The Annals of Statistics
1998, Vol. 26, No. 2, 617–644

ASYMPTOTIC BEHAVIOR OF BAYES ESTIMATES UNDER POSSIBLY INCORRECT MODELS[1]

BY OLAF BUNKE AND XAVIER MILHAUD

Humboldt University and Université Paul Sabatier

We prove that the posterior distribution in a possibly incorrect parametric model a.s. concentrates in a strong sense on the set of pseudotrue parameters determined by the true distribution. As a consequence, we obtain in the case of a unique pseudotrue parameter the strong consistency of pseudo-Bayes estimators w.r.t. general loss functions.

Further, we present a simple example based on normal distributions and having two different pseudotrue parameters, where pseudo-Bayes estimators have an essentially different asymptotic behavior than the pseudomaximum likelihood estimator. While the MLE is strongly consistent, the sequence of posterior means is strongly inconsistent and a.s. almost all its accumulation points are not pseudotrue. Finally, we give conditions under which a pseudo-Bayes estimator for a unique pseudotrue parameter has an asymptotic normal distribution.

1. Introduction. The frequentist asymptotic properties of Bayes estimators and of posterior distributions are well known and have been investigated under the assumption of a correct parametric model; see, for example, Bickel and Yahav (1969), Ibragimov and Has'minskii (1981), Strasser (1991) or Lehmann (1983). The properties are analogous to those of the MLE and it is also known that there is a higher order asymptotical equivalence between Bayes estimators and MLE [see Strasser (1981)].

The asymptotic behavior of MLE in the case of a possibly incorrect parametric model given by densities p_θ ($\theta \in \Theta$) has also been investigated in several papers [see Huber (1967), Pfanzagl (1969) or Gourieroux and Montfort (1993)]. In particular it is shown that the MLE converges a.s. to the subset Θ_G of the parameter set Θ on which the Kullback–Leibler divergence (K–L divergence) of the true distribution G against the distributions given by p_θ is minimal. The points of Θ_G are so-called pseudotrue parameters.

There are a few papers on Bayes estimators in this case. Berk (1966, 1970) showed that under regularity conditions, a.s. the posterior distribution concentrates weakly on Θ_G for increasing sample sizes n. Hanousek and Jureckova (1996) derived sufficient conditions for the consistency and asymptotic normality of the Bayes estimators of a one-dimensional parameter when there is a unique pseudotrue value. On the other hand, Diaconis and Freed-

Received June 1994; revised July 1997.

[1] Research supported in part by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 373 "Quantifikation und Simulation ökonomischer Prozesse," Berlin.

AMS 1991 *subject classifications.* 62F12, 62F15.

Key words and phrases. Consistency, asymptotic normality, incorrect parametric models, inconsistent Bayes estimates.

man (1986a,b) provide an example where Θ_G consists of two points, Bayes estimators are inconsistent and posterior distributions asymptotically a.s. weakly concentrate on Θ_G.

In our paper we investigate problems for which there is yet no satisfactory treatment.

1. Prove a stronger convergence of the posterior distribution than the a.s. weak concentration on Θ_G.
2. Obtain in the case of a k-dimensional unique pseudotrue parameter a strong consistency of Bayes estimators w.r.t. general loss functions.
3. Have a deeper insight into the possibly fluctuating behavior of Bayes estimates and of the posterior distribution in the case of inconsistency, at least for some special case, for example, determining the accumulation points of the sequence of posterior distributions or of Bayes estimators.
4. Provide sufficient general conditions for the asymptotic normality of Bayes estimators and convenient pivotal statistics needed for confidence regions.

In Theorem 2.1, under regularity conditions, we show that the posterior distribution a.s. concentrates on Θ_G in a strong sense. This leads to Theorem 2.2, which states the consistency of the Bayes estimators derived from a general loss function, when the pseudotrue value is unique. As an illustration of these results we treat, by our method, the example of a location-scale Cauchy model left open by Berk (1970). Berk (1966) and Diaconis and Freedman (1986a,b) were the first to provide examples of inconsistent asymptotical behavior of the posterior distribution. Section 3 presents a simpler example with normal distributions in which Θ_G consist of two points, where we can see that the MLE tends a.s. to the set Θ_G, while Bayes estimators fluctuate around the convex hull of Θ_G. The points in the interior of this hull are not pseudotrue but they are a.s. accumulation points of sequences of Bayes estimates. In Section 4, we state sufficient conditions for the asymptotic normality of Bayes estimators in a possibly incorrect model. In particular, the pseudotrue parameter is assumed to be unique. A corresponding asymptotically normal pivot statistic is presented.

To be more specific, we assume the observations $X_1(\omega), \dots, X_n(\omega)$ to be realizations of i.i.d. random vectors X_i: $(\Omega, \mathscr{A}) \to (R^k, \mathscr{B}^k)$, each having the distribution G. We assume G and P_ϑ, for all $\vartheta \in \Theta$, to be probability distributions (p.d.'s) on $\mathscr{B}^k$ having densities g and p_θ w.r.t. a nonnegative σ-finite measure μ on $\mathscr{B}^k$. Corresponding to the possibly incorrect model

$$X_i \sim P_\vartheta, \qquad \vartheta \in \Theta, \qquad i = 1, \dots, n, \tag{1.1}$$

we define a pseudotrue parameter ϑ_G as a value minimizing the divergence

$$K(\vartheta) := E \log[g(X_i)/p_\vartheta(X_i)], \tag{1.2}$$

over Θ. The expectation E is w.r.t. the true distribution G.

EXAMPLE 1. The location model is given by

$$g(x) = h(x - \vartheta_0), \qquad p_\vartheta(x) = f(x - \vartheta), \qquad \vartheta \in R^1, \tag{1.3}$$

where h, f are positive symmetric functions on R^1 and ϑ_0 is the true location parameter. We see that, with obvious integrability conditions on h and f, assuming $\log p_\vartheta(x)$ to be a strictly concave function of ϑ (for all fixed $x \in R^1$), the divergence (1.2) has a unique minimum at $\vartheta = \vartheta_0$ so that the pseudotrue parameter is just the location parameter. The example of Diaconis and Freedman (1986) with two pseudotrue parameters different from θ_0 is a model (1.3) with a special bimodal density g and a Cauchy density p_θ.

EXAMPLE 2. The exponential model is given by

$$p_\vartheta(x) = \exp[\vartheta' t(x) - a(\vartheta)], \qquad \vartheta \cap \Theta \subset R^d,$$

where Θ is the natural parameter space. It is easy to see that the divergence (1.2) is minimized by the uniquely determined value ϑ_G with

$$\int t(x) p_{\vartheta_G}(x)\, d\mu(x) = Et(X) = \int t(x)\, dG(x)$$

so that the pseudotrue parameter is just the true observation mean if the mean-value parametrization [see Lehman (1983)] is used.

We assume Θ to be a Borel set in R^d and denote by $\mathscr{B}_\Theta$ the class of Borel sets in Θ. A nonnegative measure ξ on $\mathscr{B}_\Theta$ is called a prior distribution, which is called improper if it is not finite. We denote by E_ξ the integral (expectation) w.r.t. ξ and by $E_{n,\xi,\omega}$ the integral w.r.t. to the posterior distribution $P_{n,\xi,\omega}$ under (1.1), which is defined by

$$(1.4)\quad dP_{n,\xi,\omega}(\vartheta) = \left[\int \prod_{i=1}^{n} p_\tau(X_i(\omega))\, d\xi(\tau)\right]^{-1} \prod_{i=1}^{n} p_\vartheta(X_i(\omega))\, d\xi(\vartheta),$$

assuming that a.s. the above integral is finite for some $n = n_0 \in N$. An estimator $\hat{\vartheta}_n$ is called a pseudo-Bayes estimator w.r.t a loss function L: $\Theta \times \Theta \to [0, \infty)$ if for almost all $\omega \in \Omega$ and $\hat{\vartheta}_n(\omega) := \hat{\vartheta}_n(X_1(\omega), \dots, X_n(\omega))$,

$$(1.5)\qquad E_{n,\xi,\omega} L(\hat{\vartheta}_n(\omega), \vartheta) = \min_{t \in \Theta} E_{n,\xi,\omega} L(t, \vartheta).$$

2. Consistency. In the following we state some relatively weak regularity conditions which will be needed for proving convergence properties of the posterior distribution:

A1: Θ is a closed (possibly unbounded) convex set in R^d with a nonempty interior, the density $p_\vartheta(x)$ is bounded over $\Theta \times R^k$ and its carrier $\{x \in R^k | p_\vartheta(x) > 0\}$ is the same for all $\vartheta \in \Theta$.

A2: For all $\vartheta \in \Theta$ there is a sphere $S[\vartheta, \eta]$ of center ϑ and radius $\eta = \eta(\vartheta) > 0$ with

$$(2.1)\qquad E \sup\{|\log[g(X)/p_t(X)]|;\, t \in S[\vartheta, \eta]\} < \infty.$$

A3: For all fixed $x \in R^k$, the density $p_\vartheta(x)$ has a continuous derivative $p'_\vartheta(x)$ w.r.t. ϑ and there are positive constants c, b_0 with

$$\int \left\| [p_\vartheta(x)]^{-1} p'_\vartheta(x) \right\|^{4(d+1)} p_\vartheta(x)\mu(dx) < c(1 + \|\theta\|^{b_0}) \tag{2.2}$$

for all $\vartheta \in \Theta$, where $\|\cdot\|$ denotes a norm in R^d.

A4: For some positive constant b_1 the affinity

$$\varrho(\vartheta) := \int [p_\vartheta(x)g(x)]^{1/2}\mu(dx) \tag{2.3}$$

has the behavior

$$\varrho(\vartheta) < c\|\theta\|^{-b_1}, \qquad \vartheta \in \Theta. \tag{2.4}$$

A5: There are positive constants b_2, b_3 so that for all $\vartheta \in \Theta$ and $r > 0$ it holds that

$$\xi(S[\vartheta, r]) \leq cr^{b_2}\big(1 + (\|\vartheta\| + r)^{b_3}\big). \tag{2.5}$$

Moreover, $\xi(S[\theta, r]) > 0$ for all $r > 0$ and $\theta \in \Theta$.

A6: Let $L\colon \Theta \times \Theta \to R^+$ be a measurable loss function with $L(\vartheta, \vartheta) = 0$ ($\vartheta \in \Theta$), c_1, c_2, c_3, b_4, b_5 be positive constants with

$$\big(c_1\|t - \vartheta\|^{b_4}\big) \wedge c_2 \leq L(t, \vartheta) \leq c_3\|t - \vartheta\|^{b_5} \tag{2.6}$$

for all $t, \vartheta \in \Theta$.

REMARK 1. Assumptions A1–A4 are fulfilled by most of the standard parametric models, provided that the true distribution G has some regularity properties. For instance, if G has a positive density w.r.t. the Lebesgue measure and has a finite second-order moment and the distributions $P_\vartheta = N(\mu, \sigma^2)$ are normal, the assumptions are fulfilled, as can be checked easily using the parameter $\theta = (\mu, \log \sigma^2) \in \Theta = R^2$.

REMARK 2. Assumptions A2 and A3 imply that for all $\vartheta \in \Theta$ and $\alpha > 0$ there is a $\eta = \eta(\vartheta, \alpha)$ with

$$E \sup\{\log[g(X)/p_t(X)] \,|\, t \in S[\vartheta, \eta]\} \leq E \log[g(X)/p_\vartheta(X)] + \alpha, \tag{2.7}$$

$$E \inf\{\log[g(X)/p_t(X)] \,|\, t \in S[\vartheta, \eta]\} \geq E \log[g(X)/p_\vartheta(X)] - \alpha. \tag{2.8}$$

The following theorem shows that the posterior a.s. concentrates on the set Θ_G of all pseudotrue parameters in a relatively strong sense.

THEOREM 2.1. *Under assumptions* A1–A5 *the K–L divergence $K(\theta)$ reaches its global minimum over Θ on a compact set Θ_G. It holds for all $p > 0$ that*

$$\lim_{n \to \infty} E_{n,\xi,\omega} d_G^p(\theta) = 0 \quad a.s. \tag{2.9}$$

where

$$d_G(\theta) = \min\{\|\theta - t\| \,|\, t \in \Theta_G\}.$$

PROOF. By A2 and A3, K is finite and continuous on Θ. Assumption A4, the well-known inequality between the affinity and the divergence, entails

$$\lim_{\|\theta\|\to\infty} K(\theta) \geq \lim_{\|\theta\|\to\infty} -2\log\rho(\theta) = +\infty. \tag{2.10}$$

Therefore K reaches its global minimum on a bounded set Θ_G. As Θ is closed and K is continuous, the set Θ_G is compact.

Without loss of generality we suppose $0 \in \Theta_G$.

We introduce the notation

$$I(\vartheta \in A) := I_A(\vartheta) = \begin{cases} 1, & \vartheta \in A, \\ 0, & \vartheta \notin A, \end{cases} \tag{2.11}$$

$$Z_n(\theta) := \prod_{i=1}^{n} p_\theta(X_i)/g(X_i), \tag{2.12}$$

$$K_0 := K(0). \tag{2.13}$$

Given $\varepsilon, \delta, \eta > 0$ we consider three parts of Θ

$$\Theta_\varepsilon^\delta = \{\theta \in \Theta | d_G(\theta) \geq \varepsilon, \|\theta\| \leq \delta\},$$
$$S_\delta^c = \{\theta \in \Theta | \, \|\theta\| > \delta\},$$
$$S_\eta = \{\theta \in \Theta | \, \|\theta\| \leq \eta\},$$

where δ is chosen large enough to have $\rho(\theta)$ small enough on $\Theta_\varepsilon^\delta$ [see (A.11) in Appendix A.1]. The following inequality obviously holds:

$$\begin{aligned} E_{n,\xi,\omega}(d_G^p(\theta)) &= E_\xi\left[d_G^p(\theta) Z_n(\theta)/E_\xi Z_n(\theta)\right] \\ &\leq \varepsilon^p + \left[(A_n + B_n)/C_n\right], \end{aligned} \tag{2.14}$$

where

$$A_n = E_\xi\left[\|\theta\|^p I(\theta \in \Theta_\varepsilon^\delta) Z_n(\theta)\right], \tag{2.15}$$

$$B_n = E_\xi\left[\|\theta\|^p I(\|\theta\| > \delta) Z_n(\theta)\right], \tag{2.16}$$

$$C_n = E_\xi\left[I(\|\theta\| < \eta) Z_n(\theta)\right]. \tag{2.17}$$

Using S_η we assure that the denominator of the posterior distribution (1.6) is not too small, proving for some $\alpha > 0$,

$$\lim_{n\to\infty} \exp(n(K_0 + [\alpha/2]))C_n = \infty \quad \text{a.s. (Lemma A.2)}. \tag{2.18}$$

On $\Theta_\varepsilon^\delta$ the tools of our proof are developed in analogy to a proof of Pfanzagl (1969) for the consistency of minimum contrast estimators under a compact parameter space and we show for conveniently chosen δ,

$$\lim_{n\to\infty} \exp(n(K_0 + 2\alpha))A_n = 0 \quad \text{a.s. (Lemma A.3)}. \tag{2.19}$$

On S_δ^c we adapt the method of Ibragimov and Has'minskii [(1981) pages 42–45, referred to henceforth as I–H], and get

$$\lim_{n\to\infty} \exp(n(K_0 + \alpha))B_n = 0 \quad \text{a.s. (Lemma A.8)}. \tag{2.20}$$

Indeed for (2.20), up to now, we do not know a proof relying directly on K–L divergence.

These convergences together with (2.14) lead to proposition (2.9). □

REMARK 3. An obvious consequence of (2.9) is on one hand the a.s. weak concentration of the posterior on Θ_G,

$$\lim_{n\to\infty} P_{n,\xi,\omega}(U) = 1 \tag{2.21}$$

for all open sets U containing Θ_G. Such a result has been obtained by Berk (1966) under a different set of somewhat weaker, more implicit assumptions. On the other hand a.s. the accumulation points P (w.r.t. weak convergence) of the sequence $\{P_{n,\xi,\omega}\}_{n\in N}$ of posterior distributions are p.d.'s P with carrier Θ_G. When the pseudotrue parameter θ_G is unique, we obviously have the weak convergence

$$P_{n,\xi,\omega} \Rightarrow \delta_{\theta_G},$$

where δ_t is the Dirac measure at point t.

REMARK 4. It is easy to see from the reasoning in the proof of Theorem 2.1 that a posterior mode $\tilde{\theta}_n$ (at which the posterior density is maximal over Θ) is strongly consistent in the sense

$$\lim_{n\to\infty} d_G(\tilde{\theta}_n) = 0 \quad \text{a.s.} \tag{2.22}$$

For the special case of the Lebesgue measure ξ on $\Theta = R^d$ as the prior, a posterior mode is a pseudo-MLE. We obtain the strong consistency (2.22) of a MLE $\tilde{\theta}_n$ for possibly nonunique pseudotrue parameters. The posterior mode may be interpreted as minimum contrast or M-estimator and the rich literature for such estimators gives conditions for its strong consistency. For example, the results of Huber (1967) hold for possibly nonunique pseudotrue parameters under weaker assumptions than those of Theorem 2.1.

REMARK 5. In the one-dimensional case and for a unique pseudotrue parameter θ_G a further consequence of Theorem 2.1 [or of (2.21)] is the strong consistency of the posterior median $\hat{\theta}_n$, which is pseudo-Bayes w.r.t. the loss $|t - \theta|$.

In Theorem 2.2 we show that pseudo-Bayes estimators are strongly consistent under assumptions A1 to A6.

THEOREM 2.2. *If the pseudotrue parameter θ_G is unique, it holds under the assumptions* A1 *to* A6 *that for all pseudo-Bayes estimators $\hat{\vartheta}_n$ w.r.t. a loss function L,*

$$\lim_{n\to\infty} \hat{\vartheta}_n = \vartheta_G \quad a.s. \tag{2.23}$$

PROOF. Because of A.6 and (1.5) we have a.s.

$$E_{n,\xi,\omega}L(\hat{\vartheta}_n(\omega),\vartheta) \le E_{n,\xi,\omega}L(\vartheta_G,\vartheta) \le c_3 E_{n,\xi,\omega}\|\vartheta_G - \vartheta\|^{b_5}. \tag{2.24}$$

With (2.9) we then obtain for almost all $\omega \in \Omega$,

$$\lim_{n\to\infty} E_{n,\xi,\omega}L(\hat{\vartheta}_n(\omega),\vartheta) = \lim_{n\to\infty} E_{n,\xi,\omega}L(\vartheta_G,\vartheta) = L(\vartheta_G,\vartheta_G) = 0. \tag{2.25}$$

Now assume that for such an ω there were an $\varepsilon \in (0, c)$ where

$$c := \left[c_1^{-1}c_2\right]^{1/b_4} \tag{2.26}$$

and a subsequence $\{n_i\}$ so that

$$\|\hat{\vartheta}_{n_i}(\omega) - \vartheta_G\| > \varepsilon, \qquad i = 1, 2 \dots . \tag{2.27}$$

This would lead to a contradiction to (2.25) using (2.9):

$$\begin{aligned}
&\limsup_{i\to\infty} E_{n_i,\xi,\omega}L(\hat{\vartheta}_{n_i}(\omega),\vartheta)\\
&\quad\ge \limsup_{i\to\infty} E_{n_i,\xi,\omega}\left\{\left(c_1\|\hat{\vartheta}_{n_i}(\omega) - \vartheta\|^{b_4}\right) \wedge c_2\right\}\\
&\quad\ge \limsup_{i\to\infty} E_{n_i,\xi,\omega}\left\{\left(c_1[\varepsilon - \|\vartheta - \vartheta_G\|]^{b_4}\right) \wedge c_2\right\}\\
&\quad= c_1 \limsup_{i\to\infty} E_{n_i,\xi,\omega}[\varepsilon - \|\vartheta - \vartheta_G\|]^{b_4}\\
&\quad\ge c_1 \limsup_{i\to\infty}\left[\varepsilon - \left(E_{n_i,\xi,\omega}\|\vartheta - \vartheta_G\|^{b_4}\right)^{1/b_4}\right]^{b_4} = c_1\varepsilon^{b_4} > 0.
\end{aligned} \tag{2.28}$$

Consequently, $\lim_{n\to\infty}\|\hat{\vartheta}_n(\omega) - \vartheta_G\| = 0$ must hold for almost all $\omega \in \Omega$. □

EXAMPLE 3. Berk (1970) ends his paper by leaving as an open problem the behavior of the posterior probability in a possibly incorrect location-scale Cauchy model:

$$p_\theta(x) = \left(\sigma\pi\left(1 + \frac{(x-\mu)^2}{\sigma^2}\right)\right)^{-1}, \qquad \theta = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+. \tag{2.29}$$

We show below that our results work in this special case. Let us assume the true distribution G of the sample to be absolutely continuous, with a positive, bounded density g ($\sup\{|g(x)| \,|\, x \in \mathbb{R}^1\} < \infty$), such that for some positive constants $0 < \alpha < 1$, $\delta > 0$ the bound $\int |x|^\delta g^{1-\alpha}(x)\,dx < \infty$ holds.

The following parametrization,

$$s = \begin{cases} \sigma - 1, & \sigma \ge 1,\\ 1 - \dfrac{1}{\sigma}, & \sigma \le 1, \end{cases} \tag{2.30}$$

leads to the new parameter $t = (\mu, s)$ and to the parameter space $\mathbb{R}^2$. Condition A1 is fulfilled. As g is bounded, some calculus with the Cauchy density show that assumption A2 is fulfilled. A3 is obviously satisfied.

For the proof of condition A4 we set $\delta^* = \inf(\delta/4, \alpha/16)$ and

$$(2.31) \quad r(\theta) = |\mu|^{\delta^*}\varrho(\theta) = \int |\mu|^{\delta^*}\sigma^{-1/2}\left[1 + (x-\mu)^2/\sigma\right]^{-1/2}\sqrt{g(x)}\,dx.$$

In Lemma A.9 we prove that

$$r(\theta) < \begin{cases} C\sigma^{-\alpha/16}, & \text{for } \sigma \geq 1, \\ C\sigma^{1/4}, & \text{for } \sigma \leq 1. \end{cases}$$

Here C is a positive constant not depending on μ and σ. Then it is easy to prove that A4 holds.

Summarizing, we have seen that all assumptions of Theorem 2.1 concerning the parametric model are fulfilled, so that we have the asymptotic behavior of the posterior given by (2.9). Moreover, if the true density is symmetric and strictly unimodal, the pseudoparameter θ_G is uniquely determined. Therefore the Bayes estimators w.r.t. convenient loss functions and a prior satisfying assumptions A5 and A6 are strongly consistent for $\theta_G = (\mu_G, \sigma_G)$.

3. Consistency of MLE versus inconsistency of Bayes estimators. In this section we want to show by an example that pseudo-MLE and pseudo-Bayes estimators are not always asymptotically equivalent, the MLE being strongly consistent, while the Bayes estimator is not. In this example, the Bayes estimates fluctuate a.s. around an interval; only the end points of it are pseudotrue parameters, but a.s. the interval is the set of accumulation points of the sequence of Bayes estimates $\{\hat{\theta}_n\}_{n \in N}$.

We assume a normal model

$$(3.1) \qquad P_\theta = N(\theta, v(\theta)), \qquad v(\theta) = a + b\theta^2,$$

in which the variance v depends on the mean θ. We assume a true distribution $G = N(0, \sigma^2)$ and

$$(3.2) \qquad 0 < a < \sigma^2, \qquad b > a/(\sigma^2 - a).$$

The prior distribution ξ is assumed to fulfill A5. We obtain the K–L divergence

$$(3.3) \qquad K(\theta) + E \log g(X) + \tfrac{1}{2}\log(2\pi) + \tfrac{1}{2}f(\theta),$$

where

$$(3.4) \qquad f(\theta) = \log v(\theta) + (\sigma^2 + \theta^2)/v(\theta).$$

Simple calculus shows that f is minimal at $\theta = \gamma$ and $\theta = -\gamma$, where

$$(3.5) \qquad \gamma = b^{-2}(b\sigma^2 - ba - a) > 0.$$

Therefore we have $\Theta_G = \{\gamma, -\gamma\}$.

Because assumptions A1–A5 as in Theorem 2.1 are fulfilled for our example, it holds that a.s for all $p > 0$,

$$\lim_{n \to \infty} E_{n,\xi,\omega}(\min\{|\theta - \gamma|, |\theta - (-\gamma)|\})^p = 0. \tag{3.6}$$

Moreover we are able to prove (see Appendix A.3) the following stronger asymptotic properties for the posterior.

1. For almost all ω the set of accumulation points (in the sense of weak convergence) of the sequence $\{P_{n,\xi,\omega}\}_{n \in N}$ of posterior distributions is the set of all the mixtures

$$\mu_\lambda = \lambda\delta_\gamma + (1 - \lambda)\delta_{-\gamma}, \qquad \lambda \in [0, 1]. \tag{3.7}$$

   This is also true for the "strong" convergence $P_n \to P$ of probability measures defined by the simultaneous validity of weak convergence $P_n \Rightarrow P$ and the convergence of all absolute moments of order $p > 0$,

$$\lim_{n \to \infty} \int |\theta|^p \, dP_n(\theta) = \int |\theta|^p \, dP(\theta). \tag{3.8}$$

2. We introduce for all $p > 0$ the "distances"

$$d_p(P_{n,\xi,\omega}, \delta_\gamma) = E_{n,\xi,\omega}|\theta - \gamma|^p \tag{3.9}$$

   between the posterior and the Dirac measure at γ. Obviously (3.9) is the infimum of $E_W|\theta - \tau|^p$ w.r.t. all joint distributions W of (θ, τ) with the marginal distributions $P_{n,\xi,\omega}$ and δ_γ. This defines the Mallows distances of order p; see Bickel and Freedman (1981). These distances and the distances $d_p(P_{n,\xi,\omega}, \delta_{-\gamma})$ will be small, each approximately with probability $\frac{1}{2}$; that is, for all sufficiently small $\varepsilon > 0$ holds that

$$\begin{aligned}\lim_{n \to \infty} P\big(\{\omega | d_p(P_{n,\xi,\omega}, \delta_\gamma) < \varepsilon\}\big) &= \lim_{n \to \infty} P\big(\{\omega | d_p(P_{n,\xi,\omega}, \delta_{-\gamma}) < \varepsilon\}\big) \\ &= \tfrac{1}{2}.\end{aligned} \tag{3.10}$$

A consequence of (1) is the inconsistency of the posterior mean $\hat{\theta}_n = E_{n,\xi,\omega}(\theta)$, because the sequence $\{\hat{\theta}_n\}_{n \in N}$ will a.s. have all points in the interval $[-\gamma, -\gamma]$ as accumulation points. On the other hand, a consequence of (2) (see also Appendix A.3) is also that the posterior mean $\hat{\theta}_n$ will be weakly consistent $(p - \lim_{n \to \infty} d_G(\hat{\theta}_n) = 0)$ and that the distribution of $\hat{\theta}_n$ (without standardization) tends weakly to the mixture $\mu_{1/2}$.

Furthermore, as discussed in Remark 4 (Section 2), the posterior mode $\tilde{\theta}_n$ and the maximum likelihood estimator $\tilde{\theta}_n = \hat{\theta}_{\mathrm{MLE}}$ (as its special case for the uniform prior $\xi(\theta) = \text{const.}$) are strongly consistent in the sense of

$$\lim_{n \to \infty} d_G(\tilde{\theta}_n) = 0 \quad \text{a.s.} \tag{3.11}$$

It is interesting that, as we show in Appendix A.3, the distribution of $\tilde{\theta}_n$ converges weakly to the mixture $\mu_{1/2}$.

4. Asymptotic normality of pseudo-Bayes estimators.

4.1. *Assumptions.* In the following, we state further conditions for the asymptotic normality of a pseudo-Bayes estimator, which complement the conditions for consistency.

A7: The pseudotrue value θ_G is unique and belongs to the interior Θ^{int} of Θ.

A8: The function $l(x, \vartheta) := \log[p_\vartheta(x)/g(x)]$ has for fixed $x \in R^k$ continuous derivatives of second-order w.r.t. ϑ in the interior Θ^{int} of Θ,

$$l'(x, \vartheta) := \frac{\partial}{\partial \vartheta} l(x, \vartheta), \qquad l''(x, \vartheta) = \frac{\partial^2}{\partial \vartheta^2} l(x, \vartheta). \tag{4.1}$$

Moreover, there is a positive function C on R^k and a positive integer b_6 with $EC(X) < \infty$,

$$\|l''(x, \vartheta) - l''(x, t)\| \le C(x)\left[1 + \|\vartheta\|^{b_6} + \|t\|^{b_6}\right]\|\vartheta - t\|, \tag{4.2}$$

$$\|l''(x, \vartheta)\| \le C(x)\left[1 + \|\vartheta\|^{b_6+1}\right], \qquad \vartheta, t \in \Theta^{\text{int}}. \tag{4.3}$$

where $\|\cdot\|$ denotes the Euclidian norm on R^d, or the analogous norm on the set of $d \times d$ matrices.

A9: We assume the integrals

$$I(\vartheta) := El'(x, \vartheta)[l'(x, \vartheta)]^T, \tag{4.4}$$

$$M(\vartheta) := -El''(x, \vartheta) = -\int l''(x, \vartheta)\, dG(x) \tag{4.5}$$

to exist and to be positive definite matrices in a neighborhood of $\vartheta = \vartheta_G$.

A10: The loss function L has in Θ^{int} continuous partial derivatives

$$L^{(i,j)}(\vartheta, t) := \frac{\partial^{i+j}}{\partial \vartheta^i\, \partial t^j} L(\vartheta, t), \qquad i, j = 1, 2. \tag{4.6}$$

Moreover we assume with $c, b_7 > 0$ and for $i, j = 1, 2$,

$$\|L^{(i,j)}(\vartheta, t)\| \le c(1 + \|\vartheta\|^{b_7} + \|t\|^{b_7}), \qquad \vartheta, t \in \Theta^{\text{int}}. \tag{4.7}$$

A11: The prior measure ξ has a density f w.r.t. the Lebesgue measure on R^d, which is continuous on R^d and fulfills for $b_8 > 0$,

$$0 < f(\vartheta) < c(1 + \|\vartheta\|^{b_8}), \qquad \vartheta \in \Theta. \tag{4.8}$$

4.2. *Asymptotic normality.*

THEOREM 4.1. *Under assumptions* A1–A11, *a pseudo-Bayes estimator* $\hat{\vartheta}$ *is asymptotically normal*:

$$\mathscr{L}\left\{\sqrt{n}\left(\hat{\vartheta} - \vartheta_G\right)\right\} \to N(0, \Lambda) \quad as\ n \to \infty, \tag{4.9}$$

where

$$\Lambda = L_2^{-1} L_1 M^{-1} I_G M^{-1} \left(L_2^{-1} L_1\right)^T, \tag{4.10}$$

$$I_G = I(\vartheta_G), \qquad M = M(\vartheta_G), \tag{4.11}$$

$$L_1 = L^{(1,1)}(\vartheta_G, \vartheta_G), \qquad L_2 = L^{(2,0)}(\vartheta_G, \vartheta_G). \tag{4.12}$$

PROOF. In the following we give a sketch of the proof, which is presented in detail in Appendix A.4. We assume w.l.o.g. that $\vartheta_G = 0$.

The Bayes estimator minimizes the posterior loss [see (1.7)]. In Lemma A.10, the expansion of the posterior loss gives us an asymptotic equivalent term $L_2^{-1} L_1 n^{1/2} A(Z_n) + G_n$ of the Bayes estimator [G_n is defined in (A.59)]. By a truncation and a dilatation argument applied on Z_n, we replace Z_n by the process Z_n^* [see (A.67)] lying in the space $\mathscr{C}$ of the continuous functions on R^d with zero limit at infinity. In Lemma A.11 we show that we can replace $n^{1/2}A(Z_n)$ by $A(Z_n^*)$ up to residual $o_P(1)$. The sequence of processes $(Z_n^*)_{n \in \mathbb{N}}$ converges in distribution in $\mathscr{C}$ to the exponential of a Gaussian process Y (Lemmas A.12, A.13, A.14).

Finally, we show that $(A(Z_n^*))_{n \in \mathbb{N}}$, thus $(A(Z_n))_{n \in \mathbb{N}}$, converges in distribution to $A(Y)$ and that G_n vanishes in probability when $n \to \infty$. The distribution of $L_n^{-1} L_1 n^{1/2} A(Y)$ is $N(0, \Lambda)$. □

REMARK 6. Extending Remark 1, we see that, for example, the normal model fulfills the assumptions of Theorem 4.1 if the true distribution G has the properties stated in Remark 1. Therefore we obtain the asymptotic normality of a Bayes estimator (e.g., a posterior mean) even if the model is incorrect. The results in the literature (see, e.g., I–H) state the asymptotic normality under the assumption of a correct model. The asymptotic covariance matrix Λ in (4.9) will in general be different from that under the correctness assumption.

If asymptotically valid statements such as the accuracy of $\hat{\theta}_n$ or asymptotical confidence regions for the pseudotrue parameter θ_G are wanted, the asymptotic covariance matrix for a pseudo-Bayes estimator $\hat{\theta}_n$ that is given in (4.10) and depends on the unknown distribution G must be estimated consistently.

Because of assumptions A8 and A10, the functions I, M and (for fixed $\omega \in \Omega$),

$$\hat{I}_n(\theta) = \hat{I}_{n,\omega}(\theta) := \frac{1}{n} \sum_{i=1}^{n} l'(X_i(\omega), \theta)\left[l'(X_i(\omega), \theta)\right]^T$$

and

$$\hat{M}_n(\theta) = \hat{M}_{n,\omega}(\theta) = -\frac{1}{n} \sum_{i=1} l''(X_i(\omega), \theta)$$

are continuous on Θ. Moreover by A8 and A9 and the strong law of large numbers, we have for some bounded neighborhood U of θ_G,

$$\lim_{n\to\infty} \sup_{\theta\in U} |\hat{I}_n(\theta) - I(\theta)| = 0 \quad \text{a.s.} \tag{4.13}$$

Theorem 2.2 gives $\lim_{n\to\infty} \hat{\theta}_n = \theta_G$ a.s. Then the estimator $\hat{I} = \hat{I}_n(\hat{\theta}_n)$ is strongly convergent to I_G. Analogously the estimators $\hat{M} = \hat{M}_n((\hat{\theta}_n)$, $\hat{L}_1 = L^{(1,1)}(\hat{\theta}_n, \hat{\theta}_n)$ and $\hat{L}_2 = L^{(2,0)}(\hat{\theta}_n, \hat{\theta}_n)$ are strongly consistent for M, L_1 and L_2, respectively.

As a consequence of Theorem 4.1, we obtain a standard pivot $T_n := \sqrt{n}\,\hat{\Lambda}^{-1/2}(\hat{\theta}_n - \theta)$, where

$$\hat{\Lambda} = \hat{L}_2^{-1}\hat{L}_1\hat{M}^{-1}\hat{I}\hat{M}^{-1}\left(\hat{L}_2^{-1}\hat{L}_1\right)^T.$$

From the pivot T_n we may derive asymptotic confidence regions and a test of a simple hypothesis on θ_G in the familiar way.

THEOREM 4.2. *Under assumptions* A1–A11 *follows*

$$\mathscr{L}\{T_n\} \underset{n\to\infty}{\to} N(0, I_d).$$

APPENDIX

A.1. Proof of Theorem 2.1.

LEMMA A.1. *For all $\varepsilon > 0$ it holds that*

$$K_\varepsilon = \inf\{K(\theta) | \theta \in \Theta, d_G(\theta) \ge \varepsilon\} > K_0 := K(0).$$

PROOF. The inequality (2.10), the compactness of Θ_G and the continuity of K entail Lemma A.1. □

Now we choose

$$\alpha = (K_\varepsilon - K_0)/4 \quad \text{for a fixed } \varepsilon > 0. \tag{A.1}$$

LEMMA A.2. *There is a $\eta > 0$ with*

$$\lim_{n\to\infty} \exp(n(K_0 + [\alpha/2]))C_n = \infty \quad a.s.$$

PROOF. With the notations of assumption A8 and setting $l_i(\theta) = l(X_i, \theta)$, we have

$$\begin{aligned} \lambda_n &:= \frac{1}{n}\inf\left\{\sum_{i=1}^n l_i(\vartheta) | \|\vartheta\| < \eta\right\} \\ &\ge -\frac{1}{n}\sum_{i=1}^n \sup\{\log[g(X_i)/p_\vartheta(X_i)] | \|\vartheta\| < \eta\}. \end{aligned} \tag{A.2}$$

Taking $\vartheta = 0$ and $\eta = \eta(0, \alpha/4)$ in (2.8) and applying the strong law of large numbers on the right-hand side of (A.2) yields

$$\liminf_{n\to\infty} \lambda_n \geq -K_0 - \alpha/4 \quad \text{a.s.} \tag{A.3}$$

Then Fatou's lemma gives

$$\begin{aligned} &\liminf_{n\to\infty} \exp(n(K_0 + [\alpha/2]))C_n \\ &\quad \geq E_\xi\Big[I(\|\vartheta\| < \eta)\liminf_{n\to\infty}\exp\big(n(K_0 + [\alpha/2] + \lambda_n)\big)\Big] = \infty \quad \text{a.s.} \end{aligned} \tag{A.4}$$ □

We set diam $(\Theta_G) = \max\{\|\theta - \theta'\| \,|\, \theta, \theta' \in \Theta_G\}$.

LEMMA A.3. *For each $\delta > \varepsilon + \operatorname{diam}(\Theta_G)$ it holds that*

$$\lim_{n\to\infty} \exp(n(K_0 + 2\alpha))A_n = 0 \quad a.s.$$

PROOF. For each $\vartheta \in \Theta_\varepsilon^\delta = \{\vartheta \in \Theta | d_G(\vartheta) \geq \varepsilon, \|\vartheta\| \leq \delta\}$, (2.8) provides a sphere $S(\vartheta, \eta)$ ($\eta = \eta(\vartheta, \alpha)$) with

$$E \sup\{l_i(t) | t \in S[\vartheta, \eta]\} \leq -K(\vartheta) + \alpha. \tag{A.5}$$

The set $\Theta_\varepsilon^\delta$ may be covered by a finite number of such spheres, say $S(\vartheta_j, \eta_j)$ ($j = 1, \dots, l$). With the notation

$$\lambda_{ij} := \sup\{l_i(t) | t \in S[\vartheta_j, \eta_j] \cap \Theta_\varepsilon^\delta\}, \tag{A.6}$$

the strong law of large numbers, (A.5) and (A.1) give for all $j = 1, \dots, l$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \lambda_{ij} \leq -K(\vartheta_j) + \alpha < -K_0 - 3\alpha \quad \text{a.s.} \tag{A.7}$$

This and the inequality

$$\zeta_n := \frac{1}{n}\sup\left\{\sum_{i=1}^n l_i(\vartheta) | \vartheta \in \Theta_\varepsilon^\delta\right\} \leq \frac{1}{n}\sup_j \sum_{i=1}^n \lambda_{ij} \tag{A.8}$$

lead to

$$\limsup_{n\to\infty} \zeta_n \leq -K_0 - 3\alpha \quad \text{a.s.} \tag{A.9}$$

Therefore,

$$\begin{aligned} &\limsup_{n\to\infty} \exp(n(K_0 + 2\alpha))A_n \\ &\quad \leq \limsup_{n\to\infty} \xi(S[0, \delta])\delta^q \exp\big(n(K_0 + 2\alpha + \zeta_n)\big) = 0 \quad \text{a.s.} \end{aligned} \tag{A.10}$$ □

Let $\delta_* = d^{-1/2}\delta - 1$; next we choose δ such that $\delta > (\varepsilon + \operatorname{diam}(\Theta_G)) \vee 2\sqrt{d}$ and

$$\begin{aligned} \sup\{\varrho(\vartheta) | \vartheta \in \Theta: \|\vartheta\| \geq \delta_*\} &\leq c(\delta_*)^{-b_1} \\ &\leq 2^{-b_1}\exp(-2d_1(K_0 + 2\alpha)). \end{aligned} \tag{A.11}$$

We introduce the notation $d_1 := d + 1$ and

$$V_n(t) := \left[Z_n(t/\sqrt{n})\right]^{1/2}, \qquad W_n = V_n^{1/d_1}, \tag{A.12}$$

$$\Theta_n := \left\{\vartheta\sqrt{n} \,|\, \vartheta \in \Theta, \|\vartheta\| > \delta_*\right\}. \tag{A.13}$$

The proof of (2.20) is now performed in successive steps and can be roughly described as follows.

Changing variable θ into $t/\sqrt{n}$, we adapt the setting of I–H (pages 42–45). Our goal is to control the trajectories of the likelihood on Θ_n. First we majorize the expectation of $W_n(t)$ using the affinity lemma (A.4). The true distribution does not belong to the parametric family and therefore we have to modify the method of I–H to majorize the continuity modulus of the stochastic process $W_n(t)$ (Lemmas A.4 through A.6). From Lemmas A.4 and A.6, we derive an upper bound for the likelihood trajectories on the pavements of Θ_n. Then a reasoning based on the Borel–Cantelli lemma leads finally to the desired convergence (2.20).

LEMMA A.4. *For each $q > 0$, there is an $n_q > 0$ such that, for $n > n_q$ and $t \in \Theta_n$, it holds that*

$$E(V_n(t)) = E\left[W_n(t)\right]^{d_1} \le \exp(-2nd_1(K_0 + \alpha))\|t\|^{-q}. \tag{A.14}$$

PROOF. Because of assumption A4 and (A.11), we have for $t \in \Theta_n$,

$$E\left[W_n(t)\right]^{d_1} \le \left[\varrho(t/\sqrt{n})\right]^n \le c^n\left[\|t\|/\sqrt{n}\right]^{-nb_1}. \tag{A.15}$$

Taking logarithms, we see that for n large enough, say $n > n_q$, the right-hand side of (A.15) is smaller than that of (A.14) if $n > n_q$. □

Let us set

$$\Delta(t, h) := E|W_n(t + h) - W_n(t)|^{d_1}. \tag{A.16}$$

LEMMA A.5. *For each $q > 0$ there are constants r and $m_q > n_q$ such that for $n > m_q$, $\|h\| < \sqrt{d}$, t and $t + h$ in Θ_n, it holds that*

$$\Delta(t, h) \le r\|h\|^{d_1} \exp(-nd_1(K_0 + \alpha))\|t\|^{-qd_1}. \tag{A.17}$$

PROOF. Let $t, t + h \in \Theta_n$ and h be as in Lemma A.5 and fixed. We set

$$S_n(u) := n^{-1/2} \sum_{i=1}^{n} h^T l_i'(t_{n,u}), \tag{A.18}$$

where

$$t_{n,u} := (t + uh)/\sqrt{n}, \tag{A.19}$$

$$l_i(\vartheta) = \log(p_i(\vartheta)/g(X_i)), \qquad l_i'(\vartheta) = \frac{\partial}{\partial\vartheta} l_i(\vartheta). \tag{A.20}$$

Then we have

$$\Delta(t,h) = E\left|\int_0^1 S_n(u) W_n(t+uh)\,du\right|^{d_1} \tag{A.21}$$

and the by Hölder inequality and the Fubini theorem,

$$\begin{aligned}
\Delta(t,h) &\le \int_0^1 \{E|S_n(u)|^{d_1} V_n^{1/2}(t+uh)\} V_n^{1/2}(t+uh)\,du \\
&\le \int_0^1 \{E|S_n|^{2d_1} V_n\}^{1/2} \{EV_n\}^{1/2}\,du \\
&\le \int_0^1 \{E|S_n(u)|^{4d_1} V_n^2(t+uh)\}^{1/4} \{EV_n(t+uh)\}^{1/2}\,du.
\end{aligned} \tag{A.22}$$

The first expectation may be written for fixed $u \in [0,1]$ as the integral

$$E|S_n|^{4d_1} V_n^2 = \int |S_n(u)|^{4d_1} \zeta_n(dx) \tag{A.23}$$

w.r.t. the probability distribution ζ_n over $(R^k)^n$ given by

$$\zeta_n(dx) := \prod_{i=1}^n p_{t_{n,u}}(x_i)\,\mu(dx_i). \tag{A.24}$$

Then $S_n(u)$ is a sum of n terms that are independent under the p.d. ζ_n. Therefore the inequality

$$\left(\sum_{i=1}^n |b_i|\right)^s \le n^{s-1} \sum_{i=1}^n |b_i|^s, \qquad s \ge 1, \tag{A.25}$$

the inequality of Burkholder [see Hall and Heyde (1980), page 23] and (2.2) give, with positive constants ψ,

$$\begin{aligned}
&E|S_n(u)|^{4d_1} V_n^2(t+uh) \\
&\quad \le \psi \|h\|^{4d_1} n^{-2d_1} \left(\int \left[\sum_{i=1}^n \|l_i'(t_{n,u})\|^2 \right]^{2d_1} d\zeta_n \right) \\
&\quad \le \psi \|h\|^{4d_1} \int \|l_1'(t_{n,u})\|^{4d_1} p_{t_{n,u}}(x_1)\,\mu(dx_1) \\
&\quad \le \lambda \|h\|^{4d_1} \left(1 + \|t_{n,u}\|^{b_0}\right).
\end{aligned} \tag{A.26}$$

For the second expectation $EV_n(t+uh)$, we apply Lemma A.4 replacing its q by $2qd_1 + (b_0+1)/2$. Recall that we have $\|h\| \le \sqrt{d}$ and $n^{-1/2}\|t\| \ge 1$. If we choose n large enough (say $n \ge m_q$) such that the following inequalities $n \ge n_q$, $n \ge n_{2(d_1 q + b_0 + 1)}$ and $(1 + \|t_{n,u}\|^{b_0}) \le \|t\|^{b_0+1}$ for all $t \in \Theta_n$ hold, then we get the conclusion of the lemma. □

From Lemma A.5 we derive an estimate of the continuity modulus of W_n. Let $t^* \in Z^d$ be a vector with integer components. We set

$$\Pi_{t^*} = \{\theta | t_j^* \le \theta_j \le t_j^* + 1, j = 1, \dots, d\},$$
$$c(\Pi_{t^*}) = \min\{\max\{|\theta_j| \,|j = 1, \dots, d\} | \theta \in \Pi_{t^*}\}.$$

The set $C(k)$ consists of the "pavements" Π_{t^*} such that $c(\Pi_{t^*}) = k$ is called a k-covering. Its cardinality is

$$\#C(k) = (2(k+1))^d - (2(k))^d \le 2^d d(k-1)^{d-1}. \tag{A.27}$$

For all θ in a pavement Π_{t^*} of a k-covering, it holds that $k \le \|\theta\| \le (k+1)\sqrt{d}$. Hence Theorem 11.1 of Ledoux and Talagrand (1991) may be applied on such a pavement using (in their denotation) $\psi = \|\cdot\|_{d_1}$, $T = \Pi_{t^*}$, $d(s,t) = \|s - t\| \exp(-n(K_0 + \alpha))k^{-q}$ and we obtain the following lemma.

LEMMA A.6. *Under the previous conditions for each* $q > 0$, *for* $n \ge m_q$, $\|h\| \le \sqrt{d}$ *and for any pavement* Π_{t^*} *of a* k-*covering* $C(k)$ *it holds that*

$$E\omega_{k,n}(h) \le c_d \exp(-n(K_0 - \alpha))k^{-q} h^{1/d_1}, \tag{A.28}$$

where

$$\omega_{k,n}(h) = \sup\{|W_n(t+h) - W_n(t)| \,|t, t+h \in \Pi_{t^*} \cap \Theta_n\},$$
$$c_d = 8(1 + \sqrt{d})^{d/d_1} d_1.$$

With the above lemmas we are now able to get a maximal inequality for the process W_n on Θ_n. We set

$$\begin{aligned} \Gamma_{k,n} &= \bigcup_{\Pi_{t^*} \in C(k)} \Pi_{t^*} \cap \Theta_n, \\ W_{k,n}^* &= \sup\{W_n(\theta) | \theta \in \Gamma_{k,n}\}. \end{aligned} \tag{A.29}$$

LEMMA A.7. *Under the previous conditions, given* ($\gamma \in (0,1)$), *we have*

$$P(W_{k,n}^* > \gamma) \le c_d' \gamma^{-d_1} \exp(-n(K_0 + \alpha))k^{-q}(k+1)^{2(d-1)} \tag{A.30}$$

with $c_d' = (2_1^d + c^d d^{1/2d_1})2^d d$.

PROOF. Let us fix a pavement $\Pi_{t^*} \in C(k)$ and $\gamma (0 < \gamma < 1)$ and we set

$$p(n, t^*) = P(\sup\{W_n(\theta) | \theta \in \Pi_{t^*} \cap \Theta_n\} > \gamma).$$

We consider a net ν consisting of the points τ_s of Θ_n,

$$\tau_s = t^* + k^{-1}s, \qquad s = (s_1, \dots, s_d), \qquad s_j = 0, \dots, k, \qquad j = 1, \dots, d.$$

Then it holds that

$$(\text{A.31}) \quad \begin{aligned} p(n,t^*) &\le P\left(\sup\{W_n(\theta)|\theta \in \nu\} > \frac{\gamma}{2}\right) + P\left(\omega_{k,n}(\sqrt{d}/k) > \frac{\gamma}{2}\right) \\ &= p_1 + p_2. \end{aligned}$$

For $m > m_q$ the inequality $p_1 \le \sum_{\theta \in \nu} P(W_n(\theta) > \gamma/2)$ and the Markov inequality with Lemma A.4 lead to

$$(\text{A.32}) \qquad p_1 \le (2/\gamma)^{d_1} \exp(-2nd_1(K_0 + \alpha))k^{-q}(k-1)^d.$$

Again the Markov inequality and Lemma A.6 give

$$(\text{A.33}) \qquad p_2 \le (2/\gamma)c_d \exp(-n(K_0 + \alpha))k^{-q}\left(\sqrt{d}/k\right)^{1/d_1}.$$

These inequalities yield

$$(\text{A.34}) \qquad p(n,t^*) \le p_1 + p_2 \le \pi(n,k,\gamma),$$

$$(\text{A.35}) \quad \pi(n,k,\gamma) = \left(2^{d_1} + 2c_d d^{1/2d_1}\right)\gamma^{-d_1} \exp(-n(K_0 + \alpha))k^{-q}(k-1)^d.$$

Obviously, we have $P(W^*_{k,n} > \gamma) \le \sum_{\Pi_{t^*} \in C(k)} p(n,t^*)$. The previous upper bound for $p(n,t^*)$ and the cardinality of $C(k)$ (see A.27) lead to the conclusion of the lemma. □

Finally, we are able to prove Lemma A.8.

LEMMA A.8. *Under the assumptions of Theorem* 2.1 *we have*

$$\lim_{n \to \infty} \exp(n(K_0 + \alpha))B_n = 0 \quad a.s.$$

PROOF. Let $\nu_n(dt)$ be the probability measure induced by $\xi(d\theta)$ and the transformation $t = \sqrt{n}\,\theta$. We have $B_n \le \sum_{k \ge \delta_* \sqrt{n}} b_{k,n}$ with

$$b_{k,n} := n^{-p/2} \int \|t\|^p W_n^{2d_1}(t) I(t \in \Gamma_{n,k}) \nu_n(dt).$$

Obviously it holds that

$$(\text{A.36}) \qquad b_{k,n} \le n^{-p/2}\left((k+1)\sqrt{d}\right)^p (W^*_{k,n})^{2d_1} \nu_n(\Gamma_{n,k}).$$

With assumption A5 we have the rough upper bounds

$$(\text{A.37}) \qquad \nu_n(\Gamma_{n,k}) \le c(k+1)^{d+b_3-1} n^{-b_2/2}$$

and

$$(\text{A.38}) \qquad b_{k,n} \le c^*(k+1)^{r_1} n^{-r_2} (W^*_{k,n})^{2d_1},$$

with the same constants $c, c^*, r_1 = p + d + b_3 - 1$, and $r_2 = (p + b_2)/2$.

Noting that

$$\sum_{k \geq \delta_* \sqrt{n}} (k+1)^{-2} \leq \left(\delta_* \sqrt{n}\right)^{-1},$$

we have the sequence of inclusions

$$\begin{aligned} &\left\{\exp(n(K_0+\alpha))B_n > \left(\delta_* \sqrt{n}\right)^{-1}\right\} \\ \text{(A.39)} \qquad &\subset \bigcup_{k \geq \delta_* \sqrt{n}} \left\{\exp(n(K_0+\alpha))b_{k,n} > (k+1)^{-2}\right\} \\ &\subset \bigcup_{k \geq \delta_* \sqrt{n}} \{W^*_{k,n} > \gamma_{k,n}\}, \end{aligned}$$

with $\gamma_{k,n} = \exp(-(n/2d_1)(K_0+\alpha))(k+1)^{-(2+r_1)/d_1} n^{r_2/d_1}$.

Using the inequality (A.7), for all $q > 0$ and $n \geq m_q$, an obvious calculation leads to

$$\begin{aligned} P(W^*_{k,n} > \gamma_{k,n}) \leq p_{k,n} &= c'_d \gamma_{k,n}^{-d_1} \exp(-n(K_0+\alpha))k^{-q}(k+1)^{2(d-1)} \\ &= c'_d \exp(-n/2(K_0+\alpha))n^{-r_2}k^{-q}(k-1)^{2+r_1+2(d-1)}. \end{aligned}$$

It is easy to see that the double series $\Sigma_{k,n} P_{k,n}$ is convergent, if q is sufficiently large. Therefore by the Borel–Cantelli theorem, for almost all ω, there is only a finite set of pairs (n_i^ω, k_i^ω) for which the inequality $\exp(n(K_0+\alpha))b_{k_i,n_i}(\omega) \geq (k_i+1)^{-2}$ holds. Then for $n > n(\omega) = \sup_i n_i^\omega$ we have $\exp(n(K_0+\alpha))b_{k,n}(\omega) \leq (k+1)^{-2}$ for all k. Hence for almost all ω and for $n > n(\omega)$ we have $\exp(n(K_0+\alpha))B_n(\omega) \leq (\delta_*)^{-1}$, thus proving the lemma. □

A.2. Proof for Example 3 of Section 2. The density of the Cauchy distribution with location and scale parameters μ and σ is $\sigma^{-1}h((x-\mu)/\sigma)$ where $h(x) = \pi^{-1}(1+x^2)^{-1}$. We introduce the function

$$\text{(A.40)} \qquad r(\theta) := \int |\mu|^{\delta^*} \sigma^{-1/2} h^{1/2}\left(\frac{x-\mu}{\sigma}\right) g^{1/2}(x)\, dx,$$

where $\theta = (\mu, \sigma)$, and $\delta^* < \inf(\delta/4, \alpha/16)$. We prove Lemma A.9.

LEMMA A.9. *Under the assumptions of Example* 3, *Section* 2, *we have the following*:

(i) $r(\theta) < C\sigma^{-\alpha/16}$ *for* $\sigma > 1$;
(ii) $r(\theta) < C\sigma^{1/4}$ *for* $\sigma < 1$.

PROOF. As we have $\delta^* < 1$, we majorize $r(\theta)$ replacing in (A.44) $|\mu|^{\delta^*}$ by $|\mu - x|^{\delta^*} + |x|^{\delta^*}$. Next for all positive constants p, q ($p^{-1} + q^{-1} = 1$) using the Hölder inequality, we obtain $r(\theta) < r_1 + r_2$ where

$$r_1 = \sigma^{-1/2+1/p+\delta^*}\left(\int\left(\frac{|\mu - x|}{\sigma}\right)^{\delta^* p} h^{p/2}\left(\frac{\mu - x}{\sigma}\right)\frac{dx}{\sigma}\right)^{1/p}\left(\int g^{q/2}(x)\,dx\right)^{1/q},$$

$$r_2 = \sigma^{-1/2+1/p}\left(\int h^{p/2}\left(\frac{\mu - x}{\sigma}\right)\frac{dx}{\sigma}\right)^{1/p}\left(\int |x|^{\delta^* q} g^{q/2}(x)\,dx\right)^{1/q}.$$

Let us set for $\sigma > 1$, $q = 2 - \alpha/2$ and for $\sigma < 1$, $q = 4$. Then a tedious calculation leads to the following majorizations of r_1 and r_2:

$$r_1, r_2 < C\sigma^{-\alpha/16} \quad \text{if } \sigma > 1,$$
$$r_1, r_2 < C\sigma^{1/4} \quad \text{if } \sigma < 1. \qquad \square$$

A.3. Proof for the example of Section 3.

PROPERTY 1. We remark, that under (3.1) the posterior density $p_{n,\xi,\omega}$ (w.r.t. ξ) fulfills for all θ, n,

(A.41) $$p_{n,\xi,\omega}(\theta) = p_{n,\xi,\omega}(-\theta)\exp[S_n \tau(\theta)],$$

where

$$S_n(\omega) := \sum_{i=1}^{n} X_i(\omega), \qquad \tau(\theta) := 2\theta/v(\theta).$$

Obviously $\tau(\theta)$ is negative for $\theta < 0$ and positive for $\theta > 0$, vanishes for $\theta = 0$ and for $\theta \to \infty$ or $-\infty$, while it has maximum at $\theta = c := \sqrt{a/b}$ and its minimum at $\theta = -c$.

With the notation

$$\Theta_\varepsilon := \{\theta \in R^1 | d_G(\theta) < \varepsilon\}, \qquad \Theta_\varepsilon^+ = (\gamma - \varepsilon, \gamma + \varepsilon),$$
$$\Theta_\varepsilon^- = (-\gamma - \varepsilon, -\gamma + \varepsilon)$$

we obtain from (3.6) for almost all $\omega \in \Omega$ and all sufficiently small $\varepsilon > 0$,

(A.42) $$\lim_{n\to\infty} P_{n,\xi,\omega}(\Theta_\varepsilon) = \lim_{n\to\infty}\left[P_{n,\xi,\omega}(\Theta_\varepsilon^+) + P_{n,\xi,\omega}(\Theta_\varepsilon^-)\right] = 1.$$

For almost all ω the set of the accumulation points of the sequence $\{S_n(\omega)\}_{n\in\mathbb{N}}$ is $\overline{R} = R \cup \{-\infty\} \cup \{+\infty\}$ [see Chung (1976), page 272, Exercise 5,). In other words, there exists a set $\Omega^* \subset \Omega$, $P(\Omega^*) = 1$, such that for all $\omega \in \Omega^*$ and $s \in \overline{R}$, we can find a sequence $(n_{(j)} = n(j, \omega, s), j \in N)$ with (3.6), (A.42) and

(A.43) $$\lim_j S_{n_{(j)}}(\omega) = s.$$

Let $\underline{\tau}_\varepsilon = \min\{\tau(\theta) | \theta \in \Theta_\varepsilon^+\}$ and $\bar{\tau}_\varepsilon = \max\{\tau(\theta) | \theta \in \Theta_\varepsilon^+\}$. Assume as fixed a $\omega \in \Omega^*$. Then for each $\varepsilon > 0$ there is a subsequence $\{n_i\}$ of $\{n_{(j)}\}$ such that the convergence

(A.44) $$\begin{aligned} p_{\omega,s,\varepsilon} &:= \lim_{i\to\infty} P_{n_i,\xi,\omega}(\Theta_\varepsilon^+) \\ &\le \limsup_{i\to\infty} E_\xi\left[I(\theta \in \Theta_\varepsilon^+) p_{n_i}(-\theta)\right]\max\{\varepsilon^{s\tau(\theta)} | \theta \in \Theta_\varepsilon^+\} \end{aligned}$$

holds [see (A.41)] and therefore because of (A.42) we have

$$p_{\omega,s,\varepsilon} \le (1 - p_{\omega,s,\varepsilon})\max\{\exp(\underline{\tau}_\varepsilon s), \exp(\bar{\tau}_\varepsilon s)\}. \tag{A.45}$$

An analogous reasoning leads to

$$p_{\omega,s,\varepsilon} \ge (1 - p_{\omega,s,\varepsilon})\min\{\exp(\underline{\tau}_\varepsilon s), \exp(\bar{\tau}_\varepsilon s)\}. \tag{A.46}$$

As (A.45), (A.46) hold for all sufficiently small $\varepsilon > 0$, we have

$$p_{\omega,s} := \lim_{\varepsilon\to 0} p_{\omega,s,\varepsilon} = e^{s\tau(\gamma)}[1 + e^{st(\gamma)}]^{-1}. \tag{A.47}$$

Especially for $s = +\infty$ and $s = -\infty$ we have $p_{\omega,s} = 1$ and $p_{\omega,s} = 0$, respectively.

From (A.42) and (A.47) follows the weak convergence of the sequence $\{P_{n_i,\xi,\omega}\}_{i\in N}$ to the mixture $\mu_{p_{\omega,s}}$ defined by (3.7). For all $p > 0$ the inequality

$$\sup_{i>i_0} E_{n_i,\xi,\omega}\|\theta\|^p \le 2^{p-1}\left[(\sup\{\|\theta\| \,|\theta \in \Theta_G\})^p + \sup_{i>i_0} E_{n_i,\xi,\omega} d_G(\theta)^p\right] < \infty$$

holds for sufficiently large i_0 and we obtain from $P_{n_i,\xi,\omega} \Rightarrow \mu_{p_{\omega,s}}$,

$$\lim_{i\to\infty} E_{n_i,\xi,\omega}\|\theta\|^p = \int \|\theta\|^p \, d\mu_{p_{\omega,s}}(\theta).$$

Therefore we have also the strong convergence $P_{n_i,\xi,\omega} \to \mu_{p_{\omega,s}}$.

Thus we have shown that for almost all ω and each mixture μ_π with $\pi \in [0,1]$ there is a subsequence $\{n_i\}$ for which $P_{n_i,\xi,\omega}$ converges weakly and in all its abolute moments of order $p > 0$ to μ_π.

PROPERTY 2. Let ε and $\tilde{\varepsilon}$ be positive and sufficiently small. Then

$$S_n < -H := \underline{\tau}_{\tilde{\varepsilon}}^{-1} \log[\varepsilon/(1-\varepsilon)] < 0$$

has the consequence

$$\begin{aligned} p_{n\omega} &:= P_{n,\xi,\omega}(\Theta_{\tilde{\varepsilon}}^+) = E_\xi\{I(\theta \in \Theta_{\tilde{\varepsilon}}^+) p_{n,\xi,\omega}(-\theta)\exp(S_n\tau(\theta))\} \\ &\le (1 - p_{n\omega})\varepsilon/(1-\varepsilon) \end{aligned}$$

and therefore $p_{n\omega} \le \varepsilon$. We obtain

$$\begin{aligned} \liminf_{n\to\infty} P(p_{n\omega} \le \varepsilon) &\ge \liminf_{n\to\infty} P(S_n < -H) \\ &= \liminf_{n\to\infty} \Phi(-H/\sqrt{n}) = \tfrac{1}{2}. \end{aligned} \tag{A.48}$$

Analogously, it follows that

$$\liminf_{n\to\infty} P(p_{n\omega} \ge 1-\varepsilon) \ge \liminf_{n\to\infty} P(S_n > H) = \tfrac{1}{2}. \tag{A.49}$$

(A.48) and (A.49) together yield, for all $\varepsilon > 0$, the property

$$\lim_{n\to\infty} P(p_{n\omega} \le \varepsilon) = \lim_{n\to\infty} P(p_{n\omega} \ge 1-\varepsilon) = \tfrac{1}{2}. \tag{A.50}$$

The a.s. validity of (A.42) for sufficiently small ε has the consequence that (A.50) is also fulfilled for

$$\pi_{n\omega} := P_{n,\xi,\omega}((0,\infty)) = p_{n\omega} + P_{n,\xi,\omega}(\Theta_\varepsilon^c \cap (0,\infty))$$

in place of $p_{n\omega}$.

For the conditional posterior distributions $P^+_{n,\xi,\omega} := P(\cdot|\theta \in (0,\infty))$ and $P^-_{n,\xi,\omega} := P(\cdot|\theta \in (-\infty, 0])$, we have

$$E_{n,\xi,\omega} d_G(\theta)^p = \pi_{n\omega} d_p(P^+_{n,\xi,\omega}, \delta_\gamma) + (1 - \pi_{n\omega}) d_p(P^-_{n,\xi,\omega}, \delta_{-\gamma}),$$

$$d_p(P_{n,\xi,\omega}, d_r) = \pi_{n\omega} d_p(P^+_{n,\xi,\omega}, \delta_\tau) + (1 - \pi_{n\omega}) d_p(P^-_{n,\xi,\omega}, \delta_\tau), \qquad \tau = \gamma, -\gamma.$$

Because of

$$d_p(P^+_{n,\xi,\omega}, \delta_{-\gamma}) \geq \|\gamma\|^p, \qquad d_p(P^-_{n,\xi,\omega}, \delta_\gamma) \geq \|\gamma\|^p$$

(3.6) and (A.50) together with the above two equations imply (3.10) for all $\varepsilon, p > 0$, that is, Property 2.

Now we will see how the limit proeprty (A.50) influences the behavior of the posterior mean $\hat{\theta}_n = E_{n,\xi,\omega}(\theta)$. We have

$$\text{(A.51)} \qquad \hat{\theta}_n = p_{n,\omega} E_{n,\xi,\omega}(\theta | \theta \in \Theta_\varepsilon^+) + q_{n,\omega} E_{n,\xi,\omega}(\theta|\theta \in \Theta_\varepsilon^-) + r_n,$$

where ε is a sufficiently small positive number and

$$q_{n,\omega} := P_{n,\xi,\omega}(\Theta_\varepsilon^-), \qquad r_n = E_{n,\xi,\omega}\{\theta I(\theta \in \Theta_\varepsilon^c)\}.$$

Theorem 2.1 has the consequence

$$\text{(A.52)} \qquad \lim_{n\to\infty} |r_n| \leq \lim_{n\to\infty} \left[E_{n,\xi,\omega} d_G(\theta) + |\gamma| P_{n,\xi,\omega}(\Theta_\varepsilon^c)\right] = 0 \quad \text{a.s.}$$

With (A.51) and

$$\hat{\theta}_n \geq p_{n\omega}(\gamma - \varepsilon) + (1 - p_{n\omega})(-\gamma - \varepsilon) + r_n$$

we obtain from (A.50) and (A.52) that for sufficiently small ε,

$$\lim_{n\to\infty} \inf P(\hat{\theta}_n \in \Theta^+_{2\varepsilon}) \geq \lim_{n\to\infty} P(p_{n\omega} > 1 - \varepsilon(2\gamma)^{-1}) = \tfrac{1}{2}.$$

But we have analogously

$$\lim_{n\to\infty} \inf P(\hat{\theta}_n \in \Theta^-_{2\varepsilon}) \geq \tfrac{1}{2}.$$

Therefore $\lim_{n\to\infty} \inf$ is there just a convergence $\lim_{n\to\infty}$ and it follows for all open sets $U \subset R$,

$$\lim_{n\to\infty} P(\hat{\theta}_n \in U) = \mu_{1/2}(U).$$

From this we obtain the weak convergence $Q_n \Rightarrow \mu_{1/2}$ of the distribution Q_n of $\hat{\theta}_n$ to the mixture $\mu_{1/2}$ and the weak consistency $p - \lim d_G(\hat{\theta}_n) = 0$ of $\hat{\theta}_n$. The distribution P_n of the posterior mode $\tilde{\theta}_n$ also converges weakly to $\mu_{1/2}$.

This is seen in the following way:

From the convergence $\lim_{n\to\infty} P(\tilde{\theta}_n \in \Theta_\varepsilon) = 1$ for $\varepsilon > 0$, which follows from Remark 4, Section 2, and because of (A.41), we obtain

$$\lim_{n\to\infty} P\big(\tilde{\vartheta} \in \Theta_\varepsilon^+\big) \geq \lim_{n\to\infty} P\Big(\sup_{\vartheta\in\Theta_\varepsilon^+} p_{n,\xi,\omega}(\vartheta) > \sup_{\vartheta\in\Theta_\varepsilon^-} p_{n,\xi,\omega}(\vartheta)\Big) \geq \tfrac{1}{2}$$

and analogously $\lim_{n\to\infty} P(\tilde{\theta} \in \Theta_\varepsilon^-) \geq \frac{1}{2}$. Therefore we obtain for sufficiently small ε the convergences

$$\lim_{n\to\infty} P\big(\tilde{\theta}_n \in \Theta_\varepsilon^+\big) = \lim_{n\to\infty} P\big(\tilde{\theta}_n \in \Theta_\varepsilon^-\big) = \tfrac{1}{2},$$

thus proving $P_n \Rightarrow \mu_{1/2}$.

A.4. Proof of Theorem 4.1. We assume w.l.o.g., that $\vartheta_G = 0$.

The Bayes estimator minimizes the posterior loss. We perform an expansion of the posterior loss to get a simple expression which is asymptotically equivalent to the Bayes estimator. In the next lemma we give an equivalent term of the Bayes estimator.

LEMMA A.10. *Under assumptions* A1–A11, *the following expression holds*:

$$\sqrt{n}\,\hat{\theta}_n = L_2^{-1} L_1 \sqrt{n} A(Z_n) + G_n + o_P(1), \tag{A.53}$$

where G_n *is defined in* (A.59) *and with*

$$A(w) = \int t w(t)\,dt \Big/ \int w(gt)\,dt. \tag{A.54}$$

PROOF. Let $r(\theta)$ be the posterior loss [see (1.7)]. Then

$$r(\vartheta) := E_n L(\vartheta, \underline{t}) := E_{n,\xi,\omega}[L(\vartheta, \underline{t})] = \int L(\vartheta, t)\,dP_{n,\xi,\omega}(t). \tag{A.55}$$

Because of (2.9), assumption A10 gives the a.s. differentiability of r in a neighborhood of $\vartheta = 0$. The differentiation may be interchanged with the expectation. As $\hat{\vartheta} = \vartheta_n$ minimizes $r(\vartheta)$ with a Taylor expansion for $r(\cdot)$ we obtain, for some n_ω and $n \geq n_\omega$ almost surely,

$$\begin{aligned} 0 &= \sqrt{n}\, r'(\hat{\vartheta}) = \sqrt{n}\, E_n L^{(1,0)}(\hat{\vartheta}, \underline{t}) \\ &= E_n \sqrt{n}\, L^{(1,0)}(0, \underline{t}) + E_n \int_0^1 L^{(2,0)}(u\hat{\vartheta}, \underline{t})\,du \sqrt{n}\,\hat{\vartheta}. \end{aligned} \tag{A.56}$$

A further expansion gives us

$$0 = \sqrt{n}\, L^{(1,0)}(0,0) + F_n + G_n + (H_n + K_n)\sqrt{n}\,\hat{\vartheta} \tag{A.57}$$

with the reduction

$$F_n := L^{(1,1)}(0,0) E_n \sqrt{n}\,\underline{t} = L^{(1,1)}(0,0)\sqrt{n}\, A[Z_n], \tag{A.58}$$

$$G_n := \sqrt{n}\, E_n \int_0^1 L^{(1,2)}(0, u\underline{t})[\underline{t}, \underline{t}]\,du, \tag{A.59}$$

with

$$L^{(1,2)}(0,\tau)[t,t] := \frac{\partial}{\partial\vartheta} t^T L^{(1,1)}(0,\theta) t|_{\theta=\tau},$$

$$L^{(2,1)}(\tau,\tilde{\tau})[t,\tilde{t}] := \frac{\partial}{\partial\theta} t^T L^{(1,1)}(\theta,\tilde{\tau})\tilde{t}|_{\theta=\tau}. \tag{A.60}$$

$$H_n := \int_0^1 L^{(2,0)}(u\hat{\vartheta},0)\,du, \tag{A.61}$$

$$K_n := E_n \int_0^1 \int_0^1 L^{(2,1)}\big(u\hat{\vartheta}, v\underline{t}\big)\big[\sqrt{n}\,\hat{\theta}, \underline{t}\big]\,du\,dv. \tag{A.62}$$

The first term in (A.57) vanishes ($L^{(1,0)}(0,0) = 0$) because of assumption A7. By the dominated convergence theorem we have, from assumption A10 and Theorem 2.2,

$$\lim_{n\to\infty} H_n = L^{(2,0)}(0,0) \quad \text{a.s.} \tag{A.63}$$

Moreover we have, because of assumptions A10, (2.9) and Theorem 2.2,

$$\|K_n\| \le c'\Big[\big(1 + \|\hat{\vartheta}_n\|^{b_7}\big)E_n\|\underline{t}\| + E_n\|\underline{t}\|^{b_7+1}\Big]\sqrt{n}\,\|\hat{\theta}\| = o_p(1). \tag{A.64}$$

The lemma follows now from (A.57), (A.63) and (A.64).

By the way, let us note that analogously to (A.64) we have

$$\|G_n\| \le c\sqrt{n}\,E_n\big(\|\underline{t}\|^2 + \|\underline{t}\|^{b_7+2}\big). \tag{A.65}$$ □

Let us recall that $\theta_G = 0$. We denote by λ the smallest eigenvalue of the matrix M in assumption A9. We choose $\varepsilon > 0$ such that the sphere $S[0, 2\varepsilon]$ is included in Θ and

$$EC(X)\big[1 + (2\varepsilon)^{b_6}\big]\varepsilon \le \lambda/16. \tag{A.66}$$

First we replace the process Z_n by a more tractable process Z_n^*, with realizations in the space $\mathscr{C}$ of the continuous functions on R^d vanishing at infinity. Then Z_n^* is defined by

$$\begin{aligned} Z_n^*(t,\omega) &= Z_n^*(t) \\ &:= Z_n(t_n)\big\{a(t_n)f(t_n)[Z_n(0)f(0)]^{-1}\big\}, \qquad t \in R^d, \end{aligned} \tag{A.67}$$

with the scale transformation $t_n := n^{-1/2}t$, $t \in R^d$,

$$a(u) = \begin{cases} 1, & \text{for } \|u\| \le \varepsilon, \\ 2(1 - \|u\|)/2\varepsilon, & \text{for } \varepsilon < \|u\| < 2\varepsilon, \\ 0, & \text{for } \|u\| \ge 2\varepsilon. \end{cases} \tag{A.68}$$

LEMMA A.11. *Let h: $R^d \to R^k$ be a function and $q > 0$ with*

$$\|h\|_q := \sup_{t\in R^d} [1 + \|t\|^q]^{-1}\|h(t)\| < \infty. \tag{A.69}$$

Further let [*see* (1.4) *and* (A.67)]

$$(A.70)\qquad E_n h(\underline{t}) := \int h(t)\, dP_{n,\xi,\cdot}(dt),$$

$$(A.71)\quad E_n^*(h(\underline{t})) := \int h(t)\, dP_{n,\xi,\cdot}^*(t) = \int h(t) Z_n^*(t)\, dt \Big/ \int Z_n^*(t)\, dt.$$

where $P_{n,\xi,\cdot}$ *stands for the random probability measure* $\omega \to P_{n,\xi,\omega}$. *Then it holds that* [*see* (*A*.54)]

$$(A.72)\ \text{(i)}\qquad E_n^* h(\underline{t}) = E_n h(\sqrt{n}\,\underline{t}) + o_p(1),$$

$$\text{(ii)}\qquad A(Z_n^*) = A(Z_n) - o_P(1).$$

PROOF. We introduce the notation

$$(A.73)\quad \begin{aligned} N_{ij} &:= E_\xi h(\sqrt{n}\,\vartheta)^i Z_n(0)^{-1} \\ &\quad \times Z_n(\vartheta)[ja(\vartheta) + [1-j][1-a(\vartheta)]], \qquad i,j = 0,1. \end{aligned}$$

Then we may write

$$(A.74)\quad \begin{aligned} E_n h(\sqrt{n}\,\underline{t}) &= [N_{00} + N_{01}]^{-1}[N_{10} + N_{11}] \\ &= [N_{01}^{-1}N_{00} + 1]^{-1}[N_{01}^{-1}N_{10} + N_{01}^{-1}N_{11}], \end{aligned}$$

because of $N_{01} > 0$. We will now see that a.s. $N_{01}^{-1}N_{00}$ and $N_{01}^{-1}N_{10}$ tend to zero for $n \to \infty$. Choosing $\varepsilon, \alpha, \eta$ as in the proof of Theorem 2.1, we get

$$(A.75)\quad \begin{aligned} \|N_{0,1}^{-1}N_{10}\| &\le \|h\|_q [E_\xi I[\|\vartheta\| < \eta] Z_n(\vartheta)]^{-1} \\ &\quad \times E_\xi [I[\|\vartheta\| < \varepsilon](1 + \|\vartheta\|^q n^{q/2}) Z_n(\vartheta)] \end{aligned}$$

and the limits (2.18), (2.19), (2.20) allow choosing a positive k, so that a.s. for all $n \in N$,

$$(A.76)\qquad \|N_{01}^{-1}N_{10}\| \le k e^{-n\alpha/2} n^{q/2}.$$

This and a similar reasoning for $N_{01}^{-1}N_{00}$ prove the desired convergences

$$(A.77)\qquad \lim_{n\to\infty} N_{01}^{-1}N_{10} = 0, \qquad \lim_{n\to\infty} N_{01}^{-1}N_{00} = 0 \quad \text{a.s.}$$

From (A.74) and (A.77) and substituting $\sqrt{n}\,\vartheta$ by t, we finally arrive at

$$(A.78)\qquad E_n(\sqrt{n}\,\underline{t}) = N_{01}^{-1}N_{11} + o_p(1) = E_n^* h(\underline{t}) + o_p(1).$$

For the conclusion (ii), we have only to remark that $A(Z_n^*) = E_n^*(\underline{t})$ and

$$\sqrt{n}\,A(Z_n) = E_n(\sqrt{n}\,\underline{t}). \qquad \square$$

Next, the following lemmas show that $(Z_n^*)_{n\in\mathbb{N}}$ converge in distribution to the exponential function of a Gaussian process. Lemma A.12 deals with the finite-dimensional distributions of Z_n^*. In Lemmas A.13 and A.14 we show that Z_n^* is tight [see Dacunha-Castelle and Duflo (1983)]. The following lemmas will be proven under assumptions A1–A11.

LEMMA A.12. *The finite-dimensional p.d.'s of Z_n^* converge to those of*

$$Y := \exp\left[t^T I_G^{1/2} U - \tfrac{1}{2} t^T M t\right],$$

where U is a standard normal random vector in R^d.

PROOF. Let $t \in R^d$ be fixed, $n^{1/2} > \|t\|/\varepsilon$ and let us set $t_n = n^{-1/2} t$. Then we may write

$$\begin{aligned} &\log Z_n^*(t) - \log[a(t_n) f(t_n)/f(0)] \\ &\quad = \log Z_n(t_n) - \log Z_n(0) \\ &\quad = t^T \frac{1}{\sqrt{n}} S_n' - \frac{1}{2} t^T [M - R_n(t_n)] t, \end{aligned} \tag{A.79}$$

where $\log[a(t_n) f(t_n)/f(0)] = o(1)$,

$$S_n' := \sum_{i=1}^{n} l'(X_i, 0), \qquad S_n''(t) := \sum_{i=1}^{t} l''(X_i, t), \tag{A.80}$$

$$R_n(t_n) := \left(\frac{1}{n} S_n''(0) + M\right) + 2\int_0^1 (1-u) \frac{1}{n} [S_n''(u t_n) - S_n''(0)]\, du. \tag{A.81}$$

Using (4.2),

$$\begin{aligned} R_n(t_n) &\le \rho_n(t_n) \\ &:= \left\| \frac{1}{n} S_n''(0) + M \right\| + \frac{1}{n} \sum_{i=1}^{n} C(X_i)\left(1 + \|t_n\|^{b_6}\right) \|t_n\|. \end{aligned} \tag{A.82}$$

By assumption A8, the application of a central limit theorem yields

$$\mathscr{L}\left\{\frac{1}{\sqrt{n}} S_n'\right\} \to \mathscr{L}\{I_G^{1/2} U\} = N(0, I_G). \tag{A.83}$$

Next, by application of the law of large numbers we have

$$R_n(t_n) = o_p(1) \text{ for all fixed } t \in \Theta, \qquad t_n = n^{-1/2} t. \tag{A.84}$$

Then the lemma is a direct consequence of (A.79), through (A.83) and (A.84). □

Now we will give an exponential bound

$$m_\alpha(t) := c_\alpha \exp\left[a_\alpha \|t\| - \frac{\lambda}{4} \|t\|^2\right] \tag{A.85}$$

for the random functions Z_n^* and Y.

LEMMA A.13. *Let λ be the smallest eigenvalue of the matrix M. For all $\alpha > 0$ there exist positive c_α, a_α such that for all n large enough (say, $n > n_\alpha$) it holds that*

$$P\left(Z_n^*(t) \le m_\alpha(t), Y(t) \le m_\alpha(t) \text{ for } t \in R^d\right) \ge 1 - \alpha. \tag{A.86}$$

PROOF. $Z_n^*(t)$ vanishes outside $\Theta_n = \{t \in \Theta | \, \|t\| \le 2\epsilon\sqrt{n}\}$. From (A.79) we get

$$Z_n^{-1}(0)Z_n(t_n) = \exp\left[t_n^T S_n' - \tfrac{1}{2}t^T(M - R_n(t_n))t\right], \tag{A.87}$$

and from (A.82) we have $\sup_{\Theta_n} R_n(t_n) \le \rho_n(2\epsilon)$.

Using the notation

$$C_n^1 := \left\{\|S_n'\| < \sqrt{n}\, b_\alpha\right\}, \qquad C_n^2 := \left\{R_n(2\epsilon) \le \lambda/2\right\}, \tag{A.88}$$

the application of the CLT gives positive constants b_α and a n_α^1 with

$$P\left(C_n^1\right) > 1 - \frac{\alpha}{3} \quad \text{for all } n \ge n_\alpha^1. \tag{A.89}$$

Because of (A.66) and the SLLN, there is an $n_\alpha > n_\alpha^1$ with

$$P\left(C_n^2\right) > 1 - \frac{\alpha}{3} \quad \text{for all } n > n_\alpha. \tag{A.90}$$

For $\omega \in C_n^1 \cap C_n^2$ we have

$$Z_n^{-1}(0)Z_n(t_n)(\omega) \le \exp\left[b_\alpha\|t\| - \frac{\lambda}{4}\|t\|^2\right], \qquad t \in \Theta_n \tag{A.91}$$

and moreover, for $n > n_\alpha$,

$$P\left(C_n^1 \cap C_n^2\right) > 1 - \frac{2\alpha}{3}. \tag{A.92}$$

On the other hand, we have

$$\sup_{t \in \Theta,\, n \in N} a(t_n)f(t_n) \le k_\alpha := \sup_{\|t\| < 2\epsilon} f(t), \tag{A.93}$$

and as $Z_n^*(t)$ vanishes for $\|t\| > \varepsilon\sqrt{n}$ and $n \ge n_\alpha$, we obtain, if $\omega \in C_n^1 \cap C_n^2$,

$$Z_n^*(t, \omega) \le k_\alpha \exp\left[b_\alpha\|t\| - \frac{\lambda}{4}\|t\|^2\right] \quad \text{for all } t \in \Theta. \tag{A.94}$$

From the definition of Y follows obviously the existence of constants $c_\alpha > k_\alpha$ and $a_\alpha > b_\alpha$ such that for all $n \in N$,

$$P\left(Y(t) \le m_\alpha(t) := c_\alpha \exp\left[a_\alpha\|t\| - \frac{\lambda}{4}\|t\|^2\right]\right) > 1 - \frac{\alpha}{3}. \tag{A.95}$$

This together with (A.92) and (A.91) proves the lemma. □

Next we will study the equicontinuity of the trajectories of Z_n^*.

LEMMA A.14. *For each* $\alpha > 0$ *there is a set* $H_\alpha \subset \mathscr{C}$ *of equicontinuous functions with*:

(i) $P(Z_n^* \in H_\alpha) > 1 - \alpha$ *for all* $n \ge n_\alpha$.

(ii) *The sequence of the processess* $(Z_n^*)_{n \in N}$ *is tight in* $\mathscr{C}$.

PROOF. From the previous lemma it is sufficient to prove that for all $T > 0$ the sequence $(Z_n^*)_{n\in N}$ is tight in the set $\mathscr{C}(0,T)$ of the continuous functions on the closed sphere $S(0,T)$. Obviously the sequence of functions $\{a(t_n)f(t_n);\ n \in N\}$ are equicontinuous on $\mathbb{R}^d$. The product of two sequences of equibounded and equicontinuous functions is a sequence of equicontinuous functions. Then $a(t_n)b(t_n)$ vanishes outside $\Theta_n = \{t \in \Theta |\ \|t\| \le 2\epsilon\sqrt{n}\}$. With the notations of the previous lemma, we have only to prove that the derivatives

$$(\text{A.96})\qquad J_n(t) := \frac{d}{dt}\left\{Z_n(0)^{-1}Z_n(t_n)\right\} = Z_n(0)^{-1}Z_n(t_n)\frac{1}{\sqrt{n}}S'(t_n)t$$

are bounded for $n \ge n_\alpha$, $t \in \Theta_n$, $\omega \in C_n^1 \cap C_n^2$.

A Taylor expansion yields

$$(\text{A.97})\qquad J_n(t) = Z_n(0)^{-1}Z_n(t_n)\left[\frac{1}{\sqrt{n}}S_n' + (M + R_n(t_n))t\right]$$

On $C_n^1 \cap C_n^2$ and for $n \ge n_\alpha$, $t \in \Theta_n$ the bracket is majorized by

$$q_\alpha^1 := b_\alpha + \left(\|M\| + \frac{\lambda}{2}\right)\|t\| \quad [\text{see (A.88)}]$$

and $Z^{-1}(0)Z_n(t_n)$ by

$$q_\alpha^2(t) := \exp\left[b_\alpha\|t\| - \frac{\lambda}{4}\|t\|^2\right].$$

Then setting $q_\alpha := \sup_t q_\alpha^1(t)q_\alpha^2(t)$, we get

$$(\text{A.98})\qquad P\left(\sup_{t\in\Theta_n}\|J_n(t)\| \le q_\alpha\right) \ge P\left(C_n^1 \cap C_n^2\right) \ge 1 - 2\alpha/3.$$

Conclusion (i) is proved. Conclusion (ii) is a consequence of (i) and of the fact that $Z_n^*(0) = 1$. □

Final reasoning in the proof of Theorem 4.1. Recall that $\mathscr{C}$ is the set of the continuous real functions on R^d. Because of Lemma A.12, the majorization of Z_n^* and Y in Lemma A.13 and Lemma A.14 for all functions $h \in \mathscr{C}$ satisfying (A.69), we have the convergence in distribution

$$(\text{A.99})\qquad \left(\int h(t)Z_n^*(t)\,dt, \int Z_n^*(t)\,dt\right) \to_d \left(\int h(t)Y(t)\,dt, \int Y(t)\,dt\right)$$

and $\int Y(t)\,dt \ne 0$ a.s. Setting $h(t) = t$, we get $A(Z_n^*) \to_d A(Y)$. The distribution of $A(Y)$ is Gaussian $N(0, M^{-1}I_G M^{-1})$. On the other hand, from (A.65) and Lemma A.11 we have

$$\|G_n\| \le c\left(n^{-1/2}E_n^*\|t\|^2 + n^{-(b_7+1)/2}E_n^*\|t\|^{b_7+2}\right) + o_P(1).$$

Then setting $h(t) = \|t\|^2$ or $\|t\|^{b_7+2}$ in (A.71), we can see with (A.65) and (A.99) that $G_n = o_p(1)$. Then (A.53) proves the theorem.

REFERENCES

BERK, R. H. (1966). Limiting behaviour of posterior distributions when the model is incorrect. *Ann. Math. Statist.* **37** 51–58.

BERK, R. H. (1970). Consistency of a posteriori. *Ann. Math. Statist.* **41** 894–906.

BICKEL, P. J. and FREEDMAN, D. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

BICKEL, P. J. and YAHAV, J. A. (1969). Some contributions to the asymptotic theory of Bayes solutions. *Z. Wahrsch. Verw. Gebiete* **11** 257–276.

CHUNG, K. L. (1974). *A Course in Probability Theory*, 2nd ed. Academic Press, New York.

DACUNHA-CASTELLE, D. and DUFLO, M. (1983). *Probabilités et Statistiques* **2**. *Problèmes à Temps Mobile*. Masson, Paris.

DIACONIS, P. and FREEDMAN, D. (1986a). On the consistency of Bayes estimates. *Ann. Statist.* **14** 1–26.

DIACONIS, P. and FREEDMAN, D. (1986b). On inconsistent Bayes estimates of location. *Ann. Statist.* **14** 68–87.

GOURIEROUX, C. and MONTFORT, A. (1993). Pseudo likelihood methods. In *Handbook of Statistics* (S. Kotz and N. L. Johnson, eds.) **11**, Chapter 12. North-Holland, Amsterdam.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

HANOUSEK and JURECKOVA, J. (1996). Unpublished manuscript.

HUBER, P. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press, Berkeley.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation. Asymptotic Theory*. Springer, New York.

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*. Springer, Berlin.

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

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

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

INSTITUT FÜR MATHEMATIK
HUMBOLDT UNIVERSITY
UNTER DEN LINDEN 6
10099 BERLIN
GERMANY
E-MAIL: bunke@mathematik.hu-berlin.de

LABORATOIRE DE STATISTIQUE
ET PROBABILITÉS
118 ROUTE DE NARBONNE
31062 TOULOUSE CEDEX 5
FRANCE
E-MAIL: milhaud@cict.fr