



On posterior pointwise convergence rate of a Gaussian signal under a conjugate prior

Alexandra Babenko, Eduard Belitser

► To cite this version:

Alexandra Babenko, Eduard Belitser. On posterior pointwise convergence rate of a Gaussian signal under a conjugate prior. *Statistics and Probability Letters*, 2009, 79 (5), pp.670. <10.1016/j.spl.2008.10.019>. <hal-00508917>

HAL Id: hal-00508917

<https://hal.science/hal-00508917v1>

Submitted on 7 Aug 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Accepted Manuscript

On posterior pointwise convergence rate of a Gaussian signal under a conjugate prior

Alexandra Babenko, Eduard Belitser

PII: S0167-7152(08)00496-3
DOI: [10.1016/j.spl.2008.10.019](https://doi.org/10.1016/j.spl.2008.10.019)
Reference: STAPRO 5248

To appear in: *Statistics and Probability Letters*

Received date: 11 July 2008
Revised date: 8 October 2008
Accepted date: 19 October 2008

Please cite this article as: Babenko, A., Belitser, E., On posterior pointwise convergence rate of a Gaussian signal under a conjugate prior. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.10.019

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



On posterior pointwise convergence rate of a Gaussian signal under a conjugate prior

Alexandra Babenko and Eduard Belitser *

Department of Mathematics, Utrecht University, PO Box 80 010, 3508 TA Utrecht, the Netherlands

8th October 2008

Abstract

We consider the problem of Bayes estimation of a linear functional of the signal in the Gaussian white noise mode, under the assumption that the unknown signal is from a Sobolev smoothness class. We propose a family of conjugate (Gaussian) priors and prove that the resulting Bayes estimators are rate minimax from both frequentist and Bayes perspectives. Finally, we show that the posterior distribution of the functional concentrates around the true value of the functional with the minimax rate uniformly over the Sobolev class.

Keywords: Minimax risk; Prior; Bayes risk; Bayes estimator; Posterior convergence rate.

1 Introduction

Suppose we observe $X = X^{(n)} = (X_1, X_2, \dots)$:

$$X_i = \theta_i + n^{-1/2}\xi_i, \quad i = 1, 2, \dots, \quad (1)$$

where ξ_i are independent $\mathcal{N}(0, 1)$ noises, $\theta = (\theta_1, \theta_2, \dots)$ is an unknown infinite dimensional parameter, which we will call signal, parameter $n \rightarrow \infty$ reflects the increase of information. The goal is to make an inference on the linear functional of the signal θ in the asymptotic setup as $n \rightarrow \infty$. For a given fixed $\theta \in \ell_2$, denote the corresponding probability measure of X by $P_\theta = P_\theta^{(n)}$. In case θ is a stochastic element, P_θ denotes a conditional probability measure of X given θ . Many quantities will depend on information parameter n , but for the sake of notation simplicity we will often skip this dependence.

*Corresponding author. Fax: +31-30-2518394.
E-mail address: e.belitser@uu.nl (E. Belitser).

The interest in model (1) is mostly motivated by its well known equivalence to the white noise model (see Belitser and Levit (1995), Johnstone (1999)):

$$dX_\epsilon(t) = f(t)dt + \epsilon dW(t), \quad 0 \leq t \leq 1, \quad (2)$$

where $X_\epsilon(\cdot)$ is an observation process, $f(\cdot) \in L_2[0, 1]$ is an unknown signal, $W(t)$ is a standard Wiener process, $\epsilon > 0$ is the noise intensity. The problem is studied in the asymptotic setup $\epsilon \rightarrow 0$. Indeed, suppose $\{\phi_i, i \in \mathbb{N}\}$ is an orthonormal basis in $L_2[0, 1]$. Then the problem can be translated into an equivalent *sequence model* (1) with observations $X_i = \int_0^1 \phi_i(t) dX_\epsilon(t)$, the unknown Fourier coefficients $\theta_i = \int_0^1 \phi_i(t) f(t) dt$, independent Gaussian noises $\xi_i = \int_0^1 \phi_i(t) dW(t)$ and $n = \epsilon^{-2}$. Since there is an isometric (with respect to the scalar products in ℓ_2 and L_2) correspondence between θ and $f(\cdot)$, we call both θ and $f(\cdot)$ signal. Being interesting in communication theory on its own, model (2) also provides a good approximation to a variety of curve estimation problems; see Nussbaum (1996) for density estimation; Brown and Low (1996) for non-parametric regression.

Suppose we are interested in estimating the signal $f(t)$ from the model (2) at a fixed point $t \in [0, 1]$. Then, in terms of parameter θ , this problem corresponds to the estimation of the linear functional $\Phi = \Phi(\theta) = \sum_{i=1}^\infty b_i \theta_i$ in the model (1), where $b_i = \phi_i(t)$. The above two problems are equivalent if the pointwise convergence $f(t) = \sum_{i=1}^\infty \theta_i \phi_i(t) = \Phi(\theta)$ holds, which can be provided by some conditions on $f(\cdot)$. For example, for the standard trigonometrical basis, continuity and Dini condition at point t would do.

In this paper, we study the problem of estimation of the linear functional $\Phi(\theta)$ in the model (1) by using a Bayesian approach. Assume that $|b_i| \leq 1$ for all $i \in \mathbb{N}$ and the signal θ belongs to a Sobolev class of certain smoothness (to be defined later). We put a simple conjugate prior on the signal θ and investigate the performance of the resulting Bayesian procedure. Li and Zhao (2002) studied this problem and showed that a certain choice of normal prior leads to the Bayes estimator which attains the minimax rate over the Sobolev ball. However, as Li and Zhao (2002) demonstrate, this prior (and posterior) assigns zero mass to the underlying Sobolev space. In fact, they show that there does not exist a Gaussian prior on Sobolev space such that the corresponding Bayes estimator attains the optimal minimax rate. Related results about the Bayesian estimation of the entire signal θ in ℓ_2 -norm can be found in Zhao (2000) and in Belitser and Ghosal (2003).

In this note, we continue the investigation line started by Li and Zhao (2002). We propose a family of conjugate priors and establish that any prior from this family leads to the minimax rate Bayes estimator of the functional $\Phi(\theta)$. The prior from Li and Zhao (2002) is the only member from this family which does not depend on the information parameter n . Next we show that the Bayes risk is of the same order as the minimax risk. However, again the prior and the corresponding posterior mass ‘escape’ the parameter space of interest. Despite this undesirable property, the posterior still does the ‘right thing’ as to the estimation of the linear functional $\Phi(\theta)$: using the results on the frequentist and Bayesian performance of our Bayes estimator, we finally establish that the posterior

distribution of the functional $\Phi(\theta)$ concentrates around the true value of the functional with the minimax rate, uniformly over the Sobolev class.

2 Preliminaries

For a smoothness parameter $\beta > 0$, introduce a Sobolev space $\Theta_\beta = \{\theta : \sum_{i=1}^{\infty} i^{2\beta} \theta_i^2 < \infty\}$ and an ℓ_2 -ellipsoid (which we call Sobolev class)

$$\Theta_\beta(Q) = \{\theta : \sum_{i=1}^{\infty} i^{2\beta} \theta_i^2 \leq Q\}.$$

Denote by $\hat{\Phi} = \hat{\Phi}(X)$ an estimator of linear functional $\Phi = \Phi(\theta) = \sum_{i=1}^{\infty} b_i \theta_i$, a measurable function of observation X from model (1). Assume throughout that $\theta \in \Theta_\beta(Q)$, $\beta > 1/2$ and $|b_i| \leq 1$, $i \in \mathbb{N}$. The functional $\Phi(\theta)$ is then well defined. Indeed, by Cauchy-Schwartz inequality

$$(\Phi(\theta))^2 \leq \left(\sum_{i=1}^{\infty} |b_i \theta_i| \right)^2 \leq \sum_{i=1}^{\infty} b_i^2 i^{-2\beta} \sum_{i=1}^{\infty} \theta_i^2 i^{2\beta} < \infty.$$

For an estimator $\hat{\Phi}$, the estimation quality is measured by the risk function

$$R(\hat{\Phi}) = R(\hat{\Phi}, \Phi(\theta)) = E_\theta (\hat{\Phi} - \Phi(\theta))^2.$$

It is known that, under appropriate assumptions on $b = (b_1, b_2, \dots)$, the minimax risk

$$r_n(\Theta_\beta(Q)) = \inf_{\hat{\Phi}} \sup_{\theta \in \Theta_\beta(Q)} R(\hat{\Phi}, \Phi(\theta))$$

is of order $n^{-(2\beta-1)/2\beta}$ and it is sharp (see Donoho and Low (1992)), i.e. there exist $0 < C_1 \leq C_2 < \infty$, such that

$$C_1 \leq \liminf_{n \rightarrow \infty} n^{(2\beta-1)/2\beta} r_n(\Theta_\beta(Q)) \leq \limsup_{n \rightarrow \infty} n^{(2\beta-1)/2\beta} r_n(\Theta_\beta(Q)) \leq C_2.$$

The lower bound in the above relations holds only for the so called ‘nonparametric’ $b = (b_1, b_2, \dots)$; a simple example is $|b_i| \geq \kappa > 0$, $i \in \mathbb{N}$.

Remark 1. This lower bound holds also for the minimax risk over the ellipsoid $\tilde{\Theta}_\beta(Q) = \{\theta : \sum_{k=1}^{\infty} k^{2\beta} (\theta_{2k}^2 + \theta_{2k+1}^2) \leq Q\}$ and sequence $b = (b_1, b_2, \dots)$ such that $b_{2k}^2 + b_{2k+1}^2 = 2$, $k \in \mathbb{N}$. The reason to look at this case is that the estimation of functional $\Phi(\theta)$ over class $\tilde{\Theta}_\beta(Q)$ for $\beta \in \mathbb{N}$ corresponds to the pointwise estimation of signal $f(t)$ in the model (2), where signal f belongs to a periodic Sobolev functional class of smoothness β (which is a subset of the usual Sobolev functional class, so that the same lower bound holds for the Sobolev class), see Evromovich (1999) and Tsybakov (2004). In this case one takes

the basis $\{\phi_k, k \in \mathbb{N}\}$ to be standard trigonometric: $\phi_1(t) = 1$, $\phi_{2k}(t) = \sqrt{2} \cos(2\pi kt)$, $\phi_{2k+1}(t) = \sqrt{2} \sin(2\pi kt)$, $k \in \mathbb{N}$, so that the above relations on $b_k = \phi_k(t)$, $k \in \mathbb{N}$, are indeed fulfilled.

In this paper we use a Bayesian approach. Namely, we put a prior π on signal θ and study the performance of the resulting Bayes procedure. The prior should be chosen in such a way that it adequately models the deterministic condition $\theta \in \Theta_\beta(Q)$. We have however to specify exactly what we mean by ‘adequately models’. Our prime statistical inference problem under study is the estimation of a linear functional Φ of signal θ and our main performance criterion is the convergence rate of the risk for this estimation problem. We propose a family of priors and look at the performance of the resulting Bayes procedures from the two different perspectives: frequentist (maximal risk) and Bayesian (Bayes risk). The priors from this family are shown to be adequate in the sense that for each prior from this family the corresponding Bayes estimator of the functional turns out to have the minimax convergence rate, simultaneously under both Bayesian and frequentist formulations. Finally, we show that these results lead to the adequate behavior of the posterior: the posterior distribution of the functional $\Phi(\theta)$ concentrates around the true value of the functional with the minimax rate, uniformly over the Sobolev class.

To be precise, let $\theta_0 \in \Theta_\beta(Q)$ be the true value of unknown deterministic parameter θ . Now we look at the posterior distribution of $r_n^{-1/2}|\Phi(\theta) - \Phi(\theta_0)|$, given X , for some appropriate positive normalising sequence $r_n \rightarrow 0$ as $n \rightarrow \infty$, from the point of view of the ‘true’ distribution $X \sim P_{\theta_0}$. We say that the posterior of the functional Φ converges with rate r_n if for any $M_n \rightarrow \infty$ as $n \rightarrow \infty$,

$$\pi(r_n^{-1/2}|\Phi(\theta) - \Phi(\theta_0)| \geq M_n | X) \rightarrow 0$$

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

If we put a prior π on signal θ , we can then regard the measure P_θ as the conditional distribution of X given θ , so that we get a joint distribution on (X, θ) . Denote by E_θ the expectation with respect to the conditional distribution of X given θ , by E_π the expectation with respect to the prior distribution of θ and by E the expectation with respect to the joint distribution of (X, θ) . Introduce the Bayesian mean square risk

$$R_\pi(\hat{\Phi}) = E_\pi E_\theta |\hat{\Phi} - \Phi(\theta)|^2 = E(\hat{\Phi} - \Phi(\theta))^2. \quad (3)$$

3 Main results

For a given smoothness parameter $\beta > 1/2$, introduce the family of priors $\{\pi_{\delta, \beta}, \delta < 2\beta - 1\}$, with $\pi_{\delta, \beta} = \pi_{\delta, \beta, n}$ such that if $\theta \sim \pi_{\delta, \beta}$, then

$$\theta_i \stackrel{ind}{\sim} \mathcal{N}(0, \tau_i^2(\delta, \beta)), \quad \tau_i^2(\delta) = \tau_i^2(\delta, \beta) = \tau_i^2(\delta, \beta, n) = i^{-2\beta + \delta} n^{-\delta/(2\beta)}, \quad i \in \mathbb{N}. \quad (4)$$

Remark 2. In terms of the continuous white noise model (2), the prior on signal $f(\cdot)$ represents a Gaussian random process $\eta(t) = \sum_{i=1}^{\infty} \theta_i \phi_i(t)$, $t \in [0, 1]$ with zero mean and covariance $\text{Cov}(\eta(t), \eta(s)) = \sum_{i=1}^{\infty} \tau_i^2(\delta, \beta) \phi_i(t) \phi_i(s)$. In case $\{\phi_k, k \in \mathbb{Z}\}$ is the standard Fourier basis, i.e. $\phi_k(t) = e^{2\pi k t i}$, this random process is stationary.

Remark 3. Notice that there is only one prior from the family which does not depend on the information parameter n : $\pi_{0,\beta}$ (i.e. when $\delta = 0$), the one considered in Li and Zhao (2002). According to the (pure) Bayesian tradition, one should use a prior which does not depend on n . In order to allow more modelling flexibility, we, however, do not impose this requirement. After all, using such priors does not affect formal constructions.

Remark 4. As discussed by Li and Zhao (2002), the prior $\pi_{0,\beta}$ assigns zero mass to the Sobolev space Θ_β since series $\sum_{i=1}^{\infty} i^{2\beta} \tau_i^2(0, \beta)$ diverges. Of course, we can take prior $\pi_{\delta,\beta}$ with $\delta < -1$ to get a convergent series $\sum_{i=1}^{\infty} i^{2\beta} \tau_i^2(\delta, \beta) = n^{-\delta/(2\beta)} \sum_{i=1}^{\infty} i^\delta = C(\delta) n^{-\delta/(2\beta)}$. The above undesirable property of the prior essentially remains though: the prior mass ‘escapes’ Θ_β as $n \rightarrow \infty$.

Recall the following simple fact: if $Z|Y \sim \mathcal{N}(Y, \sigma^2)$ and $Y \sim \mathcal{N}(\mu, \tau^2)$, then

$$Y|Z \sim \mathcal{N}\left(\frac{Z\tau^2 + \mu\sigma^2}{\tau^2 + \sigma^2}, \frac{\tau^2\sigma^2}{\tau^2 + \sigma^2}\right).$$

Thus, the prior (4) leads to the following posterior distribution $\pi_{\delta,\beta}(\theta|X)$:

$$\theta_i|X \stackrel{\text{ind}}{\sim} \mathcal{N}\left(\frac{\tau_i^2(\delta, \beta)X_i}{\tau_i^2(\delta, \beta) + n^{-1}}, \frac{\tau_i^2(\delta, \beta)n^{-1}}{\tau_i^2(\delta, \beta) + n^{-1}}\right) \quad (5)$$

Therefore the classical Bayes estimator of functional $\Phi(\theta)$ is

$$\begin{aligned} \tilde{\Phi} &= \tilde{\Phi}(\delta) = \tilde{\Phi}(\delta, \beta) = \tilde{\Phi}(\delta, \beta, X) = E(\Phi|X) = \sum_{i=1}^{\infty} b_i E(\theta_i|X) = \sum_{i=1}^{\infty} b_i \tilde{\theta}_i, \\ \tilde{\theta}_i &= \tilde{\theta}_i(\delta) = \tilde{\theta}_i(\delta, \beta) = E(\theta_i|X) = E(\theta_i|X_i) = \frac{\tau_i^2(\delta)X_i}{\tau_i^2(\delta) + n^{-1}}, \quad i \in \mathbb{N}. \end{aligned} \quad (6)$$

For $p, q > 0$, $r > -1$, $pq > r + 1$, introduce the function

$$B(p, q, r) = \int_0^{\infty} \frac{u^r}{(1+u^p)^q} du = p^{-1} \text{Beta}\left(q - \frac{r+1}{p}, \frac{r+1}{p}\right), \quad (7)$$

where $\text{Beta}(\alpha, \beta) = \int_0^1 u^{\alpha-1} (1-u)^{\beta-1} du$ is the beta function. From the properties of the beta function it is known that

$$B(p, 1, r) = p^{-1} \text{Beta}\left(1 - \frac{r+1}{p}, \frac{r+1}{p}\right) = \frac{\pi}{p^2 \sin(\pi(r+1)/p)}. \quad (8)$$

We will need the following technical lemma which is a slightly more general version of the corresponding result from Belitser and Enikeeva (2008).

Lemma 1. Suppose $0 < p < \infty$, $0 < q < \infty$, $-1 < r < pq - 1$. Let $B(p, q, r)$ be defined by (7). For $r \geq 0$ define $D(p, q, r) = r^{r/p}(pq - r)^{q-(r/p)}(pq)^{-q} = \left(1 - \frac{r}{pq}\right)^q \left(\frac{pq}{r} - 1\right)^{-r/p}$, with the convention $0^0 = 1$. Let positive $\gamma_n \rightarrow \infty$ as $n \rightarrow \infty$. Then

$$\sum_{i=1}^{\infty} \frac{i^r}{(\gamma_n + i^p)^q} = B(p, q, r) \gamma_n^{\frac{r+1}{p}-q} + \phi_n,$$

where $\phi_n = \phi_n(p, q, r)$ such that $|\phi_n| \leq D(p, q, r) \gamma_n^{-q+r/p}$ if $r \geq 0$ and $|\phi_n| \leq \gamma_n^{-q}$ if $r < 0$.

Proof. Denote $g(u) = \frac{u^r}{(\gamma_n + u^p)^q}$, $u \in \mathbb{R}_+ = \{u : u > 0\}$. If $r \geq 0$, the function $g(u)$ is increasing on $u \in [0, u_{\max}]$ and decreasing on $[u_{\max}, \infty)$ with $u_{\max} = (r\gamma_n/(pq - r))^{1/p}$. If $-1 < r < 0$, the function $g(u)$ is decreasing on \mathbb{R}_+ and in this case we let $u_{\max} = 1$. Now we have that

$$\int_0^{\infty} \frac{u^r du}{(\gamma_n + u^p)^q} - g(u_{\max}) \leq \sum_{i=1}^{\infty} \frac{i^r}{(\gamma_n + i^p)^q} \leq \int_0^{\infty} \frac{u^r du}{(\gamma_n + u^p)^q} + g(u_{\max}),$$

with $g(u_{\max}) = D(p, q, r) \gamma_n^{(r/p)-q}$ if $r \geq 0$ and $|g(u_{\max})| \leq \gamma_n^{-q}$ if $r < 0$, which establishes the lemma. \square

Remark 5. If $r \leq -1$, then for some $C_1 = C_1(p, q, r)$ and $C_2 = C_2(p, q, r)$ we can get a trivial asymptotic relation

$$C_1 \gamma_n^{-q} \leq \sum_{i=1}^{\infty} \frac{i^r}{(\gamma_n + i^p)^q} \leq C_2 \gamma_n^{-q},$$

which is not really informative.

The following theorem illustrates that the chosen prior adequately reflects the requirement $\theta \in \Theta_{\beta}(Q)$: both frequentist and Bayes risks of the Bayes estimator $\tilde{\Phi}$ have minimax convergence rate uniformly over Sobolev ellipsoid $\Theta_{\beta}(Q)$.

Theorem 1. Let $\pi = \pi_{\delta, \beta}$ be a prior from the family (4) with $\delta < \min\{2\beta - 1, \beta + 1/2\}$, $\tilde{\Phi} = \tilde{\Phi}(\delta, \beta, X)$ be the corresponding Bayes estimator defined by (6), function $B(p, q, r)$ be defined by (7) and $|b_i| \leq 1$, $i \in \mathbb{N}$. Then

$$\limsup_{n \rightarrow \infty} n^{\frac{2\beta-1}{2\beta}} \sup_{\theta \in \Theta_{\beta}(Q)} R(\tilde{\Phi}, \Phi(\theta)) \leq B(2\beta - \delta, 2, 0) + QB(2\beta - \delta, 2, 2\beta - 2\delta)$$

and

$$\limsup_{n \rightarrow \infty} n^{\frac{2\beta-1}{2\beta}} R_{\pi}(\tilde{\Phi}) \leq \frac{\pi}{(2\beta - \delta)^2 \sin(\pi/(2\beta - \delta))}.$$

Proof. Split the risk into the sum of the bias and variance terms:

$$\begin{aligned} R(\tilde{\Phi}, \Phi(\theta)) &= E_{\theta}(\tilde{\Phi} - \Phi(\theta))^2 = E_{\theta}\left(\sum_{i=1}^{\infty} \frac{b_i \tau_i^2(\delta) X_i}{\tau_i^2(\delta) + n^{-1}} - b_i \theta_i\right)^2 \\ &= \left(\sum_{i=1}^{\infty} \frac{n^{-1} b_i \theta_i}{\tau_i^2(\delta) + n^{-1}}\right)^2 + \sum_{i=1}^{\infty} \frac{n^{-1} b_i^2 \tau_i^4(\delta)}{(\tau_i^2(\delta) + n^{-1})^2}. \end{aligned}$$

First evaluate the variance term. Since $\delta < 2\beta - 1$, by Lemma 1 we have

$$\begin{aligned} \sum_{i=1}^{\infty} \frac{n^{-1} b_i^2 \tau_i^4(\delta)}{(\tau_i^2(\delta) + n^{-1})^2} &\leq \sum_{i=1}^{\infty} \frac{n^{1-\delta/\beta}}{(n^{1-\delta/(2\beta)} + i^{2\beta-\delta})^2} \\ &= n^{-\frac{2\beta-1}{2\beta}} B(2\beta - \delta, 2, 0)(1 + o(1)). \end{aligned}$$

By Cauchy-Schwartz inequality and Lemma 1, we obtain the following upper bound for the bias term: uniformly over $\theta \in \Theta_{\beta}(Q)$

$$\begin{aligned} \left(\sum_{i=1}^{\infty} \frac{n^{-1} b_i \theta_i}{\tau_i^2(\delta) + n^{-1}}\right)^2 &\leq \sum_{i=1}^{\infty} \theta_i^2 i^{2\beta} \sum_{i=1}^{\infty} \frac{n^{-2} b_i^2}{i^{2\beta} (\tau_i^2(\delta) + n^{-1})^2} \\ &\leq \sum_{i=1}^{\infty} \frac{Q i^{2\beta-2\delta}}{(n^{1-\delta/(2\beta)} + i^{2\beta-\delta})^2} \\ &= n^{-\frac{2\beta-1}{2\beta}} Q B(2\beta - \delta, 2, 2\beta - 2\delta)(1 + o(1)), \end{aligned} \tag{9}$$

since we assumed also $\delta < \beta + 1/2$; see remark on this condition below. The first relation is proved.

Using (8) and Lemma 1, we derive the asymptotic expression for the Bayes risk

$$\begin{aligned} R_{\pi}(\tilde{\Phi}) &= \sum_{i=1}^{\infty} \frac{b_i^2 \tau_i^2(\delta) n^{-1}}{\tau_i^2(\delta) + n^{-1}} \leq \sum_{i=1}^{\infty} \frac{n^{-\delta/(2\beta)}}{n^{1-\delta/(2\beta)} + i^{2\beta-\delta}} \\ &= n^{-\frac{2\beta-1}{2\beta}} B(2\beta - \delta, 1, 0)(1 + o(1)) \\ &= \frac{n^{-\frac{2\beta-1}{2\beta}} \pi}{(2\beta - \delta)^2 \sin(\pi/(2\beta - \delta))} (1 + o(1)), \end{aligned}$$

as $n \rightarrow \infty$. □

Remark 6. Clearly the condition $\delta < 2\beta - 1$ is needed to ensure the convergence of the variance term of the frequentist risk and the Bayes risk. Let us explain why we need the condition $\delta < \beta + 1/2$. If we had $\delta > \beta + 1/2$, then the upper bound for the bias term (see (9) and Remark 5) would be as follows

$$C_2 n^{-2+\delta/\beta} \geq \sum_{i=1}^{\infty} \frac{Q i^{2\beta-2\delta}}{(n^{1-\delta/(2\beta)} + i^{2\beta-\delta})^2} \geq C_1 n^{-2+\delta/\beta} \gg n^{-\frac{2\beta-1}{2\beta}}$$

as $n \rightarrow \infty$, i.e. of a worse order as compared to the optimal rate $n^{-\frac{2\beta-1}{2\beta}}$.

Notice that if $\delta = \beta + 1/2$, the bias term is of the correct order $n^{-\frac{2\beta-1}{2\beta}}$, but with some other constant. So the theorem holds also under milder condition $\delta \leq \beta + 1/2$ with adjusted constant for the case $\delta = \beta + 1/2$.

Using the above results, it is not difficult to derive our final result on the posterior convergence rate of the linear functional of the signal. The theorem below says essentially that the posterior distribution of $\Phi(\theta)$ behaves correctly in the sense that it concentrates around the true value $\Phi(\theta_0)$ with the correct rate $r_n = r_n(\beta) = n^{-(2\beta-1)/(2\beta)}$. In fact, using (5), it is not difficult to derive the posterior distribution of $\Phi(\theta)$ given X :

$$\Phi(\theta)|X \sim \mathcal{N}(\tilde{\Phi}(\delta, \beta, X), \text{Var}(\Phi(\theta)|X)), \quad \text{Var}(\Phi(\theta)|X) = \sum_{i=1}^{\infty} \frac{b_i^2 \tau_i^2(\delta) n^{-1}}{\tau_i^2(\delta) + n^{-1}}.$$

Theorem 2. Let $\beta > 1/2$, $\delta < \min\{2\beta - 1, \beta + 1/2\}$ and $r_n = r_n(\beta) = n^{-\frac{2\beta-1}{2\beta}}$. Then for any $M_n \rightarrow \infty$,

$$\pi(r_n^{-1/2}|\Phi(\theta) - \Phi(\theta_0)| \geq M_n|X) \rightarrow 0$$

in P_{θ_0} -probability as $n \rightarrow \infty$, uniformly over $\theta_0 \in \Theta_\beta(Q)$.

Proof. Let $\tilde{\Phi}$ be the Bayes estimator of Φ , defined by (6). By the conditional Chebyshev inequality, we have

$$\begin{aligned} \pi(r_n^{-1/2}|\Phi(\theta) - \Phi(\theta_0)| \geq M_n|X) &\leq \frac{r_n^{-1}E(\Phi(\theta) - \Phi(\theta_0))^2|X)}{M_n^2} \\ &\leq \frac{r_n^{-1}\text{Var}(\Phi(\theta)|X)}{M_n^2} + \frac{r_n^{-1}(\tilde{\Phi} - \Phi(\theta_0))^2}{M_n^2}. \end{aligned}$$

It is enough to show that the expectation of the right hands side of the last relation with respect to distribution P_{θ_0} tends to zero as $n \rightarrow \infty$. Since $r_n^{-1} = n^{\frac{2\beta-1}{2\beta}}$ and

$$E_{\theta_0}[\text{Var}(\Phi(\theta)|X) + (\tilde{\Phi} - \Phi(\theta_0))^2] = R_\pi(\tilde{\Phi}) + R(\tilde{\Phi}, \Phi(\theta)),$$

the result follows from the previous theorem. \square

Remark 7. The above theorem guarantees good frequentist properties of credible intervals for the functional $\Phi(\theta)$, constructed on the basis of posterior distribution.

References

- [1] Belitser, E., Enikeeva, F., 2008. Empirical Bayesian test for the smoothness. Math. Meth. Statist. 17, 1–18.

- [2] Belitser, E., Ghosal, S., 2003. Adaptive Bayesian inference on the mean of an infinite dimensional normal distribution. *Ann. Statist.* 31, 536–559.
- [3] Belitser, E., Levit, B., 1995. On minimax filtering over ellipsoids. *Math. Meth. Statist.* 3, 259–273.
- [4] Brown, L. D., Low, M. G., 1996. Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* 24, 2384–2398.
- [5] Donoho, D. L., Low, M. G., 1992. Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* 20, 944–970.
- [6] Evromovich, S., 1999. *Nonparametric Curve Estimation*. Springer, New York.
- [7] Johnstone, I., 1999. *Function Estimation in Gaussian Noise: Sequence models*. Monograph draft, <http://www-stat.stanford.edu/~imj/>
- [8] Li, X., Zhao, L. H., 2002. Bayesian nonparametric point estimation under a conjugate prior. *Statist. Prob. Lett.* 58, 23–30.
- [9] Nussbaum, M., 1996. Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* 24, 2399–2430.
- [10] Tsybakov, A., 2004. *Introduction à l'estimation non-paramétrique*. Springer, Berlin.
- [11] Zhao, L. H., 2000. Bayesian aspects of some nonparametric problems. *Ann. Statist.* 28, 532–552.