Ann. Inst. Statist. Math.
34 (1982), Part A, 399-409

UNBIASED ESTIMATORS IN THE SENSE OF LEHMANN AND THEIR DISCRIMINATION RATES (II): MULTI-PARAMETER CASES

HISATAKA KUBOKI*

(Received Feb. 4, 1982; revised Sept. 9, 1982)

Summary

The notion of *discrimination rate* of any unbiased estimator in the sense of Lehmann is, as defined by the author (1982, *Ann. Inst. Statist. Math.*, 34, A, 19-37), extended to multi-parameter cases.

1. Introduction

For any L-unbiased (unbiased in the sense of Lehmann [2]) estimator, the notion of its discrimination rate, which is an index of its performance, has been introduced by the present author (Kuboki [1]). Let $g(\theta)$, $\theta \in \Theta$ $(\subset \boldsymbol{R}^1)$ be a parametric function to be estimated, and T be an estimator of it. Let $W(\cdot, T)$ be a loss incurred by T. We assume that T is L-unbiased with respect to W, that is,

$$\mathrm{E}_\theta W(\theta, T) \leqq \mathrm{E}_\theta W(\tau, T), \qquad \text{for all } \theta, \tau \in \Theta .$$

The discrimination rate $D(\theta; T, W)$ of T at θ is a measure for evaluating the rate of change of $\mathrm{E}_\theta W(\tau, T)$ to small changes in τ at θ, which is defined by

$$D(\theta; T, W) = 2 \varliminf_{\tau \to \theta} \Lambda(\tau, \theta; T, W)/|\tau - \theta| ,$$

where

$$\Lambda(\tau, \theta; T, W) = \mathrm{E}_\theta \{W(\tau, T) - W(\theta, T)\} / [\mathrm{Var}_\theta \{W(\tau, T) - W(\theta, T)\}]^{1/2} .$$

The estimator is powerful to discriminate any wrong value of $g(\cdot)$ from the correct one when the discrimination rate is large.

This paper is concerned with an extension of the notion of discrimination rate to multi-parameter cases. In Section 2, we discuss a method

* This research was supported in part by a Grant-in-Aid for Scientific Research of the Japanese Ministry of Education, Science and Culture.

of extension and establish an upper bound of discrimination rates, which is an intrinsic sensitivity of a family of distributions under consideration. In Section 3, our extended result is applied to some typical loss functions. We investigate the relations between risk functions and discrimination rates when loss functions are of the quadratic form. As an example, we consider the estimation of parameters of a simple normal linear regression model.

2. Extension to multi-parameter cases

Let $X_1, \cdots, X_n$ be independent and identically distributed random vectors with common distribution $P_{\boldsymbol{\theta}}$, $\boldsymbol{\theta}=(\theta_1, \cdots, \theta_k)' \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is an open subset of $\boldsymbol{R}^k$. Let $f(\cdot, \boldsymbol{\theta})$ be the density of $P_{\boldsymbol{\theta}}$ relative to a σ-finite measure μ. We assume that utilizing a statistic $\boldsymbol{T}$ to estimate $\boldsymbol{g}(\boldsymbol{\theta})$ yields a loss $W(\boldsymbol{\theta}, \boldsymbol{T})$, where $\boldsymbol{T}$: $\boldsymbol{X}=(X_1, \cdots, X_n) \rightarrow \boldsymbol{R}^m$ and $\boldsymbol{g}$: $\boldsymbol{\theta}=(\theta_1, \cdots, \theta_k)' \rightarrow \boldsymbol{R}^m$, and that $\boldsymbol{T}$ is L-unbiased with respect to W, that is,

$$\mathrm{E}_{\boldsymbol{\theta}}\, W(\boldsymbol{\theta}, \boldsymbol{T}) \leqq \mathrm{E}_{\boldsymbol{\theta}}\, W(\boldsymbol{\tau}, \boldsymbol{T}) , \qquad \text{for all } \boldsymbol{\theta}, \boldsymbol{\tau} \in \boldsymbol{\Theta} .$$

Our method of extension is simple: reducing the multi-parameter to one parameter and applying the result of one-parameter cases to the reduced parameter. For every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and every $\boldsymbol{h}=(h_1, \cdots, h_k)'$ with unit length, we consider the case where $\boldsymbol{\tau}$ approaches $\boldsymbol{\theta}$ in the direction $\boldsymbol{h}$, that is, $\boldsymbol{\tau}=\boldsymbol{\theta}+v\boldsymbol{h}$, where v is a scalar such that $v \rightarrow 0$. By this restriction, we can reduce the multi-parameter to a one-parameter case, that is, $v \in V(\boldsymbol{\theta}, \boldsymbol{h})=\{v\colon \boldsymbol{\theta}+v\boldsymbol{h} \in \boldsymbol{\Theta}\}$. Then we naturally extend the definitions of discrimination rate and sensitivity as follows.

DEFINITION 2.1. For every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and every $\boldsymbol{h} \in \boldsymbol{R}^k$ with unit length, we shall define the *discrimination rate* of $\boldsymbol{T}$ in the direction $\boldsymbol{h}$ at $\boldsymbol{\theta}$ by

$$D(\boldsymbol{\theta}; \boldsymbol{T}, W|\boldsymbol{h})=2 \varliminf_{v \rightarrow 0} \Lambda(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{\theta}; \boldsymbol{T}, W)/|v| ,$$

where

$$\begin{aligned}&\Lambda(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{\theta}; \boldsymbol{T}, W)\\ &\quad=\mathrm{E}_{\boldsymbol{\theta}}\{W(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T})-W(\boldsymbol{\theta}, \boldsymbol{T})\}/[\mathrm{Var}_{\boldsymbol{\theta}}\{W(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T})-W(\boldsymbol{\theta}, \boldsymbol{T})\}]^{1/2}\end{aligned}$$

and $\Lambda(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{\theta}; \boldsymbol{T}, W)=0$ whenever the numerator vanishes.

DEFINITION 2.2. For every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and every $\boldsymbol{h} \in \boldsymbol{R}^k$ with unit length, we shall define the *sensitivity* of the family $\{P_{\boldsymbol{\theta}};\ \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ in the direction $\boldsymbol{h}$ at $\boldsymbol{\theta}$ by

$$s(\boldsymbol{\theta}\,|\,\boldsymbol{h})=2 \varliminf_{v \rightarrow 0} \rho(P_{\boldsymbol{\theta}+v\boldsymbol{h}}, P_{\boldsymbol{\theta}})/|v| ,$$

where

$$\rho(P_{\boldsymbol{\theta}+v\boldsymbol{h}}, P_{\boldsymbol{\theta}})=\left[\int \{\sqrt{f(x, \boldsymbol{\theta}+v\boldsymbol{h})}-\sqrt{f(x, \boldsymbol{\theta})}\}^2 d\mu\right]^{1/2} .$$

Similarly, the sensitivity of the family $\{Q_{\boldsymbol{\theta}};\ \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ of probability measures induced by $\boldsymbol{T}$ is defined as

$$s(\boldsymbol{\theta}; \boldsymbol{T}|\boldsymbol{h})=2 \lim_{v\to 0} \rho(Q_{\boldsymbol{\theta}+v\boldsymbol{h}}, Q_{\boldsymbol{\theta}})/|v| .$$

We now establish an inequality similar to that of Theorem 3.1 of [1]. In what follows, we assume the following conditions for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and every $\boldsymbol{h} \in \boldsymbol{R}^k$ with unit length:

(i) $\mathrm{E}_{\boldsymbol{\theta}+v_1\boldsymbol{h}} W^2(\boldsymbol{\theta}+v_2\boldsymbol{h}, \boldsymbol{T})$ is finite, for all $v_1, v_2 \in V(\boldsymbol{\theta}, \boldsymbol{h})$.

(ii) $W(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T})$ is twice continuously differentiable with respect to v in $V(\boldsymbol{\theta}, \boldsymbol{h})$, for all values of $\boldsymbol{T}$. The first and second derivatives at each $v \in V(\boldsymbol{\theta}, \boldsymbol{h})$ will be denoted by $\dot{W}(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T}|\boldsymbol{h})$ and $\ddot{W}(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T}|\boldsymbol{h})$, respectively.

The following is a key formula in our discussion (Lemma 2.1 of [1]).

$$\begin{aligned}(2.1)\quad &\mathrm{E}_{\boldsymbol{\theta}} \{W(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T})-W(\boldsymbol{\theta}, \boldsymbol{T})\} \\ &\quad +\mathrm{E}_{\boldsymbol{\theta}+v\boldsymbol{h}} \{W(\boldsymbol{\theta}, \boldsymbol{T})-W(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T})\} \\ &\leqq [\mathrm{E}_{\boldsymbol{\theta}} \{W(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T})-W(\boldsymbol{\theta}, \boldsymbol{T})\}^2/2 \\ &\quad +\mathrm{E}_{\boldsymbol{\theta}+v\boldsymbol{h}} \{W(\boldsymbol{\theta}, \boldsymbol{T})-W(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T})\}^2/2]^{1/2}[4\rho^2(Q_{\boldsymbol{\theta}+v\boldsymbol{h}}, Q_{\boldsymbol{\theta}})]^{1/2} \\ &\leqq [\mathrm{E}_{\boldsymbol{\theta}} \{W(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T})-W(\boldsymbol{\theta}, \boldsymbol{T})\}^2/2 \\ &\quad +\mathrm{E}_{\boldsymbol{\theta}+v\boldsymbol{h}} \{W(\boldsymbol{\theta}, \boldsymbol{T})-W(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T})\}^2/2]^{1/2}[4n\rho^2(P_{\boldsymbol{\theta}+v\boldsymbol{h}}, P_{\boldsymbol{\theta}})]^{1/2} ,\end{aligned}$$

for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\boldsymbol{h} \in \boldsymbol{R}^k$ with unit length and $v \in V(\boldsymbol{\theta}, \boldsymbol{h})$. Then, using (2.1), we can establish the following theorem, which is straightforward from Theorems 2.1 and 3.1 of [1].

THEOREM 2.1. *Suppose that for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and every $\boldsymbol{h} \in \boldsymbol{R}^k$ with unit length, both $\mathrm{E}_{\boldsymbol{\theta}+v_1\boldsymbol{h}} \dot{W}^2(\boldsymbol{\theta}+v_2\boldsymbol{h}, \boldsymbol{T}|\boldsymbol{h})$ and $\mathrm{E}_{\boldsymbol{\theta}+v_1\boldsymbol{h}} \ddot{W}(\boldsymbol{\theta}+v_2\boldsymbol{h}, \boldsymbol{T}|\boldsymbol{h})$ are continuous functions of $v_1, v_2 \in V(\boldsymbol{\theta}, \boldsymbol{h})$. Then we have*

$$(2.2)\qquad D(\boldsymbol{\theta}; \boldsymbol{T}, W|\boldsymbol{h})\leqq s(\boldsymbol{\theta}; \boldsymbol{T}|\boldsymbol{h})\leqq \sqrt{n}\, s(\boldsymbol{\theta}|\boldsymbol{h}) ,$$

and

$$D(\boldsymbol{\theta}; \boldsymbol{T}, W|\boldsymbol{h})=\mathrm{E}_{\boldsymbol{\theta}} \ddot{W}(\boldsymbol{\theta}, \boldsymbol{T}|\boldsymbol{h})/[\mathrm{E}_{\boldsymbol{\theta}} \dot{W}^2(\boldsymbol{\theta}, \boldsymbol{T}|\boldsymbol{h})]^{1/2} .$$

Note that the following condition (ii)′ implies condition (ii):

(ii)′ $W(\boldsymbol{\theta}, \boldsymbol{T})$ possesses continuous partial derivatives of the second order with respect to $\boldsymbol{\theta}$ at each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, for all values of $\boldsymbol{T}$.

Then it is easily seen that

$$\dot{W}(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T}|\boldsymbol{h})=\boldsymbol{h}'\dot{\boldsymbol{W}}(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T})$$

and

$$\ddot{W}(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T}|\boldsymbol{h})=\boldsymbol{h}'\ddot{\boldsymbol{W}}(\boldsymbol{\theta}+v\boldsymbol{h}, \boldsymbol{T})\boldsymbol{h} ,$$

where

$$\dot{\boldsymbol{W}}(\boldsymbol{\theta}, \boldsymbol{T})=(\partial W(\boldsymbol{\theta}, \boldsymbol{T})/\partial\theta_1, \cdots, \partial W(\boldsymbol{\theta}, \boldsymbol{T})/\partial\theta_k)'$$

and

$$\ddot{\boldsymbol{W}}(\boldsymbol{\theta}, \boldsymbol{T})=[\partial^2 W(\boldsymbol{\theta}, \boldsymbol{T})/\partial\theta_i\partial\theta_j;\ i, j=1, \cdots, k] .$$

Therefore, under conditions (i) and (ii)$'$, we have the following multi-parameter version of Theorem 3.1 of [1].

THEOREM 2.1$'$. *Suppose that for every $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ and every $\boldsymbol{h}\in\boldsymbol{R}^k$ with unit length, both matrices $\mathrm{E}_{\boldsymbol{\theta}+v_1\boldsymbol{h}}\dot{\boldsymbol{W}}(\boldsymbol{\theta}+v_2\boldsymbol{h}, \boldsymbol{T})\dot{\boldsymbol{W}}(\boldsymbol{\theta}+v_2\boldsymbol{h}, \boldsymbol{T})'$ and $\mathrm{E}_{\boldsymbol{\theta}+v_1\boldsymbol{h}}$ $\ddot{\boldsymbol{W}}(\boldsymbol{\theta}+v_2\boldsymbol{h}, \boldsymbol{T})$ are continuous functions of $v_1, v_2\in V(\boldsymbol{\theta}, \boldsymbol{h})$. Then*

$$D(\boldsymbol{\theta}; \boldsymbol{T}, W|\boldsymbol{h})=\boldsymbol{h}'\,\mathrm{E}_{\boldsymbol{\theta}}\,\ddot{\boldsymbol{W}}(\boldsymbol{\theta}, \boldsymbol{T})\boldsymbol{h}/[\boldsymbol{h}'\,\mathrm{E}_{\boldsymbol{\theta}}\,\dot{\boldsymbol{W}}(\boldsymbol{\theta}, \boldsymbol{T})\dot{\boldsymbol{W}}(\boldsymbol{\theta}, \boldsymbol{T})'\boldsymbol{h}]^{1/2}$$

and inequalities of (2.2) *hold.*

Theorem 2.1 implies that the discrimination rate of any L-unbiased estimator cannot exceed the intrinsic sensitivity of the family of probability measures under consideration. In [1], we have discussed the attainability of the upper bound. We now extend the definition of efficiency of an L-unbiased estimator as follows.

DEFINITION 2.3. For an L-unbiased estimator $\boldsymbol{T}$ with respect to W, we shall define its *efficiency* in the direction $\boldsymbol{h}$ at $\boldsymbol{\theta}$ by

$$e(\boldsymbol{\theta}; \boldsymbol{T}, W|\boldsymbol{h})=D(\boldsymbol{\theta}; \boldsymbol{T}, W|\boldsymbol{h})/\{\sqrt{n}\,s(\boldsymbol{\theta}|\boldsymbol{h})\} ,$$

if $s(\boldsymbol{\theta}|\boldsymbol{h})\neq 0$, where $s(\boldsymbol{\theta}|\boldsymbol{h})$ is the sensitivity of $\{P_{\boldsymbol{\theta}};\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\}$ under consideration.

Remark 2.1. We say that the family $\{P_{\boldsymbol{\theta}};\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\}$ is *smooth* at $\boldsymbol{\theta}$ (Pitman [3], p. 50) whenever for every $\boldsymbol{h}\in\boldsymbol{R}^k$ with unit length, the one-parameter family $\{P_{\boldsymbol{\theta}+v\boldsymbol{h}};\ v\in V(\boldsymbol{\theta}, \boldsymbol{h})\}$ is smooth at $v=0$, that is, if

$$4\lim_{v\to 0}\rho^2(P_{\boldsymbol{\theta}+v\boldsymbol{h}}, P_{\boldsymbol{\theta}})/v^2=\int\left\{\frac{\partial}{\partial v}f(x, \boldsymbol{\theta}+v\boldsymbol{h})|_{v=0}\right\}^2\Big/f(x, \boldsymbol{\theta})d\mu<\infty .$$

Define

$$\begin{aligned}\boldsymbol{I}(\boldsymbol{\theta})&=[\boldsymbol{I}_{ij}(\boldsymbol{\theta});\ i, j=1, \cdots, k]\\ &=\left[\int\frac{\partial}{\partial\theta_i}f(x, \boldsymbol{\theta})\frac{\partial}{\partial\theta_j}f(x, \boldsymbol{\theta})/f(x, \boldsymbol{\theta})d\mu;\ i, j=1, \cdots, k\right] ,\end{aligned}$$

the information matrix. Pitman gave the following sufficient conditions for smoothness:

(A-1) $f(x,\boldsymbol{\theta})$ is smooth with respect to each θ_i, $i=1,\cdots,k$ at every point in $\boldsymbol{\Theta}$,

(A-2) each partial derivative $(\partial/\partial\theta_i)f(x,\boldsymbol{\theta})$, $i=1,\cdots,k$ is *loosely* continuous in $\boldsymbol{\theta}$ (see p. 98 of [3] for definition of "loosely"),

(A-3) each $I_{ii}(\boldsymbol{\theta})$, $i=1,\cdots,k$ is a continuous function of $\boldsymbol{\theta}$ in $\boldsymbol{\Theta}$.

Under these conditions, the following relation holds:

$$s(\boldsymbol{\theta}\,|\,\boldsymbol{h})=[\boldsymbol{h}'\boldsymbol{I}(\boldsymbol{\theta})\boldsymbol{h}]^{1/2} \tag{2.3}$$

(see Theorem in Chapter 6 of [3]). Using arguments similar to those in Section 2, Remark (i) of [1] and pp. 50–51 of [3], we can prove smoothness and (2.3) under conditions (A-3) and

(B-1) each partial derivative $(\partial/\partial\theta_i)f(x,\boldsymbol{\theta})$, $i=1,\cdots,k$ is continuous in $\boldsymbol{\theta}$ for μ-almost all x.

Note that (A-3) implies (A-1) when (B-1) is satisfied.

3. Applications

In this section, we apply the result of the preceding section to loss functions of quadratic form. Let g: $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_k)'\to\boldsymbol{R}^1$ be a parametric function to be estimated and let it be partially differentiable. Let $W_0(\boldsymbol{\theta},T)=\alpha\{T-g(\boldsymbol{\theta})\}^2$ be the loss incurred by an estimator T, where α is a positive constant. Then we have the following result.

COROLLARY 3.1. *Suppose that the estimator T is L-unbiased with respect to W_0, and that $\mathrm{E}_{\boldsymbol{\theta}}T\in\{g(\boldsymbol{\tau});\ \boldsymbol{\tau}\in\boldsymbol{\Theta}\}$ for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$. If for every $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ and every $\boldsymbol{h}\in\boldsymbol{R}^k$ with unit length, $\mathrm{E}_{\boldsymbol{\theta}+v\boldsymbol{h}}T^2$ is a continuous function of $v\in V(\boldsymbol{\theta},\boldsymbol{h})$, then*

$$D(\boldsymbol{\theta};T,W_0\,|\,\boldsymbol{h})=\sqrt{\alpha}\,|\boldsymbol{h}'\nabla g(\boldsymbol{\theta})|/[\mathrm{E}_{\boldsymbol{\theta}}W_0(\boldsymbol{\theta},T)]^{1/2}\leqq\sqrt{n}\,s(\boldsymbol{\theta}\,|\,\boldsymbol{h})\ , \tag{3.1}$$

where we assume that

$$\nabla g(\boldsymbol{\theta})=(\partial g(\boldsymbol{\theta})/\partial\theta_1,\cdots,\partial g(\boldsymbol{\theta})/\partial\theta_k)'\neq 0\ .$$

Furthermore, under suitable conditions which yield $s(\boldsymbol{\theta}\,|\,\boldsymbol{h})=[\boldsymbol{h}'\boldsymbol{I}(\boldsymbol{\theta})\boldsymbol{h}]^{1/2}$ (see Remark 2.1), if $\boldsymbol{I}(\boldsymbol{\theta})$ is positive definite, then we have

$$\max_{\boldsymbol{h}} e(\boldsymbol{\theta};T,W_0\,|\,\boldsymbol{h})=\left[\frac{\alpha\nabla g(\boldsymbol{\theta})'\boldsymbol{I}(\boldsymbol{\theta})^{-1}\nabla g(\boldsymbol{\theta})}{n\,\mathrm{E}_{\boldsymbol{\theta}}W_0(\boldsymbol{\theta},T)}\right]^{1/2}, \tag{3.2}$$

and the maximum is attained at

$$\boldsymbol{h}=\lambda\boldsymbol{I}(\boldsymbol{\theta})^{-1}\nabla g(\boldsymbol{\theta})\ ,$$

where λ is a normalizing constant. Consequently, under suitable condi-

tions, we have the Cramér-Rao inequality

$$\mathrm{E}_{\boldsymbol{\theta}} W_0(\boldsymbol{\theta}, T) \geqq \frac{\alpha}{n} \nabla g(\boldsymbol{\theta})' \boldsymbol{I}(\boldsymbol{\theta})^{-1} \nabla g(\boldsymbol{\theta}) .$$

PROOF. The proof of (3.1) is similar to that of Corollary 3.1 in [1]. This is also derived from Theorem 2.1′ if $g(\boldsymbol{\theta})$ possesses partial derivatives of the second order. Indeed, using the relations

$$\dot{W}_0(\boldsymbol{\theta}, T) = -2\alpha\{T - g(\boldsymbol{\theta})\} \nabla g(\boldsymbol{\theta}) ,$$

$$\ddot{W}_0(\boldsymbol{\theta}, T) = 2\alpha \nabla g(\boldsymbol{\theta}) \nabla g(\boldsymbol{\theta})' - 2\alpha\{T - g(\boldsymbol{\theta})\} \times [\partial^2 g(\boldsymbol{\theta})/\partial\theta_i \partial\theta_j ;\ i, j = 1, \cdots, k]$$

and $\mathrm{E}_{\boldsymbol{\theta}}\{T - g(\boldsymbol{\theta})\} = 0$, we can easily see that (3.1) holds. The result (3.2) is readily obtained by the Cauchy-Schwarz inequality, that is,

$$|\boldsymbol{h}' \nabla g(\boldsymbol{\theta})| \leqq [\boldsymbol{h}' \boldsymbol{I}(\boldsymbol{\theta}) \boldsymbol{h}]^{1/2} [\nabla g(\boldsymbol{\theta})' \boldsymbol{I}(\boldsymbol{\theta})^{-1} \nabla g(\boldsymbol{\theta})]^{1/2} ,$$

with equality when $\boldsymbol{h} \propto \boldsymbol{I}(\boldsymbol{\theta})^{-1} \nabla g(\boldsymbol{\theta})$. Thus we complete the proof.

Now we discuss several features of discrimination rates of multi-parameter cases. Let T_i, $i=1, \cdots, m$ be estimators of $g_i(\boldsymbol{\theta})$, $i=1, \cdots, m$, respectively, where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)'$. Put $\boldsymbol{T} = (T_1, \cdots, T_m)'$ and $\boldsymbol{g}(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), \cdots, g_m(\boldsymbol{\theta}))'$. We now consider the following two loss functions,

$$W_0(\boldsymbol{\theta}, \boldsymbol{T}) = \boldsymbol{c}' \boldsymbol{M}(\boldsymbol{\theta}, \boldsymbol{T}) \boldsymbol{c}$$

and

$$W_1(\boldsymbol{\theta}, \boldsymbol{T}) = \boldsymbol{c}' \boldsymbol{D}(\boldsymbol{\theta}, \boldsymbol{T}) \boldsymbol{c} ,$$

for $\boldsymbol{c} = (c_1, \cdots, c_m)' \in \boldsymbol{R}^m$ such that $\boldsymbol{c} \neq 0$, where

$$\boldsymbol{M}(\boldsymbol{\theta}, \boldsymbol{T}) = (\boldsymbol{T} - \boldsymbol{g}(\boldsymbol{\theta}))(\boldsymbol{T} - \boldsymbol{g}(\boldsymbol{\theta}))'$$

and

$$\boldsymbol{D}(\boldsymbol{\theta}, \boldsymbol{T}) = \mathrm{diag}\,[\{T_1 - g_1(\boldsymbol{\theta})\}^2, \cdots, \{T_m - g_m(\boldsymbol{\theta})\}^2] .$$

We assume the following conditions:

(C-1) $\mathrm{E}_{\boldsymbol{\theta}} T_i \in \{g_i(\boldsymbol{\tau});\ \boldsymbol{\tau} \in \boldsymbol{\Theta}\}$, $i=1, \cdots, m$, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

(C-2) $\mathrm{E}_{\boldsymbol{\theta}+v\boldsymbol{h}} T_i T_j$, $i, j = 1, \cdots, m$ are all continuous functions of $v \in V(\boldsymbol{\theta}, \boldsymbol{h})$, for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and every $\boldsymbol{h} \in \boldsymbol{R}^k$ with unit length,

(C-3) $g_i(\boldsymbol{\theta})$, $i=1, \cdots, m$ are all partially differentiable at every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Suppose that $\boldsymbol{T}$ is L-unbiased with respect to both W_0 and W_1, for every $\boldsymbol{c} \in \boldsymbol{R}^m$. Then condition (C-1) implies that L-unbiasedness of $\boldsymbol{T}$ is equivalent to $\mathrm{E}_{\boldsymbol{\theta}} T_i = g_i(\boldsymbol{\theta})$, $i=1, \cdots, m$. Define

$$\dot{G}(\boldsymbol{\theta})=[\partial g_j(\boldsymbol{\theta})/\partial\theta_i;\ i=1,\cdots,k,\ j=1,\cdots,m]\ .$$

Applying Corollary 3.1 to $W_0(\boldsymbol{\theta},\boldsymbol{T})=\{\boldsymbol{c}'\boldsymbol{T}-\boldsymbol{c}'\boldsymbol{g}(\boldsymbol{\theta})\}^2$, we have

$$D(\boldsymbol{\theta};\boldsymbol{T},W_0|\boldsymbol{h})=|\boldsymbol{h}'\dot{G}(\boldsymbol{\theta})\boldsymbol{c}|/[\boldsymbol{c}'\,\mathrm{E}_{\boldsymbol{\theta}}\,\boldsymbol{M}(\boldsymbol{\theta},\boldsymbol{T})\boldsymbol{c}]^{1/2}\leqq\sqrt{n}\,s(\boldsymbol{\theta}|\boldsymbol{h})\ . \tag{3.3}$$

Moreover, if $s(\boldsymbol{\theta}|\boldsymbol{h})=[\boldsymbol{h}'\boldsymbol{I}(\boldsymbol{\theta})\boldsymbol{h}]^{1/2}$ and $\boldsymbol{I}(\boldsymbol{\theta})$ is positive definite, it follows from (3.2) that when $\boldsymbol{h}=\lambda\boldsymbol{I}(\boldsymbol{\theta})^{-1}\dot{G}(\boldsymbol{\theta})\boldsymbol{c}$,

$$\max_{\boldsymbol{h}} e(\boldsymbol{\theta};\boldsymbol{T},W_0|\boldsymbol{h})=\left[\frac{\boldsymbol{c}'\dot{G}(\boldsymbol{\theta})'\boldsymbol{I}(\boldsymbol{\theta})^{-1}\dot{G}(\boldsymbol{\theta})\boldsymbol{c}}{n\boldsymbol{c}'\,\mathrm{E}_{\boldsymbol{\theta}}\,\boldsymbol{M}(\boldsymbol{\theta},\boldsymbol{T})\boldsymbol{c}}\right]^{1/2}.$$

From this, the Cramér-Rao inequality

$$\boldsymbol{c}'\,\mathrm{E}_{\boldsymbol{\theta}}\,\boldsymbol{M}(\boldsymbol{\theta},\boldsymbol{T})\boldsymbol{c}\geqq\boldsymbol{c}'\dot{G}(\boldsymbol{\theta})'\boldsymbol{I}(\boldsymbol{\theta})^{-1}\dot{G}(\boldsymbol{\theta})\boldsymbol{c}/n\ ,\qquad\text{for all }\boldsymbol{c}\in\boldsymbol{R}^m$$

is derived. Now we treat $W_1(\boldsymbol{\theta},\boldsymbol{T})$. Using an argument similar to that in Corollary 3.1 of [1] (or using Theorem 2.1′ if each $g_i(\boldsymbol{\theta})$, $i=1,\cdots,m$ possesses the partial derivatives of order two), we have

$$\begin{aligned}D(\boldsymbol{\theta};\boldsymbol{T},W_1|\boldsymbol{h})&=\boldsymbol{h}'\dot{G}(\boldsymbol{\theta})C\dot{G}(\boldsymbol{\theta})'\boldsymbol{h}/[\boldsymbol{h}'\dot{G}(\boldsymbol{\theta})C\,\mathrm{E}_{\boldsymbol{\theta}}\,\boldsymbol{M}(\boldsymbol{\theta},\boldsymbol{T})C\dot{G}(\boldsymbol{\theta})'\boldsymbol{h}]^{1/2}\\&\leqq\sqrt{n}\,s(\boldsymbol{\theta}|\boldsymbol{h})\ ,\end{aligned} \tag{3.4}$$

where

$$C=\mathrm{diag}\,[c_1^2,c_2^2,\cdots,c_m^2]\ .$$

Let $\boldsymbol{S}=(S_1,\cdots,S_m)'$ be another L-unbiased estimator of $\boldsymbol{g}(\boldsymbol{\theta})$, which satisfies conditions (C-1) and (C-2). Furthermore we assume that

$$\mathrm{rank}\,\dot{G}(\boldsymbol{\theta})=m\ ,\qquad\text{for every }\boldsymbol{\theta}\in\boldsymbol{\Theta}\ .$$

Then from (3.3) and (3.4), we can easily see that

$$D(\boldsymbol{\theta};\boldsymbol{T},W_i|\boldsymbol{h})\geqq D(\boldsymbol{\theta};\boldsymbol{S},W_i|\boldsymbol{h})\ ,\qquad\text{for every }\boldsymbol{h}\in\boldsymbol{R}^k\text{ with unit length}$$

implies

$$\mathrm{E}_{\boldsymbol{\theta}}\,W_i(\boldsymbol{\theta},\boldsymbol{T})\leqq\mathrm{E}_{\boldsymbol{\theta}}\,W_i(\boldsymbol{\theta},\boldsymbol{S})\ ,$$

for each $i=0,1$. It is obvious that the converse is also true when $i=0$. In the case of $i=1$, however, the converse does not hold in general. The relation between discrimination rates and risk functions under general situations is still open to research.

We now apply the above discussion to estimation of parameters of a simple normal linear regression model.

Example 3.1. Let $\boldsymbol{Y}=(Y_1,\cdots,Y_n)'$ be a random vector distributed according to $N(\boldsymbol{X\theta},\sigma^2\boldsymbol{I})$, where

$$X=\begin{pmatrix}1, \cdots, 1\\ x_1, \cdots, x_n\end{pmatrix}' \quad \text{and} \quad \boldsymbol{\theta}=(\alpha, \beta)' \in \boldsymbol{R}^2 .$$

We assume that $\sum_{i=1}^{n} x_i=0$ and $V=\sum_{i=1}^{n} x_i^2 \neq 0$. Let $\boldsymbol{T}=(T_1, T_2)'$ be an estimator of $\boldsymbol{g}(\boldsymbol{\theta})=\boldsymbol{\theta}$. For each x, we consider the following two loss functions:

$$W_0(\boldsymbol{\theta}, \boldsymbol{T})=\{(T_1+T_2x)-(\alpha+\beta x)\}^2$$

and

$$W_1(\boldsymbol{\theta}, \boldsymbol{T})=(T_1-\alpha)^2+x^2(T_2-\beta)^2 .$$

Assume that $\mathrm{E}_{\boldsymbol{\theta}} \boldsymbol{T}=\boldsymbol{\theta}$ and $\mathrm{E}_{\boldsymbol{\theta}} T_i^2<\infty$, $i=1, 2$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, which guarantee the conditions (C-1) and (C-2). Let $\mathcal{U}$ be a class of all such estimators. We now consider the least squares estimator $\hat{\boldsymbol{\theta}}$, that is,

$$\hat{\boldsymbol{\theta}}=\left(\sum_{i=1}^{n} Y_i/n,\ \sum_{i=1}^{n} Y_i x_i/V\right)' .$$

Obviously, for each $i=0, 1$,

$$\mathrm{E}_{\boldsymbol{\theta}} W_i(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})=\sigma^2/n+x^2\sigma^2/V$$

and

$$\mathrm{E}_{\boldsymbol{\theta}} W_i(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) \leqq \mathrm{E}_{\boldsymbol{\theta}} W_i(\boldsymbol{\theta}, \boldsymbol{T}) , \qquad \text{for all } \boldsymbol{T} \in \mathcal{U} .$$

On the other hand, using (3.3) and (3.4), two discrimination rates of $\hat{\boldsymbol{\theta}}$ are obtained as follows:

$$D(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}, W_0 \mid \boldsymbol{h})=|h_1+h_2x|/[\sigma^2/n+x^2\sigma^2/V]^{1/2}$$

and

$$D(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}, W_1 \mid \boldsymbol{h})=(h_1^2+h_2^2x^2)/[h_1^2\sigma^2/n+h_2^2x^4\sigma^2/V]^{1/2} ,$$

where $\boldsymbol{h}=(h_1, h_2)'$ and $h_1^2+h_2^2=1$. We can easily verify that for each $i=0, 1$,

$$D(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}, W_i \mid \boldsymbol{h}) \geqq D(\boldsymbol{\theta}; \boldsymbol{T}, W_i \mid \boldsymbol{h}) , \qquad \text{for all } \boldsymbol{T} \in \mathcal{U} .$$

Thus for each $i=0, 1$ the least squares estimator $\hat{\boldsymbol{\theta}}$ is best in the sense of both criteria, that is, risk and discrimination rate. Now we examine the behavior of each $D(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}, W_i \mid \boldsymbol{h})$, $i=0, 1$ in terms of efficiency. The sensitivity $s(\boldsymbol{\theta} \mid \boldsymbol{h})$ of $N(\boldsymbol{X\theta}, \sigma^2\boldsymbol{I})$ is given by

$$s(\boldsymbol{\theta} \mid \boldsymbol{h})=[h_1^2n/\sigma^2+h_2^2V/\sigma^2]^{1/2} .$$

From Definition 2.3, we have

$$e(\boldsymbol{\theta}\,;\hat{\boldsymbol{\theta}},W_0|\boldsymbol{h})=\frac{|h_1+h_2x|}{[(h_1+h_2x)^2+(h_1x\sqrt{n/V}-h_2\sqrt{V/n})^2]^{1/2}}$$

and

$$e(\boldsymbol{\theta}\,;\hat{\boldsymbol{\theta}},W_1|\boldsymbol{h})=\frac{h_1^2+h_2^2x^2}{[(h_1^2+h_2^2x^2)^2+h_1^2h_2^2(x^2\sqrt{n/V}-\sqrt{V/n})^2]^{1/2}}\,.$$

In Figures 3.1–3.3, we describe graphs of both efficiencies as functions of h_2 under several situations. It is interesting to note the behavior

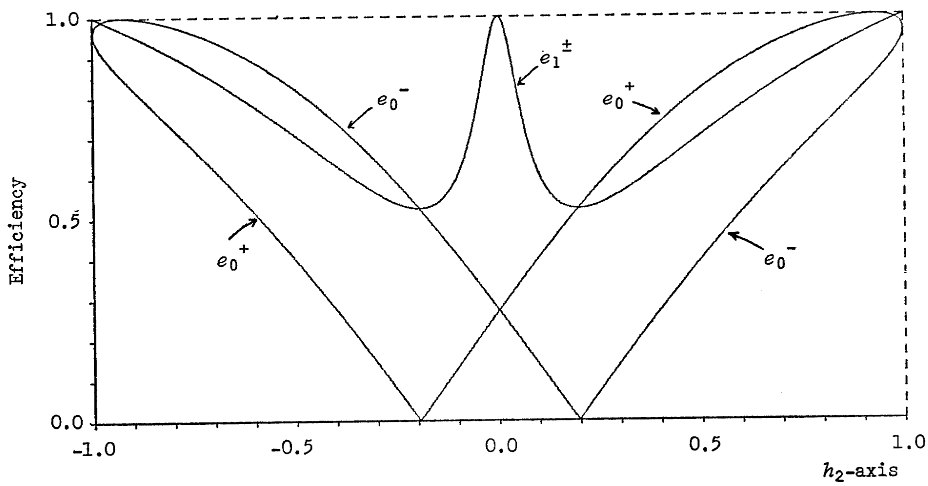

Fig. 3.1. $x=5.0$, $V=20.0$

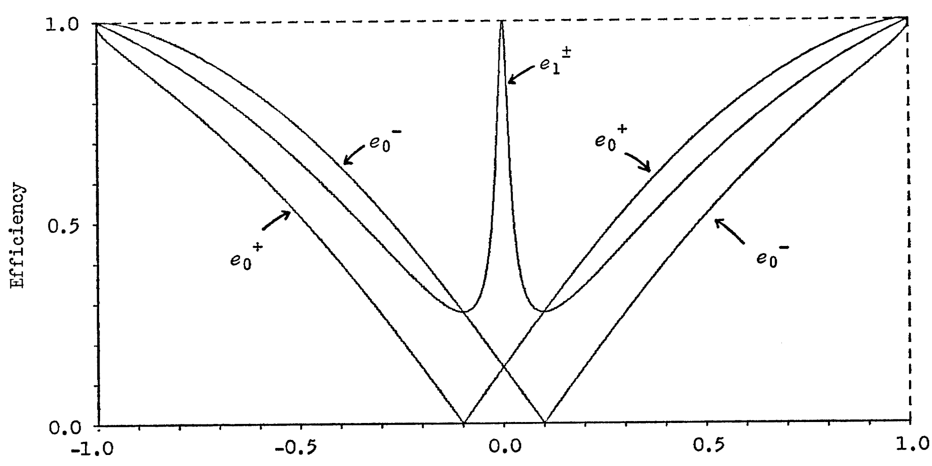

Fig. 3.2. $x=10.0$, $V=20.0$

Fig. 3.1.-3.3. The efficiencies $e(\boldsymbol{\theta};\hat{\boldsymbol{\theta}},W_0|\boldsymbol{h})$ and $e(\boldsymbol{\theta};\hat{\boldsymbol{\theta}},W_1|\boldsymbol{h})$ when $n=10$.
Define $e_i^+(h_2)=e(\boldsymbol{\theta};\hat{\boldsymbol{\theta}},W_i|\boldsymbol{h})$, when $h_1=\sqrt{1-h_2^2}$, $-1\leqq h_2\leqq 1$ and $e_i^-(h_2)=e(\boldsymbol{\theta};\hat{\boldsymbol{\theta}},W_i|\boldsymbol{h})$, when $h_1=-\sqrt{1-h_2^2}$, $-1\leqq h_2\leqq 1$, where $i=0,1$.

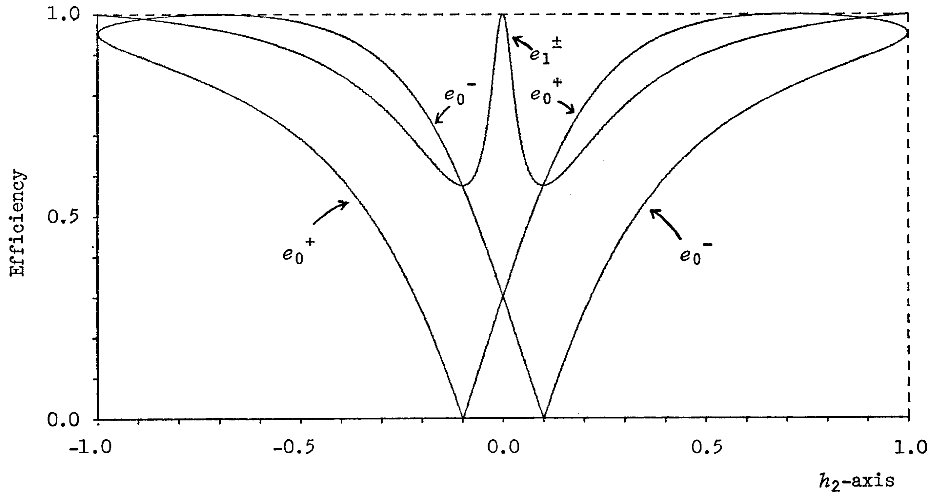

Fig. 3.3. $x=10.0$, $V=100.0$

of $e(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}, W_1|\boldsymbol{h})$ when x^2 or V changes. We can easily verify that as a function of x^2, $e(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}, W_1|\boldsymbol{h})$ monotonically increases when $x^2 \leqq V/n$ and monotonically decreases when $x^2 \geqq V/n$ for every fixed $\boldsymbol{h}$ (Fig. 3.1 and Fig. 3.2), and that as a function of V, $e(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}, W_1|\boldsymbol{h})$ monotonically increases when $V \leqq nx^2$ and monotonically decreases when $V \geqq nx^2$ for every fixed $\boldsymbol{h}$ (Fig. 3.2 and Fig. 3.3). When $x^2 = V/n$, $e(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}, W_1|\boldsymbol{h}) \equiv 1$ for all $\boldsymbol{h}$. On the other hand, $\mathrm{E}_{\boldsymbol{\theta}}\, W_1(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ is an increasing function of x^2 and is a decreasing function of V. The behavior of $e(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}, W_0|\boldsymbol{h})$ as a function of x or V is not so simple as that of $e(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}, W_1|\boldsymbol{h})$.

4. Concluding comment

For any L-unbiased estimator, we can consider two indices of its performance, that is, the risk function and the discrimination rate. The estimator is desirable when its risk function is small and its discrimination rate is large. In general, these two indices are not equivalent except when the loss is of the form $W_0(\boldsymbol{\theta}, \cdot) = \alpha\{\cdot - g(\boldsymbol{\theta})\}^2$. Every discrimination rate is bounded by the sensitivity of the family $\{P_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ under consideration. The Cramér-Rao inequality is also interpreted as an inequality with respect to the discrimination rate when the loss is W_0.

Acknowledgment

The author wishes to thank the referee for his careful reading and useful comments.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Kuboki, H. (1982). Unbiased estimators in the sense of Lehmann and their discrimination rates, *Ann. Inst. Statist. Math.*, **34**, A, 19-37.
[2] Lehmann, E. L. (1959). *Testing Statistical Hypotheses*, John Wiley & Sons, New York.
[3] Pitman, E. J. G. (1979). *Some Basic Theory for Statistical Inference*, Chapman and Hall, London.