LOCAL ASYMPTOTIC NORMALITY FOR INDEPENDENT NOT IDENTICALLY DISTRIBUTED OBSERVATIONS IN SEMIPARAMETRIC MODELS

by

BYEONG U. PARK

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

October 1987

ABSTRACT

A set of conditions ensuring local asymptotic normality for independent but not necessarily identically distributed observations in semiparametric models is presented here. The conditions are turned out to be more direct and easier to verify than those of Oosterhoff and van Zwet (1979) in semiparametric models. Examples considered include the simple linear regression model and Cox's proportional hazards model without censoring where the covariates are not random.

AMS 1980 Subject Classification: Primary 62 G 05.

Key words and phrases: Local asymptotic normality, semiparametric models, uniform Hellinger differentiability.

1. Introduction.

In the theory of asymptotic estimation and testing hypotheses, local asymptotic normality (LAN for short) of a family of distributions has had an important role since its importance was introduced by Le Cam (1960). In particular, Hájek-Le Cam's convolution theorem (see Hájek (1970) and Le Cam (1972)) and asymptotic minimax theorem (see Hájek (1972) and Le Cam (1972), (1979)) was derived from LAN property of a family of distributions although Le Cam's results covered families which may not be locally asymptotically normal. Another but recent results of this sort were established by Begun et al. (1983) in semiparametric models.

In view of this importance of the notion of LAN, many authors presented sets of conditions for LAN. Very broad conditions ensuring LAN for independent identically distributed observations in parametric models were given in Hájek (1972). Begun et al. (1983) gave a set of conditions for LAN of independent identically distributed observations in semiparametric models and using LAN they established the representation theorem and the asymptotic minimax lower bounds for regular estimates.

However, in many statistical models, the observations are not homogenous even though they are independent. Examples include various regression models where the covariates are constant. In this paper we present a set of conditions ensuring LAN for independent but not necessarily identically distributed observations in semiparametric models. As Begun et al. (1983) did it, we will be able to use LAN for establishing the appropriate representation theorem and the asymptotic minimax theorem in this inhomogenous semiparametric situation. We shall return to this point in a subsequent paper.

For independent but non-identically distributed observations in parametric models, conditions ensuring LAN were given by Kushnir (1968), Phillipou and Roussas (1973) and Ibragimov and Khas'minskii (1975). Le Cam (1960), Oosterhoff and van Zwet (1979) also considered the independent not identically distributed case in general situations but their conditions are not so transparent to verify in semiparametric models.

Section 2 of the paper is devoted to recall the necessary definitions and notations. The main results including some examples will be provided in section 3 and the proof of Theorem 3.1 in section 3 will be given in section 4.

2. Definitions, Notation and Assumptions

For notational ease, first we will deal with the simplest but most important type of semiparametric model with a parametric component $\theta \in \Theta$, where Θ is an open set in R^1, and a single nonparametric component $g \in \mathcal{G}$ where $\mathcal{G}$ is a specified set of density functions. Our conditions and results for this one-dimensional parametric component can be extended for a multi-dimensional parametric component in the obvious manner, which will be sketched in the remarks at the end of section 3.

Let Θ be an open subset of R^1 and $\mathcal{G}$ be a specified set of density functions with respect to a σ-finite measure ν on some measurable space $(\mathcal{Z},\mathcal{D})$, let $P_{j,\theta,g}$, $j=1,\dots,n$ $(n=1,2,\dots)$ be a probability measures on the measurable space $(\mathcal{X}_j,\mathcal{B}_j)$. We assume that there is a σ-finite measure μ_j on $\mathcal{B}_j$ such that μ_j dominates $P_{j,\theta,g}$, $\theta \in \Theta$, $g \in \mathcal{G}$, $j \geq 1$, and let $f_j(\cdot,\theta,g) = d\,P_{j,\theta,g}/d\mu_j$ for a specified version of the Radon-Nikodym derivative involved. Define $(\mathcal{Y},\mathcal{A}) = \Pi_{j=1}^{\infty} (\mathcal{X}_j,\mathcal{B}_j)$ and let $P_{\theta,g}$ be the product measure of $P_{j,\theta,g}$, $j \geq 1$, induced on $\mathcal{A}$. Let $X_1,X_2,\dots,X_j \in \mathcal{X}_j$ be independent observations, the j-th of which has density $f_j(\cdot,\theta,g)$ with respect to μ_j on $\mathcal{B}_j$. $P^{(n)}_{\theta,g}$ is the projection of $P_{\theta,g}$ onto $\Pi^n_{j=1}(\mathcal{X}_j,\mathcal{B}_j)$ and $E^{(n)}_{\theta,g}$ is the corresponding expectation. The norm $\|\cdot\|_j = \|\ \|_{\mu_j}$ denote the usual $L^2(\mu_j)$-norm. In the sequel, for brevity of notation, we set $f_j(\theta,g)$ for the random variable $f_j(X_j,\theta,g)$.

For $\theta,\theta^* \in \Theta$ and $g,\ g^* \in \mathcal{G}$, let

(2.1) $$r_j(\theta,\theta^*;g,g^*) = r_j(\theta,\theta^*;g,g^*;X_j) = 2\left[\frac{f_j^{1/2}(\theta^*,g^*)}{f_j^{1/2}(\theta,g)} - 1\right].$$

These r_j's are by definition functions from $\mathcal{X}_j$ to $[-2,\infty]$. The value ∞ is taken on the singular part of P_{j,θ^*,g^*} with respect to $P_{j,\theta,g}$. Obviously, $r_j(\theta,\theta^*;g,g^*) \in L^2(P_{j,\theta,g})$. These terms r_j's will be useful in proving the main theorem in section 4. Now define

(2.2) $$\Lambda_n(\theta,\theta^*;g,g^*) = \log \frac{d\,P^{(n)}_{\theta^*,g^*}}{d\,P^{(n)}_{\theta,g}} = \log \Pi^n_{j=1} \frac{f_j(\theta^*,g^*)}{f_j(\theta,g)} .$$

In section 3 we will establish the asymptotic normality for $\Lambda_n(\theta,\theta_n : g,g_n)$ (so LAN) for the sequence (θ_n,g_n) such that

$$|n^{1/2}(\theta_n-\theta) - h| \to 0 \quad\text{and}\quad \|n^{1/2}(g_n^{1/2}-g^{1/2}) - \beta\|_\nu \to 0 \quad\text{as } n\to\infty$$

for some $h \in R^1$ and $\beta \in L^2(\nu)$ respectively. Let $\mathcal{B}$ be the collection of all such $\beta \in L^2(\nu)$ i.e.

(2.3) $$\mathcal{B} = \{\beta \in L^2(\nu) \mid \|n^{1/2}(g_n^{1/2}-g^{1/2}) - \beta\|_\nu \to 0 \quad \text{as } n\to\infty \text{ for some sequence } g_n \text{ with all } g_n \in \mathcal{G}\}.$$

throughout the paper we will rely on the following Assumption about $\mathcal{B}$ as it appears in Begun et al. (1983).

<u>ASSUMPTION S</u>. The set $\mathcal{B}$ defined in (2.3) is a subspace of $L^2(\nu)$.

In addition to the comments on this Assumption in Begun et al. (1983) we can say that this is equivalent to the condition that the tangent space is equal to the tangent set in Theorem 1, section 4, Chapter 3 of Bickel et al. (1986) (see Pfanzagl (1982) or Bickel et al. for the definition of tangent space and tangent set). Assumption S ensures that we can approach to g along two opposite directions and this fact will be used in the proof of Lemma 3.1 which concerns about the rate of convergence for the singular parts of $f_j(\theta_n,g_n)$ with respect to $f_j(\theta,g)$.

We conclude this section with the following definition of the uniform Hellinger differentiability of $\{f_j^{1/2}(\theta,g)\}$.

<u>DEFINITION 2.1</u>. (Uniform Hellinger differentiability of $\{f_j^{1/2}(\theta,g)\}$). The sequence of the root densities $\{f_j^{1/2}(\theta,g)\}$ is said to be uniformly Hellinger-differentiable at $(\theta,g) \in \Theta \times \mathcal{G}$ if there exist random functions $\rho_{j,\theta} \in L^2(\mu_j)$ and bounded linear operators B_j; $L^2(\nu) \to L^2(\mu_j)$ such that

$$\text{(2.4)} \quad \frac{\sup_{1\le j\le n} \|f_j^{1/2}(\theta_n,g_n)-f_j^{1/2}(\theta,g) - \{\rho_{j,\theta}(\theta_n-\theta) + B_j(g_n^{1/2}-g^{1/2})\}\|_{\mu_j}}{|\theta_n-\theta| + \|g_n^{1/2}-g^{1/2}\|_\nu} \to 0 \quad \text{as } n \to \infty$$

for all sequences (θ_n,g_n) such that $\theta_n \to \theta$ and $g_n^{1/2} \to g^{1/2}$ in $L^2(\nu)$ where $g_n \in \mathcal{G}$ for all $n \ge 1$.

Note that when $f_j \equiv f$, (2.4) reduces to Hellinger differentiability of $f^{1/2}$ defined in Begun et al. (1983).

3. Main Results

Let (θ_n,g_n) be any sequence such that

$$\text{(3.1)} \qquad |n^{1/2}(\theta_n-\theta)-h| \to 0 \quad \text{and} \quad \|n^{1/2}(g_n^{1/2}-g^{1/2})-\beta\|_\nu \to 0 \quad \text{as } n \to \infty$$

for some $h \in R^1$ and $\beta \in L^2(\nu)$ respectively. Then the following proposition is an immediate consequence of uniform Hellinger differentiability of $\{f_j^{1/2}(\theta,g)\}$.

PROPOSITION 3.1. Suppose $\{f_j^{1/2}(\theta,g)\}$ is uniformly Hellinger-differentiable at $(\theta,g) \in \Theta \times \mathcal{G}$. Then

(3.2) $$\Sigma_{j=1}^{n} \|f_j^{1/2}(\theta_n,g_n) - f_j^{1/2}(\theta,g) - n^{-1/2}\alpha_j\|_{\mu_j}^2 \to 0 \quad \text{as} \quad n \to \infty$$

where $\alpha_j \in L^2(\mu_j)$ is given by

(3.3) $$\alpha_j = h\,\rho_{j,\theta} + B_j\,\beta\ .$$

Note that if $f_j \equiv f$, then this proposition is reduced to Proposition 2.1 in Begun et al. (1983). In addition to (3.2) we have also

(3.4) $$\langle \alpha_j, f_j^{1/2} \rangle = 0 \quad \text{for any } j \geq 1$$

with $f_j^{1/2} \equiv f_j^{1/2}(\theta,g)$ under the uniform Hellinger differentiability condition. Our conditions for LAN will be based on (3.2) and (3.4) because those are weaker than the uniform Hellinger differentiability condition. Here are our conditions ensuring LAN for the family $P_{\theta,g}^{(n)}$.

(C1) (3.2) and (3.4) hold with α_j given by (3.3).

(C2) The random functions $\alpha_j f_j^{-1/2}$, where α_j is given by (3.3), satisfy Lindeberg's conditions

(3.5) $$\lim_{n\to\infty} \frac{1}{n} \Sigma_{j=1}^{n} E_{\theta,g}^{(n)}\{\alpha_j^2 f_j^{-1} I(|\alpha_j f_j^{-1/2}| > n^{1/2}\epsilon)\} = 0$$

for any $\epsilon > 0$. And

(3.6) $$\lim_{n\to\infty} \frac{1}{n} \Sigma_{j=1}^{n} \|\alpha_j\|_{\mu_j}^2 = \sigma^2\ .$$

Remark 3.1. For independent and identically distributed case (C2) is automatically satisfied.

Remark 3.2. The condition (C2) is closely related to the condition (3.7) in Oosterhoff and van Zwet (1979) which amounts to the Lindeberg's conditions applied to $r_j(\theta,\theta_n;g,g_n)$ defined by (2.1). However (C2) is easier and more direct to verify than (3.7) in their conditions when we know our derivatives α_j's. In fact, $\rho_{j,\theta}$ is typically just the usual parametric score function for θ, i.e. $\frac{1}{2}\left[\left(\frac{\partial}{\partial\theta}\right) \log f_j(\theta,g)\right] f_j^{1/2}(\theta,g)$ and $A_j\beta$ can be found heuristically by calculating the first derivative of $\log[f_j(\theta,g_\eta)]$ with respect to η at $\eta=0$ times $f_j^{1/2}(\theta,g)$ where $g_\eta = g + \eta\beta g^{1/2}$.

Remark 3.3. (A3) and (A4) in Phillipour and Roussas (1973), even though their results are for parametric models, are much stronger than (C2).

Lemma 3.1. Under Assumption S and (C1), we have

$$\lim_{n\to\infty} \Sigma_{j=1}^{n} \int_{\{x;\, f_j(x,\theta,g)=0\}} f_j(x,\theta_n,g_n)\mu_j(dx) = 0. \tag{3.7}$$

Proof. From Assumption S there exist $g_n^* \in \mathcal{G}$ for all $n \leq 1$ such that

$$\|n^{1/2}(g_n^{*1/2} - g^{1/2}) + c\beta\|_\nu \to 0 \quad \text{for some } c > 0.$$

Also since Θ is an open subset of R^1, $\theta_n^{**} = \theta+c(\theta-\theta_n) \in \Theta$ for sufficiently large n. Hence we can find $\theta_n^* \in \Theta$ for all $n \geq 1$ such that $|n^{1/2}(\theta_n^*-\theta) + ch| \to 0$. Hence from (C1) we get

$$\Sigma_{j=1}^{n} \|f_j^{1/2}(\theta_n^*,g_n^*) - f_j^{1/2}(\theta,g) + n^{-1/2}\, c\, \alpha_j\|_{\mu_j}^2 \to 0 \quad \text{as} \quad n \to \infty. \tag{3.8}$$

This together with (3.2) implies

$$\Sigma_{j=1}^{n} \|f_j^{1/2}(\theta_n,g_n) - f_j^{1/2}(\theta,g) + \tfrac{1}{c}\{f_j^{1/2}(\theta_n^*,g_n^*) - f_j^{1/2}(\theta,g)\}\|_{\mu_j}^2 \to 0 \qquad \text{as } n \to \infty$$

Now in view of Lemma 5 in Chapter 17, section 2 of Le Cam (1986), (3.7) follows immediately from (3.8). ∎

In the proof of Lemma 3.1 we can notice that Assumption S amounts to the richness of the contingent at g in the spirit of Theorem 3.1 in Le Cam (1986). Also note that (3.7) in Oosterhoff and van Zwet (1979) implies (3.7) in the present paper. Now we state our main theorem as follows.

Theorem 3.1. Under Assumption S and conditions (C1), (C2), we have, for any $\epsilon > 0$,

$$P_{\theta,g}^{(n)}\{|\Lambda_n - 2n^{-1/2} \sum_{j=1}^{n} \alpha_j(X_j) f_j^{-1/2}(X_j,\theta,g) + 2\sigma^2| > \epsilon\} \to 0 \quad \text{as } n \to \infty \tag{3.9}$$

Thus, under $P_{\theta,g}^{(n)}$,

$$\Lambda_n \to_d N(-2\sigma^2,\ 4\sigma^2) \quad \text{as } n \to \infty$$

and the sequences $\{\Pi_{j=1}^{n} f_j(x_j,\theta_n,g_n)\}$ and $\{\Pi_{j=1}^{n} f_j(x_j,\theta,g)\}$ are contiguous.

Thus Theorem 3.1 asserts that the condition (C1) and (C2) are sufficient for the LAN of the family $P_{\theta,g}^{(n)}$ at $(\theta,g) \in \Theta \times \mathcal{G}$ under Assumption S. The proof will be given in section 4 and is mostly benefited by Ibragimov and Khas'minskii (1975).

Remark 3.4. When we have a triangular array of independent observations $X_{n1},\ldots,X_{nn},\ldots$, Theorem 3.1 remains valid under the corresponding conditions on this triangular array. We formulate this as follows: Suppose X_{nj} has a density $f_{nj}(\cdot,\theta,g)$ with respect to μ_{nj}. We define all the terms necessary using f_{nj} instead of f_j as we did it in section 2. Then we have the following theorem without any difficulty.

Theorem 3.2. Under Assumption S and the corresponding conditions with f_{nj}, we have, for any $\epsilon > 0$,

$$(3.10) \quad P_{\theta,g}^{(n)}\{|\Lambda_n - 2\, n^{-1/2} \sum_{j=1}^{n} \alpha_{nj}(X_{nj}) f_{nj}^{-1/2}(X_{nj},\theta,g) + 2\,\sigma^2| > \epsilon\} \to 0 \quad \text{as } n \to \infty$$

Thus, under $P_{\theta,g}^{(n)}$,

$$\Lambda_n \to_d N(-2\,\sigma^2,\ 4\,\sigma^2) \quad \text{as} \quad n \to \infty$$

and the sequences $\{\Pi_{j=1}^{n} f_{nj}(x_{nj},\theta_n,g_n)\}$ and $\{\Pi_{j=1}^{n} f_{nj}(x_{nj},\theta,g)\}$ are contiguous.

Remark 3.5. When we have a parametric component $\theta \in \Theta$, where Θ is an open set in $\mathbb{R}^k$, Definition 2.1, Proposition 3.1, Lemma 3.1 and Theorem 3.1 must be altered as follows: with $|\theta_n-\theta| \to 0$, $\|g_n^{1/2}-g^{1/2}\|_{\upsilon} \to 0$ as $n \to \infty$ where $|\cdot|$ denotes the usual Euclidean norm, instead of (2.4) we require that

$$(2.4)' \quad \frac{\sup_{1\le j\le n} \|f_j^{1/2}(\theta,g_n)-f_j^{1/2}(\theta,g) - \{\rho_{j,\theta}(\theta_n-\theta) + B_j(g_n^{1/2}-g^{1/2})\}\|_{\mu_j}}{|\theta_n-\theta| + \|g_n^{1/2}-g^{1/2}\|_{\upsilon}} \to 0 \quad \text{as } n \to \infty$$

for some k-vector $\rho_{j,\theta}$ of function in $L^2(\mu_j)$ and B_j a bounded linear operator as in Definition 2.1. The required change in Proposition 3.1 is that (3.3) must be replaced by

(3.3)' $$\alpha_j = h \cdot \rho_{j,\theta} + B_j \beta$$

where $|n^{1/2}(\theta_n - \theta) - h| \to 0$ and $\|n^{1/2}(g_n^{1/2} - g^{1/2}) - \beta\|_\upsilon \to 0$ as $n \to \infty$. Lemma 3.1 and Theorem 3.1 also must be understood with these changed θ_n, θ and α_j.

Example 3.1. In this example we consider the simplest regression model as follows:

$$Y_j = \theta x_j + \epsilon_j$$

where ϵ_j's are i.i.d. and have an unknown common density g with respect to lebesgue measure υ on R^1 and $\theta \in R^1$ and x_j's are not random. Then Y_j has a density $f_j(\cdot,\theta,g) = g(\cdot - \theta x_j)$. If the Fisher information I_g is finite, $\sup_j |x_j| \leq C$ for some $C > 0$ and $\frac{1}{n} \Sigma_{j=1}^n x_j^2 \to a > 0$, then (C1) and (C2) are satisfied with $\rho_{j,\theta}(\cdot) = -\frac{1}{2} x_j \dot{g} g^{-1/2}(\cdot - \theta x_j)$, $(B_j\beta)(\cdot) = \beta(\cdot - \theta x_j)$ and $\mathfrak{B} = \{\beta \in L^2(\upsilon) : \langle \beta, g^{1/2} \rangle = 0\}$. Hence Assumption S also holds and Λ_n is asymptotically normal with mean $-\frac{1}{2} aI_g$ and variance aI_g.

Example 3.2. In this example we consider Cox's regression model with constant covariates. For the purpose of simplicity, we treat the case without censoring. We observe T_j having hazard function given by

$$\lambda_j(t) = \lambda_o(t) \exp(\theta Z_j) \qquad \theta \in R$$

where $\lambda_o(\cdot) = g \bar{G}^{-1}(\cdot)$, $\bar{G}(\cdot) = 1 - G(\cdot) = \int_\cdot^\infty g \, d\upsilon$, υ is Lebesgue measure on R^+, Z_j's $\in R$ are covariates and constant in constrast to that they were treated as iid random variables with known density in Efron (1977), Tsiatis (1981), Begun et al. (1983), Bickel et al. (1986) and elsewhere. Let

$$\mathcal{G} = \{\text{all densities with respect to Lebesgue measure } \upsilon \text{ on } R^+\}.$$

Then

$$P[T_j \geq t] = \bar{G}^{r_j}(t)$$

where $r_j = \exp[\theta Z_j]$

Hence the density of T_j with respect to υ is

$$f_j(t,\theta,g) = r_j\, g(t)\bar{G}^{r_j-1}(t).$$

Now natural candidates for B_j and $\rho_{j,\theta}$ are as follows:

$$B_j\,\beta = \left\{\beta\, g^{-1/2} + (r_j-1)\,\frac{\int_{\cdot}^{\infty} \beta\, g^{1/2}\, dx}{\bar{G}(\cdot)}\right\} f_j^{1/2} \tag{3.11}$$

$$\rho_{j,\theta} = \frac{1}{2}\, Z_j(1 + \log \bar{G}^{r_j})f_j^{1/2}$$

As Begun et al. (1983) pointed it out for random Z, (3.2) fails with B_j and $\rho_{j,\theta}$ defined by (3.11) if we approach g having bounded support along with g_n's which have support outside that of g. To see this, let

$$g_n^{1/2} = [(1-\frac{1}{\sqrt{n}})g^{1/2} + \frac{1}{\sqrt{n}}\, g''^{1/2}] \,/\, c_n^{1/2} \quad \text{where} \quad c_n = 1 - \frac{2}{\sqrt{n}} + \frac{2}{n}$$

and g", g have supports S and S^c respectively.

Then for certain fixed θ

$$\sum_{j=1}^{n} \|f_j^{1/2}(\theta,g_n) - f_j^{1/2}(\theta,g) - B_j\,\beta\|^2$$

$$\geq \sum_{j=1}^{n} \int_S f_j(t,\theta,g_n)dt \geq 1$$

if $e^{\theta Z_j} \leq 1$ for all j. This happening can be avoided if we restrict ourselves to $\mathscr{B} = \{\beta \in L^2(\upsilon) \mid \langle \beta, g^{1/2} \rangle = 0,\ \text{support}\ (\beta) \subset \text{support}\ (g)\}$ as in Begun et al. (1983). Now it is not so hard to see that (C1) and (C2) hold with $\rho_{j,\theta}$ and B_j given by (3.11) if $\sup_j |Z_j| \leq C$ and $\frac{1}{n}\,\Sigma_{j=n}^{n}\, Z_j^2\, e^{\theta Z_j} \to a$ for some C and a > 0.

4. Proof of Theorem 3.1

By expanding Λ_n by Taylor's formula for $\max_{1\leq j\leq n} |r_{nj}| < \epsilon$ with $r_{nj} \equiv r_j(\theta,\theta_n;g,g_n)$, we obtain

$$\Lambda_n = \Sigma_{j=1}^{n}\, r_{nj} - \frac{1}{4}\,\Sigma_{j=1}^{n}\, r_{nj}^2 + \frac{1}{8}\,\Sigma_{j=1}^{n}\, w_{jn}|r_{nj}|^3$$

where $|w_{jn}| \leq 1$. By the fact that $\langle \alpha_j, f_j^{1/2} \rangle = 0$ for any $j \geq 1$, $2\, n^{-1/2} \Sigma_{j=1}^{n} \alpha_j(x_j) f_j^{-1/2}(x_j)$ has the asymptotic normal distribution with parameter $(0, 4\sigma^2)$ under (C2). From this fact it suffices to show the equalities

$$(4.1) \quad \lim_{n\to\infty} P_{\theta,g}^{(n)} \left\{ \max_{1\leq j\leq n} |r_{nj}| > \epsilon \right\} = 0$$

$$(4.2) \quad \lim_{n\to\infty} P_{\theta,g}^{(n)} \left\{ \left| \sum_{j=1}^{n} r_{nj}^2 - 4\sigma^2 \right| > \epsilon \right\} = 0$$

$$(4.3) \quad \lim_{n\to\infty} P_{\theta,g}^{(n)} \left\{ \left| \sum_{j=1}^{n} r_{nj} - 2\, n^{-1/2} \sum_{j=1}^{n} \alpha_j f_j^{-1/2} + \sigma^2 \right| > \epsilon \right\} = 0$$

$$(4.4) \quad \lim_{n\to\infty} P_{\theta,g}^{(n)} \left\{ \left| \sum_{j=1}^{n} |r_{nj}|^3 > \epsilon \right. \right\} = 0$$

Note that (4.4) is a direct consequence of (4.1) and (4.2) so that we need to prove only (4.1), (4.2) and (4.3).

Proof of (4.1). The following inequalities are obvious.

$$P_{\theta,g}^{(n)} \left\{ \max_{1\leq j\leq n} |r_{nj}| > \epsilon \right\}$$

$$\leq \sum_{j=1}^{n} P_{\theta,g}^{(n)} \{ |r_{nj}| > \epsilon \}$$

$$\leq \sum_{j=1}^{n} P_{\theta,g}^{(n)} \left\{ \left| \sum_{j=1}^{n} r_{nj} - 2\, n^{-1/2} \alpha_j f_j^{-1/2} \right| > \frac{\epsilon}{2} \right\} + \sum_{j=1}^{n} P_{\theta,g}^{(n)} \left\{ |\alpha_j f_j^{-1/2}| > \frac{n^{1/2}\epsilon}{4} \right\} .$$

By Chebyshev's inequality, we see that the first sum on the right tends to zero by Proposition 3.1 and the second one does so by (C2).

Proof of (4.2). Using the useful inequality

$$(4.5) \qquad |ab| \leq \alpha\, a^2/2 + b^2/2\alpha \qquad \alpha > 0,$$

we obtain

$$P^{(n)}_{\theta,g}\left\{\left|\sum_{j=1}^{n}\left(r^2_{nj} - \frac{4}{n}\alpha^2_j f^{-1}_j\right)\right| > \epsilon\right\}$$

$$\leq \frac{1}{\epsilon}\sum_{j=1}^{n} E^{(n)}_{\theta,g}\left|r^2_{nj} - \frac{4}{n}\alpha^2_j f^{-1}_j\right|$$

$$\leq \frac{\alpha}{2\epsilon}\sum_{j=1}^{n} E^{(n)}_{\theta,g}\left|r_{nj} - 2n^{-1/2}\alpha_j f_j^{-1/2}\right|^2$$

$$+ \frac{1}{\alpha\epsilon}\left[\frac{4}{n}\sum_{j=1}^{n}\|\alpha_j\|^2_{\mu_j} + \sum_{j=1}^{n}\|r_{nj}\|^2_{\rho_{j,\theta,g}}\right]$$

for any $\alpha > 0$. By Proposition 3.1 and the condition that $n^{-1}\sum_{j=1}^{n}\|\alpha_j\|^2_{\mu_j} \to \sigma^2$, the right hand side of the last inequality can be made arbitrarily small if we let $n \to \infty$ first and then let $\alpha \to \infty$. The relative stability of $\frac{1}{n}\sum_{j=1}^{n}\alpha^2_j f_j^{-1}$, which is guaranteed by (C2) (cf. Gnedenko and Kolmogorov (1968), Corollary 2 on p. 141), ensures

$$\frac{1}{n}\sum_{j=1}^{n}\alpha^2_j f^{-1}_j \to \sigma^2 \quad \text{in} \quad P^{(n)}_{\theta,g} \text{ - probability}$$

since by Lemma 3.1 we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\int_{\{f_j=0\}}\alpha^2_j\, d\mu_j = 0. \tag{4.6}$$

<u>Proof of (4.3)</u>. From Proposition 3.1

$$\sum_{j=1}^{n}\|r_{nj}\|^2_{P_{j,\theta,g}} = \frac{4}{n}\sum_{j=1}^{n}\|\alpha_j f_j^{-1/2}\|^2_{P_{j,\theta,g}} + 2\sum_{j=1}^{n}\langle r_{nj} - 2n^{-1/2}\alpha_j f_j^{-1/2},\ 2n^{-1/2}\alpha_j f_j^{-1/2}\rangle_{P_{j,\theta,g}} + O(1)$$

where $\langle\cdot\rangle_{P_{j,\theta,g}}$ is the usual inner product in $L^2(P_{j,\theta,g})$.

Using (4.5) again we can see

$$\left|\sum_{j=1}^{n}\langle r_{nj} - 2n^{-1/2}\alpha_j f_j^{-1/2},\ 2n^{-1/2}\alpha_j f_j^{-1/2}\rangle_{P_{j,\theta,g}}\right| \tag{4.7}$$

$$\le \frac{\alpha}{2} \sum_{j=1}^{n} \left\| r_{nj} - 2\, n^{½}\, \alpha_j\, f_j^{½} \right\|^2_{P_{j,\theta,g}} + \frac{1}{2\alpha} \frac{4}{n} \sum_{j=1}^{n} \|\alpha_j\|^2_{\mu_j}$$

Letting $n \to \infty$ and then $\alpha \to 0$ ensures that both of two terms on the right hand side of (4.7) tend to zero, which entails

$$\lim_{n\to\infty} \sum_{j=1}^{n} \|r_{nj}\|^2_{P_{j,\theta,g}} = 4\,\sigma^2 \tag{4.8}$$

because of (4.6). Now Lemma 3.1 together with (4.8) implies that

$$\lim_{n\to\infty} E^{n}_{\theta,g} \left(\sum_{j=1}^{n} r_{nj} \right) = -\sigma^2 \tag{4.9}$$

after going over to expectations in the identity

$$\sum_{j=1}^{n} r^2_{nj} = 4 \sum_{j=1}^{n} \left[\frac{f_j(x_j,\theta_n,g_n)}{f_j(x_j,\theta,g)} - 1 \right] - 4 \sum_{j=1}^{n} r_{nj}.$$

Hence, for sufficiently large n, we get

$$P^{(n)}_{\theta,g}\left\{ \left| \sum_{j=1}^{n} r_{nj} - 2\, n^{-½} \sum_{j=1}^{n} \alpha_j\, f_j^{-½} + \sigma^2 \right| > \epsilon \right\}$$

$$\le P^{(n)}_{\theta,g}\left\{ \left| \sum_{j=1}^{n} (r_{nj} - E^{(n)}_{\theta,g}\, r_{nj}) - 2\, n^{-½} \sum_{j=1}^{n} \alpha_j\, f_j^{-½} \right| > \frac{\epsilon}{2} \right\}$$

$$\le \frac{4}{\epsilon^2} E^{(n)}_{\theta,g} \left[\sum_{j=1}^{n} (r_{nj} - E^{(n)}_{\theta,g}\, r_{nj}) - 2\, n^{-½} \sum_{j=1}^{n} \alpha_j\, f_j^{-½} \right]^2$$

$$\le \frac{4}{\epsilon^2} \sum_{j=1}^{n} \left\| r_{nj} - 2\, n^{-½}\, \alpha_j\, f_j^{-½} \right\|^2_{P_{j,\theta,g}} .$$

Therefore (4.3) follows from Proposition 3.1. ∎

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.

Bickel, P.J., Klaassen, C.J., Ritov, Y. and Wellner, J.A. (1986). Efficient and adaptive inference in semiparametric models. Preprint.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. Journal of the American Statist. Assoc., 72, 557-565.

Gnedenko, B.V. and Kolmogorov, A.N. (1968). Limit Distributions for Sums of Independent Random Variables. Addison-Wesley, Reading, Mass.

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.

Ibragimov, I.A. and Khas'minskii, R.Z. (1975). Local asymptotic normality for non-identically distributed observations. Theory of Prob. Applications, 20, 246-260.

Kushnir, A.F. (1968). Asymptotically optimal tests for a regression problem of testing hypotheses. Theory of Prob. Applications, 13, 647-666.

Le Cam, L. (1960). Locally asymptotically normal families of distributions. Univ. of California Publications in Statistics. Vol. 3, No. 2, 37-98.

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. Hajek. In Contribution to Statistics; Jaroslav Hajek Memorial Volume, Ed. J. Jureckova, 119-135. Reidel, Dordrecht.

Le Cam, L. (1986). Asymptotic Methods in Statistical Decision Theory. Springer, New York.

Oosterhoff, J. and van Zwet, W.R. (1979). A note on contiguity and Hellinger distance. In Contribution to Statistics; Jaroslav Hájek Memorial Volume, Ed. J. Jureckova, 157-166. Reidel, Dordrecht.

Pfanzagl, J. (1982). Contribution to A General Asymptotic Statistical Theory. Lecture Notes in Statistics, Springer, New York.

Phillipou, A.N. and Roussas, G.G. (1973). Asymptotic distribution of the likelihood function in the independent not identically distributed case. Ann. Statist., 1, 454-471.

Tsiatis, A.A. (1981). A large sample study of Cox's regression model. Ann. Statist., 9, 93-108.