

The Estimating Equations Induced from the Minimum Distance Estimation¹⁾

Ro Jin Pak²⁾

Abstract

This article presents a new family of the estimating functions related with minimum distance estimations, and discusses its relationship to the family of the minimum density power divergence estimating equations. Two representative minimum distance estimations; the minimum L_2 distance estimation and the minimum Hellinger distance estimation are studied in the light of the theory of estimating equations. Despite of the desirable properties of minimum distance estimations, they are not widely used by general researchers, because theories related with them are complex and are hard to be computationally implemented in real problems. Hopefully, this article would be a help for understanding the minimum distance estimations better.

Keywords : Density power divergence; estimating equations; Hellinger distance

1. Background and Motivation

In most of the previous works, the minimum distance estimation (MDE) has been studied by focusing on estimators, but in this article we focus rather on estimating equations which produce estimators. The theory of estimating equations has been well established since Godambe (1960), and has been a hot topic in the field of estimation for a while (Godambe and Thomson (1974), Godambe (1976)). By investigating the MDE in the estimating equation's point of view, we are

1) The present research was conducted by the research fund of Dankook University in 2002.

2) Associate Professor, Department of Computer Sciences and Statistics, Dankook University, Seoul, Korea, 140-714.
E-mail : rjpak@dankook.ac.kr

given a new tool to tackle the MDE estimation to enhance its generality and applicability. In the section 2.1 we provide a general form of estimating equations based on MDE covering both the minimum Hellinger distance estimation (MHD estimation) and the minimum L2 distance estimation (ML2D estimation). The MHD estimation (section 2.2) and the ML2D estimation (section 2.3) are studied one by one in the light of the theories of estimating equations. We also show that the proposed estimating equations turn out to be as same as those obtained by minimizing the density power divergence (Basu et al., 1998).

2. Properties of estimation equations by minimum distance estimations

2.1 General Form

Given an i.i.d. random sample, X_1, X_2, \dots, X_n , having a density $f_\theta(x)$, let $f_n(x)$ be a density estimator for $f_\theta(x)$ such as

$$f_n(x) = \frac{1}{n} \sum \frac{1}{h} k\left(\frac{x - X_i}{h}\right),$$

where $k(\cdot)$ is a kernel and h is a bandwidth. Define a smoothed density as $f_\theta^*(x) = \int f_\theta(x-y)k(y)dy$ and consider a family of distances between a smoothed density and a corresponding density estimator indexed by a parameter β as

$$\left\{ \int (f_n^{1/\beta}(x) - f_\theta^{*1/\beta}(x))^2 dx, 1 \leq \beta \leq 2 \right\}$$

The distance we define here is differ from an usual distance by means of using a smoothed density in place of a density itself. Since $E[f_n(x)] = f_\theta^*(x)$, it really helps us to reduce fair amount of mathematical complexity in theories.

The Hellinger distance ($\beta=2$) and the squared distance (or L_2 distance ($\beta=1$)) are members of this family. The minimum distance estimator ($\hat{\theta}$) is defined as a solution to

$$\nabla_\theta \int (f_n^{1/\beta}(x) - f_\theta^{*1/\beta}(x))^2 dx = -2 \int (f_n^{1/\beta}(x) - f_\theta^{*1/\beta}(x)) \frac{1}{\beta} \frac{\nabla_\theta f_\theta^*(x)}{f_\theta^{*(1-1/\beta)}(x)} dx = 0,$$

where ∇_θ represents a derivative w. r. t. θ .

Theorem 2.1. For $1 \leq \beta \leq 2$

$$\int (f_n^{1/\beta}(x) - f_\theta^{*1/\beta}(x)) \frac{1}{\beta} \frac{\nabla_\theta f_\theta^*(x)}{f_\theta^{*(1-1/\beta)}(x)} dx \quad (1)$$

can be reduced to

$$\frac{1}{\beta} \int (f_n(x) - f_\theta^*(x)) f_\theta^{*(-1+1/\beta)}(x) \frac{\nabla_\theta f_\theta^*(x)}{f_\theta^{*(1-1/\beta)}(x)} dx.$$

Proof. Consider the following series expansion;

$$\begin{aligned} f_n^{1/\beta}(x) - f_\theta^{*1/\beta}(x) &= \frac{1}{\beta} (f_n(x) - f_\theta^*(x)) f_\theta^{*(-1+1/\beta)}(x) \\ &\quad + \frac{1}{2\beta} \left(-1 + \frac{1}{\beta}\right) (f_n(x) - f_\theta^*(x))^2 f_\theta^{*(-2+1/\beta)}(x) \\ &\quad + o((f_n(x) - f_\theta^*(x))^3). \end{aligned} \quad (2)$$

and put (2) into (1) then the second term in the resulting integral,

$$\begin{aligned} &\int \frac{1}{2\beta} \left(-1 + \frac{1}{\beta}\right) (f_n(x) - f_\theta^*(x))^2 f_\theta^{*(-2+1/\beta)}(x) \frac{\nabla_\theta f_\theta^*(x)}{f_\theta^{*(1-1/\beta)}(x)} dx \\ &= \frac{1}{2\beta} \left(-1 + \frac{1}{\beta}\right) \int \left(\frac{f_n(x) - f_\theta^*(x)}{f_\theta^*(x)} \right)^2 f_\theta^{*(-1+2/\beta)}(x) \nabla_\theta f_\theta^*(x) dx \\ &\rightarrow 0 \text{ in probability as } n \rightarrow \infty, \end{aligned}$$

since both $f_n(x) - f_\theta^*(x) / f_\theta^*(x) \rightarrow 0$ as $n \rightarrow \infty$ and that $f_\theta^{*(-1+2/\beta)}(x) \nabla_\theta f_\theta^*(x)$ is bounded for $-1 + 2/\beta \geq 0$. Similarly, we can show that the higher order terms go to 0 in probability. \square

Also, we can rewrite the second integral in Theorem 2.1,

$$\frac{1}{\beta} \int (f_n(x) - f_\theta^*(x)) f_\theta^{*(-1+1/\beta)}(x) \frac{\nabla_\theta f_\theta^*(x)}{f_\theta^{*(1-1/\beta)}(x)} dx$$

as

$$\frac{1}{n\beta} \sum \int \left\{ \frac{1}{h} k\left(\frac{x - X_i}{h}\right) - f_\theta^*(x) \right\} f_\theta^{*(-1+2/\beta)}(x) \nabla_\theta \log f_\theta^*(x) dx,$$

which leads to the following definition.

Definition 2.1. Define a general form of the estimating equation for $1 \leq \beta \leq 2$ as

$$g(y) = \int \left\{ \frac{1}{h} k\left(\frac{x - y}{h}\right) - f_\theta^*(x) \right\} f_\theta^{*(-1+2/\beta)}(x) \nabla_\theta \log f_\theta^*(x) dx. \quad (3)$$

If $\beta = 2$ then (3) becomes (5), which is an estimating equation producing efficient but non-robust estimators (section 2.2). If $\beta = 1$ then (3) becomes the estimating equation (6), which is an M-estimating equation producing a robust but not fully efficient estimator (section 2.3). \square

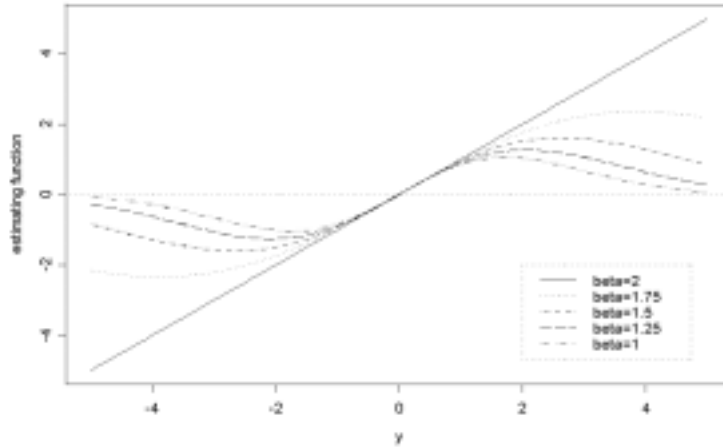


Figure 1. Estimating equations for μ .

Figure 1 displays the estimating equation $g(y)$ for different values of β , when a kernel is Gaussian and a density is $N(\mu, \sigma^2)$ with σ and h being fixed. When β is 2, $g(y) = y$, which is known to produce a non-robust but fully efficient estimator (so called, least squares estimator / MLE under Normal distribution). For $1 \leq \beta < 2$, $g(y)$ displays re-descending nature producing robust but not fully efficient estimators. It can be claimed that the value of β controls trade-off between efficiency and robustness.

Remark 2.1. The estimating equation based on a density power divergence (Basu, et al., 1998) can be reproduced by the estimating equation (3) as follows:

Consider the first part in (3). Let $z = (x - y)/h$ and do Taylor Series expansion about y , then we have

$$\begin{aligned} & \int \frac{1}{h} k\left(\frac{x-y}{h}\right) f_{\theta}^{*(-1+2/\beta)}(x) \nabla_{\theta} \log f_{\theta}^*(x) dx \\ &= \int k(z) f_{\theta}^{*(-1+2/\beta)}(y + hz) \nabla_{\theta} \log f_{\theta}^*(y + hz) dz \\ &= \int [k(z) f_{\theta}^{*(-1+2/\beta)}(y) + hz \nabla_y f_{\theta}^{*(-1+2/\beta)}(y) + \dots] [\nabla_{\theta} \log f_{\theta}^*(y) + hz \nabla_y \nabla_{\theta} \log f_{\theta}^*(y) + \dots] dz \\ &= f_{\theta}^{*(-1+2/\beta)}(y) \nabla_{\theta} \log f_{\theta}^*(y) \int k(z) dz + o(h) \\ &= f_{\theta}^{*(-1+2/\beta)}(y) \nabla_{\theta} \log f_{\theta}^*(y) + o(h). \end{aligned}$$

Hence, $g(y)$ in Definition 2.1 turns out

$$f_{\theta}^{*(-1+2/\beta)}(y) \nabla_{\theta} \log f_{\theta}^*(y) + \int f_{\theta}^{*(2/\beta)}(x) \nabla_{\theta} \log f_{\theta}^*(x) dx + o(h).$$

Since $f_{\theta}^*(x) \rightarrow f_{\theta}(x)$ as $h \rightarrow 0$, then $g(y)$ becomes

$$u_{\theta}(y) f_{\theta}^{(-1+2/\beta)}(y) - \int u_{\theta}(y) f_{\theta}^{2/\beta}(y) dy,$$

where $u_{\theta}(y)$ is the maximum likelihood score function. The above equation turns out to be as same as the $\phi(y, \theta)$ (Basu, et al., 1998) with $2/\beta = 1 + \alpha$:

$$\phi(y, \theta) = u_{\theta}(y) f_{\theta}^{\alpha}(y) - \int u_{\theta}(y) f_{\theta}^{1+\alpha}(y) dy.$$

Basu, et al. (1998) claimed that the degree of compromise between efficiency and robustness would be controlled by the tuning parameter α . Choices of α near zero, i. e. β near two, are known to afford considerable robustness while retaining efficiency close to that of maximum likelihood. \square

2.2 Hellinger Distance : $\beta=2$

The Minimum Hellinger distance estimator based on Basu and Lindsay (1994) is a solution to

$$\nabla_{\theta} \int (f_n^{1/2}(x) - f_{\theta}^{*1/2}(x))^2 dx = -2 \int (f_n^{1/2}(x) - f_{\theta}^{*1/2}(x)) \frac{\nabla_{\theta} f_{\theta}^*(x)}{f_{\theta}^{*1/2}(x)} dx = 0. \quad (4)$$

For $b \geq 0, a > 0$ we have an algebraic identity

$$b^{1/2} - a^{1/2} = (b - a) / [2a^{1/2} + (b - a)^2 / (2a^{1/2}(b^{1/2} + a^{1/2}))^2].$$

By combining this identity with (4), we have

$$\int (f_n^{1/2}(x) - f_{\theta}^{*1/2}(x)) \frac{\nabla_{\theta} f_{\theta}^*(x)}{f_{\theta}^{*1/2}(x)} dx = \frac{1}{2} \int (f_n(x) - f_{\theta}^*(x)) \frac{\nabla_{\theta} f_{\theta}^*(x)}{f_{\theta}^*(x)} dx + R_n,$$

where

$$R_n = - (f_n(x) - f_{\theta}^*(x))^2 / [2f_{\theta}^{*1/2}(x) (f_n^{1/2}(x) + f_{\theta}^{*1/2}(x))^2].$$

Suppose $R_n = o_p(1)$, which is in fact true by Beran (1977), then

$$\int (f_n^{1/2}(x) - f_{\theta}^{*1/2}(x)) \frac{\nabla_{\theta} f_{\theta}^*(x)}{f_{\theta}^{*1/2}(x)} dx \quad \text{and} \quad \frac{1}{2} \int (f_n(x) - f_{\theta}^*(x)) \frac{\nabla_{\theta} f_{\theta}^*(x)}{f_{\theta}^*(x)} dx$$

are asymptotically equivalent, and we have

$$\int (f_n(x) - f_{\theta}^*(x)) \frac{\nabla_{\theta} f_{\theta}^*(x)}{f_{\theta}^*(x)} dx = \frac{1}{n} \sum \int \frac{1}{h} k\left(\frac{x - X_i}{h}\right) \nabla_{\theta} \log f_{\theta}^*(x) dx.$$

Remark 2.2. The estimating equation based on the Hellinger distance,

$$\begin{aligned} g(y) &= \int \frac{1}{h} k\left(\frac{x-y}{h}\right) \nabla_{\theta} \log f_{\theta}^*(x) dx \\ &= \int \left\{ \frac{1}{h} k\left(\frac{x-y}{h}\right) - f_{\theta}^*(x) \right\} \nabla_{\theta} \log f_{\theta}^*(x) dx, \end{aligned} \quad (5)$$

which is the $g(y)$ in Definition 2.1 with $\beta=2$, is an unbiased optimal estimating equation when $h \rightarrow 0$. Proofs are in Appendix. \square

Example 2.1. Suppose that a kernel is Gaussian, and that the model density is $N(\mu, \sigma^2)$, then $f_{\theta}^*(x)$ is $N(\mu, h^2 + \sigma^2)$. Therefore, we have

$$g(y) = \int \frac{1}{h} k\left(\frac{x-y}{h}\right) \nabla_{\theta} \log f_{\theta}^*(x) dx = \left[\begin{array}{c} (y-\mu)/(h^2 + \sigma^2) \\ (-h^2 - \sigma^2 + (y-\mu)^2)/2(h^2 + \sigma^2)^2 \end{array} \right]$$

which is a vector of score functions of $N(\mu, \sigma^2)$, which is unbiased and optimal, in case of $h \rightarrow 0$. \square

2.3 L_2 Distance : $\beta=1$

The minimum L_2 distance estimator is solution to

$$\nabla_{\theta} \int (f_n(x) - f_{\theta}^*(x))^2 dx = -2 \int (f_n(x) - f_{\theta}^*(x)) \nabla_{\theta} f_{\theta}^*(x) dx = 0,$$

and we have

$$\int (f_n(x) - f_{\theta}^*(x)) \nabla_{\theta} f_{\theta}^*(x) dx = \sum \int \left\{ \frac{1}{h} k\left(\frac{x-X_i}{h}\right) - f_{\theta}^*(x) \right\} \nabla_{\theta} f_{\theta}^*(x) dx.$$

Remark 2.3. The estimating equation based on minimization of L_2 distance as

$$g(y) = \int \left\{ \frac{1}{h} k\left(\frac{x-y}{h}\right) - f_{\theta}^*(x) \right\} f_{\theta}^*(x) \nabla_{\theta} \log f_{\theta}^*(x) dx. \quad (6)$$

which is the $g(y)$ in Definition 2.1 with $\beta=1$, is a robust estimating function, but asymptotic optimality is not naturally given as shown as follows:

Define the matrices $J_g = E[gg^T]$ and $M_g = E[\partial g/\partial \theta]$, and then the efficiency matrix is $\text{Eff}(g) = M_g^{-1} J_g (M_g^{-1})^T$.

The expectation of the first derivative of the estimating equation in (6),

$$\begin{aligned} M_g &= E \left[\int \left\{ \frac{1}{h} k\left(\frac{x-y}{h}\right) - f_{\theta}^*(x) \right\} \nabla_{\theta}^2 f_{\theta}^*(x) dx - \int \nabla_{\theta} f_{\theta}^*(x) \nabla_{\theta} f_{\theta}^*(x) dx \right] \\ &= \int \left[E \left\{ \frac{1}{h} k\left(\frac{x-y}{h}\right) \right\} - f_{\theta}^*(x) \right] \nabla_{\theta}^2 f_{\theta}^*(x) dx - \int \nabla_{\theta} f_{\theta}^*(x) \nabla_{\theta} f_{\theta}^*(x) dx \\ &\rightarrow 0 - \int \nabla_{\theta} f_{\theta}^2(x) dx, \end{aligned}$$

which is not equal to $-I(\theta) (= E[\nabla_{\theta} \log f_{\theta}^*(x)^2])$ as $h \rightarrow 0$. \square

Example 2.2. Suppose that a kernel is Gaussian, and that f is $N(\mu, \sigma^2)$. After

dropping unnecessary constant factors, we have

$$g(y) = \left[\begin{array}{c} (y - \mu) \exp \left\{ \frac{-(y - \mu)^2}{2(2h^2 + \sigma^2)} \right\} \\ \left\{ -1 + \frac{(y - \mu)^2}{2h^2 + \sigma^2} \right\} \exp \left\{ \frac{-(y - \mu)^2}{2(2h^2 + \sigma^2)} \right\} + \left\{ \frac{2h^2 + \sigma^2}{2h^2 + 2\sigma^2} \right\}^{3/2} \end{array} \right].$$

We can not attain optimal asymptotic efficiency with $g(y)$ even if $h \rightarrow 0$, because $g(y)$ is far from the score function of the normal density, due to the factor $\exp(\cdot)$. □

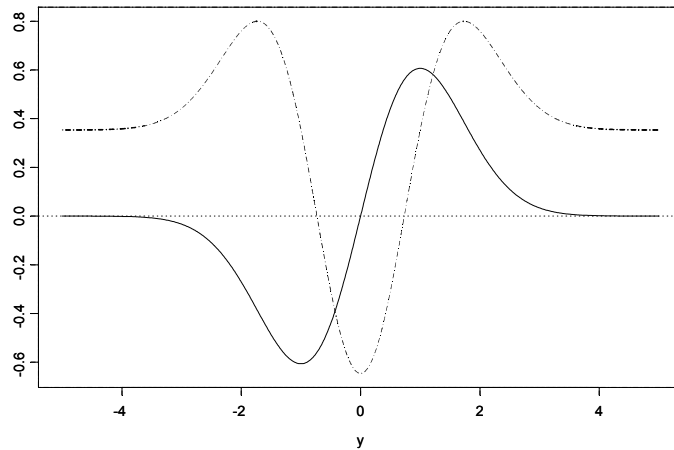


Figure 2. Estimating equations for μ (solid) and σ^2 (dotted).

Figure 2 displays estimating equations given in Example 2.2 with h and σ being fixed. They clearly show redescending nature of M-estimating functions. Each equation is of redescending types of M-estimating equations producing robust estimators for μ and σ^2 , respectively. Robustness of the estimators by the above estimating equation $g(y)$ is ensured, while losing some efficiency.

Remark 2.4. If we redefine $g_s(y) = g(y) E[\partial g(y)/\partial \theta]^{-1}$ then we have $E[g_s(y)] = 0$, since $E[h^{-1}k\{h^{-1}(x - y)\}] = f_\theta^*(x)$. Therefore, $g_s(y)$ is an unbiased estimating equation. Optimality is trivial because the standardized estimating equation is the optimal estimating equation, among unbiased estimating equations (Godambe, 1960). □

2.4 Optimality for a bandwidth

As we can notice from Example 2.1 and 2.2, the estimators based on $g(y)$ are M-estimators. Recall that an M-estimator is optimal V-robust (Hampel et al., 1986) when it minimizes $V(\phi, F)$ for a given upper bound on $\chi^*(\phi, F)$ (change-of-variance sensitivity) where ϕ is an M-estimating equation. Hall and Marron (1991) have shown that the best possible relative rate of convergence to the optimum for the data driven bandwidth selector is $n^{-1/2}$. Moreover, Fan and Marron (1992) have obtained the 'Fisher like' lower bound of the relative errors of bandwidth selector, which is given by

$$\sigma^2(f) = \frac{4}{25} \left[\frac{\int (f^{(4)})^2 f}{\{\int (f'')^2\}^2} - 1 \right].$$

Kim, Park and Marron (1994) give bandwidth selectors, h_n , which achieve the above best rate and best constant in the sense that for a fixed bandwidth h_0

$$n^{1/2}(h/h_0 - 1) \Rightarrow N(0, \sigma^2(f)).$$

Since $V(g(y, h), F)$ is in terms of h , we can discuss the optimality of h in the light of density estimation. Expand the asymptotic variance of an estimating equation in terms of the bandwidth for a fixed bandwidth h_0 ;

$$n\{V(g(y, h), F) - V(g(y, h_0), F)\} = (h - h_0) V'(g(y, h), F)|_{h=h_0} + \frac{1}{2}(h - h_0)^2 V''(g(y, h), F)|_{h=h_0} + \dots.$$

Suppose we neglect higher terms except the first order term, we have

$$n^{1/2} h_0^{-1} V(g(y, h), F) - V(g(y, h_0), F) \cong n^{1/2} (h/h_0 - 1) V'(g(y, h), F)|_{h=h_0} \Rightarrow N(0, \sigma^2(f) V'^2(g(y, h), F)|_{h=h_0}).$$

The variance is in the form of 'constant times $\sigma^2(f)$ ', so that optimality of h in our problem can be consider in the same manner as we deal with the optimality in density estimation. Therefore, for example, the plug-in type bandwidth proposed by Sheather and Jones (1991) would be a good optimal choice for h .

3. Conclusions

A new family of the estimating equations indexed by a parameter β covers various estimating equations producing robust and/or efficient estimators. The larger the β is, the more we get efficiency, and the smaller the β is, the more

we get robustness. The two representative minimum distance estimations; the minimum L_2 distance estimation and the minimum Hellinger distance estimation are studied by the proposed family of the estimating equations. The relation between the minimum distance estimation and the estimation by minimizing a density power divergence by Basu, et al. (1998) is presented. We hope that this article would be a little help for people to understand minimum distance methodology.

4. Appendix

First, the expectation of the estimating equation (5),

$$\begin{aligned} E[g(y)] &= E\left[\int \frac{1}{h} k\left(\frac{x-y}{h}\right) \nabla_{\theta} \log f_{\theta}^*(x) dx\right] = \int E\left[\frac{1}{h} k\left(\frac{x-y}{h}\right)\right] \nabla_{\theta} \log f_{\theta}^*(x) dx \\ &= \int f_{\theta}^*(x) \frac{\nabla_{\theta} f_{\theta}^*(x)}{f_{\theta}^*(x)} dx = \int f_{\theta}^*(x) \nabla_{\theta} \log f_{\theta}^*(x) dx \\ &= \int \nabla_{\theta} f_{\theta}^*(x) dx = \nabla_{\theta} \int f_{\theta}^*(x) dx = 0, \end{aligned}$$

which implies that the estimation function is unbiased.

Furthermore, since we have $f_{\theta}^*(x) \rightarrow f_{\theta}(x)$ as $h \rightarrow 0$

$$\begin{aligned} M_g &= E[\partial g / \partial \theta] = E\left[\int \frac{1}{h} k\left(\frac{x-y}{h}\right) \nabla_{\theta}^2 \log f_{\theta}^*(x) dx\right] \\ &= \int f_{\theta}^*(x) \nabla_{\theta}^2 \log f_{\theta}^*(x) dx = \int f_{\theta}^*(x) \left(\frac{\nabla_{\theta}^2 f_{\theta}^*(x) f_{\theta}^*(x) - (\nabla_{\theta} f_{\theta}^*(x))^2}{f_{\theta}^{*2}(x)}\right) dx \\ &= \int \nabla_{\theta}^2 f_{\theta}^*(x) dx - \int \left(\frac{\nabla_{\theta} f_{\theta}^*(x)}{f_{\theta}^*(x)}\right)^2 f_{\theta}^*(x) dx = 0 - \int \left(\frac{\nabla_{\theta} f_{\theta}^*(x)}{f_{\theta}^*(x)}\right)^2 f_{\theta}^*(x) dx \\ &\rightarrow - \int \left(\frac{\nabla_{\theta} f_{\theta}(x)}{f_{\theta}(x)}\right)^2 f_{\theta}(x) dx = - E[(\nabla_{\theta} \log f_{\theta}(x))^2] = - I(\theta) \end{aligned}$$

as $h \rightarrow 0$.

Hence, if $J_g = E[gg^T] \rightarrow I(\theta)$ as $h \rightarrow 0$, then $\text{Eff}(g) \rightarrow I^{-1}(\theta)$, so that $g(y)$ becomes an unbiased optimal estimating equation. \square

References

1. Basu, A. and Lindsay, B. G. (1994). Minimum disparity estimation for continuous models: efficiency, distributions and robustness, *Annals of the Institute of Statistica Mathematica*, 46, 683-705.
2. Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust

- and efficient estimation by minimizing a density power divergence, *Biometrika*, 85, 549–559.
3. Beran, R. J. (1977). Minimum Hellinger distance estimates for parametric models, *Annals of Statistics*, 5, 445–463.
 4. Fan, J. and Marron, J. S. (1992). On optimal data-based bandwidth selection in kernel density estimation, *Biometrika*, 78, 263–269.
 5. Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation, *Annals of Mathematical Statistics*, 31, 1208–1212.
 6. Godambe, V. P. (1976). Conditional likelihood and unconditional optimum estimating equations, *Biometrika*, 63, 277–284.
 7. Godambe, V. P. and Thomson, M. E. (1974). Estimating equations in the presence of a nuisance parameter, *Annals of Statistics*, 2, 568–571.
 8. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics, the Approach Based on Influence Functions*, John Wiley & Sons, New York.
 9. Hall, P. and Marron, J. S. (1991). Lower bounds for bandwidth selection in density estimation, *Probability Theory and Related Fields*, 90, 149–17.
 10. Kim, W. C., Park, B. U. and Marron, J. S. (1994). Asymptotically best bandwidth selectors in kernel density estimation, *Statistics and Probability Letters*, 19, 119–127.
 11. Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation, *Journals of the Royal Statistical Society, Series. B*, 53, 683–690.

[received date : Mar. 2003, accepted date : Jul. 2003]