HANS RUDOLF LERCHE

Fluctuations of Posterior Distributions of Quadratically
Differentiable Experiments

Fluctuations of Posterior Distributions of Quadratically Differentiable Experiments^{*}

Hans Rudolf Lerche

Institut für Angewandte Mathematik der Universität Heidelberg,
Im Neuenheimer Feld 294, D-6900 Heidelberg, Federal Republic of Germany

Summary. The law of the iterated logarithm for posterior distributions and related Bayes estimators is studied for models which are differentiable in quadratic mean at the true parameter.

1. Introduction and Summary

The asymptotic behaviour of posterior distributions and Bayes estimators, is extensively studied in the literature for the case of independent, identically distributed random variables. Consistency theorems assert the convergence of the posterior distributions and Bayes estimators to the true parameter (cf. [10]), while Bernstein-von Mises theorems describe the asymptotic normality of the posterior distributions around the true parameter (cf. [2, 4, 8]). In this paper we study the law of the iterated logarithm (LIL), which describes the almost sure fluctuations of the posterior distributions around the true parameter. Let $F_{x,n}$ be the posterior distribution given the first n observations, θ_0 the true parameter, and $B_\theta(a)$ the closed ball centered at θ with radius a in $\mathbb{R}^k$. Then the LIL has the following form

$$\limsup_{n \rightarrow \infty} F_{x,n} \left(B_{\theta_0}^c \left(d \sqrt{\frac{\log \log n}{n}} \right) \right) = \begin{cases} 0 & \text{if } d > \sqrt{2} \\ 1 & \text{if } d \leq \sqrt{2} \end{cases} \quad (1.1)$$

P_{θ_0} -almost surely. This is easy to prove in the case of the normal shift family. In the paper [7] we proved (1.1) for exponential families by approximating the posterior by normal distributions. This technique works also for models which satisfy the Cramér conditions. (A modification of the Bernstein-von Mises theorem in [2], Theorem 7, yields (1.1).)

The main part of this paper deals with the study of (1.1) for models which are differentiable in quadratic mean around the true parameter. This is a more

^{*} Partially supported by the Deutsche Forschungsgemeinschaft

general situation than that described by the Cramér conditions and also more natural from the statistical point of view (cf. [4], p. 100). The main result of the paper, Theorem 4.1, states that, in the independent identically distributed case, under the condition of differentiability in quadratic mean (more precisely, under conditions (1)–(6) of Sect. 2)

$$\lim_{n \rightarrow \infty} F_{x,n} \left(B_{\theta_0}^c \left(\gamma_0 \sqrt{\frac{\log \log n}{n}} \right) \right) = 0 \quad (1.2)$$

P_{θ_0} -almost surely, for some positive constant γ_0 . This is the half of the LIL which sharpens consistency. Related results hold also for the posterior mean and the maximum probability estimator (Theorem 4.6 and 4.7).

We do not know whether the other half of the LIL holds:

$$\limsup_{n \rightarrow \infty} F_{x,n} \left(B_{\theta_0}^c \left(\gamma_1 \sqrt{\frac{\log \log n}{n}} \right) \right) = 1 \quad (1.3)$$

P_{θ_0} -almost surely, under the assumptions of (1.2). In the case of the Cramér-conditions (1.2) and (1.3) can be deduced from an almost-sure Bernstein-von Mises theorem. This is not so in the present case: the assumptions of (1.2) yield only convergence in probability (cf. [4], p. 131). Under somewhat stronger conditions, fulfilled, for instance, by the double exponential distributions, we can deduce (1.3) for one-dimensional parameter spaces from an almost sure Bernstein-von Mises theorem as in the Cramér-condition case. This will be published elsewhere.

The plan of the present paper is as follows. Section 2 contains notations and assumptions. In Sect. 3 we derive consistency results, in part (a) for probability priors and in part (b) for σ -finite priors. In Sect. 4 the laws of the iterated logarithm are proved; in part (a) for posterior distributions, in part (b) for Bayes estimators.

The main technical tools used in the paper are those of Le Cam and Schwartz (cf. [6] and [10]). Several methods and ideas of Le Cam, which are contained in the unpublished manuscript [5], are further developed and applied. One might conjecture that one can obtain the results by applying the methods of Dudley and Kuelbs for deriving functional LIL-theorems. But on closer inspection this appears far from obvious.

The results of this paper have implications for the theory of tests with power one, developed by Darling, Robbins, Siegmund et al. (cf. [9]). The laws of the iterated logarithm of Sect. 4 enable us to construct tests with power one in the independent identically distributed case for nearly any reasonable model. To explain the construction, let us consider the testing problem $H_0: \theta = \theta_0$ against $H_1: \theta \neq \theta_0$, where Θ is an open subset of $\mathbb{R}^k$. Recall that a level α -test with power one, $0 < \alpha < 1$, is given by a stopping time T for the sequence $x_1, \dots, x_n, \dots$ with values in the natural numbers, which satisfies

- (i) $P_{\theta_0}(T < \infty) \leq \alpha$,
- (ii) $P_{\theta}(T < \infty) = 1$ for $\theta \neq \theta_0$.

We assume that the conditions (1)–(6) of Sect. 2 are fulfilled. Hence statement (1.2) holds. Let ψ be a positive increasing function $\psi: \mathbb{N} \rightarrow \mathbb{R}$ with $\gamma_0^2 \log \log n \leq \psi(n) = o(n)$ as $n \rightarrow \infty$, where $\gamma_0 = e(k+1)(1+\delta)/(2c_0^2)$ with $\delta > 0$ and c_0 the constant of Lemma 4.2. For $\varepsilon > 0$

$$T = \inf \left\{ n \geq 1 \mid F_{x,n} \left(B_{\theta_0}^c \left(\sqrt{\frac{\psi(n)}{n}} \right) \right) > 1 - \varepsilon \right\}$$

defines a test with power one. Theorem 4.1 guarantees that condition (i) can be satisfied and the consistency Theorem 3.6 implies (ii). The level of these tests can be computed approximately using an invariance principle for the posterior distributions. This result states that the posterior distributions, considered as a sequential observation process, properly rescaled, converge to a “Brownian motion in the space of measures”; that is, the process of posterior distributions for the Brownian motion with unknown drift and with Lebesgue prior. The details will be studied in a subsequent paper.

2. Notations and Assumptions

Let $\mathcal{E} = \{P_\theta; \theta \in \Theta\}$ be a set of probability measures defined on a measurable space $(\mathcal{X}, \mathcal{A})$ where Θ is a Borel subset of $\mathbb{R}^k$. As usual $\mathcal{E}$ is called a *model* or *experiment*. Let $\mathcal{B}$ denote the Borel σ -algebra of Θ . Let ν be a σ -finite measure which dominates the experiment $\mathcal{E}$. We assume that densities $f_\theta(x) = \frac{dP_\theta}{d\nu}(x)$ exist which are measurable with respect to $\mathcal{A} \otimes \mathcal{B}$. Let μ be a σ -finite measure on Θ . The *posterior distribution* of $\mathcal{E}$ with respect to μ , given the first n observations $z_n = (x_1, \dots, x_n) \in \mathcal{X}^n$, is

$$F_{z_n}(B) = \frac{\int \prod_{i=1}^n f_\theta(x_i) \mu(d\theta)}{\int \prod_{i=1}^n f_\theta(x_i) \mu(d\theta)}, \quad B \in \mathcal{B}.$$

For a probability measure μ this quantity is always well defined. We consider the posterior distributions as functions defined on the infinite product of the $\mathcal{X}$'s. Consequently we write $F_{x,n}$ instead of F_{z_n} where x is an element of the infinite product and coincides with z_n in the first n components. Let $x_1, x_2, \dots$ be independent random variables which are distributed according to $P_{\theta_0} \in \mathcal{E}$. The posterior distributions computed from these variables are studied below.

We denote by P_θ^n the n -fold product measure of P_θ and by $\mathcal{A}^n$ the product σ -algebra. Instead of the product densities $\prod_{i=1}^n f_\theta(x_i)$ we write simply $f_\theta^n(x)$. For a set $B \subset \Theta$ with $\mu(B) > 0$ we put $P_B^n = \frac{1}{\mu(B)} \int_B P_\theta^n \mu(d\theta)$. The *affinity* between two positive σ -finite measures Q, R is defined by $\gamma(Q, R) = \int \sqrt{gh} d\eta$, where g and h are densities of Q and R with respect to a dominating measure η .

$B_\theta(a)$ denotes the closed ball with center θ and radius a in $\mathbb{R}^k$ and λ denotes the Lebesgue measure in $\mathbb{R}^k$. If t is an $\mathbb{R}^k$ -column vector then t^T denotes the transposed $\mathbb{R}^k$ -row vector.

For a set $B \subset \Theta$ define

$$\pi_j(B) = \inf_{\varphi} \left[\int (1 - \varphi) dP_{\theta_0}^j + \sup_{\theta \in B} \int \varphi dP_{\theta}^j \right],$$

where the infimum is taken over all $\mathcal{A}^j$ -measurable functions φ with $0 \leq \varphi \leq 1$.

Condition (C). A set B satisfies condition (C) at θ_0 if there exists a natural number ℓ with $\pi_{\ell}(B) < 1$.

Remark. A set which satisfies condition (C) does not contain θ_0 . Condition (C) is equivalent to the existence of uniformly consistent tests as considered by Schwartz in [4].

We need also the following condition on the prior measure.

Condition (R). There exists a version of the density of μ with respect to Lebesgue-measure which is positive at θ_0 and for which θ_0 is a Lebesgue-point.

In addition we usually impose the following **regularity assumptions**.

- (1) θ_0 is an interior point of Θ .
- (2) If $\theta \neq \theta'$, then $P_{\theta} \neq P_{\theta'}$.
- (3) There exists a compact neighbourhood K_{θ_0} of θ_0 whose complement satisfies condition (C).
- (4) The function $\theta \rightsquigarrow P_{\theta}$ is L^1 -continuous in K_{θ_0} .
- (5) The map $\theta \rightsquigarrow X(\theta) = \sqrt{\frac{dP_{\theta}}{d\nu}}$ of Θ into $L^2(\nu)$ is Frechet-differentiable in θ_0 and the matrix $\Gamma(\theta_0) = 4E_{\nu} X'(\theta_0) X'(\theta_0)^T$ is regular.
- (6) The measure μ is σ -finite and satisfies condition (R).

We need also a weaker form of assumption (5):

- (5') In a neighbourhood of θ_0 we have $\|P_{\theta} - P_{\theta_0}\|_1 \leq c_{\theta_0} |\theta - \theta_0|^{\alpha}$ with $c_{\theta_0} > 0$ and $0 < \alpha < 1$.

3a. Consistency Theorems for Probability Prior Measures

In this chapter we state some generalisations of the consistency results of L. Schwartz [10], which we need below. Schwartz considered the parameter space as a subset of $[0, 1]^N$ and proved consistency for posterior means and distributions. Her proof for the posterior mean does not carry over directly when the parameter space is a subset of $\mathbb{R}^k$. In this section we use a different method, involving martingale arguments, to show consistency of the posterior mean and posterior distributions in the $\mathbb{R}^k$ -case. The results hold also for σ -finite prior measures. Our method gives explicit bounds for the speed of convergence of the considered quantities. The main ideas of our approach can be seen in the following result and its proof.

As in Sect. 2, $\pi_\ell(B)$ denotes the minimum sum of the error probabilities after ℓ observations of testing " $\theta = \theta_0$ " versus " $\theta \in B$ ".

Proposition 3.1. *Let $B \subset \Theta$ satisfy condition (C) with a certain positive integer $\ell \geq 1$. Let μ be a probability measure concentrated on B . Then*

$$\frac{dP_B^n}{dP_{\theta_0}^n} \leq |\pi_\ell(B)(2 - \pi_\ell(B))|^{n/2\ell} \quad (3.1)$$

for all sufficiently large n , P_{θ_0} -almost surely.

This result is a consequence of the following lemmas. The first is a slight generalisation of Lemma 3 in [6] for σ -finite measures. We prove this assuming that $f_\theta(x)$ is jointly measurable. The lemma holds also when the maps $\theta \mapsto P_\theta(A)$ are measurable for every $A \in \mathcal{A}$. We denote by Q_μ^n the measure $\int P_\theta \mu(d\theta)$ where μ is a σ -finite measure.

Lemma 3.2. *Let α and β be positive σ -finite measures, which are concentrated on disjoint Borel subsets A and B of Θ . Assume that a positive integer m exists with $\gamma(Q_\alpha^m, Q_\beta^m) < \infty$. Then for all $n \geq 1$ $\gamma(Q_\alpha^{m+n}, Q_\beta^{m+n}) < \infty$ and*

$$\gamma(Q_\alpha^{m+n}, Q_\beta^{m+n}) \leq \gamma(Q_\alpha^m, Q_\beta^m) \sup_{\alpha', \beta'} \gamma(Q_{\alpha'}^n, Q_{\beta'}^n), \quad (3.2)$$

where the supremum is taken over all pairs of probability measures (α', β') concentrated on A respectively B .

Proof. Denote by $F_{x,\ell}^\alpha$ the posterior measure after ℓ observations with respect to α and by $K_\ell^\alpha(x) = \int f_\theta^\ell(x) \alpha(d\theta)$ its norming constant. Then Fubini's theorem yields

$$\begin{aligned} \gamma(Q_\alpha^{m+n}, Q_\beta^{m+n}) &= \int (K_{m+n}^\alpha(z))^{\frac{1}{2}} (K_{m+n}^\beta(z))^{\frac{1}{2}} v^{m+n}(dz) \\ &= \int [(\int f_\theta^n(y) F_{x,m}^\alpha(d\theta))^{\frac{1}{2}} (K_m^\alpha(x))^{\frac{1}{2}}] \\ &\quad \cdot [(\int f_\theta^n(y) F_{x,m}^\beta(d\theta))^{\frac{1}{2}} (K_m^\beta(x))^{\frac{1}{2}}] v^n(dy) v^m(dx) \\ &= \int [(\int f_\theta^n(y) F_{x,m}^\alpha(d\theta))^{\frac{1}{2}} (\int f_\theta^n(y) F_{x,m}^\beta(d\theta))^{\frac{1}{2}} v^n(dy)] \\ &\quad \cdot (K_m^\alpha(x))^{\frac{1}{2}} (K_m^\beta(x))^{\frac{1}{2}} v^m(dx) \\ &\leq \gamma(Q_\alpha^m, Q_\beta^m) \cdot \sup_{\alpha', \beta'} \gamma(Q_{\alpha'}^n, Q_{\beta'}^n). \end{aligned}$$

Lemma 3.3. *Let $B \subset \Theta$ satisfy condition (C) with a certain natural number ℓ . Let β be a probability measure concentrated on B and let g be a β -integrable function. Put $a = (\pi_\ell(B)(2 - \pi_\ell(B)))^{\frac{1}{2}}$. Then*

$$P_{\theta_0} \left(\left\{ \max_{m \leq n} \left| \int g(\theta) \frac{f_\theta^n(x)}{f_{\theta_0}^n(x)} \beta(d\theta) \right| > a^{m+1} \right\} \right) \leq (\int |g(\theta)| \beta(d\theta))^{\frac{1}{2}} a^{\frac{m-1}{2}}. \quad (3.3)$$

Proof. Let $\rho = \int |g(\theta)| \beta(d\theta) < \infty$ and $\beta^* = \rho^{-1} |g| \beta$. Since $\left(\int \frac{f_\theta^n}{f_{\theta_0}^n} \beta^*(d\theta) \right)^{\frac{1}{2}}$, $n \geq 1$, is a positive P_{θ_0} -supermartingale, Doob's inequality and (3.2) yield

$$\begin{aligned}
P_{\theta_0} & \left(\left\{ \max_{m\ell \leq n} \left| \int g(\theta) \frac{f_{\theta}^n(x)}{f_{\theta_0}^n(x)} \beta(d\theta) \right| > \varepsilon \right\} \right) \\
& \leq \sqrt{\rho/\varepsilon} \int \left(\frac{\int f_{\theta}^{m\ell}(x) \beta^*(d\theta)}{f_{\theta_0}^{m\ell}(x)} \right)^{\frac{1}{2}} f_{\theta_0}^{m\ell}(x) \nu^{m\ell}(dx) \\
& = \sqrt{\rho/\varepsilon} \gamma(Q_{\beta^*}^{m\ell}, P_{\theta_0}^{m\ell}) \\
& \leq \sqrt{\rho/\varepsilon} [\sup_{\eta} \gamma(Q_{\eta}^{\ell}, P_{\theta_0}^{\ell})]^m.
\end{aligned}$$

Here the supremum runs over all probability measures η concentrated on B . As

$$\gamma(P_{\eta}^{\ell}, P_{\theta_0}^{\ell}) \leq (\|P_{\theta_0}^{\ell} \wedge P_{\eta}^{\ell}\|_1 (2 - \|P_{\theta_0}^{\ell} \wedge P_{\eta}^{\ell}\|_1))^{\frac{1}{2}} \quad (\text{cf. [6]})$$

we get further

$$\begin{aligned}
& \leq \sqrt{\rho/\varepsilon} [\sup_{\eta} \|P_{\theta_0}^{\ell} \wedge P_{\eta}^{\ell}\|_1 (2 - \|P_{\theta_0}^{\ell} \wedge P_{\eta}^{\ell}\|_1)]^{\frac{m}{2}} \\
& \leq \sqrt{\rho/\varepsilon} [\pi_{\ell}(B)(2 - \pi_{\ell}(B))]^{\frac{m}{2}}.
\end{aligned}$$

Choosing $\varepsilon = (\pi_{\ell}(B)(2 - \pi_{\ell}(B)))^{\frac{m+1}{2}}$ yields the result.

Proof of Proposition 3.1. Put

$$A_m = \left\{ \max_{m\ell \leq n \leq (m+1)\ell} \frac{dP_B^n}{dP_{\theta_0}^n} > a^{m+1} \right\},$$

then (3.3) yields $\sum_{m=1}^{\infty} P_{\theta_0}(A_m) < \infty$. Thus

$$\max_{m\ell \leq n \leq (m+1)\ell} \frac{dP_B^n}{dP_{\theta_0}^n} \leq a^{m+1}$$

for all sufficiently large m P_{θ_0} -almost surely.

As $m+1 \geq n/\ell$, we get

$$\frac{dP_B^n}{dP_{\theta_0}^n} \leq a^{n/\ell} = (\pi_{\ell}(B)(2 - \pi_{\ell}(B)))^{n/2\ell}.$$

It is now possible to sharpen Schwartz's Theorem 6.1 of [10], including a statement of speed of convergence in terms of condition (C).

Theorem 3.4. Let μ be a probability measure. Let B satisfy condition (C). Define $H(\theta) = E_{\theta_0} \log \frac{f_{\theta}(x)}{f_{\theta_0}(x)}$. Assume that for every $\varepsilon > 0$ B^c contains a subset W_{ε} with $\mu(W_{\varepsilon}) > 0$ and $H(\theta) > -\varepsilon$ on W_{ε} . Then for all $b < \frac{1}{2\ell} |\log(\pi_{\ell}(B)(2 - \pi_{\ell}(B)))|$

$$\lim_{n \rightarrow \infty} e^{bn} F_{x,n}(B) = 0 \quad P_{\theta_0}\text{-almost surely.}$$

Proof. We argue as in the proof of Theorem 6.1 in [10] but use Proposition 3.1 instead of Lemma 6.1 there.

Combining Lemma 3.3 with ideas of Schwartz we get the next two results, which correspond to Theorem 6.3 in [10].

Proposition 3.5. *Let condition (5') hold. Let μ be a probability measure which satisfies condition (6) and let $B \subset \Theta$ satisfy condition (C). Let $g: \Theta \rightarrow \mathbb{R}^p$ be a μ -integrable function. Then*

$$\lim_{n \rightarrow \infty} e^{bn} \int_B g(\theta) F_{x,n}(d\theta) = 0$$

P_{θ_0} -almost surely for

$$0 \leq b < |\log(\pi_\ell(2 - \pi_\ell))|/2\ell.$$

Proof. Let $W_n = \{\theta \in \Theta \mid \|P_\theta - P_{\theta_0}\|_1 \leq c_{\theta_0} n^{-\alpha\delta}\}$ with $\delta > 0$,

$$q_n = \mu(W_n)^{-1} \int_{W_n} f_\theta^n(x) \mu(d\theta)$$

and

$$d = \log(\pi_\ell(B)(2 - \pi_\ell(B)))/2\ell.$$

Then, for every $0 < \beta < 1$, (3.3) yields

$$\int_B g(\theta) F_{x,n}(d\theta) \leq \frac{\int_B g(\theta) \frac{f_\theta^n(x)}{f_{\theta_0}^n(x)} \mu(d\theta)}{\mu(W_n) \frac{q_n(x)}{f_{\theta_0}^n(x)}} \leq \left(\frac{e^{\beta d}}{\mu(W_n)^{1/n}} \right)^n e^{n(1-\beta)d} \left(\frac{q_n(x)}{f_{\theta_0}^n(x)} \right)^{-1} \quad (3.4)$$

for all sufficiently large n , P_{θ_0} -almost surely. Now

$$\liminf_{n \rightarrow \infty} \mu(W_n)^{1/n} = 1. \quad (3.5)$$

This follows from the assumptions (5') and (6) since

$$\begin{aligned} \mu(W_n) &\geq \mu(\{\theta \mid |\theta - \theta_0| \leq n^{-\delta}\}) \\ &\geq b_0 \lambda(\{\theta \mid |\theta - \theta_0| \leq n^{-\delta}\}) \\ &\geq b_1 n^{-\delta k}, \end{aligned}$$

with constants b_0 and b_1 .

Furthermore

$$\lim_{n \rightarrow \infty} \frac{q_n(x)}{f_{\theta_0}^n(x)} = 1 \quad P_{\theta_0}\text{-almost surely.} \quad (3.6)$$

This follows by assumption (5') from

$$P_{\theta_0} \left(\left| \frac{q_n(x)}{f_{\theta_0}^n(x)} - 1 \right| > \varepsilon \right) < \frac{1}{\varepsilon \mu(W_n)} \int_{W_n} \|P_\theta^n - P_{\theta_0}^n\|_1 \mu(d\theta) \leq \frac{c_{\theta_0}}{\varepsilon} n^{-\alpha\delta}.$$

Choosing $\delta > 2/\alpha$ yields

$$\sum_{n \geq 1} P_{\theta_0} \left(\left\{ \left| \frac{q_n(x)}{f_{\theta_0}^n(x)} - 1 \right| > \varepsilon \right\} \right) < \infty$$

for an arbitrary $\varepsilon > 0$. From (3.4), (3.5) and (3.6) the result follows.

As a consequence of Proposition 3.5 we get a consistency theorem for posterior distributions and conditional expectations.

Theorem 3.6. *Let μ be a probability measure and let the conditions (1)–(4), (5') and (6) hold. Then for every neighbourhood V of θ_0 there exists a constant $b_v > 0$ such that*

$$\lim_{n \rightarrow \infty} e^{b_v n} F_{x,n}(V^c) = 0 \quad P_{\theta_0}\text{-almost surely.} \quad (3.7)$$

If g is a μ -integrable function which is continuous at θ_0 , then

$$\lim_{n \rightarrow \infty} \int g(\theta) F_{x,n}(d\theta) = g(\theta_0) \quad P_{\theta_0}\text{-almost surely.} \quad (3.8)$$

Proof. We argue as Schwartz in the proof of Theorem 6.4 in [10].

Let $\{V_k; k \geq 1\}$ be a base of open neighbourhoods of θ_0 contained in K_{θ_0} , where K_{θ_0} is the compact set of condition (4). Then V_k^c satisfies condition (C) for every k . By Proposition 3.5 we get the result.

3b. Consistency Theorems for σ -finite Prior Measures

Our methods lead easily to extensions of Proposition 3.5 and Theorem 3.6 to σ -finite prior measures and to consistency results for the maximum probability estimator.

Theorem 3.7. *Let μ be a σ -finite prior measure on Θ and let the conditions (1)–(4), (5') and (6) hold. Assume further that for a positive integer j*

$$\gamma(P_{\theta_0}^j, \int (|g(\theta)| \vee 1) P_{\theta}^j \mu(d\theta)) < \infty. \quad (3.9)$$

Then $\int g(\theta) F_{x,n}(d\theta)$ is defined for $n \geq j$, P_{θ_0} -almost surely.

(i) *For every neighbourhood V of θ_0*

$$\lim_{n \rightarrow \infty} e^{bn} \int_{V^c} g(\theta) F_{x,n}(d\theta) = 0 \quad P_{\theta_0}\text{-almost surely for a certain } b > 0.$$

(ii) *If g is continuous in θ_0 , then*

$$\lim_{n \rightarrow \infty} \int g(\theta) F_{x,n}(d\theta) = g(\theta_0) \quad P_{\theta_0}\text{-almost surely.}$$

Proof. The assumptions, together with Lemma 3.2, imply that for $n \geq j$ $\gamma(P_{\theta_0}^n, \int (|g(\theta)| \vee 1) P_{\theta}^n \mu(d\theta)) < \infty$. Therefore the integrals $\int g(\theta) f_{\theta}^n(x) \mu(d\theta)$ and $\int f_{\theta}^n(x) \mu(d\theta)$ are finite P_{θ_0} -almost surely. To estimate the integral

$\int_{V^c} |g(\theta)| \frac{f_{\theta}^n(x)}{f_{\theta_0}^n(x)} \mu(d\theta)$ we proceed on the same line as in the proof of Proposition 3.1, using a generalisation of Lemma 3.3 for σ -finite measures. This

obviously follows from Lemma 3.2 by noting that

$$\left(\int_{V^c} \frac{f_\theta^n}{f_{\theta_0}^n} |g(\theta)| \mu(d\theta) \right)^{\frac{1}{2}} \quad \text{is a } P_{\theta_0}\text{-supermartingale for } n \geq j.$$

Theorem 3.7 implies the consistency of the maximum probability estimator, which is defined as follows. Let Θ be an open subset of $\mathbb{R}^k$. Assume that the posterior distributions with respect to the Lebesgue measure λ exist. For a given number $a > 0$, the *maximum probability estimator* $\hat{\theta}_\ell$ is defined as the value of θ which maximizes $F_{x,\ell}(B_\theta(a/\sqrt{\ell}))$. (The balls are rescaled with $1/\sqrt{\ell}$, as this is the contraction speed around the true parameter of the posterior distributions when the number of observations increases.)

Corollary 3.8. *Let Θ be an open subset of $\mathbb{R}^k$ with Lebesgue measure as prior distribution. Let the conditions (2)–(4) and (5') hold. Assume that $\gamma(P_{\theta_0}^j, \int P_\theta^j \lambda(d\theta)) < \infty$ for some positive integer j . Then the maximum probability estimator $\hat{\theta}_n$ is well defined for sufficiently large n , P_{θ_0} -almost surely and $\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta_0$ P_{θ_0} -almost surely.*

Remark. For an improvement see Theorem 4.6.

Proof. By Theorem 3.7 the posterior distributions are well defined for all sufficiently large n and the functions $\psi_{x,n}: \theta \mapsto F_{x,n}(B_\theta(a/\sqrt{n}))$ are continuous P_{θ_0} -almost surely. Again by Theorem 3.7 for every $d > 0$ there exists a $b > 0$ with

$$F_{x,n}(B_{\theta_0}^c(d)) = o(e^{-bn}) \quad P_{\theta_0}\text{-almost surely.} \quad (3.10)$$

As the ball $B_{\theta_0}(d)$ can be covered by $O(n^{k/2})$ disjoint cubes of side length $a/\sqrt{n}$, the maximum of the function $\psi_{x,n}$ must be larger than $\text{const.} \cdot n^{-k/2}(1 - o(e^{-bn}))$ by (3.10). This, together with (3.10), implies that the maximum of the function $\psi_{x,n}$ is attained in the ball $B_{\theta_0}(d + a/\sqrt{n})$ for all sufficiently large n , P_{θ_0} -almost surely. As $d > 0$ is arbitrary, this yields the result.

4a. The Law of the Iterated Logarithm for Posterior Distributions

In this section we prove one half of the LIL for posterior distributions. The estimates are based on a recent construction of local uniformly consistent tests given by Le Cam in [5]. (See Lemma 4.2 and 4.3 below.)

Theorem 4.1. *Let μ be a probability measure. Let condition (1)–(6) hold. Then there exists a positive constant γ_0 such that*

$$\lim_{n \rightarrow \infty} F_{x,n} \left(B_{\theta_0}^c \left(\gamma_0 \sqrt{\frac{\log \log n}{n}} \right) \right) = 0 \quad P_{\theta_0}\text{-almost surely.} \quad (4.1)$$

Remarks. a) The constant γ_0 is not explicitly known. For normal models our method of proof yields $\gamma_0 \geq \sqrt{8\pi e}$.

b) We do not know whether conditions (1)–(6) imply the existence of a positive constant γ_1 with

$$\limsup_{n \rightarrow \infty} F_{x,n} \left(B_{\theta_0}^c \left(\gamma_1 \sqrt{\frac{\log \log n}{n}} \right) \right) = 1 \quad P_{\theta_0}\text{-almost surely} \quad (4.2)$$

(the other half of the LIL).

c) If the parameter set Θ is one-dimensional, (4.2) holds under somewhat stronger conditions, which are, for instance, fulfilled by the double exponential distributions.

Proof. For simplicity let $\theta_0 = 0$ and $\log_2 n \equiv \log \log n$. Let V_0 denote the neighbourhood of $\theta_0 = 0$ introduced in Lemma 4.2. From the consistency theorem (cf. [10, p. 22] or Theorem 3.6) we get $\lim_{n \rightarrow \infty} F_{x,n}(V_0^c) = 0$ P_0 -almost surely. Therefore it is sufficient to prove

$$\begin{aligned} \lim_{n \rightarrow \infty} F_{x,n}(V_0 \cap B_0^c(\gamma \sqrt{\log_2 n/n})) &= 0 \quad P_0\text{-almost surely or equivalently} \\ \sup_{m \geq n} F_{x,m}(V_0 \cap B_0^c(\gamma \sqrt{\log_2 m/m})) &\rightarrow 0 \quad \text{in } P_0\text{-probability.} \end{aligned}$$

Dividing the sequence into blocks and denoting by C_j the set $V_0 \cap B_0^c(\gamma \sqrt{\log j/e^{j+1}})$ (where e is the base of the natural logarithm) we obtain the following inequalities

$$\begin{aligned} \sup_{e^j \leq m \leq e^{j+1}} F_{x,m}(V_0 \cap B_0^c(\gamma \sqrt{\log_2 m/m})) \\ \leq \sup_{e^j \leq m \leq e^{j+1}} F_{x,m}(C_j) \\ \leq I_{W_j} \varphi_j \sup_{e^j \leq m \leq e^{j+1}} F_{x,m}(C_j) + I_{W_j^c} + (1 - \varphi_j) \end{aligned} \quad (4.3)$$

where φ_j are certain tests specified below, and I_{W_j} is the indicator function of the set $W_j = \left\{ \inf_{e^j \leq m \leq e^{j+1}} \frac{dP_{B_i}^m}{dP_0^m} > \beta \right\}$. Here B_j denotes the ball $B_0(e^{-j/2} j^{-(1+\delta)})$ with $\delta > 0$. By Lemma 4.5, for every $0 < \beta < 1$, we have

$$\sum_{j \geq 1} P_0(W_j^c) \leq \frac{de^{\frac{1}{2}}}{1-\beta} \sum_{j \geq 1} j^{-(1+\delta)} < \infty. \quad (4.4)$$

The tests φ_j are those of Lemma 4.3, where $a_n = \frac{\gamma}{e^{\frac{1}{2}}} \left(\frac{\log_2 n}{n} \right)^{\frac{1}{2}}$ and $n = e^j$. The tests φ_j satisfy the following inequalities

$$\int (1 - \varphi_j) dP_0^{e^j} \leq c_1 j^{-2c_0^2 \gamma^2/e}, \quad (4.5)$$

$$\int \left(\int_{C_j} \varphi_j dP_0^{e^j} \right) d\theta \leq c_2 e^{-kj/2} j^{-2c_0^2 \gamma^2/e} (\log j)^{k/2-1}. \quad (4.6)$$

The regularity assumption (6) together with (4.6) imply

$$\int \varphi_j dP_{C_j}^{e^j} \leq c_3 e^{-kj/2} j^{-2c_0^2 \gamma^2/e} (\log j)^{k/2-1}. \quad (4.7)$$

Choosing γ large enough we get from (4.5)

$$\sum_{j \geq 1} \int (1 - \varphi_j) dP_0 < \infty.$$

This statement, together with (4.4), implies that the sum $\sum_{j \geq j_0} \int (I_{W_j^c} + (1 - \varphi_j)) dP_0$ is arbitrarily small, if j_0 is chosen large enough. Therefore we need only to show that

$$\sup_{j \geq j_1} I_{W_j} \varphi_j \sup_{e^j \leq m \leq e^{j+1}} F_{x,m}(C_j) \rightarrow 0$$

in P_0 -probability as $j_1 \rightarrow \infty$. It is sufficient to show that for every $\varepsilon > 0$ we have

$$\sum_{j > 1} P_0 \{ I_{W_j} \varphi_j \sup_{e^j \leq m < e^{j+1}} F_{x,m}(C_j) > \varepsilon \} < \infty. \quad (4.8)$$

Observe first that

$$\begin{aligned} & I_{W_j} \varphi_j \sup_{e^j \leq m \leq e^{j+1}} F_{x,m}(C_j) \\ & \leq I_{W_j} \varphi_j \sup_{e^j \leq m \leq e^{j+1}} \frac{dP_{C_j}^m}{dP_0^m} \left(\mu(B_j) \frac{dP_{B_j}^m}{dP_0^m} \right)^{-1} \\ & \leq \frac{1}{\beta a_1} e^{kj/2} j^{k(1+\delta)} \varphi_j \sup_{e^j \leq m < e^{j+1}} \frac{dP_{C_j}^m}{dP_0^m}. \end{aligned} \quad (4.9)$$

The last inequality follows by using the definition of the sets W_j and the inequality

$$\mu(B_j) \geq a_1 e^{-kj/2} j^{-k(1+\delta)}, \quad \text{which is a consequence of assumption (6).}$$

Since $\left(\frac{dP_{C_j}^m}{dP_0^m}; m \geq 1 \right)$ is a supermartingale we get from Doob's inequality

$$\begin{aligned} & P_0 \left\{ \varphi_j \sup_{e^j \leq m \leq e^{j+1}} \frac{dP_{C_j}^m}{dP_0^m} > \varepsilon \right\} \\ & \leq \frac{1}{\varepsilon} \int \varphi_j dP_{C_j}^{e^j} \\ & \leq \frac{c_3}{\varepsilon} e^{-kj/2} j^{-2c_0^2\gamma^2/e} (\log j)^{k/2-1}. \end{aligned} \quad (4.10)$$

The inequalities (4.9) and (4.10) now imply that, for j large enough,

$$\begin{aligned} & P_0 \{ I_{W_j} \varphi_j \sup_{e^j \leq m \leq e^{j+1}} F_{x,m}(C_j) > \varepsilon \} \\ & \leq \frac{c_3}{\varepsilon \beta a_1} j^{-(2c_0^2\gamma^2/e - k(1+2\delta))}. \end{aligned} \quad (4.11)$$

If γ is chosen large enough, the right hand side of (4.11) is summable. This implies (4.8) and finally the theorem.

Lemma 4.2 (Le Cam). *Let the conditions (1), (2) and (5) hold. Then there exists a neighbourhood V_0 of $\theta_0 = 0$ and a measurable function $T: \mathcal{X} \rightarrow [0, 1]$ such that*

the image experiment of $\{P_\theta; \theta \in V_0\}$ under T , $\{H_\theta; \theta \in V_0\}$, has the following properties:

There exists a constant $c_0 > 0$ with

$$\|H_\theta - H_0\|_K \geq 2c_0|\theta| \quad \text{for every } \theta \in V_0, \quad (4.12)$$

where

$$\|F - G\|_K := \sup_{x \in [0, 1]} |F([0, x]) - G([0, x])|$$

for positive finite measures F and G .

(ii) There exist estimates S_n of H_θ which fulfil the following inequalities with a positive constant c_1 :

$$P_\theta \{x \mid \|S_n(x) - H_\theta\|_K > \tau/n^{1/2}\} \leq c_1 \exp(-2\tau^2) \quad (4.13)$$

for all $\tau > 0$ and for every $\theta \in V_0$.

Proof. The proof of statement (4.12) can be found in [3] and [4, p.106]. Statement (4.13) is the Dvoretzky-Kiefer-Wolfowitz inequality for empirical distribution functions (cf. [1]), if one chooses S_n as the composition of the function φ and the empirical distribution function on $[0, 1]$.

Following Le Cam ([5]) this lemma is used now to construct local uniformly consistent tests. Lemma 4.3 is a quantified version of Lemma 12 in [5].

Lemma 4.3. *Let the conditions (1), (2) and (5) hold. Let $(a_n; n \geq 1)$ be a sequence of real numbers with $\lim_{n \rightarrow \infty} n^{1/2} a_n = \infty$. Let V_0 be the neighbourhood of $\theta_0 = 0$ introduced in Lemma 4.2 and let $D_n = V_0 \cap B_0^c(a_n)$. Then for all n there exist $\mathcal{A}^n$ -measurable tests ψ_n such that, for all integers $i \geq 0$,*

$$\int (1 - \psi_n) dP_0^n \leq c_1 \exp(-2c_0^2 n a_n^2), \quad (4.14)$$

$$\int_{D_n} |\theta|^i (\int \psi_n dP_\theta^n) d\theta \leq c_{2,i} \frac{a_n^{k+i}}{n a_n^2} \exp(-2c_0^2 n a_n^2) \quad (4.15)$$

with positive constants c_1 and $c_{2,i}$.

Proof (Le Cam). Let $U_n = \{x \mid \|S_n(x) - H_0\|_K \leq c_0 a_n\}$ and let $\psi_n = I_{U_n}$, where c_0 is the constant of Lemma 4.2. From Lemma 4.2 we get

$$\int (1 - \psi_n) dP_0^n = P_0 \{x \mid \|S_n(x) - H_0\|_K > c_0 a_n\} \leq c_1 e^{-2c_0^2 n a_n^2},$$

which yields inequality (4.14).

To show the inequality (4.15), note that for $\xi \in D_n$

$$U_n \subset \{x \mid \|S_n(x) - H_\xi\|_K \geq c_0 |\xi|\}.$$

This follows, since for $x \in U_n$ and $\xi \in D_n$ by (4.12)

$$\begin{aligned} \|S_n(x) - H_\xi\|_K &\geq \|H_\xi - H_0\|_K - \|S_n(x) - H_0\|_K \\ &\geq 2c_0 |\xi| - c_0 a_n \\ &\geq c_0 |\xi|. \end{aligned}$$

Hence by (4.13)

$$\begin{aligned} P_\xi^n(U_n) &\leq P_\xi^n\{x \mid \|S_n(x) - H_\xi\|_K \geq c_0 |\xi|\} \\ &\leq c_1 e^{-2n(c_0 |\xi|)^2} \end{aligned}$$

and therefore, by a direct computation,

$$\begin{aligned} \int_{D_n} (|\xi|^i \int \psi_n dP_\xi^n) d\xi &\leq c_1 \int_{\{|\xi| > a_n\}} |\xi|^i e^{-2n(c_0 |\xi|)^2} d\xi \\ &= c'_1 n^{-(k+i)/2} \int_{\sqrt{n}a_n}^{\infty} z^{k+i-1} e^{-2(c_0 z)^2} dz \\ &= c'_1 a_n^{k+i} e^{-2c_0^2 n a_n^2} \int_0^{\infty} (1+s)^{\frac{k+i}{2}-1} e^{-2c_0^2 n a_n^2 s} ds. \end{aligned}$$

The integral on the right hand side is of the order $\frac{1}{2c_0^2 n a_n^2}$. This yields (4.15) and completes the proof.

The following lemma uses the technique of [3].

Lemma 4.4. *Let the conditions (1), (2) and (5) hold. Then for every $b > 0$ there exists a constant $d_b > 0$ such that, uniformly in n ,*

$$\|P_0^n - P_{\xi/\sqrt{n}}^n\|_1 \leq d_b |\xi| \quad \text{for all } |\xi| \leq b. \quad (4.16)$$

Proof. The differentiability of $X(\theta)$ implies

$$\frac{n}{|\xi|^2} \|X(\xi/\sqrt{n}) - X(0)\|_2^2 \rightarrow \left\| \frac{\xi^T}{|\xi|} X'(0) \right\|_2^2 \quad (4.17)$$

uniformly in $|\xi| \leq b$. Together with the identity

$$\|X(\xi/\sqrt{n}) - X(0)\|_2^2 = 2(1 - \gamma(P_0, P_{\xi/\sqrt{n}}))$$

we get from (4.17)

$$\begin{aligned} \gamma(P_0, P_{\xi/\sqrt{n}})^n &\geq \left[1 - \frac{|\xi|^2}{2n} \left(\sup_{|\eta|=1} \|\eta^T X'(0)\|_2^2 + o(1) \right) \right]^n \\ &\geq \exp \left(-\frac{|\xi|^2}{2} (K + o(1)) \right) \end{aligned}$$

uniformly in $|\xi| \leq b$ where $K = \sup_{|\eta|=1} \|\eta^T X'(0)\|_2^2$.

This implies

$$\begin{aligned} \frac{1}{2} \|P_0^n - P_{\xi/\sqrt{n}}^n\|_1 &\leq 1 - \gamma(P_0, P_{\xi/\sqrt{n}})^n \\ &\leq 1 - \exp \left(-\frac{1}{2} |\xi|^2 (K + o(1)) \right) \\ &\leq \frac{1}{2} (K + o(1)) |\xi|^2 \end{aligned} \quad (4.18)$$

uniformly in $|\xi| \leq b$.

The Lipschitz-continuity for small n follows from the differentiability assumption. This, together with (4.18), implies the lemma.

Lemma 4.5. *Let the conditions (1), (2), (5) and (6) hold. Let $\delta > 0$ and denote by B_j the ball $B_0(e^{-j/2} j^{-(1+\delta)})$ where e is the base of the natural logarithm. Then for every $0 < \varepsilon < 1$ and every $j > 0$*

$$P_0 \left\{ \inf_{e^j \leq m \leq e^{j+1}} \frac{dP_{B_j}^m}{dP_0^m} < 1 - \varepsilon \right\} < \frac{de^{\frac{1}{2}}}{\varepsilon} j^{-(1+\delta)},$$

where d is a positive constant.

Proof. The sequence $\left\{ \frac{dP_0^m}{dP_{B_j}^m}; m \geq 1 \right\}$ is a martingale, since by Lemma 4.4 P_0^m is absolutely continuous with respect to $P_{B_j}^m$. Then Doob's inequality for submartingales yields

$$\begin{aligned} P_{B_j} \left\{ \inf_{1 \leq m \leq e^{j+1}} \frac{dP_{B_j}^m}{dP_0^m} < 1 - \varepsilon \right\} \\ \leq P_{B_j} \left\{ \sup_{1 \leq m \leq e^{j+1}} \left| \frac{dP_0^m}{dP_{B_j}^m} - 1 \right| > \frac{\varepsilon}{1 - \varepsilon} \right\} \\ \leq \frac{1 - \varepsilon}{\varepsilon} \|P_0^{e^{j+1}} - P_{B_j}^{e^{j+1}}\|_1. \end{aligned}$$

This, together with Lemma 4.4, implies

$$\begin{aligned} P_0 \left\{ \inf_{1 \leq m \leq e^{j+1}} \frac{dP_{B_j}^m}{dP_0^m} < 1 - \varepsilon \right\} \\ \leq \|P_0^{e^{j+1}} - P_{B_j}^{e^{j+1}}\|_1 + P_{B_j} \left\{ \inf_{1 \leq m \leq e^{j+1}} \frac{dP_{B_j}^m}{dP_0^m} < 1 - \varepsilon \right\} \\ \leq \frac{1}{\varepsilon} \|P_0^{e^{j+1}} - P_{B_j}^{e^{j+1}}\|_1 \\ \leq \frac{1}{\varepsilon \mu(B_j)} \int_{B_j} \|P_0^{e^{j+1}} - P_\theta^{e^{j+1}}\|_1 \mu(d\theta) \\ \leq \frac{de^{\frac{1}{2}}}{\varepsilon} j^{-(1+\delta)}, \end{aligned}$$

where d is the constant $d_{\sqrt{e}}$ of Lemma 4.4.

4b. The Law of the Iterated Logarithm for Bayes Estimators

We now prove results analogous to Theorem 4.1 for the maximum probability estimator and the conditional expectation. For the definition of the maximum-probability estimator see Sect. 3b.

Theorem 4.6. *Let Θ be an open subset of $\mathbb{R}^k$. Let the conditions (2)–(6) hold, where the Lebesgue measure is the prior measure. Assume that $\gamma(P_\theta^j, \int P_\theta^j \lambda(d\theta)) < \infty$ for a positive integer j . Then the maximum probability esti-*

mator satisfies

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{n}{\log \log n}} |\hat{\theta}_n - \theta_0| \leq \gamma_2 \quad P_{\theta_0}\text{-almost surely} \quad (4.19)$$

with a constant $\gamma_2 > 0$.

Proof. The proof is similar to that of Corollary 3.8. The proof of Theorem 4.1, especially the statements (4.4), (4.5) and (4.11) together with Theorem 3.7(i) yield

$$F_{x,n} \left(B_{\theta_0}^c \left(\gamma \sqrt{\frac{\log \log n}{n}} \right) \right) = o((\log n)^{-\delta}) \quad (4.20)$$

P_{θ_0} -almost surely with $\gamma > 0$ and $\delta > 0$. Now the maximum of the function $\psi_{x,n}: \theta \mapsto F_{x,n}(B_\theta(a/\sqrt{n}))$ must be larger than $\text{const} \cdot (\log_2 n)^{-\frac{k}{2}} (1 - o((\log n)^{-\delta}))$ by (4.20). This, together with (4.20), imply that the maximum of $\psi_{x,n}$ is attained in the ball $B_{\theta_0}((\gamma \sqrt{\log_2 n} + a)/\sqrt{n})$. This proves the result.

Theorem 4.7. Let μ be a probability measure with $\int |\theta| \mu(d\theta) < \infty$. Let the conditions (1)–(6) hold. Then the conditional expectation $\hat{\theta}_n = \int \theta F_{x,n}(d\theta)$ satisfies

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{n}{\log \log n}} |\hat{\theta}_n - \theta_0| \leq \gamma_3 \quad (4.21)$$

P_{θ_0} -almost surely with a constant $\gamma_3 > 0$.

Proof. For simplicity let $\theta_0 = 0$ and $\log_2 n \equiv \log \log n$. Then

$$\begin{aligned} \sqrt{\frac{n}{\log_2 n}} |\hat{\theta}_n| &\leq \sqrt{\frac{n}{\log_2 n}} \left(\int_{\{|\theta| \leq \gamma \sqrt{\log_2 n/n}\}} |\theta| F_{x,n}(d\theta) \right. \\ &\quad \left. + \int_{\{\gamma \sqrt{\log_2 n/n} < |\theta| \leq a\}} |\theta| F_{x,n}(d\theta) + \int_{\{|\theta| > a\}} |\theta| F_{x,n}(d\theta) \right). \end{aligned}$$

The last term of the right hand side converges to zero P_{θ_0} -almost surely by Proposition 3.5. The same holds for the second term. To show this, we proceed on the same lines as in the proof of Theorem 4.1. Using the same notations as there, we show that

$$\sup_{m \geq n} \sqrt{m} \int_{\{\gamma \sqrt{\log_2 m/m} < |\theta| \leq a\}} |\theta| F_{x,m}(d\theta) \rightarrow 0 \quad (4.22)$$

in P_0 -probability.

Dividing the sequence into blocks we get

$$\begin{aligned} &\sup_{e^j \leq m \leq e^{j+1}} \sqrt{m} \int_{\{\gamma \sqrt{\log_2 m/m} < |\theta| \leq a\}} |\theta| F_{x,m}(d\theta) \\ &\leq \sup_{e^j \leq m \leq e^{j+1}} e^{\frac{j+1}{2}} \int_{C_j} |\theta| F_{x,m}(d\theta), \end{aligned} \quad (4.23)$$

if a is chosen small enough. Now we argue as in the proof of Theorem 4.1. Note that the tests φ_j introduced in (4.5) and (4.6) are indicator functions $\varphi_j = I_{A_j}$ by construction. As $\sum_{j \geq 1} (P_0(W_j^c) + P_0(A_j^c)) < \infty$, we have to estimate the

right hand side of (4.23) only on the sets $A_j \cap W_j$. As in (4.9) we get on these sets

$$\sup_{e^j \leq m \leq e^{j+1}} e^{j/2} \int_{C_j} |\theta| F_{x,m}(d\theta) \leq \frac{1}{\beta a_1} e^{\frac{(k+1)j}{2}} j^{k(1+\delta)} \int_{C_j} |\theta| \frac{dP_\theta^m}{dP_0^m} \mu(d\theta).$$

By (4.15) we have

$$\int_{C_j} |\theta| (\int \varphi_j dP^{e^j}) d\theta \leq \tilde{c}_{2,1} e^{-(k+1)j/2} (\log j)^{(k+1)/2} j^{-2c_0^2 \gamma^2/e}. \quad (4.24)$$

We estimate now as in (4.11). For j large enough we get

$$\begin{aligned} P_0(A_j \cap W_j \cap \{ \sup_{e^j \leq m \leq e^{j+1}} e^{j/2} \int_{C_j} |\theta| F_{x,m}(d\theta) > \varepsilon \}) \\ \leq \frac{\tilde{c}_{2,1}}{\varepsilon \beta a_1} j^{-(2c_0^2 \gamma^2/e - k(1+2\delta))}. \end{aligned} \quad (4.25)$$

If γ is chosen large enough the right hand side of (4.25) is summable. This yields (4.22) and finally the result.

Acknowledgement. This paper is based on parts of the author's doctoral dissertation, which was written at the University of Heidelberg under the guidance of D.W. Müller, to whom the author is deeply indebted for his constant encouragement and support.

References

1. Dvoretzky, A., Kiefer, J., Wolfowitz, J.: Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Stat.* **27**, 642–669 (1956)
2. Le Cam, L.: On the asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of Cal. Publ. Stat.* **1**, 277–330 (1953)
3. Le Cam, L.: Likelihood functions for large number of independent observations. *Festschrift for J. Neyman*, ed. F.N. David. New York: Wiley 1966
4. Le Cam, L.: *Théorie Asymptotique de la Décision statistique*. Montréal 1969
5. Le Cam, L.: Remarks on the Bernstein-von Mises theorem. Unpublished manuscript
6. Le Cam, L.: Convergence of estimates under dimensionality restrictions. *Ann. Stat.* **1**, 38–53 (1974)
7. Lerche, H.R.: The law of the iterated logarithm for posterior distributions. *Proc. of VI. Conf. on Probability Theory*, Brasov, Romania, 1979, 347–355
8. von Mises, R.: *Mathematical theory of probability and statistics*. New York: Academic Press 1966
9. Robbins, H.: Statistical methods related to the law of the iterated logarithm. *Ann. Math. Stat.* **41**, 1397–1409 (1970)
10. Schwartz, L.: On Bayes procedures. *Zeitschr. f. Wahrscheinlichkeitstheorie verw. Gebiete* **4**, 10–26 (1964)

Received April 6, 1981