Adaptive Bayesian credible bands in regression with a Gaussian process prior

Suzanne Sniekers and Aad van der Vaart
Leiden University, Leiden, The Netherlands

Abstract

A credible band is the set of all functions between a lower and an upper bound that are constructed so that the set has prescribed mass under the posterior distribution. In a Bayesian analysis such a band is used to quantify the remaining uncertainty on the unknown function in a similar manner as a confidence band. We investigate the validity of a credible band in the nonparametric regression model with the prior distribution on the function given by a Gaussian process. We show that there are many true regression functions for which the credible band has the correct order of magnitude to be used as a confidence set. We also exhibit functions for which the credible band is misleading.

AMS (2000) subject classification. 62G15, 62G05, 62G20.

Keywords and phrases. Credible band, Coverage, Uncertainty quantification, Nonparametric Bayes

1 Introduction and main results

Suppose that we observe a vector $\mathbf{Y}_n := (Y_{1,n}, \dots, Y_{n,n})^T$ with coordinates distributed according to

$$Y_{i,n} = f(x_{i,n}) + \varepsilon_{i,n}, \quad i \in \{1, \dots, n\}. \quad (1.1)$$

Here the parameter is a function $f : [0, 1] \rightarrow \mathbb{R}$, 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. In this paper we investigate a nonparametric Bayesian method to estimate the regression function f , based on a Gaussian process prior. We are interested in the usefulness of the resulting posterior distribution for quantifying the remaining uncertainty about the function. More precisely, the posterior distribution allows

Research supported by the Netherlands Organization for Scientific Research (NWO) and the European Research Council under ERC Grant Agreement 320637.

to construct a *credible band*: a ball for the (weighted) uniform norm around the posterior mean of prescribed posterior probability. We investigate to what extent such a band has similar properties as a (frequentist) confidence band.

The performance of a credible band depends strongly on the combination of prior and true regression function. Furthermore, for priors that “adapt” to the true function through a regularity parameter it depends on the method of adaptation. In this paper we illustrate this with a case study of a special prior, namely scaled Brownian motion. For simplicity we choose the design points $x_{i,n}$ to be equally spaced and equal to $x_{i,n} = i/n_+$, where $n_+ = n + 1/2$. For $W = (W_t, t \in [0, 1])$ a standard Brownian motion we take $\sqrt{c}W$ as a prior for f , where the scaling parameter $c > 0$ will be set by an empirical or hierarchical Bayes method. Our insistence on a band rather than a ball for the L_2 -norm distinguishes this paper from earlier work, such as Szabó et al. (2015a) & Sniekers and van der Vaart (2015a). From the point of view of visualisation bands are preferable for data-analysis.

In the Bayesian setup the observations are distributed according to the model, where W and the errors $\varepsilon_{i,n}$ are independent,

$$Y_{i,n} = \sqrt{c}W_{x_{i,n}} + \varepsilon_{i,n}, \quad i \in \{1, \dots, n\}. \quad (1.2)$$

By definition the posterior distribution of f given $\mathbf{Y}_n$ and c is the conditional distribution of $f = \sqrt{c}W$ given $\mathbf{Y}_n$ and c in this model; we denote it by $\Pi_n(\cdot | \mathbf{Y}_n, c)$. By standard properties of Gaussian distributions this can be seen to be the distribution of a Gaussian process. It has a version with continuous sample paths and we can consider the posterior distribution formally as a Borel law on $C[0, 1]$. We denote its mean and covariance function by

$$\hat{f}_{n,c}(x) = \mathbb{E}(f(x) | \mathbf{Y}_n, c), \quad (1.3)$$

$$\sigma_n(x, y, c) = \text{cov}(f(x), f(y) | \mathbf{Y}_n, c). \quad (1.4)$$

It is natural to center a credible set for the function f at the posterior mean. We shall see that the posterior variance $\sigma_n(x, x, c)$ hardly depends on x , so that equal width intervals for different x are natural, yielding a band if applied simultaneously. We shall consider a *credible band* of the form:

$$C_n(c, L) = \left\{ f : \|f - \hat{f}_{n,c}\|_\infty < Lw_n(c) \right\}, \quad (1.5)$$

where $\|f\|_\infty = \sup_{x \in J_n} |f(x)|$ is the uniform norm over the interval $J_n = [1/l_n, 1 - 1/l_n]$, where $l_n \rightarrow \infty$ is a fixed sequence with $l_n \ll \sqrt{\log n / \log \log n}$,

L is a constant, and $w_n(c)$ is a posterior quantile of the uniform norm of $f - \hat{f}_{n,c}$: for some $\eta \in (0, 1)$,

$$\Pi_n \left(\|f - \hat{f}_{n,c}\|_\infty < w_n(c) \mid \mathbf{Y}_n, c \right) = \eta. \quad (1.6)$$

We restrict to the subinterval $J_n \subset [0, 1]$ to avoid boundary effects of the Brownian prior, and have inserted a constant L in definition Eq. 1.5 in order to make up for a possible discrepancy between a Bayesian credible level and frequentist confidence level. The credible level η will be fixed throughout. For practice we recommend to use the value $L = 1$, since this corresponds to the true Bayesian procedure. It will be apparent from our results that frequentist coverage of a given parameter f is only ensured if L is large enough relative to parameters such as ϵ in Eq. 1.13. There is no way to estimate this from the data, and we can only be assured that the *order* of magnitude of the Bayesian band is also correct in the frequentist sense.

That the distribution of the observations $\mathbf{Y}_n$ depends on f only through its values at the design points $x_{i,n}$ motivates to consider also “discrete bands”, where the argument x is restricted to the design points, of the form

$$C_n^d(c, L) = \left\{ f : \max_{i \in \mathcal{J}_n} |f(x_{i,n}) - \hat{f}_{n,c}(x_{i,n})| < L w_n^d(c) \right\}, \quad (1.7)$$

where $\mathcal{J}_n = \{i : x_{i,n} \in J_n\}$ and $w_n^d(c)$ is determined so that

$$\Pi_n \left(\max_{i \in \mathcal{J}_n} |f(x_{i,n}) - \hat{f}_{n,c}(x_{i,n})| < w_n^d(c) \mid \mathbf{Y}_n, c \right) = \eta. \quad (1.8)$$

As the design points form a grid with mesh width of the order $1/n$, one may expect the bands Eqs. 1.5 and 1.7 not to differ much, but the difference depends both on the prior and the true function f .

It will be shown below that the widths of the bands satisfy

$$w_n(c) \asymp w_n^d(c) \asymp \sqrt{\log(cn)} \left(\frac{c}{n} \right)^{1/4}.$$

Here the logarithmic factor arises because of the uniform norm, while the factor $(c/n)^{1/4}$ gives the order of magnitude of the posterior standard deviation $\sigma_n(x, x, c)^{1/2}$ of $f(x)$. This shows that for fixed c the band will never be narrower than $n^{-1/4}$, which is disappointing if f is smooth. It also suggests that for fixed c the band will not cover if f is too rough, as one cannot expect to estimate a very rough function at nearly $n^{-1/4}$ precision. In practice one

tries to overcome these problems by choosing a suitable value of c from the data. Two standard methods are, for $I_n = [(\log n)/n, n/\log n]$,

$$\hat{c}_n = \operatorname{argmin}_{c \in I_n} [\log \det \Sigma_{n,c} + \mathbf{Y}_n^T \Sigma_{n,c}^{-1} \mathbf{Y}_n] \quad (1.9)$$

and

$$\hat{c}_n = \operatorname{argmin}_{c \in I_n} [\operatorname{tr}((I - \Sigma_{n,c}^{-1})^2) - \operatorname{tr}(\Sigma_{n,c}^{-2}) + \mathbf{Y}_n^T \Sigma_{n,c}^{-2} \mathbf{Y}_n]. \quad (1.10)$$

Here $\Sigma_{n,c} = I + cU_n$, where U_n is the $n \times n$ covariance matrix of standard Brownian motion at the design points: $(U_n)_{i,j} = x_{i,n} \wedge x_{j,n}$. The first method (1.9) is exactly the maximum likelihood estimator of c based on the Bayesian marginal distribution of $\mathbf{Y}_n$: the distribution of the right side of Eq. 1.2, with c viewed as the only unknown parameter. The second method corresponds to minimizing an unbiased estimate of the quadratic risk function of the estimator $(\hat{f}_{n,c}(x_{1,n}), \dots, \hat{f}_{n,c}(x_{n,n}))$, and goes back to the literature on penalized estimation. See Wahba (1983) and Sniekers and van der Vaart (2015a) for further discussion. Given either of these estimators one may construct a credible band by simply substituting $\hat{c}_n$ in Eq. 1.5 or Eq. 1.7.

An alternative to these empirical Bayes methods is the hierarchical Bayes method, which equips c with a (hyper)prior. Following Sniekers and van der Vaart (2015a) we shall take an inverse gamma prior truncated to the interval I_n : the prior density of c satisfies, for some fixed $\kappa, \lambda > 0$,

$$\pi(c) \propto c^{-1-\kappa} e^{-\lambda/c}, \quad c \in I_n. \quad (1.11)$$

This leads to a posterior distribution for c , from which we may extract two nontrivial quantiles $\hat{c}_{1,n}$ and $\hat{c}_{2,n}$ and next use the credible set $\cup_{\hat{c}_{1,n} < c < \hat{c}_{2,n}} C_n(c, L)$ or $\cup_{\hat{c}_{1,n} < c < \hat{c}_{2,n}} C_n^d(c, L)$. It was shown in Sniekers and van der Vaart (2015a) that the hierarchical Bayes method is closely linked to the likelihood-based empirical Bayes method Eq. 1.9 in that the posterior distribution of c will concentrate near the likelihood-based empirical Bayes estimator $\hat{c}_n$.

One can devise methods to adapt the scaling parameter c to the data that are targeted especially to the uniform norm (see Yoo and van der Vaart 2018), but we shall not consider them in this paper. Our interest here is to stick strictly to the Bayesian paradigm, as best embraced by the hierarchical Bayes and likelihood-based empirical Bayes methods. We then ask for which true regression functions f the resulting credible bands work and for which not.

It was shown in Low (1997), Juditsky and Lambert-Lacroix (2003), Cai and Low (2004), Cai and Low (2006), Robins and van der Vaart (2006), &

Hoffmann and Nickl (2011) that nonparametric adaptive confidence sets can only be honest (i.e. possess coverage uniformly in the parameter f) if the true regression function possesses special properties. Bayesian credible sets can of course not beat this fundamental limitation, and hence we need to impose conditions on the true regression function. For deterministic $c = c_n$ it is enough that the prior is not smoother than the true regression function (i.e. c is not too small). This case was discussed in Sniekers and van der Vaart (2015b) (and Knapik et al. (2011)) for credible intervals, but the results generalize to bands. The finding can be understood as a consequence of the bias variance trade-off: a smooth prior will make the band narrow (small variance), but give a large bias on a rough true function; if these are not traded off properly, then coverage fails. In the present paper we consider the more interesting and more complicated case of a data-based choice of c . In this case coverage will hold only if the true f satisfies additional conditions that prevent $\hat{c}_n$, or the location of the posterior distribution of c , to be too small, which would make the bias bigger than the posterior spread. We consider two types of such conditions: first a combination of self-similarity and a Hölder condition, and second functions f characterized as realizations from a prior.

The first type of assumption is in terms of the eigenbasis of Brownian motion, given by

$$e_j(x) = \sqrt{2} \sin \left[\left(j - \frac{1}{2} \right) \pi x \right], \quad j = 1, 2, \dots \quad (1.12)$$

We call a function f *self-similar* of order $\beta > 0$ if its sequence of Fourier coefficients (f_j) with respect to this basis satisfies, for some positive constants M, ρ, ε and every m ,

$$\sup_{j \geq 1} j^{1/2+\beta} |f_j| \leq M, \quad \text{and} \quad \sum_{j=m}^{\rho m} f_j^2 \geq \varepsilon \frac{M^2}{m^{2\beta}}. \quad (1.13)$$

Functions such that $f_j \asymp j^{-1/2-\beta}$ are simple examples. The condition is the same as in Szabo et al. (2015a, b) and similar to conditions introduced in Hoffmann and Nickl (2011), Giné and Nickl (2010), Bull (2012), & Bull and Nickl (2013). It requires that the “total energy” in every sufficiently large block of “frequencies” is at least a fraction of the “total possible energy” in the signal.

Let $\mathcal{F}_{\alpha, \beta, \varepsilon, \rho}$ be the set of all functions that are self-similar of order β for given ε and ρ and some M and possess Hölder norm of order α smaller than M : i.e. $\|f\|_\infty \leq M$ and $|f(x) - f(y)| \leq M|x - y|^\alpha$ if $\alpha \leq 1$ and $|f'(x) - f'(y)| \leq M|x - y|^{\alpha-1}$ if $\alpha \in (1, 2]$.

Theorem 1 (Coverage). *For $C_n(c, L)$ given in Eq. 1.5, let $\hat{C}_n(L)$ be equal to $C_n(\hat{c}_n, L)$ for $\hat{c}_n$ defined by the likelihood-based or risk-based empirical Bayes method Eqs. 1.9 or 1.10. If $2 \geq \alpha \geq \beta > \frac{1}{2}$, then for sufficiently large L ,*

$$\liminf_{n \rightarrow \infty} \inf_{f \in \mathcal{F}_{\alpha, \beta, \varepsilon, \rho}} \Pr \left(f \in \hat{C}_n(L) \right) \geq \eta.$$

The same is true if $\hat{C}_n(L)$ is equal to $\cup_{\hat{c}_{1,n} < c < \hat{c}_{2,n}} C_n(c, L)$, for $\hat{c}_{1,n} < \hat{c}_{2,n}$ satisfying $\Pi_n(\hat{c}_{1,n} < c < \hat{c}_{2,n} | \mathbf{Y}_n) = \eta \in (0, 1)$ given a prior density satisfying Eq. 1.11, and $\beta < 1$.

The proof of the theorem is given in Sections 4 and 6.

The theorem shows that empirical or hierarchical Bayes credible bands cover the true function if the Hölder smoothness α of the function f is at least the order of self-similarity β . As we shall see later in the proof, this is due to the fact that the behaviour of the estimator $\hat{c}_n$ is determined by self-similarity, but the bias for the uniform norm by the Hölder exponent. A value $\beta > \alpha$ would lead to a choice of $\hat{c}_n$ that corresponds to overestimating the Hölder smoothness of the function, which leads to poor coverage.

For nice functions the self-similarity index β is equal to the Hölder smoothness α , but the two indices are not related in general. The self-similarity measures the speed of decrease of the Fourier series, and has an L_2 character, whereas the Hölder smoothness refers to the function in the time domain.

The restriction of the self-similarity constant β to be bigger than $1/2$ is due to the discrete design. It makes it possible to link the infinite sequence (f_j) to the coefficients $(f_{i,n})$ of the vectors $\mathbf{f}_n$ relative to the discretized eigen basis, defined in Eq. 3.2, below. We define a function f to be *discretely self-similar* of order $\beta > 0$ if for some positive constants M, ρ, ε and every $m \leq n$ and n ,

$$\sup_{1 \leq i \leq n} i^{1/2+\beta} \frac{|f_{i,n}|}{\sqrt{n_+}} \leq M, \quad \text{and} \quad \frac{1}{n_+} \sum_{i=m}^{\rho m \wedge n} f_{i,n}^2 \geq \varepsilon \frac{M^2}{m^{2\beta}}.$$

With self-similarity replaced by discrete self-similarity, Theorem 1 is true for $\beta > 0$. For $\beta > 1/2$ self-similarity implies discrete self-similarity (Sniekers and van der Vaart, 2015a).

The second set of functions for which we shall show coverage has a Bayesian flavour. According to the prior the function f is a multiple of a sample path of Brownian motion. Hence by the Karhunen-Loève theorem the function can be expanded as a multiple of the infinite sum $(\sqrt{2}/\pi) \sum_{j=1}^{\infty} Z_j e_j / (j -$

$1/2$), for i.i.d. standard normal variables $Z_1, Z_2, \dots$ and e_j given in Eq. 1.12. If the Bayesian paradigm works at all, the credible band should have the correct order of magnitude for “most of the realisations from the prior”. The following theorem shows that this is indeed true. In fact, coverage pertains for almost every realization of any process of the form, for some $\gamma > 0$, $\alpha > 0$ and $\delta \in \mathbb{R}$,

$$\gamma \sum_{j=1}^{\infty} \frac{Z_j e_j(x)}{(j + \delta)^{1/2+\alpha}}, \quad x \in [0, 1], \quad (1.14)$$

where $Z_1, Z_2, \dots$ are independent standard normal random variables. The Brownian motion prior corresponds to $\alpha = 1/2$ and $\delta = -1/2$ and $\gamma = \sqrt{2c}/\pi$.

Theorem 2 (Coverage). *For $C_n^d(c, L)$ given in Eq. 1.7, let $\hat{C}_n(L)$ be equal to $C_n^d(\hat{c}_n, L)$ for $\hat{c}_n$ defined by the likelihood-based or risk-based empirical Bayes method Eq. 1.9 or Eq. 1.10. Let $\gamma > 0$ and $\delta \in \mathbb{R}$ be given constants. For $\alpha \in (0, 1)$ and the likelihood-based method Eq. 1.9, and for $\alpha \in (0, 2)$ and the risk-based method Eq. 1.10, and for almost every realisation f of the process Eq. 1.14, we have that for sufficiently large L ,*

$$\Pr_f \left(f \in \hat{C}_n(L) \right) \rightarrow 1.$$

The same is true if $\hat{C}_n(L)$ is equal to $\cup_{\hat{c}_{1,n} < c < \hat{c}_{2,n}} C_n^d(c, L)$, for $\hat{c}_{1,n} < \hat{c}_{2,n}$ such that $\Pi_n(\hat{c}_{1,n} < c < \hat{c}_{2,n} | \mathbf{Y}_n) = \eta \in (0, 1)$ given a prior satisfying Eq. 1.11, and $\alpha < 1$.

The proof of the theorem can be found in Sections 5 and 6. It proceeds by showing that the random functions Eq. 1.14 belong with probability one to a deterministic set, for which coverage is guaranteed.

The preceding theorems show that the credible bands cover the true regression function in some generality, and hence justify the use of the posterior distribution as an expression of remaining uncertainty. The following theorem shows that this coverage is not due to these bands being overly wide.

Theorem 3 (Diameter). *For $\beta < 1$ the width of the credible band $\hat{C}_n(L)$ in Theorem 1 is $O_P(\sqrt{\log n} n^{-\beta/(1+2\beta)})$; for the risk-based empirical Bayes method this is true for $\beta < 2$. The same is true for the width of the band in Theorem 2 and every $\beta < \alpha < 2$.*

The proof of this theorem is incorporated in the proofs of Theorems 1 and 2.

If the credible band covers the true function, then its width gives the rate of estimation by the posterior mean for the (discrete) uniform norm. The minimax rate of estimation for functions that are Hölder smooth of order α is known to be of the order $(n/\log n)^{-\alpha/(1+2\alpha)}$ (see Stone 1982). Thus in the reasonable case that the order of self-similarity β is equal to the Hölder smoothness α , the width of the credible bands is close to minimax. However, the width is suboptimal up to a logarithmic factor: the factor $(\log n)^{\alpha/(1+2\alpha)}$ in the minimax rate is replaced by $\sqrt{\log n}$ in Theorem 3. This is caused by the fact that the present methods of choosing c are linked to the empirical L_2 -norm of f rather than the uniform norm. This is immediate for the risk-based empirical based method Eq. 1.10, as it is set up to minimize the L_2 -risk. It is also true for the likelihood-based empirical Bayes method Eq. 1.9 and the hierarchical Bayes method, due to the fact that the likelihood is linked to the L_2 -norm of the values $f(x_{i,n})$. It was shown in Sniekers and van der Vaart (2015a) that these methods do choose an optimal value of c , but from the point of view of L_2 -loss. For a true function that is regular of order β in an appropriate L_2 -sense they choose a value of c that balances squared bias and variance in the form

$$\left(\frac{1}{\hat{c}n}\right)^\beta \asymp \sqrt{\frac{\hat{c}}{n}}.$$

This yields a rate of contraction relative to the L_2 -norm of $(\hat{c}/n)^{1/4} = n^{-\beta/(2\beta+1)}$. (For the Brownian motion prior this is limited to $\beta \leq 2$. See Ghosal et al. (2000), Ghosal and van der Vaart (2007), van der Vaart and van Zanten (2007), Szabo et al. (2013), & Ghosal and van der Vaart (2017) for derivations of the rate in the Bayesian setup.) For the uniform norm the variance term incurs an extra logarithmic factor, and the correct trade-off would be

$$\left(\frac{1}{\hat{c}_\infty n}\right)^\beta \asymp \log n \sqrt{\frac{\hat{c}_\infty}{n}},$$

leading to the minimax rate $(\log n/n)^{\beta/(2\beta+1)}$. However, the Bayesian methods of choosing the scaling in the present paper are not informed about the loss function, and make the trade-off dictated by the likelihood or L_2 -risk. The smaller L_2 -variance term makes that $\hat{c} \asymp n^{(1-2\beta)/(1+2\beta)}$ is a logarithmic factor bigger than $\hat{c}_\infty$, and leads to the suboptimal rate $\sqrt{\log n} n^{-\beta/(2\beta+1)}$. Since the loss relative to the minimax rate $(\log n/n)^{\beta/(2\beta+1)}$ is only a logarithmic factor, this is not too bothersome. One gains a unified methodology.

For the question of coverage that is central to the present paper it is important that the loss occurs in the variance term and not in the bias term. Coverage requires that the bias is not too large relative to the variance: therefore the fact that $\hat{c} \gg \hat{c}_\infty$ helps for coverage. This explains that the Bayesian methods of the present paper, although linked to the L_2 -norm, can still work for uncertainty quantification relative to the uniform norm.

The positive results on the coverage of credible sets evoked in the preceding theorems are surprising in the light of earlier findings on nonparametric credible sets. In particular, in the papers (Cox, 1993; Freedman, 1999; Johnstone, 2010) credible sets are shown to have zero coverage almost surely. This discrepancy is due to considering nonadaptive credible sets for priors that oversmooth the true functions. On the positive side (Wahba, 1983) gave simulations and heuristic arguments that suggested promising results for credible intervals at the design points. The true functions used in these simulations satisfy the conditions imposed in Theorem 1.

The paper is organised as follows. In Section 2 we collect properties of the posterior mean, in particular its bias relative to the uniform distance, and in Section 3 we discuss the behaviour of the empirical Bayes estimators. Next Sections 4 and 5 contain the proofs of the main result in the empirical Bayes case, where in Section 5 a result of independent interest is obtained, and Section 6 gives the proof in the hierarchical case. Section 7 presents a (counterexample) of a true function, and some pictures of bands. For easy reference and completeness a supplement (Sections 8, 10, and 11) contains adaptations and extensions of results from our earlier papers. The proof of a technical lemma related to these results can be found in Section 9, also in the supplement.

Throughout we denote the interval $[\log n/n, n/\log n]$ by I_n and $[1/l_n, 1 - 1/l_n]$ by J_n , where $l_n \rightarrow \infty$ is a fixed sequence with $l_n \ll \sqrt{\log n}/\log \log n$. The symbol $\lesssim$ is used to denote “less than up to a multiplicative constant that is universal or fixed within the context”.

2 Posterior mean, spread and quantiles

The posterior distribution of $f(x)$ for a given x was studied in Sniekers and van der Vaart (2015b). In this section we present versions of a few results from the latter paper that are uniform in x and c , and we characterise posterior quantiles of the uniform norm of f minus its expectation. For easy reference and completeness, proofs that follow the same lines as in Sniekers and van der Vaart (2015b) are given in Section 8 of the supplement.

The posterior mean of $f(x)$ is the conditional expectation $\hat{f}_{n,c}(x) = E(\sqrt{c}W_x | Y_{1,n}, \dots, Y_{n,n})$, and can be written as a linear combination of the observations:

$$\hat{f}_{n,c}(x) = \sum_{i=1}^n a_{i,n}(x, c) Y_{i,n} = \mathbf{Y}_n^T \mathbf{a}_n(x, c).$$

The vector of coefficients $\mathbf{a}_n(x, c) = (a_{i,n}(x, c)) \in \mathbb{R}^n$ is characterised concretely in Proposition 15, in Section 8. The coefficients are essentially a sliding exponential filter of bandwidth $(cn)^{-1/2}$: for $|x - x_{i,n}| \lesssim (cn)^{-1/2}$,

$$a_{i,n}(x, c) \sim \frac{1}{2n} \sqrt{cn} e^{-|x - x_{i,n}| \sqrt{cn}}. \quad (2.1)$$

The covariance function of the posterior mean can be expressed in the coefficients $\mathbf{a}_n$ as:

$$\tau_n(x, y, c) = \text{cov}_f(\hat{f}_{n,c}(x), \hat{f}_{n,c}(y)) = \mathbf{a}_n(x, c)^T \mathbf{a}_n(y, c).$$

Normality and orthogonality imply the independence of the residual $\sqrt{c}W_x - \hat{f}_{n,c}(x)$ and $\mathbf{Y}_n$ in the Bayesian setup. Hence the posterior covariance function as in Eq. 1.4 is equal to the unconditional covariance function of the process $\sqrt{c}W_x - \hat{f}_{n,c}(x) = \sqrt{c}W_x - \sqrt{c}\mathbf{W}_n^T \mathbf{a}_n(x, c) - \varepsilon_n^T \mathbf{a}_n(x, c)$ and can be written

$$\sigma_n(x, y, c) = \text{ccov}(W_x - \mathbf{W}_n^T \mathbf{a}_n(x, c), W_y - \mathbf{W}_n^T \mathbf{a}_n(y, c)) + \mathbf{a}_n(x, c)^T \mathbf{a}_n(y, c).$$

For technical reasons we also introduce a slight adaptation of this function, given by

$$\bar{\sigma}_n(x, y, c) = \text{ccov}(W_x 1^T \mathbf{a}_n(x, c) - \mathbf{W}_n^T \mathbf{a}_n(x, c), W_y 1^T \mathbf{a}_n(y, c) - \mathbf{W}_n^T \mathbf{a}_n(y, c)) + \mathbf{a}_n(x, c)^T \mathbf{a}_n(y, c).$$

The numbers $1^T \mathbf{a}_n(x, c) = \sum_{i=1}^n a_{i,n}(x, c)$ will be seen to be close to 1, so that $\bar{\sigma}_n \approx \sigma_n$.

The following lemma lists the most important properties of these quantities; its proof can be found in Section 9.

Lemma 4. *Fix an arbitrary sequence $\eta_n \rightarrow 0$. The approximation Eq. 2.1 is valid uniformly in i and x and c such that $|x - x_{i,n}| \sqrt{cn} < 1/(2\eta_n)$ and $c/n \leq \eta_n$ and $(x \wedge (1 - x)) \sqrt{cn} \geq 1/\eta_n$. Furthermore, for every i the coefficients $a_{i,n}(x, c)$ are nonnegative, and uniformly in $c/n \leq \eta_n$ and $x \in [1/n_+, 1]$,*

$$a_{i,n}(x, c) \lesssim \frac{1}{n} \sqrt{cn} e^{-|x - x_{i,n}| \sqrt{cn}/2}, \quad (2.2)$$

$$\left| \sum_{i=1}^n a_i(x, c) - 1 \right| \lesssim e^{-x\sqrt{cn}/2}. \quad (2.3)$$

Furthermore, uniformly in c and x with $c/n \leq \eta_n$ and $(x \wedge (1-x)) \sqrt{cn} \geq 2 \log n$,

$$\left| \sum_{i=1}^n a_i(x, c)(x - x_{i,n}) \right| \lesssim \frac{1}{n}. \quad (2.4)$$

Moreover, uniformly in c and x with $c/n \leq \eta_n$ and $(x \wedge (1-x)) \sqrt{cn} \geq 1/\eta_n$,

$$\tau_n(x, x, c) \sim \frac{1}{4} \sqrt{\frac{c}{n}}. \quad (2.5)$$

Also, uniformly in c and $x \leq y$ with $c/n \leq \eta_n$ and $x \geq 1/n_+$,

$$\tau_n(x, y, c) \lesssim \frac{1}{n} \sqrt{cn} [1 + |x - y| \sqrt{cn}] e^{-|x-y|\sqrt{cn}/2}, \quad (2.6)$$

$$\bar{\sigma}_n(x, y, c) \lesssim \frac{1}{n} \sqrt{cn} [1 + |x - y| \sqrt{cn}] e^{-|x-y|\sqrt{cn}/2}. \quad (2.7)$$

Finally, uniformly in $c \leq d$ and $x \leq y$ with $d/n \leq \eta_n$ and $(x \wedge (1-x)) \sqrt{cn} \geq 1/\eta_n$,

$$\|\mathbf{a}_n(x, c) - \mathbf{a}_n(y, d)\|^2 \lesssim |x - y|^2 d \sqrt{dn} + \frac{(c - d)^2}{d \sqrt{dn}}, \quad (2.8)$$

$$\bar{\sigma}_n(x, x, c) + \bar{\sigma}_n(y, y, c) - 2\bar{\sigma}_n(x, y, c) \lesssim |x - y|^2 c \sqrt{cn}. \quad (2.9)$$

With the help of the preceding lemma we can characterise the order of magnitude of the posterior quantiles used in the construction of the credible bands.

Proposition 5. *The functions w_n and w_n^d defined by Eqs. 1.6 and 1.8 satisfy $w_n(c) \asymp w_n^d(c) \asymp \sqrt{\log(cn)}(c/n)^{1/4}$, uniformly in $c \in I_n$.*

PROOF. The centered posterior process given c is Gaussian with covariance function $\sigma_n(x, y, c)$. It can be represented as $\sqrt{c}W_x - \sqrt{c}\mathbf{W}_n^T \mathbf{a}_n(x, c) - \varepsilon_n^T \mathbf{a}_n(x, c)$. The mean zero Gaussian process $\mathcal{G}$ obtained by replacing the first term $\sqrt{c}W_x$ by $\sqrt{c}W_x 1^T \mathbf{a}_n(x, c)$ has covariance function $\bar{\sigma}_n$. By Eq. 2.3 it differs in uniform norm on $x \in J_n$ no more than of the order $\sup_{x \in J_n} \sqrt{ce^{-x\sqrt{cn}/2}} \leq \sqrt{ce^{-\sqrt{cn}/(2l_n)}}$ from the posterior process. This is bounded above by $(c/n)^{1/4}$ if $(cn)^{1/2}e^{-\sqrt{cn}/l_n} \leq 1$, in which case the uniform distance between $\mathcal{G}$ and the posterior process would tend to zero faster than $(c/n)^{1/4}$ and it would suffice to prove that the quantiles of the uniform norm of the process

$\mathcal{G}$ behave as claimed. Now the function $s \mapsto se^{-s/l_n}$ is decreasing for $s \geq l_n$. For $c \in I_n$ we have $s_n := \sqrt{cn} \geq \sqrt{\log n} \gg l_n \log \log n$, by the assumption on l_n and hence $(cn)^{1/2} e^{-\sqrt{cn}/l_n} \leq l_n \log \log n e^{-l_n \log \log n / l_n} = l_n \log \log n / \log n \leq 1$, uniformly in $c \in I_n$. We conclude that indeed we may consider $\mathcal{G}$ instead of the posterior process.

By Eq. 2.7 the posterior variance $\bar{\sigma}_n(x, x, c)$ of $\mathcal{G}$ is of the order $\sqrt{c/n}$, uniformly in $x \in J_n$, while the posterior covariance $\bar{\sigma}_n(x, y, c)$ is much smaller if $|x - y| \gg (cn)^{-1/2}$.

For a lower bound on the two quantiles $w_n(c)$ and $w_n^d(c)$, we select a subset $t_{1,m} < \dots < t_{m,m}$ of points from J_n with $t_{i+1,m} - t_{i,m} \geq K(cn)^{-1/2}$, for every i , for a sufficiently large constant K so that $\bar{\sigma}_n(t_{i,m}, t_{j,m}, c) \leq \delta \bar{\sigma}_n(t, t, c)$, for every i, j and t and some small δ . In the case of Eq. 1.8 we choose the points $t_{i,m}$ from the grid points $x_{i,n}$. By Eq. 2.7 we can choose m of the order $(cn)^{1/2}$. We can lower bound the quantiles of $\max_i |\mathcal{G}(t_{i,m})|$ by the expression in the proposition with the help of Lemma 6 below. The quantiles of the supremum of the process $\mathcal{G}$ over its continuous argument are not smaller and hence lower bounded in the same way.

For an upper bound on the quantiles it suffices to show that the mean of the variable $\|\mathcal{G}\|_\infty$ is of the order $\sqrt{\log(cn)}(c/n)^{1/4}$. For $x \in J_n$ and $c \in I_n$, we have $x\sqrt{cn} \geq l_n^{-1}\sqrt{\log n} \rightarrow \infty$, by assumption; similarly $(1-x)\sqrt{cn} \rightarrow \infty$. Therefore by Eq. 2.9 the square of the intrinsic metric of $\mathcal{G}$ satisfies,

$$d^2(x, y) = \bar{\sigma}_n(x, x, c) + \bar{\sigma}_n(y, y, c) - 2\bar{\sigma}_n(x, y, c) \lesssim |x - y|^2 c \sqrt{cn}.$$

It follows that the diameter for d of an interval of Euclidean length proportional to $(cn)^{-1/2}$ is of the order $(c/n)^{1/4}$ and the ε -covering number is bounded above by $(c/n)^{1/4}/\varepsilon$. Therefore, for $\|\cdot\|_{\psi_2}$ the Orlicz norm relative to $\psi_2(x) = e^{x^2} - 1$, by Corollary 2.2.5 from van der Vaart and Wellner (1996), for any t ,

$$\left\| \sup_{x: |x-t| < (cn)^{-1/2}} |\mathcal{G}(x) - \mathcal{G}(t)| \right\|_{\psi_2} \lesssim \int_0^{(c/n)^{1/4}} \sqrt{\log((c/n)^{1/4}/\varepsilon)} d\varepsilon \lesssim (c/n)^{1/4}.$$

Fix a grid (t_i) with mesh width $(nc)^{-1/2}$ over J_n . Then by the triangle inequality

$$\mathbb{E} \sup_{x \in J_n} |\mathcal{G}(x)| \lesssim \mathbb{E} \max_i |\mathcal{G}(t_i)| + \mathbb{E} \max_i \sup_{|x-t_i| < (cn)^{-1/2}} |\mathcal{G}(x) - \mathcal{G}(t_i)|.$$

Here the Orlicz norm of the Gaussian variable $\mathcal{G}(t_i)$ is bounded by a multiple of its standard deviation $\bar{\sigma}_n(t_i, t_i, c)^{1/2} \lesssim (c/n)^{1/4}$. Since there are $(cn)^{1/2}$ points t_i in the grid, the preceding display is bounded by $\sqrt{\log(cn)}(c/n)^{1/4}$ by two applications of Lemma 2.2.2 from van der Vaart and Wellner (1996).

Lemma 6. *If $(Z_{1,m}, \dots, Z_{m,m})$ possesses a zero-mean multivariate-normal distribution with $a_m^2 \leq \mathbb{E} Z_{i,m}^2 \lesssim a_m^2$ and $\text{cov}(Z_{i,m}, Z_{j,m}) \leq \delta a_m^2$ uniformly in (i, j) for some $a_m \rightarrow 0$ and $\delta < 1$, then $\Pr(\max_{1 \leq i \leq m} |Z_{i,m}| \leq w_m) = \eta \in (0, 1)$ implies that $w_m \asymp a_m \sqrt{\log m}$, as $m \rightarrow \infty$.*

PROOF. By Sudakov's inequality $\mathbb{E} \max_i |Z_{i,m}| \gtrsim a_m \sqrt{\log N(ca_m)}$, for $N(\varepsilon)$ the number of balls of radius ε needed to cover the set $\{1, 2, \dots, m\}$ relative to the metric with square $d^2(i, j) = \text{var}(Z_{i,m} - Z_{j,m})$ and any constant c . From the assumptions this metric can be seen to satisfy $d^2(i, j) \geq 2a_m^2(1 - \delta)$, whence any ball of radius ca_m , for $c^2 = (1 - \delta)/2$, contains at most one element. Thus $N(ca_m) \geq m$ and $\mathbb{E} \max_i |Z_{i,m}| \gtrsim a_m \sqrt{\log m}$. By the subGaussian maximal inequality this inequality can also be reversed. Furthermore, by Borell's inequality, for $x > 0$,

$$\Pr \left(\left| \max_{1 \leq i \leq m} |Z_{i,m}| - \mathbb{E} \max_{1 \leq i \leq m} |Z_{i,m}| \right| \geq xa_m \right) \leq 2e^{-x^2/2}.$$

Consequently, for a sufficiently large x the distribution of $\max_i |Z_{i,m}|$ gives mass arbitrarily close to 1 to an interval of width a_m located in the range $a_m \sqrt{\log m}$. Then its nontrivial quantiles must also be in this range.

3 Empirical Bayes estimators

The main result of this section is Proposition 9, which shows that the empirical Bayes estimators $\hat{c}_n$ are contained in an interval around some deterministic value $\tilde{c}_n(f)$, with probability tending to one. This assertion is a slight strengthening of results obtained in Sniekers and van der Vaart (2015a). Its statement and proof require to link the Fourier expansion of a function f on the continuous domain $[0, 1]$ to its discrete expansion on the design points.

When evaluated at the design points $x_{i,n}$, the eigenbasis e_j in Eq. 1.12 gives rise to an orthogonal basis of $\mathbb{R}^n$, which after rescaling to unit length takes the form

$$e_{j,n} = \frac{1}{\sqrt{n + 1/2}} (e_j(x_{1,n}), \dots, e_j(x_{n,n}))^T, \quad j = 1, \dots, n. \quad (3.1)$$

The discretisation $\mathbf{f}_n = (f(x_{1,n}), \dots, f(x_{n,n}))$ of a function f at the design points can be uniquely represented in terms of this basis as $\mathbf{f}_n = \sum_{j=1}^n f_{j,n} e_{j,n}$, for the coefficients $f_{j,n} := \mathbf{f}_n^T e_{j,n}$. If the Fourier series $f(x) =$

$\sum_{j=1}^{\infty} f_j e_j(x)$ of the continuous function f in terms of the e_j converges point-wise, then the discrete coefficients can be expressed in terms of the coefficients f_j as

$$f_{j,n} = \sqrt{n_+} \sum_{l=0}^{\infty} (f_{(2n+1)l+j} - f_{(2n+1)l+2n+2-j}). \quad (3.2)$$

This formula arises through aliasing of higher frequencies: when discretised to the grid of design points each of the continuous basis functions e_j for $j > n$ coincides with plus or minus $\sqrt{n_+}$ times one of the vectors Eq. 3.1 obtained from frequencies $j \leq n$; see Section 4.1 in Sniekers and van der Vaart (2015a) for details. In the latter paper it is shown that the behaviour of the empirical Bayes estimators $\hat{c}_n$ given in Eqs. 1.9 and 1.10 is determined by the coefficients $(f_{j,n})$.

Both estimators $\hat{c}_n$ minimise a criterium of the form

$$c \mapsto D_{1,n}(c, f) + D_{2,n}(c) + R_n(c, f),$$

where $D_{1,n}$ and $D_{2,n}$ are deterministic, and R_n is a stochastic remainder. With a superscript L referring to the likelihood-based and a superscript R for the risk-based functions, the deterministic functions are given by

$$\begin{aligned} D_{1,n}^L(c, f) &= \sum_{j=1}^n \frac{f_{j,n}^2}{(1 + c\lambda_{j,n})}, & D_{2,n}^L(c) &= \sum_{j=1}^n \left[\log(1 + c\lambda_{j,n}) - \frac{cl_{j,n}}{1 + c\lambda_{j,n}} \right], \\ D_{1,n}^R(c, f) &= \sum_{j=1}^n \frac{f_{j,n}^2}{(1 + c\lambda_{j,n})^2}, & D_{2,n}^R(c) &= \sum_{j=1}^n \frac{(c\lambda_{j,n})^2}{(1 + c\lambda_{j,n})^2}, \end{aligned}$$

where the $\lambda_{j,n}$ are the eigenvalues of the covariance matrix U_n of standard Brownian motion at the design points; these satisfy $\lambda_{j,n} \asymp n/j^2$ (Sniekers and van der Vaart (2015a), Example 5). In Sniekers and van der Vaart (2015a) it was proved that for both methods the estimator $\hat{c}_n$ minimizes the deterministic part $D_{n,1} + D_{n,2}$ within a multiplicative constant that can be chosen arbitrarily close to 1. This is true for general true regression functions f . Here we need a more precise result under the assumption that the true function f satisfies the *discrete polished tail condition*. For the Brownian motion prior this is the assumption that there exist constants L and ρ such that, for all sufficiently large m ,

$$\sum_{j=m}^n f_{j,n}^2 \leq L \sum_{j=m}^{\rho m \wedge n} f_{j,n}^2. \quad (3.3)$$

Lemma 7. *If f is self-similar of order $\beta > \frac{1}{2}$ with constants $(M, \varepsilon_1, \rho_1)$, then f satisfies the discrete polished tail condition Eq. 3.3 with constants (L, ρ) , where L and ρ depend on (ε_1, ρ_1) only. Moreover, uniformly for $c \in I_n$ and for proportionality constants that depend on (ε_1, ρ_1) only,*

$$\begin{aligned} D_{1,n}^L(c, f) &\asymp M^2 n \left(\frac{1}{cn}\right)^\beta, & \text{if } \beta < 1, \\ D_{1,n}^R(c, f) &\asymp M^2 n \left(\frac{1}{cn}\right)^\beta, & \text{if } \beta < 2. \end{aligned}$$

The lemma is essentially contained in Sniekers and van der Vaart (2015a); Section 10 of the supplement provides a full proof.

Functions that satisfy the discrete polished tail condition are nicely behaved in the sense that they satisfy the “good bias condition”. The following is Lemma 13 in Sniekers and van der Vaart (2015a).

Lemma 8 (Good bias condition). *If f satisfies the discrete polished tail condition with constants (L, ρ) , then f satisfies the good bias condition relative to both $D_{1,n}^L$ and $D_{1,n}^R$: there exists a constant $a > 0$ such that for $c \in I_n$,*

$$D_{1,n}(Kc, f) \leq K^{-a} D_{1,n}(c, f), \quad \text{for all } K > 1. \quad (3.4)$$

The constant a can be taken equal to $a = 2\rho(1 + \rho)^{-1}(1 + 4L)^{-1}$.

The functions $D_{2,n}$ do not depend on f , are increasing and behave asymptotically like $\sqrt{cn}$, uniformly in $c \in I_n$ (Sniekers and van der Vaart (2015a), Lemma 14). The functions $D_{1,n}$ are clearly decreasing. Let $\tilde{c}_n(f)$ be the unique solution to the equation

$$D_{1,n}(c, f) = D_{2,n}(c). \quad (3.5)$$

We have the following result on the location of the empirical Bayes estimators $\hat{c}_n$, when the true regression function satisfies the good bias condition and hence in particular when f is self-similar.

Proposition 9. *If f satisfies the good bias condition and $\tilde{c}_n(f) \in I_n$, then there are positive constants $k < K$ such that for $\hat{c}_n$ given in Eqs. 1.9 or 1.10*

$$\Pr_f(\hat{c}_n \in [k\tilde{c}_n(f), K\tilde{c}_n(f)]) \rightarrow 1.$$

The constant k depends on the constant a in the good bias condition only and is increasing in a , while K is universal. In particular, this is true if f is self-similar of order $\beta > 1/2$; in this case $\tilde{c}_n(f) \asymp M^{4/(1+2\beta)} n^{(1-2\beta)/(1+2\beta)}$, where the proportionality constant depends on the constants (L, ρ) of self-similarity only.

PROOF. Set $D_n(c, f) = D_{1,n}(c, f) + D_{2,n}(c)$, and for given $\varepsilon > 0$ let $\Lambda_n(\varepsilon)$ be the set of c such that $D_n(c, f) \leq (1 + \varepsilon) \inf_{c \in I_n} D_n(c, f)$. By Theorem 12 of Sniekers and van der Vaart (2015b) $\Pr_f(\hat{c}_n \in \Lambda_n(\varepsilon)) \rightarrow 1$, for every $\varepsilon > 0$. Since $\tilde{c}_n(f) \in I_n$ by assumption, it follows that $D_n(\hat{c}_n, f) \leq (1 + \varepsilon) D_n(\tilde{c}_n(f), f)$ with probability tending to one. For the Brownian motion prior it holds that $D_{2,n}(c) \asymp \sqrt{cn}$ (Sniekers and van der Vaart (2015a), Lemma 14). For f satisfying the good bias condition, the displayed result then follows by Lemma 17 in Section 11 in the supplement.

If f is self-similar of order β , then it satisfies the good bias condition by Lemmas 7 and 8. Furthermore, $D_{1,n}(c, f) \asymp M^2 n(cn)^{-\beta}$, by the first lemma. By monotonicity of the functions $D_{1,n}$ and $D_{2,n}$ it then follows that $\tilde{c}_n(f)$ is asymptotically of the same order as the solution in c of $M^2 n(cn)^{-\beta} = \sqrt{cn}$, which is of the order $M^{4/(1+2\beta)} n^{(1-2\beta)/(1+2\beta)}$.

4 Proof of Theorem 1: empirical Bayes case

Denote the centered posterior mean and its bias by

$$T_n(x, c) = \hat{f}_{n,c}(x) - \mathbb{E}_f \hat{f}_{n,c}(x), \quad (4.1)$$

$$\mu_n(x, c) = \mathbb{E}_f \hat{f}_{n,c}(x) - f(x). \quad (4.2)$$

The proof of Theorem 1 is based on the following two results.

Lemma 10. *Let $\epsilon > 0$. If f is Hölder of order $\alpha \in (0, 2]$ with constant M , then uniformly in $c \in [n^{-1+\epsilon}, n/\log n]$ and $x \in J_n$,*

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

PROOF. By the definition of the coefficients $a_{i,n}$ we have $\mu_n(x, c) = \sum_{i=1}^n a_{i,n}(x, c) f(x_{i,n}) - f(x)$. By Eq. 2.3 in Lemma 4 we have that $\sum_{i=1}^n a_{i,n}(x, c) = 1 + o((cn)^{-\alpha/2})$ for $x \gg \alpha(cn)^{-1/2} \log(cn)$, which is valid if $x \in J_n$ and $c \in I_n$. Therefore it suffices to bound the function $\tilde{\mu}_n(x, c) = \sum_{i=1}^n a_{i,n}(x, c) (f(x_{i,n}) - f(x))$.

For $\alpha \in (0, 1]$ and f with α -Hölder norm bounded by M the absolute value of $\tilde{\mu}_n(x, c)$ is bounded above by

$$\begin{aligned} \sum_{i=1}^n a_{i,n}(x, c) |f(x_{i,n}) - f(x)| &\leq M \sum_{i=1}^n a_{i,n}(x, c) |x_{i,n} - x|^\alpha \\ &\lesssim \int \sqrt{cn} e^{-|s-x|\sqrt{cn}/2} |s-x|^\alpha ds + \frac{1}{n} \sup_s \sqrt{cn} e^{-|s-x|\sqrt{cn}/2} |s-x|^\alpha, \end{aligned}$$

by Eq. 2.2 and Lemma 18, uniformly in $c \in I_n$ and $x \in J_n$. This is of the order $(cn)^{-\alpha/2}$ and proves the result.

If f is α -Hölder for $\alpha = 1 + \delta$ and some $\delta \in (0, 1]$, then by the mean value theorem there exist $\xi_{i,n}$ between $x_{i,n}$ and x so that $f(x_{i,n}) - f(x) = f'(\xi_{i,n})(x_{i,n} - x)$, and hence

$$\begin{aligned} |\tilde{\mu}_n(x, c)| &\leq \left| \sum_{i=1}^n a_{i,n}(x, c) (f'(\xi_{i,n}) - f'(x)) (x_{i,n} - x) + f'(x) \sum_{i=1}^n a_{i,n}(x, c) (x_{i,n} - x) \right| \\ &\leq M \sum_{i=1}^n a_{i,n}(x, c) |x_{i,n} - x|^{1+\delta} + M \left| \sum_{i=1}^n a_{i,n}(x, c) (x_{i,n} - x) \right|. \end{aligned}$$

By the argument in the preceding paragraph the first term can be seen to be bounded by a multiple of $(cn)^{-(1+\delta)/2}$. The second term is bounded by $1/n$, by Eq. 2.4 of Lemma 4, for any $c \geq n^{-1+\varepsilon}$ and $x \in J_n$.

Proposition 11. *If f satisfies the good bias condition and $\tilde{c}_n(f) \in I_n$, then for both the risk-based and likelihood-based empirical Bayes estimators $\hat{c}_n$,*

$$\sup_{x \in J_n} \frac{|T_n(x, \hat{c}_n)|}{\sqrt{\log(\hat{c}_n n)} (\hat{c}_n/n)^{1/4}} = O_{P_f}(1).$$

This statement is uniform in f such that the constant a in the good bias condition is bounded.

PROOF. Since f satisfies the good bias condition and $\tilde{c}_n := \tilde{c}_n(f)$ is contained in I_n , it follows by Proposition 9 that $\Pr_f(\hat{c}_n \in [k\tilde{c}_n, K\tilde{c}_n]) \rightarrow 1$, for some positive constants $k < K$. The constant k depends on the constant a in the good bias condition only and is bounded in a , while K is universal. Thus it suffices to show that the variables $\sup_{x \in J_n} \sup_{c \in [k\tilde{c}_n, K\tilde{c}_n]} |T_n(x, c)|$ are of the order $\sqrt{\log(\tilde{c}_n n)} (\tilde{c}_n/n)^{1/4}$ in probability.

Denote by $\|\cdot\|_{\psi_2}$ the Orlicz norm corresponding to the function $\psi_2(x) = e^{x^2} - 1$. Let $t_{1,n} < t_{2,n} < \dots < t_{m,n}$ be a minimal set of points over J_n that includes the two endpoints and hence meshwidth bounded by $1/\sqrt{\tilde{c}_n n}$; hence $m \sim \sqrt{\tilde{c}_n n}$. By Lemma 2.2.2 from van der Vaart and Wellner (1996),

$$\begin{aligned} E \sup_{x \in J_n} \sup_{c \in [k\tilde{c}_n, K\tilde{c}_n]} |T_n(x, c)| &\lesssim \sqrt{\log m} \max_{i \in \{1, \dots, m-1\}} \\ &\left\| \sup_{x \in (t_{i,n}, t_{i+1,n}]} \sup_{c \in [k\tilde{c}_n, K\tilde{c}_n]} |T_n(x, c)| \right\|_{\psi_2}. \end{aligned}$$

It suffices to show that the norms on the right are of the order $(\tilde{c}_n/n)^{1/4}$. The stochastic process T_n is zero-mean Gaussian with

$$\text{var}[T_n(x_1, c_1) - T_n(x_2, c_2)] = \|\mathbf{a}_n(x_1, c_1) - \mathbf{a}_n(x_2, c_2)\|^2 \lesssim \frac{(c_1 - c_2)^2}{\tilde{c}_n^{3/2} \sqrt{n}} + \sqrt{n}(x_1 - x_2)^2 \tilde{c}_n^{3/2},$$

by Eq. 2.8 of Lemma 4, for every $x_1 < x_2 \in (t_{i,n}, t_{i+1,n}]$ and $c_1 < c_2 \in [k\tilde{c}_n, K\tilde{c}_n]$. Let $d((x_1, c_1), (x_2, c_2))$ be the root of the right side. Since $\tilde{c}_n/n \rightarrow 0$, the diameter of the set $T := (t_{i,n}, t_{i+1,n}] \times [k\tilde{c}_n, K\tilde{c}_n]$ for d is bounded above by a multiple of $(\tilde{c}_n/n)^{1/4}$, and the ε -covering number of T is bounded by

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

Applying Corollary 2.2.5 in van der Vaart and Wellner (1996), we obtain, for every i ,

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

Combining this with the fact that $\|T_n(x_0, c_0)\|_{\psi_2} \lesssim \sqrt{\text{var} T_n(x_0, c_0)} \lesssim (\tilde{c}_n/n)^{1/4}$, for any fixed (x_0, c_0) in T , by Eq. 2.5, we see that $\left\| \sup_{(x,c) \in T} |T_n(x, c)| \right\|_{\psi_2} \lesssim (\tilde{c}_n/n)^{1/4}$, and hence it has the desired order of magnitude.

We are ready for the proof of Theorem 1 and the first assertion of Theorem 3.

In view of Proposition 5, the function f is contained in $\hat{C}_n(L)$ for a sufficiently large constant L if and only if we have $\sup_{x \in J_n} |f(x) - \hat{f}_{\hat{c}_n}(x)| < L\sqrt{\log(\hat{c}_n n)} (\hat{c}_n/n)^{1/4}$. By the triangle inequality this is certainly the case if

$$\frac{\sup_{x \in J_n} |T_n(x, \hat{c}_n)| + \sup_{x \in J_n} |\mu_n(x, \hat{c}_n)|}{L\sqrt{\log(\hat{c}_n n)} (\hat{c}_n/n)^{1/4}} < 1. \quad (4.3)$$

If f is self-similar of order $\beta > \frac{1}{2}$, then the array $(f_{i,n})$ is discrete polished tail by Lemma 7, and hence f satisfies the good bias condition by Lemma 8. Moreover, by Proposition 9 we have $\tilde{c}_n(f) \asymp M^{4/(1+2\beta)} n^{(1-2\beta)/(1+2\beta)}$. Since $\tilde{c}_n \in I_n$, it follows by Proposition 11 that the first term of Eq. 4.3 can be made arbitrarily small by choice of L . Moreover, by Proposition 9 there are positive constants $k < K$ such that $\hat{c}_n \in [k\tilde{c}_n(f), K\tilde{c}_n(f)]$ with probability tending to one. Consider the second term in Eq. 4.3. Since $f \in C^\alpha[0, 1]$, we have

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

uniformly for $c \in [k\tilde{c}_n, K\tilde{c}_n]$ and $x \in J_n$ by Lemma 10. We see that for $\alpha \geq \beta$ we have

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

with probability tending to one. It follows that the second term in Eq. 4.3 tends to 0 with probability tending to one. This concludes the proof of Theorem 1 for the empirical Bayes intervals.

Since $(\tilde{c}_n(f)/n)^{1/4} \asymp n^{-\beta/(1+2\beta)}$, the diameter of the credible band is of the order $\sqrt{\log n} n^{-\beta/(1+2\beta)}$ with probability tending to one, in view of Proposition 5. This proves the first assertion of Theorem 3.

5 Proof of Theorem 2: empirical Bayes case

We shall obtain Theorem 2 as a corollary of the following theorem, which guarantees coverage for a deterministic set of functions that will be shown to contain the random functions in the theorem with probability one.

For given $L_0 > 0$ let $\mathcal{F}_{L_0, a}$ be the set of all functions f that satisfy the good bias condition Eq. 3.4 with constant a and for which there exists $N \in \mathbb{N}$ such that for every $n \geq N$ the point $\tilde{c}_n(f)$ that equalises $D_{1,n}(c, f)$ and $D_{2,n}(c)$ (see Eq. 3.5) is contained in I_n , and such that

$$\max_{j \in \mathcal{J}_n} \frac{1}{n_+} \left(\sum_{i=1}^n \frac{f_{i,n} e_i(x_{j,n})}{1 + \tilde{c}_n(f) \lambda_{i,n}} \right)^2 < L_0 \log(\tilde{c}_n(f)n) \sqrt{\frac{\tilde{c}_n(f)}{n}}. \quad (5.1)$$

In the proof below the expression on the left side is shown to be the maximum of the square bias at the design points. Thus the condition simply assumes that the bias is smaller than the posterior deviation (a property that we proved in the preceding section under a Hölder condition on f , but will be seen to hold also for functions generated from the prior).

We shall describe the width of the credible band using the norms, for $\beta > 0$,

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

Theorem 12. *For every $L_0 > 0$ and for both the risk-based and likelihood-based empirical Bayes methods there exists L such that the credible sets $C_n^d(\hat{c}_n, L)$ as in Eq. 1.7 satisfy*

$$\liminf_{n \rightarrow \infty} \inf_{f \in \mathcal{F}_{L_0, a}} \Pr_f \left(f \in C_n^d(\hat{c}_n, L) \right) \geq \eta.$$

Furthermore, for given $\beta < 1$ the diameter of the credible set $C_n^d(\hat{c}_n, L)$ is of the order $O_P(\sqrt{\log n} n^{-\beta/(1+2\beta)})$ 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 also true for $\beta < 2$.

PROOF. As before let $T_n(x, c)$ be the posterior mean $\hat{f}_{n,c}(x)$ minus its expectation and let $\mu_n(x, c)$ the bias of the posterior mean (see Eq. 4.1–Eq. 4.2). The function f is covered by $\hat{C}_n^d(\hat{c}_n, L)$ for a sufficiently large L if we have

$$\frac{\max_{i \in \mathcal{J}_n} |T_n(x_{i,n}, \hat{c}_n)| + \max_{i \in \mathcal{J}_n} |\mu_n(x_{i,n}, \hat{c}_n)|}{L \sqrt{\log(\hat{c}_n n)} (\hat{c}_n/n)^{1/4}} < 1. \quad (5.2)$$

The term involving $T_n(x_{i,n}, \hat{c}_n)$ can be bounded by L^{-1} times the supremum in Proposition 11, and hence gives a contribution that is arbitrarily small if L is sufficiently large. We need to control the second term involving the bias at the grid points.

The vector $\mathbf{f}_n$ has prior $N(0, cU_n)$. Therefore its coordinates $f_{i,n}$ relative to the eigenbasis $e_{j,n}$ of U_n are a priori independent, and have $N(0, cl_{i,n})$ -priors. The coordinates $\tilde{Y}_{i,n}$ of the observation $\mathbf{Y}_n$ relative to the same basis are independent $N(f_{i,n}, 1)$ variables. It follows that under the posterior distribution the coordinates $f_{i,n}$ are again independent, and have normal distributions with means $cl_{i,n}/(1 + cl_{i,n})\tilde{Y}_{i,n}$. Hence the expectation of the posterior mean of $f_{i,n}$ is equal to $cl_{i,n}/(1 + cl_{i,n})f_{i,n}$.

For a grid point $x_{j,n}$ we have the representation $f(x_{j,n}) = \sum_{i=1}^n f_{i,n}(e_{i,n})_j = n_+^{-1/2} \sum_{i=1}^n f_{i,n} e_i(x_{j,n})$. The posterior mean of $f(x_{j,n})$ is obtained by replacing the $f_{i,n}$ by their posterior means. In combination with the preceding paragraph this shows that the bias at the grid point $x_{j,n}$ can be written in terms of the coefficients $(f_{i,n})$ as

$$\mu_n(x_{j,n}, c) = E_f \sum_{i=1}^n \left[\frac{cl_{i,n}}{1 + cl_{i,n}} \tilde{Y}_{i,n} - f_{i,n} \right] (e_{i,n})_j = -\frac{1}{\sqrt{n_+}} \sum_{i=1}^n \frac{f_{i,n} e_i(x_{j,n})}{1 + c\lambda_{i,n}}.$$

Assumption Eq. 5.1 entails that the square of this expression at $c = \tilde{c}_n(f)$ is bounded uniformly in $j \in \mathcal{J}_n$ by $L_1 \log(\tilde{c}_n(f)n) \sqrt{\tilde{c}_n(f)/n}$, for some constant L_1 . This is then also true uniformly for $c \in [k\tilde{c}_n, K\tilde{c}_n]$, for any constants $K > k > 0$. By Proposition 9 $\hat{c}_n$ falls in this interval with probability tending to one. It follows that the second term in Eq. 5.2 can be made arbitrarily small with probability tending to one, by taking L to be sufficiently large.

For functions f with $\|f\|_{n,\beta} \lesssim 1$ or $\|f\|_{n,\beta,\infty} \lesssim 1$ we have that $D_{1,n}^R(c, f) \lesssim n(cn)^{-\beta}$ if $\beta < 2$, while $D_{1,n}^L(c, f) \lesssim n(cn)^{-\beta}$ if $\beta < 1$. (See Examples 19

and 22 in Sniekers and van der Vaart (2015a) or Eqs. 10.3 and 10.4.) Since $D_{2,n}(c) \asymp \sqrt{cn}$, this implies that $\tilde{c}_n(f) \lesssim n^{(1-2\beta)/(1+2\beta)}$. Then the assertions on the diameter follow from Proposition 5, since $\hat{c}_n \asymp c_n(f)$.

We can now prove Theorem 2 by showing that W is in $\mathcal{F}_{L_0,a}$ for some L_0, a , almost surely. We give the proof for the risk-based empirical Bayes method; the proof for the likelihood-based method is analogous. The following lemma is Proposition 36 in Sniekers and van der Vaart (2015a).

Lemma 13. *The coordinates $W_{i,n}$ of the restriction $\mathbf{W}_n$ of the process Eq. 1.14 to the grid points relative to the basis $e_{i,n}$ are independent normal random variables with zero mean and $\text{var}(W_{i,n}) \asymp ni^{-1-2\alpha}$. Moreover, for sufficiently large L and ρ almost every realisation of W is discrete polished tail Eq. 3.3 with constants (L, ρ) .*

By Lemma 13 almost every realisation of W is discrete polished tail, and hence satisfies the good bias condition, by Lemma 8. We must prove that almost surely there is an N and L_0 such that $\tilde{c}_n(W) \in I_n$ and such that Eq. 5.1 holds for $n \geq N$.

For the proof of the first consider the stochastic process

$$V_n(c) = \frac{1}{n(cn)^{-\alpha}} (D_{1,n}^R(c, W) - \mathbb{E}D_{1,n}^R(c, W)) = \frac{1}{n(cn)^{-\alpha}} \sum_{i=1}^n \frac{W_{i,n}^2 - \mathbb{E}W_{i,n}^2}{(1 + c\lambda_{i,n})^2}.$$

In view of Lemma 13,

$$\begin{aligned} \mathbb{E}D_{1,n}^R(c, W) &\asymp \sum_{i=1}^n \frac{ni^{3-2\alpha}}{(i^2 + cn)^2} \asymp n(cn)^{-\alpha}, \\ \text{var}[V_n(c)] &\asymp (cn)^{2\alpha} \sum_{i=1}^n \frac{i^{6-4\alpha}}{(i^2 + cn)^4} \asymp \begin{cases} (cn)^{-1/2}, & \text{if } \alpha < 7/4, \\ (cn)^{-1/2} \log(cn), & \text{if } \alpha = 7/4, \\ (cn)^{2\alpha-4}, & \text{if } \alpha > 7/4. \end{cases} \end{aligned}$$

Therefore there exist constants $0 < \gamma_1 < \Gamma_1$ such that, for $c \in I_n$,

$$n(cn)^{-\alpha} (\gamma_1 + V_n(c)) \leq D_{1,n}^R(c, W) \leq n(cn)^{-\alpha} (\Gamma_1 + V_n(c)).$$

Also (see Lemma 14 in Sniekers and van der Vaart (2015a)) there exist constants $0 < \gamma_2 < \Gamma_2$ such that

$$\gamma_2 \sqrt{cn} \leq D_{2,n}^R(c) \leq \Gamma_2 \sqrt{cn}.$$

Set $b := (1-2\alpha)/(1+2\alpha)$ and consider the event $E_n = \{\tilde{c}_n(W) \notin [kn^b, Kn^b]\}$. Because $D_{1,n}$ is nonincreasing and $D_{2,n}$ is nondecreasing,

$$\Pr(\tilde{c}_n(W) < kn^b) \leq \Pr(D_{1,n}^R(kn^b, W) < D_{2,n}^R(kn^b))$$

$$\begin{aligned}
&\leq \Pr \left(k^{-\alpha} n^{1/(1+2\alpha)} \left[\gamma_1 + V_n \left(kn^b \right) \right] < \Gamma_2 \sqrt{kn}^{1/(1+2\alpha)} \right) \\
&= \Pr \left(V_n \left(kn^b \right) < -a \right),
\end{aligned}$$

where $a := \gamma_1 - \Gamma_2 k^{1/2+\alpha}$ is positive for sufficiently small k . Denote by $\|\cdot\|_{\psi_1}$ the Orlicz norm corresponding to the function $\psi_1(x) = e^x - 1$. Applying Proposition A.1.6 in van der Vaart and Wellner (1996) with $X_i = (W_{i,n}^2 - EW_{i,n}^2)/(n(cn)^{-\alpha}(1 + c\lambda_{i,n})^2)$, $S_n = V_n(c)$ and $p = 1$, followed by Lemma 2.2.2 from the same book, we see that

$$\begin{aligned}
\|V_n(c)\|_{\psi_1} &\lesssim \|V_n(c)\|_1 + \left\| \max_{i \in \mathcal{J}_n} |X_i| \right\|_{\psi_1} \\
&\lesssim \sqrt{\mathbb{E}[V_n(c)]^2} + \frac{\log n}{n(cn)^{-\alpha}} \max_{i \in \mathcal{J}_n} \frac{\|W_{i,n}^2 - EW_{i,n}^2\|_{\psi_1}}{(1 + c\lambda_{i,n})^2}.
\end{aligned}$$

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 $\|W_{i,n}^2 - EW_{i,n}^2\|_{\psi_1} \leq 2\|W_{i,n}^2\|_{\psi_1} \lesssim ni^{-1-2\alpha}$. We conclude that

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

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

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

Therefore, by Markov's inequality,

$$\Pr(V_n(kn^b) < -a) \leq \frac{1}{\psi_1(a/\|V_n(kn^b)\|_{\psi_1})} \leq \begin{cases} \left(e^{\frac{a}{C} n^{\frac{1}{4}(1+b)}} - 1 \right)^{-1}, & \text{if } \alpha < 7/4, \\ \left(e^{\frac{a}{C \log n} n^{(2-\alpha)(1+b)}} - 1 \right)^{-1}, & \text{if } \alpha \geq 7/4. \end{cases}$$

We conclude that there are constants $\tilde{K}, \tilde{k}, \gamma > 0$ such that for $\alpha \in (0, 2)$ it holds that $\Pr(V_n(kn^b) < -a) \leq \tilde{K}e^{-\tilde{k}n^\gamma}$. The probability $\Pr(\tilde{c}_n(W) > Kn^b)$ can be bounded by a similar argument. Consequently,

$$\sum_{n=1}^{\infty} \Pr(E_n) = \sum_{n=1}^{\infty} \Pr(\tilde{c}_n(W) < kn^b) + \sum_{n=1}^{\infty} \Pr(\tilde{c}_n(W) > Kn^b) \lesssim \sum_{n=1}^{\infty} e^{-\tilde{k}n^\gamma} < \infty.$$

Then the Borel-Cantelli lemma gives that almost surely there is an N such that $\tilde{c}_n(W) \in [kn^b, Kn^b]$ and hence $\tilde{c}_n(W) \in I_n$ for $n \geq N$.

Finally we prove that condition Eq. 5.1 is satisfied almost surely. The right side of this condition is of the order $\log n \sqrt{\tilde{c}_n(W)/n} \lesssim (\log n) n^{-\alpha/(1+2\alpha)}$. Define

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

By Lemma 13, uniformly for $x \in (0, 1)$, $c \in I_n$, and $s < t \in I_n$

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

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 see that, for $s < t \in [kn^b, Kn^b]$ and uniformly for $x \in (0, 1)$,

$$\|U_n(x, s) - U_n(x, t)\|_{\psi_2}^2 \lesssim \text{var}[U_n(x, s) - U_n(x, t)] \lesssim n^{-2b} n^{-\alpha(1+b)} (s-t)^2.$$

It then follows by Corollary 2.2.5 in van der Vaart and Wellner (1996) with $T = [kn^b, Kn^b]$ that

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

Applying Lemma 2.2.2 from van der Vaart and Wellner (1996) and noting that $\|U_n(x, c)\|_{\psi_2} \lesssim \sqrt{\text{var} U_n(x, c)} \lesssim n^{-\alpha/(1+2\alpha)}$ for any fixed $c \in [kn^b, Kn^b]$, we see that there exists a constant $\tilde{C} > 0$ with

$$\mathbb{E} \max_{j \in \mathcal{J}_n} \sup_{c \in [kn^b, Kn^b]} |U_n(x_{j,n}, c)| \leq \tilde{C} \sqrt{\log n} \cdot n^{-\frac{\alpha}{1+2\alpha}}.$$

By Borell's inequality the variable $S_n = \max_{j \in \mathcal{J}_n} \sup_{c \in [kn^b, Kn^b]} |U_n(x_{j,n}, c)|$ satisfies

$$\Pr \left(|S_n - \mathbb{E} S_n| > t \sqrt{\log n} \tilde{C} (cn)^{-\alpha/2} \right) \leq 2e^{-t^2 \log n}.$$

For sufficiently large t the right side is summable over n , whence the lim-sup of the events has probability zero. Since $(cn)^{-\alpha/2} \sim n^{-\alpha/(1+2\alpha)}$, for $c \in [kn^b, Kn^b]$, combination of the two preceding displays gives that $S_n \lesssim \sqrt{\log n} n^{-\alpha/(1+2\alpha)}$ eventually, almost surely. Because $\tilde{c}_n(W)$ is of polynomial order in n , the factors $\log n$ and $\log(\tilde{c}_n(W)n)$ are equivalent up to constants. This implies Eq. 5.1, and concludes the proof of Theorem 2 for the empirical Bayes choice of c .

By Lemma 13 the square norm $\|W\|_{n,\beta}^2$ of the stochastic process W has expectation of the order $\sum_j j^{2\beta-1-2\alpha}$, which is finite for $\beta < \alpha$. Together with the last assertion of Theorem 12 this implies the last assertion of Theorem 3.

6 Proof of Theorems 1 and 2: hierarchical Bayes case

In this section we extend the proofs of the main results to the hierarchical Bayes method. The key is the following result from Sniekers and van der Vaart (2015a) (see Theorem 25), which shows that the posterior distribution of the smoothing parameter c concentrates near the likelihood-based empirical Bayes estimator. Recall that the prior density for c is given by Eq. 1.11.

Proposition 14. *Suppose that there is a minimiser $c_n(f)$ of $c \mapsto D_n^L(c, f) + 2\lambda/c$ over $c \in (0, \infty)$ that satisfies $c_n(f) \in I_n$ and $2c_n(f) \in I_n$. Then there exist constants $0 < k < K < \infty$ such that*

$$\Pi_n(c : c \in [kc_n(f), Kc_n(f)] \mid \mathbf{Y}_n) \xrightarrow{P_{\mathcal{I}}} 1. \quad (6.1)$$

We have that $f \in \hat{C}_n(L)$ as in Theorem 1 if and only if there exists $c \in (\hat{c}_{1,n}, \hat{c}_{2,n})$ such that f satisfies

$$\sup_{x \in J_n} |f(x) - \hat{f}_{n,c}(x)| < L\sqrt{\log(cn)} \left(\frac{c}{n}\right)^{1/4}. \quad (6.2)$$

In Section 4 this display was seen to be valid for every c in an interval $[k\tilde{c}_n(f), K\tilde{c}_n(f)]$ around the point $\tilde{c}_n(f)$ that equalises the functions $c \mapsto D_{n,1}^L(c, f)$ and $c \mapsto D_{n,2}^L(c)$. Thus it suffices to show that an interval of this form contains a point from $(\hat{c}_{1,n}, \hat{c}_{2,n})$. Since the interval $(\hat{c}_{1,n}, \hat{c}_{2,n})$ has positive posterior mass by construction, it certainly suffices to show that Eq. 6.1 is valid with $c_n(f)$ replaced by $\tilde{c}_n(f)$.

Now for every $f \in \mathcal{F}_{\alpha,\beta}$ Lemma 7 gives that

$$D_{n,1}^L(c, f) + D_{n,2}^L(c) + \frac{2\lambda}{c} \asymp n \left(\frac{1}{cn}\right)^\beta + \sqrt{cn} + \frac{2\lambda}{c}.$$

By definition $\tilde{c}_n(f)$ equalises the first two terms. From the explicit expression on the right side it is seen that the minimiser of the sum of the first two terms is of the same order as the equaliser and both are of the order $n^{(1-2\beta)/(1+2\beta)}$. Since the last term is of smaller order in $\tilde{c}_n(f)$, we conclude that the minimiser $c_n(f)$ of the entire expression is of the same order as well. Thus $c_n(f) \asymp \tilde{c}_n(f)$ and hence the desired result follows by Proposition 14.

For the proof in the case of Theorem 2 we apply the same argument, but now show that Eq. 6.1 is valid with $c_n(f)$ replaced by $\tilde{c}_n(W)$, almost surely. This is true as both $\tilde{c}_n(W)$ and $c_n(W)$ are of the order $n^b = n^{(1-2\alpha)/(1+2\alpha)}$, almost surely. This follows because $D_{1,n}^L(c, W)$ behaves almost surely as its mean $ED_{1,n}^L(c, W)$, which is of the order $n(cn)^{-\alpha}$. This was shown for the risk-based function $D_{1,n}^R$ in the preceding section, and extends to the likelihood-based function.

7 Counterexample

Theorem 1 shows that the credible band will cover a function that is Hölder of order at least equal to its order of self-similarity. If the two measures of smoothness do not match, then the empirical or hierarchical Bayes method to choose the scaling of the prior will choose a value that does not balance the square bias and variance, which can result in credible bands that are too narrow or too wide.

The two types of smoothness do not agree in general. For instance, for the Hölder exponent of the function $f_\alpha(t) = |x - t|^\alpha$, where $\alpha \in (0, 1)$ and x is a fixed value in $(1/2, 1)$, is α . Below we show that it is self-similar of order $\beta = 1$. In this section we give an example that is Hölder of order $\alpha \in (0, 1)$, but for which the order of self-similarity is equal to $1/2$. Thus the Hölder exponent can be both smaller or larger than the order of self-similarity. For $\alpha < 1/2$ this will result in credible bands that are too narrow and hence give a mistaken impression of remaining uncertainty, while for $\alpha > 1/2$ the credible bands will cover but be unnecessarily wide. Thus despite the positive results expressed in Theorems 1 and 2, credible bands are not necessarily accurate.

We prove below that the Fourier coefficients of f_α satisfy $f_j \asymp j^{-1}$, for any $\alpha \in (0, 1)$, which implies that it is self-similar of order $\beta = \frac{1}{2}$. The bias of the posterior mean at the point x has the exact order $(cn)^{-\alpha/2}$, by Corollary 3.1 in Sniekers and van der Vaart (2015b). The self-similarity suggests that $\tilde{c}_n$ and $\hat{c}_n$ will be of the order $\tilde{c}_n \asymp 1$. Then for $\alpha < 1/2$ the bias $\sup_{x \in J_n} |\mu_n(x, \tilde{c}_n)| \gtrsim n^{-\alpha/2}$, which is much bigger than the width of the credible band $\sqrt{\log n} (\tilde{c}_n/n)^{1/4}$. Figure 1 illustrates this with a simulation for the true function with $\alpha = 1/4$.

Figure 1: Credible band for the function $f(x) = |x - 3/4|^{1/4}$, for $n = 500$ and $n = 12500$, for $\hat{c}_n$ set by the likelihood-based empirical Bayes method

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 (1988). For $\alpha > 1/2$ it can be shown that a $D_{1,n}(c, f) \asymp n(cn)^{-1/2}$ on I_n . From this it follows that the credible sets $\hat{C}_n(L)$ have asymptotic coverage one if $\alpha > \frac{1}{2}$, but the diameter is suboptimal; it is of the order $\sqrt{\log n} n^{-1/4} \gg (n/\log n)^{-\alpha/(1+2\alpha)}$.

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

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

where $r = j + 1/2$. We may write

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

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

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

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

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

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

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

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

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

Here we use $\xi_j \geq x - (b+4k)/(2r)$, but we can apply the same argument with $\xi_j \leq x - (a+4k)/(2r)$ 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

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

and from below by

$$\frac{1}{\pi} \sum_{k=0}^{\tilde{m}_0} \left(x - \frac{1+4k}{2r} \right)^\alpha - \left(x - \frac{2+4k}{2r} \right)^\alpha + \left(x - \frac{2+4k}{2r} \right)^\alpha - \left(x - \frac{3+4k}{2r} \right)^\alpha$$

$$\geq -\frac{1}{4\pi} \left[\left(x - \frac{1+4\tilde{m}_0}{2r} \right)^\alpha - \left(x - \frac{1}{2r} \right)^\alpha + \left(x - \frac{2+4\tilde{m}_0}{2r} \right)^\alpha - \left(x - \frac{2}{2r} \right)^\alpha \right] \rightarrow \frac{1}{2\pi} x^\alpha.$$

For the other sum we have

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

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

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

Here we use $\xi_j \geq (2j - 4k - d)/(2r) - x$, but applying the same argument with $\xi_j \leq (2j - 4k - c)/(2r) - x$, we obtain a lower bound (changing the upper limit of the integral to $\tilde{m}_1$) that 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 $-(1-x)^\alpha/(2\pi)$ and an upper bound that converges to $(1-x)^\alpha/(2\pi)$. Since $x > 1/2$, 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 $f_j \sim (\sqrt{2}/\pi)x^\alpha/(j+1/2)$.

8 Posterior mean

The following proposition generalizes Theorem 2.2 of Sniekers and van der Vaart (2015b). It is proved by the same arguments.

Proposition 15. *Let $i_n(x) = \max\{i : x_{i,n} < x\}$ and $\lambda_+ = 1 + \sqrt{c/n_+}$, $\sqrt{1+c/(4n_+)} + \frac{1}{2}c/n_+$, and fix an arbitrary sequence $\eta_n \downarrow 0$. The coefficients $a_{i,n}(c, x)$ satisfy, uniformly in c and x such that $c/n \leq \eta_n$ and $(x \wedge (1-x))\sqrt{cn} \geq 1/\eta_n$,*

$$a_{i,n}(x, c) \sim \frac{1}{2} \sqrt{\frac{c}{n}} \begin{cases} \lambda_+^{-i_n(x)} [\lambda_+^i - \lambda_+^{-i}], & \text{for } i \leq i_n(x), \\ \lambda_+^{i_n(x)} [\lambda_+^{-i+1} + \lambda_+^{-2n+i}], & \text{for } i > i_n(x). \end{cases} \quad (8.1)$$

Furthermore, the coefficients $a_{i,n}(x, c)$ are nonnegative, and for every $x \in (1/n_+, 1]$ bounded by thrice the expression on the right side, which is in turn bounded above by $4\sqrt{c/n} l_+^{-|i-i_n(x)|}$, uniformly in $c/n \leq \eta_n$. Moreover, for $x \geq 1/n_+$,

$$1 - \sum_{i=1}^n a_{i,n}(x, c) = \frac{n_+}{c} a_{1,n}(x, c). \quad (8.2)$$

PROOF. Abbreviate $a_{i,n}(x, c)$ to $a_{i,n}$ and $i_n(x)$ to i_n . Since $\mathbf{Y}_n^T \mathbf{a}_n$ is the projection of $\sqrt{c} W_x$ onto the observations $Y_{i,n}$, the coefficients $a_{i,n}$ satisfy the orthogonality relations, for $i = 1, \dots, n$,

$$0 = E[(\sqrt{c} W_x - \mathbf{Y}_n^T \mathbf{a}_n) Y_{i,n}] = c E[(W_x - \mathbf{W}_n^T \mathbf{a}_n) W_{x_{i,n}}] - a_{i,n}.$$

Evaluating this with $i = 1$ and noting that $EW_x W_{x_{1,n}} = 1/n_+$, for every $x \geq x_{1,n}$, which includes $x = x_{i,n}$ for every i , we readily find Eq. 8.2.

The n projection equations can be written as a linear system with rows $i = 1, \dots, n$. Replacing first for $i = 2, \dots, n$ the i th equation by the difference of the i th and $(i-1)$ th equations, and next replacing in the resulting system for $i = 1, \dots, n-1$ the i th equation by the difference of the i th and $(i+1)$ th equations, we obtain the simplified linear system

$$\begin{pmatrix} 2 + c/n_+ & -1 & 0 & \cdots & 0 \\ -1 & 2 + c/n_+ & -1 & \cdots & 0 \\ 0 & -1 & \ddots & & \vdots \\ \vdots & \vdots & & 2 + c/n_+ & -1 \\ 0 & 0 & \cdots & -1 & 1 + c/n_+ \end{pmatrix} \begin{pmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ \vdots \\ a_{n,n} \end{pmatrix} = c \begin{pmatrix} 0 \\ \vdots \\ 0 \\ x_{i_n+1,n} - x \\ x - x_{i_n,n} \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (8.3)$$

The equations in rows $2, 3, \dots, i_n-1$ and $i_n+2, \dots, n-1$ yield the recurrence relation, for $i \in \{3, \dots, i_n\}$ and $i \in \{i_n+3, \dots, n\}$,

$$a_{i,n} = (2 + c/n_+) a_{i-1,n} - a_{i-2,n}. \quad (8.4)$$

This has characteristic polynomial $\lambda^2 - (2 + c/n_+) \lambda + 1$. The roots l_+ and l_- of the corresponding characteristic equation possess product $l_+ l_- = 1$ and sum $l_+ + l_- = 2 + c/n_+$.

For $i \in \{1, \dots, i_n\}$ the general solution takes the form $a_{i,n} = A \lambda_+^i + B \lambda_-^i$, for two constants A and B . The first equation in the system Eq. 8.3 is $(l_+ + l_-) a_{1,n} - a_{2,n} = 0$. Substituting the general solutions $a_{1,n} = A l_+ + B l_-$ and $a_{2,n} = A l_+^2 + B l_-^2$ in this equation readily gives that $A = -B$. Hence

$$a_{i,n} = A(\lambda_+^i - \lambda_-^i), \quad i \in \{1, \dots, i_n\}. \quad (8.5)$$

By Eq. 8.4 the numbers $b_{i,n} = a_{n-i+1,n}$ satisfy the same recurrence relation $b_{i,n} = (2 + c/n_+) b_{i-1,n} - b_{i-2,n}$, for $i \in \{3, \dots, n - i_n\}$, and hence $b_{i,n} = \tilde{A}\lambda_+^i + \tilde{B}\lambda_-^i$, for two constants $\tilde{A}$ and $\tilde{B}$ and every $i \in \{1, \dots, n - i_n\}$. The last row of the system of Eqs. 8.3 gives $-b_{2,n} + (l_+ + l_- - 1)b_{1,n} = 0$. Substituting the general solutions for $b_{2,n}$ and $b_{1,n}$ into this equation readily gives that $\tilde{B} = l_+ \tilde{A} = \tilde{A}/l_-$. Translating back from $b_{i,n}$ to $a_{i,n}$, we conclude that

$$a_{i,n} = \tilde{A}(\lambda_+^{n-i+1} + \lambda_-^{n-i}), \quad i \in \{i_n + 1, \dots, n\}. \quad (8.6)$$

The i_n th and $(i_n + 1)$ th equations of Eq. 8.3 can be written

$$-a_{i_n-1,n} + (l_+ + l_-)a_{i_n,n} - a_{i_n+1,n} = c(x_{i_n+1,n} - x), \quad (8.7)$$

$$-a_{i_n,n} + (l_+ + l_-)a_{i_n+1,n} - a_{i_n+2,n} = c(x - x_{i_n,n}). \quad (8.8)$$

We substitute the general solutions Eqs. 8.5 and 8.6 to find, after simplification, that the constants A and $\tilde{A}$ are the solutions of the linear system

$$\begin{aligned} A(\lambda_+^{i_n+1} - \lambda_-^{i_n+1}) - \tilde{A}(\lambda_+^{n-i_n} + \lambda_-^{n-i_n-1}) &= c(x_{i_n+1,n} - x), \\ -A(\lambda_+^{i_n} - \lambda_-^{i_n}) + \tilde{A}(\lambda_+^{n-i_n+1} + \lambda_-^{n-i_n}) &= c(x - x_{i_n,n}). \end{aligned}$$

The determinant of this linear system can be calculated to be $\Delta_n = l_+^n(l_+^2 - 1) - l_-^{n-1}(l_-^2 - 1) = l_+^n(l_+^2 - 1)(1 + l_+^{-2n-1})$. Then

$$\begin{pmatrix} l_+^{i_n} A \\ l_+^{n-i_n} \tilde{A} \end{pmatrix} = \frac{cl_+^n}{\Delta_n} \begin{pmatrix} l_+(1 + l_+^{-2(n-i_n)-1}) & 1 + l_+^{-2(n-i_n)+1} \\ 1 - l_+^{-2i_n} & l_+(1 - l_+^{-2i_n-2}) \end{pmatrix} \begin{pmatrix} x_{i_n+1,n} - x \\ x - x_{i_n,n} \end{pmatrix}. \quad (8.9)$$

Here $l_+ - 1 \sim \sqrt{c/n} \rightarrow 0$ and $l_+^2 - 1 \sim 2\sqrt{c/n}$ uniformly in $c/n \rightarrow 0$, so that $cl_+^n/\Delta_n \sim \sqrt{cn}/2$. The four entries of the matrix are all smaller than $1 + l_+ \sim 2$ uniformly in $c/n \rightarrow 0$. The coordinates of the vector on the far right are nonnegative and add up to $1/n_+$, as $x_{i_n,n} < x \leq x_{i_n+1,n}$, by the definition of i_n . Together with Eqs. 8.5 and 8.6 this shows that thrice the expression on the right side of Eq. 8.1 is an upper bound on $a_{i,n}(x, c)$.

Since $i_n < xn_+ \leq i_n + 1$, we have that $(i_n \wedge (n - i_n))\sqrt{c/n} \rightarrow \infty$, uniformly in x and c such that $(x \wedge (1 - x))\sqrt{cn} \geq l_n \rightarrow \infty$ and $c \leq n$. Since $\log l_+^k \sim k\sqrt{c/n}$, this shows that the 2×2 matrix in Eq. 8.9 converges to the matrix with all four entries equal to 1 uniformly in the same set of x and c . This gives the asymptotic equivalence Eq. 8.1.

9 Proof of Lemma 4

PROOF OF EQ. 2.2. By Proposition 15 every coefficient $a_{i,n}(x, c)$ is nonnegative and bounded above by $6\sqrt{c/n}l_+^{-|i_n-i|}$. Here $l_+ = 1 + \sqrt{c/n} +$

$O(c/n)$ is bounded below by $e^{\sqrt{c/n}/2}$ uniformly in $c/n \leq \eta_n \rightarrow 0$, since $1 + 3x/4 \geq e^{x/2}$ for $x \in [0, 1]$, and hence $l_+^{i_n-i} \leq e^{-|i_n-i|\sqrt{c/n}/2}$. Also $|i_n/n - x|\sqrt{cn} \leq \sqrt{c/n} \rightarrow 0$, uniformly for $c/n \leq \eta_n \rightarrow 0$, so that $e^{(i_n/n-x)\sqrt{cn}}$ is bounded.

PROOF OF EQ. 2.1. By Proposition 15, for $i \leq i_n(x)$,

$$\log \left[\frac{2na_{i,n}(x, c)}{\sqrt{cn} e^{-(x-x_{i,n})\sqrt{cn}}} \right] = -(i_n - i) \log l_+ + \sqrt{cn}(x - x_{i,n}) + \log(1 - l_+^{2i}) + o(1).$$

Here $\log l_+ = \sqrt{c/n} + O(c/n)$, uniformly in $c/n \leq \eta_n \rightarrow 0$, so that the first term is equivalent to $-(x_{i_n,n} - x_{i,n})\sqrt{cn}$ up to a remainder of order $|x_{i_n,n} - x_{i,n}|c$, which tends to zero uniformly in $|x - x_{i,n}|c \leq \eta_n$, since $|x_{i_n,n} - x|c \leq 2c/n$. The approximation $-(x_{i_n,n} - x_{i,n})\sqrt{cn}$ combines with the second term on the right side of the display to $-(x_{i_n,n} - x)\sqrt{cn}$, which is of the order $\sqrt{c/n} \leq \sqrt{\eta_n} \rightarrow 0$. For $x\sqrt{cn} \geq \eta_n^{-1}$ and $|x_{i,n} - x|\sqrt{cn} < (1/2)\eta_n^{-1}$, we have $i\sqrt{c/n} \rightarrow \infty$, whence $l_+^{2i} \rightarrow 0$, and the last term on the right tends to zero. For $i > i_n(x)$ the proof is similar.

PROOF OF EQ. 2.3. By Proposition 15, uniformly in $x \geq 1/n_+$ and $c/n \leq \eta_n \rightarrow 0$,

$$1 - \sum_{i=1}^n a_{i,n}(x, c) = \frac{n_+}{c} a_{1,n}(x, c) \lesssim \sqrt{\frac{n}{c}} l_+^{-i_n} (l_+ - l_+^{-1}) \sim l_+^{-i_n} \left(1 + O(\sqrt{c/n})\right).$$

The logarithm of this is bounded above by $-i_n \log l_+ \sim -i_n \sqrt{c/n} \leq -(x - 1/n_+)\sqrt{cn}/2 \leq -x\sqrt{cn}/2 + o(1)$.

PROOF OF EQ. 2.4. If $(x \wedge (1-x))\sqrt{cn} \geq 2 \log n$, then $i_n \wedge (n - i_n) \geq 2 \log n \sqrt{n/c}(1 - o(1))$, uniformly in $c \leq n$. Thus we can choose j_n with $(3/2) \log n \sqrt{n/c} \leq j_n \leq (7/4) i_n \wedge (n - i_n)$, and decompose $\sum_{i=1}^n a_i(x, c)(x_{i,n} - x)$ as

$$\begin{aligned} & \sum_{j=1}^{j_n} [a_{i_n-j+1,n}(x, c)(x - x_{i_n-j+1,n}) + a_{i_n+j,n}(x, c)(x - x_{i_n+j,n})] \\ & + \sum_{\substack{i \leq i_n - j_n \\ \text{or } i > i_n + j_n}} a_{i,n}(x, c)(x - x_{i,n}). \end{aligned} \quad (9.1)$$

For $x > 1/n_+$ and $c/n \leq \eta_n \rightarrow 0$, the second term is bounded above by a multiple of

$$\sqrt{\frac{c}{n}} \sum_{i \leq i_n - j_n} l_+^{-(i_n-i)} + \sqrt{\frac{c}{n}} \sum_{i > i_n + j_n} 2l_+^{-(i-i_n)} \lesssim \sqrt{\frac{c}{n}} \frac{l_+^{j_n}}{1 - l_+^{-1}}.$$

Since $1 - l_+^{-1} \sim \sqrt{c/n}$, uniformly in $c/n \leq \eta_n \rightarrow 0$, this is bounded above by a multiple of $1/n$ if $j_n \log l_+ \geq \log n$. Since $\log l_+ \sim \sqrt{c/n}$, uniformly in $c/n \leq \eta_n \rightarrow 0$, this is certainly the case under the assumption that $j_n \geq 3/2 \log n \sqrt{n/c}$. To bound the first term of Eq. 9.1, we first note that for $i_n \wedge (n - i_n) \gg \sqrt{n/c}$, which we have assumed, the four entries in the 2×2 matrix in Eq. 8.9 are $1 + O(\sqrt{c/n})$, so that the quotient of the two coordinates of the vector on the left is $1 + O(\sqrt{c/n})$ as well, uniformly in $c/n \leq \eta_n \rightarrow 0$ and $(x \wedge (1 - x)) \sqrt{cn} \geq 2 \log n$. Combining this with Eq. 8.5 and Eq. 8.6, we see

$$\frac{a_{i_n-j+1,n}}{a_{i_n+j,n}} = \frac{A}{\bar{A}} \frac{l_+^{i_n-j+1} - l_-^{i_n-j+1}}{l_+^{n-i_n-j+1} + l_-^{n-i_n-j}} = l_+^{n-2i_n} \left[1 + O\left(\sqrt{\frac{c}{n}}\right) \right] \frac{l_+^{i_n-j+1}(1 - l_+^{-2i_n+2j-2})}{l_+^{n-i_n-j+1}(1 + l_+^{-2n+2i_n+2j-1})}.$$

The powers of l_+ cancel and $1 - l_+^{-2i_n+2j-2}$ and $1 + l_+^{-2n+2i_n+2j-1}$ are $1 + O(\sqrt{c/n})$ for $j \leq j_n$, since $((i_n - j_n) \wedge (n - i_n - j_n)) \geq (1/4) \log n \sqrt{n/c}$. We conclude that $a_{i_n-j+1,n} = a_{i_n+j,n}(1 + R_{j,n})$, for a remainder satisfying $|R_{j,n}| \lesssim \sqrt{c/n}$, so that the first term in Eq. 9.1 is bounded above by

$$\begin{aligned} & \sum_{j=1}^{j_n} [a_{i_n+j,n}(x, c) (2x - x_{i_n-j+1,n} - x_{i_n+j,n}) + (x - x_{i_n-j+1,n}) R_{j,n}] \\ & \lesssim \frac{1}{n} \sum_{j=1}^{j_n} a_{i_n+j,n}(x, c) + \sum_{j=1}^{j_n} \frac{c}{n} l_+^{-j+1} \frac{j}{n}. \end{aligned}$$

The first term is bounded by $1/n$ as the sum of the coefficients is bounded by 1. The second term is bounded above by a multiple of $(c/n^2) \sum_{j=0}^{\infty} (j l_+^{-j})$ and is also bounded by a multiple of $1/n$, uniformly in $c/n \leq \eta_n \rightarrow 0$.

PROOF OF EQ. 2.5. For given x and $M_n \rightarrow \infty$ we have, by Eq. 2.2,

$$\sum_{i: |x_{i,n} - x| > M_n / \sqrt{cn}} a_{i,n}^2(x, c) \lesssim \frac{1}{n} \sqrt{cn} e^{-M_n/2} \sum_{i=1}^n a_{i,n}(x, c) \leq \frac{1}{n} \sqrt{cn} e^{-M_n/2}.$$

For $M_n \rightarrow \infty$ this is of smaller order than $\sqrt{c/n}$. On the other hand, if $c/n \leq \eta_n \rightarrow 0$ and $M_n = 1/\sqrt{\eta_n}$, then $M_n/\sqrt{cn} \leq \sqrt{\eta_n}/c$ and hence by Eq. 2.1 (applied with $\sqrt{\eta_n}$ in the role of $2\eta_n$), uniformly in $(x \wedge (1 - x)) \sqrt{cn} \geq 1/\eta_n \geq 1/\sqrt{\eta_n}$,

$$\begin{aligned} \sum_{i: |x_{i,n} - x| \leq M_n / \sqrt{cn}} a_{i,n}^2(x, c) & \sim \frac{c}{4n} \sum_{i: |x_{i,n} - x| \leq M_n / \sqrt{cn}} e^{-2|x - x_{i,n}| \sqrt{cn}} \\ & \sim \frac{c}{4} \int_{x - M_n / \sqrt{cn}}^{x + M_n / \sqrt{cn}} e^{-2|x - s| \sqrt{cn}} ds = \frac{1}{4} \sqrt{\frac{c}{n}} \int_{-M_n}^{M_n} e^{-2|s|} ds \sim \frac{1}{4} \sqrt{\frac{c}{n}}. \end{aligned}$$

By Lemma 18 the error made in the second step by approximating the sum by the integral is smaller than (c/n) times the maximum value of the integrand, which is 1, which is indeed of smaller order than the right side of the display.

PROOF OF EQ. 2.6. For given x, y we have, by Eq. 2.2,

$$\begin{aligned} \sum_i a_{i,n}(x, c) a_{i,n}(y, c) &\lesssim \frac{c}{n} \sum_{i=1}^n e^{-|x-x_{i,n}|\sqrt{cn}/2 - |y-x_{i,n}|\sqrt{cn}/2} \\ &\lesssim c \int_0^1 e^{-|x-s|\sqrt{cn}/2 - |y-s|\sqrt{cn}/2} ds + \frac{c}{n} \sup_s e^{-|x-s|\sqrt{cn}/2 - |y-s|\sqrt{cn}/2}. \end{aligned}$$

The integral can be evaluated to be no bigger than $(3\sqrt{c/n} + c|x-y|)e^{-|x-y|\sqrt{cn}/2}$; the second term is negligible.

PROOF OF EQ. 2.7. The function $\bar{\sigma}_n(x, y, c)$ is the sum of $\tau_n(x, y, c)$ and the covariance function of the process $\sum_{i=1}^n a_{i,n}(x, c)(W_x - W_{x_{i,n}})$. For $x = x_{i_n+1,n}$ the increments $W_x - W_{x_{i,n}}$ can be written as a sum (for $i \leq i_n$) or negative sum (for $i > i_n$) of the increments $V_j = W_{x_{j,n}} - W_{x_{j-1,n}}$ over the grid points. Substituting these sums in $\sum_{i=1}^n a_{i,n}(x, c)(W_x - W_{x_{i,n}})$ and exchanging the order of the resulting double sums yields that this sum is equal to

$$\sum_{i \leq i_n} \left[a_{i,n}(x, c) \sum_{i < j \leq i_n+1} V_j \right] - \sum_{i > i_n+1} \left[a_{i,n}(x, c) \sum_{i_n+1 < j \leq i} V_j \right] = \sum_{j=2}^n A_{j,n}(x, c) V_j,$$

where

$$A_{j,n}(x, c) = \begin{cases} \sum_{i=1}^{j-1} a_{i,n}(x, c), & \text{if } j \leq i_n(x) + 1, \\ -\sum_{i=j}^n a_{i,n}(x, c), & \text{if } j > i_n(x) + 1. \end{cases}$$

By the independence of the increments it follows that

$$\text{cov} \left(\sum_{i=1}^n a_{i,n}(x, c)(W_x - W_{x_{i,n}}), \sum_{i=1}^n a_{i,n}(y, c)(W_y - W_{x_{i,n}}) \right) = \frac{1}{n} + \sum_{j=1}^n A_{j,n}(x, c) A_{j,n}(y, c).$$

For $j \leq i_n(x) + 1$ we have, by Eq. 2.2,

$$A_{j,n}(x, c) \lesssim \sqrt{\frac{c}{n}} \sum_{i=1}^{j-1} e^{-(x-x_{i,n})\sqrt{cn}/2} \leq \sqrt{cn} \int_0^{x_{j,n}} e^{-(x-s)\sqrt{cn}/2} ds \leq 2e^{-|x-x_{j,n}|\sqrt{cn}/2}.$$

For $j > i_n(x) + 1$ the same bound is valid. This bound is of the same form as the bound on $a_j(x, c)$, except for a factor $\sqrt{c/n}$. It follows by the same arguments as for the proof of Eq. 2.6 that $(c/n) \sum_{j=1}^n A_{j,n}(x, c) A_{j,n}(y, c)$ is bounded as $\sum_{j=1}^n a_{j,n}(x, c) a_{j,n}(y, c) = \tau_n(x, y, c)$.

For x not equal to a grid point, the exact representation of $\sum_{i=1}^n a_{i,n}(x, c)$ ($W_x - W_{x_{i,n}}$) in terms of the increments V_j is retained if the definitions of V_{i_n+1} and V_{i_n+2} are modified to $W_x - W_{x_{i_n,n}}$ and $W_{x_{i_n+2,n}} - W_x$. The variances of these variables are also bounded above by a multiple of $1/n$, and hence the preceding derivation goes through.

PROOF OF EQ. 2.8. The orthogonality of the residual $\sqrt{c}W_x - (\sqrt{c}\mathbf{W}_n + \varepsilon_n)^T \mathbf{a}_n$ and $(\sqrt{c}\mathbf{W}_n + \varepsilon_n)^T \mathbf{a}_n$ gives that $cU(x, \mathbf{x}_n) = (cU_n + I)\mathbf{a}_n$, for U_n the covariance matrix of $\mathbf{W}_n$ and $U(x, \mathbf{x}_n)$ the vector with coordinates $\text{cov}(W_x, W_{x_{i,n}}) = x \wedge x_{i,n}$. Therefore $(c^{-1}I + U_n)\mathbf{a}_n(x, c) = U(x, \mathbf{x}_n)$ is free of c and hence

$$\|\mathbf{a}_n(x, c) - \mathbf{a}_n(x, d)\| = \left\| \left[(c^{-1}I + U_n)^{-1}(d^{-1}I + U_n) - I \right] \mathbf{a}_n(x, d) \right\| \leq \kappa \|\mathbf{a}_n(x, d)\|,$$

for κ the largest eigenvalue of the matrix $(c^{-1}I + U_n)^{-1}(d^{-1}I + U_n) - I$. The eigenvalues of this matrix are given by $(c - d)/(d(1 + c\lambda_{j,n}))$, for $\lambda_{j,n} \asymp n/j^2$ the eigenvalues of U_n , whence $\kappa \leq |c - d|/d$. Since $\|\mathbf{a}_n(x, d)\|^2 = \tau_n(x, x, d) \asymp \sqrt{d/n}$ by Eq. 2.5, we obtain the bound $|c - d|/(d^{3/4}n^{1/4})$ on the preceding display.

To complete this to a proof of Eq. 2.8 we combine this with a bound on $\|\mathbf{a}_n(x, c) - \mathbf{a}_n(y, c)\|$. Since $(cU_n + I)\mathbf{a}_n = cU(x, \mathbf{x}_n)$ and U is continuous, the coefficients $a_{i,n}$ depend continuously on x . Furthermore, Eq. 8.3 shows that $\mathbf{a}_n$ is differentiable with respect to x in every interval $(x_{i,n}, x_{i+1,n}]$, as $i_n(x)$ is constant in such an interval and x appears only in the right side of Eq. 8.3. Differentiating across Eq. 8.3 we see that the derivatives $\mathbf{a}'_n$ satisfy the same equation, except that the vector on the far right must be replaced by its derivative, which has -1 and 1 as its i_n st and $(i_n + 1)$ st coordinates and zeros elsewhere. The same analysis as in the proof of Proposition 15 shows that

$$a'_{i,n} = \begin{cases} A_1(l_+^i - l_-^i), & i \in \{1, \dots, i_n\}, \\ \tilde{A}_1(l_+^{n-i+1} + l_-^{n-i}), & i \in \{i_n + 1, \dots, n\}, \end{cases}$$

where A_1 and $\tilde{A}_1$ are constants satisfying the analogue of Eq. 8.9 given by

$$\begin{pmatrix} l_+^{i_n} A_1 \\ l_+^{n-i_n} \tilde{A}_1 \end{pmatrix} = \frac{cl_+^n}{\Delta_n} \begin{pmatrix} l_+(1 + l_+^{-2(n-i_n)-1}) & 1 + l_+^{-2(n-i_n)+1} \\ 1 - l_+^{-2i_n} & l_+(1 - l_+^{-2i_n-2}) \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

As noted before, the four entries of the 2×2 matrix in the display tend to 1, uniformly in $(x \wedge (1 - x))\sqrt{cn} \rightarrow \infty$ and $c/n \rightarrow 0$. Since this matrix

maps the vector $(-1, 1)^T$ to 0, we expand the right side of the display more precisely as

$$\frac{\sqrt{cn}}{2} (1 + o(1)) \begin{pmatrix} l_+ + o(c/n) & 1 + o(c/n) \\ 1 + o(c/n) & l_+ + o(c/n) \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \sim \frac{\sqrt{cn}}{2} (l_+ - 1) \begin{pmatrix} -1 \\ 1 \end{pmatrix} \sim \frac{c}{2} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

We conclude that $A_1 \sim -\sqrt{cn}A$ and $\tilde{A}_1 \sim \sqrt{cn}\tilde{A}$, for A and $\tilde{A}$ given in the proof of Proposition 15, so that $|a'_{i,n}| \sim \sqrt{cn}a_{i,n}$, for $x \notin \{x_{1,n}, \dots, x_{n,n}\}$, where the sign is negative if $i \leq i_n(x)$ and positive otherwise. Therefore, for $x < y$,

$$\|\mathbf{a}_n(x, c) - \mathbf{a}_n(y, c)\|^2 = \sum_i \left(\int_x^y a'_{i,n}(s) ds \right)^2 \leq cn|y - x| \int_x^y \sum_i a_{i,n}^2(s, c) ds \leq cn|y - x|^2 \sqrt{\frac{c}{n}},$$

since $\|\mathbf{a}_n(x, c)\|^2 \lesssim \sqrt{c/n}$, uniformly in its argument, in view of Eq. 2.5.

PROOF OF EQ. 2.9. The left side of the inequality is the second moment of the increment over $[x, y]$ of the process with covariance function $\bar{\sigma}_n(x, y, c)$. By the representation as used in the proof of Eq. 2.7,

$$cE \left[\sum_{i=1}^n a_{i,n}(x, c)(W_x - W_{x_{i,n}}) - \sum_{i=1}^n a_{i,n}(y, c)(W_y - W_{x_{i,n}}) \right]^2 \sim \frac{c}{n} \sum_{j=2}^n (A_{j,n}(x, c) - A_{j,n}(y, c))^2.$$

The functions $A_{j,n}$ are continuous in x , and differentiable with respect to x except at grid points, with derivatives satisfying $A'_{j,n} = \sum_{i=1}^{j-1} a'_{i,n}$, for $j \leq i_n$ and $A'_{j,n} = \sum_{i=j}^n a'_{i,n}$, for $j > i_n$. By the result of the preceding paragraph we have in both cases that $|A'_{j,n}| \lesssim \sqrt{cn}A_{j,n}$. Therefore, by the same argument as in the preceding paragraph the preceding display is bounded above by $cn|y - x|^2 \sup_x \sum_j A_{j,n}^2(x, c)$. By Eq. 2.7 applied with $x = y$, this is bounded by the right side of Eq. 2.9.

10 Proof of Lemma 7

The first part of the following proof is adapted from Example 34 and the last parts from Examples 22 and 23 in Sniekers and van der Vaart (2015a).

Since $|f_j| \leq Mn^{-1/2-\beta}$, for every j , we have for $\ell \geq 1$,

$$|f_{(2n+1)\ell+i}| \vee |f_{(2n+1)\ell+2n+2-i}| \leq \frac{M}{n^{1/2+\beta}\ell^{1/2+\beta}}.$$

Since $\sum_{l \geq 1} l^{-1/2-\beta} =: C_\beta < \infty$, this shows that the series Eq. 3.2 that defines the aliased coefficients converges. Because the term for $l = 0$ of the series is $f_j - f_{2n+2-j}$ and $|f_{2n+2-j}| \leq M(n+2)^{-1/2-\beta}$ for every $j \leq n$, we see that

the rescaled coefficients $\tilde{f}_{i,n} = f_{i,n}/\sqrt{n_+}$ satisfy $|\tilde{f}_{i,n} - f_i| \leq 2C_\beta M n^{-1/2-\beta}$, so that $|\tilde{f}_{i,n}| \leq 3C_\beta M i^{-1/2-\beta}$ and the left side of Eq. 3.3 satisfies

$$\sum_{i=m}^n \tilde{f}_{i,n}^2 \leq 18C_\beta^2 \frac{M^2}{m^{2\beta}}. \quad (10.1)$$

We wish to show that the right side of Eq. 3.3 is lower bounded by the expression on the right, where we may assume that m satisfies $\rho m \leq n$, because otherwise there is nothing to prove. First we note that

$$|\tilde{f}_{i,n}^2 - f_i^2| = |\tilde{f}_{i,n} - f_i| |\tilde{f}_{i,n} + f_i| \leq \frac{2C_\beta M}{n^{1/2+\beta}} \left(2|f_i| + \frac{2C_\beta M}{n^{1/2+\beta}} \right) \leq \frac{8C_\beta^2 M^2}{n^{1/2+\beta} i^{1/2+\beta}}.$$

It follows that, for f self-similar with constants (ε_1, ρ_1) and any $\rho \geq \rho_1$ and $\rho m \leq n$,

$$\sum_{i=m}^{\rho m \wedge n} \tilde{f}_{i,n}^2 \geq \sum_{i=m}^{\rho m} f_i^2 - \frac{8C_\beta^2 M^2}{n^{1/2+\beta}} \sum_{i=m}^{\rho m} \frac{1}{i^{1/2+\beta}} \gtrsim \frac{M^2}{m^{2\beta}} \left(\varepsilon_1 - 8C_\beta^2 \frac{\rho - 1}{\rho^{1/2+\beta}} \right). \quad (10.2)$$

For sufficiently large ρ the constant on the right is positive. It follows that f is discretely self-similar with constants $M, \varepsilon = \varepsilon_1 - 8C_\beta^2(\rho - 1)/\rho^{1/2+\beta}$ and $\rho \geq \rho_1$ large enough that $\varepsilon > 0$. Combining this with Eq. 10.1 we also see that f satisfies the discrete polished tail condition Eq. 3.3 with constants $L = 18C_\beta^2/\varepsilon$ and ρ .

Since $1 + c\lambda_{j,n} \leq 1 + \rho^2$, for $j \geq \sqrt{cn}/\rho$ we have for discretely self-similar f ,

$$D_{1,n}^R(c, f) = \sum_{i=1}^n \frac{f_{i,n}^2}{(1 + c\lambda_{i,n})^2} \geq \frac{1}{(1 + \rho^2)^2} \sum_{i=\sqrt{cn}/\rho}^n f_{i,n}^2 \gtrsim \frac{n\varepsilon M^2 \rho^{2\beta}}{(1 + \rho^2)^2} \left(\frac{1}{cn} \right)^\beta.$$

For $D_{1,n}^L$ the same inequality is true, but with the factor $(1 + \rho^2)^2$ replaced by $1 + \rho^2$. Finally

$$\frac{1}{M^2} D_{1,n}^R(c, f) \lesssim n \sum_{j=1}^n \frac{j^{-2\beta-1}}{(1 + c\lambda_{j,n})^2} \lesssim n \sum_{j=1}^n \frac{j^{3-2\beta}}{(j^2 + cn)^2} \lesssim \begin{cases} n(cn)^{-\beta} & \text{if } \beta < 2 \\ n(cn)^{-2} \log(cn) & \text{if } \beta = 2 \\ n(cn)^{-2} & \text{if } \beta > 2. \end{cases} \quad (10.3)$$

The first case follows directly by Lemma 16, the second by writing

$$n \sum_{j=1}^n \frac{j^{3-2\beta}}{(j^2 + cn)^2} = n \sum_{j=1}^{\sqrt{cn}} \frac{j^{3-2\beta}}{(j^2 + cn)^2} + n \sum_{j=\sqrt{cn}+1}^n \frac{j^{3-2\beta}}{(j^2 + cn)^2}$$

and applying a variant of the lemma to the second sum. The third case follows immediately by using $j^2 + cn > cn$. For the likelihood-based method we have

$$\frac{1}{M^2} D_{1,n}^L(c, f) \lesssim n \sum_{j=1}^n \frac{j^{-2\beta-1}}{1+c\lambda_{j,n}} \lesssim n \sum_{j=1}^n \frac{j^{1-2\beta}}{j^2+cn} \lesssim \begin{cases} n(cn)^{-\beta} & \text{if } \beta < 1 \\ c^{-1} \log(cn) & \text{if } \beta = 1 \\ c^{-1} & \text{if } \beta > 1. \end{cases} \quad (10.4)$$

This concludes the proof of Lemma 7.

11 Technical results for easy reference

For easy reference we state technical results from earlier papers.

Lemma 16 (Sniekers and van der Vaart (2015a), Lemma 43). *Let $\gamma > -1$, $m \geq 1$ and $\nu \in \mathbb{R}$ such that $\gamma - m\nu < -1$. Then*

$$\sum_{j=1}^n \frac{j^\gamma}{(j^m + cn)^\nu} = C_{\gamma, \nu, m}(cn)^{\gamma/m - \nu + 1/m} (1 + o(1)) \quad (11.1)$$

uniformly for $c \in [l_n/n, n^{m-1}/l_n]$ as $n \rightarrow \infty$, for any $l_n \rightarrow \infty$. The constant is given by

$$C_{\gamma, \nu, m} = \int_0^\infty \frac{u^\gamma}{(u^m + 1)^\nu} du.$$

Furthermore, the left side of (11.1) has the same order as the right side uniformly in $c \in [l_n/n, n^{m-1}]$, for any $l_n \rightarrow \infty$, possibly with a smaller constant.

Lemma 17 (Sniekers and van der Vaart (2015a), Lemma 42). *Let $D_1 : I_n \rightarrow (0, \infty)$ be a decreasing function and $D_2 : I_n \rightarrow (0, \infty)$ an increasing function. Suppose that there exist $a, b, B, B' > 0$ such that*

$$\begin{aligned} D_1(Kc) &\leq K^{-a} D_1(c), & \text{for any } K > 1, \\ B' k^b D_2(c) &\geq D_2(kc) \geq B k^b D_2(c) & \text{for any } k < 1. \end{aligned}$$

Let $\tilde{c}$ satisfy $D_1(\tilde{c}) = D_2(\tilde{c})$, and for a given constant $E \geq 1$, define $\Lambda = \{c : (D_1 + D_2)(c) \leq E(D_1 + D_2)(\tilde{c})\}$. Then

$$(i) \quad D_1(c) \leq B^{-1} (2E)^{1+b/a} D_2(c), \text{ for every } c \in \Lambda.$$

$$(ii) \quad \Lambda \subset [(2E)^{-1/a} \tilde{c}, (2EB')^{1/b} \tilde{c}].$$

Lemma 18. *If $f : [0, 1] \rightarrow \mathbb{R}_{\geq 0}$ is increasing on $[0, m]$ and decreasing on $[m, 1]$, then $\int_0^1 f(x) dx - f(m)/n_+ \leq n_+^{-1} \sum_{i=1}^n f(i/n_+) \leq \int_0^1 f(x) dx + f(m)/n_+$.*

PROOF. This is elementary analysis.

Acknowledgements. Part of the results in this paper were first obtained in the PhD thesis (Sniekers, 2015) of the first author.

Open Access. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- BULL, A. (2012). Honest adaptive confidence bands and self-similar functions. *Electron. J. Statist.* **6**, 1490–1516. <https://doi.org/10.1214/12-EJS720>. <http://projecteuclid.org/euclid.ejs/1346421602>.
- BULL, A. and NICKL, R. (2013). Adaptive confidence sets in l^2 . *Probab. Theory Relat. Fields* **156**, 3–4, 889–919. <https://doi.org/10.1007/s00440-012-0446-z>.
- CAI, T.T. and LOW, M.G. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32**, 5, 1805–1840. <https://doi.org/10.1214/009053604000000049>. <http://projecteuclid.org/euclid.aos/1098883773>.
- CAI, T.T. and LOW, M.G. (2006). Adaptive confidence balls. *Ann. Statist.* **34**, 1, 202–228. <https://doi.org/10.1214/009053606000000146>.
- COX, D.D. (1993). An analysis of bayesian inference for nonparametric regression. *Ann. Statist.* **21**, 2, 903–923. <https://doi.org/10.1214/aos/1176349157>. <http://projecteuclid.org/euclid.aos/1176349157>.
- FREEDMAN, D. (1999). On the Bernstein-von Mises theorem with infinite-dimensional parameters. *Ann. Statist.* **27**, 4, 1119–1140. <http://projecteuclid.org/getRecord?id=euclid.aos/1017938917>.
- GHOSAL, S., GHOSH, J.K. and VAN DER VAART, A.W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28**, 2, 500–531. <https://doi.org/10.1214/aos/1016218228>.
- GHOSAL, S. and VAN DER VAART, A. (2007). Convergence rates of posterior distributions for noniid observations. *Ann. Statist.* **35**, 1, 192–223. <https://doi.org/10.1214/009053606000001172>. <http://projecteuclid.org/euclid.aos/1181100186>.
- GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference Cambridge Series in Statistical and Probabilistic Mathematics*, 44. Cambridge University Press, Cambridge. <https://doi.org/10.1017/9781139029834>.
- GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38**, 2, 1122–1170. <https://doi.org/10.1214/09-AOS738>. <http://projecteuclid.org/euclid.aos/1266586625>.
- HOFFMANN, M. and NICKL, R. (2011). On adaptive inference and confidence bands. *Ann. Statist.* **39**, 5, 2383–2409. <https://doi.org/10.1214/11-AOS903>.
- JOHNSTONE, I.M. (2010). High dimensional bernstein-von mises: simple examples. *IMS Collections* **6**, 87–98. <https://doi.org/10.1214/10-IMSCOLL607>. <http://projecteuclid.org/euclid.imsc/1288099014>.
- JUDITSKY, A. and LAMBERT-LACROIX, S. (2003). On nonparametric confidence set estimation. *Math. Meth. of Stat* **19**, 4, 410–428. <http://membres-timc.imag.fr/Sophie.Lambert/papier/JuditskyLambert2003.pdf>.

- KNAPIK, B., VAN DER VAART, A.W. and VAN ZANTEN, J.H. (2011). Bayesian inverse problems with gaussian priors. *Ann. Statist.* **39**, 5, 2626–2657. <https://doi.org/10.1214/11-AOS920>. http://www.few.vu.nl/bt.knapik/research/inverse_aos.pdf.
- LOW, M.G. (1997). On nonparametric confidence intervals. *Ann. Statist.* **25**, 6, 2547–2554. <https://doi.org/10.1214/aos/1030741084>. <http://projecteuclid.org/euclid.aos/1030741084>.
- ROBINS, J. and VAN DER VAART, A.W. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.* **34**, 1, 229–253. <https://doi.org/10.1214/009053605000000877>. <http://projecteuclid.org/euclid.aos/1146576262>.
- SNIEKERS, S. (2015). Credible sets in nonparametric regression. PhD Dissertation Leiden University. <https://openaccess.leidenuniv.nl/handle/1887/36587>.
- SNIEKERS, S. and VAN DER VAART, A. (2015). Adaptive Bayesian credible sets in regression with a Gaussian process prior. *Electron. J. Stat.* **9**, 2, 2475–2527. <https://doi.org/10.1214/15-EJS1078>.
- SNIEKERS, S. and VAN DER VAART, A. (2015). Credible sets in the fixed design model with Brownian motion prior. *J. Statist. Plann. Inference* **166**, 78–86. <https://doi.org/10.1016/j.jspi.2014.07.008>.
- STONE, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10**, 4, 1040–1053. [http://links.jstor.org/sici?sici=0090-5364\(198212\)10:4<1040:OGROCF>2.0.CO;2-2&origin=MSN](http://links.jstor.org/sici?sici=0090-5364(198212)10:4<1040:OGROCF>2.0.CO;2-2&origin=MSN).
- SZABÓ, B., VAN DER VAART, A.W. and VAN ZANTEN, J.H. (2015a). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43**, 4, 1391–1428. <https://doi.org/10.1214/14-AOS1270>.
- SZABÓ, B., VAN DER VAART, A.W. and VAN ZANTEN, J.H. (2015b). Rejoinder to discussions of Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43**, 4, 1463–1470. <https://doi.org/10.1214/15-AOS1270REJ>.
- SZABO, B.T., VAN DER VAART, A.W. and VAN ZANTEN, J.H. (2013). Empirical bayes scaling of gaussian priors in the white noise model. *Electron. J. Statist.* **7**, 991–1018. <https://doi.org/10.1214/13-EJS798>. <http://projecteuclid.org/euclid.ejs/1366031048>.
- VAN DER VAART, A. and VAN ZANTEN, H. (2007). Bayesian inference with rescaled Gaussian process priors. *Electron. J. Stat.* **1**, electronic, 433–448. <https://doi.org/10.1214/07-EJS098>.
- VAN DER VAART, A.W. and WELLNER, J.A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer, New York. With applications to statistics.
- WAHBA, G. (1983). Bayesian “confidence intervals” for the cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser. B* **45**, 1, 133–150. [http://links.jstor.org/sici?sici=0035-9246\(1983\)45:1<133:B>IFTC>2.0.CO;2-B&origin=MSN](http://links.jstor.org/sici?sici=0035-9246(1983)45:1<133:B>IFTC>2.0.CO;2-B&origin=MSN).
- YOO, W.W. and VAN DER VAART, A.W. (2018). The Bayes Lepski’s method and credible bands through volume of tubular neighborhoods. arXiv:1711.06926.
- ZYGMUND, A. (1988). *Trigonometric series. Vol. I, II. Cambridge Mathematical Library*. Cambridge University Press, Cambridge. Reprint of the 1979 edition.

Publisher’s Note. Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

SUZANNE SNEKERS
AND AAD VAN DER VAART
MATHEMATICAL INSTITUTE,
LEIDEN UNIVERSITY,
P.O. BOX 9512, 2300, RA LEIDEN,
THE NETHERLANDS
E-mail: avdvaart@math.leidenuniv.nl

Paper received: 16 December 2018.