The Annals of Statistics
2001, Vol. 29, No. 3, 687–714

RATES OF CONVERGENCE OF POSTERIOR DISTRIBUTIONS

BY XIAOTONG SHEN[1] AND LARRY WASSERMAN[2]

Ohio State University and Carnegie Mellon University

We compute the rate at which the posterior distribution concentrates around the true parameter value. The spaces we work in are quite general and include infinite dimensional cases. The rates are driven by two quantities: the size of the space, as measured by bracketing entropy, and the degree to which the prior concentrates in a small ball around the true parameter. We consider two examples.

1. Introduction. Nonparametric Bayesian methods have become quite popular lately, largely because of advances in computing; see Dey, Mueller and Sinha (1998) for a recent account. Because of their growing popularity, it is important to understand the properties of these methods.

There are now many results about the asymptotic properties of posterior distributions in infinite dimensional parameter spaces. For example, see Barron, Schervish and Wasserman (1999), Cox (1993), Diaconis and Freedman (1986, 1993, 1995, 1997a, b, 1998), Doob (1948), Freedman (1963, 1999), Ghosal, Ghosh and Ramamoorthi (1997, 1999a, b), Schwartz (1965), Shen (1995), Wasserman (1998) and Zhao (1993, 1998). However, there are few general results about rates of convergence. This paper provides such results. Specifically, let $p(Y|\eta)$ be a density for the random variable Y where $\eta \in \Omega$ is an unknown parameter and let π be a prior for η. We bound the posterior probability $\pi(A_n^c|Y^n)$ where A_n is a shrinking neighborhood of the true parameter value θ and $Y^n = (Y_1, \dots, Y_n)$ are the data. We write the posterior probability as

$$\pi(A_n^c|Y^n) = \frac{\int_{A_n^c} \prod_{i=1}^n \frac{p(Y_i|\eta)}{p(Y_i|\theta)} d\pi(\eta)}{\int \prod_{i=1}^n \frac{p(Y_i|\eta)}{p(Y_i|\theta)} d\pi(\eta)}, \tag{1.1}$$

then we separately upper bound the numerator and lower bound the denominator of (1.1).

In Section 2 we introduce some notation. In Section 3 we lower bound the denominator of (1.1). In Section 4 we combine the lower bounds on the denominator with upper bounds on the numerator and give some general results on rates of convergence. Section 5 is devoted to an example involving nonparametric regression. Section 6 treats the problem of estimating infinitely

Received June 1998; revised November 2000.
[1]Supported by NSA Grant MDA904-98-1-0030 and NSF Grant DMS-00-72635.
[2]Supported by NIH Grant R01-CA54852-01 and NSF Grants DMS-93-03557 and DMS-93-57646.
AMS 2000 *subject classifications*. Primary 62A15; secondary 62E20, 62G15.
Key words and phrases. Bayesian inference, asymptotic inference, non-parametric models, sieves.

many normal means. Related literature on this problem includes Diaconis and Freedman (1997a), Freedman (1999), Zhao (1993, 1998) and Cox (1993). In Section 7 we discuss the results. Some technical lemmas are contained in the appendix.

We would like to mention that after submitting the first version of this paper, we received a report by Ghosal, Ghosh and Van der Vaart (1998) based on their independent work on the same topic. Their paper has many interesting results, and inevitably, there is some overlap.

2. Notation. The data $Y^n = (Y_1, \dots, Y_n)$ are assumed to be i.i.d. and take values in a sample space $\mathscr{Y}$ with σ-field $\mathscr{B}$. In our examples, $\mathscr{Y}$ is a subset of $\mathscr{R}^k$ and $\mathscr{B}$ is the Borel σ-field. The model is $\{P_\eta; \eta \in \Omega\}$ where each P_η is a probability on $(\mathscr{Y}, \mathscr{B})$ with density $p(y|\eta)$ with respect to a common, dominating, σ-finite measure λ. Assume that the parameter space Ω is endowed with an appropriate σ-field and let π be a prior distribution on this σ-field. Because the distributions $\{P_\eta; \eta \in \Omega\}$ are dominated by a common σ-finite measure, Bayes' theorem applies so the posterior may be written as in (1.1). Our goal is to study the behavior of $\pi(A_n^c|Y^n)$ when A_n is a small shrinking neighborhood, of the true parameter value θ.

Let $\ell(\eta, Y) = \log p(Y|\eta)$ and let

$$K(\theta, \eta) = E_\theta[\ell(\theta, Y) - \ell(\eta, Y)] \quad \text{and} \quad V(\theta, \eta) = Var_\theta[\ell(\eta, Y) - \ell(\theta, Y)]$$

be the Kullback-Leibler divergence (sometimes called Kullback-Leibler pseudo-distance) and the variance of the log-likelihood ratio based on a single observation Y, respectively. The sample Kullback-Leibler divergence is defined as

$$K_n(\theta, \eta) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{p(Y_i|\theta)}{p(Y_i|\eta)}.$$

We will use the Kullback-Leibler divergence to bound the denominator of (1.1).

We shall use two other distances: Hellinger distance, defined by

$$h(\theta, \eta) = \left\{ \int (p^{1/2}(y|\theta) - p^{1/2}(y|\eta))^2 d\lambda(y) \right\}^{1/2}.$$

and the ρ_α distance [Wong and Shen (1995)], defined by

$$\rho_\alpha(\theta, \eta) = \frac{1}{\alpha} \int \left[\left(\frac{p(y|\theta)}{p(y|\eta)} \right)^\alpha - 1 \right] dP_\theta(y)$$

where $\alpha \in (0, 1]$. The main use of the ρ_α distance is to bound the denominator of (1.1). However, the ρ_α metric gives a.s. bounds while Kullback-Leibler gives bounds in probability. Corresponding to the different distances we also define the following neighborhoods:

$$S(t) = \{\eta;\ \max[K(\theta, \eta), V(\theta, \eta)] \le t\},$$
$$S_\alpha(t) = \{\eta;\ \rho_\alpha(\theta, \eta) \le t\}.$$

It is convenient to express the posterior probability as

$$\pi(A|Y^n) = \frac{m_n(A, Y^n)}{m_n(\Omega, Y^n)} \tag{2.1}$$

where

$$m_n(B, Y^n) = \int_B \exp\{-nK_n(\theta, \eta)\} d\pi(\eta).$$

We will bound $\pi(A|Y^n)$ by upper bounding the numerator $m_n(A, Y^n)$ and lower bounding the denominator $m_n(\Omega, Y^n)$.

Now we recall a result due to Wong and Shen (1995). Let $\mathscr{P}$ be a set of density functions. A set of pairs of functions $\{(f_j^L, f_j^U), j = 1, \ldots, N\}$ is a Hellinger u-bracketing of $\mathscr{P}$ if $h(f_j^L, f_j^U) \le u$ for $j = 1, \ldots, N$ and, for every $p \in \mathscr{P}$ there is a $j \in \{1, \ldots, N\}$ such that $f_j^L \le p \le f_j^U$ a.e.-λ. The *bracketing Hellinger metric entropy* of $\mathscr{P}$, denoted by $H_B(u, \mathscr{P})$, is the logarithm of the cardinality of the u-bracketing of $\mathscr{P}$ of smallest size. If $\mathscr{P} = \{p(\cdot|\eta); \eta \in \Omega\}$ then we also write $H_B(u, \Omega)$ instead of $H_B(u, \mathscr{P})$. For $\delta > 0$ define

$$\Omega(\delta) = \{p(\cdot|\eta) \in \Omega; h(\theta, \eta) \le 2\delta\}.$$

Then, $H_B(u, \Omega(\delta))$ is called the *local bracketing entropy* of $\mathscr{F}$.

THEOREM 1 [Wong and Shen (1995)]. *Let $Y_1, \ldots, Y_n$ be i.i.d. from distribution P_0 with density p_0. There are positive constants c, a, c_1, c_2 such that, for any $\varepsilon > 0$, if*

$$\int_{\varepsilon^2/2^8}^{\sqrt{2}\varepsilon} H_B^{1/2}(u/a, \mathscr{P}) du \le c\sqrt{n}\varepsilon^2$$

then

$$P_0^*\left(\sup_{p\in A^c} \prod_{i=1}^n \frac{p(Y_i)}{p_0(Y_i)} \ge e^{-c_1 n\varepsilon^2}\right) \le 4e^{-c_2 n\varepsilon^2}$$

where

$$A = \{p; h(p_0, p) \le \varepsilon\}.$$

The constants c and c_1 may be chosen to be in $(0, 1)$. *In particular, we can take* $c = (2/3)^{5/2}/512$.

REMARK. As Wong and Shen note, the theorem is still true if $H_B(u/a, \mathscr{P})$ is replaced with the local entropy $H_B(u/a, \mathscr{P} \cap \{p; h(p_0, p) \le 2\varepsilon\})$.

Let $\rho(\cdot, \cdot)$ be a metric (or pseudo-distance) on Ω. We say that the rate of convergence of the posterior is $\varepsilon_n > 0$ if for every sequence $K_n \to \infty$, $\pi(A_n^c|Y^n)$ tends to 0 in P_θ probability where $A_n = \{\eta; \rho(\theta, \eta) \le K_n\varepsilon_n\}$. Note that in this paper we are concerned with pointwise, rather than minimax rates.

We write $a_n \sim b_n$ to mean that $a_n = O(b_n)$ and $b_n = O(a_n)$. We write $a_n \succeq b_n$ to mean that $a_n \ge cb_n$ for some $c > 0$, for all sufficiently large n.

Similarly, we write $a_n \preceq b_n$ to mean $a_n \le cb_n$ for some $c > 0$, for all sufficiently large n.

3. Lower bounds on $m_n(\Omega, Y^n)$.

To bound the posterior, we need to lower bound the denominator $m_n(\Omega, Y^n)$. The first result controls the denominator in probability.

LEMMA 1. *Let t_n be a sequence of positive numbers and let $S_n = S(t_n)$. Then,*

$$P_\theta^n\left(m_n(\Omega, Y^n) \le \frac{1}{2}\pi(S_n)e^{-2nt_n} \right) \le \frac{2}{nt_n}.$$

Consequently, if $nt_n \to \infty$, then

$$m_n(\Omega, Y^n) \ge \tfrac{1}{2}\pi(S_n)e^{-2nt_n}$$

except on a set of probability tending to 0.

PROOF. For any η we have, by Chebyshev's inequality, that

$$P_\theta^n\left(\frac{K_n(\theta,\eta) - K(\theta,\eta)}{V^{1/2}(\theta,\eta)} \ge t_n^{1/2}\right) \le P_\theta^n\left(\left(\frac{K_n(\theta,\eta) - K(\theta,\eta)}{V^{1/2}(\theta,\eta)}\right)^2 \ge t_n\right) \le \frac{1}{nt_n}.$$

Let

$$W_n = \left\{(\eta, y^n);\ \frac{K_n(\theta,\eta) - K(\theta,\eta)}{V^{1/2}(\theta,\eta)} \ge t_n^{1/2}\right\}.$$

Let $W_n(y^n) = \{\eta; (\eta, y^n) \in W_n\}$ and $W_n(\eta) = \{y^n; (\eta, y^n) \in W_n\}$. Then,

$$\begin{aligned}
m_n(\Omega, Y^n) &= \int \exp\{-nK_n(\theta,\eta)\}d\pi(\eta) \\
&\ge \int_{S_n \cap W_n^c(Y^n)} \exp\{-nK_n(\theta,\eta)\}d\pi(\eta) \\
&= \int_{S_n \cap W_n^c(Y^n)} \exp\left\{-nV^{1/2}(\theta,\eta)\frac{(K_n(\theta,\eta) - K(\theta,\eta))}{V^{1/2}(\theta,\eta)}\right\} \\
&\qquad \times \exp\{-nK(\theta,\eta)\}d\pi(\eta) \\
&\ge \pi(S_n \cap W_n^c(Y^n)) \exp\left\{-nt_n^{1/2}t_n^{1/2}\right\} \exp\left\{-nt_n\right\} \\
&= [\pi(S_n) - \pi(S_n \cap W_n(Y^n))] \exp\{-2nt_n\}.
\end{aligned}$$

By Fubini's theorem,

$$\begin{aligned}
E_\theta^n \pi(S_n \cap W_n(Y^n)) &= \int\int I(S_n \cap W_n(y^n))d\pi(\eta)dP_\theta^n(y^n) \\
&= \int\int I(S_n)I(W_n)d\pi(\eta)dP_\theta^n(y^n)
\end{aligned}$$

$$= \int I(S_n) P_\theta(W_n(\eta))\, d\pi(\eta)$$
$$\le \frac{1}{nt_n}\pi(S_n)$$

where $I(\cdot)$ is the indicator function. So,

$$\begin{aligned}
&P_\theta^n\left(m_n(\Omega, Y^n) \le \frac{1}{2}\pi(S_n)e^{-2nt_n} \right)\\
&\quad \le P_\theta^n\left([\pi(S_n) - \pi(S_n \cap W_n(Y^n))]e^{-2nt_n} \le \frac{1}{2}\pi(S_n)e^{-2nt_n} \right)\\
&\quad = P_\theta^n\left(\pi(S_n \cap W_n(Y^n)) \ge \frac{1}{2}\pi(S_n)\right)\\
&\quad \le \frac{2}{\pi(S_n)} E_\theta^n(S_n \cap W_n(Y^n)) \le \frac{2}{nt_n}.
\end{aligned}$$

This completes the proof. □

The next result controls the denominator almost surely using the ρ_α metric.

LEMMA 2. *Let t_n be a sequence of positive numbers, let $\alpha \in (0, 1]$ and let $S_n = S_\alpha(t_n)$. Then,*

$$P_\theta^n\left(m_n(\Omega, Y^n) \le \tfrac{1}{2}\pi(S_n)e^{-2nt_n}\right) \le 2e^{-n\alpha t_n}.$$

Moreover, if

$$\frac{nt_n}{\log n} > \frac{1}{\alpha}$$

for all large n then, with P_θ probability one,

$$m_n(\Omega, Y^n) \ge \tfrac{1}{2}\pi(S_n)e^{-2nt_n}$$

for all large n.

PROOF. See the Appendix.

4. General results. In this section we give some general results about rates of convergence.

4.1. *Compact sets.* In some cases it is possible to bound the global behavior of the likelihood function outside a shrinking neighborhood of the true parameter. In this section, we give results for such cases. First we give a general result that relies on regularity conditions on the log-likelihood ratios and the prior distribution. Then we give a result based on bracketing entropy. In the following, $L_n(\eta) = n^{-1}\sum_{i=1}^n \ell(\eta, Y_i)$.

THEOREM 2. *Let $d(\cdot,\cdot)$ be a distance (or pseudo-distance) on Ω. Suppose there exists a positive sequence $r_n \to 0$ and a positive constant c_1 such that, if $s_n \ge r_n$ then*

$$P^*\left(\sup_{\eta;d(\theta,\eta)\ge s_n}[L_n(\eta)-L_n(\theta)] > -c_1 s_n^2\right) = o(1). \tag{4.1}$$

Let t_n satisfy

$$\pi(S(t_n)) \succeq e^{-2nt_n}$$

and define

$$\varepsilon_n = \max\left\{r_n, t_n^{1/2}\right\}.$$

If $n\varepsilon_n^2 \to \infty$, then, for sufficiently large $K > 0$,

$$\pi(A_n^c|Y^n) \preceq \exp\left\{-\frac{nK^2c_1\varepsilon_n^2}{2}\right\}$$

except on a set of probability tending to 0, where

$$A_n = \{\eta; d(\theta,\eta) \le K\varepsilon_n\}.$$

PROOF. Note that from (4.1), the fact that $\varepsilon_n \ge r_n$, and the fact that K is large,

$$m_n(A_n^c, Y^n) \le e^{-c_1K^2n\varepsilon_n^2}$$

with probability tending to one. By Lemma 1,

$$m_n(\Omega, Y^n) \ge \tfrac{1}{2}\pi(S(t_n))e^{-2nt_n} \succeq e^{-4nt_n}$$

with probability tending to one. Now, for large $K > 0$, we have $t_n \le \varepsilon_n^2 \le c_1K^2\varepsilon_n^2/8$, so

$$\begin{aligned}\pi(A_n^c|Y^n) &= \frac{m_n(A_n^c, Y^n)}{m_n(\Omega, Y^n)}\\ &\preceq \frac{e^{-c_1K^2n\varepsilon_n^2}}{e^{-4nt_n}}\\ &\le \exp\left\{-\frac{c_1K^2n\varepsilon_n^2}{2}\right\}.\end{aligned}$$

□

To use Theorem 2, we need a way to control the likelihood ratio. We can do this using bracketing entropy.

THEOREM 3. *Suppose there exists a positive constant a and a positive sequence r_n such that*

$$\int_{r_n^2/2^8}^{\sqrt{2}r_n}\{H_B(u/a,\Omega)\}^{1/2}du \le c\sqrt{n}r_n^2 \tag{4.2}$$

where $c = (2/3)^{5/2}/512$. *Then, the rate of convergence is* ε_n *as defined in Theorem* 2 *where* $d(\cdot,\cdot) = h(\cdot,\cdot)$ *is the Hellinger distance.*

PROOF. Using Theorem 1, (4.2) implies that condition (4.1) of Theorem 2 holds. □

REMARK. The entropy $H_B(u/a, \Omega)$ can be replaced by the local entropy $H_B(u/a, \Omega(r_n))$ and the Theorem still holds. Note that if (4.2) holds for r_n and if $1 \geq s_n > r_n$ then (4.2) holds for s_n too. This follows since the entropy function is non-increasing and hence, $\int_{r^2/2^8}^{\sqrt{2}r} \{H_B(u/a, \Omega)\}^{1/2} du/r^2$ is non-increasing for $0 \leq r \leq 1$.

These results say that for any true parameter θ, as data come in, the posterior is concentrated in an ε_n-shrinking neighborhood of θ with probability tending to one. The rate ε_n (depending on θ), is governed by two things: (i) the prior assignment in the neighborhood of θ and (ii) the rate of convergence of the supremum of log-likelihood ratios r_n. When the prior assigns a small probability to the neighborhood of θ, that is, when t_n is large, the posterior distribution may converge at a suboptimal rate $\varepsilon_n \asymp t_n^{1/2}$ since then $\varepsilon_n \gg r_n$ and r_n is the best possible achievable rate by the variants of the maximum likelihood estimator (such as the sieve maximum likelihood estimate). In other words, when the prior assignment to the neighborhood is relatively small, the likelihood fails to wash out the prior as the sample size increases.

4.2. *Non-compact spaces.* When the supremum of log-likelihood ratios can be controlled then the rate of convergence of the posterior can be calculated using Theorems 2 and 3. However, when the parameter space is not compact, the supremum of log-likelihood ratios there may be infinite. Consequently, the prior needs to assign a small probability to the parameter values which may possibly give very large values to the likelihood. Thus we need a tail condition on the prior.

THEOREM 4. *Suppose there exists a positive sequence* r_n, *constants* $a > 0$ *and* $d > 0$ *and a sequence of subsets* $\Omega_1, \Omega_2, \ldots \subset \Omega$ *such that*

$$\int_{r_n^2/2^8}^{\sqrt{2}r_n} \{H_B(u/a, \Omega_n)\}^{1/2} du \leq c\sqrt{n} r_n^2 \tag{4.3}$$

where $c = (2/3)^{5/2}/512$ *and*

$$\pi(\Omega_n^c) \leq e^{-d\varepsilon_n^2 n} \tag{4.4}$$

where

$$\varepsilon_n = \max\left\{r_n, t_n^{1/2}\right\}$$

and t_n *satisfies*

$$\pi(S(dt_n/16)) \succeq e^{-dnt_n/8}.$$

Then, for $K > 0$ *sufficiently large, if* $n\varepsilon_n^2 \to \infty$,

$$\pi(A_n^c|Y^n) \preceq \exp\left\{-\frac{nK^2\varepsilon_n^2}{2}\right\} + \exp\left\{-\frac{nd\varepsilon_n^2}{4}\right\}$$

except on a set of probability tending to 0, *where*

$$A_n = \{\eta; h(\theta, \eta) \le K\varepsilon_n\}.$$

PROOF. See the Appendix.

4.3. *Bayes estimates.* In this section we provide rates for Bayes estimates. Let us focus on the case where we are interested in estimating the density $p(\cdot|\theta)$. A commonly used estimator is the predictive density

$$\hat{p}(\cdot) \equiv \int p(\cdot|\eta)d\pi(\eta|Y^n).$$

We will bound the rate of convergence of this estimator.

THEOREM 5. *Assume that the conditions of Theorem* 2 *or* 3 *hold*. *Then*,

$$h^2(p(\cdot|\theta), \hat{p}(\cdot)) \preceq \varepsilon_n^2 + 2\, e^{-c_1 n\varepsilon_n^2/2}$$

in probability.

PROOF. Let A_n be as defined in Theorem 2. Note that squared Hellinger distance is convex in both its arguments and that Hellinger distance is bounded above by $\sqrt{2}$. Then by Theorem 2 (or 3),

$$\begin{aligned}
h^2(p(\cdot|\theta), \hat{p}(\cdot)) &= h^2\Big(p(\cdot|\theta), \int p(\cdot|\eta)d\pi(\eta|Y^n)\Big) \\
&\le \int h^2(p(\cdot|\theta), p(\cdot|\eta))d\pi(\eta|Y^n) \\
&= \int_{A_n} h^2(p(\cdot|\theta), p(\cdot|\eta))d\pi(\eta|Y^n) \\
&\quad + \int_{A_n^c} h^2(p(\cdot|\theta), p(\cdot|\eta))d\pi(\eta|Y^n) \\
&\le \varepsilon_n^2 \pi(A_n|Y^n) + 2\, \pi(A_n^c|Y^n) \\
&\preceq \varepsilon_n^2 + 2\, e^{-c_1 n\varepsilon_n^2/2}
\end{aligned}$$

in probability, for large n. □

REMARK. This result can be generalized to other loss functions and more general functions of interest $g(\eta)$. The rate then involves conditions on the function g.

4.4. *Discussion.* In parametric models, Bayes' rule can yield estimates with good frequentist properties; see Section 2 of Diaconis and Freedman (1986) and Berger (1986) for discussions. It is natural to believe that Bayes' rule would yield optimal estimates in non-parametric models. As we mentioned earlier, there are published instances where the Bayes estimate can be suboptimal in the frequentist sense, that is, for some values of θ, and loss function L, $\limsup_{n\to\infty} E_\theta L(\theta, T_n^*)/E_\theta L(\theta, T_n^B) = 0$, where T_n^* can be the standard, sieve, or penalized MLEs and T_n^B is the Bayes estimate. This states that for some parameter values, eventually, the variants of MLEs have smaller risks locally. By the Bayes rule, the variants of MLEs cannot have smaller risks than the Bayes estimate in the average sense that is, $EE_\theta L(\theta, T_n^*) \geq EE_\theta L(\theta, T_n^B)$. This implies that for any $n \geq 1$, there exists a set M_n with $\pi(M_n) > 0$ such that $E_\theta L(\theta, T_n^*) \geq E_\theta L(\theta, T_n^B)$ for any $\theta \in M_n$. However, this sub-optimality phenomenon says M_n^c can be a non-negligible part of the parameter space.

It is interesting to note that the Bayes estimate, which is optimal in the average sense, may not be optimal in the local sense. This occurs in a large parameter space in which the prior assigns small probability in the neighborhood of any true parameter value.

5. Example 1: Regression. Let

$$Y_i = \eta(X_i) + e_i, \qquad i = 1, \dots, n$$

where the e_i's are independent $N(0,1)$ random variables, the X_i's are independent $U(0,1)$ random variables, and $\eta(\cdot)$ is a regression function. Let $\psi_1, \psi_2, \dots$ be an orthonormal basis for $L_2[0,1]$ such that, for some $C > 0$, $\sup_{x\in[0,1]} |\psi_j(x)| \leq C$ for all j. For example, take $\psi_1(x) = 1$, $\psi_2(x) = \sqrt{2}sin(2\pi x)$, $\psi_3(x) = \sqrt{2}cos(2\pi x)$, $\psi_4(x) = \sqrt{2}sin(4\pi x)$, $\psi_5(x) = \sqrt{2}cos(4\pi x)$, $\dots$. Let

$$\Omega = \left\{ \eta \in L_2[0,1];\ \eta(x) = \sum_{j=1}^{\infty} \eta_j \psi_j(x),\ \sum_{j=1}^{\infty} \eta_j^2 j^{2p} < \infty \right\}$$

for some fixed integer $p \geq 1$ and let

$$\Omega_M = \left\{ \eta \in \Omega,\ \sum_{j=1}^{\infty} \eta_j^2 j^{2p} < M \right\}$$

for some fixed $M > 0$. Let us emphasize that for this example, the parameter space is Ω_M, not Ω. We will construct a prior for the parameter η by constructing a prior on its Fourier coefficients $\eta_1, \eta_2, \dots$.

Let $d(\theta, \eta) = ||\theta - \eta||$ be the L_2 norm. The optimal rate of convergence for point estimators in Ω_M in this distance is known to be $r_n = n^{-p/(2p+1)}$. We will construct a prior that achieves this rate for all $\theta \in \Omega_M$. In other words, we will find a prior such that the posterior probability of A_n^c tends to 0 in probability, where $A_n = \{\eta; ||\eta - \theta|| \leq K r_n\}$. For this purpose we will first construct a prior on Ω then truncate this prior to the parameter space Ω_M. The prior on

Ω will be a sieve prior as in Zhao (1993, 1998) which means it is built from a sequence of priors that live on finite dimensional spaces. Specifically, let $\lambda_j = e^{\beta}(1-e^{-\beta})e^{-\beta j}$ for some $\beta > 0$, let $d = p + (1/2)$ and define a prior π by $\pi(A) = \sum_{j=1}^{\infty} \lambda_j \pi_j(A)$ where π_j is a prior on the coefficients $\eta_1, \eta_2, \dots$ such that the coefficients are independent and $\eta_i \sim N(0, i^{-2d})$ for $i = 1, \dots j$ and $\eta_i \equiv 0$ for $i > j$. It is easy to see that $\pi(\Omega_M) > 0$. Hence, we can define a new prior γ by $\gamma(A) = \pi(A \cap \Omega_M)/\pi(\Omega_M)$. Thus $\gamma(\Omega_M) = 1$.

The reader may wonder why we use a sieve prior instead of just using a prior directly on the infinite sequence $\eta_1, \eta_2, \dots$ such that $\eta_i \sim N(0, i^{-2d})$. The problem with such a prior is that if we set $d = p + (1/2)$ then the prior puts zero prior mass and zero posterior mass on Ω_M. If we set $d > p + (1/2)$ then the prior does give positive mass to Ω_M but appears to yield sub-optimal rates. The latter point is made precise in a related problem in Zhao (1993, 1999). A referee has pointed out that one can certainly use a prior that puts no mass on the parameter space. However, in our calculations, we prefer to use a prior that does put positive probability on the parameter space.

Now we note some properties of Ω_M and then we apply Theorem 3. First note that there is a constant a_1 such that $\sup_{\theta\in\Omega_M} \sup_{x\in[0,1]} |\theta(x)| \le a_1$. To see this, let $r(p) = \sum_{j=1}^{\infty} j^{-2p}$. Then, for any $\theta \in \Omega_M$, $\sup_{x\in[0,1]} |\theta(x)| \le C \sum_j |\theta_j| = C \sum_j |\theta_j| j^p j^{-p} \le C\sqrt{\sum_j \theta_j^2 j^{2p}}\sqrt{r(p)} \le C\sqrt{Mr(p)} \equiv a_1$. Hence, uniformly for $\theta, \eta \in \Omega_M$, $\sup_{x\in[0,1]}(\theta(x) - \eta(x))^2 \le 4a_1^2$. Direct calculations show that $K(\theta, \eta) = O(||\theta - \eta||^2)$ and $V(\theta, \eta) = O(\int(\theta(x) - \eta(x))^4 dx) = O(||\theta - \eta||^2)$. Thus, there is a constant $a_2 > 0$ such that $\overline{S}(t) \subset S(t)$ where $\overline{S}(t) = \{\eta; ||\eta - \theta||^2 \le a_2 t\}$.

We are interested in showing that the posterior concentrates on shrinking L_2 balls A_n. We next relate A_n to Hellinger neighborhoods. Recall that the marginal density of X is uniform and that the distribution of Y given X is Normal. Let f_θ denote the joint density of Y and X when the parameter is θ. Then,

$$\begin{aligned}
h^2(f_\theta, f_\eta) &= 2\int_0^1 \left[1 - \exp\left\{-\frac{1}{8}(\theta(x) - \eta(x))^2\right\}\right] dx \\
&= 2\int_0^1 \exp\left\{-\frac{1}{8}(\theta(x) - \eta(x))^2\right\}\left[\exp\left\{\frac{1}{8}(\theta(x) - \eta(x))^2\right\} - 1\right] dx \\
&\ge 2e^{-a_1^2/2}\int_0^1 \left[\exp\left\{\frac{1}{8}(\theta(x) - \eta(x))^2\right\} - 1\right] dx \\
&\ge \frac{e^{-a_1^2/2}}{4}\int_0^1 (\theta(x) - \eta(x))^2 dx \\
&= a_3^2\, ||\theta - \eta||^2
\end{aligned}$$

where $a_3^2 = e^{-a_1^2/2}/4$. Hence, $A_n^c \subset B_n^c$ where $B_n = \{\eta; h(\theta, \eta) \le a_3 K r_n\}$. Hence, to show that $\gamma(A_n^c|Z^n) = o_P(1)$, it suffices to show that $\gamma(B_n^c|Z^n) = o_P(1)$ where Z^n denotes $(X_1, Y_1), \dots, (X_n, Y_n)$.

Let $f_\theta(x, y) = f_\theta(y|x)f(x) = \{2\pi\}^{-1/2} \exp\left\{-(1/2)(y - \theta(x))^2\right\}$ denote the joint density of X and Y. Let $\mathscr{P} = \{f_\eta; \eta \in \Omega_M\}$. Note that

$$\left|\sqrt{f_\theta(x, y)} - \sqrt{f_\eta(x, y)}\right| \le M(x, y)||\eta - \theta||_\infty$$

where $||g||_\infty = \sup_x |g(x)|$ and

$$M(x, y) = \{2\pi\}^{-1/4} \left(\frac{|y| + a_1}{2}\right) \exp\left\{-\frac{1}{4}(|y| - a_1)^2\right\} \exp\left\{a_1^2/4\right\}.$$

Moreover, $\int_0^1 \int_{-\infty}^{\infty} M^2(x, y)dydx < \infty$. It follows that the Hellinger bracketing entropy of $\mathscr{P}$, $H_B(u, \mathscr{P})$, is bounded by a constant times the L_∞ metric entropy of Ω_M, denoted by $H(u, \Omega_M, ||\cdot||_\infty)$. According to Theorem 2.4 of van de Geer (2000), $H(u, \Omega_M, ||\cdot||_\infty) \le Au^{-1/p}$ for some $A > 0$; see also Birman and Solomjak (1967). Thus, $H_B(u, \mathscr{P}) \le bu^{-1/p}$ for some $b > 0$. It follows that equation (4) of Theorem 3 holds with r_n proportional to $n^{-p/(2p+1)}$.

To finish verifying the conditions of Theorem 3, it remains to be shown that $\gamma(S_n) \succeq e^{-nr_n^2}$ where $S_n = \{\eta; ||\eta - \theta||^2 \le a_2 r_n^2\}$. Now, $\gamma(S_n) = \pi(S_n \cap \Omega_M)/\pi(\Omega_M) \ge \pi(S_n \cap \Omega_M) = \sum_j \lambda_j \pi_j(S_n \cap \Omega_M) \ge \lambda_k \pi_k(S_n \cap \Omega_M)$ where k is an integer that depends on n. Let $\theta^k(x) = \sum_{j=1}^k \theta_j \psi_j(x)$. Then $||\theta - \theta^k||^2 = \sum_{j=k+1} \theta_j^2 = \sum_{j=k+1} \theta_j^2 j^{2p} j^{-2p} \le k^{-2p} \sum_{j=k+1} \theta_j^2 j^{2p} \le k^{-2p} M$. Hence, if we set $k \equiv k_n$ to be the smallest integer greater than or equal to $(2M/(a_2 r_n^2))^{1/(2p)}$ then we have that $||\theta - \theta^k||^2 \le a_2 r_n^2/2$ uniformly for all $\theta \in \Omega_M$. Note that $k = O(r_n^{-1/p})$. Thus,

$$\begin{aligned}
\pi_k(S_n \cap \Omega_M) &= \pi_k\left(\sum_{i=1}^{\infty}(\eta_i - \theta_i)^2 < a_2 r_n^2,\ \sum_{j=1}^{\infty} \eta_j^2 j^{2p} < M\right) \\
&= \pi_k\left(\sum_{i=1}^{k}(\eta_i - \theta_i)^2 + \sum_{i=k+1}^{\infty} \theta_i^2 < a_2 r_n^2,\ \sum_{j=1}^{k} \eta_j^2 j^{2p} < M\right) \\
&\ge \pi_k\left(\sum_{i=1}^{k}(\eta_i - \theta_i)^2 < a_2 r_n^2/2,\ \sum_{j=1}^{k} \eta_j^2 j^{2p} < M\right).
\end{aligned}$$

Note that, for sufficiently small $a_3 > 0$ (depending on θ) we have that

$$\left\{\eta; \sum_{i=1}^{k}(\eta_i - \theta_i)^2 \le a_3^2 r_n^2\right\} \subset \left\{\eta; \sum_{i=1}^{k}(\eta_i - \theta_i)^2 < a_2 r_n^2/2\ ,\ \sum_{j=1}^{k} \eta_j^2 j^{2p} < M\right\}.$$

To see this, note that if η is in the set on the left hand side then clearly it is in the set

$$\left\{\eta; \sum_{i=1}^{k}(\eta_i - \theta_i)^2 < a_2 r_n^2/2\right\}$$

as long as $a_3^2 < a_2/2$. Next we show that η also satisfies $\sum_{j=1}^{k} \eta_j^2 j^{2p} < M$. To show this note first that $\sum_{j=1}^{\infty} \theta_j^2 j^{2p} = M - \delta$ for some $\delta > 0$. Also note that $k^{2p} r_n^2 = (2M/a_2) \equiv c$, say. So,

$$\begin{aligned}
\sum_{j=1}^{k} \eta_j^2 j^{2p} &\le \sum_{j=1}^{k} (|\eta_j - \theta_j| + |\theta_j|)^2 j^{2p} \\
&= \sum_{j=1}^{k} (\eta_j - \theta_j)^2 j^{2p} + \sum_{j=1}^{k} \theta_j^2 j^{2p} + 2 \sum_{j=1}^{k} |\eta_j - \theta_j|\, |\theta_j| j^{2p} \\
&\le k^{2p} \sum_{j=1}^{k} (\eta_j - \theta_j)^2 + \sum_{j=1}^{\infty} \theta_j^2 j^{2p} + 2 \sqrt{k^{2p} \sum_{j=1}^{k} (\eta_j - \theta_j)^2} \sqrt{\sum_{j=1}^{\infty} \theta_j^2 j^{2p}} \\
&\le c a_3^2 + (M - \delta) + 2\sqrt{M - \delta}\sqrt{c a_3^2}.
\end{aligned}$$

The latter quantity is less than $M - \delta/2$ for a_3 sufficiently small. Hence, $\sum_{j=1}^{k} \eta_j^2 j^{2p} < M$.

So,

$$\begin{aligned}
\pi_k(S_n \cap \Omega_M) &\ge \pi_k \left(\sum_{i=1}^{k} (\eta_i - \theta_i)^2 \le a_3^2 r_n^2 \right) \\
&\ge e^{-a_4 k} F(a_3^2 r_n^2 k^{2d}; k/2) \\
&= e^{-a_4 k} F(a_3^2 c k; k/2)
\end{aligned}$$

where $F(\cdot; k/2)$ is the cdf for a Gamma random variable with shape parameter $k/2$. The last two lines follows from Lemma 4 in the Appendix and the fact that $d = p + (1/2)$.

To lower bound the Gamma cdf we proceed as follows. Note that

$$\begin{aligned}
F(b; \alpha) &= \frac{1}{\Gamma(\alpha)} \int_0^b x^{\alpha-1} e^{-x} dx \\
&\ge \frac{e^{-b}}{\Gamma(\alpha)} \int_0^b x^{\alpha-1} dx \\
&= \frac{e^{-b} b^{\alpha}}{\alpha \Gamma(\alpha)} \\
&\succeq e^{\alpha} e^{-b} b^{\alpha} \alpha^{-\alpha} \alpha^{-1/2}
\end{aligned}$$

where the last line follows from Stirling's approximation. It thus follows that $F(a_4^2 c k; k/2) > e^{-a_5 k}$ for some $a_5 > 0$.

Combining all these facts we conclude that $\gamma(S_n) \succeq e^{-a_6 k}$ for some $a_6 > 0$. From the definition of k it follows that $\gamma(S_n) \succeq e^{-a_7 n r_n^2}$ for some $a_7 >$

0. Applying Theorem 3 we conclude that $\gamma(B_n^c|Z^n) = o_P(1)$ so the rate of convergence is r_n.

REMARK. One can carry out similar calculations with an "un-sieved prior," that is, start with a prior π that makes $\eta_i \sim N(0, i^{-2d})$ for $i = 1, 2, \ldots$ then truncate the prior to Ω_M. In this case, if we want π to have positive probability on Ω_M we must take $d > p + (1/2)$. One can do similar calculations, replacing Lemma 4 with Lemma 5, and these lead to sub-optimal rates. This does not prove that the rates are sub-optimal since our results only give upper bounds on the rates. However, we conjecture that the rates are indeed suboptimal. This conjecture is based on the close similarity with the regression problem and the many normal means problem for which Zhao (1993, 1998) has shown that unsieved priors yield Bayes estimates with sub-optimal rates. We investigate the many normal means problem in the next section.

6. Example 2: Many normal means. Here we consider the problem of estimating infinitely normal means. This problem is known to be intimately related to the nonparametric regression problem. We suspect, based on the results of Brown and Low (1996), that the rates for the posteriors in this problem correspond directly to the rates in the nonparametric regression problem; see also Cox (1993). However, a rigorous proof is not yet available.

Diaconis and Freedman (1997a), Freedman (1999) and Zhao (1993, 1998) have also studied Bayes estimates for this problem. Cox and Diaconis and Freedman considered a Bernstein-Von Mises theorem for this case while Zhao investigated the minimax behavior of Bayes estimates assuming the true parameter is in a Sobolev ball. We treat the same case as Zhao though we study the behavior of the posterior (as opposed to the point estimator). Also, we will study the pointwise behavior of the posterior instead of the minimax behavior.

The model is as follows. The parameter is $\eta = (\eta_1, \eta_2, \ldots)$ and the data are $Y = (Y_1, Y_2, \ldots)$ where $Y_i = \eta_i + n^{-1/2}\varepsilon_i$ and the ε_i's are i.i.d. $N(0, 1)$. We will assume that the true value θ is in $\Omega = \{\eta; \sum_{i=1}^{\infty} \eta_i^2 i^{2p} < \infty\}$. (At the end of this section, we will show that the results still hold when the prior is truncated to the set $\Omega_M = \{\eta; \sum_{i=1}^{\infty} \eta_i^2 i^{2p} < M\}$.)

For the prior, let $\eta_i \sim N(0, i^{-2d})$ independently. Following Zhao, our goal is to use a prior that puts positive mass on the parameter space Ω. This requires that $d > p + (1/2)$. As we remarked in the last section, it is possible to set $d = p + (1/2)$ (which then yields a Bayes estimator that achieves optimal rates over the truncated Sobolev ball) but this has the undesirable property of putting zero prior (and posterior) mass on the parameter space. Unlike the last section, we shall not use a sieve prior. This leads to a posterior whose rate of convergence varies over the parameter space. Our goal is to exhibit a variety of pointwise rates as θ varies over Ω. We will find both lower and upper bounds on the rates. In this example, it is possible to compute the rates directly without recourse to the general theorems in Section 4.

By Bayes rule, $\eta_i|Y_i$ is distributed as $N(T_i, \sigma_i^2)$, where $T_i = c_i Y_i$, $\sigma_i^2 = c_i/n$ and $c_i = n/(n + i^{2d})$. Let $T = (T_1, T_2, \dots)$. Note that we have suppressed the dependence of T on n.

We now calculate the risk:

$$\begin{aligned} R_n(\theta) &= E_\theta \|T - \theta\|^2 \\ &= \sum_{i=1}^{\infty} E_\theta (T_i - \theta_i)^2 \\ &= \sum_{i=1}^{\infty} \left(\frac{i^{-2d}}{i^{-2d} + n^{-1}} \right)^2 n^{-1} + \sum_{i=1}^{\infty} \left(\frac{n^{-1}}{i^{-2d} + n^{-1}} \theta_i \right)^2 \\ &= \sum_{i=1}^{\infty} \left(\frac{i^{-2d}}{i^{-2d} + n^{-1}} \right)^2 n^{-1} + \sum_{i=1}^{\infty} \frac{i^{4d}\theta_i^2}{(i^{2d} + n)^2} \\ &= I + II \end{aligned}$$

where I is the variance and II is the squared bias.

We will show that for every b satisfying

$$\frac{p}{d} < b < 1 - \frac{1}{2d} \tag{6.1}$$

there exists a $\theta \in \Omega$ such that $R_n(\theta) \sim n^{-b}$. [Note that the optimal rate for the truncated space is $n^{-2p/(2p+1)}$ which is in the interval $(p/d, 1-(1/2d))$. Hence, the Bayes rate can be faster or slower than this rate.] To this end, consider $\theta = (\theta_1, \theta_2, \dots)$ of the form $\theta_i \sim i^{-q}$ where $q = bd + (1/2)$. Since $b > p/d$ it follows that $\sum_i \theta_i^2 i^{2p} < \infty$ so that $\theta \in \Omega$.

The variance term can be bounded as follows:

$$I \sim n^{-1} \int_0^\infty \frac{t^{-4d}}{(n^{-1} + t^{-2d})^2} dt = b_2 n^{-(1-1/2d)},$$

where $b_2 = \int_0^\infty \frac{1}{(1+t^{2d})^2} dt$. This follows from Lemma 2 of Diaconis and Freedman (1997a); see Lemma 7 of the Appendix of this paper. Now,

$$\begin{aligned} II &= \sum_{i=1}^{\infty} \frac{i^{4d}\theta_i^2}{(i^{2d} + n)^2} \\ &= \sum_{i=1}^{\infty} \frac{i^{4d-2q}}{(i^{2d} + n)^2} \\ &\sim n^{-(2q-1)/(2d)} = n^{-b}. \end{aligned}$$

The last line also follows from Lemma 2 of Diaconis and Freedman (1997a). We see that

$$R_n \sim n^{-(1-1/2d)} + n^{-b} \sim n^{-b}$$

as claimed.

These results parallel others; Cox (1993) studied the coverage of posteriors in nonparametric regression; Diaconis and Freedman (1997a) and Freedman (1999) studied the asymptotic Bayesian and sampling distribution of $||T-\theta||^2$ assuming only that θ is in L_2; Zhao (1993, 1998) studied the minimax rates of convergence for this problem.

Now consider the posterior. Let $W_n = \pi(A_n^c|Y)$ where $A_n = \{\eta; ||\theta-\eta||^2 \leq K^2\varepsilon_n^2\}$. We will show that if $\varepsilon_n^2 \sim n^{-b}$ then for some $K > 0$, W_n bounded away from 0 in probability. On the other hand, if K is replaced with any increasing sequence K_n then $W_n = o_P(1)$. The result is stated in the next theorem.

THEOREM 6. *Let* $\theta = (\theta_1, \theta_2, \ldots)$ *where* $\theta_i \sim i^{-q}$, $q = bd + (1/2)$ *and* $b \in (p/d, 1-1/(2d))$. *Let* $W_n = \pi(A_n^c|Y)$ *where* $A_n = \{\eta; ||\theta-\eta||^2 \leq K_n^2\varepsilon_n^2\}$ *where* $\varepsilon_n^2 = n^{-b}$. *If* $K_n = K > 0$ *is sufficiently small, then* $P_\theta(W_n \geq 1/2 - \delta) \to 1$ *for every* $\delta > 0$. *On the other hand, if* $K_n \to \infty$ *then* $W_n = o_P(1)$.

To prove this theorem we need some preliminary results. Recall that the data $Y = (Y_1, Y_2, \ldots)$ have a distribution that depends on n. When we need to be explicit about this, we will write the data as $Y^n = (Y_1^n, Y_2^n, \ldots)$. We could similarly subscript θ by n but, in the calculations that follow we will use the same θ for each n so this will not be necessary. Note that $Y_i^n = \theta_i + n^{-1/2}Z_i^n$ where the Z_i^n's are independent, standard Gaussian random variables. Generally, we will write probabilities, expectations and variances as P_θ, E_θ and V_θ rather than P_θ^n, E_θ^n and V_θ^n.

Let $W_n(y^n) = \Pr(||\eta-\theta|| \geq K^2\varepsilon_n^2|Y^n = y^n)$ where $\varepsilon_n^2 = n^{-b}$ and $K > 0$. Here, η denotes a draw from the posterior. The first claim is that $W_n(Y_n)$ does not tend to 0 in probability. To prove this, we need the following results. Throughout, C denotes an arbitrary positive constant but need not represent the same constant in different expressions.

PROPOSITION 1. *Let* $M_n = ||T-\theta||^2$. *Then,*

$$E_\theta(M_n) = \sum_i \theta_i^2(c_i-1)^2 + n^{-1}\sum_i c_i^2 \sim n^{-b}$$

and

$$V_\theta(M_n) = 4n^{-1}\sum_i c_i^2(c_i-1)^2\theta_i^2 + 2n^{-2}\sum_i c_i^4 \sim n^{-b-1}.$$

Moreover,

$$\frac{E_\theta|M_n - E_\theta(M_n)|^4}{V_\theta^2(M_n)} \to 0.$$

Hence, Lyapounov's condition holds and so

$$\frac{M_n - E_\theta(M_n)}{V_\theta^{1/2}(M_n)} \xrightarrow{d} N(0,1).$$

PROOF. Note that $Y_i \stackrel{d}{=} \theta_i + n^{-1/2} Z_i$ where the $Z_i's$ are i.i.d and $Z_i \sim N(0, 1)$. Also, recall that $T_i = c_i Y_i$ where $c_i = n/(n + i^{2d})$. So,

$$\begin{aligned} M_n = ||T - \theta||^2 &= \sum_i (T_i - \theta_i)^2 \\ &= \sum_i (\theta_i(c_i - 1) + c_i n^{-1/2} Z_i)^2 \\ &= \sum_i \theta_i^2 (c_i - 1)^2 + 2n^{-1/2} \sum_i c_i(c_i - 1)\theta_i Z_i + n^{-1} \sum_i c_i^2 Z_i^2. \end{aligned}$$

Taking the mean and variance of the last expression gives the claimed formulae for $E_\theta(M_n)$ and $V_\theta(M_n)$. The fact that $E_\theta(M_n) \sim n^{-b}$ follows from the earlier risk calculation. Next, note that

$$\begin{aligned} V_\theta(T_i - \theta_i)^2 &= V_\theta((T_i - c_i\theta_i) + (c_i\theta_i - \theta_i))^2 \\ &= V_\theta((T_i - c_i\theta_i)^2 + 2(T_i - c_i\theta_i)(c_i\theta_i - \theta_i) + (c_i\theta_i - \theta_i)^2) \\ &= V_\theta((T_i - c_i\theta_i)^2) + 4\theta_i^2(c_i - 1)^2 V_\theta(T_i - c_i\theta_i) \\ &= \frac{2c_i^4}{n^2} + \frac{4\theta_i^2(c_i - 1)^2 c_i^2}{n} \\ &= \frac{2}{n^2} \frac{n^4}{(n + i^{2d})^4} + \frac{4}{n} \frac{i^{-2q} i^{4d} n^2}{(n + i^{2d})^4} \end{aligned}$$

since $T_i \sim N(c_i\theta_i, c_i^2/n)$. Hence,

$$\begin{aligned} V_\theta(M_n) &= V_\theta(||T - \theta||^2) \\ &\sim n^2 \sum_i \frac{1}{(n + i^{2d})^4} + n \sum_i \frac{i^{4d} i^{-2q}}{(n + i^{2d})^4} \\ &\sim \frac{1}{n^{2-1/(2d)}} + \frac{1}{n^{b+1}} \\ &\sim \frac{1}{n^{b+1}}. \end{aligned}$$

Now we verify the Lyapounov condition. First, note that $(a+b)^4 \le 8(a^4+b^4)$. Hence,

$$\begin{aligned} \tfrac{1}{8}|M_n - E_\theta(M_n)|^4 &= \sum_i |n^{-1} c_i^2 (Z_i^2 - 1) + 2n^{-1/2} c_i(c_i - 1)\theta_i Z_i|^4 \\ &\le n^{-4} \sum_i c_i^8 |Z_i^2 - 1|^4 + 16n^{-2} \sum_i c_i^4 |c_i - 1|^4 \theta_i^4 Z_i^4. \end{aligned}$$

Thus,

$$\begin{aligned} E_\theta|M_n - E_\theta(M_n)|^4 &\le C\left\{n^{-4}\sum_i c_i^8 + n^{-2}\sum_i c_i^4(c_i-1)^4 i^{-4q}\right\} \\ &= C\left\{n^{-4}\sum_i \frac{n^8}{(n+i^{2d})^8} + n^{-2}\sum_i \frac{n^4 i^{8d} i^{-4q}}{(n+i^{2d})^8}\right\} \\ &\sim \frac{n^4}{n^{8-1/(2d)}} + \frac{n^2}{n^{8-(1+8d-4q)/(2d)}} \\ &= O\left(\frac{1}{n^{4-(1/(2d))}} + \frac{1}{n^{2+2b+1/(2d)}}\right) \\ &= O\left(\frac{1}{n^{2+2b+1/(2d)}}\right). \end{aligned}$$

We have already shown that $V_\theta(M_n) \sim n^{-b-1}$. So,

$$\frac{E_\theta|M_n - E_\theta(M_n)|^4}{V_\theta^2(M_n)} = O\left(\frac{n^{-2-2b-1/(2d)}}{n^{-2(b+1)}}\right) = O\left(\frac{1}{n^{1/(2d)}}\right) = o(1). \qquad \square$$

PROPOSITION 2. *For $K > 0$ sufficiently small,*

$$P_\theta(M_n \ge K^2\varepsilon_n^2) \to 1.$$

PROOF. Note that

$$P_\theta(K^2\varepsilon_n^2 - M_n \le 0) = P_\theta\left(\frac{M_n - E_\theta(M_n)}{V_\theta^{1/2}(M_n)} \ge \frac{K^2\varepsilon_n^2 - E_\theta(M_n)}{V_\theta^{1/2}(M_n)}\right).$$

The first term in the parentheses tends to a standard Gaussian by Proposition 1. The second term tends to $-\infty$ also by Proposition 1. Hence, the result follows. $\square$

PROPOSITION 3. *With probability tending to* 1, $W_n(Y^n) \ge \widetilde{W}_n(Y^n)$, *where*

$$\widetilde{W}_n(Y^n) = Pr\left(\sum_i \sigma_i^2 Z_i^2 + 2\sum_i \sigma_i Z_i(T_i - \theta_i) > 0\right).$$

In the above display, $Z = (Z_1, Z_2, \ldots)$ are i.i.d. standard Gaussians, independent of Y^n.

PROOF. Let C_n be the set of y^n for which $K^2\varepsilon_n^2 - M_n < 0$. According to the previous proposition, C_n has probability tending to 1. Restrict attention to this set. With respect to the posterior, we have that $\eta_i = \sigma_i Z_i + T_i$ where $Z =$

$(Z_1, Z_2, \ldots)$ denotes a sequence of i.i.d. standard normal random variables. Thus, on C_n,

$$\begin{aligned}
&\Pr(\|\eta - \theta\|^2 \geq K^2 \varepsilon_n^2 | Y^n = y^n) \\
&\quad = \Pr\left(\sum_{i=1}^{\infty} (\eta_i - \theta_i)^2 \geq K^2 \varepsilon_n^2 | Y^n = y^n\right) \\
&\quad = \Pr\left(\sum_{i=1}^{\infty} (\sigma_i Z_i + T_i - \theta_i)^2 \geq K^2 \varepsilon_n^2\right) \\
&\quad = \Pr\left(\sum_{i=1}^{\infty} \sigma_i^2 Z_i^2 + 2 \sum_{i=1}^{\infty} \sigma_i Z_i (T_i - \theta_i) + \sum_i (T_i - \theta_i)^2 \geq K^2 \varepsilon_n^2\right) \\
&\quad = \Pr\left(\sum_{i=1}^{\infty} \sigma_i^2 Z_i^2 + 2 \sum_{i=1}^{\infty} \sigma_i Z_i (T_i - \theta_i) \geq K^2 \varepsilon_n^2 - M_n\right) \\
&\quad \geq \Pr\left(\sum_{i=1}^{\infty} \sigma_i^2 Z_i^2 + 2 \sum_{i=1}^{\infty} \sigma_i Z_i (T_i - \theta_i) \geq 0\right) \\
&\quad = \widetilde{W}_n(Y^n). \qquad \square
\end{aligned}$$

The next proposition is needed to establish a central limit theorem for $\tilde{W}_n(Y_n)$.

PROPOSITION 4. *Let a be such that $4 - (1/d) < a < 6 - 1/(2d)$. Let*

$$D_n = \left\{ Z^n;\ n^2 \sum_{i=1}^{\infty} \frac{(Z_i^n)^4}{(n + i^{2d})^8} < n^{-a} \right\}.$$

Then, $\Pr(D_n) \to 1$.

PROOF. Let $c = E((Z_i^n)^4) = 3$. Then,

$$\begin{aligned}
\Pr\left(n^2 \sum_{i=1}^{\infty} \frac{(Z_i^n)^4}{(n + i^{2d})^8} > n^{-a}\right) &\leq n^{2+a} \sum_i \frac{c}{(n + i^{2d})^8} \\
&\sim n^{2+a} \frac{1}{n^{8-1/(2d)}} = o(1). \qquad \square
\end{aligned}$$

Let $\tilde{D}_n = \{y^n;\ y^n = \theta + n^{1/2} z^n,\ z^n \in D_n\}$ be the set of y^n's that correspond to $z^n \in D_n$.

PROPOSITION 5. *Let $Z = (Z_1, Z_2, \ldots)$ be i.i.d. standard Gaussians and let Q denote the distribution of Z. Let* $A_n \equiv A_n(y_n) = \sum_i \sigma_i^2 Z_i^2 + 2 \sum_i \sigma_i Z_i (T_i -$

θ_i). Then $E_Q(A_n) = \sum_i \sigma_i^2$ and $V_Q(A_n) = 2\sum_i \sigma_i^4 + 4\sum_i \sigma_i^2(T_i - \theta_i)^2$. Furthermore, for $y_n \in \tilde{D}_n$, we have that

$$\frac{A_n - E_Q(A_n)}{V_Q^{1/2}(A_n)} \xrightarrow{d} N(0,1).$$

PROOF. The mean and variance are immediate. Now we verify the Lyapounov condition. We have

$$\begin{aligned}
\frac{E_Q|A_n - E_Q(A_n)|^4}{V_Q^2(A_n)} &= \frac{E|\sum_i \sigma_i^2(Z_i^2 - 1) + 2\sum_i \sigma_i Z_i(T_i - \theta_i)|^4}{[2\sum_i \sigma_i^4 + 4\sum_i \sigma_i^2(T_i - \theta_i)^2]^2} \\
&\le C\frac{\sum_i \sigma_i^8 + \sum_i \sigma_i^4(T_i - \theta_i)^4}{[2\sum_i \sigma_i^4 + 4\sum_i \sigma_i^2(T_i - \theta_i)^2]^2} \\
&\le C\frac{\sum_i \sigma_i^8 + \sum_i \sigma_i^4(T_i - \theta_i)^4}{[\sum_i \sigma_i^4]^2} \\
&= C\frac{\sum_i \sigma_i^8}{[\sum_i \sigma_i^4]^2} + C\frac{\sum_i \sigma_i^4(T_i - \theta_i)^4}{[\sum_i \sigma_i^4]^2}.
\end{aligned}$$

Now, $\sum_i \sigma_i^8 \sim n^{-4+1/(2d)}$ and $\sum_i \sigma_i^4 \sim n^{-2+1/(2d)}$, hence,

$$\frac{\sum_i \sigma_i^8}{[\sum_i \sigma_i^4]^2} \sim \frac{1}{n^{1/(2d)}} = o(1).$$

Now, for $y_n \in \tilde{D}_n$ we have

$$\begin{aligned}
\sum_i \sigma_i^4(T_i - \theta_i)^4 &= C\sum_i \sigma_i^4[\theta_i(c_i - 1) + c_i n^{-1/2} Z_i^n]^4 \\
&\le C\sum_i \sigma_i^4[\theta_i^4(c_i - 1)^4 + c_i^4 n^{-2}(Z_i^n)^4] \\
&= C\sum_i \frac{i^{-4q}(c_i - 1)^4}{(n + i^{2d})^4} + C\frac{1}{n^2}\sum_i \frac{c_i^4(Z_i^n)^4}{(n + i^{2d})^4} \\
&= C\sum_i \frac{i^{-4q} i^{8d}}{(n + i^{2d})^8} + Cn^2\sum_i \frac{(Z_i^n)^4}{(n + i^{2d})^8} \\
&\sim \frac{1}{n^{8-(8d-4q+1)/(2d)}} + n^2\sum_i \frac{(Z_i^n)^4}{(n + i^{2d})^8} \\
&= O\left(\frac{1}{n^{8-(8d-4q+1)/(2d)}} + n^{-a}\right).
\end{aligned}$$

So,

$$\frac{\sum_i \sigma_i^4(T_i - \theta_i)^4}{[\sum_i \sigma_i^4]^2} = O\left(\frac{n^{-8+(8d-4q+1)/(2d)} + n^{-a}}{n^{-4+(1/d)}}\right) = o(1)$$

since $a > 4 - (1/d)$. $\square$

PROOF OF THEOREM 6. Finally, we are ready to prove the theorem. We have shown that $W_n(y^n) \geq \widetilde{W}_n(y^n)$ on a set of y^n with probability tending to one. Now, $\widetilde{W}_n(y^n) = \Pr(\sum_i \sigma_i^2 Z_i^2 + 2\sum_i \sigma_i Z_i(T_i - \theta_i) \geq 0) = \Pr(A_n \geq 0)$. Note that $E_Q(A_n) \geq 0$. Thus, for $y^n \in \tilde{D}_n$,

$$\begin{aligned}\widetilde{W}_n(y^n) &= \Pr(A_n \geq 0)\\ &= \Pr\left(\frac{A_n - E_Q(A_n)}{V_Q^{1/2}(A_n)} \geq \frac{-E_Q(A_n)}{V_Q^{1/2}(A_n)}\right)\\ &\geq \Pr\left(\frac{A_n - E_Q(A_n)}{V_Q^{1/2}(A_n)} \geq 0\right)\\ &\to \frac{1}{2}\end{aligned}$$

from Proposition 5. Hence, for any $\delta > 0$, $\Pr(W_n(Y^n) \geq 1/2 - \delta) \to 1$. This proves the first claim.

To prove the second, let K_n be any sequence of positive numbers tending to ∞. For the remainder of the proof, let $W_n(y^n) = \Pr(||\eta - \theta||^2 \geq K_n^2 \varepsilon_n^2 | Y^n = y^n)$. Then,

$$\begin{aligned}W_n(y^n) &= \Pr(||\eta - \theta||^2 \geq K_n^2 \varepsilon_n^2 | Y^n = y^n)\\ &\leq \frac{1}{K_n^2 \varepsilon_n^2} E(||\eta - \theta||^2 | Y^n = y^n)\\ &= \frac{\sum_i \sigma_i^2 + M_n}{K_n^2 \varepsilon_n^2}\\ &= \frac{\sum_i \sigma_i^2}{K_n^2 \varepsilon_n^2} + \frac{M_n}{K_n^2 \varepsilon_n^2}.\end{aligned}$$

Consider the first term. We see that

$$\begin{aligned}\frac{\sum_i \sigma_i^2}{K_n^2 \varepsilon_n^2} &= \frac{1}{K_n^2 \varepsilon_n^2} \sum_i \frac{1}{n + i^{2d}}\\ &\sim \frac{1}{K_n^2 \varepsilon_n^2} \frac{1}{n^{1-1/(2d)}} = o(1)\end{aligned}$$

since $\varepsilon_n^2 = n^{-b}$ and $b < 1 - (1/(2d))$. Hence,

$$\begin{aligned}E_\theta(W_n(Y^n)) &= o(1) + \frac{E_\theta(M_n)}{K_n^2 \varepsilon_n^2}\\ &= o(1) + \frac{Cn^b}{K_n^2 n^b} = o(1).\end{aligned}$$

This completes the proof. □

We now show that Theorem 6 continues to hold if we truncate the prior to the set $\Omega_M = \{\eta; \sum_i \eta_i^2 i^{2p} < M\}$. Define $\gamma(A) = \pi(A \cap \Omega_M)/\pi(\Omega_M)$. Note that $d > p + (1/2)$ implies that $\pi(\Omega_M) > 0$ so this prior is well defined.

THEOREM 7. *The results of Theorem* 6 *also hold for the prior* γ.

PROOF. From Bayes' theorem, we see that for any A,

$$\gamma(A|Y^n) = \frac{\pi(A \cap \Omega_M|Y^n)}{\pi(\Omega_M|Y^n)}.$$

Below, we will show that for every $\theta \in \Omega_M$ we have that $\pi(\Omega_M|Y^n) = 1 + o_P(1)$. Thus,

$$\gamma(A|Y^n) = \frac{\pi(A \cap \Omega_M|Y^n)}{1 + o_P(1)} = \pi(A \cap \Omega_M|Y^n) + o_P(1).$$

CASE 1. Let $A_n = \{\eta; ||\theta - \eta||^2 \le K_n^2 \varepsilon_n^2\}$ where $K_n \to \infty$ and ε_n is as defined in Theorem 6. Then, $\gamma(A_n^c|Y^n) = \pi(A_n^c \cap \Omega_M|Y^n) + o_P(1) \le \pi(A_n^c|Y^n) + o_P(1) = o_P(1)$ since $\pi(A_n^c|Y^n) = o_P(1)$ from Theorem 6.

CASE 2. Let $A_n = \{\eta; ||\theta - \eta||^2 \le K\varepsilon_n^2\}$ for $K > 0$ sufficiently small. We showed in Theorem 6 that $\pi(A_n^c|Y^n) \ge 1/2$ on a set of probability tending to 1. On this set we have $\gamma(A_n|Y^n) = \pi(A_n \cap \Omega_M|Y^n) + o_P(1) \le \pi(A_n|Y^n) + o_P(1) = 1 - \pi(A_n^c|Y^n) + o_P(1) \le 1 - (1/2) + o_P(1) = 1/2 + o_P(1)$. Hence, with probability tending to 1, we have that $\gamma(A_n^c|Y^n) \ge 1/2$.

To complete the proof, we now show that $\pi(\Omega_M|Y^n)$ converges to 1 in probability. For any $c > 0$ define $H(y^n, c) = \Pr(\sum_{i=1}^{\infty} (\eta_i - \theta_i)^2 i^{2p} > c | Y^n = y^n)$. Let $Z_1, Z_2, \dots$ denote iid $N(0, 1)$ random variables. Then,

$$\begin{aligned} H(y^n, c) &\le \frac{1}{c} \sum_i E(i^{2p}(\eta_i - \theta_i)^2 | Y^n = y^n) \\ &= \frac{1}{c} \sum_i E(i^{2p}(T_i + \sigma_i Z_i - \theta_i)^2) \\ &\le \frac{2}{c} \sum_i \sigma_i^2 i^{2p} + \frac{2}{c} \sum_i i^{2p}(T_i - \theta_i)^2. \end{aligned}$$

Now,

$$\sum_i \sigma_i^2 i^{2p} = \sum_i \frac{i^{2p}}{n + i^{2d}} \sim \frac{1}{n^{(2d-2p-1)/(2d)}} = o(1)$$

since $d > p + (1/2)$. Hence,

$$\begin{aligned} E_\theta(H(Y^n, c)) &\le \frac{2}{c} \sum_i i^{2p} E_\theta(T_i - \theta_i)^2 + o(1) \\ &= \frac{2}{c} \sum_i i^{2p} E_\theta(\theta_i(c_i - 1) + c_i n^{-1/2} Z_i)^2 + o(1) \end{aligned}$$

$$\leq \frac{4}{c}\sum_i i^{2p}\theta_i^2(c_i-1)^2 + \frac{4}{n}\sum_i i^{2p}c_i^2 + o(1)$$

$$= \frac{4}{c}\sum_i i^{2p}\theta_i^2\left(\frac{i^{2d}}{n+i^{2d}}\right)^2 + \frac{4}{n}\sum_i \frac{i^{2p}n^2}{(n+i^{2d})^2} + o(1)$$

$$\sim o(1) + \frac{1}{n^{(2d-2p-1)/(2d)}} + o(1) = o(1)$$

where the first term goes to zero by dominated convergence and the second term goes to zero since $d > p + (1/2)$. Hence, for every $c > 0$, $E_\theta(H(Y^n, c)) = o_P(1)$ which implies that $\Pr(\sum_{i=1}^\infty(\eta_i - \theta_i)^2 i^{2p} \leq c|Y^n = y^n)$ tends to 1 in probability. Finally, note that for sufficiently small c (depending on θ) we have that

$$\left\{\sum_i(\eta_i-\theta_i)^2 i^{2p} < c\right\} \subset \left\{\sum_i \eta_i^2 i^{2p} < M\right\}.$$

This follows since $\sum_i \theta_i^2 i^{2p} = M - \delta$ for some $\delta > 0$. Hence, if $\sum_i(\eta_i - \theta_i)^2 i^{2p} < c$ then,

$$\begin{aligned}
\sum_i \eta_i^2 i^{2p} &= \sum_i |(\eta_i-\theta_i)+\theta_i|^2 i^{2p}\\
&\leq \sum_i (|\eta_i-\theta_i| + |\theta_i|)^2 i^{2p}\\
&= \sum_i(\eta_i-\theta_i)^2 i^{2p} + \sum_i \theta_i^2 i^{2p} + 2\sum_i|\eta_i-\theta_i||\theta_i| i^{2p}\\
&\leq \sum_i(\eta_i-\theta_i)^2 i^{2p} + \sum_i \theta_i^2 i^{2p} + 2\sqrt{\sum_i(\eta_i-\theta_i)^2 i^{2p}}\sqrt{\sum_i \theta_i^2 i^{2p}}\\
&< c + (M-\delta) + 2\sqrt{c}\sqrt{M-\delta}\\
&< M - \frac{\delta}{2}
\end{aligned}$$

for c sufficiently small. Thus, $\Pr(\sum_{i=1}^\infty \eta_i^2 i^{2p} < M|Y^n = y^n)$ tends to 1 in probability. □

7. Discussion. In regular, finite dimensional models, Bayes estimators share the same asymptotic behavior as maximum likelihood estimators if the prior is smooth. Consequently, when the sample size is moderately large, the choice of prior is not crucial. In particular, the posterior will be consistent under very weak conditions.

In infinite dimensional models, the situation is quite different. As Freedman (1963) and Diaconis and Freedman (1986) showed, consistency is not even guaranteed. Still, it is possible to achieve consistent posteriors as long as the prior obeys certain regularity conditions; see Barron (1988) and Barron, Schervish and Wasserman (1998) for example. From the results in those papers, it is clear that there are rich sets of priors that lead to consistent posteriors.

In this paper we have studied the question of rates of convergence. We have seen that the choice of prior is much more delicate. Although it is too early to draw general conclusions, it appears that the choice of prior in an infinite dimensional problem is more difficult if one wants to achieve good rates.

APPENDIX: TECHNICAL PROOFS

PROOF OF LEMMA 2. For any $\eta \in S_n$ we have

$$\begin{aligned}
P^n_\theta(K_n(\theta, \eta) \geq 2t_n) &= P^n_\theta\left(\prod_{i=1}^n \frac{p(Y_i|\theta)}{p(Y_i|\eta)} \geq e^{2nt_n}\right) \\
&= P^n_\theta\left(\prod_{i=1}^n \left(\frac{p(Y_i|\theta)}{p(Y_i|\eta)}\right)^\alpha \geq e^{2n\alpha t_n}\right) \\
&\leq e^{-2n\alpha t_n}\left[E_\theta\left(\frac{p(Y|\theta)}{p(Y|\eta)}\right)^\alpha\right]^n \\
&= e^{-2n\alpha t_n}(1+\alpha\rho_\alpha(\theta, \eta))^n \\
&= e^{-2n\alpha t_n}e^{n\log(1+\alpha\rho_\alpha(\theta, \eta))} \\
&\leq e^{-2n\alpha t_n}e^{n\alpha\rho_\alpha(\theta, \eta)} \\
&\leq e^{-2n\alpha t_n}e^{n\alpha t_n} \\
&= e^{-n\alpha t_n}.
\end{aligned}$$

Define

$$W_n = \{(\eta, y^n);\ K_n(\theta, \eta) \geq 2t_n\}.$$

Let $W_n(y^n) = \{\eta;\ (\eta, y^n) \in W_n\}$ and $W_n(\eta) = \{y^n;\ (\eta, y^n) \in W_n\}$. Then,

$$\begin{aligned}
m_n(\Omega, Y^n) &= \int e^{-nK_n(\theta,\eta)}d\pi(\eta) \\
&\geq \int_{S_n\cap W_n^c(Y^n)} e^{-nK_n(\theta,\eta)}d\pi(\eta) \\
&\geq [\pi(S_n) - \pi(S_n \cap W_n(Y^n))]e^{-2nt_n}.
\end{aligned}$$

By Fubini's theorem,

$$\begin{aligned}
E^n_\theta \pi(S_n \cap W_n(Y^n)) &= \int\int I(S_n \cap W_n(y^n))d\pi(\eta)dP^n_\theta(y^n) \\
&= \int\int I(S_n)I(W_n)d\pi(\eta)dP^n_\theta(y^n) \\
&= \int I(S_n)P_\theta(W_n(\eta))d\pi(\eta) \\
&\leq \pi(S_n)e^{-n\alpha t_n}.
\end{aligned}$$

So,

$$P_\theta^n\left(m_n(\Omega, Y^n) \le \frac{1}{2}\pi(S_n)e^{-2nt_n} \right)$$
$$\le P_\theta^n\left([\pi(S_n) - \pi(S_n \cap W_n(Y^n))]\, e^{-2nt_n} \le \frac{1}{2}\pi(S_n)e^{-2nt_n} \right)$$
$$= P_\theta^n\left(\pi(S_n \cap W_n(Y^n)) \ge \frac{1}{2}\pi(S_n)\right)$$
$$\le \frac{2}{\pi(S_n)} E_\theta^n(S_n \cap W_n(Y^n)) \le 2e^{-n\alpha t_n}.$$

The last claim follows from the first Borel-Cantelli lemma. □

PROOF OF THEOREM 4. Write

$$\pi(A_n^c|Y^n) = \frac{m_n(A_n^c \cap \Omega_n)}{m_n(\Omega, Y^n)} + \frac{m_n(A_n^c \cap \Omega_n^c)}{m_n(\Omega, Y^n)} \equiv a_1 + a_2.$$

Now,

$$a_1 \preceq \exp\left\{-\frac{nK^2c_1\varepsilon_n^2}{2}\right\}$$

in probability, by Theorem 3. Apply Markov's inequality and Fubini's theorem to the numerator of a_2 to conclude that

$$P_\theta^n(m_n(A_n^c \cap \Omega_n^c, Y^n) > e^{-dn\varepsilon_n^2/2}) \le e^{dn\varepsilon_n^2/2} \int_{\mathscr{Y}^n} \int_{A_n^c \cap \Omega_n^c} e^{-nK_n(\theta,\eta)} d\pi(\eta) dP_\theta(y^n)$$
$$= e^{dn\varepsilon_n^2/2} \pi(A_n^c \cap \Omega_n^c) \le e^{-dn\varepsilon_n^2/2}.$$

Thus, $m_n(A_n^c \cap \Omega_n^c, Y^n) \le e^{-dn\varepsilon_n^2/2}$ in probability. Apply Lemma 1 to see that

$$m_n(\Omega, Y^n) \succeq e^{-2Bnt_n}\pi(S(Bt_n)) \succeq e^{-4Bnt_n}$$

in probability where $B = d/16$. Since $4t_nB \le 4\varepsilon_n^2 B = d\varepsilon_n^2/4$,

$$a_2 \preceq \frac{e^{-dn\varepsilon_n^2/2}}{m_n(\Omega, Y^n)} \le e^{-dn\varepsilon_n^2/4}$$

in probability. □

The following lemma may be proved by induction.

LEMMA 3. *For* $r \ge 0$ *and any integrable function* f,

$$\int_{\sum_{i=1}^n x_i^2 \le r^2} f\left(\sqrt{\sum_{i=1}^n x_i^2}\right) dx_1 \cdots dx_n = r^n \frac{\pi^{n/2}}{\Gamma(n/2)} \int_0^1 u^{n/2-1} f(r\sqrt{u}) du.$$

LEMMA 4. *Let $Z_1, \dots, Z_k$ be independent standard Normal random variables, let $a_i = i^{-d}$, $d > 0$ and let $\theta_1, \dots, \theta_k$ be real numbers. If $\delta > 0$ then*

$$\Pr\left(\sum_{i=1}^{k}(a_i Z_i - \theta_i)^2 \le \delta^2\right) \ge ce^{-k}2^{-k/2}\exp\left\{-\sum_{i=1}^{k} i^{2d}\theta_i^2\right\}F(k^{2d}\delta^2; k/2)$$

where $F(\cdot; b)$ is the distribution function for a Gamma random variable with b scale parameter and $c > 0$ is a constant.

PROOF. In what follows, we will apply the previous Lemma with the function with $f(x) = e^{-j^{2d}x^2}$. Let $W_i = a_i Z_i - \theta_i \sim N(-\theta_i, i^{-2d})$ and let $A_k = \{(w_1, \dots, w_k); \sum_{i=1}^{k} w_i^2 \le \delta^2\}$. Also, define

$$\Delta = \exp\left\{-\sum_{i=1}^{k} i^{2d}\theta_i^2\right\}.$$

Then, by the previous lemma and the fact that $x! \ge x^x e^{-x}$,

$$\begin{aligned}
&\Pr\left(\sum_{i=1}^{k} W_i^2 \le \delta^2\right)\\
&\quad = \left(\frac{1}{2\pi}\right)^{k/2}(k!)^d \int_{A_k}\exp\left\{-\frac{1}{2}\sum_{i=1}^{k} i^{2d}(w_i + \theta_i)^2\right\}dw_1\cdots dw_k\\
&\quad \ge \left(\frac{1}{2\pi}\right)^{k/2}(k!)^d \int_{A_k}\exp\left\{-\sum_{i=1}^{k} i^{2d}(w_i^2 + \theta_i^2)\right\}dw_1\cdots dw_k\\
&\quad = \Delta\left(\frac{1}{2\pi}\right)^{k/2}(k!)^d \int_{A_k}\exp\left\{-\sum_{i=1}^{k} i^{2d}w_i^2\right\}dw_1\cdots dw_k\\
&\quad \ge \Delta\left(\frac{1}{2\pi}\right)^{k/2}(k!)^d \int_{A_k}\exp\left\{-k^{2d}\sum_{i=1}^{k} w_i^2\right\}dw_1\cdots dw_k\\
&\quad = \Delta\left(\frac{1}{2\pi}\right)^{k/2}(k!)^d \frac{\delta^k\pi^{k/2}}{\Gamma((k/2))}\int_0^1 u^{(k/2)-1}\exp\{-k^{2d}\delta^2 u\}du\\
&\quad \ge \Delta\left(\frac{1}{2\pi}\right)^{k/2}k^{kd}\exp\{-kd\}\frac{\delta^k\pi^{k/2}}{\Gamma((k/2))}\int_0^1 u^{(k/2)-1}\exp\{-k^{2d}\delta^2 u\}du\\
&\quad \ge \Delta\left(\frac{1}{2\pi}\right)^{k/2}k^{kd}\exp\{-kd\}\frac{\delta^k\pi^{k/2}}{\Gamma((k/2))}k^{-dk}\delta^{-k}\int_0^{k^{2d}\delta^2} u^{(k/2)-1}\exp\{-u\}du\\
&\quad \ge \Delta\exp\{-dk\}2^{-k/2}F(k^{2d}\delta^2; k/2). \qquad \square
\end{aligned}$$

LEMMA 5. *Let $Z_1, Z_2, \dots,$ be independent standard Normal random variables, let $a_i = i^{-d}$, $d \ge p$, and let $\theta_1, \theta_2, \dots$ be real numbers such that*

$c = \sum_{i=1}^{\infty} \theta_i^2 i^{2p} < \infty$. *Let* $\delta > 0$ *and let* N *be the smallest integer such that*

$$\sum_{j=N+1}^{\infty} (i^{-2d} + \theta_i^2) \le \delta^2/4. \tag{A.1}$$

Then

$$\Pr\left(\sum_{i=1}^{\infty} (a_i Z_i - \theta_i)^2 \le \delta^2\right) \ge c' e^{-2cN^{2(d-p)}}$$

where $c' > 0$.

PROOF. Let $W_i = a_i Z_i - \theta_i \sim N(-\theta_i, i^{-2d})$. First note that

$$\sum_{j=1}^{N} \theta_i^2 i^{2d} = \sum_{j=1}^{N} \theta_i^2 i^{2p} i^{2(d-p)} \le N^{2(d-p)} \sum_{j=1}^{N} \theta_i^2 i^{2p} \le cN^{2(d-p)}.$$

Then, from Markov's inequality and Lemma 4,

$$\begin{aligned}
\Pr\left(\sum_{i=1}^{\infty} W_i^2 \le \delta^2\right) &\ge \Pr\left(\sum_{i=1}^{N} W_i^2 \le \delta^2/2\right) \Pr\left(\sum_{i=N+1}^{\infty} W_i^2 \le \delta^2/2\right) \\
&= \Pr\left(\sum_{i=1}^{N} W_i^2 \le \delta^2/2\right)\left[1 - \Pr\left(\sum_{i=N+1}^{\infty} W_i^2 > \delta^2/2\right)\right] \\
&\ge \Pr\left(\sum_{i=1}^{N} W_i^2 \le \delta^2/2\right)\left[1 - \frac{2}{\delta^2}\sum_{N+1}^{\infty} E\left(\sum_{i=N+1}^{\infty} W_i^2\right)\right] \\
&\ge \frac{1}{2}\Pr\left(\sum_{i=1}^{N} W_i^2 \le \delta^2/2\right) \\
&\ge \frac{1}{2} e^{-dN} 2^{-N/2} e^{-cN^{2(d-p)}} F(N^{2d}\delta^2; N/2) \\
&\ge \frac{1}{2} e^{-2cN^{2(d-p)}} F(N^{2d}\delta^2; N/2).
\end{aligned}$$

Now, (A.1) implies that $N^{2d-1}\delta^2 \ge 4 \ge 1/2$ so that $N^{2d}\delta^2 \ge N/2$. Thus, for large N, $F(N^{2d}\delta^2; N/2) \ge 1/4$. □

When the θ_i's =0, the last lemma gives a tight bound. This is evident from the next lemma.

LEMMA 6 [Dembo, Mayer-Wolf and Zeitouni (1995)]. *Let* $Z = \sum_{i=1}^{\infty} a_i^2 Z_i^2$, *where* $\{Z_i\}$ *are independently distributed according to density* $s(x) = C_r |x|^r \exp(-x^2/2)$ *for* $r > -1$, *and* $a_i = i^{-b}$, *where* C_r *is a normalizing constant*

$$\lim_{t \to 0} P(Z \le t^2)(\exp(I)\psi)^{-1} = (2\pi)^{-1},$$

where

$$I = c_1' t^{-2/(2b-1)}(1+o(1)),$$
$$\psi^2 = c_2' t^{-2/(2b-1)}(1+o(1)),$$
$$c_1' = (b-1/2)\left[\frac{\pi/2b}{\sin(\pi/2b)}\right]^{2b/(2b-1)}$$

and

$$c_2' = \left[\frac{\pi/2b}{\sin(\pi/2b)}\right]^{1/(2b-1)}\left[\frac{2b-1}{(2b)^2\sin((2b-1)\pi/2b)}\right].$$

The following result is Lemma 2 from Diaconis and Freedman (1997a) and is very useful in the normal means problem.

LEMMA 7 [Diaconis and Freedman (1997a)]. *Let* b, c *and* α *be such that* $1 < b < \infty$, $0 \le c < \infty$ *and* $\alpha b > c+1$. *Suppose that* $\gamma_n \to \infty$. *Then,*

$$\lim_{n\to\infty} g_n \sum_{i=1}^{\infty} \frac{i^c}{(\gamma_n + i^\alpha)^b} = \int_0^\infty \frac{u^c}{(1+u^\alpha)^b}\,du$$

where $g_n = \gamma_n^{b-(1+c)/\alpha}$.

Acknowledgments. The authors thank Larry Brown and Linda Zhao for helpful discussions and the referees for suggestions that improved the paper.

REFERENCES

BARRON, A. (1988). The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions. Technical Report 7, Dept. Statistics, Univ. Illinois, Champaign.

BARRON, A., SCHERVISH, M. and WASSERMAN, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.* **27** 536–561.

BARRON, A. and YANG, Y. (1998). An asymptotic property of model selection criteria. *IEEE Trans. Inform. Theory* **44** 95–116.

BARRON, A., BIRGÉ, L. and MASSART, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413.

BERGER, J. O. (1986). Discussion on the consistency of Bayes estimates (with discussion). *Ann. Statist.* **14** 1–67.

BIRMAN, M. S. and SOLOMJAK, M. Z. (1967). Piecewise-polynomial approximation of functions of the classes W_p. *Mat. Sbornik* **73** 295–317.

BROWN, L. and LOW, M. (1996). Asymptotic equivalence of non-parametric regression and white noise model. *Ann. Statist.* **24** 2384–2398.

COX, D. D. (1993). An analysis of Bayesian inference for non-parametric regression. *Ann. Statist.* **21** 903–924.

DEMBO, D., MAYER-WOLF, E. and ZEITOUNI, O. (1995). Exact behavior of Gaussian seminorms. *Statist. Probab. Lett.* **23** 275–280.

DEY, D., MÜLLER, P. and SINHA, D. (1998). *Practical Nonparametric and Semiparametric Bayesian Statistics.* Springer, New York.

DIACONIS, P. and FREEDMAN, D. (1986). On the consistency of Bayes estimates (with discussion). *Ann. Statist.* **14** 1–67.

DIACONIS, P. and FREEDMAN, D. (1993). Non-parametric binary regression: A Bayesian approach. *Ann. Statist.* **21** 2108–2137.

DIACONIS, P. and FREEDMAN, D. (1997a). On the Bernstein-von Mises Theorem with infinite dimensional parameters. Unpublished manuscript.

DIACONIS, P. and FREEDMAN, D. (1997b). Consistency of Bayes estimates for nonparametric regression: A review. In *Festschrift for Lucien Le Cam* (D. Pollard, E. Torgersen and G. Yang, eds.) Springer, New York.

DIACONIS, P. and FREEDMAN, D. (1998). Consistency of Bayes estimates for nonparametric regression: normal theory. *Bernoulli* **4** 411–444.

DOOB, J. L. (1948). Application of the theory of martingales. *Coll. Int. du C. N. R. S. Paris* 22-28.

FREEDMAN, D. (1963). On the asymptotic behavior of Bayes estimates in the discrete case I. *Ann. Math. Statist.* **34** 1386–1403.

FREEDMAN, D. (1999). On the Bernstein-Von Mises theorem with infinite dimensional parameters. *Ann. Statist.* **27**.

GHOSAL, S., GHOSH, J. K. and RAMAMOORTHI, R. V. (1997). Non-informative priors via sieves and packing numbers. In *Advances in Statistical Decision Theory and Applications* (S. Panchapakesan and N. Balakrishnan, eds.) 119–132. Birkhäuser, Boston.

GHOSAL, S., GHOSH, J. K. and RAMAMOORTHI, R.V. (1999a). Consistency issues in Bayesian nonparametrics. In *Asymptotics, Nonparametrics and Time Series: A Tribute to Madan Lal Puri* (S. Ghosh, ed.) 639–667. Dekker, New York

GHOSAL, S., GHOSH, J. K. and RAMAMOORTHI, R.V. (1999b). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27** 143–158.

GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. (1998). Convergence rates of posterior distributions. Technical report, Free Univ. Amsterdam.

GRENANDER U. (1981) *Abstract Inference*. Wiley, New York.

KUO, H. H. (1975). *Gaussian Measures on Banach Spaces*. *Lecture Notes in Math.* **463**. Springer, Berlin.

SHEN, X. (1995). On the properties of Bayes procedures in general parameter spaces. Technical report, Ohio State Univ.

SHEN, X. and WONG, W. H. (1994). Convergence rate of sieve estimates. *Ann. Statist.* **22** 580–615.

STONE, C. (1982). Optimal global rates of convergence for non-parametric regression. *Ann. Statist.* **10** 1040–1053.

SCHWARTZ, L. (1965). On Bayes' procedures. *Z. Wahrsch. Verw. Gebiete* **4** 10–26.

TRIEBEL, H. (1983). *Theory of Function Spaces*. Birkhäuser, Boston.

VAN DE GEER, S. (2000). *Empirical Processes in M-Estimation*. Cambridge Univ. Press.

WASSERMAN, L. (1998). Asymptotic properties of nonparametric Bayesian procedures. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (D. Dey, P. Müller and D. Sinha, eds.) 293–304. Springer, New York.

WONG, W. H. and SHEN, X. (1995). A probability inequality for the likelihood surface and convergence rate of the maximum likelihood estimate. *Ann. Statist.* **23** 339–362.

ZEIDLER, E. (1990). *Nonlinear Functional Analysis and its Applications II/A*. Springer, New York.

ZHAO, L. (1993). Frequentist and Bayesian aspects of some nonparametric estimation. Ph.D. dissertation, Cornell Univ.

ZHAO, L. (1998). A hierarchical Bayesian approach in nonparametric function estimation. Technical report, Dept. Statistics, Univ. Pennsylvania.

DEPARTMENT OF STATISTICS
OHIO STATE UNIVERSITY
1958 NEIL AVENUE
COLUMBUS, OHIO 43210

DEPARTMENT OF STATISTICS
BAKER HALL
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15217
E-MAIL: larry@stat.cmu.edu