**Credible sets in nonparametric regression**

Sniekers, Suzanne

**Author**: Sniekers, Suzanne
**Title**: Credible sets in nonparametric regression
**Issue Date**: 2015-12-01

*Chapter 3*

# Credible bands

## 3.1 Introduction and main results

We consider estimating the regression function $f$ in the fixed design regression model, introduced in Chapter 1. We observe a vector $\vec{Y}_n := (Y_{1,n}, \dots, Y_{n,n})^T$ with coordinates

$$Y_{i,n} = f(x_{i,n}) + \varepsilon_{i,n}, \qquad i \in \{1, \dots, n\}. \tag{3.1}$$

Here the parameter $f$ is an unknown function $f : [0,1] \to \mathbb{R}$, the design points $(x_{i,n})$ are a known sequence of points in $[0,1]$ and the (unobservable) errors $\varepsilon_{i,n}$ are independent standard normal random variables.

When visualising the uncertainty of an estimate of the function $f$ at a point or several points, credible intervals around the estimated function value or even credible bands might be considered to be the most intuitive. The credible balls relative to the Euclidian norm in the space of function values of $f$ at the design points studied in Chapter 2 are more complicated to interpret or represent graphically. Therefore, in this chapter we focus on a uniform bound for the width of the credible intervals for $f$ at the design points (or at all $x$), rather than bounding an average. We return to the setting of Chapter 1, taking the design points $x_{i,n}$ to be equally spaced and equal to $x_{i,n} = i/n_+$ with $n_+ = n + 1/2$ and $f \in C^\alpha[0,1]$ for some $\alpha \in (0,2]$. It will turn out to be essential to assume some "local" smoothness of the function $f$, since we are studying the behaviour of this function and its estimators at specific points. In particular, assumptions on the behaviour of $f$ in terms of its Fourier coefficients or variants thereof, such as the polished tail condition or a Sobolev smoothness considered in Chapter 2, will turn out to be insufficient to fully determine this local behaviour.

We take scaled Brownian motion $\sqrt{c}\,W = (\sqrt{c}\,W_t, t \in [0,1])$ as a prior for $f$, independent of the sequence $(\varepsilon_{i,n})$. In the Bayesian setup the observations are then distributed according to the model

$$Y_{i,n} = \sqrt{c}\,W_{x_{i,n}} + \varepsilon_{i,n}, \qquad i \in \{1,\dots,n\}. \tag{3.2}$$

When estimating the function at a fixed point $x \in (0,1)$ and constructing a credible interval for given $c$, one considers the posterior distribution

$$f(x)\,|\,\vec{Y}_n, c \sim \mathcal{N}\big(\hat{f}_{n,c}(x), \sigma_n^2(x,c)\big). \tag{3.3}$$

The posterior mean is a linear combination of the observations and was studied in detail in Chapter 1:

$$\hat{f}_{n,c}(x) = \sum_{i=1}^{n} a_i^n(x,c) Y_{i,n}, \quad \vec{a}_n(x,c) = \Sigma_{n,c}^{-1} c \vec{v}_n(x), \quad \vec{v}_n(x) = (i/n_+ \wedge x)_i.$$

Here $\Sigma_{n,c} = I + cU_n$, where $U_n$ is the covariance matrix of standard Brownian motion at the design points. Recall that the eigenvalues $\lambda_{j,n}$ of this matrix satisfy $\lambda_{j,n} \asymp \frac{n}{j^2}$. The behaviour of the posterior variance was studied in the same chapter and can be seen to have exact asymptotic behaviour $\frac{1}{2}\sqrt{\frac{c}{n}}$, uniformly for $c \in [\log n/n, n/\log n]$ and independently of $x$. The natural pointwise credible interval of level $\eta$ for $f$ is

$$C_{n,\eta}(x) = \{f : |f(x) - \hat{f}_{n,c}(x)| < \zeta_\eta \sigma_n(x,c)\},$$

where $\zeta_\eta$ is a standard normal quantile such that $P(|Z| < \zeta_\eta) = \eta$ for $Z \sim \mathcal{N}(0,1)$.

In order to obtain a credible band rather than a credible interval, one considers functions such that the supremum of $|f(x) - \hat{f}_{n,c}(x)|$ over $x$ is bounded. For this we need a uniform result on the coefficients $a_i^n$, which generalises the results from Chapter 1. Throughout we will use results from this chapter with $k = n_+/c$. The following proposition shows that the result from Theorem 1.3 still holds for $x = x_n$ tending to zero sufficiently slowly.

**Proposition 3.1.** *Let $i_n = \max\{i : x_{i,n} < x\}$ and $\lambda_+ = 1 + c/(2n_+) + \frac{1}{2}\sqrt{c/n_+}\sqrt{4 + c/n_+}$. If we take $x = x_n$ to be a sequence of design points $j_n/n$ such that $j_n$ tends to infinity no slower than $n_+/\sqrt{\log n}$, then the coefficients $a_i^n$ satisfy*

$$a_i^n \asymp \sqrt{\frac{c}{n}} \begin{cases} \lambda_+^{-i_n}\left[\lambda_+^{i} - \lambda_+^{-i}\right] & \text{for } i \le i_n \\ \lambda_+^{i_n}\left[\lambda_+^{-i+1} + \lambda_+^{-2n+i-1}\right] & \text{for } i \ge i_n. \end{cases}$$

The proof of this proposition can be found in Section 3.5.

Throughout, we will use that results on bias and variance from Chapter 1, which are based on the behaviour of the $a_i^n$, hold uniformly in both $c$ and $x$. For $c$, this follows since close inspection of the proofs reveals that it is sufficient that we have $c/n \to 0$ and this holds on $[\log n/n, n/\log n]$. The same result follows for $x$ by arguments similar to those used in (3.9) in the appendix. It follows that $\sigma_n^2(x,c) \asymp \sqrt{\frac{c}{n}}$ uniformly in $x \geq 1/\sqrt{\log n}$ and hence the diameter of a credible band is close to the width of a credible interval for a fixed $x$. However, we will see that in order to obtain favourable coverage $\eta$, one should blow up the radius by $\sqrt{\log n}$, c.f. [Stone, 1982].

We have seen in Chapter 1 that when we consider the performance of the Bayesian credible sets under the frequentist model (3.1), the optimal scaling $c$ depends on the smoothness of the function $f$. Since this is generally unknown, we will here estimate $c$ from the data using either the likelihood-based or risk-based estimator $\hat{c}_n$ defined in Chapter 2. Recall that for the Brownian motion prior it was proved in Chapter 2 that in both settings, the estimator $\hat{c}_n$ behaves similar to the minimiser of a deterministic function $D_n$ that is the sum of a decreasing bias function $D_{1,n}$ depending on the unknown function $f$ and an increasing variance function $D_{2,n}$, defined in (2.23) and (2.23). We then study the credible sets given by

$$\hat{C}_{n,\eta} = \left\{ f : \sup_{x\in[1/\sqrt{\log n},1]} |f(x) - \hat{f}_{n,\hat{c}_n}(x)| < \sqrt{\ell_n \log n} \left(\frac{\hat{c}_n}{n}\right)^{1/4} \right\}, \qquad (3.4)$$

where $\ell_n$ is a sequence of numbers tending to infinity.

A main result we will prove is that these credible sets are honest confidence sets for a certain class of functions. This result is in some sense surprising, since many previous results on coverage of credible sets in the fixed design model are negative in the sense that the obtained credible sets have zero frequentist coverage, see e.g. [Cox, 1993, Freedman, 1999, Johnstone, 2010]. Previous work on pointwise credible sets and credible bands can be found in for example [Bull, 2012]. Here adaptive confidence bands are constructed for functions satisfying a self-similarity condition defined in terms of a wavelet basis of $L^2[0,1]$. Similarly, [Hoffmann and Nickl, 2011] treats adaptive credible bands in density estimation for Hölder smooth functions which satisfy some additional condition. In [Wahba, 1983] credible intervals at the design points are discussed; simulations and heuristic arguments are given that suggest promising results, but no formal proofs are provided. Positive results for the Gaussian white noise model are for example the papers [Cai and Low, 2004] and [Ghosal, 2015]. In the latter, $L_\infty$-credible sets with asymptotic coverage one and a nearly optimal size are obtained when the Hölder regularity of the true function is known. The same paper also includes a heuristic discussion that indicates that the same result holds when instead an empirical Bayes estimate of the smoothness is used for functions that satisfy the polished tail condition. The same applies here: we need an extra assumption to ensure that the empirical Bayes credible sets capture the truth with high probability.

Recall that we call a function *self-similar* of order $\beta > 0$ if its sequence of Fourier coefficients $(f_j)$ with respect to the eigenbasis of Brownian motion satisfies

$$\sup_{j\geq 1} j^{1/2+\beta}|f_j| \leq M, \qquad \text{and} \qquad \sum_{j=m}^{\rho m} f_j^2 \geq M^2 L m^{-2\beta}$$

for some positive constants $M, \rho, L$ and every $m$. Let $\mathcal{F}_{\alpha,\beta,M}$ be the set of all functions that are self-similar of order $\beta$ for a given $M$ and satisfy a Hölder condition with constant $M$, i.e. $|f(x)-f(y)| \leq M|x-y|^\alpha$ if $\alpha \leq 1$ and $|f'(x)-f'(y)| \leq M|x-y|^{\alpha-1}$ if $\alpha \in (1,2]$.

**Theorem 3.2.** *Let $\alpha \in (0,2]$ and consider the credible sets given in* (3.4)*. For both the risk-based and likelihood-based empirical Bayes methods we have the following result. If $\alpha \geq \beta > \frac{1}{2}$, we have*

$$\inf_{f\in\mathcal{F}_{\alpha,\beta,M}} P_f(f \in \hat{C}_{n,\eta}) \to 1$$

*for any sequence $\ell_n \to \infty$. Furthermore, for $\beta < 1$ the diameter of the credible band $\hat{C}_{n,\eta}$ is $O_P\big(\sqrt{\ell_n \log n}\, n^{-\frac{\beta}{1+2\beta}}\big)$. For the risk-based empirical Bayes method this is even true for $\beta < 2$.*

The theorem shows that empirical Bayes works if the Hölder smoothness of the function is at least the order of self-similarity. As we shall see later in the proof, this is due to the fact that the behaviour of the estimator $\hat{c}_n$ is completely determined by this latter order. A value of $\beta > \alpha$ is equivalent to overestimating the Hölder smoothness of the function, which leads to poor coverage, as was already shown in Chapter 1. For functions that are Hölder smooth of the order $\alpha$, the minimax rate of estimation is known to be of the order $(n/\log n)^{-\frac{\alpha}{1+2\alpha}}$, see [Stone, 1982]. The theorem shows that the diameter of the credible band is determined by the order of self-similarity and in order to obtain a (nearly) optimal diameter, we need equality $\alpha = \beta$. In particular, the size of the credible set adapts to the unknown Hölder smoothness if this is equal to the self-similarity, which can be viewed as a global smoothness index.

In the next section we will prove this main result and give an example where we suspect the theorem fails. We then proceed to derive another assumption on the function $f$ for which the empirical Bayes procedure leads to coverage, for a slightly weaker version of the credible band. Here we consider

$$\hat{C}_{n,\eta} = \left\{ f : \max_{i\in\{1,\ldots,n\}} |f(x_{i,n}) - \hat{f}_{n,\hat{c}_n}(x_{i,n})| < L\sqrt{\ell_n \log n}\left(\frac{\hat{c}_n}{n}\right)^{\frac{1}{4}} \right\}, \qquad (3.5)$$

where $L$ is a sufficiently large constant, and derive a set of conditions under which this credible set has favourable coverage. We then show that these conditions are satisfied for a large class of Gaussian-generated functions:

**Theorem 3.3.** *For given $\alpha > 0$ and $\delta \in \mathbb{R}$ set*

$$W_t = \sum_{j=1}^{\infty} \frac{Z_j}{(j+\delta)^{1/2+\alpha}} e_j(t), \qquad t \in [0,1],$$

*where $Z_1, Z_2, \ldots$ are independent standard normal random variables. Then for both the risk-based and likelihood-based empirical Bayes methods, for almost every realisation of $W$, we obtain coverage tending to one for the credible sets defined in (3.5), where $\ell_n \geq \tilde{L}\log n$ and $\tilde{L}$ is sufficiently large. For the risk-based method this holds if $\alpha < 2$ and for the likelihood-based method if $\alpha < 1$. Furthermore, for the appropriate values of $\alpha$, the diameter of the credible sets is of the order $O_P\big(\sqrt{\ell_n \log n}\, n^{-\frac{\alpha}{1+2\alpha}}\big)$.*

Note that this results holds in particular for Brownian motion, by the Karhunen-Loève expansion.

The third section deals with the proof of this main result. Finally, the last section contains technical proofs. Throughout, we denote the interval $[\log n/n, n/\log n]$ by $I_n$ and $[1/\sqrt{\log n}, 1]$ by $J_n$.

## 3.2 Coverage for self-similar functions

Before we prove the first main result, we recall some facts about the two types of empirical Bayes estimators $\hat{c}_n$. Both minimise a criterium of the form

$$L_n(c,f) = D_{1,n}(c,f) + D_{2,n}(c) + R_n(c,f) = D_n(c,f) + R_n(c,f),$$

where $D_n$ is deterministic and $R_n$ a stochastic remainder. We add the superscript $R$ when we specifically refer to the risk-based functions and the superscript $L$ for the likelihood-based functions. The functions $D_{1,n}$ are dependent on the vector of function values $\vec{f_n}$ through its coefficients $f_{j,n}$ with respect to eigenbasis $e_{1,n}, \ldots, e_{n,n}$ of the covariance matrix of Brownian motion, given by

$$e_{j,n} = \frac{1}{\sqrt{n+1/2}}\big(e_j(x_{1,n}), \ldots, e_j(x_{n,n})\big)^T, \qquad e_j(x) = \sqrt{2}\sin\big[\big(j - \tfrac{1}{2}\big)\pi x\big].$$

The functions $D_{2,n}$ do not depend on $f$. They are increasing and behave asymptotically like $\sqrt{cn}$. In Chapter 2 it was proved that for both methods the estimator $\hat{c}_n$ is close to the minimiser of the deterministic part $D_n$. More precisely, we have

$$P_f(\hat{c}_n \in \Lambda_n) \to 1,$$

where

$$\Lambda_n = \{c \in I_n : D_n(c, f) \leq (1 + \varepsilon) \inf_{c \in I_n} D_n(c, f)\}$$

for some $\varepsilon > 0$. In the following we will use an even more precise result on the location of $\hat{c}_n$. For this, we need an assumption on $f$, expressed in terms of the function $D_{1,n}$, that was first introduced in the previous chapter.

**Definition 3.4** (Good bias condition)**.** We say that the function $f$, or the corresponding array $(f_{j,n})$, satisfies the *good bias condition* relative to $D_{1,n}$ if there exists a constant $a > 0$ such that for $c \in I_n$

$$D_{1,n}(Kc, f) \leq C^{-a} D_{1,n}(c, f) \qquad \text{for all } C > 1.$$

If $f$ is discrete polished tail, then it also satisfies this good bias condition. For the Brownian motion prior this is equivalent to the existence of constants $L$ and $\rho$ such that

$$\sum_{j=m}^{n} f_{j,n}^2 \leq L \sum_{j=m}^{\rho m \wedge n} f_{j,n}^2$$

for all sufficiently large $m$. Now let $\tilde{c}_n$ be the unique solution to $D_{1,n}(c, f) = D_{2,n}(c)$. We have the following result on the location of $\hat{c}_n$ for functions that satisfy the good bias condition and hence for self-similar functions by Example 2.34 in Chapter 2.

**Lemma 3.5.** *Let $f$ satisfy the good bias condition and $\tilde{c}_n \in I_n$. Then there are positive constants $k < K$ such that*

$$P_f(\hat{c}_n \in [k\tilde{c}_n, K\tilde{c}_n]) \to 1$$

*for both the risk-based and likelihood-based empirical Bayes estimators $\hat{c}_n$.*

*Proof.* Since $P_f(\hat{c}_n \in \Lambda_n) \to 1$ and $\tilde{c}_n \in I_n$, it follows that $D_n(\hat{c}_n, f) \leq (1 + \varepsilon) D_n(\tilde{c}_n, f)$ with probability tending to one. Recall that for the Brownian motion prior it holds that $D_{2,n}(c) \asymp \sqrt{cn}$. Combining this with the fact that $f$ satisfies the good bias condition, the result follows by Lemma 2.42 from Chapter 2. □

The function $f$ is contained in $\hat{C}_{n,\eta}$ as defined in (3.4) if and only if $\sup_{x\in J_n}|f(x)-\hat{f}_{n,\hat{c}_n}(x)|<\sqrt{\ell_n\log n}\,(\hat{c}_n/n)^{1/4}$. Writing

$$T_n(x,c):=\hat{f}_{n,c}(x)-\mathrm{E}_f\hat{f}_{n,c}(x)\qquad\text{and}\qquad\mu_n(x,c):=f(x)-\mathrm{E}_f\hat{f}_{n,c}(x),$$

we can decompose the square of the left side as

$$\big(f(x)-\hat{f}_{n,\hat{c}_n}(x)\big)^2=\big(T_n(x,c)\big)^2+2T_n(x,c)\mu_n(x,c)+\big(\mu_n(x,c)\big)^2$$

for any $c$ and $x$. Here $\mu_n(x,c)$ is the bias of the posterior mean for fixed $x$ and the expectation of $\big(T_n(x,c)\big)^2$ is equal to the variance of the posterior mean:

$$\mathrm{E}\big(T_n(x,c)\big)^2=\mathrm{E}\left(\sum_{i=1}^n a_i^n(x,c)\varepsilon_{i,n}\right)^2=\sum_{i=1}^n a_i^n(x,c)^2=:t_n^2(x,c).$$

These quantities were studied in detail in Chapter 1. The following proposition shows that the random term $T_n$ is negligible if $f$ satisfies the good bias condition.

**Proposition 3.6.** *Let $f$ satisfy the good bias condition and $\tilde{c}_n\in I_n$. For both the risk-based and likelihood-based empirical Bayes estimators $\hat{c}_n$ it holds that*

$$\sup_{x\in[1/n,1]}\frac{|T_n(x,\hat{c}_n)|}{\sqrt{\ell_n\log n}\left(\frac{\hat{c}_n}{n}\right)^{1/4}}\overset{P_f}{\to}0,$$

*for any sequence $\ell_n\to\infty$.*

The proof of this result is given in Section 3.5. We now have a sufficient understanding of the credible set $\hat{C}_{n,\eta}$ to be able to prove our main result.

*Proof of Theorem 3.2.* A function $f$ is contained in $\hat{C}_{n,\eta}$ if and only if $\sup_{x\in J_n}|f(x)-\hat{f}_{n,\hat{c}_n}(x)|<\sqrt{\ell_n\log n}\,(\hat{c}_n/n)^{1/4}$. Since for all fixed $c$ it holds that

$$\begin{aligned}\sup_{x\in J_n}|f(x)-\hat{f}_{n,c}(x)|&\leq\sup_{x\in J_n}\big(|\hat{f}_{n,c}(x)-\mathrm{E}_f\hat{f}_{n,c}(x)|+|f(x)-\mathrm{E}_f\hat{f}_{n,c}(x)|\big)\\&=\sup_{x\in J_n}\big(|T_n(x,c)|+|\mu_n(x,c)|\big),\end{aligned}$$

we obtain coverage if

$$\frac{\sup_{x\in J_n}|T_n(x,\hat{c}_n)|+\sup_{x\in J_n}|\mu_n(x,\hat{c}_n)|}{\sqrt{\ell_n\log n}\,(\hat{c}_n/n)^{1/4}}<1\tag{3.6}$$

with probability tending to one.

If $f$ is self-similar of order $\beta > \frac{1}{2}$, then the array $(f_{i,n})$ is discrete polished tail and hence $f$ satisfies the good bias condition (see Example 2.34 of Chapter 2). Moreover, we have

$$n\left(\frac{1}{cn}\right)^{\beta} \gtrsim D_{1,n}(c,f) = \sum_{i=1}^{n} \frac{f_{i,n}^2}{(1+c\lambda_j)^{\delta}} \gtrsim \sum_{i=\sqrt{cn}}^{n} f_{i,n}^2 \gtrsim n\left(\frac{1}{cn}\right)^{\beta}$$

by the last display of Example 2.34, where we take $\delta = 1$ for $D_{1,n}^L$ and $\delta = 2$ for $D_{1,n}^R$. The first inequality holds if $\beta < 2$ for the risk-based empirical Bayes method and if $\beta < 1$ for the likelihood-based method (see Examples 2.20 and 2.23 in Chapter 2). Since $\tilde{c}_n$ balances $D_{1,n}$ and $D_{2,n}$, it follows that it satisfies

$$\left(\frac{1}{\tilde{c}_n n}\right)^{\beta} \asymp \sqrt{\frac{\tilde{c}_n}{n}},$$

hence $\tilde{c}_n \asymp n^{\frac{1-2\beta}{1+2\beta}}$. Noting that $\tilde{c}_n \in I_n$, it follows by Proposition 3.6 that the first term of (3.6) converges to zero in probability. Moreover, by Lemma 3.5 there are positive constants $k < K$ such that $P_f(\hat{c}_n \in [k\tilde{c}_n, K\tilde{c}_n]) \to 1$. Consider the second term in (3.6). Since $f \in C^{\alpha}[0,1]$, we have

$$|\mu_n(x,c)|^2 \lesssim \left(\frac{1}{cn}\right)^{\alpha}$$

uniformly for $c \in [k\tilde{c}_n, K\tilde{c}_n]$ and $x \in J_n$ by Corollary 1.7 in Chapter 1. We see that for $\alpha \geq \beta$ we have

$$\sup_{x\in J_n} |\mu_n(x,\hat{c}_n)|^2 \lesssim \left(\frac{1}{\hat{c}_n n}\right)^{\alpha} \lesssim \left(\frac{1}{\hat{c}_n n}\right)^{\beta} \asymp \left(\frac{1}{\tilde{c}_n n}\right)^{\beta} \asymp \sqrt{\frac{\tilde{c}_n}{n}} \asymp \sqrt{\frac{\hat{c}_n}{n}}$$

with probability tending to one. It follows that $P_f(f \in \hat{C}_{n,\eta}) \to 1$. Finally, note that since $(\tilde{c}_n/n)^{1/4} \asymp n^{-\frac{\beta}{1+2\beta}}$, the diameter of the credible band is bounded by $\sqrt{\ell_n \log n}\, n^{-\frac{\beta}{1+2\beta}}$ with probability tending to one. □

The theorem gives an example of a class of functions for which the empirical Bayes credible band is an honest credible set. Although this positive result is encouraging, it is not clear how rich this class of functions is. It follows by inspection of the proof that the Hölder and self-similarity conditions are both essential, since the former determines the bias and the latter the location of the empirical Bayes estimator. If the two measures of smoothness do not match, the bias and variance are not balanced, which can result in zero coverage or credible bands that are too wide. Unfortunately, one of these types of smoothness typically does not imply the other. The following is an example of a nicely behaved function of Hölder smoothness $\alpha$, for which the order of self-similarity is greater than the Hölder exponent for a whole range of values of $\alpha$. This suggests that the coverage will tend to zero for all possible values of $\alpha$ in this range.

**Example 3.7.** Consider the function $f(t) = |x-t|^\alpha$, where $\alpha \in (0,1)$ and $x > \frac{1}{2}$. We prove below that the Fourier coefficients of this function satisfy $f_j \asymp j^{-1}$ and hence this function is self-similar of the order $\beta = \frac{1}{2}$, which suggests that $\tilde{c}_n$ satisfies $\tilde{c}_n \asymp 1$. Since the square bias of this function at the point $x$ is of the order $(cn)^{-\alpha}$, it then follows that $\sup_{x\in J_n} |\mu_n(x, \tilde{c}_n)|^2 \gg \log n\,(\tilde{c}_n/n)^{1/2}$ for any $\alpha < \frac{1}{2}$ and hence $\sup_{x\in J_n} |\mu_n(x, \hat{c}_n)|^2 \gg \log n\,(\hat{c}_n/n)^{1/2}$ with probability tending to one. This suggests that the credible bands defined in (3.4) will not have favourable coverage for any $\alpha < \frac{1}{2}$.

Since $f \in C^\alpha$, we have $|\sum_{j=n+1}^\infty f_j e_j(t)| \lesssim n^{-\alpha}\log n$ by Theorem 10.8 of Chapter 2 in [Zygmund, 2002] and it can be shown that for $\alpha > \frac{1}{2}$ we have $D_{1,n}(c,f) \asymp n(cn)^{-1/2}$ on $I_n$. From this it follows that the credible sets in (3.4) have asymptotic coverage one if $\alpha > \frac{1}{2}$, but the diameter is suboptimal; it is of the order $\sqrt{\ell_n \log n}\, n^{-1/4} \gg (n/\log n)^{-\frac{\alpha}{1+2\alpha}}$.

We now prove that the Fourier coefficients of $f$ are of the order $j^{-1}$. Consider

$$f_{j+1} = \sqrt{2}\int_0^1 |x-t|^\alpha \sin(\pi r t)\,\mathrm{d}t = \frac{\sqrt{2}}{r}\int_0^r \left|x - \frac{s}{r}\right|^\alpha \sin(\pi s)\,\mathrm{d}s,$$

where $r = j + \frac{1}{2}$. We may write

$$\int_0^r \left|x - \frac{s}{r}\right|^\alpha \sin(\pi s)\,\mathrm{d}s = \sum_{k=0}^{2j} \int_{\frac{1}{2}k}^{\frac{1}{2}(k+1)} \left|x - \frac{s}{r}\right|^\alpha \sin(\pi s)\,\mathrm{d}s =: \sum_{k=0}^{2j} I_k.$$

We prove that for $r$ sufficiently large this sum is bounded from above and below by a positive constant. Note that if $k \bmod 4$ is either 0 or 1, $I_k$ is positive, otherwise it is negative. In the first case and for $k+1 < 2xr$, we have

$$\left(x - \frac{k+1}{2r}\right)^\alpha \cdot \frac{1}{\pi} \le I_k \le \left(x - \frac{k}{2r}\right)^\alpha \cdot \frac{1}{\pi}.$$

In the second case and for $k+1 < 2xr$, we have

$$-\left(x - \frac{k}{2r}\right)^\alpha \cdot \frac{1}{\pi} \le I_k \le -\left(x - \frac{k+1}{2r}\right)^\alpha \cdot \frac{1}{\pi}.$$

Similarly, in the first case and for $k > 2xr$, we have

$$\left(\frac{k}{2r} - x\right)^{\alpha} \cdot \frac{1}{\pi} \leq I_k \leq \left(\frac{k+1}{2r} - x\right)^{\alpha} \cdot \frac{1}{\pi}.$$

Finally, in the second case and for $k > 2xr$, we have

$$-\left(\frac{k+1}{2r} - x\right)^{\alpha} \cdot \frac{1}{\pi} \leq I_k \leq -\left(\frac{k}{2r} - x\right)^{\alpha} \cdot \frac{1}{\pi}.$$

Setting $k_0 := \lfloor 2xr \rfloor$, $m_0 = k_0 - 5 - (k_0 \bmod 4)$ and $m_1 = k_0 + 5 - (k_0 \bmod 4) + 2(j \bmod 2)$, we may write

$$\sum_{k=0}^{2j} I_k = \sum_{k=0}^{m_0} I_k + \sum_{k=m_1}^{2j} I_k + o(1).$$

Note that

$$\sum_{k=0}^{m_0} I_k = \sum_{k=0}^{(m_0-3)/4} \big[(I_{4k} + I_{4k+2}) + (I_{4k+1} + I_{4k+3})\big].$$

Set $\tilde{m}_0 = (m_0 - 3)/4$. Applying the mean value theorem, we see that for $b > a$ we have

$$\begin{aligned}
&\sum_{k=0}^{\tilde{m}_0} \left(\left(x - \frac{a+4k}{2r}\right)^{\alpha} - \left(x - \frac{b+4k}{2r}\right)^{\alpha}\right) \\
&= \alpha \sum_{k=0}^{\tilde{m}_0} (\xi_k)^{\alpha-1} \, \frac{b-a}{2r} \\
&\leq \frac{\alpha(b-a)}{2r} \sum_{k=0}^{\tilde{m}_0} \left(x - \frac{b+4k}{2r}\right)^{\alpha-1} \\
&\leq \frac{\alpha(b-a)}{2r} \int_0^{\tilde{m}_0+1} \left(x - \frac{b+4k}{2r}\right)^{\alpha-1} \, \mathrm{d}k \\
&= -\frac{\alpha(b-a)}{4\alpha} \left[\left(x - \frac{b+4(\tilde{m}_0+1)}{2r}\right)^{\alpha} - \left(x - \frac{b}{2r}\right)^{\alpha}\right] \to \frac{(b-a)}{4} x^{\alpha}.
\end{aligned}$$

Here we use $\xi_j \geq x - \frac{b+4k}{2r}$, but we can apply the same argument with $\xi_j \leq x - \frac{a+4k}{2r}$ to obtain a lower bound (changing the upper limit of the integral to $\tilde{m}_0$), which is asymptotically the same as the upper bound. Using this, we can bound $\sum_{k=0}^{\tilde{m}_0} I_k$ from above by

$$\begin{aligned}
&\frac{1}{\pi}\sum_{k=0}^{\tilde{m}_0}\left(x-\frac{4k}{2r}\right)^{\alpha}-\left(x-\frac{3+4k}{2r}\right)^{\alpha}+\left(x-\frac{1+4k}{2r}\right)^{\alpha}-\left(x-\frac{4+4k}{2r}\right)^{\alpha}\\
&\le -\frac{3}{4\pi}\left[\left(x-\frac{7+4\tilde{m}_0}{2r}\right)^{\alpha}-\left(x-\frac{3}{2r}\right)^{\alpha}+\left(x-\frac{8+4\tilde{m}_0}{2r}\right)^{\alpha}-\left(x-\frac{4}{2r}\right)^{\alpha}\right]\\
&\to \frac{3}{2\pi}x^{\alpha}
\end{aligned}$$

and from below by

$$\begin{aligned}
&\frac{1}{\pi}\sum_{k=0}^{\tilde{m}_0}\left(x-\frac{1+4k}{2r}\right)^{\alpha}-\left(x-\frac{2+4k}{2r}\right)^{\alpha}+\left(x-\frac{2+4k}{2r}\right)^{\alpha}-\left(x-\frac{3+4k}{2r}\right)^{\alpha}\\
&\ge -\frac{1}{4\pi}\left[\left(x-\frac{1+4\tilde{m}_0}{2r}\right)^{\alpha}-\left(x-\frac{1}{2r}\right)^{\alpha}+\left(x-\frac{2+4\tilde{m}_0}{2r}\right)^{\alpha}-\left(x-\frac{2}{2r}\right)^{\alpha}\right]\\
&\to \frac{1}{2\pi}x^{\alpha}.
\end{aligned}$$

For the other sum we have

$$\sum_{k=m_1}^{2j} I_k=\sum_{k=0}^{\tilde{m}_1}\left[(I_{2j-4k}+I_{2j-4k-2})+(I_{2j-4k-1}+I_{2j-4k-3})\right],$$

where $\tilde{m}_1=(2j-3-m_1)/4$. Applying the mean value theorem again, we see that for $d>c$ we have

$$\begin{aligned}
&\sum_{k=0}^{m}\left(\left(\frac{2j-4k-c}{2r}-x\right)^{\alpha}-\left(\frac{2j-4k-d}{2r}-x\right)^{\alpha}\right)\\
&\qquad=\alpha\sum_{k=0}^{m}(\xi_k)^{\alpha-1}\,\frac{d-c}{2r}\\
&\qquad\le\frac{\alpha(d-c)}{2r}\sum_{k=0}^{m}\left(\frac{2j-4k-d}{2r}-x\right)^{\alpha-1}\\
&\qquad\le\frac{\alpha(d-c)}{2r}\int_0^{m+1}\left(\frac{2j-4k-d}{2r}-x\right)^{\alpha-1}\,\mathrm{d}k\\
&\qquad=-\frac{(d-c)}{4}\left[\left(\frac{2j-4(m+1)-d}{2r}-x\right)^{\alpha}-\left(\frac{2j-d}{2r}-x\right)^{\alpha}\right]\\
&\qquad\to\frac{(d-c)}{4}(1-x)^{\alpha}.
\end{aligned}$$

Here we use $\xi_j \geq \frac{2j-4k-d}{2r} - x$, but applying the same argument with $\xi_j \leq \frac{2j-4k-c}{2r} - x$ we obtain a lower bound (changing the upper limit of the integral to $\tilde{m}_1$), which is asymptotically the same as the upper bound.

Proceeding as above and treating the cases $j$ even and $j$ odd separately, we obtain a lower bound for $\sum_{k=m_1}^{2j} I_k$ that converges to $-\frac{1}{2\pi}(1-x)^\alpha$ and an upper bound that converges to $\frac{1}{2\pi}(1-x)^\alpha$. Since $x > \frac{1}{2}$ we have $x^\alpha > (1-x)^\alpha$ and the result follows.

By further subdividing each of the $2j$ parts, we can obtain better upper and lower bounds for the integral. We believe that it is possible to obtain a more accurate result this way and that the Fourier coefficients can be shown to satisfy $f_j \sim x^\alpha/(2\pi)$.

## 3.3 Coverage for Gaussian-generated functions

Since the combined Hölder and self-similar conditions appear to be restrictive, in this section we discuss another class of functions for which we can obtain honest coverage. In the previous section we have seen that we obtain coverage for functions that satisfy the good bias condition, for which $\tilde{c}_n = \tilde{c}_n(f) \in I_n$ and the bias satisfies

$$\sup_{x \in J_n} |\mu_n(x, \hat{c}_n)| \lesssim \sqrt{\log n}\,(\hat{c}_n/n)^{1/4}$$

with probability tending to one. For a grid point $x_{j,n}$ we have the representation $f(x_{j,n}) = \sum_{i=1}^n f_{i,n}(e_{i,n})_j = \frac{1}{\sqrt{n_+}} \sum_{i=1}^n f_{i,n} e_i(x_{j,n})$. Writing $D_n = \operatorname{diag}\big((1 + c\lambda_{j,n})^{-1}\big)_j$ and denoting the orthogonal matrix with rows the eigenvectors $e_{j,n}$ by $O_n$, we see that we can write the bias in a grid point in terms of the coefficients $(f_{i,n})$ as

$$\begin{aligned}\mu_n(x_{j,n}, c) &= \mathrm{E}_f \hat{f}_{n,c}(x_{j,n}) - f(x_{j,n}) = (O_n^T D_n O_n \vec{f}_n)_j - f(x_{j,n}) \\ &= -\frac{1}{\sqrt{n_+}} \sum_{i=1}^n \frac{f_{i,n} e_i(x_{j,n})}{1 + c\lambda_{i,n}}.\end{aligned}$$

Hence we have nonnegligible coverage for all design points if

$$\sup_{j \in \{1,\dots,n\}} \frac{1}{n} \left( \sum_{i=1}^n \frac{f_{i,n} e_i(x_{j,n})}{1 + \hat{c}_n \lambda_{i,n}} \right)^2 \lesssim \log n \sqrt{\frac{\hat{c}_n}{n}}$$

with probability tending to one. Note that this is not enough to obtain a credible band in the supremum norm. Instead, as stated in the introduction, we consider credible sets of the form

$$\hat{C}_{n,\eta} = \left\{ f : \max_{i\in\{1,\ldots,n\}} |f(x_{i,n}) - \hat{f}_{n,\hat{c}_n}(x_{i,n})| < L\sqrt{\ell_n \log n} \left(\frac{\hat{c}_n}{n}\right)^{1/4} \right\},$$

where $\ell_n$ is a sequence of numbers tending to infinity and $L$ is a sufficiently large constant. Since the number of design points increases with $n$, this gives an increasingly accurate approximation of a credible band.

We make the above statements precise in the following proposition. Let $\mathcal{F}_{N,\ell_n}$ be the set of all functions $f$ that satisfy the good bias condition and $\tilde{c}_n \in I_n$ for $n \geq N$, for which the bias of the posterior mean at the design points satisfies

$$|\mu_n(x_{j,n}, \tilde{c}_n)|^2 = \frac{1}{n_+} \left( \sum_{i=1}^{n} \frac{f_{i,n} e_i(x_{j,n})}{1 + \tilde{c}_n \lambda_{i,n}} \right)^2 < \ell_n \log n \sqrt{\frac{\tilde{c}_n}{n}} \tag{3.7}$$

for $n \geq N$, uniformly for $j \in \{1,\ldots,n\}$. Here $\ell_n$ is a sequence of numbers tending to infinity. Let $\mathcal{F}_{\ell_n} = \bigcup_N \mathcal{F}_{N,\ell_n}$. Moreover, for $\beta > 0$ define the norms

$$\|f\|_{n,\beta}^2 = \frac{1}{n} \sum_{j=1}^{n} j^{2\beta} f_{j,n}^2,$$

$$\|f\|_{n,\beta,\infty}^2 = \frac{1}{n} \sup_{1\leq j\leq n} j^{1+2\beta} f_{j,n}^2.$$

We now have

**Proposition 3.8.** *For both the risk-based and likelihood-based empirical Bayes methods the credible sets given in* (3.5) *satisfy*

$$\inf_{f\in\mathcal{F}_{\ell_n}} P_f(f \in \hat{C}_{n,\eta}) \to 1.$$

*Furthermore, for $\beta < 1$ the diameter of the credible set $\hat{C}_{n,\eta}$ is of the order $O_P\big(\sqrt{\ell_n \log n}\, n^{-\frac{\beta}{1+2\beta}}\big)$ uniformly in $f$ with $\|f\|_{n,\beta} \lesssim 1$ or $\|f\|_{n,\beta,\infty} \lesssim 1$. For the risk-based empirical Bayes method this is even true for $\beta < 2$.*

*Proof.* We obtain coverage if

$$\frac{\max_{i\in\{1,\ldots,n\}} |T_n(x_{i,n}, \hat{c}_n)| + \max_{i\in\{1,\ldots,n\}} |\mu_n(x_{i,n}, \hat{c}_n)|}{L\sqrt{\ell_n \log n}\,(\hat{c}_n/n)^{1/4}} < 1 \tag{3.8}$$

with probability tending to one. The first term tends to zero by Proposition 3.6. Note that if assumption (3.7) is satisfied, we see that

$$|\mu_n(x_{j,n},c)|^2 = \frac{1}{n_+}\left(\sum_{i=1}^{n}\frac{f_{i,n}e_i(x_{j,n})}{1+c\lambda_{i,n}}\right)^2 \lesssim \ell_n \log n\sqrt{\frac{c}{n}}$$

is satisfied uniformly for $c \in [k\tilde{c}_n, K\tilde{c}_n]$ for certain constants $K > k > 0$. The second term can then be made arbitrarily small with probability tending to one by Lemma 3.5, by taking $L$ to be sufficiently large. The assertions on the diameter of the credible sets follow from Corollary 2.24 and examples 2.19, 2.20, 2.22 and 2.23 in Chapter 2. □

The conditions of this proposition seem hard to interpret, but Theorem 3.3 shows that there is a large class of functions for which the coverage is fine, in particular those generated by the prior. Recall that the Karhunen-Loève expansion of standard Brownian motion $W = (W_t, t \in [0,1])$ is given by

$$W_t = \sum_{j=1}^{\infty}\frac{Z_j}{(j-1/2)\pi}e_j(t),$$

where $Z_1, Z_2, \ldots$ are independent standard normal random variables. The theorem shows that we obtain coverage for $W$ almost surely. In fact we even get this result for any Gaussian series with polynomially decaying singular values relative to the eigenbasis of Brownian motion.

We may now prove this last main result.

*Proof of Theorem 3.3.* We give the proof for the risk-based empirical Bayes method. The proof for the likelihood-based method is analogous. By Proposition 3.8 it suffices to show that $W$ is in $\mathcal{F}_{\ell_n}$ almost surely. Denote the coordinates of $\vec{W}_n$ relative to the eigenbasis of $U_n$ by $W_{i,n}$. The $W_{i,n}$ are independent normal random variables with zero mean and variance satisfying $\mathrm{var}(W_{i,n}) \asymp ni^{-1-2\alpha}$, see the proof of Proposition 2.36 in Chapter 2. In the same proposition it was shown that almost every realisation of $W$ is discrete polished tail and hence satisfies the good bias condition. We now prove that almost surely there is an $N$ such that $\tilde{c}_n(W) \in I_n$ for $n \geq N$. Consider

$$\begin{aligned} V_n(c) &:= \frac{1}{n(cn)^{-\alpha}}\left(D^R_{1,n}(c,W) - \mathrm{E}D^R_{1,n}(c,W)\right) \\ &= \frac{1}{n(cn)^{-\alpha}}\sum_{i=1}^{n}\frac{W_{i,n}^2 - \mathrm{E}W_{i,n}^2}{(1+c\lambda_{i,n})^2}. \end{aligned}$$

This random variable has mean zero and variance

$$\begin{aligned}
\mathrm{E}\big[V_n(c)\big]^2 &= n^{-2}(cn)^{2\alpha}\sum_{i=1}^{n}\frac{\mathrm{var}(W_{i,n}^2)}{(1+c\lambda_{i,n})^4} \asymp (cn)^{2\alpha}\sum_{i=1}^{n}\frac{i^{6-4\alpha}}{(i^2+cn)^4}\\
&= (cn)^{2\alpha}\sum_{i=1}^{\sqrt{cn}}\frac{i^{6-4\alpha}}{(i^2+cn)^4} + (cn)^{2\alpha}\sum_{i=\sqrt{cn}+1}^{n}\frac{i^{6-4\alpha}}{(i^2+cn)^4}\\
&\asymp (cn)^{2\alpha-4}\sum_{i=1}^{\sqrt{cn}} i^{6-4\alpha} + \sqrt{\frac{1}{cn}}\\
&\asymp \begin{cases} (cn)^{-1/2} & \text{if } \alpha < 7/4\\ (cn)^{-1/2}\log(cn) & \text{if } \alpha = 7/4\\ (cn)^{2\alpha-4} & \text{if } \alpha > 7/4.\end{cases}
\end{aligned}$$

Since $\mathrm{E}D_{1,n}^R(c,W) \asymp n(cn)^{-\alpha}$, there are constants $0<\gamma_1<C_1$ such that

$$n(cn)^{-\alpha}\big(\gamma_1+V_n(c)\big) \;\le\; D_{1,n}^R(c,W) \;\le\; n(cn)^{-\alpha}\big(C_1+V_n(c)\big)$$

for $c\in I_n$. Also recall that there are constants $0<\gamma_2<C_2$ such that

$$\gamma_2\sqrt{cn} \le D_{2,n}^R(c) \le C_2\sqrt{cn}.$$

Set $\beta := \frac{1-2\alpha}{1+2\alpha}$ and consider the event $E_n = \big\{\tilde{c}_n(W)\notin\big[\varepsilon n^\beta, Mn^\beta\big]\big\}$. We have

$$\begin{aligned}
P\big(\tilde{c}_n(W) < \varepsilon n^\beta\big) &\le P\left(D_{1,n}^R\big(\varepsilon n^\beta, W\big) < D_{2,n}^R\big(\varepsilon n^\beta\big)\right)\\
&\le P\left(\varepsilon^{-\alpha}n^{-\frac{2\alpha}{1+2\alpha}}\left(\gamma_1+V_n\big(\varepsilon n^\beta\big)\right) < C_2\sqrt{\varepsilon}n^{-\frac{2\alpha}{1+2\alpha}}\right)\\
&= P\left(V_n\big(\varepsilon n^\beta\big) < -a\right),
\end{aligned}$$

where $a := \gamma_1 - C_2\,\varepsilon^{1/2+\alpha} > 0$ for $\varepsilon$ sufficiently small. Denote by $\|\cdot\|_{\psi_1}$ the Orlicz norm corresponding to the function $\psi_1(x) = e^x - 1$. Applying Markov's inequality, we have

$$\begin{aligned}
P\left(V_n\big(\varepsilon n^\beta\big) < -a\right) &\le P\left(\psi_1\left(\frac{|V_n\big(\varepsilon n^\beta\big)|}{\|V_n\big(\varepsilon n^\beta\big)\|_{\psi_1}}\right) \ge \psi_1\left(\frac{a}{\|V_n\big(\varepsilon n^\beta\big)\|_{\psi_1}}\right)\right)\\
&\le \frac{1}{\psi_1\left(a/\|V_n\big(\varepsilon n^\beta\big)\|_{\psi_1}\right)}.
\end{aligned}$$

In order to bound $\|V_n(c)\|_{\psi_1}$, we apply Proposition A.1.6 in [van der Vaart and Wellner, 1996] with $X_i = (W_{i,n}^2 - \mathrm{E}W_{i,n}^2)/(n(cn)^{-\alpha}(1+c\lambda_{i,n})^2)$, $S_n = V_n(c)$

and $p = 1$. Combining this with Lemma 2.2.2 from the same book, we see that

$$\begin{aligned}\|V_n(c)\|_{\psi_1} &\lesssim \|V_n(c)\|_1 + \Big\| \max_{i\in\{1,\dots,n\}} |X_i| \Big\|_{\psi_1} \\ &\lesssim \sqrt{\mathrm{E}\big[V_n(c)\big]^2} + \frac{\log n}{n(cn)^{-\alpha}} \max_{i\in\{1,\dots,n\}} \frac{\|W_{i,n}^2 - \mathrm{E}W_{i,n}^2\|_{\psi_1}}{(1+c\lambda_{i,n})^2}.\end{aligned}$$

Since $W_{i,n}$ is a mean zero normal random variable with variance of the order $ni^{-1-2\alpha}$, it follows by Lemma 2.2.1 from [van der Vaart and Wellner, 1996] that

$$\|W_{i,n}^2 - \mathrm{E}W_{i,n}^2\|_{\psi_1} \leq 2\|W_{i,n}^2\|_{\psi_1} \lesssim ni^{-1-2\alpha}.$$

We see that

$$\begin{aligned}\max_{i\in\{1,\dots,n\}} \frac{\|W_{i,n}^2 - \mathrm{E}W_{i,n}^2\|_{\psi_1}}{(1+c\lambda_{i,n})^2} &\lesssim n \max_{i\in\{1,\dots,n\}} \frac{i^{3-2\alpha}}{(i^2+cn)^2} \\ &\lesssim \begin{cases} n(cn)^{-1/2-\alpha} & \text{if } \alpha \leq 3/2 \\ n(cn)^{-2} & \text{if } \alpha > 3/2. \end{cases}\end{aligned}$$

Combining the above results, it follows that there is a constant $C > 0$ such that

$$\|V_n(\varepsilon n^\beta)\|_{\psi_1} \leq \begin{cases} Cn^{-\frac{1}{4}(1+\beta)} & \text{if } \alpha < 7/4 \\ Cn^{-(2-\alpha)(1+\beta)} \log n & \text{if } \alpha \geq 7/4 \end{cases}$$

and hence

$$P\left(V_n(\varepsilon n^\beta) < -a\right) \leq \begin{cases} \left(e^{\frac{a}{C} n^{\frac{1}{4}(1+\beta)}} - 1\right)^{-1} & \text{if } \alpha < 7/4 \\ \left(e^{\frac{a}{C \log n} n^{(2-\alpha)(1+\beta)}} - 1\right)^{-1} & \text{if } \alpha \geq 7/4. \end{cases}$$

We conclude that there are constants $K, k, \gamma > 0$ such that for $\alpha \in (0,2)$ it holds that $P\left(V_n(\varepsilon n^\beta) < -a\right) \leq Ke^{-kn^\gamma}$. The probability $P\big(\tilde{c}_n(W) > Mn^\beta\big)$ can be treated similarly. We see that

$$\begin{aligned}\sum_{n=1}^\infty P(E_n) &= \sum_{n=1}^\infty \left(P\big(\tilde{c}_n(W) < \varepsilon n^\beta\big) + P\big(\tilde{c}_n(W) > Mn^\beta\big)\right) \\ &\lesssim \sum_{n=1}^\infty e^{-kn^\gamma} < \infty.\end{aligned}$$

It follows by the Borel-Cantelli lemma that almost surely there is an $N$ such that $\tilde{c}_n(W) \in \left[\varepsilon n^\beta, Mn^\beta\right]$ and hence $\tilde{c}_n(W) \in I_n$ for $n \geq N$.

Finally, we prove that condition (3.7) is satisfied almost surely. Consider the normal random variable

$$U_n(x,c) := \frac{1}{\sqrt{n_+}} \sum_{i=1}^{n} \frac{W_{i,n} e_i(x)}{1 + c\lambda_{i,n}},$$

which has mean zero and variance

$$\operatorname{var}\big[U_n(x,c)\big] \lesssim \sum_{i=1}^{n} \frac{i^{3-2\alpha}}{(i^2+cn)^2} \lesssim (cn)^{-\alpha}$$

uniformly for $x \in (0,1)$ and $c \in I_n$. For a design point $x = x_{j,n}$ this random variable is exactly the bias at the point $x$ if the "true" function is $W$. For $s < t \in I_n$ we have

$$\begin{aligned}
\operatorname{var}\left[U_n(x,s) - U_n(x,t)\right] &= \frac{1}{n_+} \sum_{i=1}^{n} \operatorname{var}(W_{i,n}) e_i(x)^2 \left[\frac{1}{1+s\lambda_{i,n}} - \frac{1}{1+t\lambda_{i,n}}\right]^2 \\
&\lesssim \frac{(s-t)^2}{s^2} \sum_{i=1}^{n} i^{-1-2\alpha} \left[\frac{s\lambda_{i,n}}{(1+s\lambda_{i,n})^2}\right]^2 \\
&\lesssim \frac{(s-t)^2}{s^2} \sum_{i=1}^{n} \frac{i^{3-2\alpha}}{(i^2+sn)^2} \lesssim \frac{(s-t)^2}{s^2} (sn)^{-\alpha}.
\end{aligned}$$

Denote by $\|\cdot\|_{\psi_2}$ the Orlicz norm corresponding to the function $\psi_2(x) = e^{x^2} - 1$. Since $U_n(x,s) - U_n(x,t)$ has a normal distribution with mean zero, we then see that for $s < t \in \left[\varepsilon n^\beta, M n^\beta\right]$

$$\|U_n(x,s) - U_n(x,t)\|_{\psi_2}^2 \lesssim \operatorname{var}\left[U_n(x,s) - U_n(x,t)\right] \lesssim n^{-2\beta} n^{-\alpha(1+\beta)} (s-t)^2$$

uniformly for $x \in (0,1)$. It then follows by Corollary 2.2.5 in [van der Vaart and Wellner, 1996] with $T = \left[\varepsilon n^\beta, M n^\beta\right]$ that

$$\begin{aligned}
\left\| \sup_{s,t \in [\varepsilon n^\beta, M n^\beta]} |U_n(x,s) - U_n(x,t)| \right\|_{\psi_2} &\lesssim n^{-\beta - \frac{\alpha}{2}(1+\beta)} \int_0^{n^\beta} \sqrt{\log\left(1 + \frac{n^\beta}{\varepsilon}\right)}\, \mathrm{d}\varepsilon \\
&\le n^{-\beta - \frac{\alpha}{1+2\alpha}} \int_0^{n^\beta} \sqrt{\frac{n^\beta}{\varepsilon}}\, \mathrm{d}\varepsilon = n^{-\frac{\alpha}{1+2\alpha}}.
\end{aligned}$$

Applying Lemma 2.2.2 from [van der Vaart and Wellner, 1996] and noting that $\|U_n(x,c)\|_{\psi_2} \lesssim \sqrt{\operatorname{var} U_n(x,c)} \lesssim n^{-\frac{\alpha}{1+2\alpha}}$ for any fixed $c \in \left[\varepsilon n^\beta, M n^\beta\right]$, we see that

$$\left\| \max_{j\in\{1,\dots,n\}} \sup_{c\in[\varepsilon n^\beta, Mn^\beta]} U_n(x_{j,n},c) \right\|_{\psi_2}$$
$$\le C\sqrt{\log n} \max_{j\in\{1,\dots,n\}} \left\| \sup_{c\in[\varepsilon n^\beta, Mn^\beta]} U_n(x_{j,n},c) \right\|_{\psi_2} \le \tilde{C}\sqrt{\log n}\cdot n^{-\frac{\alpha}{1+2\alpha}}$$

for some constants $C, \tilde{C} > 0$. It follows that

$$\mathrm{E}\left[\exp\left(\frac{\max_{j\in\{1,\dots,n\}} \sup_{c\in[\varepsilon n^\beta, Mn^\beta]} U_n^2(x_{j,n},c)}{\tilde{C}^2 \log n \cdot n^{-\frac{2\alpha}{1+2\alpha}}}\right)\right] \le 2.$$

Using this, we see that

$$P\left(\max_{j\in\{1,\dots,n\}} U_n^2\big(x_{j,n},\tilde{c}_n\big) > \ell_n \log n \sqrt{\frac{\tilde{c}_n}{n}}\right)$$
$$\le P\left(\max_{j\in\{1,\dots,n\}} \sup_{c\in[\varepsilon n^\beta, Mn^\beta]} U_n^2\big(x_{j,n},c\big) > \ell_n \log n \sqrt{\varepsilon} n^{-\frac{2\alpha}{1+2\alpha}}\right)$$
$$= P\left(\exp\left(\frac{\max_{j\in\{1,\dots,n\}} \sup_{c\in[\varepsilon n^\beta, Mn^\beta]} U_n^2\big(x_{j,n},c\big)}{\tilde{C}^2 n^{-\frac{2\alpha}{1+2\alpha}} \sqrt{\varepsilon} \log n}\right) > e^{\ell_n/\tilde{C}^2}\right)$$
$$\le \frac{1}{e^{\ell_n/\tilde{C}^2}} \mathrm{E}\left[\exp\left(\frac{\max_{j\in\{1,\dots,n\}} \sup_{c\in[\varepsilon n^\beta, Mn^\beta]} U_n^2\big(x_{j,n},c\big)}{\tilde{C}^2 n^{-\frac{2\alpha}{1+2\alpha}} \sqrt{\varepsilon} \log n}\right)\right]$$
$$\lesssim \frac{1}{e^{\ell_n/\tilde{C}^2}}.$$

Hence the result follows by the Borel-Cantelli lemma if $\ell_n \ge \gamma\tilde{C}^2 \log n$ for some $\gamma > 1$. □

## 3.4 Discussion

The previous two sections give conditions under which empirical Bayes works, when using the estimator $\hat{c}_n$ based on the global risk or likelihood introduced in Chapter 2. However, these conditions appear either restrictive or intractable. An alternative might be to estimate the scaling parameter with a different method that might be more natural for the current purpose. A criterion based on a “local” risk might result in more natural conditions on the function $f$ for which we obtain coverage and for which the size of the credible sets adapts to the unknown smoothness of the function.

## 3.5 Technical proofs

In this section we give the technical details of the proofs of Propositions 3.1 and 3.6.

*Proof of Proposition 3.1.* Note that the denominator of $b_1^n$ given in (1.6) in Chapter 1 is independent of $i_n = j_n - 1$:

$$D\tilde{D} - (\alpha\lambda_+^{i_n} + \beta\lambda_-^{i_n})(\tilde{\alpha}\lambda_+^{n-i_n} + \tilde{\beta}\lambda_-^{n-i_n}) = \alpha\tilde{\alpha}\lambda_+^n\big(\lambda_+^2 - 1\big) + \beta\tilde{\beta}\lambda_-^n(\lambda_-^2 - 1)$$
$$\sim \tfrac{1}{2}\lambda_+^n.$$

Writing $\frac{1}{k} = \frac{c}{n_+}$, the numerator is given by $\frac{1}{k}D = \frac{1}{k}\alpha(\lambda_+^{j_n} - \lambda_-^{j_n})$. We can see that $\lambda_+^{\sqrt{k}} \to e$, hence for $j_n \gtrsim \sqrt{k}$ we have $b_1^n \asymp \sqrt{\frac{1}{k}}\lambda_+^{-(n-i_n)}$. The coefficient $a_1^n$ satisfies

$$a_1^n \asymp \frac{2}{k}\lambda_+^{-n}(\tilde{\alpha}\lambda_+^{n-i_n} + \tilde{\beta}\lambda_-^{n-i_n}) \sim \frac{1}{k}(\lambda_+^{-i_n} + \lambda_+^{-2n+i_n}) \sim \frac{1}{k}\lambda_+^{-i_n}$$

for $j_n \gtrsim \sqrt{k} = \sqrt{n_+/c}$. Since $\sqrt{n_+/c} \leq n_+/\sqrt{\log n}$ for any $c \in [\log n/n, n/\log n]$, the result follows. ☐

Note that for $j_n \ll \sqrt{k}$ the factor $(\lambda_+^{j_n} - \lambda_-^{j_n})$ is decreasing in $k$ and hence $b_1^n$ and consequently all the $a_i^n$ are of smaller order.

*Proof of Proposition 3.6.* Since $f$ satisfies the good bias condition and $\tilde{c}_n \in I_n$, it follows by Lemma 3.5 that $P_f(\hat{c}_n \in [k\tilde{c}_n, K\tilde{c}_n]) \to 1$. Denote by $\|\cdot\|_{\psi_2}$ the Orlicz norm corresponding to the function $\psi_2(x) = e^{x^2} - 1$. We prove that

$$\Big\| \sup_{x\in[1/n,1]} \sup_{c\in[k\tilde{c}_n,K\tilde{c}_n]} |T_n(x,c)| \Big\|_{\psi_2} \lesssim \sqrt{\log n}\left(\frac{\tilde{c}_n}{n}\right)^{1/4}.$$

First we apply Lemma 2.2.2 from [van der Vaart and Wellner, 1996] to see that

$$\Big\| \sup_{x\in[1/n,1]} \sup_{c\in[k\tilde{c}_n,K\tilde{c}_n]} |T_n(x,c)| \Big\|_{\psi_2}$$
$$\lesssim \sqrt{\log n} \max_{i\in\{1,\ldots,n\}} \Big\| \sup_{x\in(x_{i,n},x_{i+1,n}]} \sup_{c\in[k\tilde{c}_n,K\tilde{c}_n]} |T_n(x,c)| \Big\|_{\psi_2},$$

where we define $x_{n+1,n} = 1$. Now take $x_1 < x_2 \in (x_{i,n}, x_{i+1,n}]$ and $c_1 < c_2 \in [k\tilde{c}_n, K\tilde{c}_n]$. Since $T_n(x_1,c_1) - T_n(x_2,c_2)$ has a normal distribution with mean zero, we see that

$$\begin{aligned}
&\|T_n(x_1,c_1)-T_n(x_2,c_2)\|_{\psi_2}^2\\
&\lesssim \operatorname{var}\left[T_n(x_1,c_1)-T_n(x_2,c_2)\right]\\
&\lesssim \operatorname{var}\left[T_n(x_1,c_1)-T_n(x_1,c_2)\right]+\operatorname{var}\left[T_n(x_1,c_2)-T_n(x_2,c_2)\right]\\
&=\|\vec{a}_n(x_1,c_1)-\vec{a}_n(x_1,c_2)\|^2+\|\vec{a}_n(x_1,c_2)-\vec{a}_n(x_2,c_2)\|^2.
\end{aligned}$$

For the first term, note that

$$\left(c_1^{-1}I+U\right)\vec{a}_n(x_1,c_1)=\vec{v}_n(x_1)=\left(c_2^{-1}I+U\right)\vec{a}_n(x_1,c_2),$$

hence

$$\begin{aligned}
&\|\vec{a}_n(x_1,c_1)-\vec{a}_n(x_1,c_2)\|^2\\
&=\left\|\left\{\left(c_1^{-1}I+U\right)^{-1}\left(c_2^{-1}I+U\right)-I\right\}\vec{a}_n(x_1,c_2)\right\|^2\\
&\le\left\|\left(c_1^{-1}I+U\right)^{-1}\left(c_2^{-1}I+U\right)-I\right\|^2\|\vec{a}_n(x_1,c_2)\|^2.
\end{aligned}$$

The eigenvalues of the matrix are given by $\frac{c_1-c_2}{c_2(1+c_1\lambda_j)}$. We see that

$$\begin{aligned}
\|\vec{a}_n(x_1,c_1)-\vec{a}_n(x_1,c_2)\|^2&\lesssim\max_j\frac{(c_1-c_2)^2}{c_2^2\left(1+\frac{c_1n}{j^2}\right)^2}t_n^2(x_1,c_2)\\
&\lesssim\frac{(c_1-c_2)^2}{c_2^{3/2}\sqrt{n}}\lesssim\frac{(c_1-c_2)^2}{\tilde{c}_n^{3/2}\sqrt{n}}.
\end{aligned}$$

Here we use the fact that $t_n^2(x,c)\asymp\sqrt{\frac{c}{n}}$ uniformly for $c\in I_n$ and $x\in J_n$ by Lemma 1.8 in Chapter 1. Note that by the remark following the proof of Proposition 3.1, we also have $t_n^2(x,c)\lesssim\sqrt{\frac{c}{n}}$ for $x\in[1/n,1/\sqrt{\log n}]$. In order to bound the second term, we study the characterisation of the coefficients $a_i^n$ in the proof of Theorem 1.3 in Chapter 1. They are given by

$$a_i^n(x,c)=\begin{cases}a_1^n(x,c)\left(\alpha\lambda_+^i+\beta\lambda_-^i\right) & \text{for } i\le i_n(x)\\ b_1^n(x,c)\left(\tilde{\alpha}\lambda_+^{n-i+1}+\tilde{\beta}\lambda_-^{n-i+1}\right) & \text{for } i>i_n(x),\end{cases}$$

where $i_n(x)=\max\{i:i/n_+<x\}$ and $a_1^n(x,c)$ and $b_1^n(x,c)$ are given in (2.5) and the last display of the proof of Theorem 1.3. Note that for $x_1<x_2\in$

$(x_{i,n}, x_{i+1,n}]$ we have $i_n(x_1) = i_n(x_2)$, hence

$$\begin{aligned}
&|b_1^n(x_1,c_2) - b_1^n(x_2,c_2)| \\
&= \left| \frac{c_2 D\left(x_1 - x_2 + \frac{x_2 - x_1}{D}\left(\alpha\lambda_+^{i_n(x_1)} + \beta\lambda_-^{i_n(x_1)}\right)\right)}{D\tilde{D} - \left(\alpha\lambda_+^{i_n(x_1)} + \beta\lambda_-^{i_n(x_1)}\right)\left(\tilde{\alpha}\lambda_+^{n-i_n(x_1)} + \tilde{\beta}\lambda_-^{n-i_n(x_1)}\right)} \right| \\
&= \left| \frac{c_2 D\left(x_1 - x_2 + (x_2 - x_1)\left(1 + O(\sqrt{\frac{c_2}{n}})\right)\right)}{D\tilde{D} - \left(\alpha\lambda_+^{i_n(x_1)} + \beta\lambda_-^{i_n(x_1)}\right)\left(\tilde{\alpha}\lambda_+^{n-i_n(x_1)} + \tilde{\beta}\lambda_-^{n-i_n(x_1)}\right)} \right| \\
&\lesssim (x_2 - x_1)\sqrt{c_2 n}\, b_1^n(x_1, c_2).
\end{aligned}$$

We also have

$$\begin{aligned}
&|a_1^n(x_1,c_2) - a_1^n(x_2,c_2)| \\
&= \left| \frac{c_2(x_2 - x_1) + (b_1^n(x_1,c_2) - b_1^n(x_2,c_2))(\tilde{\alpha}\lambda_+^{n-i_n} + \tilde{\beta}\lambda_-^{n-i_n})}{D} \right| \\
&\lesssim (x_2 - x_1)\frac{c_2 + \sqrt{c_2 n}\, b_1^n(x_1,c_2)(\tilde{\alpha}\lambda_+^{n-i_n} + \tilde{\beta}\lambda_-^{n-i_n})}{D} \\
&\lesssim (x_2 - x_1)\sqrt{c_2 n}\, a_1^n(x_1, c_2).
\end{aligned} \tag{3.9}$$

We then see that

$$\begin{aligned}
\|\vec{a}_n(x_1,c_2) - \vec{a}_n(x_2,c_2)\|^2 &\lesssim (x_1 - x_2)^2 c_2 n\, t_n^2(x_1, c_2) \\
&\lesssim (x_1 - x_2)^2 \tilde{c}_n^{3/2}\sqrt{n}.
\end{aligned}$$

We conclude that

$$n^{1/4}\left\|T_n(x_1,c_1) - T_n(x_2,c_2)\right\|_{\psi_2} \lesssim d((x_1,c_1),(x_2,c_2))$$

for

$$d((x_1,c_1),(x_2,c_2)) = \sqrt{\frac{(c_1 - c_2)^2}{\tilde{c}_n^{3/2}} + n(x_1 - x_2)^2\tilde{c}_n^{3/2}}.$$

For this metric, using the fact that $\frac{\tilde{c}_n}{n} \to 0$, we can bound the diameter of the set $T := (x_{i,n}, x_{i+1,n}] \times [k\tilde{c}_n, K\tilde{c}_n]$ by

$$\operatorname{diam} T \lesssim \sqrt{\frac{\tilde{c}_n^2}{\tilde{c}_n^{3/2}} + n\frac{1}{n^2}\tilde{c}_n^{3/2}} \lesssim \tilde{c}_n^{1/4}$$

and the covering number by

$$N(\varepsilon, d) \lesssim \frac{\frac{1}{n}\sqrt{n}\tilde{c}_n^{3/4}}{\varepsilon}\frac{\tilde{c}_n}{\tilde{c}_n^{3/4}\varepsilon} = \frac{\tilde{c}_n}{\sqrt{n}\varepsilon^2}.$$

Applying Corollary 2.2.5 in [van der Vaart and Wellner, 1996], we obtain

$$\Big\| \sup_{(x_1,c_1),(x_2,c_2)\in T} |T_n(x_1,c_1) - T_n(x_2,c_2)| \Big\|_{\psi_2}$$
$$\lesssim n^{-\frac{1}{4}} \int_0^{\tilde{c}_n^{1/4}} \sqrt{\log\left(1 + \frac{\tilde{c}_n}{\sqrt{n}\varepsilon^2}\right)}\, \mathrm{d}\varepsilon \le n^{-\frac{1}{4}} \int_0^{\tilde{c}_n^{1/4}} \sqrt{\frac{\tilde{c}_n^{1/2}}{n^{1/4}\varepsilon}}\, \mathrm{d}\varepsilon = \left(\frac{\tilde{c}_n}{n}\right)^{\frac{3}{8}}.$$

Combining this with the fact that $\|T_n(x_0,c_0)\|_{\psi_2} \lesssim \sqrt{\operatorname{var} T_n(x_0,c_0)} \lesssim \left(\frac{\tilde{c}_n}{n}\right)^{1/4}$ for any fixed $(x_0,c_0)$ in $T$, we see that

$$\Big\| \sup_{x\in[1/n,1]} \sup_{c\in[k\tilde{c}_n, K\tilde{c}_n]} |T_n(x,c)| \Big\|_{\psi_2} \lesssim \sqrt{\log n}\left(\frac{\tilde{c}_n}{n}\right)^{\frac{1}{4}}.$$

The result follows. □