Existence of estimators with bounded risks under an asymmetric loss function

Yoshikazu Takada

(Received February 1, 1999)

1. Introduction

Let \((X_i; \ i \geq 1)\) be independent and identically distributed (i.i.d.) random variables with the probability density function \(\sigma^{-1}\left(\frac{x - \mu}{\sigma}\right)\) with respect to Lebesgue measure, where \(f\) is known and \(\theta = (\mu, \sigma) (\sigma > 0)\) is unknown. We consider estimating \(\mu\) under the loss function

\[ L(\theta, d) = \rho(d - \mu), \]

where \(\rho(u)\) is nonincreasing for \(u < 0\), nondecreasing for \(u > 0\) and \(\rho(0) = 0\). Let

\[ M_{\min} = \min(m_-, m_+), \quad M_{\max} = \max(m_-, m_+), \]

where \(m_- = \lim_{u \to -\infty} \rho(u)\) and \(m_+ = \lim_{u \to +\infty} \rho(u)\), which may be infinity. Let \(\delta(X)\) be an estimator based on \(X = (X_1, \cdots, X_n)\). Then the purpose of this paper is to find if there exists an estimator \(\delta(X)\) such that

\[ E_{\theta}(L(\theta, \delta(X))) \leq W, \quad (1.1) \]

for all \(\theta\), where \(W(> 0)\) is a given constant.

Under a symmetric loss function Lehmann [3] showed that if the sample size \(n\) is predetermined, then no estimators meet (1.1) for \(W < M_{\min} (= M_{\max})\). In Section 2 we shall show that the nonexistence is also true for an asymmetric loss function if \(W < M_{\min}\), but there does exist such an estimator if \(M_{\min} < W < M_{\max} < \infty\). However, we can not show if the result also holds when \(M_{\max} = \infty\). In Section 3 employing two-stage procedures by Stein [6], we shall show that it is possible to construct an estimator meeting the requirement (1.1) for some distributions. Section 4 is devoted to the normal distribution and discuss the property of the proposed two-stage procedure.
2. Fixed sample size estimators

In this section we assume that the sample size \( n \) is predetermined. Then we get the following results.

**Theorem 1.** If \( 0 < W < M_{\min} \), then there do not exist any estimators whose risk is bounded by \( W \) for all \( \theta \).

**Proof.** Suppose that an estimator \( \delta \) satisfies (1.1) for \( 0 < W < M_{\min} \). Then for any \( d > 0 \)

\[
W \geq E_\delta(\rho(\delta(X) - \mu)) \\
\geq \rho(d)P_\delta(\delta(X) - \mu \geq d) + \rho(-d)P_\delta(\delta(X) - \mu \leq -d) \\
\geq \min(\rho(d), \rho(-d))P_\delta(\delta(X) - \mu \geq d).
\]

Since \( \lim_{d \to \infty} \min(\rho(d), \rho(-d)) = M_{\min} \), there exists \( d \) such that \( 0 < a = W/\min(\rho(d), \rho(-d)) < 1 \). Hence

\[
P_\delta(\delta(X) - \mu < d) > 1 - a,
\]

which contradicts the fact that there do not exist fixed-width confidence intervals of \( \mu \) when the sample size is fixed (e.g. Lehmann [3] and Takada [?]). Hence the proof is completed.

**Theorem 2.** If \( M_{\min} < W < M_{\max} < \infty \), then there exists an estimator whose risk is bounded by \( W \) for all \( \theta \).

**Proof.** Without loss of generality, let us assume \( m_- < m_+ \). Let \( \bar{\mu}(X) \) and \( \delta(X) > 0 \) be any statistics satisfying

\[
\bar{\mu}(aX + b) = a\bar{\mu}(X) + b, \\
\delta(aX + b) = a\delta(X)
\]

for all \( a(>0) \) and \( b \), where \( aX + b = (ax_1 + b, \ldots, ax_n + b) \). For example, \( \bar{\mu} \) is the sample mean and \( \delta \) the sample standard deviation. Let \( \delta(X) = \bar{\mu}(X) + \hat{\sigma}(X) \). Then
Asymmetric loss function

\[ E_s(L(\theta, \delta(X))) = E_s(\rho(\tilde{\mu}(X) + t\delta(X) - \mu)) \]
\[ = E_{\tilde{\omega}, \tilde{v}}(\rho(\tilde{\mu}(\mu + \sigma X) + t\delta(\mu + \sigma X) - \mu)) \]
\[ = E_{\tilde{\omega}, \tilde{v}}(\rho(\sigma(\tilde{\mu}(X) + t\delta(X))))) \]
\[ \leq m_+ P_{\tilde{\omega}, \tilde{v}}(\tilde{\mu}(X) + t\delta(X) \geq 0) \]
\[ + m_- P_{\tilde{\omega}, \tilde{v}}(\tilde{\mu}(X) + t\delta(X) < 0) \]
\[ = m_- + (m_+ - m_-) P_{\tilde{\omega}, \tilde{v}}(\tilde{\mu}(X) + t\delta(X) \geq 0). \]

Let \( a = (W - m_-)/(m_+ - m_-) \). Then \( 0 < a < 1 \). So there exists \( t \) such that

\[ P_{\tilde{\omega}, \tilde{v}}(\tilde{\mu}(X) + t\delta(X) \geq 0) = a. \]

Hence we get

\[ E_s(L(\theta, \delta(X))) \leq m_- + (m_+ - m_-) a \]
\[ = W, \]

which completes the proof.

In the proof we used the condition that \( M_{\text{max}} < \infty \). We do not know if Theorem 2 holds without the condition. But the next example suggests that such an estimator may exist for some problems.

**Example 1.** Let \( X_1, \ldots, X_n \) be i.i.d. according to the exponential distribution \((E(\mu, \sigma))\) with density \( \sigma^{-1} \exp(-(x - \mu)/\sigma), x > \mu \). We want to estimate \( \mu \) under the loss function

\[ \rho(u) = \begin{cases} 
1, & \text{for } u > 0 \\
-u, & \text{for } u \leq 0.
\end{cases} \]

Hence \( M_{\text{min}} = 1 \) and \( M_{\text{max}} = \infty \). Let \( \delta(X) = \min(X_1, \ldots, X_n) \). Then the risk function of \( \delta \) is one since \( \delta(X) > \mu \), so that for any \( W(1 < W < \infty) \)

\[ E_s(L(\theta, \delta(X))) \leq W. \]

### 3. Two-stage estimation procedures

In this section we shall show that utilizing a two-stage procedure by Stein [6] may enable us to construct an estimator with bounded risk in some cases when there do not exist fixed sample size estimators.

For each \( n \), let \( X_n = (X_{1n}, \ldots, X_{nn}) \). We suppose that there exist two sequences of statistics \( \{t_n(X_n); n \geq n_0\} \) and \( \{s_n(X_n); n \geq n_0\} \) \((s_n(X_n) > 0)\) which satisfy the following assumptions.
Assumption 1. For any $a(>0)$, $b$ and $n \geq n_0$

$$t_n(aX_n + b) = at_n(X_n) + b,$$
$$s_n(aX_n + b) = as_n(X_n).$$

Assumption 2. There exists a positive constant $\beta$ such that

$$P_{\pi, \nu}(n^\beta t_n(X_n) \leq x) = G(x)$$

is independent of $n$.

Assumption 3. There exists a positive integer $m(\geq n_0)$ such that $t_n(X_n)$ is independent of $s_m(X_m)$ for $n \geq m$.

The similar assumptions are used in Ghurye [2] who considered an application of Stein's two-stage procedure of testing for a location parameter of the location-scale family. See also Mukhopadhyay [5].

Let $S_m = s_m(X_m)$ and we assume that for each $z > 0$

$$h_m(z) = \int \int \rho \left( \frac{X}{y} \right) G(dx) H_m(dy) < \infty,$$

(3.1)

where $H_m(y) = P_{\pi, \nu}(S_m \leq y)$. Then $h_m(z)$ is a nondecreasing function of $z > 0$ and $h_m(0) = 0$,

$$\lim_{z \to 0} h_m(z) = m_+ \int_{x \geq 0} G(dx) + m_- \int_{x < 0} G(dx)$$

$$= am_+ + (1 - a)m_-$$

with $a = \int_{x \geq 0} G(dx)$. Hence if $W < am_+ + (1 - a)m_-$, then there exists a positive constant $z_m$ such that $h_m(z_m) = W$.

Now we shall give a two-stage estimator whose risk is bounded by $W$. Let $X_n = (X_1, \cdots, X_n)$ be the first sample and calculate $S_m$. Define the total sample size $N$ by

$$N = \max \left\{ m, \left[ \left( \frac{S_m}{z_m} \right)^{1/\nu} \right] + 1 \right\},$$

(3.2)

where $[u]$ denotes the largest integer less than $u$. If $N > m$, take the second sample $X_{m+1}, \cdots, X_N$. Then $\mu$ is estimated by $T_N = t_N(X_N)$.

Theorem 3. If Assumptions 1 to 3 are satisfied and $W < am_+ + (1 - a)m_-$, then the estimator $T_N$ meets the requirement.
Proof. It follows from Assumption 3 that

\[ E_\theta[L(\theta, T_n)] = E_\theta[\rho(T_n - \mu)] \]

\[ = \sum_{n=1}^\infty E_\theta[I_{N=n}\rho(t_n(X_n) - \mu)] \]

\[ = \sum_{n=1}^\infty P_\theta(N=n)E_\theta[\rho(t_n(X_n) - \mu)], \]

where \( I_{N=n} \) denotes the indicator function of the set \( \{N=n\} \). Then from Assumptions 1 and 2 and the definition of \( N \) we get

\[ E_\theta[L(\theta, t_n(X_n))] = \sum_{n=1}^\infty P_\theta(N=n)E_\theta[\rho(at_n(X_n))] \]

\[ = \sum_{n=1}^\infty P_\theta(N=n) \int \rho\left(\frac{\alpha x}{\nu^2}\right) G(dx) \]

\[ = E_\theta \left\{ \int \rho\left(\frac{\alpha x}{N\nu}\right) G(dx) \right\} \]

\[ \leq E_\theta \left\{ \int \rho\left(\frac{2\alpha x}{\nu S_n}\right) G(dx) \right\} \]

\[ = h_\theta(z_n) = W. \]

Hence the proof is completed.

Example 2. Let \( \{X_i, i \geq 1\} \) be i.i.d. according to the exponential distribution \( E(\mu, \sigma) \), and let \( t_n(X_n) = \min(X_1, \cdots, X_n) \) and \( s_n(X_n) = \frac{1}{n-1}(X_i - t_n(X_n)) \). Then \( t_n(X_n) \) and \( s_n(X_n) \) are independently distributed as \( E(\mu, \sigma/n) \) and \( \frac{1}{\sqrt{2(n-1)}}\sigma \chi^{2}_{n-2} \), respectively, where \( \chi^{2}_{n-2} \) denotes the chi squared distribution with \( 2n-2 \) degrees of freedom. It is easy to see that Assumptions 1 to 3 are satisfied with \( \beta = 1 \), and that (3.1) is given by

\[ h_\theta(z) = \int_0^\infty \rho\left(\frac{2u}{m-1}\right)f_{\chi^{2}_{n-1}}(u)du, \]

where \( f_{\chi^{2}_{n-1}}(u) \) is the density of \( F \) distribution with \( (2, 2(m-1)) \) degrees of freedom. So we can get a two-stage estimator \( \min(X_1, \cdots, X_n) \) whose risk is bounded by \( W(\leq m) \). The total sample size \( N \) by (3.2) is

\[ N = \max\left\{ m, \left[\frac{S_n}{z_n}\right] + 1 \right\}, \]

where \( h_\theta(z_n) = W \) and \( S_n = s_n(X_n) \).
4. Normal distribution

Let \( \{X_i, i \geq 1\} \) be i.i.d. according to the normal distribution with mean \( \mu \) and variance \( \sigma^2 \), and let \( t_0(X_0) = \frac{1}{n} \sum_{i=1}^{n} X_i \), \( s_n(X_n) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2} \). Then it is easy to see that Assumptions 1 to 3 are satisfied with \( \beta = 1/2 \), and that (3.1) is given by

\[
h_n(z) = \int_{-\infty}^{\infty} \rho(zu) g_{m-1}(u) du,
\]

where \( g_{m-1}(u) \) is the density of \( t \) distribution with \( m-1 \) degrees of freedom. So we can get a two-stage estimator \( \overline{X}_n \) whose risk is bounded by \( W(<m_++m_-)/2) \). The total sample size \( N \) by (3.2) is

\[
N = \max \left\{ m, \left\lceil \frac{U_n}{\frac{z}{2n}} \right\rceil + 1 \right\},
\]

(4.1)

where \( h_n(z_n) = W \).

Now we consider properties of the two-stage procedure (4.1). Let

\[
h(z) = \int_{-\infty}^{\infty} \rho(zu) \phi(u) du,
\]

where \( \phi \) is the density function of the standard normal distribution. If \( \sigma \) were known, then

\[E_{\sigma}(\overline{X}_n - \mu) \leq W\]

if and only if \( n \geq n^* = (\sigma/z^*)^2 \), where \( h(z^*) = W \). So \( n^* \) would be the optimal sample size if \( \sigma \) were known.

Lemma 1. \( z^* \geq z_m \).

Proof. Suppose that \( X \) and \( Y \) are random variables with the standard normal distribution and \( t \) distribution with \( \nu = m-1 \) degrees of freedom, respectively. The it follows from Theorem 4 of Ghosh [1] that

\[P(|X| > c) \leq P(|Y| > c)\]

for any \( c > 0 \). That is, \( |Y| \) is stochastically larger than \( |X| \). Since \( h(z) = E_{\rho}(zX) \) and \( h_n(z) = E_{\rho}(zY) \), it can be shown that \( h_n(z) \geq h(z) \) (e.g. Lehmann [4], p. 116), from which the theorem is proved.

From (4.1) we get
Asymmetric loss function

\[
\frac{U_m}{2^{\alpha}} \leq N \leq \frac{U_m}{2^{\alpha}} + m. \tag{4.2}
\]

Hence

\[
\left(\frac{z^*}{2^{\alpha}}\right)^2 \frac{U_m}{\sigma^2} \leq \frac{N}{n^*} \leq \left(\frac{z^*}{2^{\alpha}}\right)^2 \frac{U_m}{\sigma^2} + \frac{mz^{*2}}{\sigma^2}. \tag{4.3}
\]

From the left inequality of (4.3) and Lemma 1 we get the following result.

**Theorem 4.** For any fixed \( W \) and \( m \)

\[
\frac{E_s(N)}{n^*} \geq \left(\frac{z^*}{2^{\alpha}}\right)^2 \geq 1.
\]

Next we consider the asymptotic properties of the two-stage procedure as \( W \to 0 \). We call the two-stage procedure asymptotically efficient if

\[
\lim_{W \to 0} \frac{E_s(N)}{n^*} = 1.
\]

It follows from (4.3) that

\[
\left(\frac{z^*}{2^{\alpha}}\right)^2 \leq \frac{E_s(N)}{n^*} \leq \left(\frac{z^*}{2^{\alpha}}\right)^2 + \frac{mz^{*2}}{\sigma^2}.
\]

So in order for the two-stage procedure to be asymptotically efficient it is necessary that \( m \) must be chosen such that

\[
\lim_{W \to 0} \frac{z^*}{2^{\alpha}} = 1 \tag{4.4}
\]

and

\[
\lim_{W \to 0} mz^{*2} = 0. \tag{4.5}
\]

Then the next result is obtained.

**Theorem 5.** If the initial sample size \( m \) is chosen such that (4.4) and (4.5) are satisfied, then the two-stage procedure is asymptotically efficient.

**Example 3.** Suppose that
Yoshikazu Takada

\[ \rho(u) = \begin{cases} au^2, & \text{for } u > 0 \\ -bu, & \text{for } u \leq 0, \end{cases} \]

where \( a > 0 \) and \( b > 0 \). Then straightforward calculations show that

\[ h(z) = cz + dz^2 \]
\[ h_m(z) = c_m z + d_m z^2, \]

where \( c = \frac{a}{2}, \quad d = \frac{b}{\sqrt{2\pi}}, \quad c_m = c - \frac{\nu}{2} \) and \( d_m = d \left( \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)^{\frac{1}{2}}} \right) \) with \( \nu = m - 1 \).

Hence

\[ z^* = -\frac{d + \sqrt{d^2 + 4cw}}{2c}, \quad z_m = -\frac{d_m + \sqrt{d_m^2 + 4c_mW}}{2c_m}. \]

So

\[ \frac{z^*}{z_m} = \frac{\sqrt{d^2 + 4cw} + d_m}{\sqrt{d^2 + 4cw} + d_m} \]

and

\[ z^* = \frac{2W}{\sqrt{d^2 + 4cw} + d}. \]

Note that \( \lim_{m\to\infty} c_m = c \) and \( \lim_{m\to\infty} d_m = d \). If \( m \) is chosen such that \( m = o(W^{-2}) \) as \( W \to 0 \), then (4.4) and (4.5) are satisfied and the two-stage procedure becomes asymptotically efficient.

References


Yoshikazu Takada
Department of Mathematics
Faculty of Science
Kumamoto University
Kumamoto 860-855, Japan
E-mail: takada@sci.kumamoto-u.ac.jp