

ASYMPTOTICALLY EFFICIENT ESTIMATORS OF LOCATION AND SCALE
PARAMETERS IN THE TWO-SAMPLE SEMIPARAMETRIC
LOCATION-SCALE MODEL

by

Byeong U. Park
University of North Carolina at Chapel Hill

November 1987

ABSTRACT

We observe X_1, \dots, X_m and Y_1, \dots, Y_n where X_i 's are i.i.d. with density g and Y_i 's are independent of X_i 's and i.i.d. with density $g((\cdot - \Delta)/\rho)$. This paper presents an asymptotically efficient estimator of (Δ, ρ) . The results of this paper are different from those of Park (1987) in two points. At first we consider unpaired two samples with different sample sizes, whereas Park (1987) dealt with paired sample cases. Secondly, we do not assume location-scale structure in X_i 's, while on the other hand he considered the case where X_i 's have that structure. It turns out that having location-scale structure in X_i 's or not, does not make any significant difference in terms of efficient score function but one may prefer our consideration when one is not quite sure about the structure of the first sample.

AMS 1980 Subject Classification: Primary 62G05, Secondary 62G20.

Key words and phrases: Semiparametric location-scale model, asymptotic lower bound, asymptotic linearity, efficient score function, efficient influence function.

1. Introduction

We observe X_1, \dots, X_n ; Y_1, \dots, Y_n where X_i 's are i.i.d. random variables having a common density g with respect to Lebesgue measure and Y_i 's are independent of X_i 's and i.i.d. with a common density $g\left[\frac{\cdot - \Delta}{\rho}\right]$. Our interest is in estimating the parameter (Δ, ρ) with minimal conditions on the density g . In this paper we present an asymptotically efficient estimator of (Δ, ρ) with minimal conditions on the density g .

A slightly different problem is as follows: X_i 's and Y_i 's are the same as the above except that now $g(\cdot)$ is of the form $g_0\left[\frac{\cdot - \mu}{a}\right]$, in other words, X_i 's have another common location-scale structure. This problem has a long history since Stein (1956) first investigated that one can estimate (Δ, ρ) as well with g_0 unknown as with g_0 known at the presence of unknown nuisance parameter (μ, a) , so called, adaptation is possible. Some earlier works of this sort include Weiss and Wolfowitz (1970) and Wolfowitz (1974). But they put too many conditions on the density g_0 . A completely definitive result was obtained by Park (1987) for the case $m=n$ where he used only the minimal conditions on the density g_0 , which is finite Fisher information, to get asymptotic efficiency. Problems with location shift only were considered by Bhattacharya (1967), van Eden (1970), Beran (1974, 1978) and Stone (1975).

Of course with the model described in the first paragraph in this section, adaptation is not possible because the score function for (Δ, ρ) is not orthogonal to the nuisance parameter scores (scores for g). But the efficient score function for (Δ, ρ) (for definition see Begun et al. (1983) or Bickel et al. (1986)), which determines the difficulty of the problem, is equivalent to that with the model described in the second paragraph where X_i ; i.i.d. $g_0\left[\frac{\cdot - \mu}{a}\right]$, Y_i ; i.i.d. $g_0\left[\frac{\cdot - \Delta}{a\rho} - \frac{\mu}{a}\right]$ up to constant term which depends on μ and a as we will see it in Section 2. Hence in terms of difficulty of estimating (Δ, ρ) , the two models are closely related. But in this paper we deal with the first because we may encounter the situation where we do not know even the structure of X_i 's.

Briefly speaking, our estimator of (Δ, ρ) will be constructed in section 3 as follows: First we find an initial estimator $(\hat{\Delta}_N, \hat{\rho}_N)$ which makes us be able to locate (Δ, ρ) with rate of convergence $N^{\frac{1}{2}}$, then we add an estimate of the efficient influence function to that. In fact, our estimator turns out to be asymptotically linear with the efficient influence function.

One final note is that we avoid considerable technical details in proving our main theorem because they can be referred to Park (1987), instead we will sketch the proof in section 4.

2. Preliminaries

First we start with stating the conditions we need.

(C1) $\frac{m}{m+n} \rightarrow p$ as $m, n \rightarrow \infty$.

(C2) g is absolutely continuous, has the first derivative \dot{g}

$$\text{and } \int (1+x^2)^{-2} \frac{(\dot{g})^2}{g} (x) dx < \infty .$$

Note that our model can be molded to one sample model by introducing a random indicator variable Z such that $P[Z=1]=p$. Hence, we observe i.i.d. $(X_1, Z_1), \dots, (X_N, Z_N)$ ($N=m+n$) with density function $f(x, z, \Delta, \rho, g)$ with respect to the product μ of Lebesgue measure ν and counting measure on $R^1 \times \{0, 1\}$, where

$$f(x, z, \Delta, \rho, g) = \left\{ \frac{1}{\rho} g \left[\frac{x-\Delta}{\rho} \right] \right\}^z \{g(x)\}^{1-z} p(z) ,$$

$$p(0) = p \text{ and } p(1) = 1-p = q.$$

2.1 \sqrt{N} - consistent estimate

At first we need to construct \sqrt{N} - consistent estimate $\tilde{\theta}_N = (\tilde{\Delta}_N, \tilde{\rho}_N)$ of $\theta = (\Delta, \rho)$ which satisfy

$$(2.1) \quad N^{\frac{1}{2}}(\tilde{\theta}_N - \theta) = O_{P_{\theta, g}}(1)$$

where $P_{\theta, g}$ is the probability measure having density $f(\cdot, \theta, g)$ with respect to the product measure μ . Let $\psi(x) = 2x/(1-x^2)$ and $\chi(x) = x\psi(x)-1$ and define the M-estimate $\tilde{\theta}_N = (\tilde{\Delta}_N, \tilde{\rho}_N)$ corresponding to ψ and χ to be any solution of

$$(2.2.a) \quad \frac{1}{n} \sum_{\{i; Z_i=1\}} \psi\left[\frac{X_i - \tilde{\Delta}_N}{\tilde{\rho}_N}\right] = \frac{1}{m} \sum_{\{i; Z_i=0\}} \psi\left[\frac{X_i - \tilde{\Delta}_N}{\tilde{\rho}_N}\right]$$

$$(2.2.b) \quad \frac{1}{n} \sum_{\{i; Z_i=1\}} \chi\left[\frac{X_i - \tilde{\Delta}_N}{\tilde{\rho}_N}\right] = \frac{1}{m} \sum_{\{i; Z_i=0\}} \chi\left[\frac{X_i - \tilde{\Delta}_N}{\tilde{\rho}_N}\right]$$

The choice of ψ and χ is rather arbitrary except that the system of equations

$$(2.3.a) \quad E\left[\psi\left[\frac{x-\Delta}{\rho}\right] \mid Z=1\right] = E[\psi(X) \mid Z=0]$$

$$(2.3.b) \quad E\left[\chi\left[\frac{x-\Delta}{\rho}\right] \mid Z=1\right] = E[\chi(X) \mid Z=0]$$

has a unique solution $\theta = (\Delta, \rho)$ and the estimate defined through the equation (2.2) converges to this solution with rate $N^{\frac{1}{2}}$. By Maronna (1976) these two requirements are met with our ψ and χ defined above (actually we need a simple two-sample generalization of the results of Maronna (1976)).

2.2 Efficient score function

Let \mathcal{G} be the collection of all probability densities on \mathbb{R}^1 with respect to Lebesgue measure ν and set $\mathcal{P}_2 = \{P_{\theta, g} \mid g \in \mathcal{G}\}$, the collection of all probability measures $P_{\theta, g}$, which has density $f(\cdot, \theta, g)$ with respect to μ , with g varying over \mathcal{G} and $\theta = (\Delta, \rho)$ fixed. As defined in Begun et al. (1983) and Bickel et al. (1986), the efficient score function for estimating θ is the orthogonal component of the score function for θ to the scores for nuisance parameter g . Now the score function for θ , denoted by $\dot{\ell} = (\dot{\ell}_1, \dot{\ell}_2)^\top$, is

$$\dot{\ell} = (\dot{\ell}_1, \dot{\ell}_2)^\top = \left[-\frac{z}{\rho} \frac{\dot{g}}{g}(\bar{x}), -\frac{z}{\rho} \{1 + \bar{x} \frac{\dot{g}}{g}(\bar{x})\} \right]^\top$$

where $\bar{x} = (x - \Delta)/\rho$. If we let \mathcal{J}_2 (following the notation in Bickel et al. (1986)) be the tangent space of \mathcal{P}_2 , the space of all scores for g (for definition of tangent space see Pfanzagl (1982) or Bickel et al. (1986)), it can be easily verified that

$$\mathcal{J}_2 = \{z h(\bar{x}) + (1-z)h(x) \mid \int h(x)g(x) dx = 0\}.$$

Now it simply involves a straightforward calculation to get

$$(2.4) \quad \Pi(\dot{\ell} \mid \mathcal{J}_2) = \left[\frac{-z}{\rho} \left\{ z \frac{\dot{g}}{g}(\bar{x}) + (1-z) \frac{\dot{g}}{g}(x) \right\}, \frac{-z}{\rho} \left\{ z(\bar{x} \frac{\dot{g}}{g}(\bar{x}) + 1) + (1-z)(x \frac{\dot{g}}{g}(x) + 1) \right\} \right]^\top.$$

where $\Pi(\cdot \mid \mathcal{J}_2)$ is the projection operator on \mathcal{J}_2 . Hence the efficient score function for estimating θ , denoted by $\dot{\ell}^*$, is

$$(2.5) \quad \begin{aligned} \dot{\ell}^*(x, z, \theta, g) &= \dot{\ell} - \Pi(\dot{\ell} \mid \mathcal{J}_2) \\ &= \frac{1}{\rho} \left[(p-1) z \frac{\dot{g}}{g}(\bar{x}) + p(1-z) \frac{\dot{g}}{g}(x), (p-1) z \left\{ \bar{x} \frac{\dot{g}}{g}(\bar{x}) + 1 \right\} + p(1-z) \left\{ x \frac{\dot{g}}{g}(x) + 1 \right\} \right]^\top. \end{aligned}$$

Remark 2.1. Suppose we observe i.i.d. $(X_1, Z_1), \dots, (X_N, Z_N)$ with density function $f(x, z, \mu, a, \Delta, \rho, g)$ with respect to μ where

$$f(x, z, \mu, a, \Delta, \rho, g) = \left\{ \frac{1}{a\rho} g\left[\frac{x-\Delta}{a\rho} - \frac{\mu}{a}\right] \right\}^z \left\{ \frac{1}{a} g\left[\frac{x-\mu}{a}\right] \right\}^{1-z} p(z),$$

$p(0) = p$, $p(1) = 1-p = q$ and $g \in \mathcal{G}_0$, the collection of all densities g with respect to Lebesgue measure ν which make (μ, a) identifiable. Then after going through a simple calculation we arrive at the following efficient score function.

$$(2.6) \quad \ell^*(x, z, \theta, g) = \frac{1}{\rho} \begin{bmatrix} \frac{1}{a} & 0 \\ \frac{\mu}{a} & 1 \end{bmatrix} \begin{bmatrix} (p-1)z \frac{\dot{g}}{g}(\bar{x}) + p(1-z) \frac{\dot{g}}{g}(\bar{x}) \\ (p-1)z \left\{ \bar{x} \frac{\dot{g}}{g}(\bar{x}) + 1 \right\} + p(1-z) \left\{ \bar{x} \frac{\dot{g}}{g}(\bar{x}) + 1 \right\} \end{bmatrix}$$

where $\bar{x} = (x-\mu)/a$ and $\bar{x} = (x-\Delta)/a\rho - \mu/a$. Comparing these two efficient score functions in (2.5) and (2.6), we can see that they are essentially the same except the constant matrix which depends on (μ, a) . Hence in terms of the difficulty of estimating (Δ, ρ) , the two models are closely related as one can expect it from their structures.

2.3 Asymptotic lower bound

Let

$$(2.7) \quad I_* = E_{P_{\theta, g}} \ell^* \ell^{*\top} (X, Z, \theta, g).$$

Then by Theorem 3.1 of Begun et al. (1983), which is an extended version of the representation theorem of Hájek (1970) for semiparametric cases and a special case of proposition 9 and 10 of Le Cam (1972), the limit law of every regular estimator (for definition see Begun et al. (1983) or Bickel et al. (1986)) may be represented as the convolution of $N(0, I_*^{-1})$ with some distribution which depends on (θ, g) only. Also Theorem 3.2 of Begun et al. (1983), which is a special case of Hájek-Le Cam-Millar asymptotic minimax theorem (see Hájek (1972), Le Cam (1972), (1979) and Millar (1979, 1984)), states that $E \ell(Z_*)$ is the asymptotic minimax bound for all "generalized

procedures" (see Millar (1979)) where $\ell: \mathbb{R}^2 \rightarrow \mathbb{R}^+$ is a bowl shaped loss function (i.e. $\ell(x) = \ell(-x)$, $\{x | \ell(x) \leq c\}$ is convex for every $c \geq 0$ and ℓ is continuous a.e. Lebesgue) and $Z_* \sim N(0, I_*^{-1})$. In particular, if ℓ is the squared error loss function, $\text{tr}(I_*^{-1})$ is the asymptotic minimax bound.

3. Main Result

Our main goal in this paper is to construct an estimate of $\theta = (\Delta, \rho)$ which is asymptotically normal with mean 0 and variance I_*^{-1} defined in section 2.3. This estimate will be obtained by adding an estimate of the efficient influence function $\frac{1}{N} \sum_{j=1}^N \ell^*(X_j, Z_j, \theta, g)$ to the initial estimate $\tilde{\theta}_N$. First of all we need to estimate the unknown density g . Let

$$(3.1) \quad \hat{g}_N(x, \theta) = b_N + \frac{1}{N b_N} \sum_{i=1}^N \left\{ Z_i K \left[\frac{x - X_i}{b_N} \right] + (1 - Z_i) K \left[\frac{x - X_i}{b_N} \right] \right\}$$

where $\bar{X}_i = (X_i - \Delta) / \rho$, $K(y) = e^{-y} / (1 + e^{-y})^2$, logistic density function and $\{b_N, N \geq 1\}$ is a sequence of positive constants converging to zero at a rate to be chosen later. Here the choice of the kernel function K is not critical but it should have some properties described in section 2, Chapter 1 of Park (1987). Now let

$$(3.2) \quad \hat{h}_N(x, \theta) = \left[\frac{\partial}{\partial x} \hat{g}_N(x, \theta) \right] / \hat{g}_N(x, \theta).$$

As an estimate of \dot{g}/g , \hat{h}_N may have good properties but as we can see it in the expression of ℓ^* in (2.5), \hat{h}_N also should serve as an estimate of $x(\dot{g}/g)(x)$ in the form of $x \hat{h}_N(x, \tilde{\theta}_N)$ which might be (actually turned out to be) uncontrollable even if \hat{h}_N has good properties as an estimate of \dot{g}/g . Hence we rather use a smoothly truncated version of \hat{h}_N defined by

$$(3.3) \quad \hat{h}_N(x, \theta) = \left\{ \left[\frac{\partial}{\partial x} \hat{g}_N(x, \theta) \right] / \hat{g}_N(x, \theta) \right\} \cdot b(c_N^{-1} x)$$

where $c_N \rightarrow \infty$ at a rate to be chosen later, $b(\cdot)$ is a symmetric function such that $b(x) = 1$ if $0 \leq x \leq 1$, $c(x)$ if $1 \leq x \leq 2$, 0 if $x \geq 2$ and $c(x)$ is a smooth function such that $c(1) = 1$, $c(2) = 0$, $c'(1) = c'(2) = c''(1) = c''(2) = 0$, $c'(x) \leq 0$, $|c'(x)| \leq H_1$ and $|c''(x)| \leq H_2$ for some positive numbers H_1 and H_2 whenever $1 \leq x \leq 2$. This smooth truncating function $b(\cdot)$ has been used in Park (1987). Let $\hat{\ell}^*(x, z, \theta)$ be an estimate of the efficient score function $\ell^*(x, z, \theta, g)$, defined by replacing (\hat{g}/g) with $\hat{h}_N(\cdot, \tilde{\theta}_N)$ and p with m/N in the expression of ℓ^* in (2.5). Namely,

$$(3.4) \quad \hat{\ell}^*(x, z, \theta) = \frac{1}{\rho} \left[-\frac{n}{N} z \hat{h}_N(\bar{x}, \tilde{\theta}_N) + \frac{m}{N} (1-z) \hat{h}_N(x, \theta_N), -\frac{n}{N} z \{\bar{x} \hat{h}_N(\bar{x}, \tilde{\theta}_N) + 1\} + \frac{m}{N} (1-z) \{x \hat{h}_N(x, \theta_N) + 1\} \right]^T$$

where $\bar{x} = (x - \Delta)/\rho$. With \hat{I}_* defined by

$$(3.5) \quad \hat{I}_* = \frac{1}{N} \sum_{j=1}^N \frac{\partial}{\partial \theta} \hat{\ell}^*(X_j, Z_j, \theta) \Big|_{\theta = \tilde{\theta}_N}$$

Here is our main theorem.

Theorem 3.1. If $Nb_N^{10} c_N^{-6} \rightarrow \infty$ and (C1) and (C2) are satisfied, then

$$(3.7) \quad N^{\frac{1}{2}}(\hat{\theta}_N - \theta_N) \rightarrow N(0, I_*^{-1})$$

in P_{θ_N, g_N} -probability for any sequence (θ_N, g_N) such that $|N^{\frac{1}{2}}(\theta_N - \theta) - t| \rightarrow 0$ and

$\|N^{\frac{1}{2}}(g_N^{\frac{1}{2}} - g^{\frac{1}{2}}) - h\|_{\nu} \rightarrow 0$ as $N \rightarrow \infty$ for some $t \in \mathbb{R}^2$ and $h \in L^2(\nu)$ where ν is Lebesgue measure

and $\|\cdot\|_{\nu}$ is the usual $L^2(\nu)$ -norm.

Note that

$$(3.8) \quad I_* = \frac{1}{\rho^2} p(1-p) \begin{bmatrix} I_{0,g} & I_{1,g} \\ I_{1,g} & I_{2,g} - 1 \end{bmatrix} = p(1-p) I_{**}$$

where $I_{a,g} = \int x^a \dot{g}^2/g \, dx$, $a=0,1,2$. Hence incorporating with the condition $m/(m+n) \rightarrow p$, an equivalent statement to (3.7) can be written as follows;

$$(3.9) \quad \sqrt{\frac{mn}{N}} (\hat{\theta}_N - \theta_N) \rightarrow N(0, I_{**}^{-1}).$$

The statement in (3.9) is more familiar to us than the statement in (3.7) in two-sample problems.

Remark 3.1. It is an immediate consequence of Theorem 3.1 that under the condition of Theorem 3.1, we have

$$(3.10) \quad P_{\theta,g} \{N(\hat{\theta}_N - \theta)^T \hat{I}_*(\hat{\theta}_N - \theta) \leq \chi^2(1-\alpha)\} \rightarrow 1-\alpha$$

for $0 < \alpha < 1$ where $\chi^2(1-\alpha)$ is the $100(1-\alpha)$ th percentile of chi-square distribution with degree of freedom 2. As discussed in Park (1987), the confidence region based on (3.10) is optimal in the sense that it has the smallest area asymptotically among all the confidence regions for θ based on regular estimates with the correct asymptotic coverage probability $(1-\alpha)$.

Our proof of Theorem 3.1 is based on Lemma 3.1. Before stating the lemma, we introduce an important property of estimates, namely asymptotic linearity.

Definition. An estimate $\hat{\theta}_N$ is called "asymptotically linear at (θ, g) " if

$$N^{1/2}(\hat{\theta}_N - \theta - \frac{1}{N} I_*^{-1} \sum_{j=1}^N \ell^*(X_j, Z_j, \theta, g)) = o_{P_{\theta,g}}(1)$$

where ℓ^* and I_* are defined in section 2.

Lemma 3.1. If $\hat{\theta}_N$ is asymptotically linear at (θ, g) , then the conclusion of Theorem 3.1 is true.

Proof. The lemma follows immediately from Le Cam's third lemma (see Hájek and Sidák (1967) Chapter 6). A quick alternative proof can be given as follows: From the asymptotic linearity of $\hat{\theta}_N$ it is easy to see that $(N^{\frac{1}{2}}(\hat{\theta}_N - \theta), \Lambda_N)$ converges weakly to Z with $Z \sim N(\mu, \Sigma)$ where $\Lambda_N = \sum_{j=1}^N \log \left[\frac{dP_{\theta_N, g_N}}{dP_{\theta, g}} \right] (X_j, Z_j)$ and

$$\mu = \begin{bmatrix} 0 \\ -\frac{1}{2} \sigma^2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} I_*^{-1} & t \\ t^\top & \sigma^2 \end{bmatrix}$$

for some $\sigma^2 > 0$. The characteristic function of $N^{\frac{1}{2}}(\hat{\theta}_N - \theta)$ under P_{θ_N, g_N} is

$$\begin{aligned} (3.11) \quad E_{P_{\theta_N, g_N}} \exp[i u^\top N^{\frac{1}{2}}(\hat{\theta}_N - \theta)] &= E_{P_{\theta_N, g_N}} \exp[i u^\top N^{\frac{1}{2}}(\hat{\theta}_N - \theta) - i u^\top t] + o(1) \\ &= E_{P_{\theta, g}} \exp[i u^\top N^{\frac{1}{2}}(\hat{\theta}_N - \theta) + \Lambda_N - i u^\top t] + o(1). \end{aligned}$$

By considering almost surely convergent versions of $(N^{\frac{1}{2}}(\hat{\theta}_N - \theta), \Lambda_N)$ and Z as in Beran (1977), the characteristic function in (3.11) converges to

$$(3.12) \quad E \exp[(i u^\top, 1) \cdot Z] \exp(-i u^\top t)$$

by Vitali's theorem. Since $E \exp[(i u^\top, 1) \cdot Z] = \exp[-\frac{1}{2} u^\top I_*^{-1} u + i u^\top t]$, the term in (3.12) is equal to $\exp[-\frac{1}{2} u^\top I_*^{-1} u]$. This concludes the proof. ■

4. Proof of Theorem 3.1

We avoid technical details here because the corresponding results can be easily derived from Park (1987) by putting $\mu = \tilde{\mu}_n = 0$ and $a = \tilde{a}_n = 1$ there.

Sketch of Proof. By Lemma 3.1, it is sufficient to show that $\hat{\theta}_N$ is asymptotically linear at (θ, g) . Put

$$\hat{I}(\theta) = \frac{1}{N} \sum_{j=1}^N \frac{\partial}{\partial \theta} \hat{\ell}^*(X_j, Z_j, \theta).$$

Note that $\hat{I}(\tilde{\theta}_N) = \hat{I}_*$ and

$$\begin{aligned} & N^{\frac{1}{2}} \left[\hat{\theta}_N - \theta - I_*^{-1} \frac{1}{N} \sum_{j=1}^N \ell^*(X_j, Z_j, \theta, g) \right] \\ &= \left\{ I + \hat{I}^{-1}(\tilde{\theta}_N) \frac{1}{N} \sum_{j=1}^N \frac{\partial}{\partial \theta} \hat{\ell}^*(X_j, Z_j, \theta) \Big|_{\theta = \tilde{\theta}_N} \right\} N^{\frac{1}{2}} (\tilde{\theta}_N - \theta) \\ &+ \left[\hat{I}^{-1}(\tilde{\theta}_N) - \hat{I}^{-1}(\theta) \right] N^{-\frac{1}{2}} \sum_{j=1}^N \hat{\ell}^*(X_j, Z_j, \theta) \\ &+ \hat{I}^{-1}(\theta) N^{-\frac{1}{2}} \sum_{j=1}^N \{ \hat{\ell}^*(X_j, Z_j, \theta) - \ell^*(X_j, Z_j, \theta, g) \} \\ &+ (\hat{I}^{-1}(\theta) - I_*^{-1}) N^{-\frac{1}{2}} \sum_{j=1}^N \ell^*(X_j, Z_j, \theta, g). \end{aligned}$$

Now it is easy to see that all the terms above go to zero if we show

$$(4.1) \quad N^{-\frac{1}{2}} \sum_{j=1}^N \{ \hat{\ell}^*(X_j, Z_j, \theta) - \ell^*(X_j, Z_j, \theta, g) \} \rightarrow 0,$$

$$(4.2) \quad \hat{I}(\theta) \rightarrow I_*$$

and

$$(4.3) \quad \sup_{|\theta' - \theta| \leq M} |\hat{I}(\theta') - \hat{I}(\theta)| \rightarrow 0$$

in $P_{\theta, g}$ -probability. We can easily derive

$$(4.4) \quad N^{-\frac{1}{2}} \sum_{\{Z_j=1\}} \{ \hat{\ell}^*(X_j, 1, \theta) - \ell^*(X_j, 1, \theta, g) \} \rightarrow 0,$$

and

$$(4.5) \quad N^{-\frac{1}{2}} \sum_{\{Z_j=0\}} \{ \hat{\ell}^*(X_j, 0, \theta) - \ell^*(X_j, 0, \theta, g) \} \rightarrow 0,$$

using the arguments in Lemma 3.1 ~ 3.5 in Park (1987). (4.1) follows from (4.4) and (4.5).

Now using the same arguments involved in the proofs of Lemma 4.1 ~ 4.11 and Theorem 4.1, 4.2 in Park (1987), we can get

$$(4.6) \quad -\frac{1}{N} \sum_{\{Z_j=1\}} \frac{\partial}{\partial \theta} \hat{\ell}^*(X_j, 1, \theta) = \frac{1}{\rho^2} (p-1)^2 p I_{**} + o_{P_{\theta, g}}(1)$$

and

$$(4.7) \quad -\frac{1}{N} \sum_{\{Z_j=0\}} \frac{\partial}{\partial \theta} \hat{\ell}^*(X_j, 0, \theta) = \frac{1}{\rho^2} p^2 (1-p) I_{**} + o_{P_{\theta, g}}(1)$$

where I_{**} is given through the equation (3.8). Putting (4.6) and (4.7) together leads to (4.1). Finally using the arguments in the proof of Theorem 4.6 in Park (1987), we can easily obtain

$$(4.8) \quad \sup_{|\theta' - \theta| \leq M} \left| \frac{1}{N} \sum_{\{Z_j=1\}} \frac{\partial}{\partial \theta'} \hat{\ell}^*(X_j, 1, \theta') - \frac{1}{N} \sum_{\{Z_j=1\}} \frac{\partial}{\partial \theta} \hat{\ell}^*(X_j, 1, \theta) \right| = o_{P_{\theta, g}}(1)$$

and

$$(4.9) \quad \sup_{|\theta' - \theta| \leq M} \left| \frac{1}{N} \sum_{\{Z_j=0\}} \frac{\partial}{\partial \theta'} \hat{\ell}^*(X_j, 0, \theta') - \frac{1}{N} \sum_{\{Z_j=0\}} \frac{\partial}{\partial \theta} \hat{\ell}^*(X_j, 0, \theta) \right| = o_{P_{\theta, g}}(1).$$

Now (4.3) follows from (4.8) and (4.9) immediately. ■

References

- Begun, J.M., Hall, W.J., Huang, W. and Wellner, J.A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* 11, 432-452.
- Beran, R. (1974). Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.* 2, 63-74.
- Beran, R. (1977). Robust location estimates. *Ann. Statist.* 5, 431-444.
- Beran, R. (1978). An efficient and robust adaptive estimator of location. *Ann. Statist.* 6, 292-313.
- Bhattacharya, P.K. (1967). Efficient estimation of a shift parameter from grouped data. *Ann. Math. Statist.* 38, 1770-1787.
- Bickel, P.J., Klaassen, C.J., Ritov, Y. and Wellner, J.A. (1986). Efficient and adaptive inference in semiparametric models. Preprint.
- Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. verw. Gebiete* 14, 323-330.
- Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* 1, 175-194. Univ. of California Press, Berkeley.
- Hájek, J. and Sidák, Z. (1967). *Theory of Rank Tests*, Academic, New York and Academia, Prague.
- Le Cam, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* 1, 245-261. Univ. of California Press, Berkeley.
- Le Cam, L. (1979). On a theorem of J. Hájek. In *Contribution to Statistics; Jaroslav Hájek Memorial Volume*, Ed. J. Jureckova, 119-135. Reidel, Dordrecht.
- Maronna, R. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* 4, 51-67.
- Millar, P.W. (1979). Asymptotic minimax theorems for the sample distribution function. *Z. Wahrsch. verw. Gebiete* 48, 233-252.
- Millar, P.W. (1984). A general approach to the optimality of minimum distance estimators. *Trans. of Amer. Math. Soc.* 286, 377-418.
- Park, B.U. (1987). Efficient estimation in the two-sample semiparametric location-scale model. Part of Ph.D. Thesis, University of California, Berkeley.
- Pfanzagl, J. (1982). *Contribution to A General Asymptotic Statistical Theory*. Lecture Notes in Statistics, Springer, New York.

- Stein, C. (1956). Efficient nonparametric testing and estimation. Proc. Third Berkeley Symp. Math. Statist. Probab. 1, 187-196. Univ. of California Press, Berkeley.
- Stone, C. (1975). Adaptive maximum likelihood estimators of a location parameter. Ann. Statist. 3, 267-284.
- van Eden, C. (1970). Efficiency-robust estimation of location. Ann. Math. Statist. 41, 172-181.
- Weiss, L. and Wolfowitz, J. (1970). Asymptotically efficient nonparametric estimators of location and scale parameters. Z. Wahrsch. verw. Gebiete 16, 134-150.
- Wolfowitz, J. (1974). Asymptotically efficient nonparametric estimators of location and scale parameters, II. Z. Wahrsch. verw. Gebiete 30, 117-128.