The Annals of Statistics
1982, Vol. 10, No. 3, 990–1005

QUICK CONSISTENCY OF QUASI MAXIMUM LIKELIHOOD ESTIMATORS

BY THOMAS PFAFF

University of Kaiserslautern

A family of probability measures $\mathscr{P}$ on some measurable space $(X, \mathscr{A})$ and a class of estimator sequences $\hat{P}_n : X^n \to \mathscr{P}$, $n \in \mathbb{N}$, containing maximum likelihood estimators are considered. For $P \in \mathscr{P}$ it is proved that there are numbers $c > 0$, $h_0 > 0$ fulfilling $P^n\{n^{1/2}\, d(\hat{P}_n, P) > h\} \leq \exp(-ch^2)$ for $n \in \mathbb{N}$, $h \geq h_0$, where d denotes the Hellinger distance of probability measures. Then parameterized families $\mathscr{P} = \{P(\theta) : \theta \in \Theta\}$ are considered where (Θ, Δ) is a separable and finite-dimensional metric space, and for sequences $\hat{\theta}_n : X^n \to \Theta$, $n \in \mathbb{N}$, estimating the parameter similar inequalities are derived.

1. Introduction. Let $\mathscr{P}$ denote a family of probability measures on some measurable space $(X, \mathscr{A})$. The problem is to estimate the "true" measure $P \in \mathscr{P}$ from n independent observations $x_1, \cdots, x_n \in X$. Under conditions which are basically generalizations of the conditions first discussed by Wald (1949) in proving strong consistency of maximum likelihood estimators, we prove for some class of estimator sequences $\hat{P}_n : X^n \to \mathscr{P}$, $n \in \mathbb{N}$, containing maximum likelihood estimators

$$P^n\{d(\hat{P}_n, P) > \varepsilon\} \leq e^{-c(\varepsilon)n}, \quad \text{for} \quad n \in \mathbb{N}, \tag{1.1}$$

which is then used for proving

$$P^n\{\sqrt{n}\, d(\hat{P}_n, P) > h\} \leq e^{-ch^2}, \quad \text{for} \quad n \in \mathbb{N}, \quad h \geq h_0. \tag{1.2}$$

(Here d denotes the Hellinger distance of probability measures; see Section 2.) These results are presented in Section 3. The proof of (1.1) is rather simple, whereas the proof of (1.2) is based on a fluctuation inequality for random functions (Lemma (7.4)). Relation (1.2) is valid for separable and finite-dimensional spaces $(\mathscr{P}, d)$. For parametric models $\mathscr{P} = \{P(\theta) : \theta \in \Theta\}$ with (Θ, Δ) being a finite-dimensional metric space we translate (1.1) and (1.2) into the space of parameters (see the results in Sections 4 and 5).

For parametric models with open $\Theta \subseteq \mathbb{R}^p$, the basic theory for consistent maximum likelihood estimation based on independent and identically distributed observations is due to Wald (1949). This theory was developed in the papers of Landers (1968), Pfanzagl (1969), Michel and Pfanzagl (1971), Landers (1972). Wald (1949) verified strong consistency for sequences of maximum likelihood estimators, $\hat{\theta}_n : X^n \to \Theta$, $n \in \mathbb{N}$:

$$P(\theta)^{\mathbb{N}}\{\lim_{n\in\mathbb{N}} \|\hat{\theta}_n - \theta\| = 0\} = 1. \tag{1.3}$$

Two stochastic inequalities for the concentration of the estimators about the true parameter θ are side-results in the papers of Michel and Pfanzagl (1971, Lemma 4 and Lemma 6) and of Chibisov (1973, Lemma 1), namely,

$$P(\theta)^n\{\|\hat{\theta}_n - \theta\| \geq \varepsilon\} \leq c_{s,\varepsilon,\theta} n^{-s/2}, \quad \text{for} \quad n \in \mathbb{N}, \tag{1.4}$$

$$P(\theta)^n\{\sqrt{n}\,\|\hat{\theta}_n - \theta\| \geq 2\beta(\theta)(\log n)^{1/2}\} \leq c_\theta n^{-1/2}, \quad \text{for} \quad n \in \mathbb{N}. \tag{1.5}$$

Note that (1.4) implies (1.3). Ibragimov and Khas'minskii (1973, Theorem 1 and Theorem 2) proved that with suitable $\lambda > 0$

Received December 1979; revised March 1982

AMS 1980 *subject classifications.* Primary, 62G05; secondary, 62F10.

Key words and phrases. Consistency of estimators, maximum likelihood estimation, exponential bounds for consistency.

(1.6) $$P(\theta)^n\{n^\lambda \| \hat{\theta}_n - \theta \| > h\} \le c_{b,\theta} h^{-b}, \quad \text{for} \quad n \in \mathbb{N}, \quad h \ge H_{b,\theta}, \quad b > 0.$$

Relation (1.6) is equivalent to the fact that for arbitrary $b > 0$ $\lim \sup_{n \in \mathbb{N}} n^{b\lambda} E_\theta[\| \hat{\theta}_n - \theta \|^b] < \infty$; and (1.6) was used by Gusev (1976) and by Pfaff (1977) in order to gain asymptotic expansions of moments of $n^{1/2}(\hat{\theta}_n - \theta)$.

In the present paper we present improvements of (1.4) and (1.5) (see Theorem 5.1) which hold true under very general assumptions. Michel and Pfanzagl (1971) and Chibisov (1973) used assumptions concerning continuity and differentiability of densities and concerning existence of moments of the derivatives; we avoid these conditions.

The result of Ibragimov and Khas'minskii (1973) was obtained for $\Theta \subseteq \mathbb{R}^1$ under restrictive conditions excluding e.g scale parameter families. Our result (Theorem 6.5) covers multi-dimensional parameter spaces and it is well applicable to location and scale families $\mathscr{P}$.

Quick consistency of $\hat{\theta}_n$ means that for some $\lambda > 0$, $n^\lambda \| \hat{\theta}_n - \theta \|$ stays bounded in P_θ^n-probability (see (1.5), (1.6)). For sufficiently regular families $\mathscr{P} = \{P(\theta) : \theta \in \Theta\}$ (with densities satisfying some differentiability- and moment-conditions) the largest accessible number λ is ½ because the distributions of $\sqrt{n}(\hat{\theta}_n - \theta)$ under P_θ^n are asymptotically normal; see e.g. Michel and Pfanzagl (1971). Examples of nonregular families are known where ½ $\le \lambda \le 1$; see, e.g., the example given by Akahira and Takeuchi (1981, pages 27–51). However, it seems to be common to both cases that $\sqrt{n}\ d(P(\hat{\theta}_n), P(\theta))$ stays bounded in P_θ^n-probability; c.f. LeCam's (1973) introduction, also our Remark (5.5).

Therefore, our concept seems to be reasonable, namely, starting with proving (1.1) and (1.2) for a very general non-parametric model $\mathscr{P}$ and then applying these results to parametric models.

For Bayesian estimators Strasser (1981b, 1981c) recently obtained results which are analogous to our Theorems 3.3, 3.10 and 5.1.

In the cited papers of Michel-Pfanzagl and Chibisov relations (1.4) and (1.5) were proved to hold locally uniformly in θ. It is easy to see that the inequalities established in the present paper also hold true locally uniformly under some suitable "uniform version" of our conditions. We do not treat these possible generalizations in order to simplify our presentation.

Following the concept of Landers (1972) and Strasser (1981a), we base our definition of quasi maximum likelihood estimators on separable random functions being equivalent to the likelihood functions (see Lemma 2.1 and Definition 2.2). This concept has a number of advantages: measurable estimators always exist (see Remark 2.3). The proofs are simplified because no measurability problems arise. Continuity conditions assuring separability of the likelihood functions (c.f. Pfanzagl, 1969, Lemma 3.8) are avoided. Nevertheless our theory is applicable to all important cases, where it is possible to choose versions of the densities such that the likelihood funtions become separable (then maximum likelihood estimators in our sense are maximum likelihood estimators in the usual sense).

Until we present our results in Sections 3, 4, 5 and 6, we give the basic notations and definitions in Section 2; the proofs are collected in Section 7.

2. Preliminaries. Let $\mathbb{N}$ denote the natural numbers and $\mathbb{R}$ the real numbers. Let $\nu \mid \mathscr{A}$ be a σ-finite measure dominating the family $\mathscr{P} \mid \mathscr{A}$. For $Q \in \mathscr{P}$, let h_Q be a density of $Q \mid \mathscr{A}$ relative to $\nu \mid \mathscr{A}$. For $n \in \mathbb{N}$, let $P^n \mid \mathscr{A}^n$, $\nu^n \mid \mathscr{A}^n$ denote the n-fold independent product of identical components $P \mid \mathscr{A}$, respectively $\nu \mid \mathscr{A}$.

For P, $Q \in \mathscr{P}$ the Hellinger distance $d(P, Q)$ and the affinity $a(P, Q)$ are defined as follows:

$$d(P, Q)^2 = 1 - a(P, Q) = \frac{1}{2} \int (h_Q^{1/2} - h_P^{1/2})^2 \, d\nu.$$

The metric space $(\mathscr{P}, d)$ is assumed separable. The open balls in $(\mathscr{P}, d)$ are denoted by $B(P, r)$. Let $\mathscr{B}$ be the Borel-σ-field on $\mathscr{P}$.

For $n \in \mathbb{N}$ and $\underset{\sim}{x} = (x_1, \cdots, x_n) \in X^n$ the function $h_{n,Q}(\underset{\sim}{x}) = \prod_{j=1}^n h_Q(x_j)$ is called the likelihood function. The following Lemma 2.1 proves for every $n \in \mathbb{N}$ the existence of a separable random function being equivalent to the random function $(h_{n,Q})_{Q\in\mathscr{P}}$. The notions "random function," "equivalence of random functions" and "separability of a random function" are defined in Definition 7.3.

2.1. LEMMA. *For any countable, dense subset $S \subseteq \mathscr{P}$ there exist functions $f_{n,Q}: X^n \to [0, \infty)$, $n \in \mathbb{N}$, $Q \in \mathscr{P}$, having the properties*:
(i) $\underset{\sim}{x} \to f_{n,Q}(\underset{\sim}{x})$ *is $\mathscr{A}^n$-measurable, for every $n \in \mathbb{N}$ and every $Q \in \mathscr{P}$.*
(ii) $f_{n,Q} = h_{n,Q}\nu^n$-a.e., *for every $n \in \mathbb{N}$ and every $Q \in \mathscr{P}$.*
(iii) *For every $n \in \mathbb{N}$, every $Q \in \mathscr{P}$ and every $\underset{\sim}{x} \in X^n$ there exists a sequence $(Q_m)_{m\in\mathbb{N}} \subseteq S$ satisfying*

$$\lim_{m\in\mathbb{N}} d(Q_m, Q) = 0 \quad \textit{and} \quad \lim_{m\in\mathbb{N}} f_{n,Q_m}(\underset{\sim}{x}) = f_{n,Q}(\underset{\sim}{x}).$$

PROOF. See Strasser (1981a, page 1108).

We choose a countable dense subset S and a family $(f_{n,Q})_{n\in\mathbb{N},Q\in\mathscr{P}}$ satisfying (i), (ii), (iii) and keep them fixed throughout the paper.

2.2. DEFINITION. Let $0 < \gamma \le 1$ and $P \in \mathscr{P}$. A sequence of $(\mathscr{A}^n, \mathscr{B})$-measurable functions $\hat{P}_n: X^n \to \mathscr{P}$, $n \in \mathbb{N}$, is called sequence of quasi maximum likelihood estimators for P, with coefficient γ, relative to the family $(f_{n,Q})_{n\in\mathbb{N},Q\in\mathscr{P}}$, if for every $n \in \mathbb{N}$ and for P^n-a.e. $\underset{\sim}{x} \in X^n$ we have

$$f_{n,\hat{P}_n(\underset{\sim}{x})}(\underset{\sim}{x}) \ge \gamma \sup_{Q\in\mathscr{P}} f_{n,Q}(\underset{\sim}{x}) \quad \text{or} \quad \sup_{Q\in\mathscr{P}} f_{n,Q}(\underset{\sim}{x}) = +\infty.$$

A sequence of quasi maximum likelihood estimators with coefficient 1 is called sequence of maximum likelihood estimators. The name "quasi maximum likelihood estimator with coefficient γ" was qually used by Roussas (1965).

2.3. REMARK. For every $\gamma \in (0, 1)$ there exists a sequence of quasi maximum likelihood estimators with coefficient γ; c.f. Strasser (1981a, page 1109).

2.4. DEFINITION. A subset $M \subseteq T$ of a metric space (T, ρ) is called finite-dimensional with dimension $D > 0$ if there is a constant $C < \infty$ such that for every $t_0 \in M$ and every pair $0 < r < R < \infty$ there is a finite collection of balls $B_\rho(t_\mu, r) = \{t \in T: \rho(t, t_\mu) < r\}$, $\mu = 1, \cdots, m$, covering the set $B_\rho(t_0, R) \cap M$, where $t_\mu \in M$ and $m \le C(R/r)^D$.

3. Convergence relative to the Hellinger distance. We formulate two conditions which imply relation (1.1).

3.1. CONDITION. There exist a compact subset $K \subseteq \mathscr{P}$, with $P \in K$, and numbers $n \in \mathbb{N}$ and $p > 1$, $q > 1$, with $1/p + 1/q = 1$, satisfying

$$\int (\sup_{Q\in\mathscr{P}-K} f_{n,Q}^{1/p}) f_{n,P}^{1/q} \, d\nu^n < 1.$$

3.2. CONDITION. For the compact set K in Condition 3.1 and for every $Q_0 \in K$ there exist a ball $B(Q_0, \delta)$ and numbers $n \in \mathbb{N}$ and $p > 1$, $q > 1$, with $1/p + 1/q = 1$, satisfying

$$\int (\sup_{Q\in B(Q_0,\delta)} f_{n,Q}^{1/p}) f_{n,P}^{1/q} \, d\nu^n \begin{cases} <1 & \text{if} \quad Q_0 \ne P, \\ <\infty & \text{if} \quad Q_0 = P. \end{cases}$$

Under these conditions, which will be made more transparent by Lemma 3.5 and by Remark 3.6 below, the following result is true.

3.3. LEMMA. *Let $P \in \mathscr{P}$ and assume that Conditions* 3.1 *and* 3.2 *are fulfilled for this measure P. Then every sequence $(\hat{P}_n)_{n\in N}$ of quasi maximum likelihood estimators for P has the properties*:

For arbitrary $\varepsilon > 0$ there exist numbers $n_0 \in N$ and $c > 0$ such that for every $n \geq n_0$

$$P^n\{d(\hat{P}_n, P) > \varepsilon\} \leq e^{-cn} \quad \textit{and} \quad P^n\{\sup_{Q\in\mathscr{P}} f_{n,Q} = \infty\} = 0. \tag{3.4}$$

PROOF. See Section 7.1.

3.5. LEMMA. *Conditions* 3.1 *and* 3.2 *are valid for $P \in \mathscr{P}$ if the family has the following properties*:

(i) *$(\mathscr{P}, d)$ is locally compact.*

(ii) *For P-a.e. $x \in X$ the map $Q \to h_Q(x)$ is upper semicontinuous.*

(iii) *There is $n \in N$ such that for P^n-a.e. $\underset{\sim}{x} \in X^n$ and for every $\varepsilon > 0$ exists a compact set $K \subseteq \mathscr{P}$ satisfying*

$$\sup_{Q\in\mathscr{P}-K} f_{n,Q}(x) < \varepsilon.$$

(*This condition means $\lim_{Q\to P^*} f_{n,Q}(\underset{\sim}{x}) = 0$ if $\mathscr{P}^* =. \mathscr{P} \cup \{P^*\}$ denotes the compactification of $\mathscr{P}$.*)

(iv) *For every $Q_0 \in \mathscr{P}$ there exist a ball $B(Q_0, \delta)$ and numbers $n \in N$ and $p > 1$, $q > 1$, $1/p + 1/q = 1$, satisfying*

$$\int (\sup_{Q\in B(Q_0,\delta)} f_{n,Q}^{1/p}) f_{n,P}^{1/q}\, d\nu^n < \infty.$$

(v) *There exist a compact set $K \subseteq \mathscr{P}$ and numbers $n \in N$ and $p > 1$, $q > 1$, $1/p + 1/q = 1$, satisfying*

$$\int (\sup_{Q\in\mathscr{P}-K} f_{n,Q}^{1/p}) f_{n,P}^{1/q}\, d\nu^n < \infty.$$

PROOF. See Section 7.2.

3.6. REMARKS.

1. In Lemma 3.5 (iv), (v) the integrals can be replaced by $\int \sup_Q f_{n,Q}/f_{n,P}\, dP$.
2. For compact $(\mathscr{P}, d)$ Condition 3.1 and the assumptions (i), (iii), (v) in Lemma 3.5 are obviously fulfilled. (Choose $K = \mathscr{P}$.)
3. If any of the Conditions 3.1, 3.2 and 3.5 (iii), (iv), (v) holds with $n = n_0$ then it holds for arbitrary $n \geq n_0$. (See the relations (7.1.3), (7.1.4) in the proof of Lemma 3.3.) However, it may happen that the conditions are not valid for small numbers n. Let e.g. $\mathscr{P}$ be the family of m-variate normal distributions, then (iii) and (v) are valid for $n \geq m + 1$ and they are violated for $n < m + 1$. See also the remark by Kiefer and Wolfowitz (1956).

Two additional conditions are needed for Theorem 3.10 which yields relation (1.2):

3.7. CONDITION. For some suitable $\delta' > 0$ the ball $B(P, \delta')$ is finite-dimensional with dimension $D > 0$.

3.8. CONDITION. There are numbers $\delta'' > 0$, $s \geq 2$, $C < \infty$ fulfilling $s > D$ (D is the dimension in Condition 3.7) and $\int |\log h_Q|\, dP \le C$ for every $Q \in B(P, \delta'')$ and

$$\int |\ell_{Q_1} - \ell_{Q_2}|^s \, dP \le C\, d(Q_1, Q_2)^s \tag{3.9}$$

for every pair $Q_1, Q_2 \in B(P, \delta'')$ where

$$\ell_Q = \log h_Q - \int \log h_Q \, dP \quad \text{for} \quad Q \in \mathscr{P}.$$

The idea of using a local dimensionality condition like Condition 3.7 in order to prove consistency of estimates is due to LeCam (1973); this idea was also applied by Strasser (1981c).

It should be noted that Condition 3.8 implies that every $Q \in B(P, \delta'')$ dominates the measure P.

3.10. THEOREM. *Let $P \in \mathscr{P}$ and assume that P satisfies Conditions* 3.1, 3.2, 3.7 *and* 3.8, *and let $(\hat{P}_n)_{n\in N}$ be a sequence of quasi maximum likelihood estimators for P. Then the properties* (i) *and* (ii) *come true*:

(i) *There are constants $a > 0$ and $\varepsilon_0 > 0$ such that*

$$P^n\{n^{1/2}\, d(\hat{P}_n, P) > a \log(1/\varepsilon)^{1/2}\} \le \varepsilon \tag{3.11}$$

holds for every $n \in N$ and every $\varepsilon \in (0, \varepsilon_0]$.

(ii) *There are constants $h_0 > 0$, $c > 0$ such that*

$$P^n\{n^{1/2}\, d(\hat{P}_n, P) > h\} \le e^{-ch^2} \tag{3.12}$$

for every $n \in N$ and every $h \ge h_0$.

(iii) *The assertions* (i) *and* (ii) *imply each other and, moreover, they imply that for arbitrary $s > 0$ there are constants $a > 0$, $C_s < \infty$ such that for every $n \in N$*

$$P^n\{n^{1/2}\, d(\hat{P}_n, P) > a(s \log n)^{1/2}\} \le C_s n^{-s}. \tag{3.13}$$

PROOF. See Section 7.6.

4. Verification of the conditions for parametrized families.

Let (Θ, Δ) denote a separable and locally compact metric space and let $\Theta \ni \tau \to P(\tau) \in \mathscr{P}$ be a bijective map. In this case $\mathscr{P}$ is called parametrized family and τ is called parameter of the measure $P(\tau)$.

The balls in the metric $\Delta[d]$ are denoted by $B_\Delta(\tau, \varepsilon)[B(P, \varepsilon)]$, $\mathscr{B}_\Delta[\mathscr{B}]$ denotes the Borel field on $\Theta[\mathscr{P}]$.

We establish four assumptions on the family $\mathscr{P}$; the third and fourth assumptions refer to some $\theta_0 \in \Theta$:

4.1. CONDITION. For every $\theta \in \Theta$ there are numbers $c_1(\theta) > 0$, $c_2(\theta) > 0$, $\alpha(\theta) > 0$, $\delta_1(\theta) > 0$ satisfying

$$c_1(\theta)\Delta(\sigma, \tau)^{\alpha(\theta)} \le d(P(\sigma), P(\tau)) \le c_2(\theta)\Delta(\sigma, \tau)^{\alpha(\theta)}$$

for $\sigma, \tau \in B_\Delta(\theta, \delta_1(\theta))$.

4.2. CONDITION. For every $\theta \in \Theta$ there is a compact subset $K(\theta) \subseteq \Theta$ and a number $\eta(\theta) > 0$ such that

$$d(P(\tau), P(\theta)) \ge \eta(\theta) \quad \text{for every} \quad \tau \in \Theta - K(\theta).$$

4.3. CONDITION. There are numbers $\delta_2 > 0$, $D_0 > 0$ such that $B_\Delta(\theta_0, \delta_2)$ is finite-dimensional with dimension D_0.

4.4. CONDITION. There are numbers $\delta_3 > 0$, $s \geq 2$, $C < \infty$ fulfilling $s > D_0/\alpha(\theta_0)$ and $\int |\log h_{P(\tau)}| \, dP(\theta_0) \leq C$ for $\tau \in B_\Delta(\theta_0, \delta_3)$ and

$$\int |\ell_{P(\sigma)} - \ell_{P(\tau)}|^s \, dP(\theta_0) \leq C\Delta(\sigma, \tau)^{\alpha(\theta_0)s}$$

for $\sigma, \tau \in B_\Delta(\theta_0, \delta_3)$, where $\alpha(\theta_0)$, D_0 are the numbers occuring in Conditions 4.1, 4.3 and where ℓ_Q was defined in Condition 3.8.

4.5. LEMMA. *Assume that Conditions* 4.1–4.4 *are satisfied for* $\theta_0 \in \Theta$. *Then Conditions* 3.7 *and* 3.8 *are fulfilled for* $P(\theta_0)$, *the dimension D in Condition* 3.7 *is equal to* $D_0/\alpha(\theta_0)$, *the number s in Condition* 3.8 *is equal to the number s in Condition* 4.4. *Moreover, the map* $\tau \rightarrow P(\tau)$ *is a homeomorphism.*

PROOF. See Section 7.7.

4.6. REMARK. Since (Θ, Δ) is separable and since the map $\tau \rightarrow P(\tau)$ is a homeomorphism, $(\mathscr{P}, d)$ is also separable. Let S_0 be a countable dense subset of Θ. Then, according to Lemma 2.1, for $S = \{P(\tau) : \tau \in S_0\}$ there exist separable versions $(f_{n,Q})_{Q\in\mathscr{P}}$ of the likelihood functions $(h_{n,Q})_{Q\in\mathscr{P}}$. This means that for every $n \in N$, $\theta \in \Theta$, $\underset{\sim}{x} \in X^n$ exists a sequence $(\theta_m)_{m\in N} \subseteq S_0$ fulfilling $\lim_{m\in N}\Delta(\theta_m, \theta) = 0$ and $\lim_{m\in N} f_{n,P(\theta_m)}(\underset{\sim}{x}) = f_{n,P(\theta)}(\underset{\sim}{x})$.

Since the parametrization is a homeomorphism it is quite simple to express the Conditions 3.1 and 3.2 and the Assumptions (i)–(v) of Lemma 3.5 using the metric Δ. This will be left to the reader.

For notational convenience, we shall write $f_{n,\tau}$, $h_{n,\tau}$, h_τ, ℓ_τ in place of $f_{n,P(\tau)}$, $h_{n,P(\tau)}$, $h_{P(\tau)}$, $\ell_{P(\tau)}$.

4.7. DEFINITION. Let $\hat{\theta}_n : X^n \rightarrow \Theta$, $n \in N$, be a sequence of $(\mathscr{A}^n, \mathscr{B}_\Delta)$-measurable maps, let $\gamma \in (0, 1]$ and let $\theta_0 \in \Theta$. $(\hat{\theta}_n)_{n\in N}$ is called sequence of quasi maximum likelihood estimators for θ_0 iff $(P(\hat{\theta}_n))_{n\in N}$ is a sequence of quasi maximum likelihood estimators for $P(\theta_0)$.

From Lemma 4.5 and Remark 4.6 we conclude:

4.8. COROLLARY. *Let* $\theta_0 \in \Theta$. *Assume that Conditions* 3.1 *and* 3.2 *are fulfilled for* $P(\theta_0)$, *assume that Conditions* 4.1, 4.2, 4.3, 4.4 *are satisfied for* θ_0, *and let* $(\hat{\theta}_n)_{n\in N}$ *be a sequence of quasi maximum likelihood estimators for* θ_0. *Then the assertions of Lemma* 3.3 *and of Theorem* 3.10 *are valid for* $P = P(\theta_0)$ *and for* $\hat{P}_n = P(\hat{\theta}_n)$.

We give two important examples where Conditions 4.1 and 4.4 can easily be verified:

4.9. EXAMPLE. (Families fulfilling Cramér-Wald conditions). Let Θ be an open subset of $\mathbb{R}^k$, and let $\Delta(\tau, \theta) := \|\tau - \theta\|_e$ where $\| \, \|_e$ denotes the Euclidean norm. A family $\mathscr{P} = \{P(\theta) : \theta \in \Theta\}$ of mutually absolutely continuous measures is called satisfying Cramér-Wald conditions, if the functions $\theta \rightarrow \log h_\theta(x)$ admit continuous derivatives of a sufficiently high order, if these derivatives fulfill certain moment conditions and if the Fisher information matrix $I(\theta)$ is positive definite for every θ. Under such conditions, it is possible to prove with the aid of Taylor expansion techniques (see, e.g., Ibragimov and Khas'minskii, 1972, page 452, Lemma 2.2 and Lemma 2.3, page 455, Lemma 2.5) that

$$d(P(\sigma), P(\tau))^2 = \tfrac{1}{8}(\sigma - \tau)^T \cdot I(\tau) \cdot (\sigma - \tau) + O(\| \sigma - \tau \|_e^3)$$

where the convergence is uniformly for τ in some ball $B_\Delta(\theta, \delta)$. This implies Condition 4.1 with $\alpha(\theta) = 1$. Condition 4.4 may be derived from the inequality

$$\begin{aligned} &\int |\log h_\sigma - \log h_\tau|^s \, dP(\theta) \\ (4.10) \qquad &\leq \| \sigma - \tau \|_e^s \sup_{0 \leq t \leq 1} \int |(\partial/\partial t) \log h_{\tau + t(\sigma - \tau)}(x)|^s P(\theta)(dx). \end{aligned}$$

4.11. Example. (Exponential families). A family $\mathscr{P} = \{P(\theta) : \theta \in \Theta\}$, with open $\Theta \subseteq \mathbb{R}^k$, is called exponential family if the densities h_θ admit a representation

$$h_\theta(x) = c(\theta) \exp\{\textstyle\sum_{j=1}^k \theta^{(j)} g^{(j)}\},$$

where $g = (g^{(1)}, \cdots, g^{(k)})^T : X \to \mathbb{R}^k$ is $\mathscr{A}$-measurable.

Assume that the functions $g^{(j)}$ have locally bounded moments of order three and that the covariance matrix $\mathrm{Cov}_\theta(g)$ is positive definite for every $\theta \in \Theta$. Then

$$d(P(\sigma), P(\tau))^2 = \tfrac{1}{8}(\sigma - \tau)^T \cdot \mathrm{Cov}_\tau(g) \cdot (\sigma - \tau) + O(\| \sigma - \tau \|_e^3)$$

holds uniformly for τ in some ball $B_\Delta(\theta, \delta)$ and, therefore, Condition 4.1 is satisfied with $\alpha(\theta) = 1$. Condition 4.4 may be concluded from

$$(4.12) \qquad \int |\ell_\sigma - \ell_\tau|^s \, dP(\theta) \leq \| \sigma - \tau \|_e^s \left(2 \sum_{j=1}^k \left[\int |g^{(j)}|^s \, dP(\theta) \right]^{1/s} \right)^s.$$

Clearly, the Examples 4.9 and 4.11 satisfy Condition 4.3, with $D_0 = k$ and with arbitrary $\delta_2 > 0$, and they satisfy conditions (i), (ii) in Lemma 3.5. Thus, the assertions of Theorems 3.3 and 3.10 come true if, additionally, Condition 4.2 and Lemma 3.5 (iii)–(v) are satisfied. Since $\theta \to h_\theta(x)$ is continuous for every x we may choose $f_{n,\tau} = h_{n,\tau}$.

5. Convergence of estimators for parameters. In this section we transform relations (3.4), (3.11), and (3.13) into inequalities for quasi maximum likelihood estimators for parameters.

5.1. Theorem. *Let $\theta_0 \in \Theta$ and let $(\hat{\theta}_n)_{n \in N}$ be a sequence of quasi maximum likelihood estimators for θ_0. Assume that Conditions* 3.1 *and* 3.2 *are valid for $P(\theta_0)$, and let Conditions* 4.1, 4.2, 4.3, 4.4 *be fulfilled for θ_0. Then for every $\varepsilon > 0$ there are numbers $n_0 \in N$ and $c > 0$ such that*

$$(5.2) \qquad P(\theta_0)^n \{\Delta(\hat{\theta}_n, \theta_0) > \varepsilon\} \leq e^{-cn}$$

for $n \geq n_0$. Moreover, there are numbers $n_0 \in N$, $c > 0$, $h_0 > 0$ fulfilling

$$(5.3) \qquad P(\theta_0)^n \{n^{1/2\alpha(\theta_0)} \Delta(\hat{\theta}_n, \theta_0) > h\} \leq \exp(-cn) + \exp(-ch^{2\alpha(\theta_0)})$$

for $n \geq n_0$, $h \geq h_0$. Furthermore, for arbitrary $s > 0$ there are $a > 0$, $C_s < \infty$ such that

$$(5.4) \qquad P(\theta_0)^n \{n^{1/2\alpha(\theta_0)} \Delta(\hat{\theta}_n, \theta_0) > a(s \log n)^{1/2\alpha(\theta_0)}\} \leq c_s n^{-s}$$

for $n \in N$.

Proof. See Section 7.8.

5.5. Remark. For "regular" families $\mathscr{P}$ we have $\alpha(\theta_0) = 1$; see our Examples 4.9, 4.11

and 6.6. However, there exist non-regular families where relations, similar to (5.3), hold with $\alpha(\theta_0) < 1$: Akahira and Takeuchi (1981, pages 27–51) consider the location family on $\mathbb{R}^1$ generated by the density

$$C_{u,v} \cdot (x-a)^{u-1} \cdot (b-x)^{v-1} g(x),$$

where $0 < u \leq v$, $a < b$, $g(x) > 0$ for $a < x < b$ and $g(x) = 0$ else. g is assumed to be sufficiently smooth. For $v < 2$ they prove the existence of an estimator-sequence $(\hat{\tau}_n)_{n \in N}$ fulfilling

$$P(\theta_0)^n \{n^{1/u} \Delta(\hat{\tau}_n, \theta_0) > h\} \leq \exp(-h^u) \tag{5.6}$$

for sufficiently large n and h, say $n \geq n_0$, $h \geq h_0$ (see Akahira and Takeuchi, 1981, pages 37–38). The $(\hat{\tau}_n)_{n \in N}$ are not maximum likelihood estimators. It is easy to see that there are numbers $C_1 > 0$, $C_2 > 0$, $\delta_0 > 0$ such that

$$C_1 \delta^v \leq d(P(\theta), P(\theta + \delta)) \leq C_2 \delta^u \tag{5.7}$$

for $0 \leq \delta \leq \delta_0$. Inequalities (5.6) and (5.7) imply

$$P(\theta_0)^n \{\sqrt{n}\, d(P(\hat{\tau}_n), P(\theta_0)) > h\} \leq \exp(-h^2) \tag{5.8}$$

for $n \geq n_0$, $h \geq h_0$. If a sequence $(\hat{\theta}_n)_{n \in N}$ of maximum likelihood estimators would satisfy (5.8), then relation (5.7) would imply

$$P(\theta_0)^n \{n^{1/v} \Delta(\hat{\theta}_n, \theta_0) > h\} \leq \exp(-h^{-v})$$

for $n \geq n_0$, $h \geq h_0$. This agrees with their result (Theorem 2.5.1) that $(\hat{\tau}_n)_{n \in N}$ has the maximal order of convergence $1/u$. It should be noted that for this family the behaviour of $(\hat{\theta}_n)_{n \in N}$ cannot be derived directly from Theorem 5.1 because our Condition 4.4 is not satisfied. Therefore, a separate proof would be needed for estimating $P(\theta_0)^n \{\sqrt{n}\; d(P(\hat{\theta}_n), P(\theta_0)) > h\}$.

6. Parametrized families with strong global properties. In the present section we give conditions which imply bounds Ch^{-b} for the probabilities in (5.3). This result cannot be derived from (3.4) and (3.13); a specific proof and specific conditions are needed. Conditions 6.1, 6.3, 6.4 are strong global properties of the family $\mathscr{P}$, Conditions 6.1 and 6.4 are global extensions of the local Conditions 3.7 and 3.8. The Conditions 6.1–6.4 refer to some $\theta_0 \in \Theta$.

6.1. Conditions. Θ is finite-dimensional with dimension $D_1 > 0$.

6.2. Condition. There are $\delta_4 > 0$ and $n_1 \in N$ such that

$$P(\theta_0)^{n_1} \{\sup_{\tau \in B_\Delta(\theta_0, \delta_4)} f_{n_1, \tau} = \infty\} = 0.$$

6.3. Condition. There are numbers $c_3 > 0$, $r > 0$, $\delta_5 > 0$ such that

$$c_3 \Delta(\tau, \theta_0)^r \leq a(P(\tau), P(\theta_0))^{-1}$$

for $\tau \in \Theta - B_\Delta(\theta_0, \delta_5)$.

6.4. Condition. There is $s \geq 2$, with $s > D_1/\alpha(\theta_0)$ for the constants D_1 in Condition 6.1 and $\alpha(\theta_0)$ in Condition 4.1, and there are numbers $t > 0$, $u > 0$, $C_1 < \infty$, $C_2 < \infty$ and there are functions $\theta \to \kappa(\theta) < \infty$, $\theta \in \Theta$, $\theta \to \varphi(\theta) > 0$, $\theta \in \Theta$, such that for every $\theta \in \Theta$:

(i) $\sup_{\tau \in B_\Delta(\theta, \varphi(\theta))} \int |\log h_\tau| \, dP(\theta_0) < \infty$,

(ii) $\int |\ell_{\tau_2} - \ell_{\tau_1}|^s \, dP(\theta_0) \leq \kappa(\theta) \Delta(\tau_1, \tau_2)^s$, for $\tau_1, \tau_2 \in B_\Delta(\theta, \varphi(\theta))$,

(iii) $\kappa(\theta) \leq C_1 a(P(\theta), P(\theta_0))^{-t}$,

(iv) $1/\varphi(\theta) \leq C_2 a(P(\theta), P(\theta_0))^{-u}$.

6.5. THEOREM. *Let $\theta_0 \in \Theta$ and let $(\hat{\theta}_n)_{n\in N}$ denote a sequence of quasi maximum likelihood estimators for θ_0. Assume that the Conditions* 4.1, 4.2 *hold true with $\alpha(\theta_0) \le 1$, and assume that Conditions* 6.1, 6.2, 6.3, 6.4 *are fulfilled for θ_0.*

Under these assumptions for arbitrary $b > 0$ there exist $h_0 > 0$, $C < \infty$ and $n_0 \in \mathbb{N}$ such that

$$P(\theta_0)^n\{n^{1/2\alpha(\theta_0)}\Delta(\hat{\theta}_n, \theta_0) > h\} \le Ch^{-b}$$

for $n \ge n_0$, $h \ge h_0$.

PROOF. See Section 7.9.

The assumptions of our Theorem 6.5 hold for location and scale parameter models fulfilling a few conditions:

6.6. EXAMPLE. (Families with location and scale parameters). Let h be a probability density relative to the Lebesgue measure λ^m on $\mathbb{R}^m$. We assume that h is bounded on $\mathbb{R}^m$, that $h(x) > 0$ for every x, and that h is twice continuously differentiable. Define $\ell(x) = \log h(x)$, and let $\ell^{(i)}$, $\ell^{(ij)}$ be the partial derivatives of ℓ, and let $x^{(i)}$ be the coordinates of x. Let $\mathscr{M}_m$ denote the symmetric and positive definite $m \times m$-matrices and define

$$\Theta = \{\theta = (a_\theta, A_\theta) : a_\theta \in \mathbb{R}^m, A_\theta \in \mathscr{M}_m\}.$$

Let $\| \ \|_e$ be the Euclidean norm on $\mathbb{R}^m$ and $\| \ \|_{sp}$ the spectral norm on $\mathscr{M}_m$, and let $I \in \mathscr{M}_m$ denote the unit matrix.

For $\theta = (a_\theta, A_\theta) \in \Theta$ let $P(\theta)$ be the probability measure with Lebesgue density

$$h_\theta(x) = (\det A_\theta)^{-1}h(A_\theta^{-1}(x - a_\theta)), \quad x \in \mathbb{R}^m.$$

In this way, we have defined a parametrized family $\mathscr{P} = \{P(\theta) : \theta \in \Theta\}$. We assume that this parametrization is injective.

Θ may be conceived as an open subset of $\mathbb{R}^{m(m+3)/2}$ by arranging the vector a_θ and the upper triangle of A_θ to a column in a unique manner.

We impose the following conditions on the density h:

(i) $\sup_{x\in\mathbb{R}^m} h(x) < \infty$.
(ii) For some $v > 0$, $\int \| x \|_e^v h(x)\lambda^m(dx) < \infty$.
(iii) For every $\theta \in \Theta$ the Fisher information matrix $I(\theta)$ is positive definite.
(iv) There is $s > m(m + 3)/2$ such that

$$\int | \ell^{(i)}(x) |^{\max(s,4)} h(x)\lambda^m(dx) < \infty,$$

$$\int | \ell^{(i)}(x)x^{(j)} |^{\max(s,4)} h(x)\lambda^m(dx) < \infty,$$

for $i, j = 1, \cdots, m$.

(v) There are $\delta > 0$, $C < \infty$ such that for every $\theta = (a_\theta, A_\theta) \in \Theta$ with $\| a_\theta \|_e < \delta$ and $\| A_\theta - I \|_{sp} < \delta$ we have

$$\int | \ell^{(i)}(x) | h_\theta(x)\lambda^m(dx) \le C \quad \text{for} \quad i = 1, \cdots, m.$$

(vi) $\int | \ell^{(ij)}(x) |^2 h(x)\lambda^m(dx) < \infty$, $\int | \ell^{(ij)}(x)x^{(k)} |^2 h(x)\lambda^m(dx) < \infty$, $\int | \ell^{(ij)}(x)x^{(k)}x^{(l)} |^2 h(x)\lambda^m(dx) < \infty$, for $i, j, k, l, = 1, \cdots, m$.

Under these assumptions, Conditions 4.1–4.4 and 6.1–6.4 are satisfied for every $\theta_0 \in \Theta$. The number $\alpha(\theta)$ in Condition 4.1 is equal to 1.

PROOF. See Section 7.10.

7. Proofs.

7.1. PROOF OF LEMMA 3.3. Let K be the compact set in Condition 3.1. It will be sufficient to show that the probabilities

$$P^{n'}\{\sup_{Q\in\mathscr{P}-K}(f_{n',Q}/f_{n',P}) > \gamma\} \quad \text{and} \quad P^{n'}\{\sup_{Q\in K-B(Q_0,\delta)}(f_{n',Q}/f_{n',P}) > \gamma\}$$

are bounded by $e^{-cn'}$ for every sufficiently large n', for every $Q_0 \in K - B(P, \varepsilon)$ and for some suitable $\delta > 0$ which may depend on Q_0; Cf. Wald's (1949) proof of Theorem 2.

Let n denote that number in Condition 3.1 and define

$$g_\kappa = \sup_{Q\in\mathscr{P}-K} f_{n+\kappa,Q}, \quad \kappa = 0, \cdots, n-1.$$

Note that the properties (i)–(iii) of $f_{n,Q}$ which are stated in Lemma 2.1 imply $P^{n'}$ – a.e.

$$\sup_{Q\in\mathscr{P}-K} f_{n',Q} = \sup_{Q\in(\mathscr{P}-K)\cap S} h_{n',Q}. \tag{7.1.1}$$

For $n' \geq n$ there is a unique representation $n' = m\, n + \kappa$, $m, \kappa \in N$, $\kappa < n$, and we obtain from (7.1.1) for $P^{n'}$ – a.e. $\mathbf{x} \in X^{n'}$

$$\sup_{Q\in\mathscr{P}-K} f_{n',Q}(\mathbf{x}) \leq \{\textstyle\prod_{\mu=0}^{m-2} g_0(x_{\mu n+1}, \cdots, x_{(\mu+1)n})\} g_\kappa(x_{(m-1)n+1}, \cdots, x_{n'}). \tag{7.1.2}$$

Furthermore, for $P^{n+\kappa}$ – a.e. $\mathbf{x} \in X^{n+\kappa}$ we have

$$g_{\kappa+1}(\mathbf{x})^{n+\kappa} \leq \textstyle\prod_{j=1}^{n+\kappa+1} \sup_{Q\in(\mathscr{P}-K)\cap S} \{\prod_{i=1}^{n+\kappa+1} h_{1,Q}(x_i)\}/h_{1,Q}(x_j) \tag{7.1.3}$$

which implies

$$\int (g_{\kappa+1}/f_{n+\kappa+1,P})^{(n+\kappa)/p(n+\kappa+1)}\, dP^{n+\kappa+1} \leq \int (g_\kappa/f_{n+\kappa,P})^{1/P}\, dP^{n+\kappa}, \quad \kappa = 0, \cdots, n-2, \tag{7.1.4}$$

and, therefore, we conclude from Condition 3.1 that

$$\alpha(\kappa) = \int g_\kappa^{1/p} f_{n+\kappa,P}^{1/q}\, d\nu^{n+\kappa} < 1, \quad \kappa = 0, \cdots, n-1, \tag{7.1.5}$$

with suitable chosen numbers $p > 1$, $q > 1$, $1/p + 1/q = 1$. Using (7.1.2) and (7.1.5) we get the inequalities

$$\begin{aligned} P^{n'}\{\sup_{Q\in\mathscr{P}-K} f_{n',Q} > \gamma\, f_{n',P}\} &\leq P^{n'}\{\mathbf{x} \in X^{n'}: (\textstyle\prod_{\mu=0}^{m-2} g_0(x_{\mu n+1}, \cdots, x_{(\mu+1)n})) \\ &\qquad \cdot g_\kappa(x_{(m-1)n+1}, \cdots, x_{n'})\, f_{n',P}(\mathbf{x})^{-1} > \gamma\} \\ &\leq \gamma^{-1/p}\alpha(0)^{m-1}\alpha(\kappa) \leq \gamma^{-1/p} \exp\left\{\frac{1}{2n} \log \alpha(0) n'\right\} \end{aligned} \tag{7.1.6}$$

for sufficiently large n'.

For every $Q \in K - B(P, \varepsilon)$ we conclude from Condition (3.2) that there is a ball $B(Q, \delta)$ such that (7.1.6) holds with $\mathscr{P} - K$ replaced by $B(Q, \delta)$.

Furthermore, Condition 3.1 and (7.1.3) imply that $\sup_{Q\in(\mathscr{P}-K)} f_{n',Q} < \infty$ $P^{n'}$ – a.e., for $n \leq n'$. Similarly, it follows from Condition 3.2 that for every sufficiently large n', $\sup_{Q\in K} f_{n',Q} < \infty$ $P^{n'}$ – a.e. These properties yield $P^{n'}\{\sup_{Q\in\mathscr{P}} f_{n',Q} = \infty\} = 0$. □

7.2. PROOF OF LEMMA 3.5. W.l.g. the numbers n in assumptions (iii) and (v) are equal. This can be seen from relations (7.1.3) and (7.1.4). Since $(\mathscr{P}, d)$ is separable and locally compact there is a sequence $K_m \subseteq \mathscr{P}$, $m \in N$, of compact sets having the properties: $P \in K_1$, $K_m \uparrow \mathscr{P}$ and K_m is contained in the interior of K_{m+1}. Then for the functions $g_m =$

$\sup_{Q\in\mathscr{P}-K_m} f_{n,Q}$ assumption (iii) implies P^n – a.e. $g_m \downarrow 0$ and from assumption (v) and from the monotone convergence theorem we conclude $\int g_m^{1/p} f_{n,P}^{1/q}\, d\nu^n \downarrow 0$. Thus there is a number $m \in \mathbb{N}$ satisfying $\int g_m^{1/p} f_{n,P}^{1/q}\, d\nu^n < 1$ and Condition 3.1 is proved.

For every $Q_0 \in K_m - \{P\}$ we define a sequence $g_k = \sup_{Q\in B(Q_0,1/k)} f_{n,Q}$, $k \in \mathbb{N}$. Then (ii) implies $g_k \downarrow f_{n,Q_0} P^n$ – a.e. and (iv) implies $\int g_k^{1/p} f_{n,P}^{1/q}\, d\nu^n \downarrow \int f_{n,Q_0}^{1/p} f_{n,P}^{1/q}\, d\nu^n < 1$. Thus Condition 3.2 is proved. □

Two lemmas will be needed in the proof of Theorem 3.10. First, we prove a fluctuation inequality for separable random functions. Before formulating this lemma we give a few definitions:

7.3. DEFINITION. Let $(\Omega, \mathscr{C}, P)$ be a probability space and let (T, ρ) be a metric space. A collection of $\mathscr{C}$-measurable functions $\xi_t: \Omega \to \mathbb{R}$, $t \in T$, is called random function on $(\Omega, \mathscr{C}, P)$ with parameter set T. The random function $(\xi_t)_{t\in T}$ is called separable if there exists a countable subset $T_0 \subseteq T$ such that for every $t \in T$ there is a sequence $(t_m)_{m\in\mathbb{N}} \subseteq T_0$ satisfying $\lim_{m\in\mathbb{N}} \rho(t_m, t) = 0$ and $\lim_{m\in\mathbb{N}} \xi_{t_m}(\omega) = \xi_t(\omega)$ for every $\omega \in \Omega$. The set T_0 is called separant of the random function. Clearly, every separant is a dense subset of T. Two random functions $(\xi_t)_{t\in T}$ and $(\chi_t)_{t\in T}$ are called equivalent if $\xi_t = \chi_t\, P$ – a.e. for every $t \in T$.

7.4. LEMMA. (Fluctuation inequality). *Let (T, ρ) be a metric space and assume that $t_0 \in T$ and $\varepsilon > 0$ are such that the ball $B_\rho(t_0, \varepsilon)$ is finite-dimensional with dimension $D > 0$ (cf. Definition 2.5). Assume that $(\xi_t)_{t\in T}$ is a separable random function on $(\Omega, \mathscr{C}, P)$ satisfying*

$$\int |\xi_{t_2}(\omega) - \xi_{t_1}(\omega)|^s P(d\omega) \le L\, \rho(t_1, t_2)^k \tag{7.4.1}$$

for every $t_1, t_2 \in B_\rho(t_0, \varepsilon)$, where $k > D$, $s > 0$ and $L < \infty$ are suitable constants. Then there is a constant $K < \infty$ such that

$$P\{\omega \in \Omega: \sup_{t_1,t_2\in B_\rho(t_0,\varepsilon)} |\xi_{t_2}(\omega) - \xi_{t_1}(\omega)| > \delta\} \le K\, L \varepsilon^k \delta^{-s}. \tag{7.4.2}$$

K depends only on k, s, D and on the constant C occurring in Definition 2.4; it is independent from ε, δ and L.

Lemma 7.4 is an improvement of a result of Rao (1975, page 358, Theorem 3.2). We do not apply Rao's theorem directly, since Rao's assumptions would require that the random function has continuous paths and that (7.4.1) is satisfied for $k > 4\,D$ (in place of $k > D$). Forerunners of Lemma (7.4) are the results of Neveu (1965, Proposition III. 5.3.) and of Gihman and Skorohod (1974, page 91, Lemma 1, and page 192, the remark below Theorem 6) which refer to parameter sets T being real intervals, and the result of Pfaff (1977, page 145, Lemma 6.1) where the parameter set T is the unit cube in the k-dimensional Euclidean space.

PROOF OF LEMMA 7.4. We give only a sketch because the proof is similar to the proofs of Rao (1975) and Pfaff (1977).

We define the integer number $n(\varepsilon)$ by $2^{-n(\varepsilon)-1} < \varepsilon \le 2^{-n(\varepsilon)}$. Employing the dimensionality condition, for $n \ge n(\varepsilon)$ we choose $B_\rho(t_{n,1}, 2^{-n}), \cdots, B_\rho(t_{n,m(n)}, 2^{-n})$ covering $B = B_\rho(t_0, \varepsilon)$ such that $U_n = \{t_{n,1}, \cdots, t_{n,m(n)}\} \subseteq B$, $m(n) \le C\, 2^{nD} \varepsilon^D$, $U_{n(\varepsilon)} = \{t_0\}$ and $m(n(\varepsilon)) = 1$. The set $U = \bigcup_{n\ge n(\varepsilon)} U_n$ is dense in B and from relation (7.4.1) we conclude that the random function $(\xi_t)_{t\in B}$ is uniformly P-continuous. Hence there exists an equivalent random function $(\chi_t)_{t\in B}$ which is separable with separant U (see Strasser, 1981a, page 22, Definition 3.1 and Theorem 3.2). Relation (7.4.1) is equally satisfied for $(\chi_t)_{t\in B}$ and it will be sufficient to prove relation (7.4.2) for χ instead of ξ. Proceeding similarly as Rao (1975) and Pfaff

(1977) we obtain

$$
\begin{aligned}
&P\{\omega \in \Omega\colon \sup_{t',t'' \in U} |\chi_{t''}(\omega) - \chi_{t'}(\omega)| > \delta\} \\
&\qquad \le \textstyle\sum_{n \ge n(\varepsilon)} \sum_{t \in U_{n+1}} P\{\omega \in \Omega : |\chi_{\rho_n^{(t)}}(\omega) - \chi_t(\omega)| \\
&\qquad\quad > 2\,\delta(2^{n(\varepsilon)}/2^{n+1})^{(k-d)/2s}(1 - 2^{-(k-D)/2s})^{-1}\} \\
&\qquad \le K\, L\varepsilon_k \delta^{-s},
\end{aligned}
$$

where $\rho_n : U_{n+1} \to U_n$ are maps satisfying $\rho(t, \rho_n(t)) < 2^{-n}$ for every $t \in U_{n+1}$ and for every $n \ge n(\varepsilon)$. □

The next lemma contains well known properties of the Hellinger distance and of the affinity of probability measures.

7.5. LEMMA. *Let P, Q be probability measures on $(\Omega, \mathscr{C})$ being dominated by σ-finite $\mu \mid \mathscr{C}$.*

(i)
$$d(P, Q)^2 = \frac{1}{2}\int [h_Q^{1/2} - h_P^{1/2}]^2\, d\mu = 1 - a(P, Q) \quad \textit{and}$$

$$a(P, Q) = \int h_P^{1/2}\, h_Q^{1/2}\, d\mu = \int h_Q^{1/2}\, h_P^{-1/2}\, dP,$$

where h_P and h_Q are densities of $P \mid \mathscr{C}$, respectively $Q \mid \mathscr{C}$, relative to $\mu \mid \mathscr{C}$. $d(P, Q)$ and $a(P, Q)$ are independent from the choice of μ.
(ii) *For $n \in \mathbb{N}$ we have $a(P^n, Q^n) = a(P, Q)^n$ and*

$$1 - d(P^n, Q^n)^2 \le \exp\{-\, n\, d(P, Q)^2\}.$$

PROOF. See LeCam (1973, Section 1 and proof of Lemma 1 in Section 2).

7.6. PROOF OF THEOREM 3.10. W.l.g. $\delta = \delta' = \delta''$. Let $n \in \mathbb{N}$ and $2 \le h \le n^{1/2}\delta - 1$. We have

$$
\begin{aligned}
(7.6.1)\quad P^n\{n^{1/2} d(\hat{P}_n, P) > h\} &\le P^n\{d(\hat{P}_n, P) \ge \delta - n^{-1/2}\} \\
&\quad + \textstyle\sum_{j=0}^{[n^{1/2}\delta - 1 - h]} P^n\{h + j \le n^{1/2} d(\hat{P}_n, P) < h + j + 1\}.
\end{aligned}
$$

The first probability on the right hand side can be estimated using Lemma 3.3. For estimating the remaining probabilities it will be sufficient, in the view of Wald's (1949) proof, to show that with suitable chosen constants $c > 0$, $C > 0$

$$(7.6.2)\qquad P^n\{\sup_{Q \in G(\lambda, n)} (f_{n,Q}/f_{n,P}) > \tfrac{1}{2}\gamma\} \le C\, e^{-c\lambda^2}$$

holds for every $n \in \mathbb{N}$ and every $\lambda \in [2, n^{1/2}\delta - 1]$, where

$$G(\lambda, n) = \{Q \in \mathscr{P} : \lambda \le n^{1/2} d(P, Q) < \lambda + 1\}.$$

Choose β with $0 < \beta < 1/D$, and define $\varepsilon = n^{-1/2} e^{-\beta\lambda^2}$. Since $G(\lambda, n) \subseteq B(P, \delta) \cap B(P, n^{-1/2}(\lambda + 1))$, $G(\lambda, n)$ can be covered by balls $B(Q_1, \varepsilon/2), \cdots, B(Q_m, \varepsilon/2)$ with $Q_1, \cdots, Q_m \in B(P, \delta)$ and $m \le C\, n^{-D/2} \lambda^D \varepsilon^{-D}$. W.l.g. for $\mu = 1, \cdots, m$ $B(Q_\mu, \varepsilon/2) \cap G(\lambda, n) \ne \varnothing$ and there is $P_\mu \in B(Q_\mu, \varepsilon/2) \cap G(\lambda, n)$ satisfying

$$\int \log h_{P_\mu} dP + (1/2n) \log 2 > \sup_{Q \in B(Q_\mu, \varepsilon/2) \cap G(\lambda, n)} \int \log h_Q\, dP.$$

This choice of P_μ is possible since (3.8) implies that the supremum is finite. Note that the sets $B_\mu = B(P_\mu, \varepsilon) \cap B(P, \delta)$ $\mu = 1, \cdots, m$, cover $G(\lambda, n)$. One easily verifies

(7.6.3) $$P^n\{\sup_{Q\in G(\lambda,n)} f_{n,Q}/f_{n,P} > \frac{1}{2}\gamma\} \le \sum_{\mu=1}^m P^n\{f_{n,P_\mu}^{1/2}/f_{n,P}^{1/2} > (\gamma/4)^{1/2}\} + \sum_{\mu=1}^m P^n\{\sup_{Q\in B_\mu\cap S}[\ell_{n,Q} - \ell_{n,P_\mu}] > (1/2)\log 2\},$$

where $\ell_{n,Q}(\mathbf{x}) = \log f_{n,Q}(\mathbf{x}) - n\int \log h_Q\,dP$.

Applying Lemma (7.5) we obtain

(7.6.4) $$P^n\{f_{n,P_\mu}^{1/2}/f_{n,P}^{1/2} > (\gamma/4)^{1/2}\} \le (4/\gamma)^{1/2} a(P_\mu^n, P^n) \le (4/\gamma)^{1/2} e^{-nd(P_\mu,P)^2} \le (4/\gamma)^{1/2} e^{-\lambda^2}.$$

The random function $(\ell_{n,Q})_{Q\in B(P,\delta)}$ is uniformly P^n-continuous by Condition 3.8. This implies that there is an equivalent random function $(\zeta_Q)_{Q\in B(P,\delta)}$ which is separable with separant $B(P, \delta) \cap S$ (cf. Strasser, 1981a, Theorem 3.2). We apply Lemma 7.4 to the random function $(\zeta_Q)_{Q\in B(P,\delta)}$, to the metric space $(T, \rho) = (B(P, \delta), d)$ and to the ball B_μ. Condition 3.8 and the inequality for absolute moments of order s for sums of i.i.d. variables (see Dharmadhikari and Jogdeo, 1969, page 1507, Theorem 2) yield

$$\int |\zeta_{Q_2} - \zeta_{Q_1}|^s\,dP^n = \int |\ell_{n,Q_2} - \ell_{n,Q_1}|^s\,dP^n \le Cn^{s/2}\,d(Q_1, Q_2)^s$$

for every pair $Q_1, Q_2 \in B(P, \delta)$. Thus relation (7.4.1) is valid for the random function $(\zeta_Q)_{Q\in B(P,\delta)}$ and we get the result

$$P^n\{\sup_{Q_1,Q_2\in B_\mu\cap S} |\ell_{n,Q_2} - \ell_{n,Q_1}| > (½)\log 2\} \le P^n\{\sup_{Q_1,Q_2\in B_\mu} |\zeta_{Q_2} - \zeta_{Q_1}| > (½)\log 2\} \le Cn^{s/2}\varepsilon^s.$$

Combining this relation and (7.6.3) and (7.6.4) and using the bound for m we obtain (7.6.2).

This result, Lemma 3.3 and (7.6.1) prove that

$$P^n\{n^{1/2}\,d(\hat{P}_n, P) > h\} \le \sum_{j=0}^\infty C_1\,e^{-c_1(h+j)^2} + e^{-c_2 n} \le C_2 e^{-c_1 h^2} + e^{-c_2 n}$$

for every $n \ge n_0$ and every $h \ge 2$. Easy computations show that this relation yields the assertion (ii) of Theorem 3.10. Assertion (iii) can easily be verified. □

7.7. Proof of Lemma 4.5. Since the parametrization is injective we have $d(P(\sigma), P(\tau)) > 0$ for $\sigma \ne \tau$. Furthermore, Condition 4.1 implies continuity of the maps $\sigma \to P(\sigma)$ and $\sigma \to d(P(\sigma), P(\tau))$, for every fixed $\tau \in \Theta$. These properties and Condition 4.2 imply that for every $\delta > 0$ and every $\theta \in \Theta$ there is a number $\eta > 0$ satisfying

(7.7.1) $$d(P(\sigma), P(\theta)) \ge \eta \quad \text{for} \quad \sigma \in \Theta - B_\Delta(\theta, \delta).$$

For every $\tau \in \Theta$ and every sequence$(\tau_n)_{n\in N} \subseteq \Theta$ we conclude from Condition 4.1 and (7.7.1) that $\lim_{n\in N} d(P(\tau_n), P(\tau)) = 0$ implies $\lim_{n\in N} \Delta(\tau_n, \tau) = 0$. Thus the map $\tau \to P(\tau)$ is a homeomorphism of the topologies generated by the metrics d and Δ.

Conditions 4.1–4.3 imply that there is $\delta > 0$ such that $B(P(\theta_0), \delta)$ is finite-dimensional with dimension $D/\alpha(\theta_0)$. This means that Condition 3.7 is fulfilled for the measure $P(\theta_0)$. Furthermore, Condition (4.4) implies the validity of Conditon 3.8 for $P(\theta_0)$. □

7.8. Proof of Theorem 5.1. The validity of relations (3.4) and (3.13) follows from Corollary 4.8. From (3.4), Condition 4.1, (7.7.1) we obtain for arbitrary $\varepsilon > 0$ and for $\alpha = \alpha(\theta_0)$, $\delta = \delta(\theta_0)$, $c = c_1(\theta_0)$

$$\begin{aligned} P(\theta_0)^n\{\Delta(\sigma_n, \theta_0) > \varepsilon\} &\le P(\theta_0)^n\{\Delta(\sigma_n, \theta_0) \ge \delta\} + P(\theta_0)^n\{\Delta(\sigma_n, \theta_0) > \varepsilon \wedge \Delta(\sigma_n, \theta_0) < \delta\} \\ &\le P(\theta_0)^n\{d(P(\sigma_n), P(\theta_0)) \ge \eta\} + P(\theta_0)^n\{d(P(\theta_n), P(\theta_0)) > c\,\varepsilon^\alpha\} \\ &\le 2e^{-cn}. \end{aligned}$$

This proves (5.2).

Furthermore, inequality (3.13) yields for sufficiently large n and h

$$\begin{aligned} P(\theta_0)^n\{n^{1/2\alpha}\Delta(\sigma_n, \theta_0) > h\} &\le P(\theta_0)^n\{\Delta(\sigma_n, \theta_0) \ge \delta\} \\ &\quad + P(\theta_0)^n\{n^{1/2\alpha}\Delta(\sigma_n, \theta_0) > h \wedge \Delta(\sigma_n, \theta_0) < \delta\} \\ &\le P(\theta_0)^n\{d(P(\sigma_n), P(\theta_0)) \ge \eta\} \\ &\quad + P(\theta_0)^n\{n^{1/2}\, d(P(\sigma_n), P(\theta_0)) > ch^\alpha\} \\ &\le e^{-cn} + \exp(-ch^{2\alpha}). \end{aligned}$$

This proves (5.3), and substituting $h = c^{-1/2\alpha}(s \log n)^{1/2}$ we obtain the last assertion (5.4). □

7.9. Proof of Theorem 6.5. Conditions 6.1 and 6.4 imply Conditons 4.3 and 4.4 and, therefore, we conclude from the proof of Lemma 4.5 (see Section 7.7) that Conditions 3.7 and 3.8 are fulfilled. Under these conditions the proof of Theorem 3.10 (see Section 7.6) yields, for $F_n = \{\sup_{\tau\in\Theta} f_{n,\tau} = \infty\}$ and for suitable numbers $\delta_0 > 0$, $c_0 > 0$,

$$(7.9.1)\qquad P(\theta_0)^n(\{h \le n^{1/2}\, d(P(\hat\theta_n), P(\theta_0)) \le n^{1/2}\delta_0\} - F_n) \le \exp\{-c_0h^2\} \quad \text{for } n \in \mathbb{N},\ h \ge h_0.$$

Choose $\delta > 0$ such that $\delta \le \delta_1(\theta_0)$ and $\delta \le \delta_4$ and $c_2(\theta_0)\delta^{\alpha(\theta_0)} \le \delta_0$, where $c_2(\theta_0)$, $\alpha(\theta_0)$, $\delta_1(\theta_0)$, δ_4 are the constants in Condition 4.1, respectively in Condition 6.2. W.l.g. the constant δ_5 in Condition 6.3 is less than δ. There is $\eta > 0$ such that

$$(7.9.2)\qquad d(P(\tau), P(\theta_0)) \ge \eta \quad \text{for} \quad \Delta(\tau, \theta_0) \ge \delta$$

(see (7.9.1.)]

Let $\alpha = \alpha(\theta_0)$, $P_0 = P(\theta_0)$, $\hat P_n = P(\hat\theta_n)$. Applying Conditions 4.1, 6.3 and (7.9.1) we get

$$(7.9.3)\qquad \begin{aligned} &P_0^n(\{n^{1/2\alpha}\Delta(\hat\theta_n, \theta_0) > h\} - F_n) \\ &\le P_0^n(\{n^{1/2\alpha}a(\hat P_n, P_0)^{-1/r} > c_3^{1/r}h \wedge \Delta(\hat\theta_n, \theta_0) \ge \delta\} - F_n) \\ &\quad + P_0^n(\{c_1(\theta_0)h^\alpha < n^{1/2}\, d(\hat P_n, P_0) \le n^{1/2}\delta_0\} - F_n) \\ &\le \textstyle\sum_{j=0}^\infty P_0^n\{\sup_{\tau\in G(c_3^{1/r}h+j,n)} f_{n,\tau}/f_{n,\theta_0} > \gamma/2\} + \exp(-c_0h^2) \end{aligned}$$

for $n \in \mathbb{N}$, $h \ge h_0$, where

$$G(\lambda, n) = \{\tau \in \Theta : \lambda \le n^{1/2\alpha}a(P(\tau), P_0)^{-1/r} \le \lambda + 1 \wedge \Delta(\tau, \theta_0) \ge \delta\}.$$

For any given $b > 0$ we choose $b' \ge ru$ fulfilling $(s - D)b' \ge b + rt + D$ for the numbers r, s, t, u, D in Conditions 6.1, 6.3, 6.4 (note that $s > D/\alpha(\theta_0) \ge D$ since $\alpha(\theta_0) \le 1$), and define $\varepsilon = \varepsilon(\lambda, n) = C_2^{-1}n^{-1/2\alpha}(\lambda + 1)^{-b'}$, where C_2 is the constant in Condition 6.4.

From Condition 6.3 we conclude $G(\lambda, n) \subseteq B_\Delta(\theta_0, c_3^{-1/r}n^{-1/2}(\lambda + 1))$. Hence, there are balls $B_\Delta(\tau_\mu, \varepsilon/2)$, $\mu = 1, \cdots, m$, covering $G(\lambda, n)$ where $\tau_\mu \in G(\lambda, n)$ and where $m \le K_1n^{-D/2\alpha}\lambda^D\varepsilon^{-D}$ for some suitable constant K_1. Proceeding similarly as in the proof of Theorem 3.10 (see the inequalities (7.6.3)) we obtain for suitable chosen $\sigma_\mu \in B_\Delta(\tau_\mu, \varepsilon) \cap G(\lambda, n)$

$$(7.9.4)\qquad \begin{aligned} &P_0^n\{\sup_{\tau\in G(\lambda,n)} f_{n,\tau}/f_{n,\theta_0} > \gamma/2\} \le K_2 \textstyle\sum_{\mu=1}^m a(P(\sigma_\mu)^n, P_0^n) \\ &\quad + \textstyle\sum_{\mu=1}^m P_0^n\{\sup_{\tau\in B_\Delta(\sigma_\mu,\varepsilon)} |\ell_{n,\tau} - \ell_{n,\theta_0}| > (½)\log 2\} \\ &\le m(K_3\lambda^{-b''} + K_4n^{s/2}(n^{-1/2\alpha}\lambda)^{rt}\varepsilon^s) \le K_5\lambda^{-b}. \end{aligned}$$

In (7.9.4) the last-but-one inequality follows from

$$a(P(\sigma_\mu)^n, P_0^n) = \exp\{-n \log(1/a(P(\sigma_\mu), P_0))\} \le K_3\lambda^{-b''}$$

for $n \geq n_0$ and for $b'' = b + (b' + 1)D$ (note that (7.9.2) yields $d(P(\sigma_\mu), P_0) \geq \eta$ and $a(P(\sigma_\mu), P_0)^{-1} \geq 1/(1 - \eta^2)$) and from an application of Lemma 7.4 to the metric space (Θ, Δ) and to the ball $B_\Delta(\sigma_\mu, \varepsilon)$. Note that Condition 6.4 implies $\varphi(\sigma_\mu) \geq \eta$ for $n \in N$, $\lambda > 0$, and that Condition 6.4 yields the validity of assumption (7.4.1), namely,

$$\int | \ell_{n,\sigma} - \ell_{n,\tau} |^s \, dP_0^n \leq K_6 n^{s/2} \kappa(\sigma_\mu) \Delta(\sigma, \tau)^s \leq K_7 n^{s/2} (n^{-1/2\alpha}\lambda)^{rt}$$

for $\sigma, \tau \in B_\Delta(\sigma_\mu, \varepsilon)$; cf. the proof of Theorem 3.10. The last inequality in (7.9.4) follows from the bound for m and from the definitions of ε, b', b''.

At least, we prove $P_0^n(F_n) = 0$ for $n \geq n_1$. If this is true then the assertion of the theorem follows from (7.9.3) and (7.9.4). Proceeding similarly as before we get for $\gamma > 1$

$$\begin{aligned} P_0^n\{\sup_{\tau \in \Theta - B_\Delta(\theta_0, \delta)} (f_{n,\tau}/f_{n,\theta_0}) > \gamma\} &\leq P_0^n \cup_{j=0}^\infty \{\sup_{\tau \in G(c_3^{1/r}\delta + j, n)} (f_{n,\tau}/f_{n,\theta_0}) > \gamma\} \\ &\leq K_8(\gamma^{-1/4} + (\log \gamma)^{-s})\delta^{-b} \quad \text{for} \quad n \geq n_0. \end{aligned}$$

Let $\gamma \to \infty$ and keep n fixed. Hence for $n \geq n_0$

$$\sup_{\tau \in \Theta - B_\Delta(\theta_0, \delta)} f_{n,\tau} \text{ is finite } P_0^n - \text{a.e.}$$

This property and Condition 6.2 imply for $n \geq \max(n_0, n_1) P_0^n(F_n) = 0$ because δ was chosen less then δ_4. □

7.10. Proof for Example 6.6. In place of a complete proof we give only some hints. In Example 4.9 we have shown that Condition 4.1 is valid. (The present example is a specialization of Example 4.9). Conditions 4.2, 4.4 are implicated in Conditions 6.3, 6.4. The dimension D in Condition 4.3 and Condition 6.1 is equal to $m(m + 3)/2$. Condition 6.2 is obviously true.

The main problem is the verification of Conditions 6.3 and 6.4. For this purpose we use the auxiliary function

$$\psi(\theta) = \max(1 + \| a_\theta \|_e, 1 + \Lambda_\theta, 1/\lambda_\theta),$$

$\theta = (a_\theta, A_\theta) \in \Theta$, where Λ_θ, λ_θ are the largest and the smallest eigenvalue of A_θ. A clever estimation shows that with some suitable constant $C < \infty$

$$a(P(\tau), P(\theta)) \leq C\psi(\theta)^{-v/4} \tag{7.10.1}$$

for $\theta, \tau \in \Theta$, where v is the constant in condition (ii). Relation (7.10.1) implies Condition 6.3. The moments in Condition 6.4 (i), (ii) are estimated with the aid of (4.10). Assumption (iv) is used to estimate the right hand side in (4.10). We choose $\varphi(\theta) = c\lambda_\theta$ with some suitable constant $c > 0$; then (6.4) (ii) is satisfied with $\kappa(\theta) \leq C\psi(\theta)^b$ for suitable $C < \infty$, $b > 0$. Thus Condition 6.4 (iii), (iv) follow from (7.10.1). □

REFERENCES

AKAHIRA, M., and TAKEUCHI, K. (1981). Asymptotic efficiency of statistical estimators: Concepts and higher order asymptotic efficiency. *Lecture Notes in Statistics* **7**. Springer, Berlin.

CHIBISOV, D. M. (1973). An asymptotic expansion for a class of estimators containing maximum likelihood estimators. *Theor. Probability Appl.* **18** 295–303.

DHARMADHIKARI, S. W., and JOGDEO, J. (1969). Bounds on moments of certain random variables. *Ann. Math. Statist.* **40** 1506–1509.

GIHMAN, I. I., and SKOROHOD, A. V. (1974). *The Theory of Stochastic Processes I*. Springer, Berlin.

GUSEV, S. I. (1976). Asymptotic expansions associated with some statistical estimators in the smooth case. II. Expansions of moments and distributions. *Theor. Probability Appl.* **21** 14–33.

IBRAGIMOV, I. A., and KHAS'MINSKII, R. Z. (1972). Asymptotic behavior of statistical estimators in the smooth case I. Study of the likelihood ratio. *Theor. Probability Appl.* **17** 445–462.

IBRAGIMOV, I. A., and KHAS'MINSKII, R. Z. (1973). On moments of generalized Bayesian estimators and maximum likelihood estimators. *Theor. Probability Appl.* **18** 508–520.

KIEFER, J., and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887-906.

LANDERS, D. (1968). *Existenz und Konsistenz von Maximum Likelihood Schätzern.* Inaugural-Dissertation, University of Cologne, West Germany.

LANDERS, D. (1972). Existence and consistency of modified minimum constrast estimates. *Ann. Math. Statist.* **43** 74–83.

LECAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53.

MICHEL, R., and PFANZAGL, J. (1971). The accuracy of the normal approximation for minimum contrast estimates. *Z. Wahrsch. verw. Gebiete* **18** 73–84.

NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.

PFAFF, TH. (1977). *Existenz und asymptotische Entwicklungen der Momente mehrdimensionaler maximum likelihood-Schätzer.* Inaugural-Dissertation, University of Cologne, West Germany.

PFANZAGL, J. (1969). On the measurability and consistency of minimum contrast estimates. *Metrika* **14** 249–272.

RAO, B. L. S. PRAKASA (1975). Tightness of probability measures generated by stochastic processes on metric spaces. *Bull. Inst. Math. Acad. Sinica* **3** 353–367.

ROUSSAS, G. G. (1965). Extension to Markov processes of a result by A. Wald about the consistency of maximum likelihood estimates. *Z. Wahrsch. verw. Gebiete* **4** 69–73.

STRASSER, H. (1981a). Consistency of maximum likelihood and Bayes estimates. *Ann. Statist.* **9** 1107–1113.

STRASSER, H. (1981b). Convergence of estimates. Part I. *J. Multivariate Anal.* **11** 127–151.

STRASSER, H. (1981c). Convergence of estimates. Part II. *J. Multivariate Anal.* **11** 152–172.

WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595–601.

REGIONALES HOCHSCHULRECHENZENTRUM
UNIVERSITÄT KAISERSLAUTERN
POSTFACH 30 49
D 6750 KAISERSLAUTERN
FEDERAL REPUBLIC OF GERMANY