

Asymptotic Efficiency of Estimates for Models with Incidental Nuisance Parameters



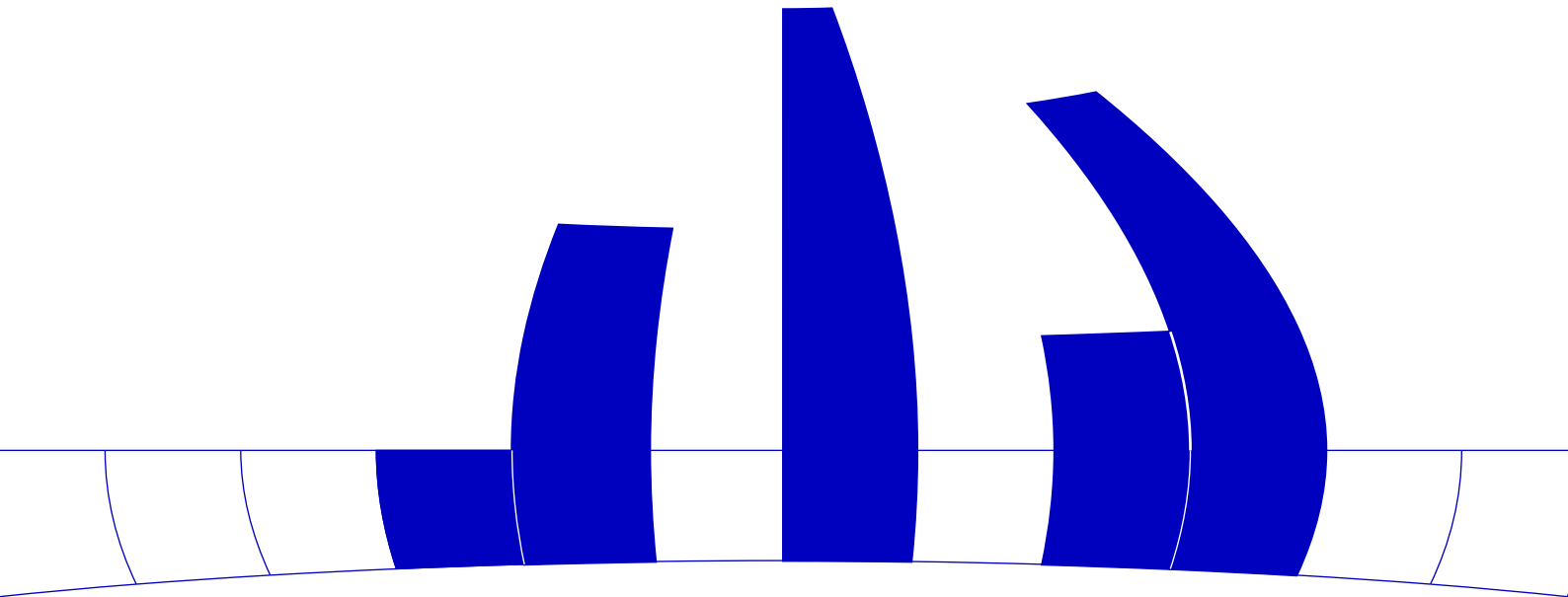
Helmut Strasser

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 34
March 1995

<http://statistik.wu-wien.ac.at/>



Asymptotic Efficiency of Estimates for Models with Incidental Nuisance Parameters

Helmut Strasser

March 1995

Abstract

In this paper we show that the well-known asymptotic efficiency bounds for full mixture models remain valid if individual sequences of nuisance parameters are considered. This is made precise both for some classes of random (i.i.d.) and non-random nuisance parameters. For the random case it is shown that superefficiency of the kind given by an example of Pfanzagl, [7], can happen only with low probability. The non-random case deals with permutation invariant estimators under one-dimensional nuisance parameters. It is shown that the efficiency bounds remain valid for individual non-random arrays of nuisance parameters whose empirical process, if it is centered around its limit and standardized, satisfies a compactness condition. The compactness condition is satisfied in the random case with high probability. The results make use of basic LAN-theory. Regularity conditions are stated in terms of L^2 -differentiability.

1 Introduction

In this section we will give an overview over the main results and the leading ideas of the paper. Moreover, we will discuss the limitations of the results.

Suppose that $\Theta \subseteq \mathbb{R}^p$ is an open set and that (Λ, \mathcal{C}) is a measurable space. Let $(P_{\theta, \eta} : \theta \in \Theta, \eta \in \Lambda)$ be a family of μ -continuous probability measures defined on a measure space $(\Omega, \mathcal{A}, \mu)$. The subject of this paper is the estimation of the parameter θ if the second parameter η is a nuisance parameter which is not observed and varies from observation to observation. Let (κ_n) be an estimator sequence and let \mathcal{W} be the set of all bounded, continuous and bowl-shaped loss functions defined on \mathbb{R}^p . We want to find local asymptotic bounds for the risks

$$\int W(\sqrt{n}(\kappa_n - \theta)) d \bigotimes_{i=1}^n P_{\theta, \eta_i} \quad (1.1)$$

where $W \in \mathcal{W}$ and $(\eta_i) \in \Lambda^n$.

A starting point of the present paper is the paper [7] by Pfanzagl, 1993. We are recommending this source also for its detailed discussion of the history of the subject. In this paper [7] bounds for the asymptotic risk of estimators are considered being valid under i.i.d. observations from mixtures of the probability measures $P_{\theta, \eta}$. The question is discussed whether these bounds remain valid if individual sequences of nuisance parameters (incidental nuisance parameters) are considered.

Before we recall the message of [7] let us describe the situation a bit more detailed.

Let Γ be a probability measure on the (Λ, \mathcal{C}) and let Y_1, Y_2, \dots, Y_n be random variables being i.i.d. according to Γ . We have to distinguish between the risks

$$\int W(\sqrt{n}(\kappa_n - \theta)) d \bigotimes_{i=1}^n P_{\theta, Y_i} \quad (1.2)$$

¹AMS 1991 subject classifications. Primary: 62A20, 62B15; secondary: 62F05, 62F12

²Key words and phrases. Regularity, differentiability in quadratic mean, sufficiency, equivalence of experiments.

for individual realizations of nuisance parameters, and the expected risks

$$E\left(\int W(\sqrt{n}(\kappa_n - \theta)) d\bigotimes_{i=1}^n P_{\theta, Y_i}\right) \quad (1.3)$$

of an estimator sequence. Let us explain that a theory for the expected risks is equivalent to a theory for i.i.d. observations from mixtures of probability measures. For this we denote

$$Q_{\theta, \Gamma}(A \times B) = \int_B P_{\theta, \eta}(A) \Gamma(d\eta), \quad A \times B \in \mathcal{A} \otimes \mathcal{C},$$

and $Q'_{\theta, \Gamma} = Q_{\theta, \Gamma}|_{\mathcal{A}}$. Then the expected risks can be written as

$$E\left(\int W(\sqrt{n}(\kappa_n - \theta)) d\bigotimes_{i=1}^n P_{\theta, Y_i}\right) = \int W(\sqrt{n}(\kappa_n - \theta)) dQ'_{\theta, \Gamma}$$

Thus, we can treat the problem of expected risks as if the observations were i.i.d. according to the mixture distributions $Q'_{\theta, \Gamma}$. For this a complete theory of local asymptotic efficiency in estimation is available and well-known (cf. Pfanzagl and Wefelmeyer, [8], section 14, Bickel, Klaassen, Ritov and Wellner, [1], section 4.5., or Strasser, [11], section 20.3). If the class of distributions of the nuisance parameters is sufficiently rich (full mixture models) then asymptotic efficient estimator sequences can be characterized by an asymptotically linear expansion.

The problem is whether the asymptotic bounds obtained from the full mixture model for the expected risks (1.3) remain valid bounds for risks (1.1) or (1.2) with incidental nuisance parameters.

Roughly speaking, Pfanzagl's paper [7] implies the following assertion (Theorem (3.1) in [7]): If an estimator sequence is asymptotically linear with an expansion being optimal for the full mixture model, then for almost all fixed realizations of a random nuisance parameter sequence the asymptotic risks (1.1) converge to the risk bound of the full mixture model. However, it must be questioned whether the asymptotic *optimality* of the risk in the mixture model remains valid for incidental nuisance parameters, too.

In his paper, Pfanzagl presents a counterexample (Example (2) in [7]): There is an estimator sequence which keeps the asymptotic risk bound of the full mixture model for almost all realizations of a random nuisance parameter, but underflows the bound for countably many further realizations. This phenomenon could be interpreted as a kind of superefficiency.

Our first main result shows that this kind of superefficiency as shown by Pfanzagl can occur only with small probability. This is Theorem (2.13) of the present paper and it is presented and proved in detail in section 2. Here we are giving only a rough statement of the assertion. The message is as follows:

(1.4) *Consider the risks (1.3). If for a stochastic array (Y_{ni}) of nuisance parameters being i.i.d. according to Γ the efficiency bound for the full mixture model is underflowed with positive probability for some loss function W , then the efficiency bound is overflowed with positive probability for some possible other loss function W .*

The result in (1.4) deserves some general comments.

(1.5) REMARK

The assertion in (1.4) and of Theorem (2.13) is an asymptotic admissibility assertion. A similar asymptotic admissibility assertion has been proved for the first time by LeCam, [5], 1953. It has been rediscovered by Hájek, [4], 1972. From a general point of view it is discussed by LeCam in [6], 1986, sections 7.4 and 7.5. We are referring to the version presented in Strasser, [10], Theorem 83.5, and in Strasser, [11].

In case $p = 1$ (1.4) and Theorem (2.13) can be stated even for a single and fixed loss function W . In case $p > 1$ a similar asymptotic admissibility assertion is not valid for a single loss function. The problem is usually circumvented by Hájek's convolution theorem, restricting the attention to regular sequences of estimators. In the present context this is the approach by Bickel and Klaassen, [2], 1986. Another

possibility which is already mentioned by Hájek, [4], 1972, is to replace the single loss function by a sufficiently large set of loss functions. Then it is possible to apply the one–dimensional admissibility assertion to every component of the estimator sequence. In this way we obtain the assertion of Theorem (2.13).

It should be noted that the assertion of Theorem (2.13) does not exclude that the risks of a non–optimal estimator sequence may underflow the risk bound with positive probability. To exclude such cases one has to impose regularity condition like those used by Bickel and Klaassen, [2], 1986.

Thus, in view of (1.4) we can show that „in probability” the local asymptotic risk bounds of the full mixture model remain valid for incidental nuisance parameters. However, it must be questioned whether the local asymptotic risk bounds of the full mixture model remain valid if fixed incidental nuisance parameters are considered. Clearly, for arbitrary sequences of estimators this cannot be true since the knowledge of the nuisance parameter sequence up to a term of order $n^{-1/2}$ makes it possible to improve the estimator sequence. We have to restrict the class of estimator sequences in such a way that they cannot be biased towards a particular sequence of nuisance parameters. Bickel and Klaassen, [2], apply a kind of regularity condition on the estimator sequence to achieve this goal. Instead of imposing regularity of the estimator sequence we prefer to restrict our attention to the subclass of permutation invariant estimators.

Our second main result deals with permutation invariant estimators in the case of non–random incidental nuisance parameters. We are going to consider triangular arrays (η_{ni}) of nuisance parameters.

(1.6) DEFINITION *Let us call a triangular array (η_{ni}) of nuisance parameters weakly distributed according to Γ if*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(\eta_{ni}) = \int_0^1 f(\eta) \Gamma(d\eta) \quad \text{for all } f \in C(\Lambda).$$

Let \mathcal{T} be a subfamily of weakly Γ –distributed arrays of nuisance parameters. It would be nice to be able to prove an assertion of the following kind for some reasonable subfamily \mathcal{T} :

(1.7) *Consider the risks (1.1) and suppose that the estimator sequence is permutation invariant. If for a triangular array in \mathcal{T} and for some loss function W the efficiency bound of the full mixture model is underflowed, then the efficiency bound is overflowed for some possible other array in \mathcal{T} and possibly other loss function W .*

In fact, we are able to prove such a result under, however, some limitations. First, it is essential for the methods we are going to apply that $\Lambda \subseteq \mathbb{R}$ is an open interval and that the distribution function of Γ is continuous and strictly increasing on Λ . An extension to nuisance parameters of dimension greater than one seems not to be straightforward with our methods.

Having accepted the restriction to the one–dimensional case it is not far to the assumption that $\Lambda = (0, 1)$ and $\Gamma = \lambda$ where λ denotes the Lebesgue measure. In fact, if the distribution function of Γ is sufficiently smooth and strictly increasing on Λ then a simple reparametrization of the family $(P_{\theta, \eta})$ with respect to η gives us the reduction to that case. (We confess not having checked whether our second main result (1.7) remains valid if the distribution function of Γ is continuous, but not smooth and strictly increasing. It seemed worth to us presenting a proof whose essentials are not obscured by being bothered about technicalities.)

In the following we assume that $\Lambda = (0, 1)$ and $\Gamma = \lambda$.

We have to specify the subfamily \mathcal{T} of triangular arrays for which (1.7) is going to be proved. Let F_{ny} be the empirical distribution function of the n –th row of the array $y = (\eta_{ni})$.

(1.8) DEFINITION *A triangular array $y = (\eta_{ni})$ is strongly equidistributed if it satisfies the following conditions:*

- (1) $\eta_{ni} \neq \eta_{nj}$ if $i \neq j$.
- (2) $(\sqrt{n}(F_{ny} - I))$ is relatively compact in $D(0, 1)$.

By \mathcal{T}_s we denote the set of all strongly equidistributed triangular arrays.

In section 3 we present and prove in detail that (1.7) is valid for the subfamily $\mathcal{T} = \mathcal{T}_s$ of strongly equidistributed arrays. This is Theorem (3.4). It could be questioned whether the arrays in \mathcal{T}_s are suited to cover interesting cases related to the nuisance parameter problem. The rest of this introductory section is devoted to a discussion of this question.

The set \mathcal{T}_s carries properties which are typical for triangular arrays of random nuisance parameters. Let us explain this in some detail.

(1.9) REMARK Let $Y = (Y_{ni})$ be a triangular array of random variables which in each row are independent and distributed according to $U(0, 1)$. Then the following facts are well-known:

(1) $\text{Prob} \bigcap_{i=1}^{n-1} \bigcap_{j=i+1}^n \{Y_{ni} \neq Y_{nj}\} = 1$ for every $n \in \mathbb{N}$.

(2) For every $\epsilon > 0$ there is a compact set $K \subseteq D(0, 1)$ such that $\text{Prob}\{\sqrt{n}(F_{nY} - I) \in K\} \geq 1 - \epsilon$ for every $n \in \mathbb{N}$.

Property (2) must not be misunderstood: Suppose that (Y_i) is a sequence of i.i.d random variables distributed according to $U(0, 1)$ and let $Y_{ni} = Y_i$ for all $n \in \mathbb{N}$ and $i = 1, 2, \dots, n$. Then (2) means that for every sample size $n \in \mathbb{N}$ the empirical processes $\sqrt{n}(F_{nY} - I)$ are in K with high probability. But it does not mean that typical sequences of realizations (η_i) are such that the empirical processes $\sqrt{n}(F_{nY} - I)$ are relatively compact. On the contrary, the law of the iterated logarithm for empirical distribution functions (see Shorack and Wellner, [9], p.530) implies, that for almost all realizations (η_i) of (Y_i) the empirical processes $\sqrt{n}(F_{nY} - I)$ leave any compact K for infinitely many $n \in \mathbb{N}$.

Hence, our results cannot be applied to study the asymptotic behaviour of estimator sequences under fixed realizations (η_i) of a sequence (Y_i) . Superefficiency for fixed sequences cannot be excluded by results on arrays in \mathcal{T}_s . This is also made explicit by Pfanzagl's counterexample.

Nevertheless, the random arrays (Y_{ni}) satisfy Property (2). Therefore, our result (1.7) is valid with arbitrary large probability if we are focusing to individual arrays of nuisance parameters. Moreover, it follows that asymptotic superefficiency of Pfanzagl's kind cannot be seen for finite sample sizes: For each finite sample size the overwhelming majority of realizations of the random nuisance parameter behaves regularly as if they give rise to strongly equidistributed arrays in \mathcal{T}_s .

(1.10) REMARK Let us illustrate the contrast between assertions for sequences and for arrays from a more familiar point of view by a parable. Assume that (X_i) is a sequence of i.i.d. random variables with mean zero and finite variance. By the law of the iterated logarithm almost all realizations (x_i) of the sequence (X_i) satisfy

$$\limsup_{n \rightarrow \infty} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \right| = \infty.$$

Nevertheless, for practical purposes we are less pessimistic and apply the central limit theorem which implies

$$\text{Prob}\left\{ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right| \leq a_\epsilon \right\} > 1 - \epsilon \quad \text{for suitable } a_\epsilon \text{ and all } \epsilon > 0.$$

Thus, for practical purposes we dispense with considering fixed realizations (x_i) . We are satisfied with the fact that the array of random variables $(\frac{1}{\sqrt{n}} X_i)$ has bounded row sums with large probability, although for different finite sample sizes different sequences of realizations are responsible for this property.

2 Efficiency bounds for random nuisance parameters

The goal of this section is the statement and the proof of Theorem (2.13) which has been summarized in (1.4). Keep the notation of section 1. We need the following conditions.

(2.1) CONDITION For every $\eta \in \Lambda$ the family $(P_{\theta, \eta} : \theta \in \Theta)$ is continuously L^2 -differentiable with Fisher's information $I(\theta, \eta)$. The family $(I(\theta, \cdot) : \theta \in \Theta)$ is uniformly Γ -integrable.

The gradient with respect to θ of the loglikelihood functions is denoted by $l_1(\cdot, \theta, \eta)$. Let $L^*(\Gamma) = \{k \in L^2(\Gamma) : E_\Gamma(k) = 0\}$ and let $E_{\theta, \Gamma}(\cdot | \mathcal{A})$ be the conditional expectation under $Q_{\theta, \Gamma}$ with respect to \mathcal{A} . Let us recall the LAN–property of mixture distributions. If $k \in L^\infty(\Gamma)$ let $U(k) = \{\tau : |\tau| < \frac{1}{\|k\|_\infty}\}$.

(2.2) THEOREM Assume that the family $(P_{\theta, \eta} : \theta \in \Theta, \eta \in \Lambda)$ satisfies condition (2.1). Then for every $k \in L^*(\Gamma) \cap L^\infty(\Gamma)$ the family $(Q'_{\theta, (1+\tau k)\Gamma} : \theta \in \Theta, \tau \in U(k))$ is continuously L^2 –differentiable and the derivative of the loglikelihood function at θ and $\tau = 0$ is given by $(E_{\theta, \Gamma}(l_1(\cdot, \theta, \cdot) | \mathcal{A}), E_{\theta, \Gamma}(k | \mathcal{A}))$.

Proof: It is sufficient to show that the families $(Q'_{\theta, (1+\tau k)\Gamma} : \theta \in \Theta)$ and $(Q'_{\theta, (1+\tau k)\Gamma} : \tau \in U(k))$ are continuously differentiable with the corresponding derivatives. For the first family $(Q'_{\theta, (1+\tau k)\Gamma} : \theta \in \Theta)$ the assertion follows from Bickel, Klaassen, Ritov and Wellner, [1], section 4.5. For the second family the assertion is valid since the families $(Q'_{\theta, (1+\tau k)\Gamma} : \tau \in U(k))$ and $(R_\tau = (1 + \tau k)\Gamma : \tau \in U(k))$ are mutually Blackwell–sufficient. (Cf. Strasser, [?]). \square

From Theorem (2.2) it follows that the family $(Q'_{\theta, (1+\tau k)\Gamma} : \theta \in \Theta, \tau \in U(k))$ satisfies the LAN–condition.

Our second condition is an identifiability condition.

(2.3) CONDITION For every $k \in L^*(\Gamma)$ the functions

$$E_{\theta, \Gamma}(k | \mathcal{A}), E_{\theta, \Gamma}(l_{1i}(\cdot, \theta, \eta) | \mathcal{A}), i = 1, 2, \dots, p,$$

are linearly independent.

Similar conditions are well–known and used by many authors dealing with the asymptotic theory of mixture models. (See Bickel, Klaassen, Ritov and Wellner, [1], section 4.5.)

The following construction leads to the asymptotic efficiency bounds for full mixture models. If Conditions (2.1) and (2.3) are satisfied then there exist orthogonal projections $l_{1i}^*(\cdot, \theta)$ of the functions $E_{\theta, \Gamma}(l_{1i}(\cdot, \theta, \eta) | \mathcal{A})$ to the orthogonal complement of the linear space $\{E_{\theta, \Gamma}(k | \mathcal{A}) : k \in L^*(\Gamma)\}$. We have $E_{\theta, \Gamma}(l_{1i}^*(\cdot, \theta)) = 0$ and denote

$$A^*(\theta) := E_{\theta, \Gamma}(l_{1i}^*(\cdot, \theta)l_{1i}^*(\cdot, \theta)^T)^{-1}. \quad (2.4)$$

This inverse matrix exists by Assumption (2.3). The matrix $A^*(\theta)$ is the efficient covariance matrix for the estimation of θ in the full mixture model. The efficient influence function is given by

$$K_\theta^* = A^*(\theta)l_1^*(\cdot, \theta).$$

Let \mathcal{W} be the set of all bounded, continuous loss functions which are bowl–shaped, i.e. whose level sets $\{W \leq \alpha\}$ are convex and centrally symmetric. This gives us a position to state the basic and essentially well-known theorem on the asymptotic efficiency bound for full mixture models. Let

$$\beta(W, \theta) := \int W d\mathcal{N}(0, A^*(\theta)), \quad W \in \mathcal{W}. \quad (2.5)$$

Our version differs from usual formulations in that we do not impose any regularity on the estimator sequence. As a compensation we have to require condition (2.7) for a sufficiently large set of loss functions. Recall Remark (1.5).

(2.6) THEOREM Assume that the family $(P_{\theta, \eta} : \theta \in \Theta, \eta \in \Lambda)$ satisfies conditions (2.1) and (2.3) for the probability measure Γ . Let (κ_n) be an estimator sequence such that

$$\limsup_{n \rightarrow \infty} \int W(\sqrt{n}(\kappa_n - \theta) - s) dQ'_{\theta+s/\sqrt{n}, (1+k/\sqrt{n})\Gamma} \leq \beta(W, \theta) \quad (2.7)$$

for all $s \in \mathbb{R}^p$, $k \in L^*(\Gamma) \cap L^\infty(\Gamma)$ and $W \in \mathcal{W}$. Then the following assertions are true:

$$\sqrt{n}(\kappa_n(\omega) - \theta) - \frac{1}{\sqrt{n}} \sum_{i=1}^n K_\theta^*(\omega_i) \rightarrow 0 \quad (Q_{\theta,\Gamma}^n), \quad (2.8)$$

$$\mathcal{L}\left(\sqrt{n}(\kappa_n - \theta) \middle| Q_{\theta+s/\sqrt{n}, (1+k/\sqrt{n})\Gamma}^n\right) \longrightarrow \mathcal{N}(s, A^*(\theta)) \quad \text{weakly, for all } s \in \mathbb{R}^p, \quad (2.9)$$

and

$$\lim_{n \rightarrow \infty} \int W(\sqrt{n}(\kappa_n - \theta) - s) dQ_{\theta+s/\sqrt{n}, (1+k/\sqrt{n})\Gamma}^n = \beta(W, \theta) \quad (2.10)$$

for all $s \in \mathbb{R}^p$, $k \in L^*(\Gamma) \cap L^\infty(\Gamma)$ and $W \in \mathcal{W}$.

Proof: The implication (2.7) \Rightarrow (2.8) is Theorem 2 in Strasser, [11], 1994. There the proof is based on the LAN-property of the mixture distributions which follows from Theorem (2.2). The implication (2.8) \Rightarrow (2.9) is obvious for $s = 0$ and $k = 0$. The general case of (2.9) follows from the LAN-property and LeCam's third lemma. \square

(2.11) REMARK If $p = 1$ then the implication (2.7) \Rightarrow (2.8) is even valid if (2.7) is satisfied for one single loss function $W \in \mathcal{W}$ such that $(W < \sup W)$ is a neighbourhood of zero. (See Remark (1.5).)

The assertion of Theorem (2.6) deals with expected risks and does not concern risks under individual random nuisance parameters. However, the following assertion shows how conclusions for individual random nuisance parameters can be drawn. We apply the obvious fact that

$$Q_{\theta,\Gamma}^n(A) = \int \left(\bigotimes_{i=1}^n P_{\theta,\eta_i} \right) (A) \Gamma^n(d\eta), \quad A \in \mathcal{A}^n. \quad (2.12)$$

(2.13) THEOREM Assume that the family $(P_{\theta,\eta} : \theta \in \Theta, \eta \in \Lambda)$ satisfies conditions (2.1) and (2.3) for the probability measure Γ . Let (κ_n) be an estimator sequence. If

$$\Gamma^n \left\{ \underline{\eta} \in \Lambda^n : \int W(\sqrt{n}(\kappa_n - \theta) - s) d \bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, \eta_i} \geq \beta(W, \theta) + \epsilon \right\} \rightarrow 0 \quad (2.14)$$

for all $s \in \mathbb{R}^p$, $\epsilon > 0$ and $W \in \mathcal{W}$, then assertions (2.7)–(2.10) are valid and it follows that

$$\Gamma^n \left\{ \underline{\eta} \in \Lambda^n : \left| \int W(\sqrt{n}(\kappa_n - \theta) - s) d \bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, \eta_i} - \beta(W, \theta) \right| > \epsilon \right\} \rightarrow 0 \quad (2.15)$$

for all $s \in \mathbb{R}^p$, $\epsilon > 0$ and $W \in \mathcal{W}$.

Proof: In the first part of the proof we show that (2.14) implies (2.7).

Suppose that (2.14) is true. Let $s \in \mathbb{R}^p$, $W \in \mathcal{W}$ and $\epsilon > 0$. Denote

$$M_{n,\epsilon}^+ = \left\{ \underline{\eta} \in \Lambda^n : \int W(\sqrt{n}(\kappa_n - \theta) - s) d \bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, \eta_i} \geq \beta(W, \theta) + \epsilon \right\}.$$

Assumption (2.14) says that $\Gamma^n(M_{n,\epsilon}^+) \rightarrow 0$. If $k \in L^*(\Gamma) \cap L^\infty(\Gamma)$ and $\Gamma_n = (1 + k/\sqrt{n})\Gamma$, then we have $(\Gamma_n^n) \triangleleft (\Gamma^n)$. This implies $\Gamma_n^n(M_{n,\epsilon}^+) \rightarrow 0$. Hence, it follows from

$$\begin{aligned} & \int W(\sqrt{n}(\kappa_n - \theta) - s) dQ_{\theta+s/\sqrt{n}, (1+k/\sqrt{n})\Gamma}^n \\ &= \int \int W(\sqrt{n}(\kappa_n - \theta) - s) d \bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, \eta_i} \Gamma_n^n(d\eta) \\ &\leq \beta(W, \theta) + \epsilon + \Gamma_n^n(M_{n,\epsilon}^+) \|W\|_u \end{aligned}$$

that

$$\limsup_{n \rightarrow \infty} \int W(\sqrt{n}(\kappa_n - \theta) - s) dQ_{\theta+s/\sqrt{n}, (1+k/\sqrt{n})\Gamma}^n \leq \beta(W, \theta) + \epsilon.$$

Since $s \in \mathbb{R}^p$, $W \in \mathcal{W}$, $k \in L^*(\Gamma) \cap L^\infty(\Gamma)$ and $\epsilon > 0$ have been chosen arbitrarily, the assertion (2.7) follows.

From Theorem (2.6) we obtain that (2.8)–(2.10) are true. In the second part of the proof we show that (2.10) and (2.14) together imply (2.15).

Let $s \in \mathbb{R}^p$, $W \in \mathcal{W}$ and $\epsilon > 0$. Denote

$$M_{n,\epsilon}^- = \left\{ \underline{\eta} \in \Lambda^n : \int W(\sqrt{n}(\kappa_n - \theta) - s) d \bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, \eta_i} \leq \beta(W, \theta) - \epsilon \right\}$$

and

$$R_n(\underline{\eta}) = \int W(\sqrt{n}(\kappa_n - \theta) - s) d \bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, \eta_i} - \beta(W, \theta).$$

From (2.10) we obtain that

$$\lim_{n \rightarrow \infty} \int W(\sqrt{n}(\kappa_n - \theta) - s) dQ_{\theta+s/\sqrt{n}, \Gamma}^n = \beta(W, \theta),$$

in other words

$$\lim_{n \rightarrow \infty} \int R_n(\underline{\eta}) \Gamma^n(d\underline{\eta}) = 0.$$

Now, the rest of the proof is almost trivial since the assertion (2.15) follows from Lemma (5.9), putting $f_n = R_n$ and $\mu_n = \Gamma^n$. \square

Let us state the assertion of Theorem (2.13) in a way indicated in (1.4).

(2.16) COROLLARY Assume that the family $(P_{\theta,\eta} : \theta \in \Theta, \eta \in \Lambda)$ satisfies conditions (2.1) and (2.3) for the probability measure Γ . Let (κ_n) be an estimator sequence and (Y_{ni}) a array of random nuisance parameters each row of which is i.i.d. according to Γ . If

$$\limsup_{n \rightarrow \infty} \text{Prob} \left\{ \int W(\sqrt{n}(\kappa_n - \theta) - s) d \bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, Y_{ni}} \leq \beta(W, \theta) - \epsilon \right\} > 0$$

for some $s \in \mathbb{R}^p$, $\epsilon > 0$ and $W \in \mathcal{W}$, then

$$\limsup_{n \rightarrow \infty} \text{Prob} \left\{ \int W(\sqrt{n}(\kappa_n - \theta) - s) d \bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, Y_{ni}} \geq \beta(W, \theta) + \epsilon \right\} > 0$$

for some possible other $s \in \mathbb{R}^p$, $\epsilon > 0$ and $W \in \mathcal{W}$.

3 Efficiency bounds for nonrandom nuisance parameters

Let $(\Omega, \mathcal{A}, \mu)$ be a σ -finite measure space and $(P_{\theta,\eta} : \theta \in \Theta, \eta \in (0, 1))$ a family of μ -continuous probability measures. Assume that $\Theta \subseteq \mathbb{R}^p$ is an open set.

In this section we consider nonrandom arrays (η_{ni}) of nuisance parameters. Recall Definition (1.8) of the family \mathcal{T}_s of strongly equidistributed arrays. For every $n \in \mathbb{N}$ let $\mathcal{S}_n \subseteq \mathcal{A}^n$ be the σ -field of the permutation invariant sets. The basic fact used to prove the main result of this section is provided by the following theorem. This theorem might be of independent interest.

(3.1) THEOREM Assume that the family $(P_{\theta,\eta} : \theta \in \Theta, \eta \in (0, 1))$ is u.c. differentiable. Let $(s_n) \subseteq \mathbb{R}^p$ be bounded. Then all sequences of probability measures $(\bigotimes_{i=1}^n P_{\theta+s_n/\sqrt{n}, \eta_{ni}} | \mathcal{S}_n)$ with $(\eta_{ni}) \in \mathcal{T}_s$ and the sequence $(Q_{\theta+s_n/\sqrt{n}, \lambda}^n | \mathcal{S}_n)$ are mutually contiguous.

Proof: In the first part of the proof we show mutual contiguity of the sequences $(\bigotimes_{i=1}^n P_{\theta+s_n/\sqrt{n}, \eta_{ni}} | \mathcal{S}_n)$. The main idea of the proof is based on a suitable parametrization of the set \mathcal{T}_s . We want to describe triangular arrays $y = (\eta_{ni}) \in \mathcal{T}_s$ by functions in $D(0, 1)$.

Let $y = (\eta_{ni})$ be a triangular array. Let $(\eta_{n:i})$ be the corresponding array of order statistics. The triangular array consisting of $\tau_{ni} := i/(n+1)$ will play a particular role. We define

$$\psi_n(y) := \sum_{i=1}^n \sqrt{n}(\eta_{n:i} - \tau_{ni}) 1_{(\frac{i-1}{n}, \frac{i}{n}]}$$

Thus, for every triangular array y we have a sequence of functions $(\psi_n(y)) \subseteq D(0, 1)$. For every $n \in \mathbb{N}$ the function $\psi_n(y)$ is the quantile function of the n^{th} row of y . The array of order statistics $(\eta_{n:i})$ can be recovered from the sequence of function $(\psi_n(y))$. For this let

$$\psi_{ni} := n \int_{(i-1)/n}^{i/n} \psi(\eta) d\eta \quad \text{if } \psi \in D(0, 1).$$

Then we have $\psi_{ni}(y) = \sqrt{n}(\eta_{n:i} - \tau_{ni})$ or $\eta_{n:i} = \tau_{ni} + \frac{\psi_{ni}(y)}{\sqrt{n}}$. This implies

$$\bigotimes_{i=1}^n P_{\theta, \eta_{ni}} | \mathcal{S}_n = \bigotimes_{i=1}^n P_{\theta, \eta_{n:i}} | \mathcal{S}_n = \bigotimes_{i=1}^n P_{\theta, \tau_{ni} + \psi_{ni}(y)/\sqrt{n}} | \mathcal{S}_n. \quad (3.2)$$

The relation between the sequence $(\psi_n(y))$ and the standardized empirical distribution functions $(\sqrt{n}(F_{ny} - I))$ of the rows of y is well-known (See Shorack and Wellner, [9], p. 86.): If $(\sqrt{n}(F_{ny} - I))$ is relatively compact in $D(0, 1)$ then $(\psi_n(y))$ is relatively compact in $D(0, 1)$, too. Thus, if $y \in \mathcal{T}_s$ then the sequence $(\psi_n(y))$ is relatively compact in $D(0, 1)$ and therefore also relatively compact in $L^2(0, 1)$. From the LAN-property stated in Theorem (4.18) we obtain that the sequences of probability measures $(\bigotimes_{i=1}^n P_{\theta+s_n/\sqrt{n}, \tau_{ni} + \psi_{ni}(y)/\sqrt{n}})$ are mutual contiguous. The same is then obviously true of their restrictions to the symmetric σ -fields \mathcal{S}_n . Thus, it follows from equation (3.2) that the sequences of probability measures $(\bigotimes_{i=1}^n P_{\theta+s_n/\sqrt{n}, \eta_{ni}} | \mathcal{S}_n)$ are mutual contiguous. This settles part one of the proof.

In the second part of the proof we will show that

$$\left(\bigotimes_{i=1}^n P_{\theta, \eta_{ni}} | \mathcal{S}_n \right) \triangleleft \left(Q_{\theta, \lambda}^n | \mathcal{S}_n \right) \quad \text{for all } (\eta_{ni}) \in \mathcal{T}_s.$$

In view of part one it is sufficient to prove the following assertion:

Suppose that $A_n \in \mathcal{A}^n$, $n \in \mathbb{N}$, is a sequence satisfying $Q_{\theta, \lambda}^n(A_n) \rightarrow 0$. Then there is at least one triangular array $y = (\eta_{ni}) \in \mathcal{T}_s$ such that $\bigotimes_{i=1}^n P_{\theta, \eta_{ni}}(A_n) \rightarrow 0$.

Let $\epsilon_n := Q_{\theta, \lambda}^n(A_n)$. By the uniform tightness of the empirical process under the uniform distribution there is a compact set $K \subseteq D(0, 1)$ such that the sets

$$M_n = \{\eta_n \in \mathbb{R}^n : \sqrt{n}(F_{n\eta_n} - I) \in K, \eta_{ni} \neq \eta_{nj} \text{ if } i \neq j\}.$$

satisfy $\lambda^n(M_n) \geq \delta > 0$ for all $n \in \mathbb{N}$. Since we have

$$\int \bigotimes_{i=1}^n P_{\theta, \eta_{ni}}(A_n) \lambda^n(d\eta) = \epsilon_n$$

it follows that

$$\inf \left\{ \bigotimes_{i=1}^n P_{\theta, \eta_{ni}}(A_n) : \eta_n \in M_n \right\} \leq \frac{\epsilon_n}{\delta}.$$

Hence, for every $n \in \mathbb{N}$ there exists $\eta_n \in M_n$ such that

$$\bigotimes_{i=1}^n P_{\theta, \eta_{ni}}(A_n) \leq \frac{\epsilon_n}{\delta} + \frac{1}{n}.$$

In this way we have obtained a triangular array $(\eta_{ni}) \in \mathcal{T}_s$ which satisfies

$$\bigotimes_{i=1}^n P_{\theta, \eta_{ni}}(A_n) \rightarrow 0.$$

This settles the second part of the proof.

In the last part of the proof we are going to show that

$$\left(\bigotimes_{i=1}^n P_{\theta, \eta_{ni}} \middle| \mathcal{S}_n \right) \triangleright \left(Q'_{\theta, \lambda} \middle| \mathcal{S}_n \right) \quad \text{for all } (\eta_{ni}) \in \mathcal{T}_s.$$

Suppose that $A_n \in \mathcal{A}^n$, $n \in \mathbb{N}$, is a sequence and there is at least one triangular array $(\eta_{ni}) \in \mathcal{T}_s$ such that $\bigotimes_{i=1}^n P_{\theta, \eta_{ni}}(A_n) \rightarrow 0$. From the first part it follows that this convergence is even true for all arrays $(\eta_{ni}) \in \mathcal{T}_s$. By the uniform tightness of the empirical process under the uniform distribution there is a compact set $K \subseteq D(0, 1)$ such that the sets M_n satisfy $\lambda^n(M_n) \geq \delta$ for all $n \in \mathbb{N}$ where δ is arbitrarily close to one. Since we have

$$\int \bigotimes_{i=1}^n P_{\theta, \eta_{ni}}(A_n) \lambda^n(d\underline{\eta}) \leq \int_{M_n} \bigotimes_{i=1}^n P_{\theta, \eta_{ni}}(A_n) \lambda^n(d\underline{\eta}) + (1 - \delta)$$

it follows that

$$\limsup_{n \rightarrow \infty} Q'_{\theta, \lambda}(A_n) \leq 1 - \delta.$$

This proves the assertion. \square

The main result of this section is the following theorem. The formal similarity with Theorem (2.6) should be noted. Recall the definition of $A^*(\theta)$ in (2.4) and of $\beta(W, \theta)$ in (2.5). The regularity conditions of Theorem (3.4) are stronger than those of Theorem (2.6) since we have to replace condition (2.1) by the following stronger condition (see (4.1)).

(3.3) **CONDITION** The family $(P_{\theta, \eta} : \theta \in \Theta, \eta \in (0, 1))$ is uniformly continuously L^2 -differentiable.

(3.4) **THEOREM** Assume that the family $(P_{\theta, \eta} : \theta \in \Theta, \eta \in (0, 1))$ satisfies conditions (3.3) and (2.3) for the probability measure $\Gamma = \lambda$. Let (κ_n) be a sequence of permutation invariant estimators such that

$$\limsup_{n \rightarrow \infty} \int W(\sqrt{n}(\kappa_n - \theta) - s) d \bigotimes_{i=1}^n P_{\theta + s/\sqrt{n}, \eta_{ni}} \leq \beta(W, \theta) \quad (3.5)$$

for all $s \in \mathbb{R}^p$, $W \in \mathcal{W}$ and $(\eta_{ni}) \in \mathcal{T}_s$. Then the following assertion are true:

$$\sqrt{n}(\kappa_n(\omega) - \theta) - \frac{1}{\sqrt{n}} \sum_{i=1}^n K_{\theta}^*(\omega_i) \rightarrow 0 \quad \left(\bigotimes_{i=1}^n P_{\theta, \eta_{ni}} \right), \quad (3.6)$$

for all $(\eta_{ni}) \in \mathcal{T}_s$,

$$\mathcal{L} \left(\sqrt{n}(\kappa_n - \theta) \middle| \bigotimes_{i=1}^n P_{\theta + s/\sqrt{n}, \eta_{ni}} \right) \longrightarrow \mathcal{N}(s, A^*(\theta)) \quad \text{weakly} \quad (3.7)$$

for all $s \in \mathbb{R}^p$ and $(\eta_{ni}) \in \mathcal{T}_s$, and

$$\lim_{n \rightarrow \infty} \int W(\sqrt{n}(\kappa_n - \theta) - s) d \bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, \eta_{ni}} = \beta(W, \theta) \quad (3.8)$$

for all $s \in \mathbb{R}^p$, $W \in \mathcal{W}$ and $(\eta_{ni}) \in \mathcal{T}_s$.

Proof: Let (κ_n) be a sequence of permutation invariant estimators which satisfies (3.5) for all $s \in \mathbb{R}^p$, $W \in \mathcal{W}$ and $(\eta_{ni}) \in \mathcal{T}_s$. As a first step we show that the sequence (κ_n) satisfies the assumption (2.14) of Theorem (2.13).

Let $\delta > 0$ and $K \subseteq D(0, 1)$ be a compact set such that the sets

$$M_n = \{\eta_n \in (0, 1)^n : \sqrt{n}(F_{n\eta_n} - I) \in K, \eta_{ni} \neq \eta_{nj} \text{ if } i \neq j\}.$$

satisfy $\lambda^n(M_n) \geq 1 - \delta$ for all $n \in \mathbb{N}$. Condition (3.5) implies

$$\lambda^n \left(M_n \cap \left\{ \int W(\sqrt{n}(\kappa_n - \theta) - s) d \bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, \eta_{ni}} \geq \beta(W, \theta) + \epsilon \right\} \right) \rightarrow 0$$

for every $\epsilon > 0$. Since $\lambda(M_n) > 1 - \delta$ for all $n \in \mathbb{N}$ and since $\delta > 0$ is arbitrarily small the assertion (2.14) follows.

From Theorem (2.13) we obtain that assertion (2.8) is valid for all $s \in \mathbb{R}^p$. Now, we apply Theorem (3.1). Since the sequence (κ_n) is permutation invariant it follows that (2.8) implies (3.6) for all $(\eta_{ni}) \in \mathcal{T}_s$.

In order to prove assertion (3.7) we apply the identity (3.2). We have to show that for $a \in \mathbb{R}^p$ and $K = a^T K_\theta^*$ the assertion

$$\lim_{n \rightarrow \infty} \mathcal{L} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n K(\omega_i) \middle| \bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, \tau_{ni} + \psi_{ni}/\sqrt{n}} \right) = \mathcal{N}(\mu, \sigma^2) \text{ weakly} \quad (3.9)$$

with $\mu = a^T s$, $\sigma^2 = a^T A^*(\theta)a$, and for all relatively compact sequences in $(\psi_n) \subseteq D(0, 1)$. It is sufficient to prove it for convergent sequences in $L^2(0, 1)$. For this we apply Theorem (4.24). This gives (3.9).

Assertion (3.8) is an immediate consequence of (3.7). \square

Let us state the assertion of Theorem (3.4) in a way indicated in (1.7).

(3.10) COROLLARY Assume that the family $(P_{\theta, \eta} : \theta \in \Theta, \eta \in \Lambda)$ satisfies conditions (3.3) and (2.3) for the probability measure $\Gamma = \lambda$. Let (κ_n) be a sequence of permutation invariant estimators. If

$$\liminf_{n \rightarrow \infty} \int W(\sqrt{n}(\kappa_n - \theta) - s) d \bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, \eta_{ni}} < \beta(W, \theta)$$

for some $s \in \mathbb{R}^p$, $(\eta_{ni}) \in \mathcal{T}_s$ and $W \in \mathcal{W}$, then

$$\limsup_{n \rightarrow \infty} \int W(\sqrt{n}(\kappa_n - \theta) - s) d \bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, \eta_{ni}} > \beta(W, \theta)$$

for some possible other $s \in \mathbb{R}^p$, $(\eta_{ni}) \in \mathcal{T}_s$ and $W \in \mathcal{W}$.

Another corollary is concerned with the existence of estimator sequences which satisfy the assumptions of Theorem (3.4).

(3.11) COROLLARY Assume that the family $(P_{\theta, \eta} : \theta \in \Theta, \eta \in \Lambda)$ satisfies conditions (3.3) and (2.3) for the probability measure $\Gamma = \lambda$. Let (κ_n) be an estimator sequence which is efficient for the full mixture model. Then for this sequence the assertions (3.5)–(3.8) are satisfied.

Proof: Since the full mixture model is an i.i.d. model the efficient sequence (κ_n) can be assumed to be permutation invariant. Efficiency for the full mixture model means that condition (2.8) is satisfied. From Theorem (3.1) we obtain (3.6) and thus (3.7) and (3.8). \square

4 LAN for non-i.i.d. observations

In this section we are going to summarize and prove some basic facts on local asymptotic normality with independent but not identically distributed observations. We attempt to give a set of conditions which is general but simple.

Let us begin with a general situation where the parameter set is an open and relatively compact set $A \subseteq \mathbb{R}^k$. The parameter is denoted by x . (Later we will specialize to the case $A = \Theta \times \Lambda$ and $x = (\theta, \eta)$.) Let $(P_x : x \in A)$ be a family of probability measures on a measurable space which is dominated by a σ -finite measure $\mu|_{\mathcal{A}}$.

(4.1) DEFINITION *The family $(P_x : x \in A)$ is uniformly continuously (u.c.) L^2 -differentiable if the mapping*

$$F : A \rightarrow L^2(\Omega, \mathcal{A}, \mu) : x \mapsto 2 \left(\frac{dP_x}{d\mu} \right)^{1/2}$$

is differentiable and if the derivative $DF : A \rightarrow L^2(\Omega, \mathcal{A}, \mu)$ is uniformly continuous on A .

Continuous differentiability is a familiar assumption to verify the LAN-property for products with identical factors (see e.g. Bickel, Klaassen, Ritov and Wellner, [1]). We need in addition uniform continuity of the derivative since we are considering likelihood ratios whose denominators are products with non-identical factors. If continuous differentiability is satisfied then on compact subsets of the parameter space even uniform continuous differentiability is fulfilled.

Assume that the family $(P_x; x \in A)$ is u.c. differentiable. Let us collect some well-known consequences.

The derivatives satisfy $DF(x) = 0$ where $\frac{dP_x}{d\mu} = 0$ μ -a.e. Therefore there are functions $l_x \in L^2(\Omega, \mathcal{A}, P_x)$, such that the derivative $DF(x)$ can be written as

$$DF(x) = l_x \cdot \left(\frac{dP_x}{d\mu} \right)^{1/2}, \quad x \in A.$$

The functions l_x may be viewed as derivatives of the loglikelihood function in μ -measure. (If the densities are positive μ -a.e., then they are in fact the derivatives of the loglikelihood functions.) It is a well-known fact that

$$E_x(l_x) = 0, \quad x \in A. \quad (4.2)$$

If the parameter space is of dimension greater than one let the derivatives $DF(x)$ and l_x be represented by row vectors of random variables. Denote the Fisher information function by

$$I(x) = \int l_x^T l_x dP_x, \quad x \in A.$$

The mapping $x \mapsto I(x)$ is uniformly continuous on A .

(4.3) LEMMA *Suppose that (P_x) is u.c. differentiable. Then for every $\epsilon > 0$ there is a $\delta(\epsilon) > 0$ (independent of $x \in A$) such that*

$$\int \left(F(x+t) - F(x) - \frac{1}{2} l_x \cdot t F(x) \right)^2 d\mu < \epsilon |t|^2, \quad \text{if } |t| < \delta(\epsilon), \quad (4.4)$$

whenever the interval between $x \in A$ and $x+t \in A$ is in A .

Proof: This is a consequence of the mean value theorem. Apply Dieudonné, [3], 8.6.2, and consider that $x \mapsto DF(x)$ can be extended continuously onto \bar{A} . \square

Among others, this implies that the Hellinger distances of $(P_x : x \in A)$ satisfy a Lipschitz condition:

$$d_2(P_x, P_y) \leq C|x-y|, \quad \text{if } x, y \in A. \quad (4.5)$$

It follows that $x \mapsto \frac{dP_x}{d\mu}$ is uniformly continuous in $L^1(\mu)$. Since A is relatively compact the family $(\frac{dP_x}{d\mu} : x \in A)$ is even uniformly integrable in $L^1(\mu)$:

$$\lim_{a \uparrow \infty} \sup_{x \in A} \int_{\frac{dP_x}{d\mu} > a} \frac{dP_x}{d\mu} d\mu = 0. \quad (4.6)$$

For the LAN-property we need uniform integrability of the loglikelihood derivatives in order to satisfy the Lindeberg condition.

(4.7) LEMMA *Suppose that (P_x) is u.c. differentiable. Then*

$$\lim_{a \uparrow \infty} \sup_{x \in A} \int_{|l_x| > a} |l_x|^2 dP_x = 0. \quad (4.8)$$

Proof: The mappings $x \mapsto F(x)$ and $x \mapsto G(x) := l_x F(x)$ can be extended continuously to \bar{A} . If we define

$$\bar{l}_x := \frac{G(x)}{F(x)} \mathbf{1}_{\{F(x) > 0\}}, \quad x \in \bar{A},$$

then $x \mapsto \bar{l}_x, x \in \bar{A}$, is a μ -continuous extension of $x \mapsto l_x, x \in A$. Hence, the mappings

$$x \mapsto \int_{|\bar{l}_x| > a} |\bar{l}_x|^2 dP_x, \quad x \in \bar{A},$$

are continuous and decreasing to zero. Since \bar{A} is compact, Dini's theorem proves (4.8). \square

Now we are in a position to state the basic theorem concerning the LAN-property. Let us denote by $\|\cdot\|_D$ the Dudley norm of measures on metric spaces.

(4.9) THEOREM *Assume that the family $(P_x : x \in A)$ is u.c. differentiable on A . Let $(x_{ni}) \subseteq A$ be an arbitrary triangular array and let $(t_{ni}) \subseteq A$ be a triangular array satisfying*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n |t_{ni}|^2 < \infty, \quad (4.10)$$

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{1}{\sqrt{n}} |t_{ni}| = 0. \quad (4.11)$$

Denote

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n t'_{ni} I(x_{ni}) t_{ni}, \quad n \in \mathbb{N}. \quad (4.12)$$

Then it follows that

$$\frac{d \bigotimes_{i=1}^n P_{x_{ni} + t_{ni}/\sqrt{n}}}{d \bigotimes_{i=1}^n P_{x_{ni}}} = \exp \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n l_{x_{ni}} \cdot t_{ni} - \frac{1}{2} \sigma_n^2 + r_n \right), \quad (4.13)$$

where $r_n \rightarrow 0$ ($\bigotimes_{i=1}^n P_{x_{ni}}$). Moreover,

$$\left\| \mathcal{L} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n l_{x_{ni}} \cdot t_{ni} \middle| \bigotimes_{i=1}^n P_{x_{ni}} \right) - \nu_{0, \sigma_n^2} \right\|_D \rightarrow 0. \quad (4.14)$$

Proof: The proof is straightforward. For instance, one may apply Corollary 74.4 in [10], putting $P_{ni} = P_{x_{ni}}$, $Q_{ni} = P_{x_{ni}+t_{ni}/\sqrt{n}}$ and $h_{ni} = n^{-1/2}l_{x_{ni}} \cdot t_{ni}$. The Lindeberg condition needed for (4.14) is a consequence of (4.8). \square

In the following we specialize to the situation considered in section 3. We need some abbreviations.

Let $A = \Theta \times (0, 1)$ where $\Theta \subseteq \mathbb{R}^p$ is an open set. The Lebesgue measure on $(0, 1)$ is denoted by λ . If $x = (\theta, \eta) \in A$, the derivative of the loglikelihood function is denoted by $l_x : \omega \mapsto l(\omega, \theta, \eta)$ and the partial derivatives with respect to θ and η are denoted by $l_1(\cdot, \theta, \eta)$ and $l_2(\cdot, \theta, \eta)$, respectively. The components of Fisher's information matrix $I(x)$ are denoted by $I_{ij}(\theta, \eta)$, i.e.

$$I_{ij}(\theta, \eta) = \int l_i(\cdot, \theta, \eta) l_j(\cdot, \theta, \eta) dP_{\theta, \eta}.$$

Let $\tau_{ni} = \frac{i}{n+1}$.

It is straightforward to specialize Theorem (4.9) to the products

$$\bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, \tau_{ni}+t_{ni}/\sqrt{n}},$$

where (t_{ni}) is a triangular array of numbers in $(0, 1)$. However, in section 3 we are considering a particular parametrization of the triangular arrays (t_{ni}) . There the parameter set is $D([0, 1])$. The LAN-property is even valid for the larger parameter set $L^2(0, 1)$. Therefore we will state the LAN-property for this set.

Let $\psi \in L^2(0, 1)$ be any square integrable function. For such a function we define a triangular array (ψ_{ni}) of numbers

$$\psi_{ni} := n \int_{(i-1)/n}^{i/n} \psi(\eta) d\eta \quad \text{if } \psi \in L^2(0, 1).$$

The products of probability measures can then be parametrized by pairs $(s, \psi) \in \mathbb{R}^p \times L^2(0, 1)$ according to

$$P_{ns\psi} := \bigotimes_{i=1}^n P_{\theta+s/\sqrt{n}, \tau_{ni}+\psi_{ni}/\sqrt{n}} \quad \text{if } (s, \psi) \in \mathbb{R}^p \times L^2(0, 1). \quad (4.15)$$

The LAN-property requires a sequence of random variables and a nonrandom quadratic function of the parameters. For convenience let us abbreviate these components by

$$X_{ns\psi} : \underline{\omega} \mapsto \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(l_1(\omega_i, \theta, \tau_{ni}) \cdot s + l_2(\omega_i, \theta, \tau_{ni}) \psi_{ni} \right), \quad \underline{\omega} \in \Omega^n \quad (4.16)$$

and

$$I_{s\psi} = \int (l_{1\theta} \cdot s + l_{2\theta} \psi)^2 dQ'_{\theta, \lambda}. \quad (4.17)$$

Now we are in a position to state that kind of LAN-property which is applied in section 3.

(4.18) **COROLLARY** *Assume that the family $(P_{\theta, \eta} : \theta \in \Theta, \eta \in (0, 1))$ is u.c. differentiable. Let $(s_n) \subseteq \mathbb{R}^p$ be bounded and $(\psi_n) \subseteq L^2(0, 1)$ be relatively compact. Then we have*

$$\frac{dP_{ns_n\psi_n}}{dP_{n00}} = \exp \left(X_{ns_n\psi_n} - \frac{1}{2} I_{s_n\psi_n} + r_n \right), \quad (4.19)$$

where $r_n \rightarrow 0$ (P_{n00}) and

$$\|\mathcal{L}(X_{ns_n\psi_n} | P_{n00}) - \mathcal{N}(0, I_{s_n\psi_n})\|_D \rightarrow 0. \quad (4.20)$$

Proof: According to Lemma (5.1) the triangular array $(s_n, t_{ni}(\psi))$ satisfies conditions (4.10) and (4.11). Denoting

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n (s_n, \psi_{ni}) I(\theta, \tau_{ni}) \begin{pmatrix} s_n \\ \psi_{ni} \end{pmatrix}$$

it follows from Theorem (4.9) that

$$\frac{dP_{ns_n\psi_n}}{dP_{n00}} = \exp \left(X_{ns_n\psi_n} - \frac{1}{2} \sigma_n^2 + r_n \right),$$

where $r_n \rightarrow 0$ (P_{n00}) and

$$\|\mathcal{L}(X_{ns_n\psi_n} | P_{n00}) - \mathcal{N}(0, \sigma_n^2)\|_D \rightarrow 0.$$

From Lemma (5.4) it follows that $\sigma_n^2 - I_{s_n\psi_n} \rightarrow 0$. This proves the assertion. \square

Our second goal in this section is concerned with the asymptotic distribution of standardized sums of a random variable under independent, but not identically distributed observations. This is applied to prove assertion (3.7) of Theorem (3.4).

Let K be a \mathcal{A} -measurable function satisfying

$$\int \int K^2 dP_{\theta,\eta} d\eta < \infty, \quad (4.21)$$

$$\int \int K dP_{\theta,\eta} d\eta = 0. \quad (4.22)$$

Denote

$$J_i(\theta, \eta) := \int Kl_i(\cdot, \theta, \eta) dP_{\theta,\eta}, \quad i = 1, 2. \quad (4.23)$$

(4.24) **THEOREM** *Suppose that the family $(P_{\theta,\eta} : \theta \in \Theta, \eta \in (0, 1))$ is u.c. differentiable. Let $\psi_n \rightarrow \psi$ in $L^2(0, 1)$ and $s_n \rightarrow s$ in \mathbb{R}^p . If K satisfies (4.21) and (4.22) then*

$$\mathcal{L} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n K(\omega_i) \middle| \bigotimes_{i=1}^n P_{\theta+s_n/\sqrt{n}, \tau_{ni}+\psi_{ni}/\sqrt{n}} \right) \rightarrow \mathcal{N}(\mu, \sigma^2) \text{ weakly,}$$

where

$$\begin{aligned} \mu &= \int J_1(\theta, \cdot) \cdot s d\lambda + \int J_2(\theta, \cdot) \psi d\lambda, \\ \sigma^2 &= \int \int K^2 dP_{\theta,\eta} d\eta - \int \left(\int K dP_{\theta,\eta} \right)^2 d\eta. \end{aligned}$$

The proof of this theorem is divided into two parts. First, we will prove a special case. This special case is a straightforward application of LeCam's third lemma. We provide a proof for the reader's convenience.

(4.25) **LEMMA** *The assertion of Theorem (4.24) is valid if K is bounded.*

Proof: Keep the notation of (4.15). Let

$$\begin{aligned} U_n(\underline{\omega}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n K(\omega_i), \\ V_n(\underline{\omega}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n l_1(\omega_i, \theta, \tau_{ni}) \cdot s, \\ W_n(\underline{\omega}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n l_2(\omega_i, \theta, \tau_{ni}) \psi_{ni}. \end{aligned}$$

First we are interested in the common limit distribution of (U_n, V_n, W_n) under $(\bigotimes_{i=1}^n P_{\theta, \tau_{ni}}) = (P_{n00})$. The expectations of V_n and W_n are zero by (4.2). The expectations of (U_n) converge to zero by Lemma (5.7), (4.5) and assumption (4.22). Since K is bounded (U_n) satisfies the Lindeberg condition. The Lindeberg conditions for (V_n) and (W_n) are implied by (4.8) and Lemma (5.1).

Thus, we obtain

$$\mathcal{L}\left(U_n, V_n, W_n \mid \bigotimes_{i=1}^n P_{\theta, \tau_{ni}}\right) \rightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{12} & c_{22} & c_{23} \\ c_{13} & c_{23} & c_{33} \end{pmatrix}\right) \text{ weakly,}$$

where by Lemma (5.4)

$$\begin{aligned} c_{11} &= \int \int K^2 dP_{\theta, \eta} d\eta - \int \left(\int K dP_{\theta, \eta} \right)^2 d\eta, \\ c_{22} &= s^2 \int I_{11}(\theta, \eta) d\eta, \\ c_{33} &= \int \psi^2(\eta) I_{22}(\theta, \eta) d\eta, \\ c_{12} &= s \int J_1(\theta, \eta) d\eta, \\ c_{23} &= s \int \psi(\eta) I_{12}(\theta, \eta) d\eta, \\ c_{13} &= \int \psi(\eta) J_2(\theta, \eta) d\eta. \end{aligned}$$

From Corollary (4.18) we obtain

$$\left| \log \frac{dP_{ns_n\psi_n}}{dP_{n00}} - V_n - W_n + \frac{1}{2}c_{22} + c_{23} + \frac{1}{2}c_{33} \right| \rightarrow 0 \quad (P_{n00}).$$

This implies

$$\mathcal{L}\left(U_n, \log \frac{dP_{ns_n\psi_n}}{dP_{n00}} \mid P_{n00}\right) \rightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ -\frac{1}{2}c_{22} - c_{23} - \frac{1}{2}c_{33} \end{pmatrix}, \begin{pmatrix} c_{11} & c_{12} + c_{13} \\ c_{12} + c_{13} & c_{22} + 2c_{23} + c_{33} \end{pmatrix}\right), \text{ weakly.}$$

By a standard argument („LeCam’s 3rd Lemma”) the assertion

$$\mathcal{L}(U_n \mid P_{ns_n\psi_n}) \rightarrow \mathcal{N}(c_{12} + c_{13}, c_{11}), \text{ weakly,}$$

follows. □

Proof: (of Theorem (4.24)) We shall apply Theorem (3.1). For every $n \in \mathbb{N}$ let $\mathcal{S}_n \subseteq \mathcal{A}^n$ be the symmetric sub- σ -field.

Let $\epsilon > 0$ be arbitrary. Since by Theorem (3.1) $(P_{ns_n\psi_n} \mid \mathcal{S}_n) \triangleleft (Q'_{\theta, \lambda} \mid \mathcal{S}_n)$ there is $\delta(\epsilon) > 0$ such that for every sequence of subsets $A_n \in \mathcal{S}_n$

$$\limsup_{n \rightarrow \infty} Q'_{\theta, \lambda}(A_n) \leq \delta(\epsilon) \Rightarrow \limsup_{n \rightarrow \infty} P_{ns_n\psi_n}(A_n) \leq \epsilon. \quad (4.26)$$

Let $f \in \mathcal{C}_b(\mathbb{R}^p)$ be uniformly continuous and choose $\delta(\epsilon)$ such that

$$|x - y| < \delta(\epsilon) \Rightarrow |f(x) - f(y)| < \epsilon, \quad x, y \in \mathbb{R}^p. \quad (4.27)$$

For every $a \geq 0$ we define

$$K_a = K \mathbf{1}_{\{|K| \leq a\}} - \int K \mathbf{1}_{\{|K| \leq a\}} dQ'_{\theta, \lambda}.$$

Each function K_a satisfies the conditions of Lemma (4.25) and therefore we have

$$\lim_{n \rightarrow \infty} \int f \left(\frac{1}{\sqrt{n}} K_a(\omega_i) \right) dP_{n s_n \psi_n} = \int f d\mathcal{N}(\mu_a, \sigma_a^2). \quad (4.28)$$

It is easy to see that

$$\lim_{a \uparrow \infty} \int (K - K_a)^2 dQ'_{\theta, \lambda} = 0.$$

This implies that

$$\lim_{a \uparrow \infty} \mu_a = \mu \quad \text{and} \quad \lim_{a \uparrow \infty} \sigma_a^2 = \sigma^2.$$

Therefore we may choose $a(\epsilon)$ such that

$$\int (K - K_{a(\epsilon)})^2 dQ'_{\theta, \lambda} < \delta^3(\epsilon)$$

and

$$\left| \int f d\mathcal{N}(a, \sigma^2) - \int f d\mathcal{N}(\mu_{a(\epsilon)}, \sigma_{a(\epsilon)}^2) \right| < \epsilon.$$

By Tschebyscheff's inequality we have for every $n \in \mathbb{N}$

$$Q'_{\theta, \lambda} \left\{ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n K(\omega_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^n K_{a(\epsilon)}(\omega_i) \right| > \delta(\epsilon) \right\} \leq \frac{1}{\delta^2(\epsilon)} \int (K - K_{a(\epsilon)})^2 dQ'_{\theta, \lambda} < \delta(\epsilon).$$

By (4.26) this implies

$$\limsup_{n \rightarrow \infty} P_{n s_n \psi_n} \left\{ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n K(\omega_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^n K_{a(\epsilon)}(\omega_i) \right| > \delta(\epsilon) \right\} \leq \epsilon. \quad (4.29)$$

Considering the inequality

$$\begin{aligned} & \left| \int f \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n K(\omega_i) \right) dP_{n s_n \psi_n} - \int f d\mathcal{N}(\mu, \sigma^2) \right| \\ & \leq \left| \int f \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n K(\omega_i) \right) dP_{n s_n \psi_n} - \int f \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n K_{a(\epsilon)}(\omega_i) \right) dP_{n s_n \psi_n} \right| \\ & \quad + \left| \int f \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n K_{a(\epsilon)}(\omega_i) \right) dP_{n s_n \psi_n} - \int f d\mathcal{N}(\mu_{a(\epsilon)}, \sigma_{a(\epsilon)}^2) \right| \\ & \quad + \left| \int f d\mathcal{N}(\mu_{a(\epsilon)}, \sigma_{a(\epsilon)}^2) - \int f d\mathcal{N}(\mu, \sigma^2) \right| \end{aligned}$$

it follows from (4.27), (4.28) and (4.29) that

$$\limsup_{n \rightarrow \infty} \left| \int f \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n K(\omega_i) \right) dP_{n s_n \psi_n} - \int f d\mathcal{N}(\mu, \sigma^2) \right| \leq (2 + \|f\|_u) \epsilon.$$

□

5 Auxiliary Lemmas

If $\psi \in L^2(0, 1)$ define

$$\psi_{ni} = n \int_{(i-1)/n}^{i/n} \psi(\eta) d\eta, \quad i = 1, 2, \dots, n.$$

(5.1) LEMMA Suppose that $(\psi_n) \subseteq L^2(0, 1)$ is uniformly L^2 -integrable. Then

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \psi_{ni}^2 < \infty, \quad (5.2)$$

and

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{1}{\sqrt{n}} |\psi_{ni}| = 0. \quad (5.3)$$

Proof: Since

$$\frac{1}{n} \sum_{i=1}^n \psi_{ni}^2 \leq \int_0^1 \psi_n^2(\eta) d\eta$$

holds, (5.2) follows. Moreover, we have

$$\frac{1}{\sqrt{n}} |\psi_{ni}| \leq \sqrt{n} \int_{(i-1)/n}^{i/n} |\psi_n(\eta)| \left(1_{\{|\psi_n| \leq \epsilon \sqrt{n}\}}(\eta) + 1_{\{|\psi_n| > \epsilon \sqrt{n}\}}(\eta) \right) d\eta \leq \epsilon + \frac{1}{\epsilon} \int_{|\psi_n| > \epsilon \sqrt{n}} \psi_n^2(\eta) d\eta.$$

This proves (5.3). \square

(5.4) LEMMA Let $f \in C_b(0, 1)$ and assume that $(\psi_n) \subseteq L^2(0, 1)$ is uniformly L^2 -integrable. Then

$$\frac{1}{n} \sum_{i=1}^n \psi_{ni} f\left(\frac{i}{n+1}\right) - \int_0^1 \psi_n(\eta) f(\eta) d\eta \longrightarrow 0, \quad (5.5)$$

$$\frac{1}{n} \sum_{i=1}^n \psi_{ni}^2 f\left(\frac{i}{n+1}\right) - \int_0^1 \psi_n^2(\eta) f(\eta) d\eta \longrightarrow 0, \quad (5.6)$$

Proof: Let

$$f_n(\eta) = \sum_{i=1}^n f\left(\frac{i}{n+1}\right) 1_{(\frac{i-1}{n}, \frac{i}{n}]}(\eta).$$

Then we have $f_n \rightarrow f$ (μ) and hence $\psi_n(f_n - f) \rightarrow 0$ (μ) and $\psi_n^2(f_n - f) \rightarrow 0$ (μ). Since both sequences are uniformly integrable the assertions follow. \square

(5.7) LEMMA Assume that $f \in C_b(0, 1)$ satisfies a Lipschitz condition. Then

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n f\left(\frac{i}{n+1}\right) - \int_0^1 f(\eta) d\eta \right) \longrightarrow 0.$$

Proof: Suppose that $|f(s) - f(t)| \leq C|s - t|$. Then we have

$$\left| \frac{1}{n} \sum_{i=1}^n f\left(\frac{i}{n+1}\right) - \int_0^1 f(\eta) d\eta \right| \leq \sum_{i=1}^n \int_{(i-1)/n}^{i/n} \left| f\left(\frac{i}{n+1}\right) - f(\eta) \right| d\eta \leq \frac{C}{n} \sum_{i=1}^n \int_{(i-1)/n}^{i/n} 1 d\eta = O\left(\frac{1}{n}\right).$$

□

The proofs of the next lemmas are straightforward and hence omitted.

(5.8) LEMMA Assume that the family $\{f_{\theta,\eta} : \theta \in \Theta, \eta \in (0, 1)\}$ is uniformly μ -integrable. Then for every probability measure $\pi \in \mathcal{B}(0, 1)$ the family $\{\int f_{\theta,\eta} d\pi : \theta \in \Theta\}$ is uniformly μ -integrable, too.

(5.9) LEMMA Let (f_n) be a uniformly bounded sequence of measurable functions and (μ_n) be a sequence of probability measures. If

$$\lim_{n \rightarrow \infty} \int f_n d\mu_n = 0$$

and

$$\lim_{n \rightarrow \infty} \mu_n\{f_n > \epsilon\} = 0 \text{ for all } \epsilon > 0,$$

then

$$\lim_{n \rightarrow \infty} \mu_n\{f_n < -\epsilon\} = 0 \text{ for all } \epsilon > 0.$$

Acknowledgements

The author thanks J. Pfanzagl for posing the problem and valuable discussions. Moreover, the author owes several substantial improvements to the referee.

References

- [1] P.J. Bickel, C.A.J. Klaassen, Y. Ritov, and J.A. Wellner. *Efficient and adaptive estimation for semi-parametric models*. Johns Hopkins Univ. Press, 1993.
- [2] P.J. Bickel and C.A.L. Klaassen. Empirical Bayes estimation in functional and structural models. *Advances in Applied Mathematics*, pages 55–69, 1986.
- [3] J. Dieudonné. *Foundations of modern analysis*. Academic Press, 1960.
- [4] J. Hájek. Local asymptotic minimax and admissibility in estimation. In *Proc. 6th Berkeley Symp. Math. Stat. Prob.*, pages 175–194, 1972.
- [5] L. LeCam. On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. Calif. Publ. Statist.*, 1:277–330, 1953.
- [6] L. LeCam. *Asymptotic Methods in Statistical Decision Theory*. Springer, 1986.
- [7] J. Pfanzagl. Incidental versus random nuisance parameters. *Annals of Statistics*, 21:1663–1691, 1993.
- [8] J. Pfanzagl and W. Wefelmeyer. *Contributions to a general asymptotic statistical theory*, volume 13 of *Lecture Notes in Statistics*. Springer, 1982.
- [9] G.R. Shorack and J.A. Wellner. *Empirical processes with applications to statistics*. John Wiley, 1986.
- [10] H. Strasser. *Mathematical theory of statistics: Statistical experiments and asymptotic decision theory*, volume 7 of *De Gruyter Studies in Mathematics*. de Gruyter, 1985.
- [11] H. Strasser. Asymptotic admissibility and uniqueness of efficient estimates in semiparametric models. In D. Pollard and G. Yang, editors, *Research papers in probability and statistics: a Festschrift for Lucien LeCam*, chapter 20. 1994.