KYBERNETIKA — VOLUME *30* (1994), NUMBER 3, PAGES 359–367

ON ADAPTIVE ESTIMATION IN NONLINEAR REGRESSION

SILVELYN ZWANZIG

To study adaptive estimators for the regression parameter we embed the usual nonlinear regression model in a semiparametric one. The parameter of interest is the finite dimensional regression parameter and the unknown density of the error distribution is the infinite dimensional nuisance parameter.

In this paper the LAN property for the semiparametric nonlinear regression model is shown. Necessary conditions for the existence of an adaptive estimator are derived and a minimax theorem is given.

The interpretation of the necessary conditions is the following: In the nonlinear model we need a symmetric error density. In the linear model adaption is also possible with asymmetric error density, if we have an asymptotic symmetric design.

1. THE SEMIPARAMETRIC MODEL

We regard the nonlinear regression model:

$$y_i = g(x_i, \vartheta) + \varepsilon_i, \quad i = 1, \ldots, n \tag{1}$$

The regression function $g(\cdot,\cdot) : X \times \Theta \to \Re$ is known. The random errors $\varepsilon_i,\ i = 1, \ldots, n$ are i.i.d. with $\mathsf{E}\varepsilon_1 = 0$, $\mathsf{Var}\varepsilon_1 = \sigma^2$ and density $p(\cdot)$ of ε_1 with respect to a sigma-finite measure μ. Let $\mathrm{L}^2(\mu)$ denote the usual L^2-space of square integrable functions and let $\|.\|_\mu$ denote the usual norm in $\mathrm{L}^2(\mu)$. Thus $\left\|\sqrt{p}\right\|_\mu = 1$. The unknown "parameter" θ of the semiparametric model has a parametric component ϑ and a nonparametric component p:

$$\theta = (\vartheta, p) \tag{2}$$

$$\begin{aligned} \vartheta \in \Theta \subseteq \Re^q &\qquad \text{parameter of interest} \\ p \in \Gamma \subseteq L^2(\mu) &\qquad \text{nuisance parameter} \end{aligned}$$

The design points $\mathrm{x}_i \in X \subseteq \Re^m$, $i = 1, \ldots, n$, are fixed and known, such that $Y = (y_1, \ldots, y_n)$ is a sample of random, independent, not identically distributed values with probability P_θ^n.

The main problem is to estimate the unknown regression parameter. If the density p is known, we can use the maximum likelihood estimator, which is asymptotically efficient. The general aim of adaptive estimating is to find an estimator, which has the same asymptotical efficiency and does not depend on p. For that reason we regard the density p as nuisance parameter.

2. THE LAN PROPERTY

In the usual nonlinear regression model the LAN property was proven under some regularity conditions on the likelihood function and the regression function (compare [2], Chapter 1). We will need such conditions and propose the following notations. Let Γ_{LAN} be the set of all densities, which fulfill (L1), (L2), (L3).

(L1) The likelihood function $l(\cdot) = \ln p(\cdot)$ is twice continuously differentiable. $l^{(k)}(\cdot)$ denotes the kth derivative.

(L2)

$$\mathsf{E}_p l^{(1)}(\varepsilon_1) = 0 \quad \mathsf{E}_p \left(\frac{p^{(2)}(\varepsilon_1)}{p(\varepsilon_1)} \right) < \infty,$$
$$0 < s^2 = -\mathsf{E}_p l^{(2)}(\varepsilon_1) = \mathsf{Var}_p l^{(1)}(\varepsilon_1) < \infty.$$

(L3) There exist a function R, with $\mathsf{E}_p R(\varepsilon_1) < \infty$, and a constant δ_0 such that for all $\delta < \delta_0$ and for all d with $d = d(\varepsilon_1)$ and $|d| < \delta$

$$\left| p^{(2)}\left(\varepsilon_1 + d\right) - p^{(2)}\left(\varepsilon_1\right) \right| < \delta\, p\left(\varepsilon_1\right) R\left(\varepsilon_1\right).$$

Let G the set of all regression functions, which fulfill (G1), (G2), (G3), (G4).

(G1) The regression function $g(x_i, \cdot) : \Theta \to \Re$ is twice differentiable uniformly respect to i and the derivatives are equicontinuous. $g^{(1)}(x_i, \cdot)$ denotes the q dimensional vector of first partial derivatives. $g^{(2)}(x_i, \cdot)$ denotes the $q \times q$ dimensional matrix of second partial derivatives.

(G2) For all ϑ

$$\frac{1}{\sqrt{n}} \max_{i=1,\ldots,n} \left\| g^{(1)}(x_i, \vartheta) \right\| = o\,(1).$$

(G3) It exists a constant c independent of i such that for all ϑ

$$\left\| g^{(2)}(x_i, \vartheta) \right\| \le c.$$

(G4) It exists a positive definite matrix $I(\vartheta)$ such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g^{(1)}(x_i, \vartheta)\, g^{(1)}(x_i, \vartheta)^{\mathrm{T}} = I\left(\vartheta\right)$$

Let us generalize the concept of local alternatives $\theta_n = (\vartheta_n, p_n)$ of the "true parameter" θ on the semiparametric case. We define:

$$\vartheta_n = \vartheta + \frac{1}{\sqrt{n}} h, \quad h \in \Re^q, \tag{3}$$

$$p_n = p + \frac{1}{\sqrt{n}} \beta, \quad \beta \in B_0 \subseteq B \subseteq L^2(\mu). \tag{4}$$

The set B in (4) respects the definition (2.2) of all possible deviations of $\sqrt{p}$ in [1]. Also contaminated models

$$p_n = \left(1 - \frac{1}{\sqrt{n}} \varepsilon\right) p + \frac{1}{\sqrt{n}} \varepsilon\, p_1, \quad p_1 \in \Gamma_{\text{LAN}}, \quad \varepsilon \in (0,1)$$

can be given in the form of (4), if we set

$$B_0 = \{\beta : \beta = \varepsilon (p_1 - p),\, p_1 \in \Gamma_{\text{LAN}},\, \varepsilon \in (0,1)\}.$$

We choose the set of deviations B in such a way, that also p_n is in Γ_{LAN} for sufficiently large n. More exactly B denotes the set of all β, which fulfill (B1), (B2), (B3).

(B1)

$$\int \beta \, \mathrm{d}\mu = 0.$$

(B2)

$$\mathsf{E}_p \left(\frac{\beta(\varepsilon_1)}{p(\varepsilon_1)}\right)^2 < \infty.$$

(B3) β is continuously differentiable and there exist a function R with $\mathsf{E}_p R(\varepsilon_1) < \infty$ and a constant δ_0 such that for all $\delta < \delta_0$ and for all d with $d = d(\varepsilon_1)$ and $|d| < \delta$

$$\left|\beta^{(1)}(\varepsilon_1 + d) - \beta^{(1)}(\varepsilon_1)\right| < \delta\, p(\varepsilon_1)\, R(\varepsilon_1), \quad \mathsf{E}_p \left(\frac{\beta^{(1)}(\varepsilon_1)}{p(\varepsilon_1)}\right)^2 < \infty.$$

Now we can formulate the theorem on the asymptotic behaviour of the likelihood quotient. Introduce the log likelihood ratio

$$L_n = \ln \prod_{i=1}^n \frac{p_n(y_i - g(x_i, \vartheta_n))}{p(y_i - g(x_i, \vartheta))}, \quad \text{if} \quad \prod_{i=1}^n p(y_i - g(x_i, \vartheta)) > 0$$

and

$$L_n = 0, \quad \text{if} \quad \prod_{i=1}^n p(y_i - g(x_i, \vartheta)) = \prod_{i=1}^n p_n(y_i - g(x_i, \vartheta_n)) = 0$$

and

$$L_n = \infty, \quad \text{if} \quad \prod_{i=1}^n p(y_i - g(x_i, \vartheta)) = 0 \quad \text{and} \quad \prod_{i=1}^n p_n(y_i - g(x_i, \vartheta_n)) > 0.$$

Theorem 1. (Local asymptotic normality) For $g \in G$ and $p \in \Gamma_{\mathrm{LAN}}$ and $\beta \in B$ and $h \in \Re^q$ it holds

$$L_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Delta_i - \frac{1}{2n} \sum_{i=1}^{n} (\Delta_i)^2 + o_\theta(1)$$

with

$$\Delta_i = \frac{\beta(\varepsilon_i)}{p(\varepsilon_i)} - \frac{p^{(1)}(\varepsilon_i)}{p(\varepsilon_i)} g^{(1)}(x_i, \vartheta)^{\mathrm{T}} h. \tag{5}$$

$o_\theta(1)$ converges with probability P_θ^n to zero.

The Δ_i are independent random variables with expected value zero and the average of their variances is bounded. Hence we have:

Corollary. For $g \in G$ and $p \in \Gamma_{\mathrm{LAN}}$ and $\beta \in B$ and $h \in \Re^q$, such that $\sigma^2(\beta, h) > 0$ with

$$\sigma^2(\beta, h) = \left\| \frac{p^{(1)}}{\sqrt{p}} m^{\mathrm{T}} h - \frac{\beta}{\sqrt{p}} \right\|_\mu^2 + s^2 h^{\mathrm{T}} \left(I(\vartheta) - m m^{\mathrm{T}} \right)$$

and

$$m = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g^{(1)}(x_i, \vartheta), \quad s^2 = \left\| \frac{p^{(1)}}{\sqrt{p}} \right\|_\mu^2$$

it holds

$$L_n \to N\left(-\frac{1}{2}\sigma^2(\beta, h), \sigma^2(\beta, h) \right).$$

and the probability measure $P_{\theta_n}^n$ is contiguous to P_θ^n.

Remark. From the contiguity it follows, that the consistency of some estimator under p implies the consistency under the contaminated model $p_n = (1 - \varepsilon_n) p + \varepsilon_n p_1$ with $\varepsilon_n = \frac{1}{\sqrt{n}} \varepsilon$ for any $p_1 \in \Gamma_{\mathrm{LAN}}$.

Proof of Theorem 1. We give an outline. Especially we are interested in the influence of the deviation β from the density p. Because of (4) we have P_θ^n a.s., that

$$L_n = \sum_{i=1}^{n} \ln \left(\frac{p(y_i - g(x_i, \vartheta_n))}{p(\varepsilon_i)} + \frac{1}{\sqrt{n}} \frac{\beta(y_i - g(x_i, \vartheta_n))}{p(\varepsilon_i)} \right).$$

Under (G1) for all i Taylor expansions of first and second order hold, such that

$$\Delta g_i = g(x_i, \vartheta_n) - g(x_i, \vartheta) = \frac{1}{\sqrt{n}} g^{(1)} \left(x_i, \overline{\vartheta}^{(1)} \right)^{\mathrm{T}} h,$$

$$\Delta g_i = \frac{1}{\sqrt{n}} g^{(1)}(x_i, \vartheta)^{\mathrm{T}} h + \frac{1}{2n} h^{\mathrm{T}} g^{(2)} \left(x_i, \overline{\vartheta}^{(2)} \right)^{\mathrm{T}} h,$$

$$\text{with} \quad \overline{\vartheta}^{(k)} \quad \text{such that} \quad \left\| \overline{\vartheta}^{(k)} - \vartheta \right\| \le \frac{1}{\sqrt{n}} \|h\| \quad \text{for } k = 1, 2.$$

We have also expansions for the density and the deviation:

$$\begin{aligned}
&p\left(\varepsilon_i-\Delta g_i\right)=p\left(\varepsilon_i\right)-p^{(1)}\left(\varepsilon_i\right)\Delta g_i+\frac{1}{2}p^{(2)}\left(\varepsilon_i+\delta_i^{(2)}\Delta g_i\right)\Delta g_i^2,\\
&\beta\left(\varepsilon_i-\Delta g_i\right)=\beta\left(\varepsilon_i\right)-\beta^{(1)}\left(\varepsilon_i+\delta_i^{(1)}\Delta g_i\right)\Delta g_i,\\
&\text{with}\quad \delta_i^{(k)}=\delta^{(k)}\left(\varepsilon_i\right)\quad\text{such that}\quad \left|\delta_i^{(k)}\right|<1\quad\text{for } k=1,2.
\end{aligned}$$

Therefore

$$L_n=\sum_{i=1}^{n}\ln\left(1+\frac{1}{\sqrt{n}}\Delta_i+\frac{1}{n}R_i^{(1)}\right), \tag{6}$$

where Δ_i is defined in (5) and

$$\begin{aligned}
2R_i^{(1)}=&\frac{p^{(2)}\left(\varepsilon_i+d_i^{(2)}\right)}{p(\varepsilon_i)}\left(g^{(1)}\left(x_i,\overline{\vartheta}^{(1)}\right)^{\mathrm{T}}h\right)^2-\frac{p^{(1)}(\varepsilon_i)}{p(\varepsilon_i)}h^{\mathrm{T}}g^{(2)}\left(x_i,\overline{\vartheta}^{(2)}\right)^{\mathrm{T}}h\\
&-\frac{\beta^{(1)}\left(\varepsilon_i+d_i^{(1)}\right)}{p(\varepsilon_i)}g^{(1)}\left(x_i,\overline{\vartheta}^{(1)}\right)^{\mathrm{T}}h
\end{aligned}$$

$$\text{with}\quad \left\|d_i^{(k)}\right\|\le\frac{1}{\sqrt{n}}\max_{i=1,..,n}\left\|g^{(1)}\left(x_i,\overline{\vartheta}^{(1)}\right)\right\|<o(1)\quad\text{for } k=1,2. \tag{7}$$

(7) is a consequence of (G2).

It holds

$$\frac{1}{\sqrt{n}}\Delta_i+\frac{1}{n}R_i^{(1)}=o_\theta\left(1\right),$$

because under (G2) and (B2) uniformly in i

$$\frac{1}{n}\mathsf{E}\left|\Delta_i\right|^2\le\frac{1}{n}s^2\left\|g^{(1)}\left(x_i,\vartheta\right)\right\|^2\|h\|^2+\frac{1}{n}\mathsf{E}_p\left(\frac{\beta\left(\varepsilon_1\right)}{p\left(\varepsilon_1\right)}\right)^2\longrightarrow 0 \tag{8}$$

and because of (G3) and (B3) uniformly in i

$$\frac{1}{n}\mathsf{E}\left|R_i^{(1)}\right|\to 0.$$

Now we apply the expansion $\ln(1+x)=x-\frac{1}{2}x^2+\Delta x^3$ with $\Delta<1$ in (6):

$$L_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Delta_i-\frac{1}{2n}\sum_{i=1}^{n}\Delta_i^2+R_n^{(2)},$$

where

$$R_n^{(2)}=\frac{1}{n}\sum_{i=1}^{n}R_i^{(1)}-\frac{1}{\sqrt{n}n}\sum_{i=1}^{n}R_i^{(1)}\Delta_i-\frac{1}{2n^2}\sum_{i=1}^{n}\left(R_i^{(1)}\right)^2+\Delta\sum_{i=1}^{n}\left(\frac{1}{n}R_i^{(1)}+\frac{1}{\sqrt{n}}\Delta_i\right)^3.$$

It remains to show, that $R_n^{(2)}$ converges in probability to zero. The first term of the rest we split again:

$$\frac{1}{n}\sum_{i=1}^{n}R_i^{(1)}=\frac{1}{n}\sum_{i=1}^{n}R_i^{(3)}+\frac{1}{n}\sum_{i=1}^{n}R_i^{(4)} \tag{9}$$

with

$$2R_i^{(3)} = \frac{p^{(2)}(\varepsilon_i)}{p(\varepsilon_i)} g^{(1)}(x_i,\vartheta)^{\mathrm{T}} h - \frac{p^{(1)}(\varepsilon_i)}{p(\varepsilon_i)} h^{\mathrm{T}} g^{(2)}(x_i,\vartheta) h - \frac{\beta^{(1)}(\varepsilon_i)}{p(\varepsilon_i)} g^{(1)}(x_i,\vartheta)^{\mathrm{T}} h. \tag{10}$$

The regularity conditions on p and β allow a change of integration and differentiation, such that from (B1) and the normalizing property of densities it follows, that the expected value of $R_i^{(3)}$ is zero. Further we get from (G3), (G4) and (L2), (B2), (B3), that the average of their variances is bounded. Then the law of large numbers implies

$$\frac{1}{n}\sum_{i=1}^{n} R_i^{(3)} = o_\theta(1). \tag{11}$$

Using the continuity arguments of (G1), (L3) and (B3) we can also show

$$\frac{1}{n}\sum_{i=1}^{n} R_i^{(4)} = o_\theta(1). \tag{12}$$

Let us discuss it for one of the terms more detailed:

$$F_n = \frac{1}{n}\sum_{i=1}^{n} \frac{\beta^{(1)}\left(\varepsilon_i + d_i^{(1)}\right) - \beta^{(1)}(\varepsilon_i)}{p(\varepsilon_i)} g^{(1)}(x_i,\vartheta)^{\mathrm{T}} h.$$

Using (7), (B3), (G2) and (G4) give us the following inequalities

$$\mathsf{E}\,|F_n| \le \max_{i=1,\ldots,n} \mathsf{E}\left|\frac{\beta^{(1)}\left(\varepsilon_i + d_i^{(1)}\right) - \beta^{(1)}(\varepsilon_i)}{p(\varepsilon_i)}\right| \frac{1}{n}\sum_{i=1}^{n}\left|g^{(1)}(x_i,\vartheta)^{\mathrm{T}} h\right|$$

.

$$\le o(1)\ \mathsf{E}_p R(\varepsilon_i)\sqrt{h^{\mathrm{T}} I(\vartheta) h} + o(1)$$

The other terms of $R_n^{(2)}$ we can estimate by arguments used above. For instance it follows from (8) and (9) with (11) and (12)

$$\begin{aligned}
& \mathsf{P}\left(\frac{1}{\sqrt{n}n}\sum_{i=1}^{n} R_i^{(1)}\Delta_i > \varepsilon\right) \\
\le\ & \mathsf{P}\left(\frac{1}{\sqrt{n}n}\sum_{i=1}^{n} R_i^{(1)}\Delta_i > \varepsilon \frown \frac{1}{\sqrt{n}}\max_{i=1,\ldots,n}\Delta_i \le \sqrt{\varepsilon}\right) \\
& + \mathsf{P}\left(\frac{1}{\sqrt{n}n}\sum_{i=1}^{n} R_i^{(1)}\Delta_i > \varepsilon \frown \frac{1}{\sqrt{n}}\max_{i=1,\ldots,n}\Delta_i > \sqrt{\varepsilon}\right) \\
\le\ & \mathsf{P}\left(\frac{1}{n}\sum_{i=1}^{n} R_i^{(1)} > \sqrt{\varepsilon}\right) + \mathsf{P}\left(\frac{1}{\sqrt{n}}\max_{i=1,\ldots,n}\Delta_i > \sqrt{\varepsilon}\right) \le o(1).
\end{aligned}$$

□

Now we can search the "least favourable" direction β^* to approach p for the problem of estimating ϑ.

$$\beta^* = \arg\min_{\beta\in B_0} \sigma^2(\beta, h) = \arg\min_{\beta\in B_0} \left\| \frac{p^{(1)}}{\sqrt{p}} m^{\mathrm{T}} h - \frac{\beta}{\sqrt{p}} \right\|_\mu^2, \tag{13}$$

$$\frac{\beta^*}{\sqrt{p}} = P\left(\frac{p^{(1)}}{\sqrt{p}}\right) m^{\mathrm{T}} h; \tag{14}$$

where $P : L^2(\mu) \to B_0$ denotes the projection. If $B_0 = B$, then $\beta^* = p^{(1)} m^{\mathrm{T}} h$. We obtain

$$\sigma^2(\beta^*, h) = s^2 h^{\mathrm{T}} I^*(\vartheta)\, h$$

with

$$I^*(\vartheta) = I(\vartheta) - mm^{\mathrm{T}} \left(\frac{\left\| P\left(\frac{p^{(1)}}{\sqrt{p}}\right)\right\|_\mu^2}{\left\|\frac{p^{(1)}}{\sqrt{p}}\right\|_\mu^2} \right). \tag{15}$$

If $B_0 = B$, then $I^*(\vartheta) = I(\vartheta) - mm^{\mathrm{T}}$. Note $I^*(\vartheta) \le I(\vartheta)$ and $I^*(\vartheta) = I(\vartheta)$ if $P\left(\frac{p^{(1)}}{\sqrt{p}}\right) = 0$.

3. NECESSARY CONDITIONS FOR ADAPTATION

First let us recall the definition for an adaptive estimator (compare [1]).

Definition. A sequence of estimators $\vartheta_n(Y)$ is said to be Γ-adaptive if under θ_n from (3) and (4) for all θ_n and all θ

$$L\left(\sqrt{n}\left(\vartheta_n(Y) - \vartheta_n\right)\right) \longrightarrow N_q\left(0, s^{-2} I(\vartheta)^{-1}\right)$$

Remember, $s^{-2} I(\vartheta)^{-1}$ is the covariance matrix of the asymptotic distribution of the maximum likelihood estimator (compare for instance [2], Chapter 1).

Further we say, that an estimator $\vartheta_n(Y)$ is regular at θ, if for every sequence θ_n the distribution of $\sqrt{n}\left(\vartheta_n(Y) - \vartheta_n\right)$ converges under θ_n to a law $L(\theta)$, which depends on θ but not on h and β. From the LAN property we get the presentation theorem for regular estimators.

Theorem 2. Let $p \in \Gamma_{\mathrm{LAN}}$, $\beta \in B_0$, $g \in G$. If $\vartheta_n(Y)$ is a regular estimator with limit distribution L, then it exists a distribution $L_1(\theta)$ such that

$$L(\theta) = N_q\left(0, s^{-2} I^*(\vartheta)^{-1}\right) * L_1(\theta)$$

The proof is omitted. It is analogously to that of Theorem 3.1. of [1].

If we want to find an adaptive estimator, then it is only possible if $I(\vartheta) = I^*(\vartheta)$. From this equality we derive the necessary conditions for the existence of adaptive estimators.

Theorem 3. Let $p \in \Gamma_{\text{LAN}}$, $\beta \in B_0$, $g \in G$. A necessary condition for the existence of an adaptive estimator in the nonlinear regression model (1) is (N1) or (N2):

(N1)

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g^{(1)}(x_i, \vartheta) = 0 \quad \text{for all } \vartheta.$$

(N2)

$$\int \frac{p^{(1)}\beta}{p} \, \mathrm{d}\mu = 0 \quad \text{for all } \ \beta \in B_0.$$

Remark. The condition (N1) is fulfilled in the linear regression model with an asymptotic symmetric design:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = 0$$

In the location model $\mathrm{g}(x_i, \vartheta) = \vartheta$ (N1) is not fulfilled. For $B = B_0$ adaption is not possible. We have $I(\vartheta) = 1$ and $I^*(\vartheta) = 0$. It is necessary to choose :

$$B_0 = \left\{ \beta : \int \frac{p^{(1)}\beta}{p} \, \mathrm{d}\mu = 0, \ \beta \in B \right\}.$$

The condition (N2) is fulfilled if p is a symmetric density and the local alternatives p_n are also only symmetric.

Proof of Theorem 3. If an adaptive estimator exists, then by Theorem 2 it must hold $I(\vartheta) = I^*(\vartheta)$. The equality holds only for $\beta^* = 0$. Remember β^* is the solution of the minimization problem (13). This implies

$$\left(\frac{p^{(1)}}{p} m^{\mathrm{T}} h - \frac{\beta^*}{\sqrt{p}} \right) \perp \frac{\beta}{\sqrt{p}} \quad \text{for all } \ \beta \in B_0.$$

If $\beta^* = 0$, then

$$\frac{p^{(1)}}{p} m^{\mathrm{T}} h \perp \frac{\beta}{\sqrt{p}} \quad \text{for all } \ \beta \in B_0.$$

Hence for all h

$$m^{\mathrm{T}} h \int \frac{p^{(1)}\beta}{p} \, \mathrm{d}\mu = 0.$$

From this it follows (N1) or (N2). □

4. THE MINIMAX BOUND

A further consequence of the LAN property is an asymptotic minimax theorem of Hájek type. It may be useful also in cases, where adaption is not possible.

Denote W the class of functions $w : \Re^q \to \Re_+$ such that $w(0) = 0, \ w(x) = w(-x)$, all sets $\{x : w(x) < c\}$ are convex and $\ln\, w(x) \le o(\|x\|^2)$.

Theorem 4. Suppose $w \in W$, $p \in \Gamma_{\text{LAN}}$, $g \in G$ and $I^*(\vartheta) \succ 0$ and

$$U_n(c) = \left\{ \theta_n = (\vartheta_n, p_n) : \|\vartheta_n - \vartheta\|^2 < c \, , \; \|p_n - p\|_\mu^2 < c \right\}.$$

Then it holds for all $c > 0$ and all estimators $\vartheta_n(Y)$

$$\lim_{n\to\infty} \inf \sup_{\theta_n \in U(c)} \mathsf{E}_{\theta_n} w\left(\sqrt{n}\left(\vartheta_n(Y) - \vartheta_n\right)\right) \geq \mathsf{E} w(z^*)$$

where z^* is $N_q\left(0, \left(s^2 I^*(\vartheta)\right)^{-1}\right)$ distributed.

P r o o f of Theorem 4. Choose the worst direction of deviation β^* in (13) of p and fix the sequence

$$\theta_n^* = (\vartheta_n, p_n^*), \qquad \text{with } \; p_n^* = p + \frac{1}{\sqrt{n}}\beta^*,$$

then

$$\sup_{\theta_n \in U(c)} \mathsf{E}_{\theta_n} w\left(\sqrt{n}\left(\vartheta_n(Y) - \vartheta_n\right)\right) \geq \sup_{\|\vartheta_n - \vartheta\|^2 < c} \mathsf{E}_{\theta_n^*} w\left(\sqrt{n}\left(\vartheta_n(Y) - \vartheta_n\right)\right)$$

The likelihood quotient

$$L_n^* = \prod_{i=1}^{n} \frac{p_n^*\left(y_i - g(x_i, \vartheta_n)\right)}{p_n\left(y_i - g(x_i, \vartheta)\right)}$$

is LAN at θ with $s^2 I^*(\vartheta)$. It is not possible to apply the Hájek Theorem (Theorem 12.1) in the book of Ibragimov, Hasminskii [3] directly, because L_n^* has a different structure as the likelihood quotient in their Definition 2.1. If we check the proof of Theorem 12.1 step by step, we see that only the LAN property is used. We omit this checking and apply this theorem on the last inequality and obtain the result. □

(Received March 3, 1994.)

REFERENCES

[1] J. M. Begun, W. J. Hall, W. Huang and J. A. Wellner: Information and asymptotic efficiency in parametric – nonparametric models. Ann. Statist. *11* (1983), 432–452.

[2] K. M. S. Humak: Statistische Methoden der Modellbildung II. Akademie Verlag, Berlin 1983.

[3] I. A. Ibragimov and R. Z. Hasminskii: Statistical Estimation: Asymptotic Theory. Springer–Verlag, New York 1981.

Dr. Silvelyn Zwanzig, University of Hamburg, Institute for Mathematical Stochastic, Bundesstraße 55, D–20 146 Hamburg. Germany.