数理解析研究所講究録
第 538 巻 1984 年 1-23

A Survey of Non-Regular Estimation, I

Kei Takeuchi(*) and Masafumi Akahira(**)

(*) Faculty of Economics, University of Tokyo, Hongo, Bunkyo-ku
Tokyo 113, Japan

(**) Department of Mathematics, University of Electro-Communications,
Chofu, Tokyo 182, Japan

1. Introduction.

"Non-regular estimation" literally means the theory of statistical estimation when some or other of the regularity conditions for the "usual" theory fail to hold. The concept itself is purely negative and it seems to be almost self-contradiction to try to establish a "general theory" of non-regular estimation. In the sample and large sample theories of estimation of real parameters, however, there are well established sets of regularity conditions, and it is worth while to examine what may follow if any one of these regularity conditions fails to hold. And there has been accumulated substantial amount of results obtained by rather many authors, though somewhat sporadic investigations, which can give some insight into the structure of non-regular estimation and can clarify the "meaning" of each of the regularity conditions by showing which part of the theorem fails to hold and it must be modified if it is not satisfied. The purpose of this paper is to review those results from some unifying viewpoint and also to point out some problems yet to be solved. Our main interest is, therefore, not to look for some strange looking "pathological" examples, but rather to contribute to the main stream of the theory of statistical estimation by clarifying the "regular" theory from the reverse side.

First we shall consider the set of regularity conditions of statistical estimation theory. In the most general abstract terms, the problem of statistical estimation can be formulated as follows. Let $(\mathcal{X}, \mathcal{A})$ be a sample space and let $\{P_\omega | \omega \in \Omega\}$ be a set of probability measures over $(\mathcal{X}, \mathcal{A})$. The index set Ω can be any abstract set. Let $\theta=\theta(\omega)$ be a real p-dimensional vector valued function of ω and is called the parameter. An estimator $\hat{\theta} = \hat{\theta}(x)$ is a measurable function from $\mathcal{X}$ into Euclidean p-space R^p. An estimator which always comes close to $\theta(\omega)$ when a random variable X is distributed according to $P\omega$, $\omega \in \Omega$, is considered to be a "good" estimator. The main set of regularity conditions usually considered as follows :

(A.1). The set of probability measures is dominated by a σ-finite measure μ over $(\mathcal{X},\mathcal{A})$, and the density function $dP\omega/d\mu = f(x,\omega)$ $(\omega \in \Omega)$ is defined.

(A.2). For each $\omega \in \Omega$, the support of $P\omega$, i.e. the set $\{x \mid f(x,\omega)>0\}$ can be defined independently of ω.

(A.3). The index space Ω itself is an open subset of R^q, where $q \geq p$.

(A.4). For almost all $x\ [\mu]$, $f(x,\omega)$ is continuous in ω.

(A.5). For almost all $x\ [\mu]$, $f(x,\omega)$ is (k-times) continuously differentiable with respect to ω.

(A.6). The Fisher information matrix

$$I_\omega = E_\omega\left[\frac{\partial}{\partial\omega} f(X,\omega) \frac{\partial}{\partial\omega'} f(x,\omega)\right], \quad \omega \in \Omega$$

is well defined and finite.

For large sample theory we have to consider a sequence of sample spaces $(\mathcal{X}_{(n)}, \mathcal{A}_{(n)})$ $(n=1,2,...)$ and that of probability measures $P_\omega^{(n)}$ $(\omega \in \Omega)$ on $(\mathcal{X}_{(n)}, \mathcal{A}_{(n)})$ $(n=1,2,...)$ with the common index space Ω, and additional regularity conditions are :

(L1) For each n, $(\mathcal{X}_{(n)}, \mathcal{A}_{(n)})$ is subset of $(\mathcal{X}_{(n+1)}, \mathcal{A}_{(n+1)})$ and $P_\omega^{(n)}$ $(\omega \in \Omega)$ is the marginal probability measure over $(\mathcal{X}_{(n)}, \mathcal{A}_{(n)})$ derived from the probability measure $P_\omega^{(n+1)}$ $(\omega \in \Omega)$ over $(\mathcal{X}_{(n+1)}, \mathcal{A}_{(n+1)})$. (Sequence of observations).

(L2) For each n $(\mathcal{X}_{(n)}, \mathcal{A}_{(n)})$ is the n-fold direct product space of sample spaces $(\mathcal{X}_i, \mathcal{A}_i)$ $(i=1, \ldots, n)$ and $P_\omega^{(n)}$ $(\omega \in \Omega)$ is the corresponding product measure $P_\omega^{(n)}$ $(\omega \in \Omega)$ on $(\mathcal{X}_{(n)}, \mathcal{A}_{(n)})$ of $P_\omega^i (\omega \in \Omega)$ on $(\mathcal{X}_i, \mathcal{A}_i)$ $(i=1, \ldots, n)$. (Independence).

(L3) For each $i=1, \ldots, n$, $\mathcal{X}_i=\mathcal{X}$, $\mathcal{A}_i=\mathcal{A}$ and $P_\omega^i=P_\omega^1$ $(\omega \in \Omega)$. (Identical distribution).

Another set of regularity conditions are :

(S1) The probability measures $P_\omega (\omega \in \Omega)$ admit a finite dimensional sufficient statistic $T=t(X)$.

(S2) The sufficient statistic is complete.

(S3) The probability measures form an exponential family, i.e., $f(x,\omega)$ can be expressed as

$$f(x,\omega)=c(\omega)h(x) \exp \{s(\omega)'t(x)\} ,$$

where $s(\omega)$ and $t(x)$ are p-dimensional real vectors.

There are still some other minor conditions for various theorems, but the most commonly discussed are included in the above.

Next we consider the theorems of estimation. For small sample situation, most of the theory deals with unbiased estimation, and main theorems deal with a) the existence of locally best unbiased estimators and b) the existence of uniformly minimum variance unbiased (UMVU) estimators. Also in small sample situation we may consider other types of "unbiasedness" condition than the usual expectation-unbiasedness, and other types of dispersion criteria than the variance. For large sample theory, the main results are concerned with a) the existence of consistency, b) the maximum order of consistency, c) the asymptotic efficiency of estimators and d) higher order asymptotic efficiency of estimators (See Akahira and Takeuchi [1]).

In the subsequent sections we consider various combinations of problems and situations.

2. Unbiased estimation .

2.1. Undominated case.

The most extremely non-regular case would be the one when the condition (A1) is not satisfied, i.e., the probability measures are not dominated. There exists, however, rather simple examples of the undominated case.

Example 2.1.1. Let $\Omega=(\omega_1,...,\omega_N)$ be a collection of N real values.

The sample space $\mathcal{X}$ consists of the pair J and Y, where J is a set of integers $(I_1, \ldots, I_n)$ with $1 \leq I_i \leq N$ $(i=1, \ldots, n)$ and Y is a set of n real values $(Y_1, \ldots, Y_n)$. And the probability measure is defined as

$$P\{(I_1, \ldots, I_n) = (i_1, \ldots, i_n)\} = p(i_1, \ldots, i_n)$$

and $Y_i = \omega_{I_i}$ $(i=1, \ldots, n)$ with probability 1 and it is assumed that $p(i_1, \ldots, i_n)$ is independent of ω . This is nearly most general formulation of the problem of survey sampling when Ω denotes the set of the values of a "finite population" and Y the set of sample values, and P the "sampling scheme". The problem is to estimate some "population" parameter $\theta=\theta(\omega_1, \ldots, \omega_N)$. In general case the sample size n may be random, and I_i's need not all be distinct. (The case of sampling with replacement.) If we assume that ω_i can be any element of an open subset of R^1, the distribution is not dominated. Then the following is known.

<u>Theorem 2.1.1.</u> ([9]). For parameter $\theta=\theta(\omega)$ with unbiased estimators, a locally minimum variance unbiased estimator at specified $\omega=\omega_0$ has zero variance at $\omega=\omega_0$. Hence there never exists an UMVU estimator unless there is one which is always of zero variance.

<u>Example 2.1.2.</u> Let $X_1, \ldots, X_n$ be independently and identically distributed (i.i.d.) and for each $i=1, \ldots, n$, $X_i - \alpha$ $(0 \leq \alpha < 1)$ is distributed according to the Poisson distribution with parameter λ , where n is fixed. The index is the pair of non-negative real constants α and λ. The class of distributions is clearly undominated, but $([\bar{X}], \bar{X} - [\bar{X}])$ is obviously sufficient and complete, where [s] denotes the largest integer less than or equal to s, and UMVU estimator exists for any estimable parameter.

Example 2.1.3 Let $X_1, \ldots, X_n$ be i.i.d. random variables. The class of possible distributions P of X_1 is the set of the all discrete distributions over the real line. Obviously the class is not dominated. Now let $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ be the order statistic obtained from $X_1, \ldots, X_n$. We denote $(X_{(1)}, \ldots, X_{(n)})$ by Y. Y is sufficient, and we shall prove that it is complete.

Assume that

$E[\phi(Y)] = 0$ for all discrete distributions P.

Now take any possible value of Y $y=(y_1, \ldots, y_n)$, $y_1 \leq y_2 \leq \ldots \leq y_n$ of which there are k distinct values $z_1 < \ldots < z_k$ and the numbers of y_i values equal to z_j be n_j, ($\sum_{j=1}^{k} n_j = n$). Now consider the class M_z of discrete distributions with the support $(z_1, \ldots, z_k)$ and $P(X=z_j)=p_j$ $(j=1, \ldots, k)$, $p_j \geq 0$ $(j=1, \ldots, k)$ $\sum_{j=1}^{k} p_j = 1$. Then from completeness of multinomial distributions we have $E[\phi(Y)] = 0$ for all $p \in M_z$ implies $\phi(y) \equiv 0$ over the set of values of Y where all y_i's are equal to one of z_j's, and this implies that $\phi(y)=0$. Since y can be taken arbitrarily, we have $\phi(y)=0$ for all y, which completes the proof.

Consequently from the case any parameter with unbiased estimators admits a UMVU estimator.

The above examples show the undominated cases where some uniform results for the existence of UMVU estimators can be established.

There is also a simple but rather strange example of the undominated case shown below.

Example 2.1.4. Suppose that for every $n \geq 2$, $X_1, \ldots, X_n$ are i.i.d. random variables according to the probability distribution with the probability mass equal to 1/2 concentrated at the point θ, and the rest uniformly distributed over the interval (0,1). The parameter θ is assumed to be unknown in the interval (0,1). The class of probability distributions is undominated, but we can construct an unbiased estimator by the following : Let θ_0 be any constant in the interval (0,1). When two or more of X_i's have identical values, let the value by X^* and

$$\hat{\theta}_n = \frac{1}{1-(n+1)s^{-n}} (X^* - \theta_0) + \theta_0 ,$$

and $\hat{\theta}_n = \theta_0$ otherwise.

Then it is obvious that $V_{\theta_0}(\hat{\theta}_n)=0$. Hence the variance of the locally minimum variance unbiased (LMVU) estimator is equal to zero.

This example may be considered to be the limiting case of the regular situation with the following density $f(x,\theta)$ as ε tends to zero :

$$f(x,\theta) = \begin{cases} \frac{1}{2} & , |x-\theta|>\varepsilon ; \\ \frac{1}{2} + \frac{1}{\varepsilon} |x-\theta| , & |x-\theta| \le \varepsilon , \end{cases}$$

where $0<\varepsilon<\theta < 1-\varepsilon$.

Then the amount I_ε of the information is given by

$$\begin{aligned} I_\varepsilon &= \int_{-\infty}^{\infty} \{ \frac{\partial}{\partial\theta} \log f(x,\theta) \}^2 f(x,\theta) dx \\ &= \int_{\theta-\varepsilon}^{\theta+\varepsilon} \frac{1}{\varepsilon^2 f(x,\theta)} dx \\ &= \frac{4}{\varepsilon} \int_0^\varepsilon \frac{1}{2x+\varepsilon} dx = \frac{2\log 3}{\varepsilon} . \end{aligned}$$

Hence I_ε tends to infinity as $\varepsilon \to 0$, which implies that the lower bound of the variance of unbiased estimators goes to zero.

2.2. The support depending on the parameter.

In the case when the support of the distribution depends on the parameter, we often encounter a situation where the variance of locally minimum variance unbiased (LMVU) estimator is zero.

In other cases we have the infimum of the variance of unbiased estimator is zero, but there does not exist one with zero variance.

<u>Example 2.2.1.</u> Let X be a random variable with a density function $f(x,\theta)$, where θ is a real-valued parameter. We consider the location parameter case, i.e. $f(x,\theta)=f(x-\theta)$.

Assume that

$$f(x) > 0 \quad \text{for} \quad a < x < b \;;$$

$$f(x) = 0 \quad \text{for} \quad x \leq a, \quad x \geq b \;,$$

and further that $f(x)$ is continuously differentiable in the open interval (a,b) and

$$\lim_{x \to a+0} f(x) = A_0 \;, \; \lim_{x \to b-0} f(x) = B_0$$

exist, including the case where either or both A_0 and B_0 are infinity.

Then it is shown that

$$\min_{\hat{\theta}:\ \text{unbiased}} V_{\theta_0}(\hat{\theta}(x)) = 0$$

for any specified value of θ_0 ,

Here an estimator $\hat{\theta}_0$ with zero variance at $\theta = \theta_0$ can be constructed as follows.

Let $\hat{\theta}_0(x) = \theta_0$ for $\theta_0 + a < x < \theta_0 + b$. And for $\theta_0 + b < x \leq \theta_0 + 2b - a$, $\hat{\theta}_0$ is determined by the equation

$$(2.2.1) \qquad \int_{\theta_0+b}^{\theta+b} \hat{\theta}_0(x)dx = \theta - \theta_0 \int_{\theta}^{\theta_0+b} \hat{\theta}_0(x)dx$$

which is obtained from the condition

$$E_{\theta}[\hat{\theta}_0(X)] = \theta \quad \text{for} \quad \theta_0 < \theta \leq \theta_0 + (b-a) \;.$$

It is easily seen that the equation (2.2.1) has a solution for $\hat{\theta}_0$. Repeating the similar process, we can get $\hat{\theta}_0(x)$ for all values of x.

Example 2.2.2 Let X be a random variable according to a triangular distribution with a density function

$$f(x,\theta) = \begin{cases} 1-|x-\theta| & \text{for } |x-\theta|<1 \ ; \\ 0 & \text{for } |x-\theta|\geq 1 \ . \end{cases}$$

Then it is shown that

$$\inf_{\hat{\theta}\,:\,\text{unbiased}} V_{\theta_0}(\hat{\theta}(x)) = 0 \ ,$$

but the lower bound is not attained.

Fix $\theta_0 = 0$. Let us define an estimator of the type

$$\hat{\theta}_\varepsilon(x) = \begin{cases} \frac{1}{f(x,0)} = \frac{c}{1-x} & \text{for } 0<x<1-\varepsilon \ ; \\ -\frac{1}{f(x,0)} = -\frac{c}{1-x} & \text{for } 1+\varepsilon<x<1 \ ; \\ 0 & \text{for } 1-\varepsilon<x<1,\ -1<x<-1+\varepsilon \ , \end{cases}$$

where $0<\varepsilon<1$.

By the similar process as with $\hat{\theta}_0$ in Example 2.2.1 we can so construct $\hat{\theta}_\varepsilon(x)$ for all values of x that it is unbiased.

And we can prove that

$$V_{\theta=0}(\hat{\theta}_\varepsilon) \to 0 \quad \text{as} \quad \varepsilon \to 0$$

which implies that $\inf_{\hat{\theta}\,:\,\text{unbiased}} V_0(\hat{\theta})=0$. Note that here we can not let $\varepsilon=0$, because we can not construct an unbiased estimator outside the interval (-1,1) of x,

Morimoto and Sibuya [7] discussed the estimation of a so-called selection parameter. Let $X_1, \ldots, X_n$ be i.i.d. random variables with the density function(with respect to a non-atomic measure μ) of the type

$$f(x,\theta) = \begin{cases} c(\theta)f(x) & \text{for } x\in A_\theta \ ; \\ 0 & \text{for } x\notin A_\theta \ , \end{cases}$$

where $f(x)$ is a function independent of the real parameter θ.
Let us define $T_1=t_1(x_1,\dots,x_n)$ and $T_2=t_2(x_1,\dots,x_n)$ by

$$T_1=\inf \{\theta \mid (x_1,\dots,x_n) \in A_\theta \}$$

and

$$T_2=\sup \{\theta \mid (x_1,\dots,x_n) \in A_\theta \} .$$

Obviously, the pair $(T_1,T_2)(T_1 \le T_2)$ is sufficient for θ (with proper measurability condition etc.). Then assuming that the distribution of (T_1, T_2) is absolutely continuous w.r.t. the Lobesgue measure, the density function is given as

$$f^*(t_1,t_2,\theta) = \begin{cases} c(\theta)\, g(t_1, t_2), & a(\theta) < t_1 \le t_2 < b(\theta) \quad ; \\ 0 & , \text{ otherwise} . \end{cases}$$

Generally, (T_1, T_2) is complete if $\theta \ge c$ with $a(c) < b(c)$ and $a(\theta)$ is monotone decreasing and $b(\theta)$ is monotone increasing, and any estimable function $\eta(\theta)$ admits an UMV estimator. And if both $a(\theta)$ and $b(\theta)$ is monotone increasing, (T_1, T_2) is not complete, and under some regularity conditions there exists an unbiased estimator $\hat{\theta}_0$ of θ which has zero variance at any specified value θ_0.

<u>Example 2.2.3</u> ([7]). Let $X_1, \dots X_n$ be i.i.d. random variables with a density function

$$f(x,\theta) = \begin{cases} f(x)/F(\theta) & , \quad 0 \le x \le \theta, \\ 0 & , \quad \theta < x , \end{cases}$$

where $f(x) > 0$, a.e. and $F(\theta)=\int_0^\theta f(x)dx < \infty$.

We consider the estimation of the selection parameter θ of the family $\{ f(x,\theta) : 0 < \theta < \infty \}$. A minimal sufficient statistic for θ is $T=\max (x_1,\dots,x_n)$ and its density function is given by

$$g(t,\theta) = \begin{cases} g(t)/G(\theta) , & 0 \le t \le \theta ; \\ 0 , & \theta < t , \end{cases}$$

where $g(t)=nF^{n-1}(t)f(t)$ and $G(\theta)=F^n(\theta)$. T is complete, and if $\eta(\theta)$ is an absolutely continuous function defined on $(0, \infty)$ and if $\lim_{\varepsilon\to+0} \eta(\varepsilon)G(\varepsilon)=0$, the UMVU estimator of $\eta(\theta)$ is given by

$$\phi(t)=\frac{1}{g(t)}\{\eta(t)G(t)\}'=\eta(t)+\eta'(t)\frac{G(t)}{g(t)}=\eta(t)+\eta'(t)\frac{F(t)}{nf(t)} .$$

The estimator $\phi(T)$ has the variance

$$V_\theta(\phi(T))=\frac{1}{G(\theta)}\int_0^\theta \frac{\{\eta'(t)G(t)\}^2}{g(t)}dt=\frac{1}{nF^n(\theta)}\int_0^\theta\{\eta'(t)\}^2\frac{F^{n+1}(t)}{f(t)}dt .$$

Example 2.2.4 ([7]). Let $X_1, \ldots, X_n$ be i.i.d. random variables with a density function

$$f(x, \theta) = \begin{cases} f(x)/F(\theta), & \theta \le x \le b(\theta), \\ 0, & \text{otherwise}, \end{cases}$$

where $f(x)>0$ a.e. and $F(\theta)=\int_\theta^{b(\theta)} f(x)dx < \infty$.

A minimal sufficient statistic for $(\theta, b(\theta))$ is the pair of $T_1=\min(X_1,\ldots,X_n)$ and $T_2=\max(X_1,\ldots,X_n)$. Then the family of its densities $g_\theta(t)=g(t_1,t_2)/G(\theta)$ with

$$g(t_1,t_2) = n(n-1)\left\{\int_{t_1}^{t_2} f(x)dx\right\}^{n-2} f(t_1)f(t_2)$$

and

$$G(\theta) = \left\{\int_\theta^{b(\theta)} f(x)dx\right\}^n ,$$

is not complete.

Assume that $b(\theta)$ $(>\theta)$ is a strictly increasing and a.e. differentiable function. It will be possible to construct an unbiased estimator $\phi(x)$ of $\eta(\theta)$, based on a single observation X, with zero variance at a given value θ_0 of the parameter.

Obviously, an estimator with this property must satisfy

$$\phi(x) = \eta(\theta_0) \text{ a.e.,} \quad \theta_0 \le x \le b(\theta_0) .$$

And from its unbiasedness

$$E_\theta [\phi(x)] = \int_\theta^{b(\theta)} \phi(x) \frac{f(x)}{F(\theta)} dx = \eta(\theta) \quad \text{for all } \theta ,$$

which implies the relation

$$b'(x) \phi(b(x))f(b(x)) - \phi(x)f(x) = [\eta(x)F(x)]'$$

for almost all x.

As a special case in the example let $b(\theta)=2\theta$, $F(\theta)=\theta$, $f(x)=1$ and $\eta(\theta)=\theta$. Then we have

(2.2.2) $$2\phi(2x) - \phi(x) = 2x .$$

Letting $\theta_0=1$, we obtain $\phi(x)=1$ for $1\le x< 2$. From (2.2.2) we have

$$\phi(x) = \frac{1}{2}(x+1) \quad \text{for} \quad 2\le x< 4 \ ;$$

$$\phi(x) = \frac{1}{2}(\frac{5}{4} x + \frac{1}{2}) \quad \text{for} \quad 4\le x< 8 \ ;$$

$$\phi(x) = \frac{1}{8}(\frac{21}{4} x + 1) \quad \text{for} \quad 8\le x< 16 , \ldots .$$

Similarly we obtain from (2.2.2)

$$\phi(x) = 2-2x \quad \text{for} \quad \frac{1}{2}\le x< 1 \ ;$$

$$\phi(x) = 4 - 10x \quad \text{for} \quad \frac{1}{4}\le x< \frac{1}{2} \ ;$$

$$\phi(x) = 8 - 42x \quad \text{for} \quad \frac{1}{8}\le x< \frac{1}{4} , \ldots .$$

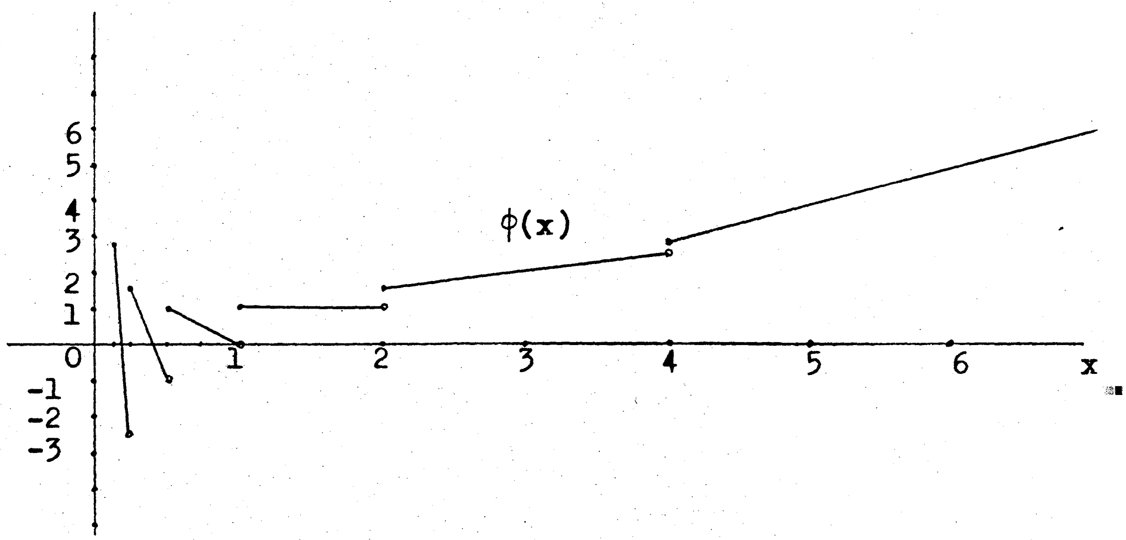

Figure 2.2.1. The unbiased estimator $\phi(x)$ given in the above way in the case of the uniform distribution.

In other cases of non-constant support, we have the Cramér-Rao-Bhattacharyya bound.

Even when the support depends on θ, we sometimes have the Cramér-Rao type bounds for unbiased estimators. Suppose that a random variable X is distributed with the density function $f(x-\theta)$ with respect to the Lebesgue measure where f(x) is defined as

$$f(x) = \begin{cases} c(1-x^2)^2 , & |x| \le 1 , \\ 0 , & \text{otherwise}, \end{cases}$$

where c is some constant.

Then for any unbiased estimator $\hat{\theta}(x)$ we have

$$\int_{\theta-1}^{\theta+1} \hat{\theta}(x) f(x-\theta) dx = \theta.$$

By differentiating with respect to θ we have

$$\hat{\theta}(\theta+1)f(1) - \hat{\theta}(\theta-1)f(-1) - \int_{\theta-1}^{\theta+1} \hat{\theta}(x) f'(x-\theta) dx = 1 .$$

Then it must be proved that the differentiation is allowed, but we omit the detailed discussion.

Since $f(1)=f(-1)=0$ we have

$$\int_{\theta-1}^{\theta+1} \hat{\theta}(x) f'(x-\theta) dx = 1 .$$

By putting $\theta=0$ and applying the Cauchy-Schwarz inequality we have

$$\left\{ \int_{-1}^{1} \hat{\theta}^2(x) f(x) dx \right\} \left\{ \int_{-1}^{1} \left(\frac{f'(x)}{f(x)}\right)^2 f(x) dx \right\} \ge \left\{ \int_{-1}^{1} \hat{\theta}(x) f'(x) dx \right\}^2 = 1 .$$

Hence we obtain

$$V_0(\hat{\theta}(X)) \ge \frac{1}{\int_{-1}^{1} \frac{(f'(x))^2}{f(x)} dx} = \frac{3}{32} c ,$$

What is remarkable here is that the bound is sharp, that is,

$$\inf_{\hat{\theta}:\text{unbiased}} V_0(\hat{\theta}(x)) = \frac{3}{32} c .$$

Such an argument can be generalized to the Bhattacharyya type inquality.

Example 2.2.5. ([2]). We consider the location parameter case, i.e. $f(x, \theta)=f(x-\theta)$, and unbiased estimator of θ. We assume that for any $p \geq 1$, the density function $f(x)$ is given by

$$f(x) = \begin{cases} c(1-x^2)^{p-1} & \text{if } |x|<1 ; \\ 0 & \text{otherwise} , \end{cases}$$

where $c=1/B(1/2, p)$ with $B(\alpha, \beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$ $(\alpha>0, \beta>0)$.

Case (i). Let p=1. Then the distribution is uniform, and it is easy to check that

$$\min_{\hat{\theta}:\text{unbiased}} V_{\theta_0}(\hat{\theta})=0$$

for any specific value θ_0 ([10]).

Case (ii). Let p=2. In this case, it is easy to check that the Fisher information

$$\int_{-\infty}^{\infty}\left\{ \frac{f'(x)}{f(x)} \right\}^2 f(x)dx = \infty .$$

Then it is shown that

$$\inf_{\hat{\theta}:\text{unbiased}} V_{\theta_0}(\hat{\theta})=0$$

for any specified value θ_0.

Before proceeding on to the next cases, we note the following :

Let Λ be a $k \times k$ non-negative definite matrix whose elements are

$$(2.2.3) \quad \lambda_{ij}= \int_{\theta-1}^{\theta+1} \frac{1}{f(x-\theta)} \frac{\partial^i f(x-\theta)}{\partial\theta^i} \frac{\partial^j f(x-\theta)}{\partial\theta^j} dx$$

$$= \int_{-1}^{1} (-1)^{i+j} \frac{1}{f(x)} f^{(i)}(x)f^{(j)}(x)dx \; ; \; i,j=1,\dots,k.$$

If $k<p/2$, λ_{ii} $(i=1,\dots,k)$ are finite since $\int_{-1}^{1}(1-x^2)^{p-2k-1}dx<\infty$.

Then the Bhattacharyya bound of the variance of unbiased estimator $\hat{\theta}(X)$ of θ is given by

$$(2.2.4) \quad V_{\theta_0}(\hat{\theta}) \geq (1,0,\dots,0)\,\Lambda^{-1}(1,0,\dots,0)'$$

for any spacified value θ_0.

(See Akahira, Puri and Takeuchi [2]).

We obtain for $|x| < 1$,

$$
\begin{aligned}
f^{(1)}(x) &= -2c(p-1)x(1-x^2)^{p-2} \;; \\
f^{(2)}(x) &= -2c(p-1)\{(1-x^2)^{p-2} - 2(p-2)x^2(1-x^2)^{p-3}\} \;; \qquad (2.2.5)\\
f^{(3)}(x) &= 4c(p-1)(p-2)\{3x(1-x^2)^{p-3} - 2(p-2)x^3(1-x^2)^{p-4}\} \;; \\
&\cdots .
\end{aligned}
$$

If i+j is an odd number, if follows by (2.2.3) and (2.2.5) that $\lambda_{ij}=0$ since $f^{(i)}(x)f^{(j)}(x)$ is an odd function.

From (2.2.3) and (2.2.5), we have

$$\lambda_{11} = 4c(p-1)^2 B\left(\frac{3}{2}, p-2\right) \;;$$

$$\lambda_{13} = 8c(p-1)^2(p-2)\left\{2(p-3)B\left(\frac{5}{2}, p-4\right) - 3B\left(\frac{3}{2}, p-3\right)\right\} \;;$$

$$\lambda_{22} = 4c(p-1)^2\left\{B(1/2, p-2) - 4(p-2)B\left(\frac{3}{2}, p-3\right) + 4(p-2)^2 B\left(\frac{5}{2}, p-4\right)\right\} \;;$$

$$\lambda_{33} = 16c(p-1)^2(p-2)^2\left\{9B\left(\frac{3}{2}, p-2\right) - 12(p-3)B\left(\frac{5}{2}, p-5\right) + 4(p-3)^2 \cdot B\left(\frac{7}{2}, p-6\right)\right\} \;; \;\ldots .$$

Case(iii). Let p=3,4,5,6. Then it is shown that

$$\inf_{\hat{\theta}:\text{unbiased}} V_{\theta_0}(\hat{\theta}) = \frac{1}{\lambda_{11}}$$

for any specific value θ_0 .

Case(iv). Let p=7. In this case, we see that k=3. Using (2.2.3), (2.2.4) and (2.2.5), we obtain

$$V_{\theta_0}(\theta) \geq \frac{1}{|\Lambda|}\begin{vmatrix} \lambda^{22} & 0 \\ 0 & \lambda_{33} \end{vmatrix} = \left[\lambda_{11}\left(1 - \frac{\lambda_{13}^2}{\lambda_{11}\lambda_{33}}\right)\right]^{-1}$$

where

$$\Lambda = \begin{pmatrix} \lambda_{11} & 0 & \lambda_{13} \\ 0 & 22 & 0 \\ \lambda_{13} & 0 & \lambda_{33} \end{pmatrix}$$

with

$$\lambda_{11} = 144cB(\tfrac{3}{2}\ ,\ 5)\ ;$$

$$\lambda_{13} = 1440c\{\ 8B(\tfrac{5}{2}\ ,\ 3)\ -\ 3B(\tfrac{3}{2}\ ,\ 4)\}\ ;$$

$$\lambda_{22} = 144c\ \{B(\tfrac{1}{2}\ ,\ 5)\ -\ 20B(\tfrac{3}{2}\ ,\ 4)\ +\ 100B(\tfrac{5}{2}\ ,\ 3)\}\ ;$$

$$\lambda_{33} = 14400c\ \{9B(\tfrac{3}{2}\ ,\ 5)\ -\ 48B(\tfrac{5}{2}\ ,\ 2)\ +\ 64B(\tfrac{7}{2}\ ,\ 1)\}\ .$$

We also obtain

$$\frac{\lambda_{13}^2}{\lambda_{11}\lambda_{33}} = \frac{\{\ 8B(\frac{5}{2},3)-3B(\frac{3}{2},4)\}^{\ 2}}{B(\frac{3}{2},5)\{\ 9B(\frac{3}{2},5)-48B(\frac{5}{2},2)+64B(\frac{7}{2},1)\}}$$

Here again $\inf_{\hat{\theta}:\text{unbiased}} V_{\theta_0}(\hat{\theta}) = [\lambda_{11}(1-\frac{\lambda_{13}^2}{\lambda_{11}\lambda_{13}})]^{-1}$ for any specific θ_0 , i.e. we have a sharp bound.

Case (v). For $p \geq 8$ we can continue in a similar manner by choosing $k = [(p-1)/2]$, where $[s]$ denotes the largest integer less than or equal to s.

2.3. Discrete parameter set.

In some cases the parameter set itself can be discrete. Then the parameter set is either finite or countable. The case of finite parameter set can be generalized to the case of finite rank, which is defined as follows.

The class $\{P_\omega\ |\omega\ \Omega\}$ of probability distributions is said to be of rank $m(<\infty)$ if there exist ω_1 , ... , ω_m in Ω such that for any $\omega\in\Omega$ there exists a set of constants $c_1(\omega)$, ... , $c_m(\omega)$ satisfying

$P_\omega = c_1(\omega)P_{\omega_1} + \ldots + c_m(\omega)P_{\omega_m}$ for all $\omega \in \Omega$

and $P_{\omega_1}, \ldots, P_{\omega_m}$ are linearly independent in the sense that $c_1P_{\omega_1} + \ldots + c_mP_{\omega_m} = 0$ implies $c_1 = \ldots = c_m = 0$.

Then for any real-valued function $\theta = \theta(\omega)$ of ω, unbiased estimators exist if and only if

$$\theta(\omega) = c_1(\omega)\,\theta(\omega_1) + \ldots + c_m(\omega)\,\theta(\omega_m)$$

for all $\omega \in \Omega$. The class of all UMVU estimators (of any parameter) is equal to the class of measurable functions of a finite subfild $\mathfrak{M}$ of $\mathcal{A}$.

In the case of countable index set we have an LMVU estimator for any parameter point $\omega = \omega_0$ whose variance is infinity except for $\omega = \omega_0$.

<u>Example 2.3.1</u> ([6]). Let $X_1, \ldots, X_n$ be i.i.d. random variables according to a normal distribution with mean m and variance nd^2, where m is a integer-valued parameter. The sample mean $\bar{X} = \sum_{i=1}^{n} X_i/n$ is sufficient and its distribution is normal with mean m and variance d^2. Then the LMVU estimator at m=0 is given by

$$f(\bar{x}) = \sum_{u=1}^{\infty}(-1)^{u+1}\{\exp(u(u-1)/2d^2)(\exp(u/d^2)-1)\exp(u^2/2d^2)\}^{-1}$$

$$\cdot\{\exp(u\bar{x}/d^2) - \exp(-u\bar{x}/d^2)\}.$$

Its local minimum variance at m=0 is

$$V(f) = 2\sum_{u=1}^{\infty}(-1)^{u+1}u\,\{\exp(-u(u-1)/2d^2)/(\exp(u/d^2)-1)\}.$$

The LMVU estimator at m=0 has infinite variance at all $m \neq 0$.

When the sample space $\mathcal{X}$ itself is a finite set of size N, then the probability distribution P_ω can be considered to be a finite dimensional real N-vector, and if there are N linearly independent P_{ω_i} (i=1, ... , N) then the sample X is complete and any **parameter** with unbiased estimators has a unique unbiased estimator which is UMV unbiased by the definition.

Example 2.3.2. Let X be distributed according to a hypergeometric distribution

$$P(X=x) = \frac{{}_nC_x \cdot {}_{N-M}C_{n-x}}{{}_NC_M} \quad \text{for} \quad x=0,1, \dots , n ,$$

where $n < M$ and $2n < N$.

Let M be the unknown parameter and the possible values of M be 0, 1, ... , N-n. Then X is complete.

2.4. Discontinuous and non-differentiable density.

Now suppose that the density function $f(x,\theta)$ is not continuous with respect to θ while the support $A(\theta)=\{x|f(x,\theta)>0\}$ can be defined to be independent of θ . Note that this condition does not affect the existence of a LMVU estimator, because it can be shown that for any real $\gamma(\theta)$ which has an unbiased estimator, a LMVU estimator at $\theta=\theta_0$ exists if

$$\int_{A(\theta_0)} \frac{\{f(x,\theta)\}^2}{f(x,\theta_0)} d\mu < \infty$$

(e.g. see Barankin [4] and Stein [8]).

But the LMVU estimator sometimes behaves very strangely.

Example2.4.1. ([3]). Let $X_1, \dots , X_n$ be i.i.d. random variables with the following density

$$f(x,\theta) = \begin{cases} p, & 0< x<\theta \quad \text{and} \quad \theta+1< x \leq 2 , \\ q, & \theta < x< \theta + 1 , \\ 0, & \text{otherwise} , \end{cases}$$

where p and q with $0<p<q$ and $p+q=1$ are fixed constants.
Since the support of $f(x,\theta)$ is the interval $[0, 2]$, in the subsequent discussion it is enough to consider it as the domain of x.
The parameter θ has the range $0\leq\theta\leq 1$.
Then the LMVU estimator of θ at $\theta=\theta_0$ has the form

$$(2.4.1)\quad \hat{\theta}_0(x_1,\ldots,x_n) = \int_0^1 \prod_{i=1}^{n} \frac{f(x_i, \eta)}{f(x_i,\theta_0)}\, dG_n(\eta) ,$$

where $G_n(\eta)$ is some signed measure over the closed interval $[0, 1]$ (e.g. see Stein [8]).

Case (i). Let $n=1$ and $x=x_1$. Then we take a signed measure G_1^* over $[0, 1]$ satisfying $G_1(\{0\})=-1/c$, $G_1(\{1\})=1/c$, $G_1(\{\theta_0\})=\theta_0$ and $G_1((0,\theta_0)\cup(\theta_0,1))=0$. Letting $G_1=G_1^*$ in (2.4.1) we have

$$\hat{\theta}_1^*(x) = \int_0^1 \frac{f(x,\eta)}{f(x,\theta_0)}\, dG_1^*(\eta)$$

$$= \begin{cases} \frac{1}{c}(1-\frac{q}{p}) & \text{for } 0\leq x\leq\theta_0 ; \\ \frac{1}{c}(\frac{p}{q}-1) & \text{for } \theta_0 < x < 1 ; \\ \frac{1}{c}(1-\frac{p}{q}) & \text{for } 1< x < \theta_0+1 ; \\ \frac{1}{c}(\frac{q}{p}-1) & \text{for } \theta_0+1\leq x\leq 2 , \end{cases}$$

where $c=\frac{1}{pq}-4\ (>0)$.

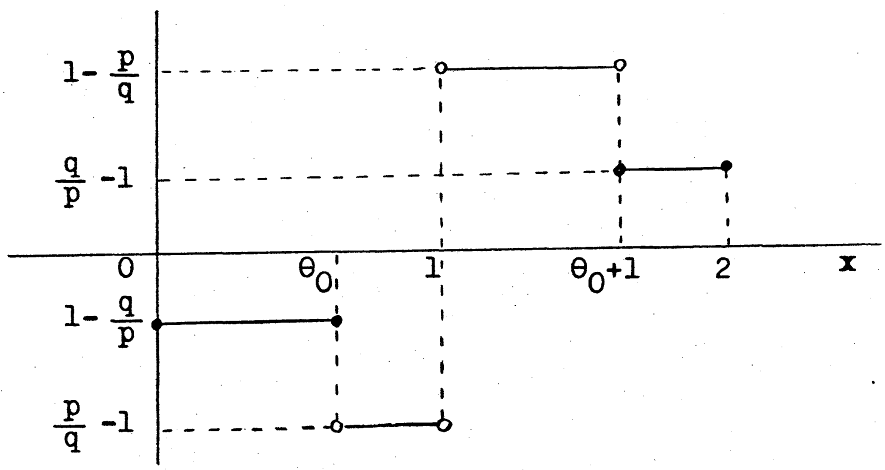

Figure 2.4.1. The values of $\hat{\theta}_1^*(x)$ for all $x\in[0, 2]$.

Then the variance of the LMVU estimator $\hat{\theta}_1^*$ is given by

$$(2.4.2) \qquad V_\theta(\hat{\theta}_1^*) = \frac{1}{c} + |\theta - \theta_0|(1 - |\theta - \theta_0|).$$

Note that

$$V_\theta(\hat{\theta}_1^*) \geq \frac{1}{c} = V_{\theta_0}(\hat{\theta}_1^*)$$

for all $\theta \in [0, 1]$.

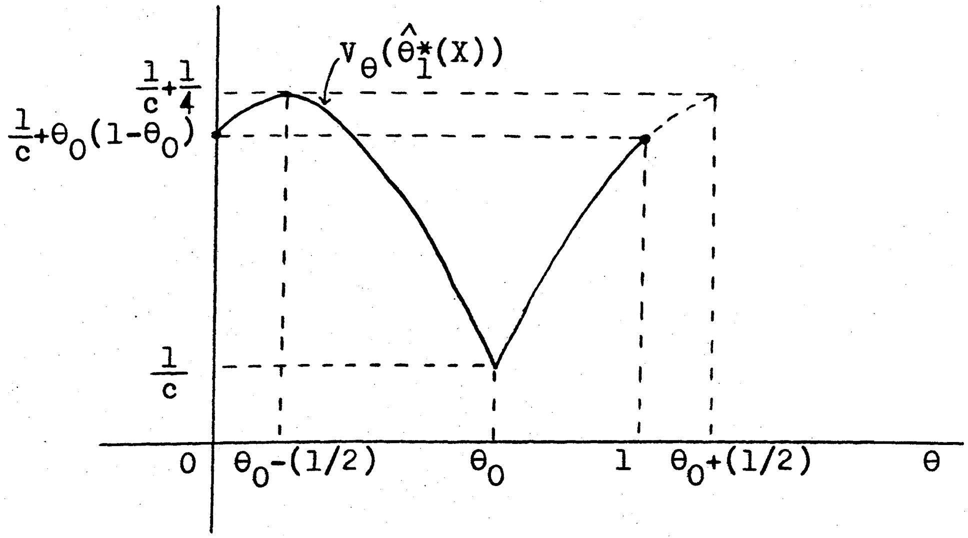

Figure 2.4.2. The variance of the LMVU estimator $\hat{\theta}_1^*$ at $\theta=\theta_0$ given by (2.4.2) when $\theta_0 > 1/2$.

Case (ii). Let $n \geq 2$. Then we take a signed measure G_n^* over $[0, 1]$ satisfying

$$\frac{dG_n^*}{d\eta} = (1 - \frac{1}{n})\frac{1}{(1+c|\eta-\theta_0|)^n} \operatorname{sgn}(\eta-\theta_0) \text{ for } 0<\eta<\theta_0 \text{ , } \theta_0<\eta<1 \quad ;$$

$$G_n^*(\{0\}) = -\frac{1}{cn(1+c\theta_0)^{n-1}} \ ;$$

$$G_n^*(\{1\}) = \frac{1}{cn(1+c(1-\theta_0))^{n-1}} \ ;$$

$$G_n^*(\{\theta_0\}) = \theta_0 .$$

It is shown that

$$\hat{\theta}_0^*(x_1,\ldots,x_n) = \int_0^1 \prod_{i=1}^{n} \frac{f(x_i,\eta)}{f(x_i,\theta_0)}\, dG_n^*(\eta)$$

is an LMVU estimator of θ which has the following variance at $\theta=\theta_0$:

$$V_{\theta_0}(\hat{\theta}_n^*) = \begin{cases} \frac{1}{2c^2}\log\{(1+c\theta_0)(1+c(1-\theta_0))\} & \text{for } n=2 ; \\ \frac{1}{c^2 n(n-2)}\left\{2-\frac{1}{(1+c\theta_0)^{n-2}} - \frac{1}{(1+c(1-\theta_0))^{n-2}}\right\} & \text{for } n>2. \end{cases}$$

In the case of discontinuous or non-differentiable density functions a lower bound of the variance of unbiased estimators was obtained by Chapman and Robbins [5] , which is given by

$$V_{\theta_0}(\hat{\theta}(x)) \geq \sup_{\eta} \frac{(\eta-\theta_0)^2}{\int_{\mathcal{X}}\left\{\frac{f(x,\eta)}{f(x,\theta_0)}-1\right\}^2 d\mu} .$$

By definition we have

$$V_{\theta_0}(\hat{\theta}(x)) \geq \left[\inf_{\eta} \frac{K_{\theta_0}(\eta)}{(\eta-\theta_0)^2}\right]^{-1} ,$$

where

$$K_{\theta_0}(\eta) = \int_{\mathcal{X}} \frac{\{f(x,\eta)-f(x,\theta_0)\}^2}{f(x,\theta_0)}\, d\mu .$$

When $X_1, \ldots, X_n$ are i.i.d. random variables,

(2.4.3) $$V_{\theta_0}(\hat{\theta}(X_1,\ldots,X_n)) \geq \left[\inf_{\eta} \frac{\{k_{\theta_0}(\eta)+1\}^n-1}{(\eta-\theta_0)^2}\right]^{-1} .$$

In the regular cases we have

$$K_{\theta_0}(\eta) = I(\theta)(\eta-\theta_0)^2 + o((\eta-\theta_0)^2)$$

and the right-hand side of (2.4.3) is attained when $\eta-\theta_0$ is of the order $1/\sqrt{n}$. In the non-regular cases we have

$$K_{\theta_0}(\eta) = I(\theta)\,|\eta-\theta_0|^{\alpha} + o(|\eta-\theta_0|^{\alpha}) ,$$

where $\alpha > 0$.

2.5. $f(x, \theta)/f(x,\theta_0)$ is not square-integrable.

There are cases when the support $S=\{x \mid f(x, \theta) > 0\}$ is independent of θ, but $\int_S \{f(x, \theta)\}^2/f(x,\theta_0)d\mu$ is infinite for some $\theta \in Ⓗ$.

Assume that $Ⓗ = \{\theta_0, \theta_1, \dots, \theta_p\}$. Then we consider to minimize

$$\int_S \{\hat{\theta}(x)\}^2 f(x, \theta_0)d\mu$$

under the unbiasedness condition : $E_{\theta_i}(\hat{\theta})=\gamma(\theta_i)=\gamma_i$ (say) $(i=1,\dots,k)$ and the condition

(A.2.5.1) $f(x, \theta_i)(i=1,\dots,p)$ are linearly independent,

$$\int_S \frac{\{f(x, \theta_i)\}^2}{f(x,\theta_0)} d\mu<\infty \qquad (i=1,\dots,k\ ;\ k \le p),$$

and

$$\int_S \frac{\{\sum_{i=1}^{p} c_i f(x, \theta_i)\}^2}{f(x,\theta_0)} d\mu < \infty$$

implies

$$c_{k+1} = \dots = c_p = 0 .$$

Let Λ be a $k\times k$ non-negative definite matrix whose elements are

$$\lambda_{ij} = \int_S \frac{f(x, \theta_i)f(x, \theta_j)}{f(x,\theta_0)} d\mu \quad (i,j=1,\dots,k) .$$

Then it is shown by Takeuchi and Akahira [11] that under the condition (A.2.5.1)

$$\inf_{\hat{\theta}:\text{unbiased}} \int_S \{\hat{\theta}(x)\}^2 f(x, \theta_0)d\mu = \underset{\sim}{\gamma}'_{(k)}\Lambda\,\underset{\sim}{\gamma}_{(k)}$$

holds, where $\underset{\sim}{\gamma}_{(k)}=(\gamma_1, \dots, \gamma_k)'$ is given in the above. Therefore in this case the lower bound of the variance of the unbiased estimator can be obtained by simply disregarding of $(\theta_{k+1}, \dots, \theta_p)$, however, it should be noted that the lower bound is not generally attained.

References

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

[2] Akahira,M., Puri,M.L., and Takeuchi,K. (1983). Bhattacharyya bound of variances of unbiased estimators in nonregular cases. (To appear).

[3] Akahira,M. and Takeuchi,K. (1984). Locally minimum variance unbiased estimator in a discontinuous density function. (To appear).

[4] Barankin, E.W. (1949). Locally best unbiased estimates. Ann. Math. Statist. 20, 477-501.

[5] Chapman,D.G. and Robbins,H. (1951). Minimum variance estimation without regularity assumptions. Ann. Math. Statist. 22, 581-586.

[6] Kojima,Y., Morimoto,H., and Takeuchi,K. (1982). Two best unbiased estimators of normal integral mean. In Statistics and Probability : Essays in Honor of C.R.Rao, North-Holland, New York, 429-441.

[7] Morimoto,H. and Sibuya,M. (1967). Sufficient statistics and unbiased estimation of restricted selection parameter. Sankhya (A) 27, 15-40.

[8] Stein,C. (1950). Unbiased estimates of minimum variance. Ann. Math. Statist. 21, 406-415.

[9] Takeuchi,K. (1961). Some remarks on unbiased estimation in sampling from finite populations. Rep. Stat. Appl. Res., JUSE, 8, 207-210.

[10] Takeuchi,K. (1961). On the fallacy of a theory of Gunner Blom. Rep. Stat. Appl. Res., JUSE, 9, 34-35.

[11] Takeuchi,K. and Akahira,M. (1984). A note on minimum variance. (To appear in Metrika).