The Annals of Statistics
2003, Vol. 31, No. 2, 536–559

ADAPTIVE BAYESIAN INFERENCE ON THE MEAN OF AN INFINITE-DIMENSIONAL NORMAL DISTRIBUTION

BY EDUARD BELITSER AND SUBHASHIS GHOSAL

Utrecht University and North Carolina State University

We consider the problem of estimating the mean of an infinite-dimensional normal distribution from the Bayesian perspective. Under the assumption that the unknown true mean satisfies a "smoothness condition," we first derive the convergence rate of the posterior distribution for a prior that is the infinite product of certain normal distributions and compare with the minimax rate of convergence for point estimators. Although the posterior distribution can achieve the optimal rate of convergence, the required prior depends on a "smoothness parameter" q. When this parameter q is unknown, besides the estimation of the mean, we encounter the problem of selecting a model. In a Bayesian approach, this uncertainty in the model selection can be handled simply by further putting a prior on the index of the model. We show that if q takes values only in a discrete set, the resulting hierarchical prior leads to the same convergence rate of the posterior as if we had a single model. A slightly weaker result is presented when q is unrestricted. An adaptive point estimator based on the posterior distribution is also constructed.

1. Introduction. Suppose we observe an infinite-dimensional random vector $\mathbf{X} = (X_1, X_2, \dots)$, where X_i's are independent, X_i has distribution $N(\theta_i, n^{-1})$, $i = 1, 2, \dots,$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \ell_2$, that is, $\sum_{i=1}^{\infty} \theta_i^2 < \infty$. The parameter $\boldsymbol{\theta}$ is unknown and the goal is to make inference about $\boldsymbol{\theta}$. Let $P_{\theta,n}$ stand for the distribution of $\mathbf{X}$; here and throughout, we suppress the dependence of $\mathbf{X}$ and $P_\theta = P_{\theta,n}$ on n unless indicated otherwise. We shall write $\|\cdot\|$ for the ℓ_2 norm throughout the paper.

We also note that $\mathbf{X}$ may be thought of as the sample mean vector of the first n observations of an i.i.d. sample $\mathbf{Y}_1, \mathbf{Y}_2, \dots,$ each taking values in $\mathbb{R}^\infty$ and distributed like $\mathbf{Y} = (Y_1, Y_2, \dots)$, where Y_i's are independent and Y_i has distribution $N(\theta_i, 1)$, $i = 1, 2, \dots$. Since the sample mean is sufficient, these two formulations are statistically equivalent. To avoid complicated notation, we shall generally work with the first formulation. However, the latter formulation will help us apply the general theory of posterior convergence developed by Ghosal, Ghosh and van der Vaart (2000).

The interest in the infinite-dimensional normal model is partly due to its equivalence with the prototypical white noise model

$$dX_\varepsilon(t) = f(t)\,dt + \varepsilon\,dW(t), \qquad 0 \le t \le 1,$$

Received February 2000; revised March 2002.
AMS 2000 subject classifications. Primary 62G20; secondary 62C10, 62G05.
Key words and phrases. Adaptive Bayes procedure, convergence rate, minimax risk, posterior distribution, model selection.

where $X_\varepsilon(t)$ is the noise-corrupted signal, $f(\cdot) \in L_2[0,1]$ is the unknown signal, $W(t)$ is a standard Wiener process and $\varepsilon > 0$ is a small parameter. The statistical estimation problem is to recover the signal $f(t)$, based on the observation $X_\varepsilon(t)$. The above model arises as the limiting experiment in some curve estimation problems such as in density estimation [Nussbaum (1996); Klemelä and Nussbaum (1998)] and nonparametric regression [Brown and Low (1996)], where $\varepsilon = n^{-1/2}$ and n is the sample size.

Suppose that the functions $\{\phi_i, i = 1, 2, \dots\}$ form an orthonormal basis in $L_2[0,1]$. Under the assumption $f(\cdot) \in L_2[0,1]$, we can reduce this problem to the problem of estimating the mean $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \ell_2$ of an infinite-dimensional normal distribution. Indeed,

$$X_i = \theta_i + \varepsilon \xi_i, \qquad i = 1, 2, \dots,$$

where $X_i = \int_0^T \phi_i(t)\,dX_\varepsilon(t)$, $\theta_i = \int_0^T \phi_i(t) f(t)\,dt$ and $\xi_i = \int_0^T \phi_i(t)\,dW(t)$, so that the ξ_i's are independent standard normal random variables. In so doing, we arrive at the infinite-dimensional Gaussian shift experiment with $\varepsilon = n^{-1/2}$. The signal $f(t)$ can be recovered from the basis expansion $f(t) = \sum_{i=1}^\infty \theta_i \phi_i(t)$ and the map relating f and $\boldsymbol{\theta}$ is an isometric isomorphism.

Coming back to the infinite-dimensional normal model, if a Bayesian analysis is intended, one assigns a prior to $\boldsymbol{\theta}$ and looks at the posterior distribution. Diaconis and Freedman (1997) and Freedman (1999) considered the independent normal prior $N(0, \tau_i^2)$ for θ_i, $i = 1, 2, \dots$, and comprehensively studied the nonlinear functional $\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2$, where $\hat{\boldsymbol{\theta}}$ is the Bayes estimator, both from the Bayesian and the frequentist perspectives. As a consequence of their main results, they established frequentist consistency of the Bayes estimator for all $\boldsymbol{\theta} \in \ell_2$ if $\sum_{i=1}^\infty \tau_i^2 < \infty$. A peculiar implication of their result is that the Bayesian and frequentist asymptotic distributions of $\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2$ differ, and hence the Bernstein–von Mises theorem fails to hold for this simple infinite-dimensional problem. A similar conclusion was reached earlier by Cox (1993) in a slightly different model.

This estimation problem was first studied in a minimax setting by Pinsker (1980). He showed that if the unknown infinite-dimensional parameter $\boldsymbol{\theta}$ is assumed to belong to an ellipsoid $\Theta_q(Q) = \{\boldsymbol{\theta} : \sum_{i=1}^\infty i^{2q}\theta_i^2 \le Q\}$, then the exact asymptotics of the minimax quadratic risk over the ellipsoid $\Theta_q(Q)$ are given by

$$\lim_{n\to\infty} \inf_{\hat{\boldsymbol{\theta}}} \sup_{\boldsymbol{\theta} \in \Theta_q(Q)} n^{2q/(2q+1)} E_\theta \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 = Q^{1/(2q+1)} \gamma(q), \tag{1.1}$$

where $\gamma(q) = (2q+1)^{1/(2q+1)}(q/(q+1))^{2q/(2q+1)}$ is the Pinsker constant [Pinsker (1980)] and E_θ denotes the expectation with respect to the probability measure P_θ generated by $\mathbf{X}$ given $\boldsymbol{\theta}$.

Zhao (2000) considered the independent normal prior with $\tau_i^2 = \tau_i^2(q) = i^{-(2q+1)}$, $i = 1, 2, \dots$ [considered also by Cox (1993) and Freedman (1999)], and showed that the posterior mean attains the minimax rate $n^{-q/(2q+1)}$. We shall write Π_q to denote this prior. We show in Theorem 2.1 that the posterior distribution $\Pi_q(\cdot|\mathbf{X})$ obtained from the prior Π_q also converges at the rate $n^{-q/(2q+1)}$ in the sense that, for any sequence $M_n \to \infty$,

$$\Pi_q\{\boldsymbol{\theta} : n^{q/(2q+1)}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n | \mathbf{X}\} \to 0$$

in P_{θ_0}-probability as $n \to \infty$, where the true value of the parameter $\boldsymbol{\theta}_0$ belongs to the linear subspace $\Theta_q = \{\boldsymbol{\theta} : \sum_{i=1}^{\infty} i^{2q}\theta_i^2 < \infty\}$. It has long been known that for finite-dimensional models the posterior distribution converges at the classical $n^{-1/2}$ rate, but for infinite-dimensional models results have only been obtained recently by Ghosal, Ghosh and van der Vaart (2000) and Shen and Wasserman (2001).

Our main goal in the present paper is to construct, without knowing the value of the smoothness parameter q, a prior such that the posterior distribution attains the optimal rate of convergence. In this case, instead of one single model $\boldsymbol{\theta} \in \Theta_q$, we have a nested collection of models $\boldsymbol{\theta} \in \Theta_q$, $q \in \mathcal{Q}$. The inference based on such a prior is therefore adaptive in the sense that the same prior gives the optimal rate of convergence irrespective of the smoothness condition. Here, besides the problem of estimation, we encounter the problem of model selection. In a Bayesian framework, perhaps the most natural candidate for a prior that gives the optimal rate over various competing models is a mixture of the appropriate priors in different models, indexed by the smoothness parameter q. So we consider a mixture of Π_q's over q, that is, a prior of the form $\sum \lambda_q \Pi_q$, where $\lambda_q > 0$, $\sum \lambda_q = 1$ and the sum is taken over some countable set. We show that the resulting hierarchical or mixture prior leads to the optimal convergence rate of the posterior simultaneously for all q under consideration. Similar results were also found by Ghosal, Lember and van der Vaart (2002) in the context of density estimation and Huang (2000) for density estimation and regression.

The problem of adaptive estimation in a minimax sense was first studied by Efromovich and Pinsker (1984). They proposed an estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ which does not assume knowledge of the smoothness parameter q, yet attains the minimum in (1.1).

The organization of the paper is as follows. In the next section we study the case when the smoothness parameter is known. Then we formulate the main result of the paper, which states that if the smoothness parameter is unknown but lies in a discrete set, the optimal rate can be achieved by a mixture prior simultaneously for different smoothness values. In Section 4, we construct an adaptive estimator based on the posterior distribution that pointwise achieves the optimal rate of convergence in the frequentist sense. The proof of the main result is based on several auxiliary lemmas among which Lemma 3.1 is of interest on its own. The proofs of these lemmas are given in Sections 5 and 6. In Section 7, we consider the case where q can take any value in the continuum and present a slightly weaker result. Uniformity of the convergence of the posterior distribution with respect to the parameter lying in an infinite dimensional ellipsoid is discussed in Section 8.

From now on, all symbols O and o refer to the asymptotics $n \to \infty$ unless otherwise specified. By $\lfloor x \rfloor$, we shall mean the greatest integer less than or equal to x.

2. Known smoothness. Recall that we have independent observations X_i distributed as $N(\theta_i, n^{-1})$, $i = 1, 2, \dots$. From now on, we assume that the unknown parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$ belongs to the set $\Theta_q = \{\boldsymbol{\theta} : \sum_{i=1}^{\infty} i^{2q} \theta_i^2 < \infty\}$, a Sobolev-type subspace of ℓ_2. Note that q measures the "smoothness" of $\boldsymbol{\theta}$, since, in the equivalent white noise model with the standard trigonometric Fourier basis, the set Θ_q essentially corresponds to the class of all periodic functions $f(\cdot) \in L_2[0, 1]$ whose $L_2[0, 1]$ norm of the qth derivative is bounded, when q is an integer (otherwise the qth Weyl derivative is meant). So, in some sense q stands for the "number of derivatives" of the unknown signal in the equivalent white noise model. For this reason, q will often be referred to as the smoothness parameter of $\boldsymbol{\theta}$.

Let $\boldsymbol{\theta}_0 = (\theta_{10}, \theta_{20}, \dots)$ denote the true value of the parameter so that we have $\boldsymbol{\theta}_0 \in \Theta_q$. Let the prior Π_p be defined as

$$\theta_i\text{'s are independent } N(0, i^{-(2p+1)}).$$

Then the posterior distribution of $\boldsymbol{\theta}$ given $\mathbf{X}$ is described by

$$\theta_i\text{'s are independent } N\left(\frac{nX_i}{n + i^{2p+1}}, \frac{1}{n + i^{2p+1}}\right).$$

The posterior mean is given by $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \dots)$, where

$$\hat{\theta}_i = \frac{nX_i}{n + i^{2p+1}}, \qquad i = 1, 2, \dots.$$

Note that the posterior distribution of θ_i depends on $\mathbf{X}$ only through X_i. In Theorem 5.1 of Zhao (2000) [and implicitly in Section 3 of Cox (1993) and Theorem 5 of Freedman (1999)], it is shown that $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}_0$ at the rate $n^{-\min(p,q)/(2p+1)}$. By a slight extension of Zhao's (2000) argument, we observe that the posterior distribution also converges at this rate.

THEOREM 2.1. *For any sequence $M_n \to \infty$,*

$$\Pi_p\{\boldsymbol{\theta} : n^{\min(p,q)/(2p+1)} \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n | \mathbf{X}\} \to 0, \tag{2.1}$$

in P_{θ_0}-probability as $n \to \infty$. In particular, if we know q, we can choose $p = q$, that is, select the prior Π_q to achieve the best possible rate $n^{-q/(2q+1)}$ of convergence.

PROOF. Applying Chebyshev's inequality, the posterior probability in (2.1) is at most

$$M_n^{-2} n^{2\min(p,q)/(2p+1)} \sum_{i=1}^{\infty} \big((\hat{\theta}_i - \theta_{i0})^2 + \operatorname{var}(\theta_i | X_i)\big).$$

It suffices to show that the P_{θ_0}-expectation of the above expression tends to zero. Now it is easy to see that

$$\sum_{i=1}^{\infty} \operatorname{var}(\theta_i | X_i) = \sum_{i=1}^{\infty} (n + i^{2p+1})^{-1} = O\big(n^{-2p/(2p+1)}\big).$$

It follows from Theorem 5.1 of Zhao (2000) that

$$\sum_{i=1}^{\infty} E_{\theta_0} (\hat{\theta}_i - \theta_{i0})^2 = O\big(n^{-2\min(p,q)/(2p+1)}\big)$$

and hence the result follows. □

REMARK 2.1. Shen and Wasserman (2001) also calculated the rate of convergence of the posterior distribution using a different method for the special case when $\theta_{i0} = i^{-p}$ and also showed that the obtained rate is sharp.

REMARK 2.2. Interestingly, as Zhao (2000) pointed out, although the prior Π_p with $p = q$ is the best choice (among independent normal priors with power–variance structure) as far as the convergence rate is concerned, Π_q has a peculiar property. Both the prior and the posterior assign zero probability to Θ_q. This is an easy consequence of the criterion for summability of a random series. It is also interesting to note that summability of $\sum_{i=1}^{\infty} i^{2q}\theta_i^2$ is barely missed by the prior since $\sum_{i=1}^{\infty} i^{2p}\theta_i^2 < \infty$ almost surely for any $p < q$ and so $q = \sup\{p : \sum_{i=1}^{\infty} i^{2p}\theta_i^2 < \infty \text{ almost surely}\}$. To fix this problem, Zhao (2000) proposed a compound prior $\sum_{k=1}^{\infty} w_k \pi_k$, where under π_k, the θ_i's are independent with distribution $N(0, i^{-(2q+1)})$ for $i = 1, \dots, k$ and degenerate at 0 otherwise, and the w_k's are weights bounded from below by a sequence exponentially decaying in k. She showed that, on the one hand, the posterior mean attains the minimax rate of convergence and, on the other hand, assigns probability 1 to Θ_q.

3. Discrete spectrum adaptation. So far our approach relies heavily on the fact that we know the smoothness, that is, we have chosen a single model as the "correct" one from the collection of possible models corresponding to different choices of $q > 0$. In general, when the parameter specifying the model is not chosen correctly, this may lead to suboptimal or inconsistent procedures, since further analysis does not take into account the possibility of other models.

In Theorem 2.1, it is appropriate to use the prior Π_p with $p = q$ only if we know the smoothness parameter q correctly, that is, we know that $\boldsymbol{\theta}_0 \in \Theta_q$ and the assertion holds for no larger value of q. If actually $\boldsymbol{\theta}_0 \in \Theta_{q'}$ for some $q' > q$, then we lose in terms of the rate of convergence since we could have obtained the better rate $n^{-q'/(2q'+1)}$ by using the prior $\Pi_{q'}$. On the other hand, if $\boldsymbol{\theta}_0 \in \Theta_{q'}$, $q' < q$ only, then the model is misspecified. In general, one expects inconsistency in this type of situation. In this case, however, since all the models are dense in ℓ_2, the posterior is nevertheless consistent. Still, there is a loss in rate of convergence again as we only get the rate $n^{-q'/(2q+1)}$ from Theorem 2.1 instead of the rate $n^{-q'/(2q'+1)}$ achievable by the prior $\Pi_{q'}$.

Therefore, we intend to present a prior Π that achieves the posterior rate of convergence $n^{-q/(2q+1)}$ at $\boldsymbol{\theta}_0$ whenever $\boldsymbol{\theta}_0 \in \Theta_q$ for different values of the smoothness parameter q. Let $\mathcal{Q} = \{\dots, q_{-1}, q_0, q_1, \dots\}$ be a countable subset of the positive real semiaxis without any accumulation point other than 0 or ∞. Such a set may be arranged in an increasing sequence that preserves the natural order, that is, $0 < \cdots < q_{-1} < q_0 < q_1 < \cdots$. We show that there is a prior Π for $\boldsymbol{\theta}$ such that whenever $\boldsymbol{\theta}_0 \in \Theta_q$, $q \in \mathcal{Q}$, the posterior converges at the rate $n^{-q/(2q+1)}$.

We may think of q as another unknown parameter. So, instead of one single model, we have a sequence of nested models parameterized by a "smoothness" parameter q ranging over the discrete set $\mathcal{Q}$. In a Bayesian approach, perhaps the most natural way to handle this uncertainty in the model is to put a prior on the index of the model or the "hyperparameter" q. The resulting prior Π is therefore the two-level hierarchical prior

$$\begin{aligned}&\text{given } q, \text{ the } \theta_i\text{'s are independent } N(0, i^{-(2q+1)});\\ &q \text{ has distribution } \lambda.\end{aligned}$$

The main purpose of this article is to show that this simple and natural procedure is rate adaptive, that is, the mixture prior achieves the convergence rate $n^{-q/(2q+1)}$ whenever $\boldsymbol{\theta}_0 \in \Theta_q$ and $q \in \mathcal{Q}$.

Let $\lambda_m = \lambda(q = q_m)$. Henceforth, we shall write Π_m for Π_{q_m}. We thus have $\Pi = \sum_{m=-\infty}^{\infty} \lambda_m \Pi_m$. Throughout, we shall assume that $\lambda_m > 0$ for all m.

Without loss of generality, we may assume that $\boldsymbol{\theta}_0 \in \Theta_{q_0}$. Otherwise, we simply relabel.

THEOREM 3.1. *For any sequence $M_n \to \infty$, the posterior probability*

$$\Pi\{\boldsymbol{\theta} : n^{q_0/(2q_0+1)} \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n | \mathbf{X}\} \to 0 \tag{3.1}$$

in P_{θ_0}-probability as $n \to \infty$.

Since q serves as a parameter, we can talk about its posterior distribution. The idea behind the proof is first to show that the posterior probability of q being smaller than q_0 is small for large n in P_{θ_0}-probability, that is, the probability of selecting a coarser model from the posterior is small. This effectively reduces the prior to the form $\sum_{m\geq 0}\lambda_m\Pi_m$. For such a prior we then show that the posterior probability of $\{\boldsymbol{\theta}:\sum_{i=1}^{\infty} i^{2q_0}\theta_i^2 > B, q > q_0\}$ is small for a sufficiently large B, while Theorem 2.1 implies that $\{\boldsymbol{\theta}: n^{q_0/(2q_0+1)}\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| > M_n, q = q_0\}$ converges to 0 in P_{θ_0}-probability. Therefore, the effective parameter space is further reduced to $\{\boldsymbol{\theta}:\sum_{i=1}^{\infty} i^{2q_0}\theta_i^2 \leq B\}$ for which the general theory developed by Ghosal, Ghosh and van der Vaart (2000) applies.

REMARK 3.1. It should be noted that the cases $q < q_0$ and $q > q_0$ are not symmetrically treated. Unlike $q < q_0$, we do not show that $\Pi(q > q_0|\mathbf{X})$ tends to 0. An obvious reason is that it may still be possible that $\boldsymbol{\theta}_0 \in \Theta_{q_m}$ for $m > 0$. For instance, if θ_{i0}^2 decreases exponentially with i, then $\boldsymbol{\theta}_0 \in \Theta_q$ for all q. In such cases, clearly it is not to be expected that $\Pi(q > q_0|\mathbf{X})$ tends to 0. Even if $\boldsymbol{\theta}_0 \notin \Theta_{q_1}$, $q_1 > q_0$, it is still possible that $\Pi(q > q_0|\mathbf{X})$ does not tend to 0. Indeed, considering two values $\{q_0, q_1\}$ and the special case $\theta_{i0} = i^{-p}$, a referee exhibited that if $p \geq q_1 + 1 - (2q_1+1)/(2(2q_0+1))$, then $\Pi(q = q_1|\mathbf{x}) \to 1$ in probability. In general, the Π_{q_1} prior leads to a slower posterior convergence rate than the required $n^{-q_0/(2q_0+1)}$ for a general $\boldsymbol{\theta}_0 \in \Theta_{q_0}$, so it seems at first glance that our mixture prior cannot simultaneously achieve the optimal rate. The apparent paradox is resolved when one observes, as the referee does, that the latter happens only when $p < q_1 + 1 - (2q_1+1)/(2(2q_0+1))$.

The following lemma shows that, given the observations, the conditional probability of misspecifying the model in favor of a coarser model (than the true one) from the posterior converges to zero.

LEMMA 3.1. *There exist an integer N and a constant $c > 0$ such that for any $m < 0$ and $n > N$*,

$$E_{\theta_0}\Pi(q = q_m|\mathbf{X}) \leq \frac{\lambda_m}{\lambda_0}\exp\bigl[-cn^{1/(q_0+q_m+1)}\bigr] \tag{3.2}$$

and, therefore,

$$\Pi(q < q_0|\mathbf{X}) \to 0 \tag{3.3}$$

in P_{θ_0}-probability as $n \to \infty$.

The proof of the Lemma 3.1 is somewhat lengthy and involves some messy calculations. We defer it to Section 5.

REMARK 3.2. Note that if the observation $\mathbf{X}$ is interpreted as the sample mean $(\mathbf{Y}_1 + \cdots + \mathbf{Y}_n)/n$, where the $\mathbf{Y}_i$'s are as in the Introduction, then (3.3) holds in the almost sure sense. We, however, do not use this fact.

According to Lemma 3.1, the posterior mass is asymptotically concentrated on $q \geq q_0$. We separate two possibilities: $q = q_0$ and $q > q_0$. For $m > 0$, introduce $\tilde{\lambda}_m = \lambda_m / \sum_{j=1}^{\infty} \lambda_j$, and $\tilde{\Pi} = \sum_{m=1}^{\infty} \tilde{\lambda}_m \Pi_m$. The next lemma shows that the posterior $\tilde{\Pi}(\cdot|\mathbf{X})$ is effectively concentrated on the set $\{\boldsymbol{\theta} : \sum_{i=1}^{\infty} i^{2q_0}\theta_i^2 \leq B\}$ for a large B. The proof of this lemma is postponed to Section 6.

LEMMA 3.2.

$$\lim_{B\to\infty} \sup_{n\geq 1} E_{\theta_0} \tilde{\Pi}\left\{\boldsymbol{\theta} : \sum_{i=1}^{\infty} i^{2q_0}\theta_i^2 > B \middle| \mathbf{X}\right\} = 0. \tag{3.4}$$

Finally, the following lemma shows that the posterior converges at rate $n^{-q_0/(2q_0+1)}$. In the proof which is given in Section 6, we exploit the general theory of posterior convergence developed by Ghosal, Ghosh and van der Vaart (2000).

For $m \geq 0$, introduce $\bar{\lambda}_m = \lambda_m / \sum_{j=0}^{\infty} \lambda_j$ and $\bar{\Pi} = \sum_{m=0}^{\infty} \bar{\lambda}_m \Pi_m$. Note that, unlike in $\tilde{\Pi}$, the possibility $q = q_0$ is not ruled out here.

LEMMA 3.3. *For any $B > 0$ and $M_n \to \infty$,*

$$\lim_{n\to\infty} E_{\theta_0} \bar{\Pi}\left\{\boldsymbol{\theta} : n^{q_0/(2q_0+1)}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n, \sum_{i=1}^{\infty} i^{2q_0}\theta_i^2 \leq B \middle| \mathbf{X}\right\} = 0. \tag{3.5}$$

With the help of Lemmas 3.1–3.3 and Theorem 2.1, the main result can be easily proved now.

PROOF OF THEOREM 3.1. For the sake of brevity, denote $r_n = r_n(q_0) = n^{q_0/(2q_0+1)}$. Clearly,

$$\begin{aligned}
&\Pi\{\boldsymbol{\theta} : r_n\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n|\mathbf{X}\} \\
&\quad = \Pi\left\{\boldsymbol{\theta} : r_n\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n, q > q_0, \sum_{i=1}^{\infty} i^{2q_0}\theta_i^2 > B \middle| \mathbf{X}\right\} \\
&\qquad + \Pi\left\{\boldsymbol{\theta} : r_n\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n, q > q_0, \sum_{i=1}^{\infty} i^{2q_0}\theta_i^2 \leq B \middle| \mathbf{X}\right\} \\
&\qquad + \Pi\{\boldsymbol{\theta} : r_n\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n, q < q_0|\mathbf{X}\} \\
&\qquad + \Pi\{\boldsymbol{\theta} : r_n\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n, q = q_0|\mathbf{X}\} \\
&\quad \leq \tilde{\Pi}\left\{\boldsymbol{\theta} : \sum_{i=1}^{\infty} i^{2q_0}\theta_i^2 > B \middle| \mathbf{X}\right\} \Pi(q > q_0|\mathbf{X}) \\
&\qquad + \Pi\left\{\boldsymbol{\theta} : r_n\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n, q \geq q_0, \sum_{i=1}^{\infty} i^{2q_0}\theta_i^2 \leq B \middle| \mathbf{X}\right\}
\end{aligned}$$

$$
\begin{aligned}
&+ \Pi\{\boldsymbol{\theta} : r_n\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n, q < q_0|\mathbf{X}\} \\
&+ \Pi_0\{\boldsymbol{\theta} : r_n\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n|\mathbf{X}\}\Pi(q = q_0|\mathbf{X}) \\
\le{}& \tilde{\Pi}\left\{\boldsymbol{\theta} : \sum_{i=1}^{\infty} i^{2q_0}\theta_i^2 > B \,\middle|\, \mathbf{X}\right\} \\
&+ \bar{\Pi}\left\{\boldsymbol{\theta} : r_n\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n, \sum_{i=1}^{\infty} i^{2q_0}\theta_i^2 \le B \,\middle|\, \mathbf{X}\right\} \\
&+ \Pi(q < q_0|\mathbf{X}) \\
&+ \Pi_0\{\boldsymbol{\theta} : r_n\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n|\mathbf{X}\}.
\end{aligned}
$$

Given $\varepsilon > 0$, B can be chosen sufficiently large to make the first term less than ε by Lemma 3.2. For this B, apply Lemma 3.3 to the second term. The third term goes to zero in P_{θ_0}-probability by Lemma 3.1, while the last term converges to zero by Theorem 2.1. The theorem is proved. □

REMARK 3.3. In place of Π_m, Zhao's (2000) mixture prior described in Remark 2.2 may also possibly be used. However, we do not pursue this approach for the following reasons. First, the expressions will be even more complicated. Second, we think that when ℓ_2 is considered as the parameter space, the property that Π_m assigns zero probability to Θ_q is not as bad as it might first appear since Θ_q is not closed and its closure, the whole of ℓ_2, obviously receives the whole mass. Indeed, both Π_m and Zhao's mixture prior have support of the whole of ℓ_2. Finally, the criterion that "Θ_q should receive the whole mass" loses much of its original motivation in the present context of adaptation, when ℓ_2 is the grand parameter space.

4. Adaptive estimator. As a consequence of the result on adaptivity of the posterior distribution, we now show that there is an estimator $\hat{\boldsymbol{\theta}}$ based on the posterior distribution that is rate adaptive in the frequentist sense. The problem of adaptive estimation in a minimax sense was first studied by Efromovich and Pinsker (1984). They proposed an estimator which attains the minimum in (1.1) without knowledge of the smoothness parameter q. Their method is based on adaptively determining optimal damping coefficients in an orthogonal series estimator.

To construct an estimator based on the posterior distribution, one may maximize the posterior probability of a ball of appropriate radius as in Theorem 2.5 of Ghosal, Ghosh and van der Vaart (2000). However, their construction requires knowledge of the convergence rate and hence does not apply to the adaptive setup of the problem. The following modification of their construction is due to van der Vaart (2000). Let

$$
\delta_n^* = \inf\{\delta_n : \Pi(\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}'\| \le \delta_n|\mathbf{X}) \ge 3/4 \text{ for some } \boldsymbol{\theta}' \in \ell_2\}. \tag{4.1}
$$

Take any point $\hat{\boldsymbol{\theta}} \in \ell_2$ which satisfies

$$\Pi(\boldsymbol{\theta} : \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\| \le \delta_n^* + n^{-1} | \mathbf{X}) \ge 3/4. \tag{4.2}$$

In words, $\hat{\boldsymbol{\theta}}$ is the center of the ball of nearly the smallest radius subject to the constraint that its posterior mass is not less than 3/4. Note that while defining $\hat{\boldsymbol{\theta}}$, we have not used the value of the smoothness parameter. The next theorem shows that $\hat{\boldsymbol{\theta}}$ has the optimal frequentist rate of convergence $n^{-q_0/(2q_0+1)}$ and hence is an adaptive estimator.

THEOREM 4.1 [van der Vaart (2000)]. *For any $M_n \to \infty$ and $\boldsymbol{\theta}_0 \in \Theta_{q_0}$,*

$$P_{\theta_0}\{n^{q_0/(2q_0+1)} \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| > M_n\} \to 0,$$

where $\hat{\boldsymbol{\theta}}$ is defined by (4.1) *and* (4.2).

PROOF. By Theorem 3.1, we have that for any $M_n \to \infty$,

$$\Pi\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le M_n n^{-q_0/(2q_0+1)} | \mathbf{X}\} \ge 3/4 \tag{4.3}$$

with P_{θ_0}-probability tending to 1.

Let $B(\tilde{\boldsymbol{\theta}}, r) = \{\boldsymbol{\theta}' : \|\boldsymbol{\theta}' - \tilde{\boldsymbol{\theta}}\| \le r\}$ denote the ball around $\tilde{\boldsymbol{\theta}}$ of size r. The balls $B(\hat{\boldsymbol{\theta}}, \delta_n^* + 1/n)$ and $B(\boldsymbol{\theta}_0, M_n n^{-q_0/(2q_0+1)})$ both have posterior probability at least 3/4, so they must intersect; otherwise, the total posterior mass would exceed 1. Therefore, by the triangle inequality,

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \le \delta_n^* + n^{-1} + M_n n^{-q_0/(2q_0+1)} \le 2M_n n^{-q_0/(2q_0+1)} + n^{-1},$$

since by the definition of δ_n^* and (4.3), $\delta_n^* \le M_n n^{-q_0/(2q_0+1)}$. The proof follows. □

5. Proof of the Lemma 3.1. We begin with a couple of preliminary lemmas. We recall that given $\boldsymbol{\theta}$, the X_i's are independent and distributed as $N(\theta_i, n^{-1})$, and given $q = q_m$, the θ_i's are independent and distributed as $N(0, i^{-(2q_m+1)})$. Therefore, given $q = q_m$, the marginal distribution of $\mathbf{X}$ is given by the countable product of $N(0, n^{-1} + i^{-(2q_m+1)})$, $i = 1, 2, \dots$. Let us denote this measure by P_m. Let $P_\lambda = \sum_{m=-\infty}^{\infty} \lambda_m P_m$, the marginal distribution of $\mathbf{X}$ when q is distributed according to λ. The true distribution of $\mathbf{X}$ is, however, P_{θ_0}, the countable product of $N(\theta_{i0}, n^{-1})$, $i = 1, 2, \dots$.

LEMMA 5.1. *For any $n \ge 1$ and $\boldsymbol{\theta}_0 \in \ell_2$, P_λ and P_{θ_0} are mutually absolutely continuous.*

PROOF. It suffices to show that for any m, P_m and P_{θ_0} are mutually absolutely continuous. Abbreviate q_m by q. Recall that the affinity between two probability densities $f(x)$ and $g(x)$ on the real line is defined by $A(f, g) = \int \sqrt{f(x)g(x)}\,dx$. The affinity between two normal densities $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ is therefore given by

$$\left[1 - \frac{(\sigma_1 - \sigma_2)^2}{\sigma_1^2 + \sigma_2^2}\right]^{1/2} \exp\left\{-\frac{(\mu_1 - \mu_2)^2}{4(\sigma_1^2 + \sigma_2^2)}\right\}.$$

Since both P_m and P_{θ_0} are product measures, according to Kakutani's criterion [see, e.g., Williams (1991), page 144], we need only to show that the infinite product

$$\prod_{i=1}^{\infty} A\big(N(0, n^{-1} + i^{-(2q+1)}), N(\theta_{i0}, n^{-1})\big)$$

or, equivalently,

$$\prod_{i=1}^{\infty} \left\{1 - \frac{((n^{-1} + i^{-(2q+1)})^{1/2} - n^{-1/2})^2}{2n^{-1} + i^{-(2q+1)}}\right\} \tag{5.1}$$

and

$$\prod_{i=1}^{\infty} \exp\left\{-\frac{\theta_{i0}^2}{4(2n^{-1} + i^{-(2q+1)})}\right\} \tag{5.2}$$

do not diverge to zero. A product of the form $\prod_{i=1}^{\infty}(1 - a_i)$, $a_i > 0$, $i = 1, 2, \dots$, converges if $\sum_{i=1}^{\infty} a_i$ converges. Also, $(a-b)^2 \le a^2 - b^2$ for $a \ge b \ge 0$. Therefore, (5.1) converges as

$$\sum_{i=1}^{\infty} \frac{i^{-(2q+1)}}{2n^{-1} + i^{-(2q+1)}} \le \frac{n}{2} \sum_{i=1}^{\infty} i^{-(2q+1)} < \infty.$$

The product in (5.2) also converges since

$$\sum_{i=1}^{\infty} \frac{\theta_{i0}^2}{8n^{-1} + 4i^{-(2q+1)}} \le \frac{n}{8} \sum_{i=1}^{\infty} \theta_{i0}^2 < \infty.$$

The lemma is proved. □

The following lemma estimates the posterior probability of the mth model.

LEMMA 5.2. *For any integers m and $n \ge 1$,*

$$E_{\theta_0} \Pi(q = q_m | \mathbf{X}) \le \frac{\lambda_m}{\lambda_0} \exp\left[\frac{1}{2} \sum_{i=1}^{\infty} \frac{(i^{-(2q_0+1)} - i^{-(2q_m+1)})(i^{-(2q_0+1)} - \theta_{i0}^2)}{i^{-2(q_0+q_m+1)} + 2n^{-1} i^{-(2q_0+1)} + n^{-2}}\right].$$

PROOF. By the martingale convergence theorem [see, e.g., Williams (1991), page 109],

$$\Pi(q = q_m|\mathbf{X}) = \lim_{k\to\infty} \Pi(q = q_m|X_1, \ldots, X_k) \qquad \text{a.s. } [P_\lambda].$$

Therefore, by Lemma 5.1, $\Pi(q = q_m|X_1, \ldots, X_k)$ converges, as $k \to \infty$, to $\Pi(q = q_m|\mathbf{X})$ a.s. $[P_{\theta_0}]$. Since these, being probabilities, are bounded by 1, the convergence also takes place in $L_1(P_{\theta_0})$. By Bayes theorem,

$$\begin{aligned}
&\Pi(q = q_m|X_1, \ldots, X_k)\\
&\quad = \frac{\lambda_m \prod_{i=1}^k (n^{-1} + i^{-(2q_m+1)})^{-1/2} \exp[-\frac{1}{2}\sum_{i=1}^k (n^{-1} + i^{-(2q_m+1)})^{-1} X_i^2]}{\sum_{l=-\infty}^{\infty} \lambda_l \prod_{i=1}^k (n^{-1} + i^{-(2q_l+1)})^{-1/2} \exp[-\frac{1}{2}\sum_{i=1}^k (n^{-1} + i^{-(2q_l+1)})^{-1} X_i^2]}\\
&\quad \le \frac{\lambda_m \prod_{i=1}^k (n^{-1} + i^{-(2q_m+1)})^{-1/2} \exp[-\frac{1}{2}\sum_{i=1}^k (n^{-1} + i^{-(2q_m+1)})^{-1} X_i^2]}{\lambda_0 \prod_{i=1}^k (n^{-1} + i^{-(2q_0+1)})^{-1/2} \exp[-\frac{1}{2}\sum_{i=1}^k (n^{-1} + i^{-(2q_0+1)})^{-1} X_i^2]}.
\end{aligned}$$

Put $a_i = (n^{-1} + i^{-(2q_m+1)})^{-1} - (n^{-1} + i^{-(2q_0+1)})^{-1}$. Exploiting the independence of the X_i's under P_{θ_0} and using

$$E\left(\exp\left[-\frac{\alpha}{2}X^2\right]\right) = \frac{1}{\sqrt{1+\alpha\sigma^2}} \exp\left[-\frac{\mu^2\alpha}{2(1+\alpha\sigma^2)}\right]$$

for X distributed as $N(\mu, \sigma^2)$ and $\alpha > -\sigma^{-2}$, we obtain

$$\begin{aligned}
&E_{\theta_0}\Pi(q = q_m|\mathbf{X})\\
&\quad = \lim_{k\to\infty} E_{\theta_0}\Pi(q = q_m|X_1, \ldots, X_k)\\
&\quad \le \frac{\lambda_m}{\lambda_0} \limsup_{k\to\infty} \prod_{i=1}^k \left\{\left(\frac{n^{-1} + i^{-(2q_0+1)}}{n^{-1} + i^{-(2q_m+1)}}\right)^{1/2} E_{\theta_0}\left(\exp\left[-\frac{a_i}{2}X_i^2\right]\right)\right\}\\
&\quad = \frac{\lambda_m}{\lambda_0} \prod_{i=1}^\infty \left\{\left(\frac{n^{-1} + i^{-(2q_0+1)}}{n^{-1} + i^{-(2q_m+1)}}\right)^{1/2} \left(1 + \frac{a_i}{n}\right)^{-1/2} \exp\left[-\frac{a_i\theta_{i0}^2}{2(1 + (a_i/n))}\right]\right\}\\
&\quad = \frac{\lambda_m}{\lambda_0} \prod_{i=1}^\infty \left\{\left(1 + \frac{a_i i^{-(2q_0+1)}}{1 + (a_i/n)}\right)^{1/2} \exp\left[-\frac{a_i\theta_{i0}^2}{2(1 + (a_i/n))}\right]\right\}\\
&\quad \le \frac{\lambda_m}{\lambda_0} \exp\left[\frac{1}{2}\sum_{i=1}^\infty \frac{a_i}{1 + (a_i/n)}(i^{-(2q_0+1)} - \theta_{i0}^2)\right]\\
&\quad = \frac{\lambda_m}{\lambda_0} \exp\left[\frac{1}{2}\sum_{i=1}^\infty \frac{(i^{-(2q_0+1)} - i^{-(2q_m+1)})(i^{-(2q_0+1)} - \theta_{i0}^2)}{i^{-2(q_0+q_m+1)} + 2n^{-1}i^{-(2q_0+1)} + n^{-2}}\right].
\end{aligned}$$

The last three steps follow from some algebra and the inequality $1 + x \le e^x$ for all x. $\square$

REMARK 5.1. By exactly the same arguments, we can also conclude that for any l,

$$E_{\theta_0}\Pi(q=q_m|\mathbf{X}) \le \frac{\lambda_m}{\lambda_l}\exp\left[\frac{1}{2}\sum_{i=1}^{\infty}\frac{(i^{-(2q_l+1)}-i^{-(2q_m+1)})(i^{-(2q_l+1)}-\theta_{i0}^2)}{i^{-2(q_l+q_m+1)}+2n^{-1}i^{-(2q_l+1)}+n^{-2}}\right].$$

PROOF OF LEMMA 3.1. Set

$$S_1=\sum_{i=1}^{\infty}\frac{(i^{-(2q_0+1)}-i^{-(2q_m+1)})i^{-(2q_0+1)}}{i^{-2(q_0+q_m+1)}+2n^{-1}i^{-(2q_0+1)}+n^{-2}}$$

and

$$S_2=-\sum_{i=1}^{\infty}\frac{(i^{-(2q_0+1)}-i^{-(2q_m+1)})\theta_{i0}^2}{i^{-2(q_0+q_m+1)}+2n^{-1}i^{-(2q_0+1)}+n^{-2}}.$$

We shall show that there exists an N not depending on m, $m<0$, such that for $n>N$,

$$S_1 \le -\tfrac{1}{12}n^{1/(q_0+q_m+1)} \tag{5.3}$$

and

$$S_2 \le \tfrac{1}{24}n^{1/(q_0+q_m+1)}. \tag{5.4}$$

Note that

$$i^{-(2q_0+1)}-i^{-(2q_m+1)} \le \begin{cases} 0, & \text{for all } i,\\ -\frac{1}{2}i^{-(2q_m+1)}, & \text{for } i>I,\end{cases}$$

where $I=2^{1/(2(q_0-q_{-1}))}$. Also if $n>N_1=2^{(q_0+q_{-1}+1)/(q_0-q_{-1})}$, then for $i\le n^{1/(q_0+q_m+1)}$, the first term in the common denominator in the expressions for S_1 and S_2 dominates the other two terms. Piecing these facts together, we see that the terms in S_1 are less than or equal to $-1/6$ for $I<i\le n^{1/(q_0+q_m+1)}$ and less than or equal to zero in general. Thus

$$S_1 \le -\tfrac{1}{6}\left(\lfloor n^{1/(q_0+q_m+1)}\rfloor - I\right) \le -\tfrac{1}{12}n^{1/(q_0+q_m+1)}$$

if $n>N_2=(2I+2)^{2q_0+1}$, and so (5.3) follows.

Let $C_0=\max(1,\sum_{i=1}^{\infty}i^{2q_0}\theta_{i0}^2)$. Now for any k,

$$\begin{aligned} S_2 &\le \sum_{i=1}^{\infty}\frac{i^{-(2q_m+1)}\theta_{i0}^2}{i^{-2(q_0+q_m+1)}+2n^{-1}i^{-(2q_0+1)}+n^{-2}}\\ &\le \sum_{i=1}^{k} i^{2q_0+1}\theta_{i0}^2 + n^2\sum_{i=k+1}^{\infty} i^{-(2q_m+1)}\theta_{i0}^2 \qquad (5.5)\\ &\le kC_0 + n^2(k+1)^{-(2q_0+2q_m+1)}\sum_{i=k+1}^{\infty} i^{2q_0}\theta_{i0}^2. \end{aligned}$$

The first term can be made less than or equal to $(48)^{-1}n^{1/(q_0+q_m+1)}$ by choosing $k = \lfloor (48C_0)^{-1}n^{1/(q_0+q_m+1)}\rfloor$. Since $k+1 \geq (48C_0)^{-1}n^{1/(2q_0+1)}$, the second term is also less than or equal to $(48)^{-1}n^{1/(q_0+q_m+1)}$ for $n > N_3$, where N_3 is the smallest n such that

$$\sum_{i \geq (48C_0)^{-1}n^{1/(2q_0+1)}} i^{2q_0}\theta_{i0}^2 < (48C_0)^{-2(2q_0+1)}. \tag{5.6}$$

Note that N_1, N_2 and N_3 do not depend on a particular $m < 0$. Thus for $n > N = \max(N_1, N_2, N_3)$, (5.3) and (5.4) follow. Equation (3.2) now follows from Lemma 5.2. Equation (3.3) is an immediate consequence of (3.2). □

6. Proof of Lemmas 3.2 and 3.3. To simplify notation, we drop the overhead tildes from $\tilde{\Pi}$ and $\tilde{\lambda}_m$'s and simply write $\Pi = \sum_{m=1}^{\infty} \lambda_m \Pi_m$, where $\sum_{m=1}^{\infty} \lambda_m = 1$.

PROOF OF LEMMA 3.2. By Chebyshev's inequality,

$$\Pi\left\{\boldsymbol{\theta} : \sum_{i=1}^{\infty} i^{2q_0}\theta_i^2 > B \middle| \mathbf{X}\right\} \leq B^{-1}\sum_{i=1}^{\infty} i^{2q_0} E(\theta_i^2|\mathbf{X}).$$

Now

$$\begin{aligned} E(\theta_i^2|\mathbf{X}) &= \sum_{m=1}^{\infty} \Pi(q = q_m|\mathbf{X}) E(\theta_i^2|\mathbf{X}, q = q_m) \\ &= \sum_{m=1}^{\infty} \Pi(q = q_m|\mathbf{X}) \left\{\frac{1}{n + i^{2q_m+1}} + \frac{n^2 X_i^2}{(n + i^{2q_m+1})^2}\right\} \\ &\leq \frac{1}{n + i^{2q_1+1}} + \frac{n^2 X_i^2}{(n + i^{2q_0+1})^2}. \end{aligned}$$

Therefore,

$$\begin{aligned} &E_{\theta_0}\Pi\left\{\boldsymbol{\theta} : \sum_{i=1}^{\infty} i^{2q_0}\theta_i^2 > B \middle| \mathbf{X}\right\} \\ &\quad\leq B^{-1}\sum_{i=1}^{\infty} i^{2q_0}\left\{\frac{1}{n + i^{2q_1+1}} + \frac{n^2 E_{\theta_0} X_i^2}{(n + i^{2q_0+1})^2}\right\} \\ &\quad= B^{-1}\left\{\sum_{i=1}^{\infty}\frac{i^{2q_0}}{n + i^{2q_1+1}} + \sum_{i=1}^{\infty}\frac{n^2 i^{2q_0}\theta_{i0}^2}{(n + i^{2q_0+1})^2} + \sum_{i=1}^{\infty}\frac{n i^{2q_0}}{(n + i^{2q_0+1})^2}\right\}. \end{aligned}$$

It therefore suffices to show that the sums in the above display are finite and bounded in n. The first two sums are clearly so because these are bounded by the convergent series $\sum_{i=1}^{\infty} i^{-1-2(q_1-q_0)}$ and $\sum_{i=1}^{\infty} i^{2q_0}\theta_{i0}^2$, respectively. We split the last sum into sums over ranges $\{i : 1 \le i \le k\}$ and $\{i : i > k\}$, with $k = \lfloor n^{1/(2q_0+1)} \rfloor$. Then

$$\sum_{i=1}^{\infty} \frac{n i^{2q_0}}{(n + i^{2q_0+1})^2} \le n^{-1} \sum_{i=1}^{k} i^{2q_0} + n \sum_{i=k+1}^{\infty} i^{-2q_0-2}$$
$$\le \frac{n^{-1} k^{2q_0+1} + n(k+1)^{-2q_0-1}}{2q_0+1},$$

which is bounded by $2/(2q_0+1)$. □

To prove Lemma 3.3, it is more convenient to view the sample as n i.i.d. observations $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n$, where each $\mathbf{Y}_j$ is distributed like $\mathbf{Y} = (Y_1, Y_2, \dots)$ with distribution P_θ: the Y_i's are independently distributed as $N(\theta_i, 1)$, $i = 1, 2, \dots$. Note that here we distinguish between this P_θ and $P_{\theta,n}$, the distribution of $\mathbf{X}$. We need some preparatory lemmas.

LEMMA 6.1. *For any* $\boldsymbol{\theta}, \boldsymbol{\theta}_0 \in \ell_2$:

(i) P_θ *is absolutely continuous with respect to* P_{θ_0}.

(ii) $\sum_{i=1}^{\infty} (Y_i - \theta_{i0})(\theta_i - \theta_{i0})$ *converges a.s.* $[P_{\theta_0}]$ *and in* $L_2(P_{\theta_0})$, *and has mean* 0 *and variance* $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2$.

(iii) $$\frac{dP_\theta}{dP_{\theta_0}}(\mathbf{Y}) = \exp\left[\sum_{i=1}^{\infty} (Y_i - \theta_{i0})(\theta_i - \theta_{i0}) - \frac{1}{2}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\right].$$

(iv) $$-\int \log \frac{dP_\theta}{dP_{\theta_0}}(\mathbf{y}) dP_{\theta_0}(\mathbf{y}) = \frac{1}{2}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2.$$

(v) $$\int \left(\log \frac{dP_\theta}{dP_{\theta_0}}(\mathbf{y})\right)^2 dP_{\theta_0}(\mathbf{y}) = \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2 + \frac{1}{4}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^4.$$

(vi) *The Hellinger distance*

$$H(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \left(\int \left(\left[\left(\frac{dP_\theta}{dP_{\theta_0}}\right)(\mathbf{y})\right]^{1/2} - 1\right)^2 dP_{\theta_0}(\mathbf{y})\right)^{1/2}$$

satisfies

$$H^2(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = 2\left(1 - \exp\left[-\tfrac{1}{8}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2\right]\right), \tag{6.1}$$

so that, in particular, $H(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \le \frac{1}{2}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|$, *and if* $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le 1$, *then* $H(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \ge e^{-1/16}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|/2$.

PROOF. Whereas P_θ and P_{θ_0} are product measures and the infinite product of affinities of their respective components converges to $\exp[-\frac{1}{8}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2]$, (i) follows from Kakutani's criterion as in the proof of Lemma 5.1. For (ii), consider the mean zero martingale $S_k = \sum_{i=1}^k (Y_i - \theta_{i0})(\theta_i - \theta_{i0})$ and note that $\sup_{k\geq 1} ES_k^2 = \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2 < \infty$. The martingale convergence theorem [see, e.g., Williams (1991), page 109] now applies. For (iii), note that on the sigma-field generated by $(Y_1, Y_2, \dots, Y_k)$, the Radon–Nikodym derivative is given by

$$\exp\left[\sum_{i=1}^k (Y_i - \theta_{i0})(\theta_i - \theta_{i0}) - \frac{1}{2}\sum_{i=1}^k (\theta_i - \theta_{i0})^2\right].$$

The rest follows from (ii) and the martingale convergence theorem. Assertions (iv) and (v) are immediate consequences of (ii) and (iii). For (vi), note that $H^2(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = 2 - 2E_{\theta_0}((dP_\theta/dP_{\theta_0})(\mathbf{Y}))^{1/2}$. To evaluate $E_{\theta_0}((dP_\theta/dP_{\theta_0})(\mathbf{Y}))^{1/2}$, consider the martingale

$$S_k = \exp\left[\frac{1}{2}\sum_{i=1}^k (Y_i - \theta_{i0})(\theta_i - \theta_{i0}) - \frac{1}{8}\sum_{i=1}^k (\theta_i - \theta_{i0})^2\right].$$

Clearly S_k is positive, has unit expectation and bounded second moment, so its limit $(dP_\theta/dP_{\theta_0})^{1/2}(\mathbf{Y})\exp[\frac{1}{8}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2]$ also has expectation 1. This implies (6.1). The next assertion in (vi) now follows from the inequality $1 - e^{-x} \leq x$ and the last from the mean value theorem. □

The following lemma, which estimates the probability of a ball under product normal measure from below, will be used in the calculation of certain prior probabilities. The result appeared as Lemma 5 in Shen and Wasserman (2001). Its simple proof is included for completeness.

LEMMA 6.2. *Let $W_1, \dots, W_N$ be independent random variables, with W_i having distribution $N(-\xi_i, i^{-2d})$, $d > 0$. Then*

$$\Pr\left\{\sum_{i=1}^N W_i^2 \leq \delta^2\right\} \geq 2^{-N/2}e^{-dN}\exp\left[-\sum_{i=1}^N i^{2d}\xi_i^2\right]\Pr\left\{\sum_{i=1}^N V_i^2 \leq 2\delta^2 N^{2d}\right\},$$

where $V_1, \dots, V_N$ are independent standard normal random variables.

PROOF. Using $(a+b)^2 \leq 2(a^2+b^2)$, $N! \geq e^{-N}N^N$ and the change of variable $v_i = \sqrt{2}N^d w_i$, we obtain

$$\begin{aligned}&\Pr\left\{\sum_{i=1}^N W_i^2 \leq \delta^2\right\}\\ &\quad= \int_{\sum_{i=1}^N w_i^2 \leq \delta^2} \prod_{i=1}^N \left\{(2\pi)^{-1/2} i^d \exp[-\tfrac{1}{2} i^{2d}(w_i + \xi_i)^2]\right\} dw_1 \cdots dw_N\end{aligned}$$

$$
\begin{aligned}
&\geq (2\pi)^{-N/2}(N!)^d \int_{\sum_{i=1}^N w_i^2 \leq \delta^2} \exp\left[-\sum_{i=1}^N i^{2d}(w_i^2+\xi_i^2)\right] dw_1\cdots dw_N \\
&\geq (2\pi)^{-N/2}(N!)^d \exp\left(-\sum_{i=1}^N i^{2d}\xi_i^2\right) \\
&\qquad \times \int_{\sum_{i=1}^N w_i^2 \leq \delta^2} \exp\left[-N^{2d}\sum_{i=1}^N w_i^2\right] dw_1\cdots dw_N \\
&\geq (N!)^d \exp\left(-\sum_{i=1}^N i^{2d}\xi_i^2\right) 2^{-N/2}N^{-dN}(2\pi)^{-N/2} \\
&\qquad \times \int_{\sum_{i=1}^N v_i^2 \leq 2N^{2d}\delta^2} \exp\left(-\tfrac{1}{2}\sum_{i=1}^N v_i^2\right) dv_1\cdots dv_N \\
&\geq 2^{-N/2}e^{-dN}\exp\left(-\sum_{i=1}^N i^{2d}\xi_i^2\right)\Pr\left\{\sum_{i=1}^N V_i^2 \leq 2N^{2d}\delta^2\right\}. \qquad \square
\end{aligned}
$$

For simplicity, we now drop the overhead bars from $\bar{\Pi}$ and $\bar{\lambda}$ and simply write $\Pi = \sum_{m=0}^\infty \lambda_m \Pi_m$, where $\sum_{m=0}^\infty \lambda_m = 1$.

LEMMA 6.3. *There exist positive constants C, c such that for all $\varepsilon > 0$,*

$$
\Pi\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \varepsilon\} \geq C\lambda_0 \exp[-c\varepsilon^{-1/q_0}]. \tag{6.2}
$$

PROOF. Clearly the left-hand side (LHS) of (6.2) is bounded from below by $\lambda_0\Pi_0\{\boldsymbol{\theta} : \sum_{i=1}^\infty(\theta_i-\theta_{i0})^2 \leq \varepsilon^2\}$. Now by independence, for any N,

$$
\begin{aligned}
&\Pi_0\left\{\boldsymbol{\theta} : \sum_{i=1}^\infty(\theta_i-\theta_{i0})^2 \leq \varepsilon^2\right\} \\
&\qquad \geq \Pi_0\left\{\boldsymbol{\theta} : \sum_{i=1}^N(\theta_i-\theta_{i0})^2 \leq \varepsilon^2/2\right\}\Pi_0\left\{\boldsymbol{\theta} : \sum_{i=N+1}^\infty(\theta_i-\theta_{i0})^2 \leq \varepsilon^2/2\right\}.
\end{aligned}
$$

Also

$$
\sum_{i=N+1}^\infty(\theta_i-\theta_{i0})^2 \leq 2\sum_{i=N+1}^\infty \theta_i^2 + 2\sum_{i=N+1}^\infty \theta_{i0}^2. \tag{6.3}
$$

The second sum in (6.3) is less than or equal to

$$
2N^{-2q_0}\sum_{i=N+1}^\infty i^{2q_0}\theta_{i0}^2 \leq 2N^{-2q_0}C_0 < \frac{\varepsilon^2}{4}
$$

whenever $N > N_1 = (8C_0)^{1/2q_0}\varepsilon^{-1/q_0}$, where $C_0 = \sum_{i=1}^{\infty} i^{2q_0}\theta_{i0}^2$. By Chebyshev's inequality, the first sum on the right-hand side (RHS) of (6.3) is less than $\varepsilon^2/4$ with probability at least

$$1 - \frac{8}{\varepsilon^2}\sum_{i=N+1}^{\infty} E_{\Pi_0}(\theta_i^2) = 1 - \frac{8}{\varepsilon^2}\sum_{i=N+1}^{\infty} i^{-(2q_0+1)} \geq 1 - \frac{4}{q_0 N^{2q_0}\varepsilon^2} > \frac{1}{2}$$

if $N > N_2 = (8/q_0)^{1/2q_0}\varepsilon^{-1/q_0}$. To bound $\Pi_0\{\boldsymbol{\theta}: \sum_{i=1}^{N}(\theta_i - \theta_{i0})^2 \leq \varepsilon^2/2\}$, we apply Lemma 6.2 with $d = q_0 + \frac{1}{2}$, $\xi_i = \theta_{i0}$ and $\delta^2 = \varepsilon^2/2$. Note that by the central limit theorem, the factor $\Pr\{\sum_{i=1}^{N} V_i^2 \leq 2\delta^2 N^{2d}\}$ on the RHS of the inequality from Lemma 6.2 is at least $\frac{1}{4}$ if $2\delta^2 N^{2d} > N$ and N is large, that is, if $N > N_3 = \varepsilon^{-1/q_0}$ and $N > N_4$. Choosing $N = \max(N_1, N_2, N_3, N_4)$ and noting that $\sum_{i=1}^{N} i^{2q_0+1}\theta_{i0}^2 \leq NC_0$, (6.2) is obtained. □

The final lemma gives the entropy estimates. Recall that for a totally bounded subset S of a metric space with a distance function d, the ε-packing number $D(\varepsilon, S, d)$ is defined to be the largest integer m such that there exist $s_1, \ldots, s_m \in S$ with $d(s_j, s_k) > \varepsilon$ for all $j, k = 1, 2, \ldots, m$, $j \neq k$. The ε-entropy is $\log D(\varepsilon, S, d)$. A bound for the entropy of Sobolev balls was obtained in Theorem 5.2 of Birman and Solomjak (1967). It may be possible to derive a bound similar to that in Lemma 6.4 from their result by a Fourier transformation. However, the exact correspondence is somewhat unclear because of the possible fractional value of the index q. Below, we present a simple proof which could be of some independent interest as well.

LEMMA 6.4. *For all* $\varepsilon > 0$,

$$\log D\left(\varepsilon, \left\{\boldsymbol{\theta}: \sum_{i=1}^{\infty} i^{2q}\theta_i^2 \leq B\right\}, \|\cdot\|\right) \leq \tfrac{1}{2}(8B)^{1/(2q)} \log\left(4(2e)^{2q}\right)\varepsilon^{-1/q}. \tag{6.4}$$

PROOF. Denote $\Theta_q(B) = \{\boldsymbol{\theta}: \sum_{i=1}^{\infty} i^{2q}\theta_i^2 \leq B\}$. If $\varepsilon^2 > 4B$, then

$$\sum_{i=1}^{\infty}\theta_i^2 \leq \sum_{i=1}^{\infty} i^{2q}\theta_i^2 \leq B < \frac{\varepsilon^2}{4}$$

and hence for any two $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta_q(B)$, $\|\boldsymbol{\theta} - \boldsymbol{\theta}'\| < \varepsilon$, implying that $D(\varepsilon, \Theta_q(B)) = 1$. Thus (6.4) is trivially satisfied for such an ε. We shall therefore assume that $\varepsilon^2 \leq 4B$.

Let $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m \in \Theta_q(B)$ be such that $\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\| > \varepsilon$, $j, k = 1, 2, \ldots, m$, $j \neq k$. For an integer N, consider the set

$$\Theta_{q,N}(B) = \left\{(\theta_1, \ldots, \theta_N, 0, \ldots): \sum_{i=1}^{N} i^{2q}\theta_i^2 \leq B\right\}.$$

With $\bar{\boldsymbol{\theta}} = (\theta_1, \dots, \theta_N, 0, \dots)$, note that for any $\boldsymbol{\theta} \in \Theta_q(B)$,

$$\|\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\|^2 = \sum_{i=N+1}^{\infty} \theta_i^2 \leq (N+1)^{-2q} B \leq \frac{\varepsilon^2}{8}$$

if $N = \lfloor (8B)^{1/(2q)} \varepsilon^{-1/q} \rfloor$. For this choice of N, we have

$$1 \leq N \leq (8B)^{1/(2q)} \varepsilon^{-1/q} < N + 1 \leq 2N. \tag{6.5}$$

Therefore, for $j, k = 1, 2, \dots, m,\ j \neq k$,

$$\varepsilon^2 < \|\boldsymbol{\theta}_j - \boldsymbol{\theta}_k\|^2 = \|\bar{\boldsymbol{\theta}}_j - \bar{\boldsymbol{\theta}}_k\|^2 + \|(\boldsymbol{\theta}_j - \bar{\boldsymbol{\theta}}_j) - (\boldsymbol{\theta}_k - \bar{\boldsymbol{\theta}}_k)\|^2 \leq \|\bar{\boldsymbol{\theta}}_j - \bar{\boldsymbol{\theta}}_k\|^2 + 4\frac{\varepsilon^2}{8},$$

implying that $\|\bar{\boldsymbol{\theta}}_j - \bar{\boldsymbol{\theta}}_k\| > \varepsilon/\sqrt{2}$. The set $\Theta_{q,N}(B)$ may be viewed as an N-dimensional ellipsoid and the $\bar{\boldsymbol{\theta}}_j$'s as N-dimensional vectors in $\mathbb{R}^N$ with the usual Euclidean distance. For $\mathbf{t} = (t_1, \dots, t_N)$ and $\delta > 0$, let $B(\mathbf{t}, \delta)$ stand for the Euclidean ball with radius δ centered at $\mathbf{t}$. The balls $B(\bar{\boldsymbol{\theta}}_j, \varepsilon/2\sqrt{2})$ are clearly disjoint. Note that if $\mathbf{t} \in \Theta_q(B)$ and $\mathbf{t}' = (t_1', \dots, t_N') \in B(\mathbf{t}, \varepsilon/2\sqrt{2})$, we have

$$\sum_{i=1}^{N} i^{2q} t_i'^2 \leq 2 \sum_{i=1}^{N} i^{2q} t_i^2 + 2 \sum_{i=1}^{N} i^{2q} (t_i' - t_i)^2 \leq 2B + 2N^{2q} \frac{\varepsilon^2}{8} \leq 4B$$

by (6.5), and hence

$$\bigcup_{j=1}^{m} B\left(\bar{\boldsymbol{\theta}}_j, \frac{\varepsilon}{2\sqrt{2}}\right) \subset \Theta_{q,N}(4B).$$

Let V_N stand for the volume of the unit ball in $\mathbb{R}^N$. Note that the volume of the ellipsoid $\{\mathbf{t} : \sum_{i=1}^N a_i^2 t_i^2 \leq A\}$ is given by $V_N A^{N/2} / \prod_{i=1}^N a_i$. Bounding the volume of $\bigcup_{j=1}^m B(\bar{\boldsymbol{\theta}}_j, \varepsilon/2\sqrt{2})$ by that of $\Theta_{q,N}(4B)$, we obtain

$$m \left(\frac{\varepsilon^2}{8}\right)^{N/2} V_N \leq (4B)^{N/2} V_N \prod_{i=1}^{N} i^{-q} = (4B)^{N/2} V_N (N!)^{-q}.$$

Since $N! \geq N^N e^{-N}$ for all $N \geq 1$, we arrive at the estimate

$$m \leq \left(\frac{32 B e^{2q}}{N^{2q} \varepsilon^2}\right)^{N/2}. \tag{6.6}$$

From (6.5), we also have $N^{2q} \varepsilon^2 \geq 2^{3-2q} B$. Substituting in (6.6), the required bound is obtained. □

To prove Lemma 3.3, we shall use the following variation of Theorem 2.1 of Ghosal, Ghosh and van der Vaart (2000).

THEOREM 6.1. *Suppose we have i.i.d. observations $Y_1, Y_2, \dots$ from a distribution P having a density p (with respect to some sigma-finite measure) belonging to a class $\mathcal{P}$. Let $H(P_0, P)$ denote the Hellinger distance between P_0 and P. Let Π be a prior on $\mathcal{P}$ and $\bar{\mathcal{P}} \subset \mathcal{P}$. Let ε_n be a positive sequence such that $\varepsilon_n \to 0$ and $n\varepsilon_n^2 \to \infty$ and suppose that*

$$\log D(\varepsilon_n, \bar{\mathcal{P}}, H) \le c_1 n \varepsilon_n^2 \tag{6.7}$$

and

$$\Pi\left\{P : -\int \left(\log \frac{p}{p_0}\right) dP_0 \le \varepsilon_n^2, \int \left(\log \frac{p}{p_0}\right)^2 dP_0 \le \varepsilon_n^2\right\} \ge c_2 e^{-c_3 n \varepsilon_n^2} \tag{6.8}$$

for some constants c_1, c_2, c_3, where p_0 stands for the density of P_0. Then for a sufficiently large constant M, the posterior probability

$$\Pi\{P \in \bar{\mathcal{P}} : H(P_0, P) \ge M\varepsilon_n | Y_1, Y_2, \dots, Y_n\} \to 0 \tag{6.9}$$

in P_0^n-probability.

PROOF OF LEMMA 3.3. Take $\mathcal{P} = \{P_\theta : \boldsymbol{\theta} \in \ell_2\}$ and $\bar{\mathcal{P}} = \{P_\theta : \sum_{i=1}^\infty i^{2q_0} \times \theta_i^2 \le B\}$. By part (vi) of Lemma 6.1, the Hellinger distance $H(P_{\theta_0}, P_\theta)$ is equivalent to the ℓ_2 distance $\|\boldsymbol{\theta}_0 - \boldsymbol{\theta}\|$, so in (6.7) and in the conclusion (6.9), the former may be replaced by the latter. Lemma 6.4 shows that (6.7) is satisfied by a multiple of $n^{-q_0/(2q_0+1)}$. For (6.8), note that by parts (iv) and (v) of Lemma 6.1, the set

$$\left\{\boldsymbol{\theta} : -E_{\theta_0}\left(\log \frac{dP_\theta}{dP_{\theta_0}}(\mathbf{Y})\right) \le \varepsilon^2, E_{\theta_0}\left(\log \frac{dP_\theta}{dP_{\theta_0}}(\mathbf{Y})\right)^2 \le \varepsilon^2\right\}$$

contains $\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2 \le \varepsilon^2/2\}$ for $\varepsilon < 1$. Therefore, by Lemma 6.3, (6.8) is also satisfied by a multiple of $n^{-q_0/(2q_0+1)}$. The result thus follows. □

7. Continuous spectrum adaptation.

Suppose now that $q_0 \in \mathcal{Q}$, where $\mathcal{Q}$ is an arbitrary subset of the positive semiaxis $\mathbb{R}_+$. Since our basic approach relies on the assumption that the smoothness parameter can only take countably many possible values, we proceed as follows. Choose a countable dense subset $\mathcal{Q}^*$ of $\mathcal{Q}$ and put a prior λ on it such that $\lambda(q = s) > 0$ for each $s \in \mathcal{Q}^*$.

THEOREM 7.1. *For any sequence $M_n \to \infty$ and for any given $\delta > 0$, the posterior probability*

$$\Pi\{\boldsymbol{\theta} : n^{q_0/(2q_0+1)-\delta}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n|\mathbf{X}\} \to 0$$

in P_{θ_0}-probability as $n \to \infty$.

Note that the prior does not depend on δ, so the obtained rate is almost optimal. The proof of the following theorem is very similar to that of Theorem 3.1. The main differences in the proofs are indicated in the outline of the proof below.

OUTLINE OF THE PROOF. The true q_0 may not belong to $\mathcal{Q}^*$, but we may choose $q^* \in \mathcal{Q}^*$, $q^* < q_0$ arbitrarily close to q_0. Choose also $\bar{q}, \tilde{q} \in \mathcal{Q}^*$ which satisfy $\bar{q} < q^* < \tilde{q} < q_0$, and hence $\boldsymbol{\theta}_0 \in \Theta_{q_0} \subset \Theta_{\tilde{q}} \subset \Theta_{q^*} \subset \Theta_{\bar{q}}$.

Lemmas 3.1 and 3.2 go through in the sense that $\Pi(q < q^*|\mathbf{X}) \to 0$ as $n \to \infty$ and

$$\sup_n \Pi\left\{\boldsymbol{\theta} : \sum_{i=1}^{\infty} i^{2\bar{q}} \theta_i^2 > B, q \geq q^* \middle| \mathbf{X}\right\} \to 0 \qquad \text{as } B \to \infty$$

in P_{θ_0}-probability. To see this, apply Lemma 3.1 with q_0 replaced by $\tilde{q}$ and q_{-1} by q^* and Lemma 3.2 with q_0 replaced by $\bar{q}$ and q_1 by q^*.

As in the proof of Theorem 3.1, we obtain the bound

$$\begin{aligned}
&\Pi\{\boldsymbol{\theta} : n^{\bar{q}/(2\bar{q}+1)} \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n |\mathbf{X}\} \\
&\quad = \Pi\{\boldsymbol{\theta} : n^{\bar{q}/(2\bar{q}+1)} \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n, q \geq q^* |\mathbf{X}\} \\
&\qquad + \Pi\{\boldsymbol{\theta} : n^{\bar{q}/(2\bar{q}+1)} \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n, q < q^* |\mathbf{X}\} \\
&\quad \leq \Pi\left\{\boldsymbol{\theta} : \sum_{i=1}^{\infty} i^{2\bar{q}} \theta_i^2 > B, q \geq q^* \middle| \mathbf{X}\right\} \\
&\qquad + \Pi\left\{\boldsymbol{\theta} : n^{\bar{q}/(2\bar{q}+1)} \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n, \sum_{i=1}^{\infty} i^{2\bar{q}} \theta_i^2 \leq B, q \geq q^* \middle| \mathbf{X}\right\} \\
&\qquad + \Pi(q < q^* |\mathbf{X}).
\end{aligned}$$

Note that the fourth term on the RHS of the series of inequalities in the proof of Theorem 3.1 has been absorbed into the first two terms on the RHS of the last inequality. Now, the second term on the RHS of the last display converges to zero by Lemma 3.3 with q_0 replaced by $\bar{q}$. Therefore all the terms go to zero. Since $\bar{q}$ can be made arbitrarily close to q_0, this means that for any given $\delta > 0$, the posterior converges at least at the rate $n^{-q_0/(2q_0+1)+\delta}$ in ℓ_2. □

REMARK 7.1. The loss of a power δ is due to the fact that the q's are not strictly separated. It is intuitively clear that the requirement of strict separation stems from the fact that the closer the possible values of the smoothness parameter are, the more difficult it is to distinguish between them.

REMARK 7.2. It is also natural to use a prior λ on the whole set $\mathcal{Q}$, for instance, a positive density $\lambda(q)$ on $\mathbb{R}_+$. This time, however, our approach based on the estimates in Lemmas 5.2 and 6.3 does not seem to work because of the absence of point masses. Nevertheless, we believe that Theorem 7.1 continues to hold in this case, but a complete analogue of Theorem 3.1 will possibly not hold. The intuition behind our belief is essentially the same as in Theorem 7.1. Since the slightest deviation in the assumed value of q from the true q_0 affects the rate by a power on n, and without a point mass at q_0, one can at best hope for a concentration of the posterior distribution of q values in a neighborhood of q_0, the rate of convergence of the posterior distribution will be affected by an arbitrarily small power of n.

8. Uniformity over ellipsoids. It is of some interest to know the extent to which the posterior convergence in Theorems 2.1, 3.1 and 7.1 is uniform over $\boldsymbol{\theta}_0$; see Brown, Low and Zhao (1997) in this context. For Theorem 2.1, it is immediately seen from the proof that (2.1) holds uniformly over an ellipsoid, that is, for any $Q \geq 0$ and $M_n \to \infty$,

$$\sup_{\boldsymbol{\theta}_0 \in \Theta_q(Q)} E_{\theta_0} \Pi_q \{\boldsymbol{\theta} : n^{q/(2q+1)} \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n | \mathbf{X}\} \to 0.$$

For Theorem 3.1, such a proposition is much more subtle. One needs to check uniformity in each of Lemmas 3.1, 3.2 and 3.3. In Theorem 8.1 below, we show that a local uniformity holds in the sense that for every $\boldsymbol{\theta}_0 \in \Theta_0$, there exists a small ellipsoid around it where the convergence is uniform.

THEOREM 8.1. *For any $\boldsymbol{\theta}^* \in \Theta_{q_0}$ there exists $\varepsilon > 0$ such that*

$$\sup_{\boldsymbol{\theta}_0 \in \mathcal{E}_{q_0}(\boldsymbol{\theta}^*, \varepsilon)} E_{\theta_0} \Pi_q \{\boldsymbol{\theta} : n^{q_0/(2q_0+1)} \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n | \mathbf{X}\} \to 0,$$

where $\mathcal{E}_{q_0}(\boldsymbol{\theta}^*, \varepsilon) = \{\boldsymbol{\theta}_0 : \sum_{i=1}^{\infty} i^{2q_0} (\theta_{i0} - \theta_i^*)^2 < \varepsilon\}$.

OUTLINE OF THE PROOF. We need to check the uniformity in every step of the proof of Theorem 3.1. Going through the proof of Lemma 3.2, it is easily seen that for any $Q > 0$,

$$\lim_{B \to \infty} \sup_{\boldsymbol{\theta}_0 \in \Theta_{q_0}(Q)} \sup_{n \geq 1} E_{\theta_0} \tilde{\Pi} \left\{ \boldsymbol{\theta} : \sum_{i=1}^{\infty} i^{2q_0} \theta_i^2 > B \middle| \mathbf{X} \right\} = 0.$$

For the uniform version of Lemma 3.3, we need to show that for any $B > 0$,

$$\text{(8.1)} \quad \lim_{n \to \infty} \sup_{\boldsymbol{\theta}_0 \in \Theta_{q_0}(Q)} E_{\theta_0} \bar{\Pi} \left\{ \boldsymbol{\theta} : r_n \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > M_n, \sum_{i=1}^{\infty} i^{2q_0} \theta_i^2 \leq B \middle| \mathbf{X} \right\} = 0.$$

Further, note that (6.9) in Theorem 6.1 could have been stated as an inequality [see the proof of Theorem 2.1 of Ghosal, Ghosh and van der Vaart (2000)],

$$E_{\theta_0} \Pi\{P \in \bar{\mathcal{P}} : H(P_0, P) \geq M\varepsilon_n | Y_1, Y_2, \ldots, Y_n\} \leq C e^{-cn\varepsilon_n^2} + \frac{1}{n\varepsilon_n^2},$$

where C and c are positive constants depending on c_1, c_2 and c_3 that appeared in the conditions of Theorem 6.1. Therefore, (8.1) holds if the conditions of Theorem 6.1 hold simultaneously for $\boldsymbol{\theta}_0 \in \Theta_{q_0}(Q)$. The entropy equation (6.7) does not depend on the true value $\boldsymbol{\theta}_0$ [see relation (6.4) in Lemma 6.4], so c_1 is free of $\boldsymbol{\theta}_0$. It is possible to choose the same c_2 and c_3 for $\boldsymbol{\theta}_0$ belonging to the ellipsoid $\Theta_q(Q)$ by the estimates given by Lemmas 6.2 and 6.3. Therefore, it remains to check uniformity in Lemma 3.1.

However, (3.3) in Lemma 3.1 does not hold uniformly over $\boldsymbol{\theta}_0 \in \Theta_{q_0}(Q)$ for a given $Q > 0$. The problem is that by varying $\boldsymbol{\theta}_0$ over $\Theta_{q_0}(Q)$, one can encounter arbitrarily slow convergence of the LHS of (5.6) to zero, making the integer N_3 defined there dependent on $\boldsymbol{\theta}_0$. Nevertheless, for a given $\boldsymbol{\theta}^* = (\theta_1^*, \theta_2^*, \dots) \in \Theta_{q_0}$, it is possible to choose an $\varepsilon > 0$ such that

$$\lim_{n\to\infty} \sup_{\boldsymbol{\theta}_0 \in \mathcal{E}_{q_0}(\boldsymbol{\theta}^*, \varepsilon)} E_{\theta_0} \Pi(q < q_0 | \mathbf{X}) \to 0. \tag{8.2}$$

The proof is largely a repetition of the arguments given in the proof of Lemma 3.1, so we restrict ourselves only to those places where modifications are needed.

First, if $0 < \varepsilon < 1$, $\boldsymbol{\theta}_0 \in \mathcal{E}(\boldsymbol{\theta}^*, \varepsilon)$ and $C_0 = \sum_{i=1}^\infty i^{2q_0}(\theta_i^*)^2$, then using $(a-b)^2 \le 2(a^2+b^2)$, we have $\sum_{i=1}^\infty i^{2q_0}\theta_{i0}^2 \le D_0$, where $D_0 = 2C_0 + 2$. Proceed as in the proof of Lemma 3.1 with $k = \lfloor (48D_0)^{-1} n^{1/(q_0+q_m+1)} \rfloor$. If we choose $\varepsilon = 4^{-1}(48D_0)^{-2(2q_0+1)}$ and N_3 the smallest integer n such that

$$\sum_{i \ge (48D_0)^{-1} n^{1/(2q_0+1)}} i^{2q_0}(\theta_i^*)^2 < 4^{-1}(48D_0)^{-2(2q_0+1)},$$

then it easily follows that for $n \ge N_3$,

$$\sum_{i \ge (48D_0)^{-1} n^{1/(2q_0+1)}} i^{2q_0}\theta_{i0}^2 < (48D_0)^{-2(2q_0+1)}.$$

The proof of (8.2) now follows as before and, hence, the theorem is proved. □

REMARK 8.1. The uniformity holds for $\boldsymbol{\theta}_0$ belonging to a compact set in the topology generated by the norm $(\sum_{i=1}^\infty i^{2q_0}\theta_i^2)^{1/2}$ on Θ_{q_0}.

REMARK 8.2. In a similar manner, one can formulate the uniform version of posterior convergence for the case of a continuous spectrum as well.

Acknowledgments. We are grateful to Professor Aad van der Vaart for kindly allowing us to reproduce his construction of the adaptive estimator in Section 4. We also wish to thank Professor Lawrence D. Brown and Professor Linda Zhao for a stimulating discussion. Suggestions by the referees led to an improvement of the presentation.

REFERENCES

BIRMAN, M. S. and SOLOMJAK, M. Z. (1967). Piecewise-polynomial approximation of functions of the class W_p^α. *Soviet Math. Sb.* **73** 295–317.

BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398.

BROWN, L. D., LOW, M. G. and ZHAO, L. H. (1997). Superefficiency in nonparametric function estimation. *Ann. Statist.* **25** 2607–2625.

COX, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.* **21** 903–923.

DIACONIS, P. and FREEDMAN, D. (1997). On the Bernstein–von Mises theorem with infinite dimensional parameters. Technical Report 492, Dept. Statistics, Univ. California.

EFROMOVICH, S. Y. and PINSKER, M. S. (1984). Learning algorithm for nonparametric filtering. *Autom. Remote Control* **45** 1434–1440.

FREEDMAN, D. (1999). On the Bernstein–von Mises theorem with infinite dimensional parameters. *Ann. Statist.* **27** 1119–1140.

GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531.

GHOSAL, S., LEMBER, J. and VAN DER VAART, A. W. (2002). On Bayesian adaptation. In *Proceedings of 8th Vilnius Conference on Probability and Statistics*. To appear.

HUANG, T. M. (2000). Convergence rates for posterior distributions and adaptive estimation. Preprint.

KLEMELÄ, J. and NUSSBAUM, M. (1998). Constructive asymptotic equivalence of density estimation and Gaussian white noise. Technical Report 53, Humboldt Univ., Berlin.

NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430.

PINSKER, M. S. (1980). Optimal filtration of square-integrable signals in Gaussian white noise. *Problems Inform. Transmission* **16** 120–133.

SHEN, X. and WASSERMAN, L. (2001). Rates of convergence of posterior distributions. *Ann. Statist.* **29** 687–714.

VAN DER VAART, A. W. (2000). Private communication.

WILLIAMS, D. (1991). *Probability with Martingales*. Cambridge Univ. Press.

ZHAO, L. H. (2000). Bayesian aspects of some nonparametric problems. *Ann. Statist.* **28** 532–552.

DEPARTMENT OF MATHEMATICS
UTRECHT UNIVERSITY
P.O. BOX 80010
3508 TA UTRECHT
THE NETHERLANDS
E-MAIL: belitser@math.uu.nl

DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
220 PATTERSON HALL
2501 FOUNDERS DRIVE
RALEIGH, NORTH CAROLINA 27695-8203
E-MAIL: ghosal@stat.ncsu.edu