The Annals of Statistics
1993, Vol. 21, No. 4, 1663–1691

INCIDENTAL VERSUS RANDOM NUISANCE PARAMETERS

BY J. PFANZAGL

University of Cologne

Let $\{P_{\vartheta,\eta}: (\vartheta,\eta) \in \Theta \times \mathrm{H}\}$, with $\Theta \subset \mathbb{R}$ and H arbitrary, be a family of mutually absolutely continuous probability measures on a measurable space $(X, \mathscr{A})$. The problem is to estimate ϑ, based on a sample $(x_1,\ldots,x_n)$ from $\times_1^n P_{\vartheta,\eta_\nu}$. If $(\eta_1,\ldots,\eta_n)$ are independently distributed according to some unknown prior distribution Γ, then the distribution of $n^{1/2}(\vartheta^{(n)} - \vartheta)$ under $P^n_{\vartheta,\Gamma}(P_{\vartheta,\Gamma}$ being the Γ-mixture of $P_{\vartheta,\eta}$, $\eta \in \mathrm{H})$ cannot be more concentrated asymptotically than a certain normal distribution with mean 0, say $N_{(0,\sigma_0^2(\vartheta,\Gamma))}$. Folklore says that such a bound is also valid if $(\eta_1,\ldots,\eta_n)$ are just unknown values of the nuisance parameter: In this case, the distribution cannot be more concentrated asymptotically than $N_{(0,\sigma_0^2(\vartheta, E^{(n)}_{(\eta_1,\ldots,\eta_n)}))}$, where $E^{(n)}_{(\eta_1,\ldots,\eta_n)}$ is the empirical distribution of $(\eta_1,\ldots,\eta_n)$. The purpose of the present paper is to discuss to which extent this conjecture is true. The results are summarized at the end of Sections 1 and 3.

1. Introduction. Let $\{P_{\vartheta,\eta}: (\vartheta,\eta) \in \Theta \times \mathrm{H}\}$, with $\Theta \subset \mathbb{R}$ and H arbitrary, be a family of mutually absolutely continuous probability measures (p-measures) on a measurable space $(X, \mathscr{A})$. Assume that H is endowed with some σ-algebra $\mathscr{B}$, and that $\eta \to P_{\vartheta,\eta}(A)$ is measurable for $\vartheta \in \Theta$, $A \in \mathscr{A}$. Under this condition, the densities $p(\cdot,\vartheta,\eta)$ of $P_{\vartheta,\eta}$ with respect to some σ-finite dominating measure, say μ, can be chosen such that $(x,\eta) \to p(x,\vartheta,\eta)$ is measurable for $\vartheta \in \Theta$. For $\nu \in \mathbb{N}$ let x_ν be a realization from P_{ϑ,η_ν}. The problem is to estimate ϑ, based on a sample $(x_1,\ldots,x_n)$, with the nuisance parameters $(\eta_1,\ldots,\eta_n)$ unknown.

For practical purposes, one is interested in estimators $\vartheta^{(n)}$ for which the distribution of $n^{1/2}(\vartheta^{(n)} - \vartheta)$ under $\times_1^n P_{\vartheta,\eta_\nu}$ is approximable by some simple distribution, usually a normal distribution. This goal has been achieved in many particular cases. It is, however, difficult to evaluate the quality of such estimators, since bounds (even asymptotic ones) for the concentration of estimators are not available for this model.

The situation is different if η_ν, $\nu = 1,\ldots,n$, are i.i.d. realizations from some p-measure Γ on $(\mathrm{H}, \mathscr{B})$. Then the interest is in the distribution of $n^{1/2}(\vartheta^{(n)} - \vartheta)$ under $P^n_{\vartheta,\Gamma}$, where $P_{\vartheta,\Gamma}$ denotes the Γ-mixture of $P_{\vartheta,\eta}$, $\eta \in \mathrm{H}$, defined by

$$P_{\vartheta,\Gamma}(A) := \int P_{\vartheta,\eta}(A)\Gamma(d\eta), \qquad A \in \mathscr{A}. \tag{1.1}$$

If $p(\cdot,\vartheta,\eta)$ is a μ-density of $P_{\vartheta,\eta}$, then $p(\cdot,\vartheta,\Gamma) := \int p(\cdot,\vartheta,\eta)\Gamma(d\eta)$ defines a μ-density of $P_{\vartheta,\Gamma}$.

Received February 1991; revised October 1992.
AMS 1991 *subject classifications*. Primary 62G05; secondary 62G20.
Key words and phrases. Estimation, nuisance parameters.

Now, the sequence $(\eta_1, \dots, \eta_n)$ of n unknown random nuisance parameters has been replaced by a single—if more complex—unknown nuisance parameter, namely the p-measure Γ.

For the purpose of illustration consider the celebrated example of Neyman and Scott (1948), page 3. For $\nu = 1, \dots, n$ let (x_ν, y_ν) be a realization from $N^2_{(\mu_\nu, \sigma^2)}$, with μ_ν and σ^2 unknown. It is straightforward to show that

$$\sigma_n^2((x_\nu, y_\nu)_{\nu=1,\dots,n}) := \frac{1}{2n}\sum_1^n (x_\nu - y_\nu)^2, \qquad n \in \mathbb{N},$$

is a reasonable estimator sequence (e.s.) for σ^2. Its distribution under $\times_1^n N^2_{(\mu_\nu, \sigma^2)}$ does not depend on $\mu_1, \dots, \mu_n$. For $n \to \infty$, $n^{1/2}(\sigma_n^2 - \sigma^2)$ is asymptotically (as.) normal with mean 0 and variance $2\sigma^4$. Despite the intuitive appeal of this estimator, the question naturally occurs whether better estimators exist. If μ_ν, $\nu \in \mathbb{N}$, is a realization from some completely unknown p-measure Γ, then σ_n^2 is as. optimal. [See Pfanzagl and Wefelmeyer (1982), pages 235–236, Example 14.3.21.] But is σ_n^2 also (approximately) optimal if $\mu_1, \dots, \mu_n$ are just unknown constants?

Following the usual terminology, we use the term "structural" to characterize the model with random nuisance parameters, and the term "functional" for the model with varying unknown nuisance parameters. (For short, we write S-model and F-model, respectively.)

If Γ is known to be a member of a certain *parametric* family, then we are back to a parametric model, now with a constant unknown (finite dimensional) nuisance parameter. In this case, a bound for the as. concentration of e.s. is available, as well as methods for the construction of e.s. attaining this bound. If Γ belongs to a given *nonparametric* family [usually the family of "all" p-measures on $(\mathrm{H}, \mathscr{B})$], the general theory based on tangent spaces can be applied to obtain as. bounds. Methods for the construction of e.s. attaining these bounds are available in certain special cases [see Bickel and Ritov (1987), Pfanzagl (1987), van der Vaart (1988), Pfanzagl (1990), and Wong and Severini (1991)].

Several authors study the problem of as. optimality for restricted classes of e.s. As an example we mention estimators depending on the data through a "model-specific" ancillary statistic (the distribution of which depends on ϑ, but not on the nuisance parameters). An advocate of such a procedure is Sprent (1966, 1976). See also Nussbaum (1979), Section 4. Another possibility is to restrict attention to estimators obtainable as solutions to an estimating equation, that is, to find a function $u(x, \vartheta)$ such that $\int u(\cdot, \vartheta)\, dP_{\vartheta,\eta} = 0$ for all $\eta \in \mathrm{H}$, and to determine $\vartheta^{(n)}(x_1, \dots, x_n)$ as a solution to $\sum_1^n u(x_\nu, \vartheta) = 0$. For a certain optimality theory for such estimators, see Godambe (1976), Kumon and Amari (1984) and Amari and Kumon (1988). [See also Pfanzagl (1990), Section 6, for a brief discussion of this approach.] It is, of course, important to know the (as.) best e.s. obtainable by a certain method. This information is but of limited value without the complementary information about the best e.s. obtainable by *any* method. [See, e.g., the remarks by Bunke (1979), pages 521–522.]

So far, the only author who tries to obtain bounds for a reasonably large class of e.s. is Andersen (1970a, b). Under suitable regularity conditions on the family $\{P_{\vartheta,\eta}\colon (\vartheta,\eta) \in \Theta \times \mathrm{H}\}$, he obtains as. bounds for e.s. which are as. normal and as. locally uniformly median unbiased. Under his assumptions, the as. variance cannot be smaller than $\lim_n \hat{\sigma}^2(\vartheta, E_\eta^{(n)})$, where $E_\eta^{(n)}$ is the empirical distribution on $\mathscr{B}$, defined by

$$E_\eta^{(n)}(B) := n^{-1} \sum_1^n 1_B(\eta_\nu), \qquad B \in \mathscr{B}, \tag{1.2}$$

and

$$\begin{aligned} &\hat{\sigma}^2(\vartheta,\Gamma) \\ &:= \Bigg[\int \Bigg(\int l^{(1)}(\cdot,\vartheta,\eta)^2 \, dP_{\vartheta,\eta} \\ &\qquad - \left(\int l^{(1)}(\cdot,\vartheta,\eta) l^{(2)}(\cdot,\vartheta,\eta)\, dP_{\vartheta,\eta}\right)^2 \Big/ \int l^{(2)}(\cdot,\vartheta,\eta)^2 \, dP_{\vartheta,\eta} \Bigg) \Gamma(d\eta)\Bigg]^{-1}. \end{aligned} \tag{1.3}$$

Straightforward examples demonstrate that the bound $\lim_n \hat{\sigma}^2(\vartheta, E_\eta^{(n)})$ is not attainable in general. Hence it seems hard to agree with Andersen's statement [Andersen (1970a), page 85] that "in situations where the lower bound is not attained, [it] provides us with a denominator for an efficiency measure."

An attainable bound can be expected only for sequences of permutation invariant estimators $\vartheta^{(n)}$. In this case, $\times_1^n P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$ is invariant under permutations of $(\eta_1, \dots, \eta_n)$. The failure of Andersen's approach results, among others, from the fact that he makes no use of this invariance.

The absence of as. bounds for the F-model is widely recognized as an important loophole in statistical theory. Moran (1971), pages 251–252, mentions among several "outstanding problems" that "there is no theory of the optimality of estimators in the functional equation case" (i.e., for the F-model). Similarly, de Leeuw and Verhelst (1986), page 193, state that "... it is already nonstandard to *define* consistency and efficiency in this [i.e., the F-] model." A competent discussion of these problems can be found in van der Vaart (1988), subsection 5.4.2, where he also suggests an estimator which might be superefficient. This estimator is obtained by splitting the sample in two parts and estimating the score function for each part separately.

Since bounds for the F-model, attainable under general regularity conditions, are not available, many authors are inclined to take the bounds for regular e.s. in the S-model as bounds for (which class of?) e.s. in the F-model. In fact, the interest in as. bounds for the S-model springs partly from the opinion that these are as. bounds for the F-model. Says Follman (1988), page 561: "It provides a statistically accepted way... making the weak assumption that [the nuisance parameter] be random." Portnoy (1985), who studies the as. behavior of M-estimators in the F-model in a related context, uses in the proof of his Theorem 3.3, page 1414, an auxiliary result, stated as Theorem 4.1, which refers to normally distributed nuisance parameters. [For a proof of the latter theorem, see Portnoy (1987), page 28, Theorem 2.1.] Amari and Kumon (1988), page 1062, confronted with the problem that an e.s. attaining their bound does not exist, "reformulate the problem by regarding the nuisance parameter as a random variable." Similarly, Skovgaard (1989), page 345: "We restrict the problem by considering the sequence of nuisance parameters to be drawn from some (unknown) distribution. Given the exchangeability of the observations in the model, this step should not be alarming to statisticians... ." Several authors express the opinion that exchangeability of the nuisance parameters virtually implies that they may be considered as a "random sample" [see, e.g., Novick, Lewis and Jackson (1973), page 20, or Hambelton and Swaminathan (1985), pages 92 and 142]. This opinion can be traced back to Lindley (1971), page 437, who refers to de Finetti's theorem without giving any details.

Presumably, these authors have the following argument in mind. If a sequence of nuisance parameters is irregular, any permutation of this sequence is of the same nature. If one thinks of $(\eta_\nu)_{\nu \in \mathbb{N}}$ as a realization from a p-measure on $\mathscr{B}^{\mathbb{N}}$, this p-measure ought to be invariant under arbitrary finite permutations, in which case it can be considered (according to an appropriate version of de Finetti's theorem) as a mixture of probability measures $\Gamma^{\mathbb{N}}|\mathscr{B}^{\mathbb{N}}$ (mixed over Γ). Hence one can think of $(\eta_\nu)_{\nu \in \mathbb{N}}$ as a realization of the following two-stage process: Having selected a p-measure $\Gamma|\mathscr{B}$ at random, one obtains $(\eta_\nu)_{\nu \in \mathbb{N}}$ as a realization from $\Gamma^{\mathbb{N}}$. What counts is that $(\eta_\nu)_{\nu \in \mathbb{N}}$ is a realization from $\Gamma^{\mathbb{N}}$, no matter where the p-measure Γ comes from.

In view of this strong implication, it appears that "exchangeability" is not the adequate mathematical counterpart of our "lack of knowledge" about $(\eta_1, \dots, \eta_n)$. More adequate is the requirement that $\vartheta^{(n)}(x_1, \dots, x_n)$ be permutation invariant. Without knowing anything about the origin of the nuisance parameters—how can we exclude the possibility that $(\eta_1, \dots, \eta_n)$ is a realization of some stationary stochastic process, hidden to us? By postulating exchangeability one tries to deduce from our lack of knowledge something about reality [here the origin of $(\eta_\nu)_{\nu \in \mathbb{N}}$].

The only paper which tries to justify the interpretation of bounds for the S-model as bounds for the F-model by theoretical arguments is Bickel and Klaassen (1986). Their Proposition 1.1(ii) says that e.s. which are regular in the F-model and as. efficient in the S-model are also as. efficient in the F-model among all regular e.s. The crucial point is their definition of regularity in the F-model, namely,

$$\bigtimes_1^n P_{\vartheta_n, \eta_\nu} * n^{1/2}\big(\vartheta^{(n)} - \vartheta_n\big) \Rightarrow N_{(0, \sigma^2(\vartheta, \Gamma))},$$

for every sequence $\vartheta_n \to \vartheta$, and every sequence $\eta \in \mathrm{H}^{\mathbb{N}}$ for which $E_\eta^{(n)} \Rightarrow \Gamma$. This requires in particular that (i) $\lim_n \bigtimes_1^n P_{\vartheta, \eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$ exists for every sequence $\eta \in \mathrm{H}^{\mathbb{N}}$ with $E_\eta^{(n)}$, $n \in \mathbb{N}$, weakly convergent, (ii) the limit depends on η through $\lim_n E_\eta^{(n)}$ only, and is (iii) a normal distribution with mean 0. Though many e.s. share these properties, they are hard to defend as a general condition to be imposed upon *all* e.s. How restrictive this condition really is can be seen from the fact that it implies a rather strong version of regularity in the S-model, namely, $P^n_{\vartheta_n, \Gamma_n} * n^{1/2}(\vartheta^{(n)} - \vartheta_n) \Rightarrow N_{(0, \sigma^2(\vartheta, \Gamma))}$ for every sequence $\vartheta_n \to \vartheta$ and every sequence $\Gamma_n \Rightarrow \Gamma$ (not only sequences with Γ_n^n and Γ^n contiguous). [See their Proposition 1.1(i).]

The property of an e.s. relevant in the S-model is the as. behavior of $P^n_{\vartheta, \Gamma} * n^{1/2}(\vartheta^{(n)} - \vartheta)$, $n \in \mathbb{N}$. From numerous examples we are used to the fact that $\times_1^n P_{\vartheta, \eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$, $n \in \mathbb{N}$, converges to a limit distribution also in the case where η_ν, $\nu \in \mathbb{N}$, is a sequence of unknown constants fulfilling certain regularity conditions. Such an as. behavior is, in fact, necessary if we wish to use asymptotic distributions as approximations for finite sample sizes. Though common in many examples, weak convergence of $\times_1^n P_{\vartheta, \eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$, $n \in \mathbb{N}$, is not what occurs in general under regularity conditions. At a surface inspection one would think that

$$\lim_n P^n_{\vartheta, \Gamma} * n^{1/2}(\vartheta^{(n)} - \vartheta) = B_{\vartheta, \Gamma} \quad \text{for some } p\text{-measure } \Gamma|\mathscr{B}$$

entails

$$\lim_n \mathop{\times}_1^n P_{\vartheta, \eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta) = B_{\vartheta, \Gamma},$$

for $\Gamma^{\mathbb{N}}$-a.a. sequences η_ν, $\nu \in \mathbb{N}$. Regrettably, this is not true. If η_ν, $\nu \in \mathbb{N}$, is a realization from $\Gamma^{\mathbb{N}}$, the distribution $\times_1^n P_{\vartheta, \eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$ itself is a random element. Even for highly regular e.s. (say e.s. which are permutation invariant and as. linear), this sequence of random distributions converges to a distribution over the set of p-measures on $(\mathbb{R}, \mathbb{B})$, the real line, endowed with the Borel algebra. This distribution is nondegenerate unless the gradient $K(\cdot, \vartheta, \Gamma)$, occurring in the representation of the e.s. [see (2.6)], fulfills $\int K(\cdot, \vartheta, \Gamma)\, dP_{\vartheta, \eta} = 0$ for Γ-a.a. $\eta \in \mathrm{H}$. This may not be alarming, since this condition is always fulfilled if the family of possible mixing distributions is full —the case which one usually has in mind. Perhaps more critical is the fact that—even in the case $\int K(\cdot, \vartheta, \Gamma)\, dP_{\vartheta, \eta} \equiv 0$—there are sequences $\eta \in \mathrm{H}^{\mathbb{N}}$ with $E_\eta^{(n)}$, $n \in \mathbb{N}$, weakly convergent, for which $\times_1^n P_{\vartheta, \eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$, $n \in \mathbb{N}$, converges to a limit distribution which is more concentrated than the optimal limit distribution in the S-model. This entails in particular that the limit distribution of $\times_1^n P_{\vartheta, \eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$, $n \in \mathbb{N}$, if it exists for some sequence $\eta \in \mathrm{H}^{\mathbb{N}}$, is not uniquely determined by the limit of $E_\eta^{(n)}$, $n \in \mathbb{N}$.

2. Asymptotic bounds for the S-model as bounds for the F-model. In this section we develop the concept of an as. bound for the S-model and discuss its possible role as an as. bound for the F-model.

The specification of an S-model includes the specification of the family of possible mixing distributions Γ, say $\mathscr{G}$. A family $\mathscr{G}$ is called "full" if it is large enough, so that the tangent space of the family $\mathscr{G}$ at Γ, say $T(\Gamma, \mathscr{G})$, equals $\mathscr{L}_*(\Gamma) = \{g \in \mathscr{L}_2(\Gamma, \mathscr{B})\colon \int g\, d\Gamma = 0\}$.

To obtain a meaningful concept of an as. bound, we have to specify the class of e.s. $\vartheta^{(n)}$, $n \in \mathbb{N}$, say $\mathscr{E}$, to which this bound refers. The obvious minimal requirement for such an e.s. is weak convergence of $P^n_{\vartheta,\Gamma} * n^{1/2}(\vartheta^{(n)} - \vartheta)$, $n \in \mathbb{N}$, for every $(\vartheta, \Gamma) \in \Theta \times \mathscr{G}$.

An as. bound for the concentration of e.s. in $\mathscr{E}$ is a family of p-measures $\{B_{\vartheta,\Gamma}|\mathbb{B}\colon (\vartheta, \Gamma) \in \Theta \times \mathscr{G}\}$ with the following property:

$$\lim_n P^n_{\vartheta,\Gamma}\{n^{1/2}(\vartheta^{(n)} - \vartheta) \in (-t, t)\} \le B_{\vartheta,\Gamma}(-t, t) \tag{2.1}$$

$$\text{for all } t > 0 \text{ and all } (\vartheta, \Gamma) \in \Theta \times \mathscr{G}.$$

The bound is *attainable* if an e.s. exists in $\mathscr{E}$ for which equality holds in (2.1), for all $t > 0$ and all $(\vartheta, \Gamma) \in \Theta \times \mathscr{G}$.

Entering (2.1) through their values on symmetric intervals only, the p-measures $B_{\vartheta,\Gamma}|\mathbb{B}$ are uniquely determined only on measurable sets symmetric about 0. Hence we presume throughout the following that $B_{\vartheta,\Gamma}$ is symmetric about 0.

The concept of an as. bound as defined in (2.1) is not meaningful if $\mathscr{E}$ is the family of all e.s. Usually one can find e.s. which are "superefficient" at a given parameter value (ϑ_0, Γ_0) to such an extent that

$$\lim_n P^n_{\vartheta_0,\Gamma_0}\left\{n^{1/2}\left(\vartheta^{(n)} - \vartheta_0\right) \in (-t, t)\right\} = 1 \quad \text{for every } t > 0.$$

Then a family $\{B_{\vartheta,\Gamma}\colon (\vartheta, \Gamma) \in \Theta \times \mathscr{G}\}$ which fulfills (2.1) for all e.s. is necessarily trivial: It fulfills $B_{\vartheta,\Gamma}\{0\} = 1$ for all $(\vartheta, \Gamma) \in \Theta \times \mathscr{G}$ and is therefore not attainable. This motivates restricting the family $\mathscr{E}$ to e.s. which are S-regular at every $(\vartheta, \Gamma) \in \Theta \times \mathscr{G}$.

S-regularity of $\vartheta^{(n)}$, $n \in \mathbb{N}$, at (ϑ, Γ) is defined as follows. $P^n_{\vartheta_n,\Gamma_n} * n^{1/2}(\vartheta^{(n)} - \vartheta_n)$, $n \in \mathbb{N}$, converges weakly to the same limit for every sequence (ϑ_n, Γ_n), $n \in \mathbb{N}$, with ϑ_n converging to ϑ at a rate $n^{-1/2}$, and Γ_n, $n \in \mathbb{N}$, converging to Γ at a rate $n^{-1/2}$ so that Γ_n^n fulfills an LAN-condition with respect to Γ^n, that is, there exists a Γ^n-density of Γ_n^n of the type

$$(\eta_1, \dots, \eta_n) \to \exp\left[n^{-1/2}\sum_1^n k(\eta_\nu) - \frac{1}{2}\int k^2\, d\Gamma + r^{(n)}(\eta_1, \dots, \eta_n)\right],$$

with $k \in T(\Gamma, \mathscr{G})$ and $r^{(n)} \to 0\ (\Gamma^n)$.

This type of "regularity" is generally accepted as a paradigm, mainly because it makes as. theory work, not so much because of its intuitive appeal.

Using the theory of tangent spaces, one can obtain as. bounds for S-regular e.s. [under mild regularity conditions on $\{P_{\vartheta,\eta}\colon (\vartheta, \eta) \in \Theta \times \mathrm{H}\}$]. One obtains that

$$B_{\vartheta,\Gamma} = N_{(0,\sigma_0^2(\vartheta,\Gamma))}, \tag{2.2}$$

with

$$\sigma_0^2(\vartheta, \Gamma) = \left(\int L_0(\cdot, \vartheta, \Gamma)^2 \, dP_{\vartheta,\Gamma}\right)^{-1}, \tag{2.3}$$

where $L_0(\cdot, \vartheta, \Gamma)$ is the orthogonal component of $l^{\cdot}(\cdot, \vartheta, \Gamma) := (\partial/\partial\vartheta)\log p(\cdot, \vartheta, \Gamma)$ with respect to

$$T_0(P_{\vartheta,\Gamma}) := \left\{x \to \frac{\int k(\eta)p(x, \vartheta, \eta)\Gamma(d\eta)}{p(x, \vartheta, \Gamma)} : k \in T(\Gamma, \mathscr{G})\right\}, \tag{2.4}$$

the tangent space of the family $\{P_{\vartheta,\Gamma}: \Gamma \in \mathscr{G}\}$. [See, for instance, Pfanzagl (1990), pages 17–18, 38–42 and 48.]

If $\mathscr{G}$ (the family of possible mixing distributions) is parametric, then these bounds are attainable under the usual regularity conditions. For nonparametric families $\mathscr{G}$, the attainability of the bound (2.2) has been established for certain types of models. [See Pfanzagl (1990) for some recent results in this field.] The paper by Ritov and Bickel (1990) emphasizes that there are severe problems with nonparametric families. It remains to be seen how these problems can be remedied by restriction to nonparametric families of "nice" p-measures. [See also Pfanzagl and Wefelmeyer (1982), page 165, Corollary 9.4.5.]

Assume now that $\{B_{\vartheta,\Gamma}: (\vartheta, \Gamma) \in \Theta \times \mathscr{G}\}$ is an attainable bound for e.s. in $\mathscr{E}$. Let $\hat{\mathrm{H}}$ denote the class of all sequences $\eta \in \mathrm{H}^{\mathbb{N}}$ with $E_\eta^{(n)}$, $n \in \mathbb{N}$, converging weakly to some p-measure, say E_η. Let $\hat{\mathrm{H}}(\mathscr{G})$ denote the class of all sequences $\eta \in \hat{\mathrm{H}}$ with $E_\eta \in \mathscr{G}$.

The intuitive claim that bounds for S-regular e.s. are bounds for e.s. in the F-model can now be written as follows: For every e.s. $\vartheta^{(n)}$, $n \in \mathbb{N}$, in $\mathscr{E}$,

$$\begin{aligned} &\limsup_n \bigtimes_1^n P_{\vartheta,\eta_\nu}\{n^{1/2}(\vartheta^{(n)} - \vartheta) \in (-t, t)\} \\ &\qquad \le B_{\vartheta, E_\eta}(-t, t) \quad \text{for all } t > 0,\ \vartheta \in \Theta \text{ and } \eta \in \hat{\mathrm{H}}(\mathscr{G}). \end{aligned} \tag{2.5}$$

Straightforward examples show that this claim cannot be true for arbitrary S-regular e.s. One can easily find S-regular e.s. which are superefficient for particular sequences $\eta \in \hat{\mathrm{H}}(\mathscr{G})$. The problem becomes more delicate if we restrict our attention to e.s. in $\mathscr{E}$ which are (i) as. efficient in the S-model and (ii) invariant under permutations of $(x_1, \dots, x_n)$. The class of these e.s. will be denoted by $\mathscr{E}_0$. If nothing is known about the sequence of nuisance parameters, using e.s. which are superefficient for a particular sequence, yet inefficient for i.i.d.-sequences, can hardly be recommended. This justifies restriction (i). Similarly, in the absence of any knowledge about the values of the nuisance parameters, there are no arguments for the use of e.s. other than permutation invariant ones (and there are easy counterexamples of superefficient e.s. in $\mathscr{E}$ which are not permutation invariant because they are adjusted to a particular sequence of nuisance parameters).

To find out whether (2.5) is true for all e.s. in $\mathscr{E}_0$, we need some more preparation.

An e.s. is *as. linear* in the S-model with *influence function* K if it admits a representation

$$n^{1/2}\left(\vartheta^{(n)}(\underline{x}) - \vartheta\right) = n^{-1/2}\sum_{1}^{n} K(x_\nu, \vartheta, \Gamma) + r^{(n)}(\underline{x}, \vartheta, \Gamma), \tag{2.6}$$

with

$$\int K(\cdot, \vartheta, \Gamma)\, dP_{\vartheta,\Gamma} = 0, \tag{2.7}$$

$$\int K(\cdot, \vartheta, \Gamma)^2\, dP_{\vartheta,\Gamma} \in (0, \infty) \tag{2.8}$$

and

$$r^{(n)}(\cdot, \vartheta, \Gamma) \to 0 \qquad (P^n_{\vartheta,\Gamma}). \tag{2.9}$$

If an e.s. is as. linear and regular in the S-model, $K(\cdot, \vartheta, \Gamma)$ is necessarily a gradient. If an e.s. is as. efficient in the S-model amongst all S-regular e.s., it is (as a consequence of the convolution theorem) as. linear, and $K(\cdot, \vartheta, \Gamma)$ is the canonical gradient, that is [see (2.3) and (2.4)], relation (2.6) holds with

$$K(\cdot, \vartheta, \Gamma) = L_0(\cdot, \vartheta, \Gamma) \Big/ \int L_0(\cdot, \vartheta, \Gamma)^2\, dP_{\vartheta,\Gamma}. \tag{2.10}$$

[A convenient reference for these results is Pfanzagl (1990), pages 4–5.]

Observe that every e.s. in $\mathscr{E}_0$ is as. linear with influence function given by (2.10). If an e.s. is called "as. efficient in the S-model," it is understood that this e.s. is also regular in the S-model.

REMARK 2.11. Many e.s. are as. linear with a remainder sequence fulfilling a stronger version of (2.9), namely,

$$r^{(n)}(\cdot, \vartheta, \Gamma) \to 0 \qquad P^{\mathbb{N}}_{\vartheta,\Gamma}\text{-a.e.} \tag{2.9$'$}$$

This condition is (by Lemma 5.1) equivalent to

$$r^{(n)}(\cdot, \vartheta, \Gamma) \to 0 \qquad \mathop{\times}_{1}^{\infty} P_{\vartheta,\eta_\nu}\text{-a.e. for } \Gamma^{\mathbb{N}}\text{-a.a. } \eta \in \mathrm{H}^{\mathbb{N}}. \tag{2.9$''$}$$

Condition (2.9) is adequate for the usual as. theory, which takes care of the as. performance of $P^n_{\vartheta,\Gamma} * n^{1/2}(\vartheta^{(n)} - \vartheta)$, $n \in \mathbb{N}$.

If the remainder sequence $r^{(n)}$, $n \in \mathbb{N}$, converges to 0 stochastically under $P^{\mathbb{N}}_{\vartheta,\Gamma}$, but fails to converge to 0 stochastically under $\times_1^\infty P_{\vartheta,\eta_\nu}$, then $\times_1^n P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$, $n \in \mathbb{N}$, will, in general, fail to converge [even if $\int K(\cdot, \vartheta, \Gamma)\, dP_{\vartheta,\eta} \equiv 0$; see Theorem 3.1]. Hence without

$$r^{(n)}(\cdot, \vartheta, \Gamma) \to 0 \qquad \left(\mathop{\times}_{1}^{\infty} P_{\vartheta,\eta_\nu}\right) \text{ for } \Gamma^{\mathbb{N}}\text{-a.a. } \eta \in \mathrm{H}^{\mathbb{N}}, \tag{2.9$'''$}$$

one cannot expect a useful result on the as. performance of

$$\times_1^n P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta),\ n \in \mathbb{N}.$$

The indispensable condition (2.9‴) is implied by (2.9′).

For parametric families, as. linear e.s. with a remainder sequence converging to 0 a.e. (rather than stochastically) can always be obtained by the one-step improvement procedure, if the initial e.s. is as. linear with a remainder sequence converging to 0 a.e. Whether this is also true for nonparametric mixture models in general remains to be seen. It is, however, true for our Examples 1 and 2.

REMARK 2.12. $\mathscr{E}_0$ was defined as the class of all e.s. which are as. efficient in the S-model amongst all e.s. which are S-regular *and* permutation invariant. $\{B_{\vartheta,\Gamma}: (\vartheta,\Gamma) \in \Theta \times \mathscr{G}\}$ was introduced as a bound for all S-regular e.s. If this bound is attainable by an S-regular e.s., then it is also attainable by an S-regular sequence of permutation invariant estimators, since any as. linear e.s. can be replaced by a sequence of permutation invariant estimators with the same as. behavior. [*Hint*: The median of all $\vartheta^{(n)}(x_{i_1},\dots,x_{i_n})$, $(i_1,\dots,i_n)$ a permutation of $(1,\dots,n)$, has this property. Since

$$n^{1/2}\big(\vartheta^{(n)}(x_{i_1},\dots,x_{i_n}) - \vartheta\big) = n^{-1/2}\sum_1^n K(x_\nu,\vartheta,\Gamma) + r^{(n)}(x_{i_1},\dots,x_{i_n},\vartheta,\Gamma),$$

this follows from Lemma L.17 in Pfanzagl (1990), page 94.]

3. The asymptotic performance of $n^{1/2}(\vartheta^{(n)} - \vartheta)$ under random nuisance parameters. Our interest still is in the as. behavior of $\times_1^n P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$, $n \in \mathbb{N}$, for sequences $\eta \in \mathrm{H}^{\mathbb{N}}$ with $E_\eta^{(n)}$, $n \in \mathbb{N}$, weakly convergent. To get an idea of what might happen, we first study the as. behavior for sequences η which are realizations from $\Gamma^{\mathbb{N}}$. In this case, the p-measure $\times_1^n P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)|\mathbb{B}$ itself is a random element.

THEOREM 3.1. *Assume that $\vartheta^{(n)}$, $n \in \mathbb{N}$ is as. linear in the S-model with influence function K and a remainder term fulfilling* (2.9‴). *Then the following holds true for* $\Gamma^{\mathbb{N}}$-*a.a.* $\eta \in \mathrm{H}^{\mathbb{N}}$:

$$\limsup_n \sup_{t\in\mathbb{R}} \Big| \times_1^n P_{\vartheta,\eta_\nu}\{n^{1/2}(\vartheta^{(n)} - \vartheta) \le t\} - N_{(\mu_n(\eta,\vartheta,\Gamma),\,\sigma^2(\vartheta,\Gamma))}(-\infty, t] \Big| = 0, \tag{3.2}$$

with

$$\mu_n(\eta,\vartheta,\Gamma) := n^{-1/2}\sum_1^n \int K(\cdot,\vartheta,\Gamma)\, dP_{\vartheta,\eta_\nu} \tag{3.3}$$

and

$$\sigma^2(\vartheta, \Gamma) := \sigma_1^2(\vartheta, \Gamma) - \sigma_2^2(\vartheta, \Gamma), \tag{3.4}$$

where

$$\sigma_1^2(\vartheta, \Gamma) := \int K(\cdot, \vartheta, \Gamma)^2 \, dP_{\vartheta, \Gamma} \tag{3.4'}$$

and

$$\sigma_2^2(\vartheta, \Gamma) := \int \left(\int K(\cdot, \vartheta, \Gamma) \, dP_{\vartheta, \eta} \right)^2 \Gamma(d\eta). \tag{3.4''}$$

As a particular consequence of (3.2), we have

$$\lim_n P^n_{\vartheta, \Gamma} * n^{1/2}(\vartheta^{(n)} - \vartheta) = N_{(0, \sigma_1^2(\vartheta, \Gamma))}.$$

ADDENDUM. *If*

$$\int K(\cdot, \vartheta, \Gamma) \, dP_{\vartheta, \eta} = 0 \quad \textit{for } \Gamma\textit{-a.a. } \eta \in \mathrm{H}, \tag{3.5}$$

then

$$\lim_n \bigtimes_1^n P_{\vartheta, \eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta) = N_{(0, \sigma_1^2(\vartheta, \Gamma))} \quad \textit{for } \Gamma^{\mathbb{N}}\textit{-a.a. } \eta \in \mathrm{H}^{\mathbb{N}}. \tag{3.6}$$

PROOF. To simplify our notation, we omit ϑ and Γ whenever possible. In particular, we write P_η, $K(x)$, σ^2 and so forth, for $P_{\vartheta, \eta}$, $K(x, \vartheta, \Gamma)$ and $\sigma^2(\vartheta, \Gamma)$. By definition, we have

$$n^{1/2}(\vartheta^{(n)} - \vartheta) = n^{-1/2} \sum_1^n K(x_\nu) + r_n(\underline{x}), \tag{3.7}$$

with $r_n \to 0$ $(\bigtimes_1^\infty P_{\vartheta, \eta_\nu})$ for $\Gamma^{\mathbb{N}}$-a.a. $\eta \in \mathrm{H}^{\mathbb{N}}$. The proof starts from the decomposition

$$n^{-1/2} \sum_1^n K(x_\nu) = \mu_n(\eta) + n^{-1/2} \sum_1^n \left(K(x_\nu) - \int K \, dP_{\eta_\nu} \right),$$

with $\mu_n(\eta) := n^{-1/2} \sum_1^n \int K \, dP_{\eta_\nu}$. For $\Gamma^{\mathbb{N}}$-a.a. $\eta \in \mathrm{H}^{\mathbb{N}}$, $n^{-1/2} \sum_1^n (K(x_\nu) - \int K \, dP_{\eta_\nu})$ is as. normal with mean 0 and variance σ^2. Moreover, μ_n is, under $\Gamma^{\mathbb{N}}$, as. normal with mean 0 and variance σ_2^2. These two components add up to render $n^{-1/2} \sum_1^n K(x_\nu)$ and therefore also $n^{1/2}(\vartheta^{(n)} - \vartheta)$ as. normal under $P_\Gamma^{\mathbb{N}}$ with mean 0 and variance $\sigma_1^2 = \sigma^2 + \sigma_2^2$.

More formally, the proof proceeds as follows.

According to Corollary 5.7, applied with $h(x) = K(x)$, the following relations hold for $\Gamma^{\mathbb{N}}$-a.a. $\eta \in \mathrm{H}^{\mathbb{N}}$:

$$\lim_n \sup_{t \in \mathbb{R}} \left| \bigtimes_1^n P_{\eta_\nu} \left\{ n^{-1/2} \sum_1^n \left(K(x_\nu) - \int K \, dP_{\eta_\nu} \right) \le t \right\} - N_{(0, \sigma^2)}(-\infty, t] \right| = 0.$$

Hence

$$\limsup_n \sup_{t\in\mathbb{R}} \left| \bigtimes_1^n P_{\eta_\nu}\left\{ n^{-1/2}\sum_1^n K(x_\nu) \le t \right\} - N_{(\mu_n(\eta),\sigma^2)}(-\infty, t] \right| = 0. \tag{3.8}$$

Moreover, (2.6) implies

$$\begin{aligned} &\bigtimes_1^n P_{\eta_\nu}\{ n^{1/2}(\vartheta^{(n)}(\underline{x}) - \vartheta) \le t\} - \bigtimes_1^n P_{\eta_\nu}\left\{ n^{-1/2}\sum_1^n K(x_\nu) \le t\right\} \\ &\quad \le \bigtimes_1^n P_{\eta_\nu}\left\{ t < n^{-1/2}\sum_1^n K(x_\nu) \le t + \varepsilon \right\} + \bigtimes_1^n P_{\eta_\nu}\{ r_n(\underline{x}) \le -\varepsilon\}. \end{aligned}$$

Since

$$\begin{aligned} &\limsup_n \sup_{t\in\mathbb{R}} \bigtimes_1^n P_{\eta_\nu}\left\{ t < n^{-1/2}\sum_1^n K(x_\nu) \le t + \varepsilon \right\} \\ &\quad \le \limsup_n \sup_{t\in\mathbb{R}} N_{(\mu_n(\eta),\sigma^2)}(t, t+\varepsilon] \le \frac{1}{\sqrt{2\pi}\,\sigma}\varepsilon, \end{aligned}$$

and, as a consequence of (2.9‴),

$$\lim_n \bigtimes_1^n P_{\eta_\nu}\{ r_n(\underline{x}) \le -\varepsilon\} = 0 \quad \text{for } \Gamma^{\mathbb{N}}\text{-a.a. } \eta \in H^{\mathbb{N}},$$

we obtain

$$\begin{aligned} &\limsup_n \sup_{t\in\mathbb{R}} \Bigg[\bigtimes_1^n P_{\eta_\nu}\{ n^{1/2}(\vartheta^{(n)}(\underline{x}) - \vartheta) \le t\} \\ &\qquad\qquad - \bigtimes_1^n P_{\eta_\nu}\left\{ n^{-1/2}\sum_1^n K(x_\nu) \le t\right\}\Bigg] \\ &\quad \le \frac{1}{\sqrt{2\pi}\,\sigma}\varepsilon, \end{aligned} \tag{3.9}$$

for $\Gamma^{\mathbb{N}}$-a.a. $\eta \in H^{\mathbb{N}}$. Since $\varepsilon > 0$ was arbitrary, the left-hand side of (3.9) is nonpositive. Together with the opposite inequality, this implies

$$\begin{aligned} &\limsup_n \sup_{t\in\mathbb{R}} \Bigg| \bigtimes_1^n P_{\eta_\nu}\{ n^{1/2}(\vartheta^{(n)}(\underline{x}) - \vartheta) \le t\} \\ &\qquad - \bigtimes_1^n P_{\eta_\nu}\left\{ n^{-1/2}\sum_1^n K(x_\nu) \le t\right\}\Bigg| = 0 \quad \text{for } \Gamma^{\mathbb{N}}\text{-a.a. } \eta \in H^{\mathbb{N}}. \end{aligned} \tag{3.10}$$

The assertion now follows from (3.8) and (3.10). □

A realization η from $\Gamma^{\mathbb{N}}$ determines a p-measure $N_{(\mu_n(\eta,\vartheta,\Gamma),\sigma^2(\vartheta,\Gamma))}$. The distribution of $\eta \to \mu_n(\eta, \vartheta, \Gamma)$, under $\Gamma^{\mathbb{N}}$, converges weakly to $N_{(0,\sigma_2^2(\vartheta,\Gamma))}$.

This means that the distribution of $\eta \to N_{(\mu_n(\eta,\vartheta,\Gamma),\sigma^2(\vartheta,\Gamma))}$, under $\Gamma^{\mathbb{N}}$, converges to a distribution over the space $\{N_{(\mu,\sigma^2(\vartheta,\Gamma))}: \mu \in \mathbb{R}\}$, which can conveniently be described by a distribution of the parameter μ over $(\mathbb{R}, \mathbb{B})$, namely, $N_{(0,\sigma_2^2(\vartheta,\Gamma))}$. Hence (3.2) has an obvious implication for the as. behavior of the sequence of distributions of $\eta \to \times_1^n P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$ under $\Gamma^{\mathbb{N}}$. This is a sequence of p-measures over the set $\mathscr{Q}$ of all p-measures on $\mathbb{B}$, converging to a p-measure which is concentrated on a subset of $\mathscr{Q}$, namely, $\{N_{(\mu,\sigma^2(\vartheta,\Gamma))}: \mu \in \mathbb{R}\}$.

REMARK 3.11. The p-measures which constitute the support of the limit of $\Gamma^{\mathbb{N}} * (\eta \to \times_1^n P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta))$, $n \in \mathbb{N}$, that is, $\{N_{(\mu,\sigma^2(\vartheta,\Gamma))}: \mu \in \mathbb{R}\}$, have, in general, a variance smaller than the as. variance bound for S-regular e.s., $\sigma_0^2(\vartheta, \Gamma)$, given in (2.3). It is, perhaps, of some interest to point to the fact that this phenomenon is met with in a more general context.

Let $\mathscr{N} = \{N_{(\mu,\sigma^2)}: (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+\}$. Let Π be a distribution over the parameter space $\mathbb{R} \times \mathbb{R}_+$. If

$$\int N_{(\mu,\sigma^2)} \Pi\big(d(\mu,\sigma^2)\big) = N_{(0,\sigma_0^2)},$$

then the distribution Π is concentrated on $\mathbb{R} \times (0, \sigma_0^2]$, that is, all p-measures $N_{(\mu,\sigma^2)}$ occurring in the mixture have a variance $\sigma^2 \le \sigma_0^2$. If, as in our case, the mixing distribution Π is concentrated on a subset $\mathbb{R} \times \{\sigma_1^2\}$, then $\sigma_1^2 < \sigma_0^2$ if Π is nondegenerate. See Teicher (1960), page 65, Theorem 6.

INTERPRETATION OF THEOREM 3.1. Our intention is to obtain information about the as. performance of $\times_1^n P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$ for "irregular" sequences of nuisance parameters η_ν, $\nu \in \mathbb{N}$. To obtain an intuitive interpretation of Theorem 3.1, we consider the following "Gedankenexperiment." We determine several i.i.d. realizations $(\eta_\nu)_{\nu\in\mathbb{N}}$ from $\Gamma^{\mathbb{N}}$, which we lay in store as "ideal models" of irregular sequences of nuisance parameters. What we expect is that $\times_1^n P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$, $n \in \mathbb{N}$, converges to the same limit distribution for any such sequence. What we find is that $\times_1^n P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$ is—for large n—approximately normal, but the means of the approximating normal distributions [given by (3.3)] are different for different sequences $(\eta_\nu)_{\nu\in\mathbb{N}}$: The means themselves are approximately normally distributed, with mean 0 and variance $\sigma_2^2(\vartheta, \Gamma)$ [given by (3.4″)]. If $\sigma_2^2(\vartheta, \Gamma) > 0$, it therefore makes no sense to speak of the as. behavior of the sequence $\times_1^\infty P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$, $n \in \mathbb{N}$, given the path $(\eta_\nu)_{\nu\in\mathbb{N}}$. There is no "limit as n tends to infinity" which could be used as an approximation to $\times_1^\infty P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$ for finite n.

One could, of course, object that, with this interpretation, one is in the wrong model, that the correct interpretation has to proceed from $P_{\vartheta,\Gamma}^{\mathbb{N}} * n^{1/2}(\vartheta^{(n)} - \vartheta)$. This is certainly true in the S-model. In the F-model, however, we have just irregular sequences of nuisance parameters, and no distribution over the space of these sequences which we could use to "average" over $\times_1^\infty P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$ for different $(\eta_\nu)_{\nu \in \mathbb{N}}$.

This is illustrated by Example 1 (given in Section 4) which describes the as. performance of $\times_1^\infty P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$ for sequences η which are realizations of a stationary autoregressive process with parameter α, say $G_\alpha | \mathscr{B}^{\mathbb{N}}$ (including the i.i.d. case for $\alpha = 0$). The distribution of $n^{1/2}(\vartheta^{(n)} - \vartheta)$ under $\int(\times_1^\infty P_{\vartheta,\eta_\nu}) G_\alpha(d\eta)$ converges to a normal limit distribution with mean 0 and a variance depending on α. Hence this limit distribution is not determined by the stationary distribution of the process which is the same for all $\alpha \in (-1, 1)$, and which agrees with $\lim_n E_\eta^{(n)}$ for G_α-a.a. $\eta \in \mathrm{H}^{\mathbb{N}}$.

The situation is different under condition (3.5). This condition is always true Γ-a.e. if the family of possible mixing distributions is full [i.e., $T(\Gamma, \mathscr{G}) = \mathscr{L}_*(\Gamma)$ for every $\Gamma \in \mathscr{G}$]. In this case, $\times_1^\infty P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$, $n \in \mathbb{N}$, does, in fact, converge to a limit distribution, for $\Gamma^{\mathbb{N}}$-a.a. $\eta \in \mathrm{H}^{\mathbb{N}}$. It is in this particular case that we could possibly expect to have convergence of $\times_1^\infty P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$ not only for $\Gamma^{\mathbb{N}}$-a.a. $\eta \in \mathrm{H}^{\mathbb{N}}$, but for a larger class of sequences $\eta \in \mathrm{H}^{\mathbb{N}}$, say all sequences with $E_\eta^{(n)}$, $n \in \mathbb{N}$, convergent. And only in such a case is it meaningful to speak of an as. bound for the concentration of e.s. in the F-model, and to ask whether it coincides with the bound for regular e.s. in the S-model. In other words, only in such a case is it meaningful to ask whether $N_{(0,\sigma_0^2(\vartheta,\Gamma))}$, the bound for i.i.d. sequences from Γ, extends to arbitrary sequences η with $\lim_n E_\eta^{(n)} = \Gamma$. The answer to this question is "No."

Example 2 (given in Section 4) exhibits for a model with a full family $\mathscr{G}$ an e.s. which is permutation invariant, strongly as. linear [see (2.9‴)] and as. efficient among all S-regular e.s., yet "superefficient" along countably many "irregular" sequences in $\hat{\mathrm{H}}$.

What are the consequences of this negative result? (i) For scholars interested in statistical theory: Find an appropriate concept of "stability" for sequences of nuisance parameters η which absorbs more attributes of an i.i.d. sequence than just weak convergence of $E_\eta^{(n)}$, $n \in \mathbb{N}$, and prove the intuitive claim, specified in (2.5), for such stable sequences. Alternatively: Prove that the "number" of sequences along which e.s. in $\mathscr{E}_0$ are superefficient is negligible (in some sense yet to be defined). (ii) For scholars interested in applications: Use e.s. which are as. efficient among all S-regular e.s., but make sure that these e.s. are as. linear with a remainder term converging stochastically $(\times_1^\infty P_{\vartheta,\eta_\nu})$ to 0 for any sequence $\eta \in \hat{\mathrm{H}}$.

4. Examples. The first example refers to a model with $T(\Gamma, \mathscr{G}) \subsetneqq \mathscr{L}_*(\Gamma)$, for which condition (3.5) is not fulfilled. The natural e.s. $\vartheta^{(n)}$, $n \in \mathbb{N}$ [see (4.6)], has the following properties:

1. For every $n \in \mathbb{N}$, $\vartheta^{(n)}$ is permutation invariant.
2. $\vartheta^{(n)}$, $n \in \mathbb{N}$, is as. efficient in the S-model.
3. If the sequences η are realizations of a certain autoregressive process $G_\alpha | \mathscr{B}^{\mathbb{N}}$, then the distribution of $\eta \to \times_1^\infty P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$ under G_α

converges to a *nondegenerate* distribution which is concentrated on $\{N_{(\mu,1)}: \mu \in \mathbb{R}\}$.

4. Under $\int(\times_1^\infty P_{\vartheta,\eta_\nu})G_\alpha(d\eta)$, the distribution of $n^{1/2}(\vartheta^{(n)} - \vartheta)$ converges to $N_{(0,2/(1+\alpha))}$. Since $\lim_n E_\eta^{(n)} = N_{(0,1)}$ for G_α-a.a. $\eta \in \mathrm{H}^{\mathbb{N}}$, for every $\alpha \in (-1,1)$, the limit distribution is not determined by $\lim_n E_\eta^{(n)}$. Its as. variance is smaller than the as. variance bound in the S-model for $\lim_n E_\eta^{(n)}$ if $\alpha \in (0,1)$.

EXAMPLE 1. Let Γ be an unknown distribution; given a realization η_ν from Γ, a realization x_ν from $N_{(\eta_\nu,1)}$ is observed. The problem is to estimate the expectation of Γ. This model can be rephrased as $P_{\vartheta,\eta} = N_{(\vartheta+\eta,1)}$, with η distributed according to an unknown p-measure Γ with $\int\eta\Gamma(d\eta) = 0$. Let $\mathscr{G}$ denote the family of all p-measures $\Gamma|\mathbb{B}$ with $\int\eta\Gamma(d\eta) = 0$ and $\int\eta^2\Gamma(d\eta) < \infty$. Then the tangent space at Γ in $\mathscr{G}$ is

$$T(\Gamma,\mathscr{G}) = \left\{k \in \mathscr{L}_*(\Gamma): \int \eta k(\eta)\Gamma(d\eta) = 0\right\}. \tag{4.1}$$

At first we shall show that the as. variance bound for S-regular e.s. for ϑ is

$$\sigma_0^2(\Gamma) := 1 + \int\eta^2\Gamma(d\eta). \tag{4.2}$$

We have [see (2.4)]

$$T_0(P_{\vartheta,\Gamma}) = \left\{x \to \frac{\int k(\eta)\varphi(x-(\vartheta+\eta))\Gamma(d\eta)}{\int\varphi(x-(\vartheta+\eta))\Gamma(d\eta)}: k \in T(\Gamma,\mathscr{G})\right\},$$

with φ denoting the density of the standard normal distribution. To obtain the as. variance bound, we have to determine the orthogonal component of $l^\cdot(\cdot,\Gamma)$ [see (4.4)] with respect to $T_0(P_{\vartheta,\Gamma})$. For this purpose we show that the orthogonal complement of $T_0(P_{\vartheta,\Gamma})$ in $\mathscr{L}_*(P_{\vartheta,\Gamma})$ is

$$\mathscr{M}(P_{\vartheta,\Gamma}) = \{x \to c(x-\vartheta): c \in \mathbb{R}\}. \tag{4.3}$$

To simplify our notation, let now $\vartheta = 0$. By Lemma 5.12 a function $f \in \mathscr{L}_*(P_{0,\Gamma})$ is orthogonal to $T_0(P_{0,\Gamma})$ iff

$$\int f(x)\int k(\eta)\varphi(x-\eta)\Gamma(d\eta)\,dx = 0 \quad \text{for all } k \in T(\Gamma,\mathscr{G}),$$

that is,

$$\int f(x)\varphi(x-\eta)\,dx = a + c\eta, \qquad \eta \in \mathbb{R}.$$

Now $f \in \mathscr{L}_*(P_{0,\Gamma})$ implies $\int\int f(x)\varphi(x-\eta)\,dx\,\Gamma(d\eta) = 0$, whence $a = 0$.

Therefore, $f \in \mathscr{L}_*(P_{0,\Gamma})$ is orthogonal to $T_0(P_{0,\Gamma})$ iff there exists $c \in \mathbb{R}$ such that

$$\int f(x)\varphi(x-\eta)\,dx = c\eta, \qquad \eta \in \mathbb{R},$$

equivalently: If

$$\int (f(x) - cx)\varphi(x - \eta)\,dx = 0, \qquad \eta \in \mathbb{R}.$$

By completeness of $\{N_{(\eta,1)}: \eta \in \mathbb{R}\}$, this holds true iff

$$f(x) = cx \qquad \lambda\text{-a.e.}$$

This proves (4.3).

With $l(x, \Gamma) := \log \int \varphi(x - \eta)\Gamma(d\eta)$, we have

$$l^{\cdot}(x - \vartheta, \Gamma) := \frac{\partial}{\partial\vartheta} l(x - \vartheta, \Gamma).$$

Hence

$$l^{\cdot}(x - \vartheta, \Gamma) = \frac{\int (x - (\vartheta + \eta))\varphi(x - (\vartheta + \eta))\Gamma(d\eta)}{\int \varphi(x - (\vartheta + \eta))\Gamma(d\eta)}. \tag{4.4}$$

The projection of $l^{\cdot}(x - \vartheta, \Gamma)$ into $\mathscr{M}(P_{\vartheta,\Gamma})$, say $L_0(x, \vartheta, \Gamma)$, is $\sigma_0^{-2}(\Gamma)(x - \vartheta)$. Since $\sigma_0^{-2}(\Gamma)(x - \vartheta)$ is in $\mathscr{M}(P_{\vartheta,\Gamma})$, it remains to be shown that $l^{\cdot}(x - \vartheta, \Gamma) - \sigma_0^{-2}(\Gamma)(x - \vartheta)$ is orthogonal to $\mathscr{M}(P_{\vartheta,\Gamma})$, equivalently that

$$\int \big(l^{\cdot}(x, \Gamma) - \sigma_0^{-2}(\Gamma)x\big)xP_{0,\Gamma}(dx) = 0. \tag{4.5}$$

This follows easily from (4.4).

From $L_0(x, \vartheta, \Gamma) = \sigma_0^{-2}(\Gamma)(x - \vartheta)$, we obtain the as. variance bound [see (2.3)]

$$\left(\int L_0(x, \vartheta, \Gamma)^2 P_{\vartheta,\Gamma}(dx)\right)^{-1} = \sigma_0^2(\Gamma),$$

a relation which was anticipated in our notation.

The canonical gradient [see (2.10)] is $K(x, \vartheta, \Gamma) = x - \vartheta$. Since $\int K(x, \vartheta, \Gamma)P_{\vartheta,\eta}(dx) = \eta$, condition (3.5) is not fulfilled.

The natural e.s. for ϑ is

$$\vartheta^{(n)}(\underline{x}) := n^{-1}\sum_1^n x_\nu. \tag{4.6}$$

We have for any sequence $\eta \in \mathbb{R}^{\mathbb{N}}$,

$$\bigtimes_1^n P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta) = N_{(\mu_n(\eta),1)}, \quad \text{with } \mu_n(\eta) := n^{-1/2}\sum_1^n \eta_\nu. \tag{4.7}$$

For $\Gamma \in \mathscr{G}$, the distribution of μ_n under $\Gamma^{\mathbb{N}}$ converges weakly to

$$N_{(0, \int \eta^2\Gamma(d\eta))}.$$

Hence the sequence of distributions of $\eta \to \bigtimes_1^n P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$ under

$\Gamma^{\mathbb{N}}$ converges weakly to a distribution over $\{N_{(\mu,1)}: \mu \in \mathbb{R}\}$, which can conveniently be described as μ being distributed according to $N_{(0, \int \eta^2 \Gamma(d\eta))}$.

To establish that $\vartheta^{(n)}$, $n \in \mathbb{N}$, is S-regular and as. efficient amongst these e.s., we have to show that

$$P^n_{\vartheta_n, \Gamma_n} * n^{1/2}\big(\vartheta^{(n)} - \vartheta_n\big) \Rightarrow N_{(0, \sigma_0^2(\Gamma))},$$

for ϑ_n converging to ϑ at a rate $n^{-1/2}$, and Γ_n^n being LAN with respect to Γ^n. Because of (4.7) it suffices to show that $\Gamma_n^n * \mu_n \Rightarrow N_{(0, \int \eta^2 \Gamma(d\eta))}$. Since Γ_n^n admits, by definition, a Γ^n-density

$$(\eta_1, \dots, \eta_n) \to \exp\left[n^{-1/2} \sum_1^n k(\eta_\nu) - \frac{1}{2}\int k^2 \, d\Gamma + r^{(n)}(\eta_1, \dots, \eta_n)\right],$$

with $r^{(n)} \to 0$ (Γ^n) and $k \in T(\Gamma, \mathscr{G})$, it is straightforward to see that the weak limit of $\Gamma_n^n * \mu_n$, $n \in \mathbb{N}$, is the same as the weak limit of $\Gamma^n * \mu_n$, the latter being $N_{(0, \int \eta^2 \Gamma(d\eta))}$. [This e.s. is, of course, not regular in the F-model in the strong sense required by Bickel and Klaassen (1986).]

Now we investigate the as. behavior of the e.s. (4.6) under various sequences of nuisance parameters. At first we consider sequences η_ν, $\nu \in \mathbb{N}$, which are realizations of a stationary autoregressive process with expectation 0. For $\alpha \in (-1, 1)$, the process $G_\alpha | \mathbb{B}^{\mathbb{N}}$ is defined inductively as follows: η_1 is distributed as $N_{(0,1)}$; the conditional distribution of $\eta_{\nu+1}$, given $(\eta_1, \dots, \eta_\nu)$, is $N_{(-\alpha\eta_\nu, 1-\alpha^2)}$. For every $\alpha \in (-1, 1)$, the process G_α has the following features:

(i) The distribution of η_ν under G_α is $N_{(0,1)}$ for every $\nu \in \mathbb{N}$.
(ii) $E_\eta^{(n)} \Rightarrow N_{(0,1)}$ for G_α-a.a. $\eta \in \mathbb{R}^{\mathbb{N}}$.

Assertion (i) follows easily from Wold (1954), pages 54–55 and 112. Assertion (ii) follows from the fact that $G_\alpha * (\eta \to (1_B(\eta_\nu))_{\nu \in \mathbb{N}})$ is an irreducible, aperiodic, finite, homogeneous Markov chain with stationary limit distribution $N_{(0,1)}$. See Langrock and Jahn (1979), pages 178 and 165.

Hence any realization η from G_α might serve as a model of a "stable" sequence of nuisance parameters. It includes an i.i.d.-model for $\alpha = 0$. To see how $\times_1^n P_{\vartheta, \eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$ behaves under such realizations η, we make use of the following fact [see Anderson (1971), page 478, Theorem 8.4.1].

(iii) The distribution of μ_n [see (4.7)] under G_α is normal with mean 0 and a variance converging to $(1-\alpha)/(1+\alpha)$. Hence it converges in the sup-metric to $N_{(0, (1-\alpha)/(1+\alpha))}$.

Therefore the sequence of distributions of

$$\mathop{\times}_1^n P_{\vartheta, \eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta), \qquad n \in \mathbb{N},$$

converges under G_α weakly to a (nondegenerate) distribution over $\{N_{(\mu,1)}: \mu \in \mathbb{R}\}$ which may conveniently be described as μ being distributed as $N_{(0, (1-\alpha)/(1+\alpha))}$.

Observe that this distribution is also nondegenerate for $\alpha = 0$, corresponding to the i.i.d. case.

The distribution of $n^{1/2}(\vartheta^{(n)} - \vartheta)$ under $\int(\times_1^n P_{\vartheta, \eta_\nu})G_\alpha(d\eta)$ converges weakly to $N_{(0, 2/(1+\alpha))}$. Despite the fact that—for every $\alpha \in (-1, 1)$—$\lim_n E_\eta^{(n)} = N_{(0,1)}$ for G_α-a.a. η, the as. distribution of the e.s. under G_α, being $N_{(0, 2/(1+\alpha))}$, depends on α. For $\alpha = 0$ (the i.i.d. case), the as. variance becomes 2, thus coinciding with the as. variance bound $\sigma_0^2(N_{(0,1)}) = 2$ [see (4.2)]. What is of interest here is that the as. variance under G_α is smaller than this as. variance bound if $\alpha \in (0, 1)$.

That the as. variance surpasses the as. variance bound for $\alpha \in (-1, 0)$ must not be interpreted as subefficiency through which we have to pay for superefficiency for $\alpha \in (0, 1)$. This interpretation would be justified if α were an unknown nuisance parameter which varies in $(-1, 1)$ (and is taken into account in the estimation procedure). Our interpretation is different: For any $\alpha \in (-1, 1)$, G_α-a.a. realizations η are reasonable models of "stable" sequences of nuisance parameters. If α happens to be in $(0, 1)$, the e.s. $\vartheta^{(n)}$, $n \in \mathbb{N}$, which is as. efficient in the S-model, is superefficient along these "stable" sequences of nuisance parameters.

There are, of course, also sequences $\eta \in \mathrm{H}$ for which $\times_1^\infty P_{\vartheta, \eta_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta)$, $n \in \mathbb{N}$, converges to a limit distribution. Let $\eta \in \mathbb{R}^\mathbb{N}$ be a sequence with $|\eta_m| \le m^{1/4}$ for $m \in \mathbb{N}$, and $E_\eta^{(n)}$, $n \in \mathbb{N}$, converging weakly to some limit distribution, say $\Gamma|\mathbb{B}$, which is symmetric about 0. [If ζ is a realization from $\Gamma^\mathbb{N}$ fulfilling $E_\zeta^{(n)} \Rightarrow \Gamma$, then $\eta_m = \zeta_m 1_{[-m^{1/4}, m^{1/4}]}(\zeta_m)$ has the required properties.] Let $\xi_{2\nu-1} = \eta_\nu$, $\xi_{2\nu} = -\eta_\nu$. Then $E_\xi^{(n)} \Rightarrow \Gamma$, but $\times_1^n P_{\vartheta, \xi_\nu} * n^{1/2}(\vartheta^{(n)} - \vartheta) \Rightarrow N_{(0,1)}$, whatever Γ. Since $1 < \sigma_0^2(\Gamma)$ for any nondegenerate Γ, the e.s. (4.6) is "superefficient" along this sequence ξ, if evaluated by comparison with the as. variance bound pertaining to the limit of $E_\xi^{(n)}$, $n \in \mathbb{N}$.

The second example refers to a model with $T(\Gamma, \mathscr{G}) = \mathscr{L}_*(\Gamma)$. In this case, condition (3.5) is necessarily fulfilled. Hence, unlike Example 1, "nondegenerate limit distributions" over $\{N_{(\mu, \sigma^2)}: \mu \in \mathbb{R}\}$ will not occur. We establish the existence of an e.s. $\hat{\vartheta}^{(n)}$, $n \in \mathbb{N}$, with the following properties:

(i) For every $n \in \mathbb{N}$, $\hat{\vartheta}^{(n)}$ is permutation invariant.

(ii) $\hat{\vartheta}^{(n)}$, $n \in \mathbb{N}$, is as. efficient in the S-model.

(iii) For every $\Gamma|\mathbb{B}$, $\times_1^n P_{\vartheta, \eta_\nu} * n^{1/2}(\hat{\vartheta}^{(n)} - \vartheta)$, $n \in \mathbb{N}$, converges for $\Gamma^\mathbb{N}$-a.a. $\eta \in \mathbb{R}^\mathbb{N}$ to the optimal limit distribution in the S-model.

(iv) There exist countably many "irregular" sequences $\eta \in \mathbb{R}^\mathbb{N}$ with $E_\eta^{(n)}$, $n \in \mathbb{N}$, weakly convergent, such that $\times_1^n P_{\vartheta, \eta_\nu} * n^{1/2}(\hat{\vartheta}^{(n)} - \vartheta)$, $n \in \mathbb{N}$, converges to a limit distribution which is more concentrated about 0 than the optimal limit distribution in the S-model pertaining to $\lim_n E_\eta^{(n)}$.

EXAMPLE 2. For $\vartheta \in \mathbb{R}$ and $\eta \in \mathrm{H} = \mathbb{R}$, let $P_{\vartheta, \eta} = N_{(\vartheta + \eta, 1)} \times N_{(\eta, 1)}$. Assume that $\mathscr{G}$ is the family of all p-measures on $\mathbb{B}$, so that $T(\Gamma, \mathscr{G}) = \mathscr{L}_*(\Gamma)$ for $\Gamma \in \mathscr{G}$.

At first we shall show that the as. variance bound for S-regular e.s. is 2 for every Γ. This can be seen as follows.

For every fixed $\vartheta \in \mathbb{R}$, the function $S((x,y),\vartheta) := (x-\vartheta)+y$ is sufficient for $\{P_{\vartheta,\eta}: \eta \in \mathbb{R}\}$, and $\{P_{\vartheta,\eta} * S(\cdot,\vartheta): \eta \in \mathbb{R}\}$ is complete [since $P_{\vartheta,\eta} * S(\cdot,\vartheta) = N_{(2\eta,2)}$]. Hence $T_0(P_{\vartheta,\Gamma})$ [see (2.4)] consists of all functions in $\mathscr{L}_*(P_{\vartheta,\Gamma})$ which depend on (x,y) through $S((x,y),\vartheta)$ only.

To apply the results of Pfanzagl (1990), pages 38–40, we write the density of $P_{\vartheta,\eta}$ as

$$q((x,y),\vartheta)p_0(S((x,y),\vartheta),\eta),$$

with

$$q((x,y),\vartheta) = \frac{1}{\sqrt{\pi}}\exp\left[-\frac{1}{4}((x-\vartheta)-y)^2\right],$$

$$p_0(s,\eta) = \frac{1}{2\sqrt{\pi}}\exp\left[-\frac{1}{4}(s-2\eta)^2\right],$$

$p_0(\cdot,\eta)$ being a density of $P_{\vartheta,\eta} * S(\cdot,\vartheta)$.

Since $((x-\vartheta)-y)$ and $S((x,y),\vartheta)$ are stochastically independent under $P_{\vartheta,\eta}$, we have $P_{\vartheta,\cdot}^{S(\cdot,\vartheta)}(q^{\cdot}(\cdot,\vartheta)/q(\cdot,\vartheta)) = 0$.

[$P_{\vartheta,\cdot}^{S(\cdot,\vartheta)}f$ denotes the conditional expectation of f, given $S(\cdot,\vartheta)$, with respect to $P_{\vartheta,\eta}$. We write $P_{\vartheta,\cdot}^{S(\cdot,\vartheta)}$ rather than $P_{\vartheta,\eta}^{S(\cdot,\vartheta)}$ to emphasize its independence of η.]

With a, b, L_0 denoting the orthogonal components with respect to $T_0(P_{\vartheta,\Gamma})$ of $q^{\bullet}/q$, $S^{\bullet}$ and $l^{\bullet}$, respectively, we obtain

$$(4.8)\quad a((x,y),\vartheta) = q^{\bullet}((x,y),\vartheta)/q((x,y),\vartheta) = ((x-\vartheta)-y)/2,$$

$$(4.9)\quad b((x,y),\vartheta) = 0,$$

and therefore

$$(4.10)\qquad L_0((x,y),\vartheta,\Gamma) = ((x-\vartheta)-y)/2.$$

Since $\int L_0(\cdot,\vartheta,\Gamma)^2\,dP_{\vartheta,\Gamma} = \frac{1}{2}$, the as. variance bound for S-regular e.s. equals 2 [see relation (2.3)].

The estimator

$$(4.11)\qquad \vartheta^{(n)}((x_\nu,y_\nu)_{\nu=1,\dots,n}) := n^{-1}\sum_1^n (x_\nu - y_\nu)$$

is permutation invariant. The sequence $\vartheta^{(n)}$, $n \in \mathbb{N}$, is S-regular and attains the as. variance bound. Moreover, $\times_1^n P_{\vartheta,\eta_\nu} * n^{1/2}(\vartheta^{(n)}-\vartheta)$, $n \in \mathbb{N}$, converges to the optimal limit distribution for $\Gamma^{\mathbb{N}}$-a.a. $\eta \in \mathbb{R}^{\mathbb{N}}$ for every $\Gamma|\mathbb{B}$.

In the following we construct an e.s. which shares these properties and is, in addition, superefficient for countably many sequences of nuisance parameters. For this purpose we need a sequence of permutation invariant remainder terms which converges stochastically to 0 for almost all i.i.d. realizations, yet behaves differently for certain sequences $\eta \in \mathbb{R}^{\mathbb{N}}$ with $E_\eta^{(n)}$, $n \in \mathbb{N}$, weakly convergent. The basic idea is the same as in the celebrated example of Hodges [see Lehmann (1983), page 405]. We use a preliminary test to discriminate asymptotically between a distinguished sequence of nuisance parameters and *arbitrary* i.i.d. sequences, and let the estimator depend on the outcome of this test. The delicate nature of this discrimination makes the proof somewhat technical.

As a first step, we define a sequence of tests which distinguishes asymptotically with probability 1 between realizations from a given p-measure, say Q, and p-measures $P \neq Q$.

For $Q|\mathbb{B}$ and $n \in \mathbb{N}$, we define [with $L(n) = \log\log n$]

$$C_n(Q) := \left\{ v \in \mathbb{R}^{\mathbb{N}} \colon \sup_{t \in \mathbb{R}} \left| \sum_1^n \left(1_{(-\infty, t]}(v_\nu) - Q(-\infty, t] \right) \right| < \sqrt{nL(n)} \right\}. \tag{4.12}$$

LEMMA 4.13. *For nonatomic p-measures P, Q on $\mathbb{B}$, the following holds true*:

$$Q^{\mathbb{N}}(\liminf C_n(Q)) = 1, \tag{4.14'}$$

$$P^{\mathbb{N}}(\limsup C_n(Q)) = 0 \quad \textit{for } P \neq Q. \tag{4.14''}$$

PROOF. By Smirnov's Glivenko–Cantelli version of the law of the iterated logarithm [see, e.g., Shorack and Wellner (1986), page 530],

$$\limsup_n \sup_{t \in \mathbb{R}} \frac{\left|\Sigma_1^n \left(1_{(-\infty, t]}(v_\nu) - Q(-\infty, t]\right)\right|}{\sqrt{nL(n)}} = \frac{1}{\sqrt{2}} \quad \text{for } Q^{\mathbb{N}}\text{-}a.a.\ v \in \mathbb{R}^{\mathbb{N}}. \tag{4.15}$$

This implies (4.14′). Since

$$\begin{aligned} &\sup_{t \in \mathbb{R}} n|P(-\infty, t] - Q(-\infty, t]| \\ &\quad \leq \sup_{t \in \mathbb{R}} \left| \sum_1^n \left(1_{(-\infty, t]}(v_\nu) - P(-\infty, t] \right) \right| \\ &\qquad + \sup_{t \in \mathbb{R}} \left| \sum_1^n \left(1_{(-\infty, t]}(v_\nu) - Q(-\infty, t] \right) \right|, \end{aligned}$$

for every $v \in \mathbb{R}^{\mathbb{N}}$, $n \in \mathbb{N}$, the relation $P \neq Q$ implies $C_n(P) \cap C_n(Q) = \emptyset$ for n sufficiently large. Hence (4.14″) follows from (4.14′). □

Throughout the following we write N_ξ for the normal distribution with mean ξ and variance 1, and N_Γ for $\int N_\xi \Gamma(d\xi)$.

Let Γ_0 denote the p-measure with characteristic function $\exp[-|t|^{1/2}]$. Recall that Γ_0 is symmetric about 0 and unimodal with bounded Lebesgue density. Moreover, the distribution of $n^{-2}\Sigma_1^n x_\nu$ under Γ_0^n is Γ_0. [See, e.g., Lukacs (1970), pages 136, 138 and 158.]

To prepare the definition of the estimator sequence, we need the following result.

LEMMA 4.16. *For any sequence* $a_n \in \mathbb{R}$, $n \in \mathbb{N}$, *we have*

$$\lim_{n\to\infty} \left| n^{-1/2} \sum_1^n y_\nu - a_n \right| = \infty \quad \textit{for } N_{\Gamma_0}^{\mathbb{N}}\textit{-a.a. } y \in \mathbb{R}^{\mathbb{N}}.$$

PROOF. By Anderson's theorem

$$\begin{aligned}
&N_{\Gamma_0}^{\mathbb{N}}\left\{ y \in \mathbb{R}^{\mathbb{N}}: \left| n^{-1/2} \sum_1^n y_\nu - a_n \right| \le n^{1/4} \right\}\\
&\quad = N_0^{\mathbb{N}} \times \Gamma_0^{\mathbb{N}}\left\{ (v, \xi) \in \mathbb{R}^n \times \mathbb{R}^n: \left| n^{-1/2} \sum_1^n (v_\nu + \xi_\nu) - a_n \right| \le n^{1/4} \right\}\\
&\quad = N_0 \times \Gamma_0\{ (v, \xi) \in \mathbb{R} \times \mathbb{R}: |v + n^{3/2}\xi - a_n| \le n^{1/4} \}\\
&\quad \le \Gamma_0\{ \xi \in \mathbb{R}: |\xi| \le n^{-5/4} \}.
\end{aligned}$$

Since Γ_0 has a bounded Lebesgue density, $\Sigma_{n=1}^\infty \Gamma_0\{\xi \in \mathbb{R}: |\xi| \le n^{-5/4}\} < \infty$. Hence, by the lemma of Borel and Cantelli,

$$N_{\Gamma_0}^{\mathbb{N}}\left(\limsup \left\{ y \in \mathbb{R}^{\mathbb{N}}: \left| n^{-1/2} \sum_1^n y_\nu - a_n \right| \le n^{1/4} \right\} \right) = 0.$$

This implies the assertion. □

Now we define a countable class of sequences $\xi^l \in \mathbb{R}^{\mathbb{N}}$, $l \in \mathbb{N}$, by the following inductive procedure. Assume that ξ^l, $l = 1, \dots, k-1$, are defined. By Lemma 4.16, applied with $a_n = n^{-1/2}\Sigma_1^n \xi_\nu^l$ for $l = 1, \dots, k-1$, we obtain that for $N_{\Gamma_0}^{\mathbb{N}}$-a.a. $y \in \mathbb{R}^{\mathbb{N}}$,

$$\lim_{n\to\infty} \left| n^{-1/2} \sum_1^n (y_\nu - \xi_\nu^l) \right| = \infty \quad \text{for } l = 1, \dots, k-1.$$

This may be rewritten as follows: For $\Gamma_0^{\mathbb{N}}$-a.a. $\xi \in \mathbb{R}^{\mathbb{N}}$,

$$N_0^{\mathbb{N}}\left\{ v \in \mathbb{R}^{\mathbb{N}}: \lim_{n\to\infty} n^{-1/2} \left| \sum_1^n (v_\nu + \xi_\nu - \xi_\nu^l) \right| = \infty \quad \text{for } l = 1, \dots, k-1 \right\} = 1. \tag{4.17}$$

Let $\xi^k \in \mathbb{R}^{\mathbb{N}}$ be a sequence with the following properties:

$$N_0^{\mathbb{N}}\left\{v \in \mathbb{R}^{\mathbb{N}}: \lim_{n\to\infty} n^{-1/2}\left|\sum_1^n (v_\nu + \xi_\nu^k - \xi_\nu^l)\right| = \infty \quad \text{for } l = 1,\dots,k-1\right\} = 1, \tag{4.18'}$$

$$\bigtimes_1^\infty N_{\xi_\nu^k}\big(\liminf C_n(N_{\Gamma_0})\big) = 1, \tag{4.18''}$$

$$E_{\xi^k}^{(n)} \Rightarrow \Gamma_0. \tag{4.18'''}$$

The existence of such a sequence follows from (4.14′), applied with $Q = N_{\Gamma_0}$, and (4.17). [(4.18‴) follows, in fact, from (4.18″). We require it as an extra condition for the purpose of saving space.] Observe that (4.18′) implies implicitly that for all $l = 1,\dots,k-1$, $\xi_\nu^k \neq \xi_\nu^l$ for infinitely many $\nu \in \mathbb{N}$. Even if we subject ξ^k to additional "irregularity conditions" which hold for realizations from $\Gamma_0^{\mathbb{N}}$ with positive probability, the existence of such a sequence is guaranteed.

For $v \in \mathbb{R}$ let

$$g(v) := v\exp[-av^2], \tag{4.19}$$

where $a > 0$ is an arbitrary constant. In the following we repeatedly use that g is bounded, and $\lim_{|v|\to\infty} g(v) = 0$.

With the help of the sequences $\xi^l \in \mathbb{R}^{\mathbb{N}}$, $l \in \mathbb{N}$, we define our estimator sequence for ϑ as follows:

$$\hat\vartheta^{(n)}((x_\nu, y_\nu)_{\nu=1,\dots,n}) := n^{-1}\sum_1^n (x_\nu - y_\nu) + n^{-1/2} r_n(y_1,\dots,y_n), \tag{4.20}$$

with

$$r_n(y_1,\dots,y_n) := \sum_{l=1}^\infty 2^{-l} g\left(n^{-1/2}\sum_1^n (y_\nu - \xi_\nu^l)\right) 1_{C_n(N_{\Gamma_0})}(y_1,\dots,y_n). \tag{4.21}$$

Observe that $\hat\vartheta^{(n)}$ is invariant under permutations of $(x_1, y_1),\dots,(x_n, y_n)$.

LEMMA 4.22. (i) *For every* $\Gamma|\mathscr{B}$,

$$r_n \to 0 \quad N_\Gamma^{\mathbb{N}}\text{-}a.e. \tag{4.23}$$

(ii) *For every* $k \in \mathbb{N}$,

$$r_n(y_1,\dots,y_n) - 2^{-k} g\left(n^{-1/2}\sum_1^n (y_\nu - \xi_\nu^k)\right) \to 0 \quad \textit{for } \bigtimes_1^\infty N_{\xi_\nu^k}\text{-}a.a.\ y \in \mathbb{R}^{\mathbb{N}}. \tag{4.24}$$

PROOF. (i) If $\Gamma \neq \Gamma_0$, we have $N_\Gamma \neq N_{\Gamma_0}$; hence $N_\Gamma^{\mathbb{N}}(\limsup C_n(N_{\Gamma_0})) = 0$ by (4.14″). This implies $r_n \to 0$ $N_\Gamma^{\mathbb{N}}$-a.e.

If $\Gamma = \Gamma_0$, we obtain from Lemma 4.16 that for $N_{\Gamma_0}^{\mathbb{N}}$-a.a. $y \in \mathbb{R}^{\mathbb{N}}$ the following relation holds for all $l \in \mathbb{N}$:

$$\lim_{n\to\infty} n^{-1/2}\left|\sum_1^n (y_\nu - \xi_\nu^l)\right| = \infty.$$

Together with (4.19) and (4.21), this implies $r_n \to 0$ $N_{\Gamma_0}^{\mathbb{N}}$-a.e.

(ii) By (4.18′), $l \neq k$ implies

$$\times_1^\infty N_{\xi_\nu^k}\left\{y \in \mathbb{R}^{\mathbb{N}}: \lim_{n\to\infty} n^{-1/2}\left|\sum_1^n (y_\nu - \xi_\nu^l)\right| = \infty\right\} = 1. \tag{4.25}$$

Since

$$\begin{aligned} r_n(y_1,\dots,y_n) &- 2^{-k} g\left(n^{-1/2}\sum_1^n (y_\nu - \xi_\nu^k)\right) \\ &= \sum_{l\neq k} 2^{-l} g\left(n^{-1/2}\sum_1^n (y_\nu - \xi_\nu^l)\right) 1_{C_n(N_{\Gamma_0})}(y_1,\dots,y_n) \\ &\quad - 2^{-k} g\left(n^{-1/2}\sum_1^n (y_\nu - \xi_\nu^k)\right)\left(1 - 1_{C_n(N_{\Gamma_0})}(y_1,\dots,y_n)\right), \end{aligned}$$

relation (4.24) follows from (4.25), (4.19) and (4.18″). □

For the following, observe that

$$\begin{aligned} n^{1/2}&\left(\hat{\vartheta}^{(n)}((x_\nu, y_\nu)_{\nu=1,\dots,n}) - \vartheta_n\right) \\ &= n^{-1/2}\sum_1^n \left((x_\nu - \vartheta_n) - y_\nu\right) + r_n(y_1,\dots,y_n). \end{aligned} \tag{4.26}$$

Moreover, for r_n, now considered as a function of (x_ν, y_ν), $\nu = 1,\dots,n$,

$$P^n_{\vartheta_n,\Gamma} * r_n = N^n_\Gamma * r_n, \tag{4.27′}$$

$$\times_1^n P_{\vartheta_n,\eta_\nu} * r_n = \times_1^n N_{\eta_\nu} * r_n. \tag{4.27″}$$

PROPOSITION 4.28. *$\hat{\vartheta}^{(n)}$, $n \in \mathbb{N}$, is as. efficient in the S-model, that is, for every $\vartheta \in \mathbb{R}$, every $\Gamma|\mathscr{B}$, every sequence $\vartheta_n \to \vartheta$ and every sequence $\Gamma_n \Rightarrow \Gamma$ with Γ_n^n, $n \in \mathbb{N}$, and Γ^n, $n \in \mathbb{N}$, contiguous,*

$$P^n_{\vartheta_n,\Gamma_n} * n^{1/2}\left(\hat{\vartheta}^{(n)} - \vartheta_n\right) \Rightarrow N_{(0,2)}. \tag{4.29}$$

Moreover,

$$\times_1^n P_{\vartheta_n,\eta_\nu} * n^{1/2}\left(\hat{\vartheta}^{(n)} \to \vartheta_n\right) \Rightarrow N_{(0,2)} \quad \text{for } \Gamma^{\mathbb{N}}\text{-a.a. } \eta \in \mathbb{R}^{\mathbb{N}}. \tag{4.30}$$

PROOF. As a consequence of (4.23), we have $r_n \to 0$ $(\times_1^\infty N_{\eta_\nu})$ for $\Gamma^{\mathbb{N}}$-a.a. $\eta \in \mathbb{R}^{\mathbb{N}}$, and $r_n \to 0$ $(N_{\Gamma_n}^n)$ for any sequence Γ_n for which Γ_n^n and Γ^n are contiguous (which implies contiguity of $N_{\Gamma_n}^n$ and N_Γ^n).

Because of (4.27′) and (4.27″), this implies that $r_n \to 0$ under $\times_1^n P_{\vartheta_n, \eta_\nu}$ for $\Gamma^{\mathbb{N}}$-a.a. $\eta \in \mathbb{R}^{\mathbb{N}}$, and also under $P_{\vartheta_n, \Gamma_n}^n$ if Γ_n^n and Γ^n are contiguous. Hence the assertion follows from (4.26). □

After having established that $\hat{\vartheta}^{(n)}$, $n \in \mathbb{N}$, behaves nicely for $\Gamma^{\mathbb{N}}$-a.a. $\eta \in \mathbb{R}^{\mathbb{N}}$, we shall show that it is superefficient if the sequence of nuisance parameters is one of the sequences $\xi^k \in \mathbb{R}^{\mathbb{N}}$, $k \in \mathbb{N}$.

PROPOSITION 4.31. (i) *For every $\vartheta \in \mathbb{R}$ and every sequence $\vartheta_n \to \vartheta$,*

$$\times_1^n P_{\vartheta_n, \xi_\nu^k} * n^{1/2}\big(\hat{\vartheta}^{(n)} - \vartheta_n\big) \Rightarrow M_k, \tag{4.32}$$

with

$$M_k = N_{(0,\Sigma)} * \big((u, v) \to u + 2^{-k} v \exp[-av^2]\big), \tag{4.33′}$$

where Σ is the matrix with elements

$$\sigma_{11} = 2, \qquad \sigma_{12} = \sigma_{21} = -1, \qquad \sigma_{22} = 1. \tag{4.33″}$$

(ii) *The limit distribution M_k is symmetric about* 0, *and more concentrated than $N_{(0,2)}$, the limit distribution in the S-model*: *We have*

$$M_k(-t', t'') > N_{(0,2)}(-t', t'') \quad \textit{for all } t', t'' > 0, \tag{4.34}$$

and

$$\lim_n \int n^{1/2}\big(\hat{\vartheta}^{(n)} - \vartheta_n\big)\, d \times_{\nu=1}^n P_{\vartheta_n, \xi_\nu^k} = 0, \tag{4.35}$$

$$\lim_n \int \Big(n^{1/2}\big(\hat{\vartheta}^{(n)} - \vartheta_n\big)\Big)^2 d \times_{\nu=1}^n P_{\vartheta_n, \xi_\nu^k} = \int w^2 M_k(dw), \tag{4.36}$$

which is less than 2, *the as. variance bound in the S-model.*

PROOF. We have

$$\times_1^n P_{\vartheta_n, \xi_\nu^k} * \left(n^{-1/2} \sum_1^n \big((x_\nu - \vartheta_n) - y_\nu\big), n^{-1/2} \sum_1^n (y_\nu - \xi_\nu^k)\right)$$
$$= (N_0 \times N_0)^n * \left(n^{-1/2} \sum_1^n (u_\nu - v_\nu), n^{-1/2} \sum_1^n v_\nu\right) = N_{(0,\Sigma)}.$$

Hence

$$\times_1^n P_{\vartheta_n, \xi_\nu^k} * \left(n^{-1/2} \sum_1^n \big((x_\nu - \vartheta_n) - y_\nu\big) + 2^{-k} g\left(n^{-1/2} \sum_1^n (y_\nu - \xi_\nu^k)\right)\right) \Rightarrow M_k,$$

with M_k defined by (4.33′).

Now (4.32) follows from (4.26), (4.24) and (4.27″).

(ii) Symmetry of M_k about 0 follows from the central symmetry of $N_{(0,\Sigma)}$.

For every $v \in \mathbb{R}$, $v \neq 0$, $r \to N_{(0,1)}(-t + rv, t + rv)$ is decreasing on $[0,\infty)$ (by Anderson's theorem). Since $0 < 1 - 2^{-k}\exp[-av^2] < 1$, this implies

$$\begin{aligned} &N_{(0,1)}(-t+v, t+v)\\ (4.37)\qquad &< N_{(0,1)}(-t + v(1-2^{-k}\exp[-av^2]), t + v(1-2^{-k}\exp[-av^2]))\\ &\qquad\qquad \text{for every } v \in \mathbb{R},\ v \neq 0. \end{aligned}$$

Since

$$N_{(0,\Sigma)} = N^2_{(0,1)} * ((u,v) \to (u-v, v)),$$

we obtain

$$\begin{aligned} M_k &= N_{(0,\Sigma)} * ((u,v) \to u + 2^{-k} v \exp[-av^2])\\ &= N^2_{(0,1)} * ((u,v) \to u - v(1 - 2^{-k}\exp[-av^2])). \end{aligned}$$

Hence

$$\begin{aligned} (4.38)\qquad M_k(-t,t) = \int N_{(0,1)}(&-t + v(1-2^{-k}\exp[-av^2]),\\ &t + v(1-2^{-k}\exp[-av^2]))N_{(0,1)}(dv). \end{aligned}$$

Since

$$(4.39)\qquad N_{(0,2)}(-t,t) = \int N_{(0,1)}(-t+v, t+v)N_{(0,1)}(dv),$$

the relation $N_{(0,2)}(-t,t) < M_k(-t,t)$ for $t > 0$ follows from (4.37)–(4.39). Since $N_{(0,2)}$ and M_k are symmetric about 0, this implies $N_{(0,2)}(-t',t'') < M_k(-t',t'')$ for arbitrary $t', t'' > 0$.

Because of (4.26)

$$\begin{aligned} &\left(n^{1/2}\left(\hat{\vartheta}^{(n)}((x_\nu, y_\nu)_{\nu=1,\dots,n}) - \vartheta_n\right)\right)^2\\ &\qquad \le 2\left(n^{-1/2}\sum_1^n((x_\nu - \vartheta_n) - y_\nu)\right)^2 + 2r_n(y_1,\dots,y_n)^2. \end{aligned}$$

Since $\times_1^n P_{\vartheta_n,\xi_\nu^k} * (n^{-1/2}\Sigma_1^n((x_\nu - \vartheta_\nu) - y_\nu)) = N_{(0,2)}$, the sequence of functions $(n^{-1/2}\Sigma_1^n((x_\nu - \vartheta_n) - y_\nu))^2$ is uniformly integrable with respect to $\times_1^n P_{\vartheta_n,\xi_\nu^k}$. The same holds true for r_n^2 (which is uniformly bounded). This implies uniform $\times_1^n P_{\vartheta_n,\xi_\nu^k}$-integrability of $(n^{1/2}(\hat{\vartheta}^{(n)} - \vartheta_n))^2$ and hence (4.35) and (4.36).

$\int w^2 M_k(dw) < 2$ follows from (4.34), applied with $t' = t''$. □

5. Auxiliary results.

LEMMA 5.1. *For any set* $A \in \mathscr{A}^{\mathbb{N}}$, $P_\Gamma^{\mathbb{N}}(A) = 1$ $[= 0]$ *is equivalent to* $\times_1^\infty P_{\eta_\nu}(A) = 1$ $[= 0]$ *for* $\Gamma^{\mathbb{N}}$*-a.a.* $\eta \in \mathrm{H}^{\mathbb{N}}$.

In particular $r^{(n)} \to 0$ $P_\Gamma^{\mathbb{N}}$-*a.e. is equivalent to* $r^{(n)} \to 0$ $\times_1^\infty P_{\eta_\nu}$-*a.e. for* $\Gamma^{\mathbb{N}}$-*a.a.* $\eta \in \mathrm{H}^{\mathbb{N}}$.

LEMMA 5.2. *Let* $(X, \mathscr{A})$ *and* $(\mathrm{H}, \mathscr{B})$ *be measurable spaces. Let* $P_\eta|\mathscr{A}$, $\eta \in \mathrm{H}$, *be a family of p-measures such that* $\eta \to P_\eta(A)$ *is* $\mathscr{B}$-*measurable for every* $A \in \mathscr{A}$.

Let $h\colon X \times \mathrm{H} \to \mathbb{R}$ *be an* $\mathscr{A} \times \mathscr{B}$-*measurable function fulfilling*

$$\int h(x,\eta) P_\eta(dx) = 0 \quad \textit{for } \Gamma\textit{-a.a. } \eta \in \mathrm{H} \tag{5.3}$$

and

$$\sigma^2 := \int\int h(x,\eta)^2 P_\eta(dx)\Gamma(d\eta) \in (0,\infty). \tag{5.4}$$

Then

$$\mathop{\times}_1^\infty P_{\eta_\nu} * \left(\underline{x} \to n^{-1/2} \sum_1^n h(x_\nu, \eta_\nu) \right) \Rightarrow N_{(0,\sigma^2)}$$

$$\textit{for } \Gamma^{\mathbb{N}}\textit{-a.a. } (\eta_\nu)_{\nu\in\mathbb{N}} \in \mathrm{H}^{\mathbb{N}}.$$

PROOF. Let

$$\sigma^2(\eta) := \int h(x,\eta)^2 P_\eta(dx)$$

and

$$\tau_n^2(\eta_1, \dots, \eta_n) := \sum_1^n \sigma^2(\eta_\nu).$$

Since $\int \sigma^2(\eta)\Gamma(d\eta) = \sigma^2 \in (0,\infty)$, we have

$$n^{-1}\tau_n^2(\eta_1,\dots,\eta_n) \to \sigma^2, \tag{5.5}$$

for all $(\eta_\nu)_{\nu\in\mathbb{N}}$ outside a $\Gamma^{\mathbb{N}}$-null set, say N'.

Now we introduce the p-measure $M|\mathscr{A} \times \mathscr{B}$ defined by

$$M(A \times B) := \int 1_B(\eta) P_\eta(A)\Gamma(d\eta), \qquad A \in \mathscr{A},\ B \in \mathscr{B}.$$

From (5.3) and (5.4) we obtain

$$\int h(x,\eta) M(d(x,\eta)) = 0 \tag{5.3$'$}$$

and

$$\int h(x,\eta)^2 M(d(x,\eta)) = \sigma^2 \in (0,\infty). \tag{5.4$'$}$$

For $r > 0$ let

$$D(r) := \{(x,\eta) \in X \times \mathrm{H}\colon |h(x,\eta)| > r\}.$$

By (5.4′), for every $\varepsilon > 0$ there exists c_ε such that

$$\int h(x,\eta)^2 1_{D(c_\varepsilon)}(x,\eta) M(d(x,\eta)) < \varepsilon.$$

Hence there exists a $\Gamma^{\mathbb{N}}$-null set N'' such that

$$\limsup_n n^{-1} \sum_1^n \int h(x,\eta_\nu)^2 1_{D(c_\varepsilon)}(x,\eta_\nu) P_{\eta_\nu}(dx) \le \varepsilon \quad \text{for } (\eta_\nu)_{\nu\in\mathbb{N}} \notin N''.$$

Let $c > 0$ be arbitrary. As a consequence of (5.5), $(\eta_\nu)_{\nu\in\mathbb{N}} \notin N'$ implies $c\tau_n(\eta_1,\dots,\eta_n) > c_\varepsilon$ for n sufficiently large. Hence

$$\limsup_n n^{-1} \sum_1^n \int h(x,\eta_\nu)^2 1_{D(c\tau_n(\eta_1,\dots,\eta_n))}(x,\eta_\nu) P_{\eta_\nu}(dx) \le \varepsilon$$

$$\text{for } (\eta_\nu)_{\nu\in\mathbb{N}} \notin N' \cup N''.$$

Since $\varepsilon > 0$ was arbitrary, this implies: For $(\eta_\nu)_{\nu\in\mathbb{N}} \notin N' \cup N''$ and $c > 0$,

$$\lim_n n^{-1} \sum_1^n \int h(x,\eta_\nu)^2 1_{D(c\tau_n(\eta_1,\dots,\eta_n))}(x,\eta_\nu) P_{\eta_\nu}(dx) = 0.$$

Using (5.5), we obtain for $(\eta_\nu)_{\nu\in\mathbb{N}} \notin N' \cup N''$ and $c > 0$,

$$(5.6) \quad \lim_n \tau_n^{-2}(\eta_1,\dots,\eta_n) \sum_1^n \int h(x,\eta_\nu)^2 1_{D(c\tau_n(\eta_1,\dots,\eta_n))}(x,\eta_\nu) P_{\eta_\nu}(dx) = 0.$$

Condition (5.6) is Lindeberg's condition. Hence $(\eta_\nu)_{\nu\in\mathbb{N}} \notin N' \cup N''$ implies

$$\bigtimes_1^n P_{\eta_\nu} * \left(\underline{x} \to \tau_n^{-1}(\eta_1,\dots,\eta_n) \sum_1^n h(x_\nu,\eta_\nu) \right) \Rightarrow N_{(0,1)}.$$

Using (5.5) again, we obtain

$$\bigtimes_1^n P_{\eta_\nu} * \left(\underline{x} \to n^{-1/2} \sum_1^n h(x_\nu,\eta_\nu) \right) \Rightarrow N_{(0,\sigma^2)}. \qquad \square$$

COROLLARY 5.7. *Let $(X,\mathscr{A})$ and $(\mathrm{H},\mathscr{B})$ be measurable spaces. Let $P_\eta|\mathscr{A}$, $\eta \in \mathrm{H}$, be a family of mutually absolutely continuous p-measures such that $\eta \to P_\eta(A)$ is $\mathscr{B}$-measurable for every $A \in \mathscr{A}$.*

Let $h\colon X \to \mathbb{R}$ be an $\mathscr{A}$-measurable function fulfilling

$$(5.8) \qquad \int h(x) P_\Gamma(dx) = 0$$

and

$$(5.9) \qquad \int h(x)^2 P_\Gamma(dx) \in (0,\infty).$$

Let

$$\sigma^2 := \int h(x)^2 P_\Gamma(dx) - \int\left(\int h(x) P_\eta(dx)\right)^2 \Gamma(d\eta). \tag{5.10}$$

Then $\sigma^2 \in (0, \infty)$ *and*

$$\bigtimes_1^\infty P_{\eta_\nu} * \left(\underline{x} \to n^{-1/2} \sum_1^n \left(h(x_\nu) - \int h(x) P_{\eta_\nu}(dx)\right)\right) \Rightarrow N_{(0,\sigma^2)}$$

for $\Gamma^{\mathbb{N}}$-*a.a.* $(\eta_\nu)_{\nu\in\mathbb{N}} \in \mathrm{H}^{\mathbb{N}}$.

PROOF. We apply Lemma 5.2 with $h(x, \eta) := h(x) - \int h\, dP_\eta$. Condition (5.3) is obvious. It remains to prove (5.4).

$\sigma^2 < \infty$ follows immediately from (5.9). By Cauchy–Schwarz,

$$\left(\int h(x) P_\eta(dx)\right)^2 \le \int h(x)^2 P_\eta(dx) \quad \text{for all } \eta \in \mathrm{H}. \tag{5.11}$$

It remains to prove $\sigma^2 > 0$. The relation $\sigma^2 = 0$ implies equality in (5.11) for all η outside a Γ-null set, say N. Because of (5.8), $\eta \notin N$ implies $h(x) = 0$ for P_η-a.a. $x \in X$ and therefore $\int h(x)^2 P_\eta(dx) = 0$. This, however, contradicts (5.9). □

LEMMA 5.12. *For* $\eta \in \mathrm{H}$, *let* $P_\eta|\mathscr{A}$ *be a p-measure admitting a density* $p(\cdot, \eta)$ *with respect to some σ-finite measure. For* $f \in \mathscr{L}_*(P_\Gamma)$, *the function* $\eta \to \int f\, dP_\eta$ *is* Γ-*orthogonal to* $T(\Gamma, \mathscr{G})$ *iff* f *is* $P_{\vartheta,\Gamma}$-*orthogonal to*

$$\left\{\frac{\int k(\eta) p(\cdot, \eta)\Gamma(d\eta)}{\int p(\cdot, \eta)\Gamma(d\eta)} : k \in T(\Gamma, \mathscr{G})\right\}.$$

ADDENDUM. If $T(\Gamma, \mathscr{G}) = \mathscr{L}_*(\Gamma)$, then

$$\int f\, dP_\eta = 0 \quad \text{for } \Gamma\text{-a.a. } \eta \in \mathrm{H}$$

iff f is $P_{\vartheta,\Gamma}$-orthogonal to

$$\left\{\frac{\int k(\eta) p(\cdot, \eta)\Gamma(d\eta)}{\int p(\cdot, \eta)\Gamma(d\eta)} : k \in \mathscr{L}_*(\Gamma)\right\}.$$

PROOF. By Fubini's theorem, $f \in \mathscr{L}_*(P_\Gamma)$ and $k \in T(\Gamma, \mathscr{G})$ implies

$$\int\int f(x) k(\eta) P_\eta(dx)\Gamma(d\eta) = \int f(x) \frac{\int k(\eta) p(x, \eta)\Gamma(d\eta)}{\int p(x, \eta)\Gamma(d\eta)} P_\Gamma(dx).$$

This implies the assertion. □

Acknowledgments. The author wishes to thank W. Wefelmeyer for helpful discussions. R. Hamböker checked the technical details and suggested various improvements. He contributed the proof of (4.34). The comments of the reviewer and an Associate Editor contributed much to improve the presentation. The discussion following a presentation of this paper at the Wirtschaftsuniversität Wien motivated the author to improve Example 2.

REFERENCES

AMARI, S. and KUMON, M. (1988). Estimation in the presence of infinitely many nuisance parameters—geometry of estimating functions. *Ann. Statist.* **16** 1044–1068.

ANDERSEN, E. B. (1970a). On Fisher's bound for asymptotic variances in the case of infinitely many incidental parameters. *Scand. Actuar. J.* **53** 78–85.

ANDERSEN, E. B. (1970b). Asymptotic properties of conditional maximum-likelihood estimators. *J. Roy. Statist. Soc. Ser. B* **32** 283–301.

ANDERSON, T. W. (1971). *The Statistical Analysis of Time-Series.* Wiley, New York.

BICKEL, P. J. and KLAASSEN, C. A. J. (1986). Empirical Bayes estimation in functional and structural models, and uniformly adaptive estimation of location. *Adv. in Appl. Math.* **7** 55–69.

BICKEL, P. J. and RITOV, Y. (1987). Efficient estimation in the errors in variables model. *Ann. Statist.* **15** 513–540.

BUNKE, O. (1979). Comment on "Asymptotic efficiency of estimators of a multivariate linear functional relation" by M. Nussbaum. *Math. Operationsforsch. Statist. Ser. Statist.* **10** 521–522.

DE LEEUW, J. and VERHELST, N. (1986). Maximum likelihood estimation in generalized Rasch models. *Journal of Educational Statistics* **11** 183–196.

FOLLMAN, D. (1988). Consistent estimation in the Rasch model based on nonparametric margins, *Psychometrika* **53** 553–562.

GODAMBE, V. P. (1976). Conditional likelihood and unconditional optimum estimating equations. *Biometrika* **63** 277–284.

HAMBLETON, R. K. and SWAMINATHAN, H. (1985). *Item Response Theory.* Kluwer Nijhoff Publishing, Boston.

KUMON, M. and AMARI, S. (1984). Estimation of a structural parameter in the presence of a large number of nuisance parameters. *Biometrika* **71** 445–459.

LANGROCK, P. and JAHN, W. (1979). *Einführung in die Theorie der Markovschen Ketten und ihre Anwendungen.* Teubner, Leipzig.

LEHMANN, E. L. (1983). *Theory of Point Estimation.* Wiley, New York.

LINDLEY, D. V. (1971). The estimation of many parameters. In *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.) 435–455. Holt, Rinehart and Winston, Toronto.

LUKACS, E. (1970). *Characteristic Functions*, 2nd ed. Griffin, London.

MORAN, P. A. P. (1971). Estimating structural and functional relations. *J. Multivariate Anal.* **1** 232–255.

NEYMAN, J. and SCOTT, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16** 1–32.

NOVICK, M. R., LEWIS, C. and JACKSON, P. H. (1973). The estimation of proportions in m groups. *Psychometrika* **38** 19–46.

NUSSBAUM, M. (1979). Asymptotic efficiency of estimators of a multivariate linear functional relation (with discussion). *Math. Operationsforsch. Statist. Ser. Statist.* **10** 505–527.

PFANZAGL, J. (1987). Bounds for the asymptotic efficiency of estimators based on functional contractions; applications to the problem of estimation in the presence of random nuisance parameters. In *Proceedings of the First World Congress of the Bernoulli Society* (Yu. Prohorov and V. V. Sazonov, eds.) *Probability Theory and Mathematical Statistics* **2** 237–248. VNU Science Press, Utrecht.

PFANZAGL, J. (1990). *Estimation in Semiparametric Models. Some Recent Developments. Lecture Notes in Statist.* **63** Springer, New York.

PFANZAGL, J. and WEFELMEYER, W. (1982). *Contributions to a General Asymptotic Statistical Theory. Lecture Notes in Statist.* **13**. Springer, New York.

PORTNOY, S. (1985). Asymptotic behavior of M estimators of p regression parameters when p^2/n is large. II. Normal approximation. *Ann. Statist.* **13** 1403–1417.

PORTNOY, S. (1987). Central limit theorem applicable to robust regression estimators. *J. Multivariate Anal.* **22** 24–50.

RITOV, Y. and BICKEL, P. J. (1990). Achieving information bounds in non- and semiparametric models. *Ann. Statist.* **18** 925–938.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

SKOVGAARD, IB. M. (1989). A review of higher order likelihood inference. *Bulletin of the International Statistical Institute* **53** 331–351.

SPRENT, P. (1966). A generalized least-squares approach to linear functional relationships. *J. Roy. Statist. Soc. Ser. B* **28** 278–297.

SPRENT, P. (1976). Modified likelihood estimation of a linear relationship. In *Studies in Probability and Statistics* (E. J. Williams, ed.) 109–119. North-Holland, Amsterdam.

TEICHER, H. (1960). On the mixture of distributions. *Ann. Math. Statist.* **31** 55–73.

VAN DER VAART, A. (1988). *Statistical Estimation in Large Parameter Spaces. CWI Tract* **44**. Math. Centrum, Amsterdam.

WOLD, H. (1954). *A Study in the Analysis of Stationary Time Series*, 2nd ed. Almqvist and Wiksells, Stockholm.

WONG, W. H. and SEVERINI, T. A. (1991). On maximum likelihood estimation in infinite dimensional parameter spaces. *Ann. Statist.* **19** 603–632.

MATHEMATISCHES INSTITUT
UNIVERSITÄT ZU KÖLN
50923 KÖLN
GERMANY