* D, *and for any given* $\varphi\in D, f(\cdot,\varphi;x)$ *is strictly concave on* C.

In fact, many log-likelihood functions are semi-concave functions. Two examples are provided in Section 4.

The problems of MLE when log-likelihood function is semi-concave can be written:

$$\max_{\theta\in C,\varphi\in D} L(\theta,\varphi;x), \tag{2}$$

where $L(\theta,\varphi;x)$ is a semi-concave function defined on $C\times D$, x is a given sample, C is a convex subset of R^{k_1} and D is a convex subset of R^{k_2}. Under some conditions, Shi et al [10] proposed AIM to solve problem (2) and shown that the AIM converged. In fact, it can been found that the estimated value is a biased , when the AIM is applied to solve problem (2). For adjusting bias, in Section 2, we give a definition of the alternating iterative maximum likelihood estimator (AIMLE) through AIM for problem (2). And we adjust AIMLE to result in asymptotically unbiased and consistent estimators by using the iterative bias correction method as in Kuk [7]. Section 3 completes the proof of Theorem 1. In Section 4, two examples and a simulation will be given to illustrate the AIMLE and the bias correction for AIMLE.

2. The AIMLE and bias correction

In this section, we propose the alternating iterative maximum likelihood estimator (AIMLE) for problem (2), and discuss the asymptotic properties of the AIMLE. According to the definition of a semi-concave function, problem (2) can be transformed into problem (1) if θ or φ is fixed. To derive the AIMLE, the following alternating iterative method will be useful:

Step (0, 1) For any $\varphi\in D$, find $\theta^{(0)}(\varphi;x)$, the maximum point of $L(\theta,\varphi;x)$ on C;

Step (0, 2) For $\theta^{(0)}(\varphi;x)$, find $\varphi^{(0)}(x)$, the maximum point of $L(\theta^{(0)}(\varphi;x),\varphi;x)$ on D;

Step $(n, 1)$ For $\varphi^{(n-1)}(x)$, find $\theta^{(n)}(x)$, the maximum point of $L(\theta,\varphi^{(n-1)}(x);x)$ on C;

Step $(n, 2)$ For $\theta^{(n)}(x)$, find $\varphi^{(n)}(x)$, the maximum point of $L(\theta^{(n)}(x),\varphi;x)$ on D.

From the above algorithm, we can obtain two point estimation value sequences $\{\theta^{(n)}(x)\}\subset C$ and $\{\varphi^{(n)}(x)\}\subset D$. It can also be seen that, for $n\geq 1$

$$L(\theta^{(n)}(x),\varphi^{(n)}(x);x)\leq L(\theta^{(n+1)}(x),\varphi^{(n)}(x);x)\leq L(\theta^{(n+1)}(x),\varphi^{(n+1)}(x);x). \tag{3}$$

Since the sequences $\{\theta^{(n)}(x)\}$ and $\{\varphi^{(n)}(x)\}$ are obtained alternately, if they converge to $\widehat{\theta}(x),\widehat{\varphi}(x)$ respectively under some regularity conditions, the estimators $\widehat{\theta}(x)$ and $\widehat{\varphi}(x)$ are both called the alternating iterative maximum likelihood estimator (AIMLE). Since, due to non-linearity, the AIMLE have no explicit expression of the sample point x. The AIMLE $\widehat{\varphi}(x)$ can also be seen as solving the following problem:

$$\max_{\varphi\in D} L(\widehat{\theta}(x),\varphi;x). \tag{4}$$

Similarly to the AIMLE $\widehat{\varphi}(x)$, the AIMLE $\widehat{\theta}(x)$ solves the following problem:

$$\max_{\theta\in C} L(\theta,\widehat{\varphi}(x);x). \tag{5}$$

Since $L(\widehat{\theta}(x),\varphi;x)$ and $L(\theta,\widehat{\varphi}(x);x)$ are obtained alternately by using AIM, which are similar to profile likelihood function, the two functions are called alternating iterative profile likelihood function (AIPLF). Some results about the profile likelihood function have been studied by some authors, for example, Bartlett [3], Barndorff-Nielsen [2] and so on.

By assumption, for any $\theta \in C$, $L(\theta,\cdot;x)$ is strictly concave on D. So, we can establish a continuous mapping relation $\varphi(\cdot)$ from C to D which satisfies $L(\theta,\varphi(\theta);x) = \max_{\varphi\in D} L(\theta,\varphi;x)$, similarly to $\varphi(\cdot)$, we can also establish a continuous mapping $\theta(\cdot)$ from D to C such that $L(\theta(\varphi),\varphi;x) = \max_{\theta\in C} L(\theta,\varphi;x)$. The proof for the continuity of $\varphi(\cdot)$ will be provided in Section 3. According to two continuous mappings, $\theta(\cdot)$ and $\varphi(\cdot)$, a continuous composite mapping $\theta\circ\varphi(\cdot)$ from C to C can be obtained. So the alternating iterative formula can be written as $\theta^{(n+1)}(x) = \theta\circ\varphi(\theta^{(n)}(x)) = \theta(\varphi(\theta^{(n)}(x))$ $(n \geq 1)$.

We next present conditions for the convergence of alternating iterative sequences $\{\theta^{(n)}(x)\}$ and $\{\varphi^{(n)}(x)\}$.

Theorem 1 (Convergence) *Let the following assumptions hold:*

(A1) The log-likelihood function $L(\theta;\varphi;x)$ is semi-concave function defined on $C\times D$, x is a given sample, C is a convex subset of R^{k_1} and D is a convex subset of R^{k_2}.

(A2) The log-likelihood function $L(\theta;\varphi;x)$ is continuous second-order partial derivatives on an open set contained in $C\times D$.

(A3) For any given $\varphi\in D$, the Hessian matrix $(\frac{\partial^2 L(\theta;\varphi;x)}{\partial\theta_i\partial\theta_j})$ of $L(\theta;\varphi;x)$ is negative definite for $\theta\in C$.

(A4) The continuous composite mapping $\theta\circ\varphi(\cdot)$ has at most countable fixed points in C.

Let the alternating iterative sequences $\{\theta^{(n)}(x)\}$ and $\{\varphi^{(n)}(x)\}$ obtained by the alternating iterative method be a sequence of solutions to the problem

$$\max_{\theta\in C,\varphi\in D} L(\theta,\varphi;x). \tag{6}$$

For $k_1 \geq 2$, then $\{\theta^{(n)}(x),\varphi^{(n)}(x)\}$ converges.

The proof of Theorem 1 is similar to that of Shi et al [10] and is provided in the section 3.

Remark For $k_1 = 1$, let assumptions (A1)–(A3) hold, without assumptions (A4), then $\{\theta^{(n)}(x), \varphi^{(n)}(x)\}$ also converges.

It is well known that the alternating iterative profile likelihood function (AIPLF) is not a true likelihood function, for example, $E_\theta\{\frac{\partial L(\theta,\widehat{\varphi}(x);x)}{\partial\theta}\} \neq 0$. The AIMLE obtained by AIPLF is a biased estimator, though alternating iterative sequences $\{\theta^{(n)}(x)\}$ and $\{\varphi^{(n)}(x)\}$ can converge to the AIMLE $\widehat{\theta}(x)$ and $\widehat{\varphi}(x)$ respectively under some conditions.

From a first order expansion of the concentrated score $\frac{\partial L(\theta,\widehat{\varphi}(x);x)}{\partial\theta}$ around the value $\theta^\star$ which is not the true value θ and called pseudo true value, we obtain the usual expression for the AIMLE $\widehat{\theta}(x)$

$$H_N\sqrt{N}(\widehat{\theta}(x)-\theta^\star)=-\frac{1}{\sqrt{N}}\frac{\partial L(\theta,\widehat{\varphi}(x);x)}{\partial\theta}|_{\theta=\theta^\star}+O_p(\frac{1}{\sqrt{N}}),\tag{7}$$

where $\theta^\star=h(\theta)$ is contained in the following equation (8) for given $\widehat{\varphi}(x)$

$$E_\theta\{\frac{\partial L(\theta,\widehat{\varphi}(x);x)}{\partial\theta}|_{\theta=\theta^\star}\}=0,\tag{8}$$

and

$$H_N=\frac{1}{N}\frac{\partial^2 L(\theta,\widehat{\varphi}(x);x)}{\partial\theta^2}|_{\theta=\theta^\star}.\tag{9}$$

A standard central limit theorem applies to the concentrated score $\frac{\partial L(\theta,\widehat{\varphi}(x);x)}{\partial\theta}|_{\theta=\theta^\star}$. From (8) we have

$$\frac{1}{\sqrt{N}}\frac{\partial L(\theta,\widehat{\varphi}(x);x)}{\partial\theta}|_{\theta=\theta^\star}\xrightarrow{d}N(\mathbf{0},V_N),\tag{10}$$

where

$$V_N=\frac{1}{N}E_\theta\{\frac{\partial L(\theta,\widehat{\varphi}(x);x)}{\partial\theta}|_{\theta=\theta^\star}(\frac{\partial L(\theta,\widehat{\varphi}(x);x)}{\partial\theta}|_{\theta=\theta^\star})^T\}.\tag{11}$$

Finally, combining (7) and (10), we have the following Theorem 2:

Theorem 2 (Asymptotic normality of AIMLE) *When sample size* $N\longrightarrow\infty$

$$\sqrt{N}(\widehat{\theta}(x)-\theta^\star)\xrightarrow{d}N(\mathbf{0},\Lambda),\tag{12}$$

where

$$\Lambda=H_N^{-1}\times V_N\times(H_N^{-1})^T.\tag{13}$$

It is well known that the estimating equation obtained by using AIPLF is generally biased so that $\theta^\star\neq\theta$. From (12), the asymptotic bias of the AIMLE $\widehat{\theta}(x)$ is $b(\theta)=\theta^\star-\theta=h(\theta)-\theta$. Kuk [7] proposed sampling-based methods to study the estimating problem for generalized linear models with random effects. Also the method given in Kuk [7] can be applied to our proposed bias correction problems. The method for correcting the bias $b(\theta)$ is the following. Let $b^{(0)}$ be an initial estimate of the bias of $\widehat{\theta}(x)$. The $k+1$ step updated estimate of bias of $\widehat{\theta}(x)$ can be written as

$$b^{(k+1)}=h(\widehat{\theta}(x)-b^{(k)})-(\widehat{\theta}(x)-b^{(k)}).\tag{14}$$

The $k+1$ step updated bias corrected estimate of $\widehat{\theta}(x)$ can be denoted by

$$\tilde{\theta}^{(k+1)}(x)=\widehat{\theta}(x)-b^{(k+1)}.\tag{15}$$

Assuming that the limit of $b^{(k)}$ exists, we can let $k\to\infty$ in equation (14) to obtain

$$b=h(\tilde{\theta}(x))-(\widehat{\theta}(x)-b),\tag{16}$$

so that

$$\tilde{\theta}(x)=h^{-1}(\widehat{\theta}(x)).\tag{17}$$

Assuming that $h(\cdot)$ is one to one and differentiable, from the above expression (12) and Slutsky's theorem, we can obtain the following Theorem 3:

Theorem 3 (Iterative bias correction of AIMLE) *When sample size* $N \longrightarrow \infty$

$$\sqrt{N}(\tilde{\theta}(x) - \theta) \xrightarrow{d} N(\boldsymbol{0}, D\Lambda D^T), \tag{18}$$

where $D = \dfrac{dh^{-1}(\theta)}{d\theta}|_{\theta=\theta^\star}$.

Thus the estimator $\widetilde{\theta}(x)$ defined by equation (17) is asymptotically unbiased and consistent, where the estimator $\widetilde{\theta}(x)$ can be understood as an estimate resulting from iterative bias correction for $\widehat{\theta}(x)$.

The equations (14) and (15) are the very core of the subject about iterative bias correction of $\widehat{\theta}(x)$. So the function $h(\cdot)$ contained in equation (8), which is used in equation (14), is very important. Except for very simple problems, the function $\theta^\star = h(\theta)$ has no explicit expression, and usually is complicated integral equation. Just for the complexity of the function $h(\cdot)$, the implementation of iterative bias correction is difficult. From (12), we can find that $h(\theta) = \theta^\star$ is the asymptotic mean of $\widehat{\theta}(x)$. We propose to approximate $h(\theta)$ by $h_M(\theta)$ which is the average of $\widehat{\theta}(x)$ over simulated samples. $h_M(\theta)$ can be written as the follows:

$$h_M(\theta) = \frac{1}{M}\sum_{i=1}^{M}\widehat{\theta}(\mathbf{x}_i), \tag{19}$$

where $\mathbf{x}_1, \ldots, \mathbf{x}_M$ are simulated from the model with the parameters set at θ and φ. Replacing $h_M(\theta)$ in equations (14) and (15), we obtain

$$b_M^{(k+1)} = h_M(\widehat{\theta}(x) - b_M^{(k)}) - (\widehat{\theta}(x) - b_M^{(k)}) \tag{20}$$

as bootstrap estimate of the bias of $\widehat{\theta}(x)$ at the $k+1$th iteration and

$$\tilde{\theta}^{(k+1)}(x) = \widehat{\theta}(x) - b_M^{(k+1)} \tag{21}$$

as the updated bootstrap iterative bias correction estimate of θ.

The bias correction method for profile likelihood is also researched by many scholars, for example, Cox and Reid [4] proposed a conditional modified profile likelihood function to reduce the bias, but it needs orthogonal reparameterizations. McCullagh and Tibshirani [8] modified a biased estimating equation to result in an unbiased estimating equation. Our method is to solve the biased estimating equation, to obtain the AIMLE without modification, then to adjust the AIMLE using the iterative bias bootstrap correction method as in Kuk [7], and to result in asymptotically unbiased and consistent estimators.

3. Proof of Theorem 1

The following Lemma 1 is well known:

Lemma 1 *Let* $\{y_n\}$ *be a bounded infinite sequence in* R^k. *If every convergent subsequence of* $\{y_n\}$ *has a common limit point* $\hat{y}$, *then* $\{y_n\}$ *converges.*

Lemma 2 *If* $L(\theta, \varphi; x)$ *satisfies (A1), (A2) and (A3) of Theorem 1, then* $\|\theta^{(n+1)}(x) - \theta^{(n)}(x)\| \longrightarrow 0$ *as* $n \longrightarrow \infty$.

The proof of Lemma 2 is similar to that of Shi et al [10].

Lemma 3 *Suppose that* $\{\theta^{(n_k)}(x), \varphi^{(n_k)}(x)\}$ *is an arbitrary convergent subsequence of* $\{\theta^{(n)}(x), \varphi^{(n)}(x)\}$, *and* $(\theta^{(n_k)}(x), \varphi^{(n_k)}(x)) \longrightarrow (\theta'(x), \varphi'(x))$ *when* $k \longrightarrow \infty$, *then*

$$L(\theta'(x), \varphi'(x); x) = \max_{\theta \in C} L(\theta, \varphi'(x); x) = \max_{\varphi \in D} L(\theta'(x), \varphi; x).$$

Proof For any $\varphi \in D$, the algorithm shows that

$$L(\theta^{(n_k)}(x), \varphi^{(n_k)}(x); x) \geq L(\theta^{(n_k)}(x), \varphi; x).$$

Therefore

$$\lim_{k \to \infty} L(\theta^{(n_k)}(x), \varphi^{(n_k)}(x); x) \geq \lim_{k \to \infty} L(\theta^{(n_k)}(x), \varphi; x),$$

i.e.,

$$L(\theta'(x), \varphi'(x); x) \geq L(\theta'(x), \varphi; x).$$

Similarly, for any $\theta \in C$ we have

$$L(\theta'(x), \varphi'(x); x) \geq L(\theta, \varphi'(x); x).$$

This completes the proof of Lemma 3. □

Lemma 4 *The mapping* $\varphi(\cdot)$ *is continuous from* C *to* D.

Proof Suppose that the sequence $\theta^{(n)}(x)$ converges to $\theta_0(x) \in C$, and let $\{\varphi(\theta^{(n_k)}(x))\}$ be a convergent subsequence of $\{\varphi(\theta^{(n)}(x))\}$ with $\varphi_0(x) = \lim_{k\to\infty}\{\varphi(\theta^{(n_k)}(x))\}$. By the definition of $\varphi(\cdot)$, for any $\varphi \in D$

$$L(\theta^{(n_k)}(x), \varphi(\theta^{(n_k)})(x); x) \geq L(\theta^{(n_k)}(x), \varphi; x).$$

And by the continuity of $L(\theta, \varphi; x)$ as $k \longrightarrow \infty$, we have

$$L(\theta_0(x), \varphi_0(x); x) \geq L(\theta_0(x), \varphi; x).$$

Namely $L(\theta_0(x), \varphi_0(x); x) = \max_{\varphi \in D} L(\theta_0(x), \varphi; x)$ or $\varphi_0(x) = \varphi(\theta_0(x))$. By Lemma 1, we have

$$\lim_{n \to \infty} \{\varphi(\theta^{(n)}(x)\} = \varphi(\theta_0(x)).$$

This completes the proof of Lemma 4.

From the alternating iterative process in Section 2, the sequence $\{\theta^{(n)}(x), \varphi^{(n)}(x)\}$ obtained by alternating iterative method is the same as the sequence $\{\theta^{(n)}(x), \varphi(\theta^{(n)}(x))\}$. By Lemma 4, we have the following Corollary 1.

Corollary 1 *The sequence* $\{\theta^{(n)}(x), \varphi^{(n)}(x)\}$ *obtained by the alternating iterative method converges if and only if* $\{\theta^{(n)}(x)\}$ *converges.*

The following Lemma 5 is given in Shi and Jiang [11]:

Lemma 5 *Let* $\{y_n\}$ *be a uniformly bounded sequence in* R^k. *If* $\|y_{n+1} - y_n\| \longrightarrow 0$ *as* $n \to \infty$, *and the sequence is not convergent, then there are infinitely many accumulation points of the sequence* $\{y_n\}$.

Since C is a convex and compact sunset of R^k, and $\theta \circ \varphi(\cdot)$ is continuous from C to C, there exists a vector $\theta_0 \in C$ such that $\theta \circ \varphi(\theta_0) = \theta_0$ (see Smart [6]). We have the following lemma.

Lemma 6 *Any accumulation point of the iterative sequence* $\{\theta^{(n)}(x)\}$ *is a fixed point of the composed mapping* $\theta \circ \varphi(\cdot)$ *in* C.

Proof Let $\theta'(x)$ be an accumulation point of $\{\theta^{(n)}(x)\}$. Then there exists a subsequence $\{\theta^{(n_k)}(x)\} \subset \{\theta^{(n)}(x)\}$ satisfying $\lim_{k\to\infty} \theta^{(n_k)}(x) = \theta'(x)$. By Corollary 1, $\{\varphi(\theta^{(n_k)}(x))\}$ converges. By Lemma 3, $L(\theta'(x), \varphi(\theta'(x)); x) = \max_{\theta\in C} L(\theta, \varphi(\theta'(x)); x)$ or $\theta'(x) = \theta \circ \varphi(\theta'(x))$. This completes the proof of Lemma 6. □

Proof of Theorem 1 By Corollary 1, it suffices to prove the sequence $\{\theta^{(n)}(x)\}$ is convergent. Since $L(\theta, \varphi; x)$ satisfies conditions (A1), (A2) and (A3), we have $\|\theta^{(n+1)}(x) - \theta^{(n)}(x)\| \longrightarrow 0$ as $n \longrightarrow \infty$ by Lemma 2. If $\{\theta^{(n)}(x)\}$ is not convergent, the conditions of Lemma 5 hold. Therefore, there are infinitely many accumulation points of the sequence $\{\theta^{(n)}(x)\}$. By Lemma 6, any accumulation point of the iterative sequence $\{\theta^{(n)}(x)\}$ is a fixed point of the composed mapping $\theta \circ \varphi(\cdot)$ in C. So $\theta \circ \varphi(\cdot)$ has at most countable fixed points in C. This contradicts the condition (A4). The contradiction implied that this problem is true. This completes the proof of Theorem 1. □

4. Examples and simulation

In this section, we provide two examples and a simulation. In Example 1, the AIMLE has an explicit expression and happens to coincide with the MLE, and we derive the bias correction for the AIMLE. In Example 2, we study the estimation problem of binary panel data model, and the AIMLE has no explicit expression. Based on Example 1, we do a Monte carlo simulation.

Example 1 Given a sample $X_1, X_2, \ldots, X_n$ from $N(\mu, \sigma^2)$, where μ and σ^2 are all unknown. The log-likelihood function is $L(\mu, \sigma^2; x) = -n \ln \sigma - \sum_{i=1}^{n} (x_i - \mu)^2 / 2\sigma^2 + c$, where the log-likelihood function $L(\mu, \sigma^2; x)$ is not concave function, c is a constant. The MLE of μ and σ^2 solve the following problem

$$\max_{\mu\in R, \sigma^2\in R^+} L(\mu, \sigma^2; x). \tag{$*$}$$

From the above, its maximum occurs at $\hat{\mu} = \bar{x}$, $\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$. The proof that the estimates $\hat{\mu}$ and $\hat{\sigma^2}$ are the MLE of μ and σ^2 can be found in Rao [5]. It can be found that $\hat{\sigma^2}$ is a biased estimator. Of course, we can replace the n in $\hat{\sigma^2}$ by $n-1$ and thus correct the bias in $\hat{\sigma^2}$, called modified MLE.

The following will give the AIMLE of μ and σ^2 by the alternating iterative method where the AIMLE of μ and σ^2 is the same as the MLE. Let $v = \frac{1}{\sigma^2}$. Replacing v in $L(\mu, \sigma^2; x)$, we can obtain

$$L^\star(\mu, v; x) = \frac{n}{2} \ln v - \frac{v}{2} \sum_{i=1}^{n} (x_i - \mu)^2 + c,$$

where c is a constant. So the solution of $(*)$ is equivalent to the solution of following:

$$\max_{\mu \in R, v \in R^+} L^\star(\mu, v; x). \tag{$**$}$$

It can be found that the log-likelihood function $L^\star(\mu, v; x)$ is semi-concave function of (μ, v) defined on $R \times R^+$. Because $\mu \in R$, i.e., $k_1 = 1$, and it is found that the log-likelihood function $L^\star(\mu, v; x)$ satisfies the conditions (A1), (A2) and (A3) of Theorem 1. So the iterative sequence $\{(\mu^{(n)}(x), v^{(n)}(x))\}$ converges and the sequence $\{(\mu^{(n)}(x), {\sigma^2}^{(n)}(x)\}$ also converges, obviously we can find that $\{\mu^{(n)}(x)\}$ converges to $\bar{x}$, the MLE of μ, and $\{{\sigma^2}^{(n)}(x)\}$ converges to $\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$, the MLE of σ^2. According to the definition of the AIMLE, the AIMLE of μ and σ^2 coincides with the MLE, which is the explicit expression of x. We find that the AIMLE of σ^2 is a biased estimator. It is well known that the asymptotic distribution of the AIMLE $\hat{\sigma^2}(x)$ is $N(\frac{n-1}{n}\sigma^2, \frac{2(n-1)^2\sigma^4}{n^3})$. Let $\sigma^{2\star} = \frac{n-1}{n}\sigma^2$. Thus we have $h(x) = \frac{n-1}{n}x$, which is an explicit expression. The AIMLE of σ^2 has an asymptotic bias given by $b(\sigma^2) = \sigma^{2\star} - \sigma^2 = h(\sigma^2) - \sigma^2$. Let $b^{(0)}$ be an initial estimate of the bias of the AIMLE $\hat{\sigma^2}(x)$. The $k+1$ step updated estimate of bias of $\hat{\sigma^2}(x)$ can be obtained by equation (14) in Section 2. The $k+1$ step updated bias corrected estimate of σ^2 can also be obtained by equation (15) in Section 2. Assuming that the limit of $b^{(k)}$ exists, we can let $k \to \infty$ to obtain $\tilde{\sigma^2}(x) = \frac{n}{n-1}\hat{\sigma^2}(x)$. Because $h(x) = \frac{n-1}{n}x$ is one to one and differentiable, from the above expression and Slutsky's theorem, we can obtain the asymptotic distribution of $\tilde{\sigma^2}(x)$, where $\tilde{\sigma^2}(x)$ is the bias correction estimate for AIMLE $\hat{\sigma^2}(x)$, and the asymptotic distribution is $N(\sigma^2, \frac{2\sigma^4}{n})$. Thus the bias correction estimator $\tilde{\sigma^2}(x)$ is asymptotically unbiased and consistent for σ^2, which corrects the bias in $\hat{\sigma^2}(x)$ as before.

Example 2 Consider the estimation problem in binary panel data model with fixed T and large N. The model is

$$y_{it} = I(\alpha + \beta x_{it} + \eta_i + v_{it} \geq 0), \quad t = 1, \dots, T;\ i = 1, \dots, N, \tag{$\dagger$}$$

where $I(\cdot)$ is indicator function, η_i is fixed effect, and v_{it} is independently identical distributed with logistic distribution $\Lambda(x) = \frac{e^x}{1+e^x}$. We are interested in estimating α and β, and therefore treat η_i as parameters to be estimated (nuisance parameters). The estimation problem can be written as

$$\max_{\theta \in R, \eta \in R^N} L(\theta, \eta; x), \tag{$\dagger\dagger$}$$

where $L(\theta, \eta; x)$ is log-likelihood function of model ($\dagger$), $\theta = (\alpha, \beta)$ and $\eta = (\eta_1, \dots, \eta_N)$. We can find that the objective function of problem ($\dagger\dagger$) is semi-concave function and satisfies the assumption conditions (A1), (A2), (A3) and (A4) of Theorem 1. So the iterative sequence $\{\theta^{(n)}(x), \eta^{(n)}(x)\}$ converges. By the definition of AIMLE, we know that the AIMLE $\hat{\theta}(x)$ of θ exists, which is a biased estimator by Theorem 2. Using Theorem 3, we can adjust the AIMLE $\hat{\theta}(x)$ to the unbiased estimator $\tilde{\theta}(x)$, which is BCAIMLE. Here $h(x)$ has no explicit expression, so we obtain the AIMLE by using equations (20) and (21).

Simulation We present a small Monte Carlo study to illustrate the usefulness of the bias correction for AIMLE. We will report the results for our bias correction estimator for AIMLE (BCAIMLE) and AIMLE. Throughout, we report the mean, SD (standard deviation) of these two estimators based on 1000 replications for each design with sample size equal to 50 and 100 respectively. The results are in Table 1 below.

$N(\mu, \sigma^2)$	N	Method	Mean	SD
$N(0,1)$	50	AIMLE	0.981	0.206
		BCAIMLE	1.001	0.212
	100	AIMLE	0.986	0.140
		BCAIMLE	0.996	0.142
$N(2,4)$	50	AIMLE	3.918	0.794
		BCAIMLE	3.994	0.812
	100	AIMLE	3.939	0.574
		BCAIMLE	3.981	0.582
$N(4,16)$	50	AIMLE	15.66	3.186
		BCAIMLE	15.98	3.266
	100	AIMLE	15.80	2.157
		BCAIMLE	15.96	2.197
$N(6,36)$	50	AIMLE	35.27	6.897
		BCAIMLE	35.97	6.998
	100	AIMLE	35.59	4.725
		BCAIMLE	35.96	4.808

Table 1 Various estimators of σ^2 in normal model

For simplicity, the simulation is based on Example 1, the data for the simulation is generated by the normal model with different mean and variance. To illustrate the bias correction for AIMLE, here we assume that the function $h(x)$ has no closed form and is only defined implicitly by equation (8) which is a complicated integral equation, even though we know that $h(x)$ has an explicit expression. For each simulated data set, we compute the AIMLE $\hat{\sigma^2}(x)$ by the alternating iterative method. The bias correction for AIMLE $\tilde{\sigma^2}(x)$ is obtained from $\hat{\sigma^2}(x)$ iteratively by using equations (20) and (21) with starting value $b^{(0)} = 0$ so that $\tilde{\sigma^2}^{(0)}(x) = \hat{\sigma^2}(x)$. To save computations, we set $M = 100$.

As expected, the mean of the BCAIMLE performs better than that of the AIMLE. In general, the simulation results confirm our proposed bias correction for AIMLE.

Acknowledgments The authors would like to thank the editor and the anonymous referees for helpful suggestions and insightful comments that greatly improved the paper.

References

[1] M. AVRIEL. *Nonlinear Programming: Analysis and Methods*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1976.

[2] O. BARNDORFF-NIELSEN. *On a formula for the distribution of the maximum likelihood estimator*. Biometrika, 1983, **70**(2): 343–365.

[3] M. S. BARTLETT. *Approximate confidence intervals, III: A bias correction*. Biometrika, 1955, **42**: 201–204.

[4] D. R. COX, N. REID. *Parameter orthogonality and approximate conditional inference*. J. Roy. Statist. Soc. Ser. B, 1987, **49**(1): 1–39.

[5] C. R. RAO. Linear Statistical Inference and its Applications, 2nd Edition. John Wiley Sons, New York-London-Sydney, 1973.

[6] D. R. SMART. Fixed Point Theorems. Cambridge University Press, 1974.

[7] A. Y. C. KUK.*Asymptotically unbiased estimation in generalized linear models with random effects*. J. Roy. Statist. Soc. Ser. B, 1995, **57**(2): 395–407.

[8] P. MCCULLAGH, R. TIBSHIRANI. *A simple method for the adjustment of profile likelihoods*. J. Roy. Statist. Soc. Ser. B, 1990, **52**(2): 325–344.

[9] A. L. PERESSINI, F. E. SULLIVAN, J. J. UHL. *The Mathematics of Nonlinear Programming*. Springer-Verlag, New York, 1988.

[10] Ningzhong SHI, Guorong HU, Qing CUI. *An alternating iterative method and its application in statistical inference*. Acta Math. Sin. (Engl. Ser.), 2008, **24**(5): 843–856.

[11] Ningzhong SHI, Hua JIANG. *Maximum likelihood estimation of isotonic normal means with unknown variance*. J. Multivariate Anal., 1998, **64**(2): 183–195.

[12] J. STORE, C. WITZGALL. *Convexity and Optimization in Finite Dimension (I)*. Springer-Verlag, New York-Berlin, 1970.