This is to certify that the

dissertation entitled

On adaptive estimation

presented by

Mr. Anton Schick

has been accepted towards fulfillment
of the requirements for

PhD degree in Statistics

Vaclav Fabian
Major professor

Date July 20, 1983

ON ADAPTIVE ESTIMATION

By

Anton Schick

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1983

ABSTRACT

ON ADAPTIVE ESTIMATION

By

Anton Schick

A general method for the construction of locally asymptotically minimax (LAM) - adaptive estimates is given under conditions weaker than those in Bickel (1982). In particular, we show that Bickel's condition S* is not necessary for LAM-adaptive estimation and replace it by a weaker condition. This new condition is found to be necessary and sufficient for a class of estimates to be regular, a property which implies LAM-adaptivity under Stein's (1956) necessary condition for LAM-adaptive estimation and which coincides with Bickel's notion of adaptivity. We demonstrate our method by constructing an LAM-adaptive estimate in a situation where condition S* fails.

To my parents, my wife Jeanette and my son Andreas

ACKNOWLEDGEMENT

I wish to express my sincere thanks to Professor Václav Fabian for his guidance in the preparation of this dissertation. The advice and encouragement he gave are greatly appreciated. Also, I would like to thank Professor H. Koul for awakening my interest in adaptive estimation and for his review of my thesis and Professors J. Hannan and S. Axler for serving on my committee.

Finally, I wish to thank my parents for their constant support throughout my studies and my wife Jeanette for her understanding and encouragement.

TABLE OF CONTENTS

	Page
1. Introduction	1
2. A necessary and sufficient condition	3
3. An example	14
4. An auxiliary result	21
Bibliography	25

1. INTRODUCTION

In a recent paper Fabian and Hannan (1982) use results on locally asymptotically minimax (LAM) estimates for locally asymptotically normal (LAN) families to reformulate Stein's (1956) heuristic arguments on adaptive estimation. The authors define LAM adaptivity of estimates, prove that a condition S, due to Stein, is necessary for the existence of such estimates, and give a sufficient property - regularity - for estimates to be LAM adaptive (see their Theorem 7.10).

Bickel (1982) formulates a condtition S^*, stronger than S. He then constructs regular (and thus LAM adaptive) estimates under Condition S^* and when estimates of the nuisance parameter are available. He uses this result to obtain regular estimates in several important cases.

Bickel states that S^* is "heuristically necessary" for the existence of LAM adaptive estimates (preceding Conditon C, following Condition H). We give a simple counterexample to the necessity of S^* and obtain results on the existence of regular estimates without S^*. It is seen that if S^* does not hold regular estimates are more difficult to construct in that a certain rate of convergence for the estimate of the nuisance parameter is required.

We consider estimates of a certain type and obtain a necessary and sufficient condition for such an estimate to be regular. The class of estimates we consider includes estimates considered by Bickel, but it is a larger class.

The results described above are derived in the case of i.i.d. observations under weaker conditons than Bickel's regularity conditions (see Remark 4.4).

Some notation will be introduced next. If P, Q are probabilities on a σ-algebra $\underline{X}$ and Q_+ is the absolutely continuous, with respect to P, part of Q, then any Radon-Nikodym derivative of Q_+ with respect to P will be called a pseudodensity of Q with respect to P, and also a pseudodensity of $\int \cdot dQ$ with respect to $\int \cdot dP$. We shall talk frequently about expectations using terms which make sense when applied to the probabilities. If E and F are expectations on a σ-algebra $\underline{X}$, then dF/dE denotes the set of all pseudodensities of F with respect to E which are non-negative and finite valued.

$\mathbb{R}^k$ denotes the k-dimensional Euclidean space, $\underline{B}_k$ the σ-algebra of the Borel subsets of $\mathbb{R}^k$, $\mathbb{R} = \mathbb{R}^1$, $\underline{B} = \underline{B}_1$. $< >$ will be used to denote finite or infinite sequences, and, in particular, points in $\mathbb{R}^k$. In matrix calculations, points in $\mathbb{R}^k$ are columns. $\underline{1}$ denotes the identity matrix. The dimension is not displayed in $\underline{1}$ but will be clear from the context.

If H_n is an expectation on a σ- algebra $\underline{X}_n, g_n$ an $<\underline{X}_n, \underline{B}_K>$ measurable transformation for each $n = 1,2,\ldots,$ and if $c \in \mathbb{R}^k$, then (i) we write $g_n \to c$ in $<H_n>$-prob. if $H_n \chi_{\{\|g_n - c\| > \varepsilon\}} \to 0$ for every $\varepsilon > 0$ and (ii) we say $<g_n>$ is bounded in $<H_n>$-prob. if $<F_n>$ is tight, where F_n is the distribution of g_n under H_n.

2. A NECESSARY AND SUFFICIENT CONDITION

We begin by specifying the asymptotic estimation problem we shall consider throughout this paper.

1. Assumption. Θ is a non empty set satisfying $\Theta = \Theta_1 \times \Theta_2$ with Θ_1 an open subset of $\mathbb{R}^m$. $\theta = \langle\theta_1,\theta_2\rangle$ is a point in Θ. $\underline{X}$ is a σ-algebra and for every $\delta \in \Theta$, E_δ is an expectation on $\underline{X}$. $X_1, X_2, \ldots$ is a sequence of s-dimensional random vectors which are independent and identically distributed with distribution F_δ under E_δ, for each $\delta \in \Theta$. There exists a σ-finite integral J such that each F_δ has a density f_δ with respect to J. For every $\delta = \langle t,v\rangle \in \Theta$ the map $u \in \Theta_1 \to f^{\frac{1}{2}}_{\langle u,v\rangle}$ is differentiable at t in $L_2(J)$ with derivative $\dot{h}_\delta = \dot{h}(\cdot,t,v)$. Furthermore $J\,\dot{h}_\theta\dot{h}_\theta^T$ is nonsingular and the map $u \in \Theta_1 \to \dot{h}(\cdot,u,\theta_2)$ is componentwise continuous at θ_1 in $L_2(J)$.

2. Notation. For every $\delta \in \Theta$ and $n = 1,2,\ldots$ we denote by $\underline{X}_n$ the σ-algebra generated by $X_1,\ldots,X_n$ and by $E_{n\delta}$ the restriction of E_δ to $\underline{X}_n$. We set $\mathcal{E} = \langle E_{n\delta}, \delta \in \Theta\rangle$ and $\mathcal{E}_v = \langle E_{n\langle t,v\rangle}, t \in \Theta_1\rangle$ for $v \in \Theta_2$. For convenience in notation we shall often write $f(\cdot,t,v)$ instead of $f_{\langle t,v\rangle}$ and similarly for other functions g_δ, $\delta \in \Theta$. We also set for $\delta \in \Theta$ and $n = 1,2,\ldots$

$$M(\delta) = 4\,J\,\dot{h}_\delta\dot{h}_\delta^T \tag{1}$$

(2) $$\dot{\ell}_\delta = 2\,\frac{\dot{h}_\delta}{f_\delta^{\frac{1}{2}}}\,\chi_{\{f_\delta > 0\}}$$

and

(3) $$\gamma_{n\delta} = (nM(\delta))^{-\frac{1}{2}} \sum_{j=1}^{n} \dot{\ell}_\delta(X_j)$$

By $\tilde{\Theta}$ we denote the set of all $\delta = \langle t,v \rangle$ in Θ for which $M(\delta)$ is nonsingular and $u \in \Theta_1 \to \dot{h}(\cdot,u,v)$ is componentwise continuous at t in $L_2(J)$.

3. Remark. We are interested in estimating the first component δ_1 of an unknown point δ in Θ. Θ_1 specifies our knowledge about this component while Θ_2 summarizes our knowledge about the second component, the nuisance parameter. If $\delta \in \tilde{\Theta}$ we obtain from Theorem 4.8 in Fabian and Hannan (1980) that E_{δ_2} satisfies condition $LAN\langle\delta_1, nM(\delta), \gamma_{n\delta}\rangle$. In this case we want our estimate to be $LAMA(A_\delta,\delta)$ where A_δ is the class of all subproblems which are $LAN\langle n\tilde{M}(\delta), \tilde{\gamma}_{n\delta}\rangle$ for some $\tilde{M}(\delta)$ and $\tilde{\gamma}_{n\delta}$ and satisfy Stein's necessary condition $(\tilde{M}(\delta))_{12} = 0$, see Fabian and Hannan (1982), Section 7. By Theorem 7.10 in the same paper this can be done by constructing an estimate $\langle Z_n \rangle$ which is regular at δ, i.e. satisfies

(1) $$(nM(\delta))^{\frac{1}{2}}(Z_n-\delta_1) - \gamma_{n\delta} \to 0 \quad \text{in} \quad \langle E_{n\delta}\rangle\text{ - prob. .}$$

Since δ is unknown, this suggests to construct an estimate which is (globally) regular, i.e. regular at each point in $\tilde{\Theta}$. In some cases, however, estimates which are regular at just one point, say θ, are also of interest. For this reason we restrict ourselves to the construction of an estimate regular at θ. This will facilitate the treatment and the reader will have no difficulties to see under what conditions this estimate is globally regular.

4. Remark. Bickel (1982) defines adaptivity at θ for an estimate $\langle Z_n \rangle$ by

(1) For every sequence $\langle t_n \rangle$ in Θ_1 such that $\langle n^{\frac{1}{2}}(t_n - \theta_1) \rangle$ is bounded the distribution of $(nM(\theta))^{\frac{1}{2}}(Z_n - t_n)$ under $E_{n \langle t_n, \theta_2 \rangle}$ converges weakly to the m-dimensional standard normal distribution.

Condition (1) is equivalent to (3.1). This follows from Theorem 6.3 in Fabian and Hannan (1982), Theorem 6.1 in Bickel (1982) and the note thereafter. Hence an estimate which is adaptive at θ in Bickel's sense is regular at θ.

Bickel claims that the existence of a regular estimate implies that each subproblem obeying his regularity condition R satisfies Stein's condition $(\tilde{M}(d))_{12} = 0$, for every regular point d. But the proof of this claim is incorrect due to an inappropriate reference to Hájek (1972): Bickel considers only local alternatives for the paramenter of interest and not local alternatives of both the parameter of interest and the nuisance parameter as needed in Hájek's Theorem 4.2. Thus it remains an open question whether Bickel's claim is indeed true.

Next we define a map Q from Θ into $\mathbb{R}^m$ by

$$Q(t,v) = \int \dot{\ell}(\cdot,t,v) f(\cdot,t,\theta_2)$$

if the integral is well defined and 0 otherwise.

Bickel's condition S* is

(S*) $$Q = 0$$

The example below shows that (S*) is not necessary for the construction of regular estimates. Another example is given in Section 3.

5. Example. Let $\Theta_1 = (0,\infty)$, $\Theta_2 = \mathbb{R}$ and let the X_i's be normal random variables with mean μ and standard deviation σ under $E_{<\sigma,\mu>}$, i.e. we want to estimate the standard deviation in the presence of an unknown mean. Easy calculations show that assumption 1 holds and that

(1) $$\dot{\ell}(\cdot,\sigma,\mu) = -\sigma^{-1} + \sigma^{-3}(\cdot-\mu)^2$$

(2) $$\tilde{\Theta} = \Theta$$

and

(3) $$Q(\sigma,\mu) = \sigma^{-3}(\mu-\theta_2)^2 \text{ , for } \sigma > 0,\ \mu,\theta_2 \in \mathbb{R}$$

Furthermore for every $\delta \in \Theta$, the full problem E satisfies condition $\mathrm{LAN}<\delta, n\tilde{M}(\delta), \tilde{\gamma}_{n\delta}>$, where with $\delta = <\sigma,\mu>$

(4) $$\tilde{M}(\delta) = \sigma^{-2}\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

and

(5) $$\tilde{\gamma}_{n\delta} = <(2n)^{-\frac{1}{2}} \sum_{j=1}^{n} \left(\left(\frac{X_j-\mu}{\sigma}\right)^2 - 1\right),\ n^{-\frac{1}{2}} \sum_{j=1}^{n} \frac{X_j-\mu}{\sigma}>$$

and the estimate $<\tilde{Z}_n>$ defined by

(6) $$\tilde{Z}_n = \langle (n^{-1} \sum_{j=1}^{n} (X_j - \bar{X}_n)^2)^{\frac{1}{2}} , \bar{X}_n \rangle$$

with $\bar{X}_n$ the sample average satisfies

(7) $$(n\tilde{M}(\delta))^{\frac{1}{2}}(\tilde{Z}_n - \delta) - \tilde{\gamma}_{n\delta} \to 0 \text{ in } E_\delta \text{ - prob..}$$

Verifications of the above are easy. We refer the reader to example 9.2.12 and Theorem 9.4.33 in Fabian and Hannan (1983). From (4) and (7) we obtain that Stein's necessary condition for LAM-adaptive estimation holds and that $\langle Z_n \rangle$, with Z_n the first component of $\tilde{Z}_n$, is a globally regular estimate.

6. Remark. We now motivate a natural choice for an estimate regular at θ and then give a necessary and sufficient conditon for this type of estimate to be regular at θ. We begin with a definition.

7. Definition. We say $\langle U_n \rangle$ is an auxiliary estimate at θ if each U_n is a Θ_1-valued $\underline{X}_n$ measurable random vector and $\langle n^{\frac{1}{2}}(U_n - \theta_1) \rangle$ is bounded in $\langle E_{n\theta} \rangle$-prob..

We say $\langle W_n \rangle$ is a consistent estimate of the information matrix at θ if the W_n are positive definite matrix valued random vectors on $\underline{X}_n$ such that $M(\theta)^{-\frac{1}{2}} W_n M(\theta)^{-\frac{1}{2}} \to \underline{1}$ in $\langle E_{n\theta} \rangle$ -prob..

We say $\langle t_n \rangle$ is a local sequence if $\langle t_n \rangle$ is a sequence in Θ_1 and $\langle n^{\frac{1}{2}}(t_n - \theta_1) \rangle$ is bounded.

8. Remark. In Section 4 we prove that

(1) $$Z_n(U_n, \theta_2) = U_n + (nW_n)^{-1} \sum_{j=1}^{n} \dot{\ell}(X_j, U_n, \theta_2)$$

is regular at θ if $<U_n>$ is a discrete auxiliary estimate at θ and $<W_n>$ is a consistent estimate of the information matrix at θ. For a discussion and the use of discrete estimates we refer to Fabian and Hannan (1982) and Bickel (1982). The estimate in (1) is of limited practical value since it presupposes the knowledge of θ_2, but it suggests an obvious candidate for a regular estimate. Simply replace θ_2 by an estimate. A different method consists in estimating the score-function $\dot{\ell}(\cdot,\cdot,\theta_2)$ directly as Bickel (1982) does. But the present method serves us better for the purpose of illustration. Substituting an estimate for θ_2 in (1) has to be done with some care. For technical reasons we adopt an idea Bickel (1982) uses, but modify it to obtain better estimates of the nuisance parameter. Recall that Bickel splits the sample in two unequal parts, estimates the nuisance parameter based on the observations in the smaller subsample and evaluates the scorefunction only at observations of the larger part. We divide the sample in two equal parts, obtain an estimate of the nuisance parameter from each part and when evaluating the scorefunction with an observation of the first part we use the estimate of the nuisance parameter based on the second part and vice versa. Thus our estimates of the nuisance parameter are based on half the sample and not just on a small proportion of the sample. This improvement is vital, since it turns out that the estimate described above is regular at θ if we can construct an estimate of the nuisance parameter with a certain rate of convergence.

To have the above method well defined we make the following assumption.

9. Assumption. Assumption 1 holds. Θ_2 is a topological space. The map $\dot{\ell}(\cdot,t,\cdot)$ is measurable for each $t \in \Theta_1$ and

$$(1) \qquad \int \|\dot{\ell}(\cdot,t,v) - \dot{\ell}(\cdot,t,\theta_2)\|^2 f(\cdot,t,\theta_2) \to 0$$

as $\langle t,v\rangle$ in Θ converges to θ.

For every $n = 1,2,\ldots$ there is a measurable map h_n from $(\mathbb{R}^s)^n$ into Θ_2 such that $V_n = h_n(X_1,\ldots,X_n)$ converges to θ_2 in E_θ-prob.. $\langle U_n\rangle$ is an auxiliary estimate at θ and $\langle W_n\rangle$ is a consistent estimate of the information matrix at θ.

10. Remarks and Notation. Note that Assumption 9 implies (3.11) of condition H' in Bickel (1982) with $\hat{\dot{\ell}}_n(\cdot,\cdot,X_1,\ldots,X_n) = \dot{\ell}(\cdot,\cdot,V_n)$. Indeed, we have

$$(1) \qquad \int \|\dot{\ell}(\cdot,t_n,V_{k_n}) - \dot{\ell}(\cdot,t_n,\theta_2)\|^2 f(\cdot,t_n,\theta_2) \to 0$$

in E_θ-prob. for every sequence $\langle t_n\rangle$ in Θ_1 converging to θ_1 and every sequence of integers $\langle k_n\rangle$ tending to infinity.

Also observe that under Assumption 9 $\int \dot{\ell}(\cdot,t,v)f(\cdot,t,\theta_2)$ is well defined for $\langle t,v\rangle$ in a neighborhood of θ.

The estimate described in Remark 8 is formally defined by

$$(2) \qquad \hat{Z}_n(\bar{U}_n) = \bar{U}_n + (nW_n)^{-1}\Big(\sum_{j=1}^{m_n} \dot{\ell}(X_j,\bar{U}_n,V_{n2}) + \sum_{j=m_n+1}^{n} \dot{\ell}(X_j,\bar{U}_n,V_{n1})\Big)$$

with $\langle \bar{U}_n\rangle$ a discrete auxiliary estimate at θ, m_n the integer part of $n/2$ and

$$(3) \qquad V_{n1} = h_{m_n}(X_1,\ldots,X_{m_n}) \text{ and } V_{n2} = h_{n-m_n}(X_{m_n+1},\ldots,X_n) .$$

Typically $\langle\bar{U}_n\rangle$ will be a discretized version of $\langle U_n\rangle$. But we do not want the regularity at θ of $\langle\hat{Z}_n(\bar{U}_n)\rangle$ to depend on the way we discretize. In other words, we want $\langle\hat{Z}_n(\bar{U}_n)\rangle$ to be regular at θ for each discrete auxiliary estimate $\langle\bar{U}_n\rangle$ at θ. We now give a necessary and sufficient condition for this to happen.

11. Theorem. Suppose Assumption 9 holds. Then the following are equivalent.

(1) $\langle\hat{Z}_n(\bar{U}_n)\rangle$ is regular at θ for every discrete auxiliary estimate $\langle\bar{U}_n\rangle$ at θ.

(2) For every local sequence $\langle t_n\rangle$

$$n^{\frac{1}{2}}Q(t_n,V_n) \to 0 \quad \text{in} \quad E_\theta\text{-prob.}.$$

Proof: Note that (1) is equivalent to

(3) $\langle\hat{Z}_n(t_n)\rangle$ is regular at θ for every local sequence $\langle t_n\rangle$.

Also recall (see Remark 8) that

(4) $\langle Z_n(t_n,\theta_2)\rangle$ is regular at θ for every local sequence $\langle t_n\rangle$.

We shall show that for every local sequence $\langle t_n\rangle$

(5) $n^{\frac{1}{2}}(\hat{Z}_n(t_n) - Z_n(t_n,\theta_2) - W_n^{-1}R_n(t_n)) \to 0$ in E_θ prob.,

where $R_n(t) = n^{-1}(m_nQ(t,V_{n2}) + (n-m_n)Q(t,V_{n1}))$.

Combining the above shows that (1) is equivalent to

(6) $n^{\frac{1}{2}}W_n^{-1}R_n(t_n) \to 0$ in E_θ-prob. for every local sequence $<t_n>$.

By the consistency of $<W_n>$ and the independence of V_{n1} and V_{n2} (6) is equivalent to (2). Thus we are left to verify (5). Again by the consistency of $<W_n>$, (5) is equivalent to

(7) $n^{-\frac{1}{2}}(\sum_{j=1}^{m_n} T_{n2}(X_j,t_n) + \sum_{j=m_n+1}^{n} T_{n1}(X_j,t_n)) \to 0$ in E_θ-prob.,

where $T_{ni}(\cdot,t) = \dot{\ell}(\cdot,t,V_{ni}) - \dot{\ell}(\cdot,t,\theta_2) - Q(t,V_{ni})$ for $i = 1,2$ and $t \in \Theta_1$.

Now fix a local sequence $<t_n>$. Abbreviate $E_{n<t_n,\theta_2>}$ by $\bar{E}_n$ and note that $<\bar{E}_n>$ and $<E_{n\theta}>$ are mutually contiguous. Next observe that for $j = 1,2,\ldots,m_n$

(8) $\bar{E}_n(T_{n2}(X_j,t_n)|X_{m_n+1},\ldots,X_n) = 0$ a.e. $\bar{E}_n$

and thus

(9) $$\bar{E}_n(\|n^{-\frac{1}{2}} \sum_{j=1}^{m_n} T_{n2}(X_j,t_n)\|^2|X_{m_n+1},\ldots,X_n)$$

$$= n^{-1} \sum_{j=1}^{m_n} \bar{E}_n(\|T_{n2}(X_j,t_n)\|^2|X_{m_n+1},\ldots,X_n) \text{ a.e. } \bar{E}_n$$

$$\leq J \|\dot{\ell}(\cdot,t_n,V_{n2}) - \dot{\ell}(\cdot,t_n,\theta_2)\|^2 f(\cdot,t_n,\theta_2) \text{ a.e. } \bar{E}_n$$

by a property of conditional variances. Using the mutual contiguity of $<\bar{E}_n>$ and $<E_{n\theta}>$ and (10.1) we find that (10) converges to zero in $<\bar{E}_n>$-prob.. This shows that

(10) $n^{-\frac{1}{2}} \sum_{j=1}^{m_n} T_{n2}(X_j,t_n) \to 0$ in $<\bar{E}_n>$ -prob..

In the same way we obtain that

(11) $n^{-\frac{1}{2}} \sum_{j=m_n+1}^{n} T_{n1}(X_j,t_n) \to 0$ in $\langle \bar{E}_n \rangle$-prob..

Using the mutual contiguity of $\langle \bar{E}_n \rangle$ and $\langle E_{n\theta} \rangle$ we conclude from (10) and (11) that (7) holds. This completes the proof.

12. Remarks. Note that (11.2) is trivially satisfied if (S^*) holds and in this case consistency of $\langle V_n \rangle$ guarantees the existence of a estimate regular at θ. This is Bickel's (1982) result. But if (S^*) fails consistency of $\langle V_n \rangle$ alone does not suffice to construct an estimate regular at θ. In this sense LAM-adaptive estimation is more difficult in cases when (S^*) fails.

Assume for the moment that Θ_2 is an open subset of $\mathbb{R}^p$ for some positive integer p and that the whole problem E satisfies conditon LAN $\langle \delta, n\tilde{M}(\delta), \tilde{\gamma}_{n\delta} \rangle$ for each $\delta \in \Theta$. Also assume regularity conditions which allow the Taylor expansion

$$Q(t,v) = Q(t,\theta_2) - (v-\theta_2)^T \tilde{M}_{12}(\theta) + o(\|t-\theta_1\|) + O(\|v-\theta_2\|^2)$$

as $\langle t,v \rangle \to \theta$. In this case the necessary conditon for LAM-adaptive estimation $\tilde{M}_{12}(\theta) = 0$ implies that

(1) $$Q(t,v) = o(\|t-\theta_1\|) + O(\|v-\theta_2\|^2)$$

Note that $Q(t,\theta_2) = 0$. Thus (1) shows that (11.2) is satisfied if

(2) $$n^{\frac{1}{4}}(V_n-\theta_2) \to 0 \quad \text{in} \quad E_\theta\text{-prob..}$$

Obviously (2) is weaker than

(3) $n^{\frac{1}{2}}(V_n - \theta_2)$ is bounded in E_θ-prob.,

a condition which together with the existence of an auxiliary estimate at θ and of a consistent estimate of the information matrix at θ suffices to construct LAM-adaptive estimates if Stein's condition holds (c.f. Theorem 6.15 and Theorem 7.10 in Fabian and Hannan (1982)).
We remind the reader that (1) is satisfied in example 5.

3. AN EXAMPLE

1. Description of the example

We consider the regression model

(1) $$Y_j = \theta_1 + \theta_2(T_j) + \varepsilon_j \qquad j = 1,2,\ldots$$

where $T_1, T_2, \ldots$ are i.i.d. random variables with uniform distribution on $[0,1]$, $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. random variables with Lebesgue density g and independent of $T_1, T_2, \ldots, \theta_1$ is a real number and θ_2 is a real valued absolutely continuous function on $[0,1]$ with square integrable derivative θ_2' and $\int_0^1 \theta_2(t)dt = 0$. We suppose that the density g satisfies the following conditions

(2) $\int x \, g(x)dx = 0$

(3) $\int x^2 \, g(x)dx = \sigma^2 < \infty$

(4) g is absolutely continuous with derivative g' and has finite Fisher information

$$I(g) = \int \frac{(g'(x))^2}{g(x)} \, dx$$

(5) With $L = -\frac{g'}{g} \chi_{\{g > 0\}}$ we have

(5a) $$\int_0^1 \int_{-\infty}^{\infty} (L(x + v(t)) - L(x))^2 \, g(x)dxdt \to 0$$

and

$$\text{(5b)} \qquad \int_0^1 \int_{-\infty}^{\infty} L(x-v(t))g(x)dx\,dt = 0\left(\int_0^1 v^2(t)dt\right)$$

$$\text{for} \int_0^1 v(t)dt = 0 \text{ and } \int_0^1 v^2(t)dt \to 0 .$$

Note that (5) is satisfied if L is twice continuously differentiable with bounded derivatives L' and L''.

2. Remark. The above regression model satisfies Assumption 2.1 with $\Theta_1 = \mathbb{R}$, Θ_2 the family of all realvalued functions v on $[0,1]$ which are absolutely continuous with square integrable derivative and satisfy $\int_0^1 v(t)dt = 0$, $X_j = \langle Y_j, T_j \rangle$, J the integral induced by the Lebesgue measure on the Borel field of $\mathbb{R} \times [0,1]$ and f_δ and h_δ defined by $f(x,t,v) = g(x_1-t-v(x_2))$ and $h(x,t,v) = \frac{1}{2} L(x_1-t-v(x_2))g^{\frac{1}{2}}(x_1-t-v(x_2))$ with $x = \langle x_1,x_2 \rangle$ in $\mathbb{R} \times [0,1]$ and $\delta = \langle t,v \rangle$ in Θ. The differentiability in $L_2(J)$ follows from (1.4) and Lemma A.3 in Hájek (1972), while the required continuity of the derivative is a consequence of Theorem 9.5 in Rudin (1974). Note also that $\tilde{\Theta} = \Theta$ by the translation invariance of the Lebesgue measure.

Furthermore, if we endow Θ_2 with the topology induced by the norm $\|\cdot\|_2$ defined by $\|v\|_2^2 = \int_0^1 v^2(t)dt$ for v in Θ_2, then (2.9.1) follows from (1.5). Also observe that the sample average $\langle n^{-1} \sum_{j=1}^{n} Y_j \rangle$ is an auxiliary estimate at θ and that the sequence $\langle I(g), I(g), \ldots \rangle$ is a consistent estimate of the information matrix at θ.

Next it is easily checked that Q satisfies

$$Q(t,v) = \int_0^1 \int_{-\infty}^{\infty} L(x-(v(u)-\theta_2(u)))g(x)dxdu$$

This and (3) show that (2.11.2) is satisfied if $\langle V_n \rangle$ satisfies

(1) $$n^{\frac{1}{2}} \| V_n - \theta_2 \|_2^2 \to 0 \quad \text{in} \quad E_\theta\text{-prob.}.$$

We shall now construct such an estimate.

3. Construction of the estimate $\langle V_n \rangle$.

We let $\langle a_n \rangle$ denote a seouence of positive integers and set $b_n = a_n^{-1}$. For each $n = 1,2,\ldots$ we partition the unit interval $[0,1]$ in a_n intervals I_{ni}, $i = 1,\ldots,a_n$ of equal length b_n. We let m_{ni} denote the midpoint of I_{ni} and χ_{ni} the indicator of I_{ni}. Furthermore we assume that the intervals I_{ni} are numbered in such a way that $m_{nj} < m_{nk}$ for $1 \le j < k \le a_n$. Next we set

(1) $$U_n = n^{-1} \sum_{j=1}^{n} Y_j$$

and

(2) $$Y_{ni} = (nb_n)^{-1} \sum_{j=1}^{n} Y_j \chi_{ni}(T_j) \quad , \quad i = 1,\ldots,a_n$$

and define V_n by

(3) $$V_n(t) = \begin{cases} Y_{n1} - U_n & 0 \le t \le m_{n1} \\ Y_{ni} - U_n + \dfrac{t - m_{ni}}{b_n}(Y_{ni+1} - Y_{ni}) & m_{ni} \le t < m_{ni+1} \\ Y_{na_n} - U_n & m_{na_n} \le t \le 1 \end{cases}$$

It is easily verified that V_n is a Θ_2-valued random vector, e.g.

(4) $$\int_0^1 V_n(t)dt = \sum_{i=1}^{a_n} b_n Y_{ni} - U_n = 0$$

4. Lemma. If the sequence $\langle a_n \rangle$ is chosen such that

(1) $$nb_n^4 \to 0 \quad \text{and} \quad nb_n^2 \to \infty$$

then $$n^{\frac{1}{2}} E_\theta \| V_n - \theta_2 \|_2^2 \to 0$$

Proof: For $i = 1,2,\ldots,a_n$ set

(2) $$C_{ni} = a_n \int_0^1 \chi_{ni}(u)\theta_2(u)du$$

and note that

(3) $$E_\theta Y_{ni} = \theta_1 + C_{ni}$$

Easy calculations show that

(4) $$E_\theta (Y_{ni} - \theta_1 - C_{ni})^2 \leq 3(nb_n)^{-1}(\theta_1^2 + \|\theta_2\|_2^2 + \sigma^2)$$

and

(5) $$E_\theta (U_n - \theta_1)^2 \leq n^{-1}(\sigma^2 + \|\theta_2\|_2^2)$$

Next note that by the Schwarz inequality for $0 \leq u_1 < u_2 \leq 1$

(6) $$(\theta_2(u_2) - \theta_2(u_1))^2 = \left(\int_{u_1}^{u_2} \theta_2'(x)dx\right)^2 \leq (u_2 - u_1) \int_{u_1}^{u_2} (\theta_2'(x))^2 dx$$

Using this and the Schwarz inequality we obtain

(7) $$\int (\theta_2(t) - C_{ni})^2 \chi_{ni}(t)dt$$
$$= \int \left(a_n \int (\theta_2(t) - \theta_2(u))\chi_{ni}(u)du\right)^2 \chi_{ni}(t)dt$$
$$\leq a_n \iint (\theta_2(t) - \theta_2(u))^2 \chi_{ni}(u)du\chi_{ni}(t)dt$$
$$\leq b_n^2 \|\theta_2' \chi_{ni}\|_2^2$$

and

(8) $$(C_{ni+1}-C_{ni})^2 \leq a_n^2 \int_0^1 \int_0^1 (\theta_2(t)-\theta_2(u))^2 \chi_{ni+1}(u)\chi_{ni}(t)dudt$$

$$\leq 2b_n \|(\chi_{ni+1} + \chi_{ni})\theta_2'\|_2^2$$

Combining (4) and (8) shows that for some constant C

(9) $$\sum_{i=1}^{a_n-1} E_\theta(Y_{ni+1}-Y_{ni})^2 \leq C(n^{-1}b_n^{-2}+ b_n)$$

Next define

(10) $$\overline{V}_n = \sum_{i=1}^{a_n} Y_{ni}\chi_{ni} - U_n$$

and

(11) $$\tilde{V}_n = \sum_{i=1}^{a_n} C_{ni}\chi_{ni}$$

It follows form (1) and (9) that

(12) $$n^{\frac{1}{2}}E_\theta\|V_n-\overline{V}_n\|_2^2 \leq n^{\frac{1}{2}} \sum_{i=1}^{a_n-1} E_\theta(Y_{ni+1}-Y_{ni})^2 b_n \to 0$$

and from (1), (4) and (5) that

(13) $$n^{\frac{1}{2}}E_\theta\|\overline{V}_n-\tilde{V}_n\|_2^2 \to 0 .$$

Furthermore by (1) and (7)

(14) $$n^{\frac{1}{2}}\|\tilde{V}_n-\theta_2\|_2^2 \to 0 .$$

Combining (12) to (14) gives the desired result.

5. Remark. Remark 2 and Lemma 4 show that Assumption 2.10 and condition (2.11.2) are satisfied. Therefore it follows from Theorem 2.11 that

$$(1) \qquad \bar{U}_n + (nI(g))^{-1}\Big(\sum_{j=1}^{m_n} L(Y_j-\bar{U}_n-V_{n2}(T_j))+ \sum_{j=m_n+1}^{n} L(Y_j-\bar{U}_n-V_{n1}(T_j))\Big)$$

is a regular estimate, where $\langle \bar{U}_n \rangle$ is a discretized version of $\langle U_n \rangle$ and m_n, V_{n1} and V_{n2} are as described in Remark 2.10.

Next observe that conditon (S^*) does not hold for a proper choice of g, e.g. with $g(x) = \frac{1}{2} e^{-|x|}$ we obtain for $v = \theta_2 + r$ in Θ_2 by easy calculations

$$\begin{aligned}(2) \qquad Q(t,v) &= \int_0^1\int_{-\infty}^{\infty} \operatorname{sign}(x - r(u))g(x)dx\, du \\ &= \int_0^1 \operatorname{sign}(r(u))(e^{-|r(u)|} - 1)du \\ &= \int_0^1 \operatorname{sign}(r(u))(e^{-|r(u)|} - 1 - |r(u)|)du \\ &= O(\|r\|_2^2)\end{aligned}$$

and from (2) it is easily seen that Q is not identically zero. This shows that also in an infinite-demensional nuisance parameter space Θ_2 conditon S^* is not necessary for the existence of a regular estimate.

6. Remark. The estimate $\langle Z_n \rangle$ defined by (5.1) is LAM-adaptive (A_θ, θ) where A_θ is the class of all LAN $\langle n\tilde{M}(\theta), \tilde{\gamma}_{n\theta} \rangle$ subproblems with $\tilde{M}(\theta)_{12} = 0$ (see Remark 2.3). We now describe a class of subproblems which belong to A_θ:

Let Γ denote the family of all one to one maps γ from an open neighborhood S around 0 in $\mathbb{R}^p$ into Θ_2 for some positive integer p satisfying $\gamma(0) = \theta_2$ and

(1) $$\|\gamma(a)-\gamma(0) - a^T\psi\|_2 = O(\|a\|) \quad \text{as} \quad a \to 0$$

for some vector $\psi = \langle\psi_1,\ldots,\psi_p\rangle$ such that ψ_i is in Θ_2 for $i = 1,\ldots,p$ and $W = \int_0^1 \psi(u)\psi^T(u)du$ is nonsingular.

For γ in Γ we define the subproblem $\langle\Theta_0,\alpha\rangle$ by $\Theta_0 = \Theta_1 \times \gamma[S]$ and $\alpha(t,\gamma(a)) = \langle t,a\rangle$ for $\langle t,a\rangle \in \mathbb{R} \times S$ (for the defintion of subproblem see Definition 7.3 in Fabian and Hannan (1982)). This subproblem satisfies condition LAN $\langle n\tilde{M},\tilde{\gamma}_n\rangle$ with

(2) $$\tilde{M} = I(g) \begin{bmatrix} 1 & 0 \\ 0 & W \end{bmatrix}$$

and

(3) $$\tilde{\gamma}_n = (n\tilde{M})^{-\frac{1}{2}} \sum_{j=1}^{n} L(Y_j-\theta_1-\theta_2(T_j)) \langle 1,\psi(T_j)\rangle$$

and hence satisfies $\tilde{M}_{12} = 0$.

The above follows from Theorem 4.8 in Fabian and Hannan (1980), since the map $\langle t,a\rangle \in \mathbb{R} \times S \to f^{\frac{1}{2}}(\cdot,t,\gamma(a))$ is differentiable in $L_2(J)$ at $\langle\theta_1,0\rangle$ with derivative λ given by

(4) $$\lambda(x) = \dot{h}(x_1,\theta_1,\theta_2) \langle 1,\psi(x_2)\rangle \quad \text{for} \quad x = \langle x_1,x_2\rangle \quad \text{in} \quad \mathbb{R}\times[0,1]$$

This is easily verified using (1.4), the arguments in the proof of Lemma A.3 in Hájek (1972), Theorem 9.5 in Rudin (1974) and the properties of γ.

4. AN AUXILIARY RESULT

1. Remark and Notation. In this section we shall prove that the estimate $<Z_n(U_n,\theta_2)>$ as given in (2.8.1) is regular at θ. To simplify notation we abbreviate $M(\theta)$ by M and $\gamma_{n\theta}$ by γ_n. Also we shall use $E_t, E_{nt}, M(t)$ and γ_{nt} short for $E_{<t,\theta_2>}, E_{n<t,\theta_2>}, M(t,\theta_2)$ and $\gamma_{n<t,\theta_2>}$, with $t \in \Theta_1$.

2. Lemma. Suppose Assumption 2.1 holds, $<t_n>$ and $<u_n>$ are local sequences and $g_n \in dE_{nu_n}/dE_{nt_n}$. Then

$$\log g_n - w_n^T \tilde{\gamma}_{nt_n} + \tfrac{1}{2}\|w_n\|^2 \to 0 \quad \text{in } E_{\theta_1}\text{-prob.},$$

with $\quad w_n = (nM)^{\frac{1}{2}}(u_n - t_n)$ and $\tilde{\gamma}_{nt_n} = (nM)^{-\frac{1}{2}} \sum_{j=1}^{n} \dot{\ell}(X_j, t_n, \theta_2)$.

Proof: Let $s_n = t_n - \theta_1$ and $T_n = \{t \in \Theta_1 : t + s_n \in \Theta_1\}$ and set for $t \in T_n$, $H_{nt} = E_{nt+s_n}$, $h_{nt} = f^{\frac{1}{2}}(\cdot, t+s_n, \theta_2)$ and $\dot{h}_{nt} = \dot{h}(\cdot, t+s_n, \theta_2)$.

Using the fact that the map $t \in \Theta_1 \to f^{\frac{1}{2}}(\cdot, t, \theta_2)$ is continuously differentiable at θ_1 in $L_2(J)$ we obtain for every local sequence $<\delta_n>$ with δ_n in T_n and for every $\varepsilon > 0$

(1) $$n\,J(h_{n\delta_n}-h_{n\theta_1}-(\delta_n-\theta_1)^T\dot{h}_{n\theta_1})^2 \to 0$$

(2) $$J\|\dot{h}_{n\theta_1}\|^2\chi_{\{\|\dot{h}_{n\theta_1}\| > n^{\frac{1}{2}}\varepsilon h_{n\theta_1}\}} \to 0$$

and

(3) $$M(t_n) \to M$$

From (3) we obtain that $M(t_n)$ is invertible for all $n \geq n_0$, for some integer n_0. We now obtain by Theorem 4.5 in Fabian and Hannan (1980) that the family $\langle H_{nt}, t \in T_n, n \geq n_0\rangle$ satisfies condition LAN $\langle nM(t_n), \gamma_{nt_n}\rangle$. This shows that

(4) $$\log g_n - \tilde{w}_n^T \gamma_{nt_n} + \tfrac{1}{2}\|\tilde{w}_n\|^2 \to 0 \quad \text{in } \langle H_{n\theta_1}\rangle\text{-prob.},$$

with $\tilde{w}_n = (nM(t_n))^{\frac{1}{2}}(u_n-t_n)$. Now note that

(5) $$w_n^T \tilde{\gamma}_{nt_n} = \tilde{w}_n^T \gamma_{nt_n}$$

and that by (3)

(6) $$\|\tilde{w}_n\|^2 - \|w_n\|^2 \to 0$$

The desired result follows now from (4), (5) and (6) and the mutual contiguity of $\langle H_{n\theta_1}\rangle$ and $\langle E_{n\theta_1}\rangle$.

3. <u>Proposition</u>. Suppose Assumption 2.1 holds, $\langle U_n\rangle$ is a discrete auxiliary estimate at θ and $\langle W_n\rangle$ is a consistent estimate of the information matrix at θ. Then

$$Z_n = U_n + (nW_n)^{-1}\sum_{j=1}^{n} \dot{\ell}(X_j, U_n, \theta_2)$$

is regular at θ.

Proof: We have to show that

(1) $$(nM)^{\frac{1}{2}}(Z_n-\theta_1) - \gamma_n \to 0 \text{ in } E_{\theta_1}\text{-prob..}$$

By the discreteness of $\langle U_n \rangle$ and the consistency of $\langle W_n \rangle$ it suffices to show that

(2) $$(nM)^{\frac{1}{2}}(t_n-\theta_1) + \tilde{\gamma}_{nt_n} - \gamma_n \to 0 \text{ in } E_{\theta_1}\text{-prob.,}$$

for every local sequence $\langle t_n \rangle$.

Let $\langle t_n \rangle$ be a local sequence and set $u_n = t_n + (nM)^{-\frac{1}{2}}u$ for $u \in \mathbb{R}^m$. With $g_n \in dE_{nu_n}/dE_{nt_n}$ we obtain from Lemma 4.3 in Fabian and Hannan (1982) and the mutual contiguity of $\langle E_{nt_n} \rangle$ and $\langle E_{n\theta_1} \rangle$

(3) $$\log g_n - u^T(\gamma_n-\hat{t}_n) + \tfrac{1}{2}\|u\|^2 \to 0 \text{ in } E_{\theta_1}\text{-prob.,}$$

with $\hat{t}_n = (nM)^{\frac{1}{2}}(t_n-\theta_1)$. On the other hand Lemma 2 shows that

(4) $$\log g_n - u^T \tilde{\gamma}_{nt_n} + \tfrac{1}{2}\|u\|^2 \to 0 \text{ in } E_{\theta_1}\text{-prob..}$$

Combining (3) and (4) shows that

(5) $$u^T(\hat{t}_n + \tilde{\gamma}_{nt_n} - \gamma_n) \to 0 \text{ in } E_{\theta_1}\text{-prob.}$$

for every $u \in \mathbb{R}^m$. From this (2) follows which concludes the proof.

4. Remark. Bickel (1982) constructs an estimate as in Proposition 3 under stronger conditions than ours. It is easily checked that his regularity conditions R(i), R(ii) and UR(iii) imply continuous differentiability at θ_1 in $L_2(J)$. Also note that we can choose W_n to

be $M(U_n)$ if the latter is nonsingular and $\underline{1}$ otherwise. However, estimates which are regular at θ and do require the knowledge of θ_2 can be constructed under weaker conditions than ours; see Theorem 6.15 in Fabian and Hannan (1982). The estimates constructed there are based on difference quotients rather than on the "derivative" $\dot{\ell}$. Since the use of $\dot{\ell}$ facilitates our treatment we have chosen to work with Assumption 2.1.

BIBLIOGRAPHY

BIBLIOGRAPHY

Bickel, P.J. (1982). On adaptive estimation. Ann. Statist. 10, 647-671.

Fabian, V. and Hannan, J. (1980). Sufficient conditions for local asymptotic normality. RM-403. Department of Statistics and Probability, Michigan State University.

Fabian, V. and Hannan, J. (1982). On estimation and adaptive estimation for locally asymptotically normal families. Z. Wahrscheinlichkeitstheorie verw. Gebiete 59, 459-478.

Fabian, V. and Hannan, J. (1983). Introduction to Probability and Mathemathical Statistics. (Forthcoming book)

Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. Proc. Sixth Berkeley Sympos. Math. Statist. Probab. I, 175-194. Univ. Calif. Press (1972).

Rudin, W. (1974). Real and complex analysis, (2nd ed.). McGraw Hill, New York.

Stein, C. (1956). Efficient nonparametric testing and estimation. Proc. Third Berkeley Synpos. Math. Statist. and Probab. I, 187-195. Univ. Calif. Press.