Journal of the Korean
Statistical Society
Vol. 22. No. 2. 1993

Asymptotically Admissible and Minimax Estimators of the Unknown Mean†

Andrew L. Rukhin[1] Woo-Chul Kim[2]

ABSTRACT

An asymptotic estimation problem of the unknown mean is studied under a general loss function. The proof of this result is based on the asymptotic expansion of the risk function. Also conditions for second order admissibility and minimaxity of a class of estimators depending only on the sample mean are established.

KEYWORDS : Asymptotic expansion, Admissibility, Minimaxity, Second order Bayes risk.

1. INTRODUCTION

Let $x_1, \cdots, x_n$ be a random sample from a distribution P whose mean μ is unknown. Here we assume that μ is the only unknown parameter. We study the point estimation problem of this parameter under a general loss function $W(\delta, \mu)$. Our goal is to investigate asymptotic admissibility (second order admissibility) and asymptotic minimaxity of estimators which are smooth functions of the sample mean $\bar{x} = \sum_1^n x_j/n$.

This problem has been studied mainly for the quadratic loss $W(\delta, \mu) = (\delta - \mu)^2$ or for scaled versions of this loss (see Lehmann (1983) pp 258-259 for minimaxity

[1] Department of Mathematics and Statistics, UMBC, Baltimore, MD 21228 USA.

[2] Department of Computer Science and Statistics, Seoul National University, Seoul, 151-742, Korea.

†Research of the first author was supported by NSF Grant DMS-9124009.

results and Akahira and Takeuchi (1981) for asymptotically complete class results). Asymptotic variance estimation was considered by Joshi and Rukhin (1990).

In Section 2 an asymptotic expansion of the risk function for such estimators is obtained. It turns out that under mild regularity assumptions this quantity admits an asymptotic expansion with the leading order n^{-1} term which does not depend on the estimator under consideration. The second order n^{-2} term is defined by a differential expression which can be treated as a new risk function with the corresponding definitions and conditions of admissibility and minimaxity conditions being provided in Section 3.

Similar asymptotic expansions for a loss function of the form $W(\sqrt{n}(\delta-\mu))$ were obtained by Chibisov (1973) and Levit (1982).

2. ASYMPTOTIC EXPANSION OF RISK FUNCTION

We consider here the estimators of μ having the form

$$\delta(x_1,\cdots,x_n)=\bar{x}+g(\bar{x})/n \tag{2.1}$$

where g is a smooth function. Estimators of the form (2.1) (with $\bar{x}$ playing the role of the maximum likelihood estimator) were studied by Pfanzagl and Wefelmeyer (1978) who showed that for any estimator δ_1 not necessarily of this form there exists δ such that (2.1) holds and $R(\delta_1,\mu)-R(\delta,\mu)=o(n^{-1})$. As a matter of fact results of this paper apply to a more general class of estimators of the form $\bar{x}+g(\bar{x})/n+\epsilon_n(\bar{x})$ where ϵ_n is bounded and tends to zero faster than n^{-1}.

Let X denote a random variable with distribution P so that

$$EX=\mu.$$

We make the following assumptions:

A1. Let $m_k=EX^k$ for $k=1,2,\cdots,5$. It is assumed that

$$E|X|^5<\infty.$$

A2. The loss function $W(\delta,\mu)$ is a unimodal nonnegative function of δ with a unique minimum at $\delta=\mu$. Denote $W^{(k)}(\mu)$ the k-th derivative of this function at $\delta=\mu$. It is assumed that these derivatives exist up to $k=5$ and the 5th derivative is a continuous bounded function. Thus $W(\mu)=0=W'(\mu)$ and $W^{(2)}(\mu)>0$.

A3. The function g is bounded and differentiable with continuous bounded derivative, and the risk function of (2.1) is finite.

The main object of our interest here is an asymptotic expansion of risk function

$$R(\delta, \mu) = EW(\delta, \mu)$$

where δ is of the form (2.1). In the sequel we denote the variance of x_i by σ^2 and by $\gamma = m_3/\sigma^3$ the standardized cumulant of order 3.

Theorem 1. Under Assumptions A1-A3 one has the following asymptotic expansion

$$R(\delta, \mu) = n^{-1}\sigma^2 W^{(2)}(\mu)/2 + n^{-2}[\{g^2(\mu) + 2\sigma^2 g'(\mu)\}W^{(2)}(\mu)/2 \\ +\{\sigma^2 g(\mu) + \gamma\sigma^3/3\}W^{(3)}(\mu)/2 + \sigma^4 W^{(4)}(\mu)/8] + o(n^{-2}). \quad (2.2)$$

Proof. With $z = \sqrt{n}(\bar{x} - \mu)/\sigma$, the estimator of the form (2.1) can be written as

$$\delta = \mu + n^{-1/2}\sigma z + n^{-1} g(\mu + n^{-1/2}\sigma z).$$

It follows from the assumption A2 that $W(\delta, \mu)$ can be expanded as follows;

$$W(\delta, \mu) = W^{(2)}(\mu)(\delta - \mu)^2/2 + W^{(3)}(\mu)(\delta - \mu)^3/6 \\ +W^{(4)}(\mu)(\delta - \mu)^4/24 + c_{1n}(\delta - \mu)^5, \quad (2.3)$$

with a uniformly bounded random variable c_{1n}. The assumption A3 also implies that

$$g(\mu + n^{-1/2}\sigma z) = g(\mu) + n^{-1/2}\sigma z g'(\mu) + c_{2n} n^{-1/2}\sigma z, \quad (2.4)$$

with a uniformly bounded random variable c_{2n} converging to 0 in probability. Furthermore, it is well known (see, for example, Serfling (1980) p. 68) that

$$E|Z|^k = O(1) \qquad k = 1, 2, \cdots, 5, \quad (2.5)$$

under the assumption A1.

Therefore it follows from (2.3), (2.4),(2.5) and A3 that

$$R(\delta, \mu) = n^{-1/2} W^{(2)}(\mu)\sigma^2 E(Z^2)/2 + n^{-3/2} W^{(3)}(\mu)\sigma^3 E(Z^3)/6 \\ +n^{-2}\{W^{(2)}(\mu)\sigma^2 g'(\mu)E(Z^2) + W^{(2)}(\mu)g^2(\mu)/2 \\ +W^{(3)}(\mu)\sigma^2 g(\mu)E(Z^2)/2 + W^{(4)}(\mu)\sigma^4 E(Z^4)/24\}$$

$$+o(n^{-2}),$$

which implies (2.2) by the relation between the moments and cumulants. ∎

It is worth noticing that the leading term of the risk does not depend on the estimator (2.1), and that the second order term has the form

$$(\Gamma g)(\mu) = \{g^2(\mu) + 2\sigma^2 g'(\mu)\}W^{(2)}(\mu) + \{\sigma^2 g(\mu) + \gamma\sigma^3/3\}W^{(3)}(\mu) + \sigma^4 W^{(4)}(\mu)/4. \quad (2.6)$$

3. ASYMPTOTIC ADMISSIBILITY AND MINIMAXITY

Theorem 1 shows that an estimator (2.1) is asymptotically better than

$$\delta_0 = \bar{x} + \tfrac{1}{n} g_0(\bar{x})$$

if

$$(\Gamma g)(\mu) \leq (\Gamma g_0)(\mu) \quad (3.1)$$

with strict inequality for some μ. In other terms δ_0 is asymptotically inadmissible within the class (2.1) if for some g (3.1) holds for all μ with strict inequality for some μ.

It is convenient to consider (2.6) as defining a general differential operator Γ defined on piecewise continuously twice differentiable functions g such that

$$(\Gamma g)(t) = 2h(t)g(t) + 2\ell(t)g'(t) + k(t)g^2(t) + r(t) \quad (3.2)$$

for some continuous functions h, ℓ, k and r defined on some interval with end points $\underline{t}$ and $\bar{t}$. We assume throughout this section that k and ℓ are strictly positive. In our main application to (3.1), we have

$$h = \sigma^2 W^{(3)}/2, \ell = \sigma^2 W^{(2)}, k = W^{(2)}$$

and

$$r = \gamma\sigma^3 W^{(3)}/3 + \sigma^4 W^{(4)}/4$$

so that this assumption means that $W^{(2)}(\mu) > 0$. It should be noted that σ^2 and γ may depend on μ.

Declare a function g_0 satisfying A3 to be *permissible* if the inequality

$$(\Gamma g_0)(t) \geq (\Gamma g)(t) \tag{3.3}$$

valid with some g, satisfying A3, for all values of t, implies that the inequality in (3.3) becomes the equality for all t. In other terms g_0 is permissible if it is admissible for a new risk function $(\Gamma g)(\mu)$ with the decision space consisting of all piecewise continuous twice differentiable functions of real argument t taking values in the interval $I = (\underline{t}, \bar{t})$.

We obtain now a necessary and sufficient condition of permissibility particular cases of which were derived by Stein in his unpublished lecture notes on statistical decision theory at Stanford University and by Ghosh and Sinha (1981) who were mainly interested in the case $g_0 = 0$.

Theorem 2. Assume that the differential operator Γ is given by (3.2) with positive functions k and ℓ. A function g_0 is permissible if and only if for any (or some) inner point s of interval I with end points $\underline{t}$ and $\bar{t}$

$$\int_s^{\bar{t}} \exp\Big[-\int_s^t [k(y)g_0(y) + h(y)]\ell^{-1}(y)dy\Big] k(t)/\ell(t)\; dt = \infty$$

and

$$\int_{\underline{t}}^s \exp\Big[\int_t^s [k(y)g_0(y) + h(y)]\ell^{-1}(y)dy\Big] k(t)/\ell(t)\; dt = \infty.$$

Proof. Notice first of all that if operator Γ is defined by (3.2) then the inequality (3.1) can be rewritten in the following form

$$f^2(t) + 2\Delta(t)f'(t) \leq f_0^2(t) + 2\Delta(t)f_0'(t) \tag{3.4}$$

where $f(t) = g(t) + h(t)/k(t), f_0(t) = g_0(t) + h(t)/k(t)$ and $\Delta(t) = \ell(t)/k(t)$.

Putting

$$f(t) = f_0(t) + \phi(t), \quad \text{and} \quad \rho(t) = \phi^{-1}(t),$$

we have the following inequality equivalent to (3.4)

$$\Delta^{-1}(t) + 2f_0(t)\Delta^{-1}(t)\rho(t) - 2\rho'(t) \leq 0. \tag{3.5}$$

Now suppose that the first integral in the condition of Theorem 2 converges, i.e.,

$$\int_s^{\bar{t}} \Delta^{-1}(t)\exp\Big\{-\int_s^t f_0(y)\Delta^{-1}(y)dy\Big\}dt = K < \infty.$$

Then, equating the left hand side of (3.5) to $-\Delta^{-1}(t)$, one obtains a linear differential equation

$$\rho'(t) = f_0(t)\Delta^{-1}(t)\rho(t) + \Delta^{-1}(t),$$

solution of which provides a strictly negative solution to (3.5). Solving this linear differential equation in ρ one obtains

$$\rho(t) = \exp\{\int_s^t f_0(y)\Delta^{-1}(y)dy\}[\rho(s) + \int_s^t \Delta^{-1}(y)\exp\{-\int_s^y f_0(u)\Delta^{-1}(u)du\}dy].$$

Choosing now $\rho(s)$ so that $\rho(s) < -K$ we see that $\rho(t)$ is negative for all t. If the second integral in the conditions of Theorem 2 converges one obtains

$$\rho(t) = \exp\{\int_s^t f_0(y)\Delta^{-1}(y)dy\}[\rho(s) - \int_t^s \Delta^{-1}(y)\exp\{\int_y^s f_0(u)\Delta^{-1}(u)du\}dy],$$

which provides a solution to (3.5) and can be taken to be positive.

Next, let us prove that if inequality (3.5) has a nontrivial solution then one of the conditions of Theorem 2 must fail. Assume first that $\phi(s) < 0$. Define

$$t_1 = \min(\inf\{t : \phi(t) \geq 0\}, \bar{t}).$$

Then $\phi(t)$ is negative if $s < t < t_1$ and one can proceed as before to derive the following inequality with a nonnegative function z, which in fact is $-\Delta$ times the left hand side of (3.5),

$$\begin{aligned} 0 &> \rho(t)\exp\{-\int_s^t f_0(y)\Delta^{-1}(y)dy\} \\ &= \rho(s) + \int_s^t (1 + z(y))\Delta^{-1}(y)\exp\{-\int_s^y f_0(u)\Delta^{-1}(u)du\}dy/2 \\ &\geq \rho(s) + \int_s^t \Delta^{-1}(y)\exp\{-\int_s^y f_0(u)\Delta^{-1}(u)du\}dy/2. \end{aligned} \tag{3.6}$$

If $t_1 = \bar{t}$ this completes the proof. If $t_1 < \bar{t}$ continuity of ϕ implies that

$$\lim_{t \to t_1} \rho(t) = -\infty$$

which contradicts (3.6). The situation when $\phi(s) > 0$ is treated similarly. ∎

It is instructive to contrast Theorem 2 with the Stein (1955) condition according to which any admissible procedure must be approximable by Bayes rules. What is the form of a Bayes procedure if the risk is defined by differential operator Γ ? Let $\lambda(t)$ be a smooth prior density with compact support which vanishes at the end-points of I. Integration by parts establishes the following formula

$$\int (\Gamma g)(t)\lambda(t)dt = \int \{-2(\ell(t)\lambda(t))'g(t) + \lambda(t)[k(t)g^2(t) + 2h(t)g(t) + r(t)]\}dt,$$

which can be regarded as the second order Bayes risk by (2.6). Therefore the rule $g = g_\lambda$ which minimizes this second order Bayes risk has the form

$$g_\lambda = \frac{(\ell(t)\lambda(t))' - h(t)\lambda(t)}{k(t)\lambda(t)}. \tag{3.7}$$

The resulting estimator (2.1) with $g_\lambda(\bar{x})$ can be regarded as the second order Bayes rule. Limits of functions of form (3.7) have the same form with λ being a positive function. Theorem 2 can be interpreted as follows: function g_0 of form (3.7) is permissible if and only if

$$\int_s^{\bar{t}} \frac{k(t)}{\ell^2(t)\lambda(t)}dt = \int_{\underline{t}}^s \frac{k(t)}{\ell^2(t)\lambda(t)}dt = \infty. \tag{3.8}$$

Indeed it follows from (3.7) that with some constant C

$$\lambda(t) = C\exp\Big\{\int_s^t [k(y)g_0(y) + h(y)]\ell^{-1}(y)dy\Big\}\ell^{-1}(t).$$

Note that condition (3.8) resembles the Zidek's (1970) admissibility condition in the classical estimation problem.

For example the traditional estimator $\bar{x}$ of the unknown normal mean under loss function W is asymptotically admissible if and only if

$$\int_s^{\bar{t}} (W^{(2)}(t))^{-1/2}dt = \int_{\underline{t}}^s (W^{(2)}(t))^{-1/2}dt = \infty. \tag{3.9}$$

Here we used the facts that in this problem

$$\ell(t) = \sigma^2 W^{(2)}(t) = \sigma^2 k(t)$$

and

$$h(t) = \ell'(t)/2$$

so that this result follows from Theorem 2 as permissibility condition for $g_0 \equiv 0$.

Brown (1988) obtains a representation which generalizes (3.7) to operator Γ defined on vector functions $g(t), t \in R^p$

$$\Gamma g = 2\mathcal{D}g + f \cdot Bg.$$

Here $\mathcal{D}$ is the first order linear differential operator

$$\mathcal{D}g = \sum a_i g_i + \sum a_{ij} \frac{\partial}{\partial t_j} g_i$$

with continuously differentiable functions a_i and a_{ij}, and a positive symmetric definite matrix valued function B. The argument as above shows that any permissible g has the form similar to (3.7).

Brown also proved that a function g is permissible in this situation if and only if there is a sequence $\lambda_k, k = 1, 2, \ldots$ of nonnegative functions such that

$$\begin{aligned}
&\lambda_k = 1 \qquad \|t\| \leq 1, \\
&\lambda_k = 0 \qquad \|t\| \geq k, \\
&r_k = \int (A(\nabla\lambda_k)) \cdot B^{-1} A \nabla \lambda_k f \, dt \to 0 \text{ as } k \to \infty,
\end{aligned}$$

where A is the matrix formed by function a_{ij}. The quantity r_k here plays the same role as the difference of posterior risks in the classical admissibility proofs (cf for instance Stein, 1959).

Assume now that $W^{(2)}(\mu)\sigma^2 = const$, i.e. the leading term in (2.2) is parameter free. Then the estimator δ_o is asymptotically minimax (second order minimax) if

$$\sup_t (\Gamma g_0)(t) = \inf_g \sup_t (\Gamma g)(t). \tag{3.10}$$

The next result is an analogue of the classical minimaxity condition (see Lehmann, 1983, p 256) for a function g_0.

Theorem 3. Suppose that $\lambda_m, m = 1, 2, \ldots$ is a sequence of differentiable prior densities which vanish at the end-points of interval I and such that

$$\sup_t (\Gamma g_0)(t) \leq \limsup \int_I \lambda_m(t) \Big[f(t) - \frac{[\ell(t)\lambda'_m(t)/\lambda_m(t) + \ell'(t) - h(t)]^2}{k(t)} \Big] dt.$$

Then g_0 is minimax in sense (3.10).

Proof. For any g from the domain of Γ and any fixed k because of (3.7)

$$\begin{aligned}
\sup_t (\Gamma g)(t) &\geq \int_I (\Gamma g)(t) \lambda_m(t) dt \geq \int_I (\Gamma g_{\lambda_m})(t) \lambda_m(t) dt \\
&= \int_I \lambda_m(t) \Big[f(t) - \frac{[\ell(t)\lambda'_m(t)/\lambda_m(t) + \ell'(t) - h(t)]^2}{k(t)} \Big] dt.
\end{aligned}$$

Hence

$$\sup_t (\Gamma g_0)(t) \leq \sup_t (\Gamma g)(t)$$

and g_0 is minimax. ∎

Returning to the normal example observe that estimator $\bar{x}$ is asymptotically minimax if $I = R^1, W^{(2)}(\mu)\sigma^2 = w, W^{(3)}(\mu) = 0$ and $W^{(4)}(\mu) = const.$ Indeed this fact follows from Theorem 3 with λ_m being normal densities with variance m. The example of linex loss function

$$W(\delta, \mu) = \exp\{\alpha(\delta - \mu)\} - \alpha(\delta - \mu) - 1$$

with $\alpha \neq 0$ shows that if for instance $W^{(3)}(\mu) \neq 0$ then estimator $\bar{x}$ may not be minimax for any sample size (see Rojo, 1987 and Sadooghi-Alvendi and Rematollah, 1989). This estimator is asymptotically admissible if $I = R^1$ because of (3.9).

REFERENCES

(1) M. Akahira and K. Takeuchi(1981). *Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency.* Lecture Notes in Statistics, 7, Springer, New York.

(2) L.D. Brown(1988). The differential inequality of a statistical estimation problem. In S. S. Gupta and J. O. Berger, editors, *Statistical Decision Theory and Related Topics IV.* Springer, NY, pp. 299-324.

(3) D.M. Chibisov(1973). An asymptotic expansion for a class of estimators containing maximum likelihood estimators. *Theory of Probability and Its Applications*, 18, 295-303.

(4) J.K. Ghosh and B.K. Sinha(1981). A necessary and sufficient condition for second order admissibility with applications to Berkson's bioassay problem. *Annals of Statistics*, 9, 1334-1338.

(5) S. Joshi and A. L. Rukhin(1990). Asymptotic estimation of variance. *Theory of Probability and Its Applications*, 35, 777-784.

(6) E.L. Lehmann(1983). *Theory of Point Estimation.* Wiely, New York.

(7) B. Levit(1982). Minimax estimation and positive solutions of elliptic equations. *Theory of Probability and Its Applications*, 27, 525-546.

(8) J. Pfanzagl and W. Wefelmeyer(1978). A third-order optimum property of the maximum likelihood estimator. *Journal of Multivariate Analysis*, 8, 1-29.

(9) J. Rojo(1987). On the admissibility of $c\bar{x} + d$ with respect to the linex loss function. *Communications in Statistics, Part A-Theory and Methods*, 16, 3745-3748.

(10) S.M. Sadooghi-Alvendi and R. Rematollahi(1989). A note on the admissibility of $c\bar{x} + d$ relative to the linex loss function. *Communications in Statistics, Part A-Theory and Methods*, 18, 1871-1873.

(11) R.J. Serfling(1980). *Approximation Theorems of Mathematical Statistics.* John Wiely & Sons, New York.

(12) C. Stein(1955). A necessary conditions for admissibility. *Annals of Mathematical Statistics*, 26, 518-522.

(13) C. Stein(1959). The admissibility of Pitman's estimator of a single location parameter. *Annals of Mathematical Statistics*, 30, 970-979.

(14) J.V. Zidek(1970). Sufficient conditions for the admissibility under squared error loss of formal Bayes estimators. *Annals of Mathematical Statistics*, 41, 446-456.