Ann. Inst. Statist. Math.
Vol. 45, No. 3, 499–510 (1993)

ESTIMATING FUNCTION WITH ASYMPTOTIC BIAS AND ITS ESTIMATOR

YOSHIJI TAKAGI[1] AND NOBUO INAGAKI[2]

[1] *College of Integrated Arts and Science, University of Osaka Prefecture, Sakai, Osaka 593, Japan*
[2] *Department of Mathematical Science, Faculty of Engineering Science, Osaka University, Toyonaka, Osaka 560, Japan*

(Received September 9, 1991; revised September 11, 1992)

Abstract. We introduce the estimating function with asymptotic bias and investigate the asymptotic behavior of the estimator based on it by using their relationship. The estimator based on the estimating function with asymptotic bias has the asymptotic normality with asymptotic bias. We show that this theory has several interesting applications in practical statistics.

Key words and phrases: Estimating function, asymptotic bias, asymptotic normality, order statistics, ridge estimator function, maximum posterior likelihood estimator.

1. Introduction

Let $X_n = (X_{n1}, X_{n2}, \dots, X_{nn})$ be an observation vector for each positive integer n elements of which are independent but not necessarily identically distributed (i.n.i.d.) and let $\theta = (\theta^{(1)}, \dots, \theta^{(k)})$ be a k-dimensional unknown parameter vector and there are no nuisance parameters throughout this paper. We consider the k-dimensional estimating function $\Psi_n(\theta) = (\Psi_n^{(1)}(\theta), \dots, \Psi_n^{(k)}(\theta))$ as follows:

$$\Psi_n(\theta) = \Psi_n(X_n, \theta) = \frac{1}{n}\sum_{i=1}^{n} \psi_{ni}(X_{ni}, \theta), \tag{1.1}$$

where $\psi_{ni} = (\psi_{ni}^{(1)}, \dots, \psi_{ni}^{(k)})$ is a k-dimensional score function of observation X_{ni} and parameter θ which is noted to be dependent of n for $i = 1, \dots, n$. We call $\tilde{\theta}_n = \tilde{\theta}_n(X_n)$ the estimator of θ based on $\Psi_n(\theta)$ when

$$\Psi_n(\tilde{\theta}_n) = 0. \tag{1.2}$$

We also use the following notation:

$$\xi_n(\theta) = \sqrt{n}\Psi_n(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi_{ni}(X_{ni}, \theta). \tag{1.3}$$

Several important properties of the estimator, for example, consistency and asymptotic normality, are derived from the relationship with the estimating function (see Wilks (1962), Huber (1967) and Inagaki (1973)).

Our aim of the present paper is to introduce the estimating function with asymptotic bias and to investigate the asymptotic behavior of the estimator based on it by considering the effect of the asymptotic bias. We shall discuss several interesting applications to the practical problems in statistics. Our results are closely connected with those of Inagaki (1973), where the asymptotically unbiased estimating function is discussed.

In Section 2, notations, assumptions and preliminary lemmas are stated. In Section 3, we introduce an estimating function with asymptotic bias and show that its estimator is asymptotically biased. In Section 4, we consider the applications of the asymptotically biased estimating function to the practical problems in statistics.

2. Notations, assumptions and preliminary lemmas

In this section, we prepare the notations and assumptions, (A1)–(A3) and (A7) of which are fundamental but (A4)–(A6) are technical. These assumptions are similar to those in Inagaki (1973) and supposed throughout this paper. Assumption (A2) expresses the asymptotic bias of the estimating function.

Notations. $(\mathcal{X}, \mathcal{A}, P)$: a probability space,

Θ: a parameter space which is a subset of the k-dimensional Euclidean space R^k such that for any $M > 0$, $\Theta \cap \{\|\theta\| \leq M\}$ is closed,

θ_0: the true parameter which is unknown but fixed and exists in the interior of Θ,

$\|\cdot\|$: the maximum norm, i.e. $\|\theta\| = \max\{|\theta^{(1)}|, \ldots, |\theta^{(k)}|\}$,

$\mathcal{L}(Y)$: the distribution of a random vector Y under the probability measure P,

A': the transposed matrix of the matrix A.

ASSUMPTIONS.

(A1) $\psi_{ni}(x, \theta)$, $i = 1, \ldots, n$ are $\mathcal{A} \times \mathcal{B}$-measurable, where $\mathcal{B}$ is the σ-field of Borel subsets of Θ, and separable when considered as a process in θ.

(A2) Expected values

$$\lambda_{ni}(\theta) = (\lambda_{ni}^{(1)}(\theta), \ldots, \lambda_{ni}^{(k)}(\theta))' = E\{\psi_{ni}(X_{ni}, \theta)\} \qquad i = 1, \ldots, n,$$

exist for all $\theta \in \Theta$ and their arithmetic mean converges:

$$\lambda_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \lambda_{ni}(\theta) \to \lambda(\theta), \qquad \text{as} \quad n \to \infty, \tag{2.1}$$

where $\lambda(\theta_0) = 0$ and $\lambda(\theta) \neq 0$ if $\theta \neq \theta_0$. Furthermore, there is a k-dimensional constant vector β satisfying

$$\sqrt{n}\lambda_n(\theta_0) \to \beta. \tag{2.2}$$

(A3) For each $i = 1, \dots, n$, let $\lambda_{ni}(\theta)$ be continuously differentiable in some neighborhood of θ_0 with the differential coefficient matrix

$$\Lambda_{ni}(\theta) = \frac{\partial \lambda_{ni}(\theta)}{\partial \theta} = \left(\frac{\partial \lambda_{ni}^{(l)}(\theta)}{\partial \theta^{(m)}} \right), \qquad (l, m = 1, \dots, k) \tag{2.3}$$

respectively, and their arithmetic mean be denoted by

$$\Lambda_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \Lambda_{ni}(\theta).$$

Then, $\Lambda_n(\theta)$ converges to $\Lambda(\theta)$ uniformly in the neighborhood of θ_0, and $\Lambda_0 = \Lambda(\theta_0)$ (say) is nonsingular.

(A4) (a) There exists a positive constant λ_∞ such that

$$\underline{\lim}_{n\to\infty} \underline{\lim}_{\|\theta\|\to\infty} \|\lambda_n(\theta)\| \geq \lambda_\infty > 0. \tag{2.4}$$

(b) A positive function

$$b_n(\theta) = \max\left(\|\lambda_n(\theta)\|, \frac{\lambda_\infty}{3} \right) \tag{2.5}$$

satisfies

$$\overline{\lim}_{n\to\infty} E\left\{ \sup_{\theta} \frac{\|\psi_{ni}(X_{ni}, \theta) - \lambda_{ni}(\theta)\|}{b_n(\theta)} \right\} < \infty. \tag{2.6}$$

(c) Letting

$$W_{ni,M}(X_{ni}) = \sup_{\|\theta\|>M} \frac{\|\psi_{ni}(X_{ni}, \theta) - \lambda_{ni}(\theta)\|}{b_n(\theta)}, \tag{2.7}$$

there exists a positive number M satisfying the following conditions:

$$\overline{\lim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\{W_{ni,M}(X_{ni})\} < 1, \tag{2.8}$$

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}\{W_{ni,M}(X_{ni})\} = 0. \tag{2.9}$$

(A5) For each $i = 1, \dots, n$, $E\{\|\psi_{ni}(X_{ni}, \theta) - \lambda_{ni}(\theta)\|^2\}$ exists and

$$\frac{1}{n^2} \sum_{i=1}^{n} E\{\|\psi_{ni}(X_{ni}, \theta) - \lambda_{ni}(\theta)\|^2\} \to 0. \tag{2.10}$$

(A6) Let

$$u_{ni}(X_{ni}, \theta, d) = \sup_{\|\tau-\theta\|\le d} \|\psi_{ni}(X_{ni}, \tau) - \psi_{ni}(X_{ni}, \theta)\|. \tag{2.11}$$

For every compact set $C \subset \Theta$, there are positive numbers d_0, H_1, and $H_2 > 0$ such that

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{i=1}^{n} E u_{ni}(X_{ni}, \theta, d) < H_1 d, \tag{2.12}$$

and

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{i=1}^{n} E\{u_{ni}(X_{ni}, \theta, d)\}^2 < H_2 d, \tag{2.13}$$

the last two convergences in (2.12) and (2.13) being uniform for $d \le d_0$ and $\theta \in C$.

(A7) Denote $S_{ni} = \mathrm{Var}(\psi_{ni}(X_{ni}, \theta_0))$, the covariance matrix for each $i = 1, \dots, n$, respectively. Then, their arithmetic mean S_n converges to a positive definite matrix S:

$$S_n = \frac{1}{n} \sum_{i=1}^{n} S_{ni} \to S. \tag{2.14}$$

Furthermore, there exists $E\|\psi_{ni}(X_{ni}, \theta_0) - \lambda_{ni}(\theta_0)\|^3$, $i = 1, \dots, n$, and

$$\lim_{n\to\infty} \frac{1}{n^{3/2}} \sum_{i=1}^{n} E\|\psi_{ni}(X_{ni}, \theta_0) - \lambda_{ni}(\theta_0)\|^3 = 0. \tag{2.15}$$

The following three lemmas are straightforward extentions of the corresponding propositions due to Inagaki (1973) (essentially Huber (1967)) and thus, their proofs are omitted (see Takagi and Inagaki (1991)).

LEMMA 2.1. *If a sequence of estimators, $\{\hat\theta_n\}$, satisfies the condition:*

$$\Psi_n(\hat\theta_n) = \frac{1}{n} \sum_{i=1}^{n} \psi_{ni}(X_{ni}, \hat\theta_n) \to 0 \quad \textit{in } P \quad \textit{as } n \to \infty, \tag{2.16}$$

then it converges to θ_0 in P:

$$\hat\theta_n \to \theta_0 \quad \textit{in } P \quad \textit{as } n \to \infty.$$

LEMMA 2.2. *If $\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)]\}$ is relatively compact (see Inagaki (1973)), the following asymptotic expansion holds:*

$$\xi_n(T_n) - \xi_n(\theta_0) - \Lambda_0 \sqrt{n}(T_n - \theta_0) \to 0 \quad \textit{in } P, \quad \textit{as } n \to \infty. \tag{2.17}$$

LEMMA 2.3. *In order that $\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)]\}$ is relatively compact, it is necessary and sufficient that $\{\mathcal{L}[\xi_n(T_n)]\}$ is relatively compact.*

3. Asymptotically biased estimator

In this section, we show the asymptotic normality of the estimator $\hat{\theta}_n$ based on the estimating function $\xi_n(\theta)$. Furthermore, we construct the asymptotically equivalent estimator to $\hat{\theta}_n$ by the so-called "one-step estimator".

THEOREM 3.1. *The random vector $\xi_n(\theta_0)$ is asymptotically normal with asymptotic bias in the following sense:*

$$\mathcal{L}[\xi_n(\theta_0)] \to N_k(\beta, S) \quad \textit{in law.} \tag{3.1}$$

PROOF. From Assumption (A7), we have that

$$\begin{aligned} &\mathcal{L}[\xi(\theta_0) - \sqrt{n}\lambda_n(\theta_0)] \\ &= \mathcal{L}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\psi_{ni}(X_{ni}, \theta_0) - \lambda_{ni}(\theta_0)\}\right] \to N_k(0, S) \quad \text{in law} \end{aligned} \tag{3.2}$$

(see Loève (1978), p. 287), and thus, from Assumption (2.2) in (A2) that the result of this theorem holds. □

THEOREM 3.2. *If a sequence of estimators, $\{\hat{\theta}_n\}$, satisfies the condition:*

$$\xi_n(\hat{\theta}_n) \to 0 \quad \textit{in } P, \tag{3.3}$$

then, the estimator $\hat{\theta}_n$ is consistent and asymptotically normally distributed:

$$\mathcal{L}[\sqrt{n}(\hat{\theta}_n - \theta_0)] \to N_k[\ \beta_\Lambda,\ S_\Lambda\ ] \quad \textit{in law,} \tag{3.4}$$

where $\beta_\Lambda = -({\Lambda_0}^{-1})\beta$ and $S_\Lambda = ({\Lambda_0}^{-1})S({\Lambda_0}^{-1})'$.

PROOF. Since (3.3) implies (2.16), we have the consistency of $\hat{\theta}_n$ by Lemma 2.1. It follows from (3.3) and Lemma 2.3 that $\{\mathcal{L}[\sqrt{n}(\hat{\theta}_n - \theta_0)]\}$ is relatively compact, and therefore by Lemma 2.2 that

$$\xi_n(\hat{\theta}_n) - \xi_n(\theta_0) - \Lambda_0\sqrt{n}(\hat{\theta}_n - \theta_0) \to 0 \quad \text{in } P. \tag{3.5}$$

Thus, we have from (3.1), (3.3) and (3.5) that

$$L[\Lambda_0\sqrt{n}(\hat{\theta}_n - \theta_0)] \to N_k(-\beta, S) \quad \text{in law.}$$

Hence, the result of this theorem is proved. □

THEOREM 3.3. *If an estimator T_n satisfies the condition:*

$$\mathcal{L}[\xi_n(T_n)] \to G, \quad \textit{in law,} \tag{3.6}$$

where G is a probability distribution, then it holds that

$$\mathcal{L}[\Lambda_0\sqrt{n}(T_n - \hat{\theta}_n)] \to G, \qquad \text{in law.} \tag{3.7}$$

The converse is also true.

PROOF. Standard techniques show that the condition (3.6) or (3.7) implies the relative compactness of $\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)]\}$. Therefore, it follows by Lemma 2.2 that T_n satisfies

$$\xi_n(T_n) - \xi_n(\theta_0) - \Lambda_0\sqrt{n}(T_n - \theta_0) \to 0 \quad \text{in } P. \tag{3.8}$$

We have already in the proof of Theorem 3.2 that

$$-\xi_n(\theta_0) - \Lambda_0\sqrt{n}(\hat{\theta}_n - \theta_0) \to 0 \quad \text{in } P. \tag{3.9}$$

These lead to

$$\xi_n(T_n) - \Lambda_0\sqrt{n}(T_n - \hat{\theta}_n) \to 0 \quad \text{in } P. \tag{3.10}$$

Hence, either one of (3.6) and (3.7) derives the other. □

THEOREM 3.4. *Suppose that $\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)\}$ is relatively compact. Put*

$$T_n^* = T_n - \Lambda^{-1}(T_n)\Psi_n(T_n), \tag{3.11}$$

where $\Lambda^{-1}(T_n)$ means that θ_0 in $\Lambda^{-1}(\theta_0)$ is taken place by T_n. Then, it holds that

$$\xi_n(T_n^*) \to 0 \quad \text{in } P. \tag{3.12}$$

This implies by Theorem 3.3 that T_n^* and $\hat{\theta}_n$ are asymptotically equivalent:

$$\sqrt{n}(T_n^* - \hat{\theta}_n) \to 0 \quad \text{in } P, \tag{3.13}$$

and furthermore, by Theorem 3.2 that

$$\mathcal{L}[\sqrt{n}(T_n^* - \theta_0)] \to N_k[\,\beta_\Lambda,\ S_\Lambda\,] \quad \text{in law.} \tag{3.14}$$

PROOF. (3.11) is equivalent to

$$\sqrt{n}(T_n^* - \theta_0) = \sqrt{n}(T_n - \theta_0) - \Lambda^{-1}(T_n)\xi_n(T_n). \tag{3.15}$$

The relative compactness of $\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)]\}$ and the continuity of $\Lambda(\theta)$ imply that

$$\Lambda^{-1}(T_n) \to \Lambda^{-1}(\theta_0) = \Lambda_0{}^{-1} \quad \text{in } P, \tag{3.16}$$

and thus, from (3.15) that $\{\mathcal{L}[\sqrt{n}(T_n^* - \theta_0)\}$ is relatively compact. The last leads to

$$\xi_n(T_n^*) - \xi_n(\theta_0) - \Lambda_0\sqrt{n}(T_n^* - \theta_0) \to 0 \quad \text{in } P. \tag{3.17}$$

Of course, (3.8) holds, too. Therefore, it follows by (3.15) and (3.17) that

$$\xi_n(T_n^*) + \{(\Lambda_0\Lambda^{-1}(T_n))\xi_n(T_n) - \xi_n(\theta_0) - \Lambda_0\sqrt{n}(T_n - \theta_0)\} \to 0 \quad \text{in } P. \tag{3.18}$$

(3.8) and (3.16) implies that the part of { } in (3.18) converges to 0 in P. This and (3.18) conclude (3.12). The proof is complete. □

By Theorem 3.4, we can construct the so-called one-step estimator T_n^*, which is asymptotically equivalent to $\hat{\theta}_n$ even though it is not explicitly gotten as the solution of the estimating function.

4. Applications

4.1 *Central order statistics*

Let observations X_i, $i = 1, 2, \ldots$ be independent and identically distributed according to the distribution function F with the density function f. Let the parameter space $\Theta = R^1$. We suppose that $0 < F(\theta_0) = p < 1$, $f(x)$ is continuous at θ_0 and $f(\theta_0) > 0$. Put

$$\psi_{ni}(x,\theta) = \begin{cases} p_n = p + \dfrac{\beta}{\sqrt{n}}, & \text{if } x > \theta, \\ q_n = \dfrac{\beta}{\sqrt{n}}, & \text{if } x = \theta, \\ r_n = -(1-p) + \dfrac{\beta}{\sqrt{n}}, & \text{if } x < \theta, \end{cases}$$

which is not depend of i. Then, we have

$$\lambda_n(\theta) = E\psi_{ni}(X_i, \theta) = p - F(\theta) + \frac{\beta}{\sqrt{n}}.$$

It is easy to see that Assumptions (A1)–(A7) are satisfied, where

$$\begin{aligned} \lambda(\theta) &= p - F(\theta), \\ \Lambda(\theta) &= \Lambda_n(\theta) = -f(\theta), \\ S &= S_n(\theta_0) = p(1-p). \end{aligned}$$

Now, we denote the order statistics of $X_1, \ldots, X_n$ by $X_{n:1} \le \cdots \le X_{n:n}$. For a positive integer, ν_n, which satisfies

$$\frac{\nu_n}{n} = p + \frac{\beta}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right), \quad \text{as} \quad n \to \infty, \tag{4.1}$$

we consider an order statistic $X_{n:\nu_n}$. Then, we see that

$$\begin{aligned}\xi_n(X_{n:\nu_n}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{ni}(X_i, X_{n:\nu_n}) \\ &= \frac{1}{\sqrt{n}}\{p_n(n-\nu_n)+q_n+r_n(\nu_n-1)\} \\ &= \frac{1}{\sqrt{n}}\{1-p+o(\sqrt{n})\} \to 0 \quad \text{as} \quad n\to\infty.\end{aligned}$$

Thus, by Theorem 3.2 we have the following theorem (see Serfling (1980)).

THEOREM 4.1. *The ν_n-th order statistic satisfying* (4.1) *is asymptotically normally distributed*:

$$\mathcal{L}[\sqrt{n}(X_{n:\nu_n}-\theta_0)] \to N(\beta_\Lambda,\ S_\Lambda), \tag{4.2}$$

with

$$\beta_\Lambda = \frac{\beta}{f(\theta_0)} \qquad \textit{and} \qquad S_\Lambda = \frac{p(1-p)}{f^2(\theta_0)}. \tag{4.3}$$

4.2 *The ridge-type estimating function*

We consider the linear regression model, $GM(Y_n; X_n\theta, \sigma^2 I_n)$, where $Y_n = (Y_{n1},\dots,Y_{nn})'$ is the vector of independent observations, $X_n = \{x_{ij}\}$, $i=1,\dots,n$, $j=1,\dots,k$ the design matrix, and $\theta = (\theta_1,\dots,\theta_k)'$ the parameter vector. Then, the normal equation for the least square estimator is $X_n'(Y_n - X_n\theta) = 0$. It is well known that the least square estimator is the best linear unbiased estimator. Hoerl and Kennard (1970) propose the ridge estimator

$$T_n = (X_n'X_n + \nu_n I_k)^{-1}X_n'Y_n, \tag{4.4}$$

as the solution of the equation $X_n'(Y_n - X_n\theta) = \nu_n\theta$. Now, suppose

$$\begin{aligned} &\nu_n = \sqrt{n}\nu, \\ &\frac{1}{n}X_n'X_n \to \Gamma, \qquad \text{as} \quad n\to\infty, \end{aligned} \tag{4.5}$$

where Γ is nonsingular. The ridge estimating function is

$$\Psi_n(\theta) = \frac{1}{n}\{X_n'(Y_n - X_n\theta) - \sqrt{n}\nu\theta\}.$$

Then, we have that

$$\lambda_n(\theta) = \frac{1}{n}\{X_n'X_n(\theta_0-\theta) - \sqrt{n}\nu\theta\} \to \lambda(\theta) = \Gamma(\theta_0-\theta),$$

and thus, that $\lambda(\theta_0) = 0$, $\sqrt{n}\lambda_n(\theta_0) = -\nu\theta_0 = \beta$ (say), and

$$\Lambda_n(\theta) = -\frac{1}{n}X_n'X_n - \frac{1}{\sqrt{n}}\nu I \to \Lambda = -\Gamma.$$

Furthermore, we have

$$S_n = \sigma^2\frac{1}{n}X_n'X_n \to S = \sigma^2\Gamma.$$

THEOREM 4.2. *The ridge estimator T_n under Assumption (4.5) is asymptotically normally distributed:*

$$\mathcal{L}[\sqrt{n}(T_n - \theta_0)] \to N(\beta_\Lambda,\ S_\Lambda), \tag{4.6}$$

with

$$\beta_\Lambda = \Gamma^{-1}\beta = -\nu\Gamma^{-1}\theta_0 \quad \text{and} \quad S_\Lambda = \sigma^2\Gamma^{-1}. \tag{4.7}$$

4.3 *The maximum posterior likelihood estimator*

Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with the density function $f(x \mid \theta)$. Then, the joint density function of $X = (X_1, \ldots, X_n)$ is

$$f_n(x \mid \theta) = \prod_{i=1}^{n} f(X_i \mid \theta).$$

We deal with the Baysian problem that the parameter θ has the prior density function $\pi_n(\theta)$, which is supposed to satisfy the following conditions:

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta_0}\log \pi_n(\theta_0) = \beta, \\ &\lim_{n\to\infty} \frac{1}{n}\frac{\partial^2}{\partial\theta^2}\log \pi_n(\theta) = 0. \end{aligned} \tag{4.8}$$

We consider the estimator defined by the mode of posterior distribution,

$$p_n(\theta \mid x) = \frac{f_n(x \mid \theta)\pi_n(\theta)}{\int f_n(x \mid \theta)\pi_n(\theta)d\theta},$$

which is called the maximum posterior likelihood estimator (MPLE). In our situation, the estimating function is

$$\Psi_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i \mid \theta) + \frac{1}{n}\frac{\partial}{\partial\theta}\log \pi_n(\theta).$$

Then, under the so-called regularity conditions (for example, see Cramér (1946)), we have

$$\lambda_n(\theta) - \lambda(\theta) = \frac{1}{n}\frac{\partial}{\partial\theta}\log \pi_n(\theta) \to 0, \qquad \text{as} \quad n \to \infty,$$

$$\Lambda_n(\theta) - \Lambda(\theta) = \frac{1}{n}\frac{\partial^2}{\partial^2\theta}\log \pi_n(\theta) \to 0, \qquad \text{as} \quad n \to \infty,$$

where

$$\lambda(\theta) = E_{\theta_0}\left\{\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right\},$$
$$\Lambda(\theta) = E_{\theta_0}\left\{\frac{\partial^2}{\partial^2\theta}\log f(X \mid \theta)\right\}.$$

Furthermore, we see $\lambda(\theta_0) = 0$ and $\Lambda(\theta_0) = -I(\theta_0)$, where $I(\theta)$ is the Fisher information. Thus, we have the following theorems:

THEOREM 4.3. *The maximum posterior likelihood estimator under Assumption* (4.8) *is asymptotically normally distributed:*

$$\mathcal{L}[\sqrt{n}(T_n - \theta_0)] \to N(\beta_\Lambda,\ S_\Lambda), \tag{4.9}$$

with

$$\beta_\Lambda = I(\theta_0)^{-1}\beta \quad \textit{and} \quad S_\Lambda = I(\theta_0)^{-1}. \tag{4.10}$$

THEOREM 4.4. *For the MLE $\hat\theta_n$ satisfying the likelihood equation:*

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i \mid \hat\theta_n) = 0,$$

we have, by Theorem 3.4, *the one-step estimator T_n^*:*

$$T_n^* = \hat\theta_n - \Lambda^{-1}(\hat\theta_n)\frac{1}{n}\frac{\partial}{\partial\theta}\log \pi_n(\hat\theta_n). \tag{4.11}$$

For example, in the Bernoullis trials with the success probability θ, the least favorable prior density function is the following Beta density function:

$$\pi_n(\theta) = \frac{1}{B\left(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right)}\theta^{\sqrt{n}/2-1}(1-\theta)^{\sqrt{n}/2-1},$$

which satisfies the conditions (4.8) with

$$\beta = \frac{\frac{1}{2} - \theta_0}{\theta_0(1-\theta_0)}.$$

Since the Fisher information is $I(\theta) = 1/\theta(1-\theta)$, we obtain that the maximum posterior likelihood estimator:

$$T_n = \frac{\bar{X}_n + \frac{1}{2\sqrt{n}} - \frac{1}{n}}{1 + \frac{1}{\sqrt{n}} - \frac{2}{n}} \tag{4.12}$$

is asymptotically normally distributed in the following form:

$$L[\sqrt{n}(T_n - \theta_0)] \to N\left(\frac{1}{2} - \theta_0, \theta_0(1-\theta_0)\right). \tag{4.13}$$

On the other hand, by Theorem 4.4, we have the one-step estimator:

$$T_n^* = \left(1 - \frac{1}{\sqrt{n}} + \frac{2}{n}\right)\bar{X}_n + \frac{1}{2\sqrt{n}} - \frac{1}{n}. \tag{4.14}$$

5. Discussion

We have that the estimator with asymptotic bias is inferior to the estimator with asymptotically unbiased under mean square error of the asymptotic distribution. However, we seem that it is worth noting the justice and utility of the estimator with asymptotic bias in any practical situation where the sample size may be large but not indefinitely large.

For example, in Subsection 4.3, the maximum posterior likelihood estimator T_n has the mean square error

$$\frac{\left(\theta - \frac{1}{2}\right)^2 \left\{\left(1 - \frac{2}{\sqrt{n}}\right)^2 - 1\right\} + \frac{1}{4}}{n\left(1 + \frac{1}{\sqrt{n}} - \frac{2}{n}\right)^2}.$$

On the other hand, the MLE $\bar{X}_n$ has the mean square error $\theta(1-\theta)/n$. Now T_n is better than $\bar{X}_n$ for value of θ such that

$$\frac{\left(\theta - \frac{1}{2}\right)^2 \left\{\left(1 - \frac{2}{\sqrt{n}}\right)^2 - 1\right\} + \frac{1}{4}}{n\left(1 + \frac{1}{\sqrt{n}} - \frac{2}{n}\right)^2} < \frac{\theta(1-\theta)}{n}$$

or in the interval

$$\theta \in \left(\frac{1}{2} - a, \frac{1}{2} + a\right) \qquad \text{where}$$

$$a = \frac{1}{2}\left(\frac{\left(1 + \frac{1}{\sqrt{n}} - \frac{2}{n}\right)^2 - 1}{\left(1 + \frac{1}{\sqrt{n}} - \frac{2}{n}\right)^2 - 1 + \left(1 - \frac{2}{\sqrt{n}}\right)^2}\right)^{1/2}.$$

This interval tends to zero as $n \to \infty$. This fact corresponds to the fact that T_n has the asymptotic bias. But in our finite sample, T_n will be superior to $\bar{X}_n$ in the neighbourhood of $\theta = 1/2$. In fact, when $n = 100$, T_n is better than $\bar{X}_n$ in the interval $\theta \in (0.273, 0.727)$.

In the case of Subsection 4.2, the same fact is showed. It is well known that in general, the ridge estimator is a shrinkage estimator toward zero and has a bias but is superior to the least square estimator under the total mean square error. Therefore, in our situation, the ridge estimator of (4.4) has the asymptotic bias but will be better than $(X_n' X_n)^{-1} X_n Y_n$ in the neighbourhood of $\theta = 0$.

Acknowledgements

The authors are grateful to the referees for their valuable comments and useful suggestions.

References

Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press, Princeton.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for non-orthogonal problems, *Technometrics*, **12**, 55–67.

Huber, P. J. (1967). The behavior of maximum likelihood estimators under nonstandard conditions, *Proc. Fifth Berkelay Symp. on Math. Statist. Prob.*, Vol. 1, 221–233, Univ. of California Press, Berkley.

Inagaki, N. (1973). Asymptotic relations between the likelihood estimating function and the maximum likelihood estimator, *Ann. Inst. Statist. Math.*, **25**, 1–26.

Loève, M. (1978). *Probability Theory, I, II*, 4th. ed., Springer, New York.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Takagi, Y. and Inagaki, N. (1991). Asymptotic biased estimating function and its applications, Reseach Reports on Statistics No. 31, Osaka University.

Wilks, S. (1962). *Mathematical Statistics*, Wiley, New York.