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

*Chapter 1*

# Pointwise credible sets

## 1.1 Introduction and main result

We consider estimating the regression function $f$ in the fixed design regression problem, where we have data

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

Here $(x_{i,n})$ is a known sequence of points in the interval $[0,1]$, and $(\varepsilon_{i,n})$ is a sequence of unobservable i.i.d. standard normal random variables. We take a nonparametric Bayesian approach, using a Gaussian process prior $W = (W_t : t \in [0,1])$ on $f$, and are interested in the resulting credible sets. These are sets of prescribed posterior probability, which in the Bayesian paradigm are used to quantify the remaining uncertainty of the statistical analysis. We investigate the coverage of these sets when treating them as confidence sets in the non-Bayesian setting. Specifically we focus on credible intervals for $f(x)$, the function $f$ evaluated at a given point $x$, which can be derived from the marginal posterior distribution of $W_x$.

As a prior for $f$ we consider the distribution of a scaled Brownian motion. Thus we are given a mean-zero Gaussian process $W = (W_t : t \in [0,1])$ with covariance function $\operatorname{cov}(W_s, W_t) = c_n(s \wedge t)$, for given scale factors $c_n > 0$. We take this process to be independent of the sequence $(\varepsilon_{i,n})$. In the Bayesian setup the observations are distributed according to the model

$$Y_{i,n} = W_{x_{i,n}} + \varepsilon_{i,n}.$$

Furthermore, the posterior distribution of $f(x)$ is the conditional distribution of $W_x$ given $Y_{1,n}, \dots, Y_{n,n}$. In this Gaussian model the posterior distribution is also Gaussian and hence is characterised by its posterior mean $\hat{f}_n(x) = \mathrm{E}(W_x|Y_{1,n}, \dots, Y_{n,n})$ and posterior variance $\sigma_n^2 = \mathrm{var}(W_x|Y_{1,n}, \dots, Y_{n,n})$. The natural credible interval with level $\eta$ for $f(x)$ is the central interval

$$C_\eta = \big(\hat{f}_n(x) - \sigma_n\zeta_\eta, \hat{f}_n(x) + \sigma_n\zeta_\eta\big),$$

where $\zeta_\eta$ is a standard normal quantile such that $P(|Z| < \zeta_\eta) = \eta$ for $Z \sim \mathcal{N}(0,1)$. The coverage of this interval in the frequentist setting is the probability $P_f(f(x) \in C_\eta)$, where $P_f$ refers to the distribution of $Y_{1,n}, \dots, Y_{n,n}$ in the original model (1.1), where a "true" $f$ is given.

This model has been widely studied in the literature. In [Kimeldorf and Wahba, 1970], Kimeldorf and Wahba showed that the posterior mean is the solution to a penalized smoothing problem. Rates of contraction of the posterior distribution $W \mid \vec{Y}_n$ relative to the $L_2$-metric were obtained in [van der Vaart and van Zanten, 2007], [van der Vaart and van Zanten, 2008] and [van der Vaart and van Zanten, 2011]. In this chapter we study the marginal posterior distribution $W_x \mid \vec{Y}_n$. The results on posterior mean and variance can be used to obtain rates of contraction for this marginal posterior. Bayesian credible sets for the function $f$ in some infinite-dimensional space were considered in [Wahba, 1983], [Cox, 1993], [Leahu, 2011], [Knapik et al., 2011], but only in a heuristic discussion and simulation study, without proofs, or only for the white noise model. (Estimation of a smooth functional of $f$ is a different problem, which may be studied using Bernstein-von Mises theorems.) The present treatment extends this to pointwise credible sets in the regression model. Scaling factors $c_n$ in the variance were introduced in [van der Vaart and van Zanten, 2007] with the purpose of adapting the prior to the smoothness of the underlying regression function. These authors show that the rescaled Brownian motion with $c_n = n^{(1-2\alpha)/(1+2\alpha)}$ is a suitable prior for a true function $f$ of Hölder smoothness $\alpha$, where $\alpha \in (0,1]$. In this chapter, we obtain a similar result in the marginal setting for $\alpha \in (0,2]$.

The prior $W$ considered here takes the value $W_0 = 0$ at the origin. This could be remedied by adding an independent normal variable to $W$, but as we consider the performance of the posterior distribution at fixed $x > 0$, this will be irrelevant in the following.

The present model permits a fairly explicit solution. We will consider the case where the design points are given by $x_{i,n} = i/n_+$ with $n_+ = n + 1/2$. The exact formulas cannot easily be extended to a more general choice of design points. Furthermore, we take the scaling factors equal to $c_n = n_+^\beta$ for some $\beta \in (-1,1)$.

Define $C^\alpha[0,1]$ as the space of Hölder continuous functions with exponent $\alpha \in (0,2]$. The main result of the chapter is the following theorem.

**Theorem 1.1.** *Define $\xi_\beta := \frac{1-\beta}{2(1+\beta)}$. The following holds for the coverage $c_\eta^f := P_f\big(f(x) \in C_\eta\big)$:*

- *If $\alpha > \xi_\beta$, we have $c_\eta^f \to P(|U| < \zeta_\eta) =: p_\eta > \eta$ for all $f \in C^\alpha[0,1]$, where $U \sim \mathcal{N}(0, 1/2)$.*
- *If $\alpha = \xi_\beta$, then for each $p \in (0, p_\eta]$ there exists $f \in C^\alpha[0,1]$ such that $c_\eta^f \to p$.*
- *If $\alpha < \xi_\beta$, there exists $f \in C^\alpha[0,1]$ such that $c_\eta^f \to 0$.*

In the first of the three cases the credible interval is a conservative confidence set (i.e $p_\eta > \eta$). Although it is wider than necessary for coverage, its width shrinks to zero at the same order of magnitude as the frequentist confidence interval based on the posterior mean, which would use the frequentist standard deviation of the posterior mean, rather than the standard deviation of the posterior distribution. This follows from the fact that $p_\eta$ is strictly smaller than 1. As $\xi_\beta \downarrow 0$ as $\beta \uparrow 1$, the range of $\alpha$ for which this favourable conclusion holds can be made arbitrarily large by choice of $\beta$. However, we shall see that

$$\sigma_n \asymp n^{(\beta-1)/4}.$$

Therefore using a large value of $\beta$ will also increase the width of the credible set, even by an order of magnitude.

In the third case the credible interval is too narrow to give positive coverage for all functions of given Hölder smoothness. The standard deviation of the posterior distribution is of smaller order than the bias of the posterior mean in this case. This is due to oversmoothing of the true function by the prior, the Bayesian way of choosing too large a bandwidth in a smoothing method.

Without scaling (i.e. $\beta = 0$) the cut-off between good and bad performance of the credible sets is at $\xi_0 = 1/2$. This can be viewed as the smoothness of Brownian motion itself. In this case, functions of smoothness bigger than $\frac{1}{2}$ yield credible sets with positive coverage, whereas functions that are rougher than Brownian motion do not.

Inspection of the proof shows that the assumption $f \in C^\alpha[0,1]$ can be relaxed to Hölder continuity in an arbitrarily small neighbourhood of $x$.

In the next section, we gain insight into the posterior mean by analysing its coefficients as an $L^2$ projection. In the third section, we study the bias and variance of the posterior mean as a frequentist estimator, as well as the posterior variance. Combining these results, we arrive at our main theorem. Throughout, we use $A \lesssim B$ to mean $A \leq cB$ and $A \asymp B$ to mean $A \lesssim B$ and $B \lesssim A$.

## 1.2 Understanding the posterior mean

In order to be able to analyse credible sets, we will need to know more about the posterior mean $\hat{f}_n(x) = \mathrm{E}(W_x \mid Y_{1,n}, \dots, Y_{n,n})$. Since conditional expectations correspond to $L^2$ projections, we may write

$$\mathrm{E}(W_x \mid Y_{1,n}, \dots, Y_{n,n}) = \sum_{i=1}^{n} a_i^n Y_{i,n} = \vec{Y}_n^T \vec{a}_n,$$

where the $(a_i^n)$ are coefficients in $\mathbb{R}$. Our aim in this section is to study the asymptotic behaviour of these coefficients. We require the following technical lemma:

**Lemma 1.2.** *Let* $\lambda_\pm = 1 + \frac{1}{2k} \pm \frac{1}{2\sqrt{k}}\sqrt{4 + \frac{1}{k}}$ *for* $k \in \mathbb{R}$. *Then* $\lambda_+\lambda_- = 1$ *and* $\lambda_- \uparrow 1$ *and* $\lambda_+ \downarrow 1$ *as* $k \to \infty$. *Furthermore, for each* $\gamma \in \mathbb{R}$, *there exists* $C_\gamma > 0$ *such that* $\lambda_+^{k^\gamma} \geq e^{C_\gamma k^{\gamma - 1/2}}$.

*Proof.* Since $\log z \geq \frac{1}{2}(z-1)$ for $z \in [1,2]$, we have

$$k^\gamma \log \lambda_+ \gtrsim k^\gamma \left( \frac{1}{2k} + \frac{1}{2\sqrt{k}}\sqrt{4 + \frac{1}{k}} \right) \gtrsim k^{\gamma - 1/2}. \qquad \square$$

In the following, we will use the rescaled index $k = n_+/c_n = n_+^{1-\beta}$, since this turns out to be computationally convenient. Note that we have $c_n = k^{\frac{\beta}{1-\beta}}$. We will study asymptotics of the sequence $(a_i^n)$ in terms of $k$.

Let $i_n = \max\{i : i/n_+ < x\}$, the index $i$ such that $x_{i,n} < x$ is closest to $x$. The following theorem shows that coefficients $a_i$, where $i$ is far from $i_n$, tend to zero exponentially fast. Specifically, applying the above lemma, it can be seen that $a_i^n$ tends to zero exponentially fast if $|i - i_n| \gg \sqrt{k}$.

**Theorem 1.3.** *We have*

$$2\sqrt{k}\, a_i^n = \begin{cases} \lambda_+^{-i_n}\left[A_n \lambda_+^{i} - B_n \lambda_+^{-i}\right] & \text{for } i \leq i_n \\ \lambda_+^{i_n}\left[\tilde{A}_n \lambda_+^{-i+1} + \tilde{B}_n \lambda_+^{-2n+i-1}\right] & \text{for } i \geq i_n, \end{cases}$$

*where* $A_n, B_n, \tilde{A}_n, \tilde{B}_n \to 1$. *In particular,* $a_i^n \geq 0$.

*Proof.* The coefficients $a_i^n$ satisfy the projection relations

$$0 = \left\langle W_x - \vec{Y}_n^T \vec{a}_n, Y_{i,n} \right\rangle = \left\langle W_x - \vec{Y}_n^T \vec{a}_n, W_{x_{i,n}} \right\rangle - a_i^n.$$

Expanding this yields

$$\begin{aligned}
0 &= \operatorname{cov}(W_x, W_{x_{i,n}}) - \sum_{j=1}^{n} a_j^n \operatorname{cov}(W_{x_{j,n}}, W_{x_{i,n}}) - a_i^n \\
&= c_n \left( x \wedge \frac{i}{n_+} - \sum_{j=1}^{i} \frac{j}{n_+} a_j^n - \frac{i}{n_+} \sum_{j=i+1}^{n} a_i^n \right) - a_i^n .
\end{aligned}$$

These equations can be written in matrix form:

$$c_n \begin{pmatrix}
\frac{1}{c_n} + \frac{1}{n_+} & \frac{1}{n_+} & \frac{1}{n_+} & \cdots & \frac{1}{n_+} \\
\frac{1}{n_+} & \frac{1}{c_n} + \frac{2}{n_+} & \frac{2}{n_+} & \cdots & \frac{2}{n_+} \\
\frac{1}{n_+} & \frac{2}{n_+} & \ddots & & \vdots \\
\vdots & \vdots & & \frac{1}{c_n} + \frac{n-1}{n_+} & \frac{n-1}{n_+} \\
\frac{1}{n_+} & \frac{2}{n_+} & \cdots & \frac{n-1}{n_+} & \frac{1}{c_n} + \frac{n}{n_+}
\end{pmatrix}
\begin{pmatrix} a_1^n \\ a_2^n \\ \vdots \\ \\ \vdots \\ a_n^n \end{pmatrix}
= c_n \begin{pmatrix} \frac{1}{n_+} \\ \vdots \\ \frac{i_n}{n_+} \\ x \\ \vdots \\ x \end{pmatrix} . \tag{1.2}$$

Using $\frac{c_n}{n_+} = \frac{1}{k}$ and applying elementary matrix operations, we obtain

$$\begin{pmatrix}
1 + \frac{1}{k} & \frac{1}{k} & \frac{1}{k} & \cdots & \frac{1}{k} \\
-1 & 1 + \frac{1}{k} & \frac{1}{k} & \cdots & \frac{1}{k} \\
0 & -1 & \ddots & & \vdots \\
\vdots & \vdots & & 1 + \frac{1}{k} & \frac{1}{k} \\
0 & 0 & \cdots & -1 & 1 + \frac{1}{k}
\end{pmatrix}
\begin{pmatrix} a_1^n \\ a_2^n \\ \vdots \\ \\ \vdots \\ a_n^n \end{pmatrix}
= c_n \begin{pmatrix} \frac{1}{n_+} \\ \vdots \\ \frac{1}{n_+} \\ x - x_{i_n,n} \\ 0 \\ \vdots \\ 0 \end{pmatrix} ,$$

which can be further simplified to

$$\begin{pmatrix}
2 + \frac{1}{k} & -1 & 0 & \cdots & 0 \\
-1 & 2 + \frac{1}{k} & -1 & \cdots & 0 \\
0 & -1 & \ddots & & \vdots \\
\vdots & \vdots & & 2 + \frac{1}{k} & -1 \\
0 & 0 & \cdots & -1 & 1 + \frac{1}{k}
\end{pmatrix}
\begin{pmatrix} a_1^n \\ a_2^n \\ \vdots \\ \\ \vdots \\ a_n^n \end{pmatrix}
= c_n \begin{pmatrix} 0 \\ \vdots \\ 0 \\ x_{i_n+1,n} - x \\ x - x_{i_n,n} \\ 0 \\ \vdots \\ 0 \end{pmatrix} .$$

For $i \in \{3, \ldots, i_n\}$ and $i \in \{i_n + 3, \ldots, n\}$ we obtain the recurrence relation

$$a_i^n = \big(2 + \tfrac{1}{k}\big) a_{i-1}^n - a_{i-2}^n , \tag{1.3}$$

while the middle two rows yield

$$-a^n_{i_n-1} + \big(2+\tfrac{1}{k}\big)a^n_{i_n} - a^n_{i_n+1} = c_n(x_{i_n+1,n} - x), \tag{1.4a}$$
$$-a^n_{i_n} + \big(2+\tfrac{1}{k}\big)a^n_{i_n+1} - a^n_{i_n+2} = c_n(x - x_{i_n,n}). \tag{1.4b}$$

The recurrence (1.3) has characteristic polynomial $\lambda^2 - \big(2+\frac{1}{k}\big)\lambda + 1$ and hence we have for $i \in \{1,\dots,i_n\}$ the general solution

$$a^n_i = A\lambda^i_+ + B\lambda^i_-, \quad \lambda_\pm = \frac{1}{2}\left(2 + \frac{1}{k} \pm \sqrt{\frac{1}{k^2} + \frac{4}{k}}\right) = 1 + \frac{1}{2k} \pm \frac{1}{2\sqrt{k}}\sqrt{4 + \frac{1}{k}}.$$

Using the first row, we have

$$\begin{aligned} A\lambda_+ + B\lambda_- &= a^n_1, \\ A\lambda^2_+ + B\lambda^2_- &= a^n_2 = \big(2+\tfrac{1}{k}\big)a^n_1, \end{aligned}$$

from which we find

$$A = a^n_1 \frac{\frac{1}{k} + 2 - \lambda_-}{(\lambda_+ - \lambda_-)\lambda_+} =: a^n_1\alpha, \qquad B = a^n_1 \frac{\lambda_+ - 2 - \frac{1}{k}}{(\lambda_+ - \lambda_-)\lambda_-} =: a^n_1\beta.$$

Note that $\beta = -\alpha$. Using (1.4a), we find

$$\begin{aligned} &c_n(x_{i_n+1,n} - x) + a^n_{i_n+1} \\ &\quad = -a^n_{i_n-1} + \big(2+\tfrac{1}{k}\big)a^n_{i_n} \\ &\quad = A\lambda^{i_n-1}_+\big(\big(2+\tfrac{1}{k}\big)\lambda_+ - 1\big) + B\lambda^{i_n-1}_-\big(\big(2+\tfrac{1}{k}\big)\lambda_- - 1\big). \end{aligned}$$

This yields

$$\begin{aligned} a^n_1 &= \frac{c_n(x_{i_n+1,n} - x) + a^n_{i_n+1}}{\alpha\lambda^{i_n-1}_+\big(\big(2+\frac{1}{k}\big)\lambda_+ - 1\big) + \beta\lambda^{i_n-1}_-\big(\big(2+\frac{1}{k}\big)\lambda_- - 1\big)} \\ &=: \frac{c_n(x_{i_n+1,n} - x) + a^n_{i_n+1}}{D}. \end{aligned} \tag{1.5}$$

For $i \in \{i_n+1,\dots,n-2\}$, we have

$$a^n_i = \big(2+\tfrac{1}{k}\big)a^n_{i+1} - a^n_{i+2}.$$

Writing $b^n_i = a^n_{n-i+1}$, we see that for $i \in \{3,\dots,n-i_n\}$ we again have the recurrence

$$b^n_i = \big(2+\tfrac{1}{k}\big)b^n_{i-1} - b^n_{i-2},$$

which has the same solution $b^n_i = \tilde{A}\lambda^i_+ + \tilde{B}\lambda^i_-$ for $i \in \{1,\dots,n-i_n\}$, where

$$\tilde{A} = b^n_1 \frac{\frac{1}{k} + 1 - \lambda_-}{(\lambda_+ - \lambda_-)\lambda_+} =: b^n_1\tilde{\alpha}, \qquad \tilde{B} = b^n_1 \frac{\lambda_+ - 1 - \frac{1}{k}}{(\lambda_+ - \lambda_-)\lambda_-} =: b^n_1\tilde{\beta}.$$

We apply (1.4b) to find

$$\begin{aligned} b_1^n &= \frac{c_n(x - x_{i_n,n}) + b^n_{n-i_n+1}}{\tilde{\alpha}\lambda_+^{n-i_n-1}\big((2+\frac{1}{k})\lambda_+ - 1\big) + \tilde{\beta}\lambda_-^{n-i_n-1}\big((2+\frac{1}{k})\lambda_- - 1\big)} \\ &=: \frac{c_n(x - x_{i_n,n}) + b^n_{n-i_n+1}}{\tilde{D}}. \end{aligned}$$

Note that $b^n_{n-i_n+1} = a^n_{i_n} = a_1^n(\alpha\lambda_+^{i_n} + \beta\lambda_-^{i_n})$. We substitute this in the above and similarly we substitute $a^n_{i_n+1} = b^n_{n-i_n} = b_1^n(\tilde{\alpha}\lambda_+^{n-i_n} + \tilde{\beta}\lambda_-^{n-i_n})$ in (1.5). This yields two linear equations in $a_1^n$ and $b_1^n$. Solving for $b_1^n$, we obtain

$$b_1^n = c_n \frac{D\left(x - x_{i_n,n} + \frac{x_{i_n+1,n}-x}{D}(\alpha\lambda_+^{i_n} + \beta\lambda_-^{i_n})\right)}{D\tilde{D} - (\alpha\lambda_+^{i_n} + \beta\lambda_-^{i_n})(\tilde{\alpha}\lambda_+^{n-i_n} + \tilde{\beta}\lambda_-^{n-i_n})}. \tag{1.6}$$

Our aim is to determine the asymptotic behaviour of $b_1^n$ and $a_1^n$. Note that $\lambda_\pm \to 1$ and $\lambda_+ - \lambda_- = \frac{1}{\sqrt{k}}\sqrt{4 + \frac{1}{k}} \sim \frac{2}{\sqrt{k}}$, hence $\alpha \sim \frac{1}{2}\sqrt{k}$, $\beta \sim -\frac{1}{2}\sqrt{k}$ and $\tilde{\alpha} \sim \tilde{\beta} \sim \frac{1}{2}$. Furthermore, applying Lemma 1.2, we see that

$$D \sim \tfrac{1}{2}\sqrt{k}\lambda_+^{i_n}, \qquad \tilde{D} \sim \tfrac{1}{2}\lambda_+^{n-i_n},$$

where we ignore the exponentially small terms involving $\lambda_-$. Similarly, we have $\frac{\alpha\lambda_+^{i_n}+\beta\lambda_-^{i_n}}{D} = 1 + O(1/\sqrt{k})$, since

$$\begin{aligned} \frac{\alpha\lambda_+^{i_n} + \beta\lambda_-^{i_n}}{D} - 1 &= \frac{\alpha\lambda_+^{i_n} + \beta\lambda_-^{i_n} - D}{D} \\ &= \frac{\alpha\lambda_+^{i_n-1}\big(\lambda_+ - (2+\frac{1}{k})\lambda_+ + 1\big) + \cdots}{D} \sim -\frac{1}{\sqrt{k}}. \end{aligned}$$

Finally, it can be seen that

$$\begin{aligned} D\tilde{D} - &(\alpha\lambda_+^{i_n} + \beta\lambda_-^{i_n})(\tilde{\alpha}\lambda_+^{n-i_n} + \tilde{\beta}\lambda_-^{n-i_n}) \\ &= \alpha\tilde{\alpha}\lambda_+^{n-2}\Big(\big[(2+\tfrac{1}{k})\lambda_+ - 1\big]^2 - \lambda_+^2\Big) + \cdots \sim \tfrac{1}{2}\lambda_+^n. \end{aligned}$$

From this we conclude $b_1^n \sim \frac{1}{\sqrt{k}}\lambda_+^{-(n-i_n)}$. Additionally, we find

$$\begin{aligned} a_1^n &= \frac{c_n(x_{i_n+1,n} - x) + b_1^n(\tilde{\alpha}\lambda_+^{n-i_n} + \tilde{\beta}\lambda_-^{n-i_n})}{D} \sim \frac{\frac{1}{\sqrt{k}}\lambda_+^{-(n-i_n)}\frac{1}{2}\lambda_+^{n-i_n}}{\frac{1}{2}\sqrt{k}\lambda_+^{i_n}} \\ &= \tfrac{1}{k}\lambda_+^{-i_n}. \end{aligned}$$

The result follows. □

This theorem gives us insight into the behaviour of the coefficients $(a_i^n)$. We can further see that they asymptotically sum to one:

**Corollary 1.4.** *We have* $1 - \sum_{i=1}^n a_i^n \sim \lambda_+^{-i_n}$.

*Proof.* By the first row of (1.2), we have

$$\frac{1}{k}\sum_{i=1}^n a_i^n = \frac{1}{k} - a_1^n \quad \Rightarrow \quad \sum_{i=1}^n a_i^n = 1 - k a_1^n. \qquad \square$$

For the further analysis in the next section, a technical result is useful.

**Lemma 1.5.** *Let*

$$\Pi(r,m) = \sum_{i=1}^m \lambda_+^i (r-i)^\alpha$$

*where* $r \geq m$. *Then*

$$\begin{aligned}\lambda_+^{-1}\int_{(r-m)\log\lambda_+}^{(r-1)\log\lambda_+} e^{-v}v^\alpha\, dv &\leq (\log\lambda_+)^{\alpha+1}\lambda_+^{-r}\Pi(r,m)\\ &\leq \lambda_+\int_0^{(r+1)\log\lambda_+} e^{-v}v^\alpha\, dv.\end{aligned}$$

*Proof.* Note that $\lambda_+^t(r-t+1)^\alpha \geq \lambda_+^i(r-i)^\alpha$ for $t \in [i, i+1]$, hence

$$\begin{aligned}\sum_{i=1}^m \lambda_+^i(r-i)^\alpha &\leq \int_1^{m+1}\lambda_+^t(r-t+1)^\alpha\,\mathrm{d}t \leq \lambda_+^{r+1}\int_0^{r+1} e^{-u\log\lambda_+}u^\alpha\,\mathrm{d}u\\ &= (\log\lambda_+)^{-\alpha-1}\lambda_+^{r+1}\int_0^{(r+1)\log\lambda_+} e^{-v}v^\alpha\,\mathrm{d}v.\end{aligned}$$

Similarly, since $\lambda_+^t(r-t-1)^\alpha \leq \lambda_+^i(r-i)^\alpha$ for $t \in [i-1, i]$, we see that

$$\sum_{i=1}^m \lambda_+^i(r-i)^\alpha \geq (\log\lambda_+)^{-\alpha-1}\lambda_+^{r-1}\int_{(r-m)\log\lambda_+}^{(r-1)\log\lambda_+} e^{-v}v^\alpha\,\mathrm{d}v. \qquad \square$$

## 1.3 Bias and posterior variance

In this section, we will study the bias and variance of the posterior distribution using Theorem 1.3. In the proofs, we will encounter the coefficients $A, B, \tilde{A}$ and $\tilde{B}$ as seen in the proof of Theorem 1.3, given by $A = \frac{1}{2\sqrt{k}}\lambda_+^{-i_n}A_n$, $B =$

$-\frac{1}{2\sqrt{k}}\lambda_+^{-i_n} B_n$, $\tilde{A} = \frac{1}{2\sqrt{k}}\lambda_+^{-(n-i_n)}\tilde{A}_n$ and $\tilde{B} = \frac{1}{2\sqrt{k}}\lambda_+^{-(n-i_n)}\tilde{B}_n$. For the most part, we will ignore terms involving $B_n$ and $\tilde{B}_n$, since they are exponentially small. Indeed, setting $\gamma = \frac{1}{2}\left(\frac{1}{1-\beta} - \frac{1}{2}\right)$, we can apply Lemma 1.2 to see that all terms involving these quantities are $O\left(e^{-k^\gamma}\right)$.

To understand the behaviour of credible sets for this model, we will consider first the bias $\mu_n(f) := \mathrm{E}\hat{f}_n(x) - f(x)$. Note that

$$\mathrm{E}\hat{f}_n(x) = \sum_{i=1}^n a_i^n f(x_{i,n}).$$

Applying the above result, we obtain

**Corollary 1.6.** *For $f(t) = |x-t|^\alpha$, we have $\mu_n(f) \sim \Gamma(\alpha+1)k^{\left(\frac{1}{2}-\frac{1}{1-\beta}\right)\alpha}$.*

*Proof.* Applying Lemma 1.5, we have

$$\begin{aligned}
\mu_n(f) &= \sum_{i=1}^n a_i^n f(x_{i,n}) \\
&= A\sum_{i=1}^{i_n} \lambda_+^i (x - x_{i,n})^\alpha + \tilde{A}\sum_{i=1}^{n-i_n} \lambda_+^i (x_{n-i+1,n} - x)^\alpha + O\big(e^{-k^\gamma}\big) \\
&= A n_+^{-\alpha}\sum_{i=1}^{i_n}\lambda_+^i (n_+ x - i)^\alpha + \tilde{A} n_+^{-\alpha}\sum_{i=1}^{n-i_n}\lambda_+^i\left[n_+(1-x) - i + \tfrac{1}{2}\right]^\alpha + O\big(e^{-k^\gamma}\big) \\
&= A n_+^{-\alpha}\Pi(n_+ x, i_n) + \tilde{A} n_+^{-\alpha}\Pi\big(n_+(1-x) + \tfrac{1}{2}, n - i_n\big) + O\big(e^{-k^\gamma}\big) \\
&\sim \frac{1}{\sqrt{k}}(\log\lambda_+)^{-\alpha-1} n_+^{-\alpha}\Gamma(\alpha+1) \sim \Gamma(\alpha+1)k^{\left(\frac{1}{2}-\frac{1}{1-\beta}\right)\alpha}.
\end{aligned}$$

This gives the desired result. □

Using this specific result, we see that more generally we have

**Corollary 1.7.** *Let $f$ be Hölder continuous of order $\alpha \in (0,2]$ at $x$. Then $|\mu_n(f)| \lesssim k^{\left(\frac{1}{2}-\frac{1}{1-\beta}\right)\alpha}$.*

*Proof.* Let $\delta_n = k^{-\epsilon}$ and write

$$\mu_n(f) = \sum_{|x_{i,n}-x|>\delta_n} a_i^n f(x_{i,n}) + \sum_{|x_{i,n}-x|\le\delta_n} a_i^n f(x_{i,n}) - f(x) =: \gamma_n(f) + \tilde{\mu}_n(f).$$

Note that $|\gamma_n(f)| \leq \|f\| \sum_{|x_{i,n}-x|>\delta_n} a_i^n$. By Theorem 1.3, we have $a_i^n \lesssim A\lambda_+^{n_+(x-\delta_n)}$ for $i < n_+(x-\delta_n)$ and $a_i^n \lesssim (\tilde{A}+\tilde{B})\lambda_+^{n_+(1-x-\delta_n)}$ for $i > n_+(x+\delta_n)$. Hence

$$\begin{aligned}\sum_{|x_{i,n}-x|>\delta_n} a_i^n &\lesssim nA\lambda_+^{n_+(x-\delta_n)} + n(\tilde{A}+\tilde{B})\lambda_+^{n_+(1-x-\delta_n)} \\ &\lesssim c_n\sqrt{k}\lambda_+^{-n_+\delta_n} = c_n\sqrt{k}\lambda_+^{-k^{\frac{1}{1-\beta}-\epsilon}},\end{aligned}$$

which is exponentially small if $\epsilon < \frac{1}{1-\beta} - \frac{1}{2}$ by Lemma 1.2. Finally, by Corollary 1.4 we have $\sum_{|x_{i,n}-x|\leq\delta_n} a_i^n \approx 1 - \sum_{|x_{i,n}-x|>\delta_n} a_i^n \approx 1$ and hence for $\alpha \in (0,1)$ we have

$$\begin{aligned}|\tilde{\mu}_n(f)| = \left|\sum_{|x_{i,n}-x|\leq\delta_n} a_i^n f(x_{i,n}) - f(x)\right| &\lesssim \sum_{|x_{i,n}-x|\leq\delta_n} a_i^n |f(x_{i,n}) - f(x)| \\ \lesssim \sum_{|x_{i,n}-x|\leq\delta_n} a_i^n |x_{i,n}-x|^\alpha &\leq \sum_{i=1}^n a_i^n |x_{i,n}-x|^\alpha = \mu_n(g_\alpha),\end{aligned}$$

where $g_\alpha(t) = |t-x|^\alpha$. The result now follows for $\alpha \in (0,1)$ by applying Corollary 1.6. For $\alpha \in [1,2]$, we give a refinement of the characterisation of the coefficients $a_i^n$. Consider

$$\begin{aligned}\frac{A}{\tilde{A}} = \frac{\alpha a_1^n}{\tilde{\alpha} b_1^n} &= \frac{\alpha\big(c_n(x_{i_n+1,n}-x) + b_1^n(\tilde{\alpha}\lambda_+^{n-i_n} + \tilde{\beta}\lambda_-^{n-i_n})\big)}{D\tilde{\alpha}b_1^n} \\ &= \frac{\alpha c_n(x_{i_n+1,n}-x)}{D\tilde{\alpha}b_1^n} + \frac{\alpha b_1^n(\tilde{\alpha}\lambda_+^{n-i_n} + \tilde{\beta}\lambda_-^{n-i_n})}{D\tilde{\alpha}b_1^n}.\end{aligned}$$

We have

$$D = \alpha\lambda_+^{i_n+1} + \beta\lambda_-^{i_n+1}$$

since $\left(2+\frac{1}{k}\right)\lambda_\pm - 1 = \lambda_\pm^2$ by the characteristic equation. Hence the second term is equal to

$$\frac{\alpha\tilde{\alpha}\lambda_+^{n-i_n}\left(1+O\left(e^{-k^\gamma}\right)\right)}{\tilde{\alpha}\alpha\lambda_+^{i_n+1}\left(1+O\left(e^{-k^\gamma}\right)\right)} = \lambda_+^{n-2i_n-1}\Big(1+O\big(e^{-k^\gamma}\big)\Big).$$

Now consider the first term. We have

$$\frac{\alpha c_n(x_{i_n+1,n}-x)}{D\tilde{\alpha}b_1^n} \lesssim \frac{\sqrt{k}\frac{1}{k}}{\sqrt{k}\lambda_+^{i_n+1}\frac{1}{\sqrt{k}}\lambda_+^{-(n-i_n)}} = \frac{1}{\sqrt{k}}\lambda_+^{n-2i_n-1}.$$

We see that

$$\frac{A}{\tilde{A}} = \lambda_+^{n-2i_n-1}\left(1+O\left(\frac{1}{\sqrt{k}}\right)\right).$$

Using this, we see that we can bound

$$\sum_{i=1}^{n} a_i^n(x_{i,n}-x) = A\sum_{i=1}^{i_n}\lambda_+^i(x_{i,n}-x)+\tilde{A}\sum_{i=1}^{n-i_n}\lambda_+^i(x_{n-i+1,n}-x)+O\big(e^{-k^\gamma}\big)$$

by a constant times

$$\begin{aligned}
&\tilde{A}\lambda_+^{n-2i_n-1}\left(1+O\left(\frac{1}{\sqrt{k}}\right)\right)\sum_{i=1}^{i_n}\lambda_+^i(x_{i,n}-x)+\tilde{A}\sum_{i=1}^{n-i_n}\lambda_+^i(x_{n-i+1,n}-x)\\
&\lesssim \frac{1}{\sqrt{k}}\left(\sum_{i=1}^{i_n}\lambda_+^{i-i_n-1}(x_{i,n}-x)+\sum_{i=i_n+1}^{n}\lambda_+^{i_n-i+1}(x_{i,n}-x)\right)\\
&\quad+\tilde{A}\lambda_+^{n-2i_n-1}\frac{1}{\sqrt{k}}\sum_{i=1}^{i_n}\lambda_+^i(x_{i,n}-x)\\
&\lesssim \frac{1}{\sqrt{k}}\left(\sum_{j=1}^{i_n}\lambda_+^{-j}(x_{i_n-j+1,n}-x)+\sum_{j=0}^{n-i_n-1}\lambda_+^{-j}(x_{i_n+j+1,n}-x)\right)\\
&\quad+\lambda_+^{-i_n-1}\frac{1}{k}\sum_{i=1}^{i_n}\lambda_+^i(x_{i,n}-x).
\end{aligned}$$

Note that the absolute value of the second term is given by

$$\lambda_+^{-i_n-1}\frac{1}{k}\frac{1}{n_+}\sum_{i=1}^{i_n}\lambda_+^i(n_+x-i)=\lambda_+^{-i_n-1}\frac{1}{k}\frac{1}{n_+}\Pi(n_+x,i_n)\lesssim\frac{1}{n}.$$

Now suppose that $i_n \le n-i_n-1$ (the case $n-i_n-1\le i_n$ is similar). Then the first term is equal to

$$\begin{aligned}
&\frac{1}{\sqrt{k}}\left(2\sum_{j=1}^{i_n}\lambda_+^{-j}\left(\frac{i_n+1}{n_+}-x\right)+(x_{i_n+1,n}-x)+\sum_{j=i_n+1}^{n-i_n-1}\lambda_+^{-j}(x_{i_n+j+1,n}-x)\right)\\
&\asymp\frac{1}{\sqrt{k}}\left(\frac{1}{n}\frac{1-\lambda_+^{-i_n}}{1-\lambda_+^{-1}}+\frac{1}{n}+O\big(e^{-k^\gamma}\big)\right)\asymp\frac{1}{\sqrt{k}}\left(\frac{\sqrt{k}}{n}+\frac{1}{n}+O\big(e^{-k^\gamma}\big)\right)\asymp\frac{1}{n}.
\end{aligned}$$

We conclude that

$$\sum_{i=1}^{n}a_i^n(x_{i,n}-x)\lesssim\frac{1}{n}.$$

Since $\sum_{|x_{i,n}-x|>\delta_n} a_i^n$ is exponentially small, we then also see that $\sum_{|x_{i,n}-x|\leq\delta_n} a_i^n(x_{i,n}-x) \lesssim \frac{1}{n}$. Hence if $f \in C^\alpha$ where $\alpha = 1+\delta$ for some $\delta \in (0,1)$, then

$$\begin{aligned}
&\left|\sum_{|x_{i,n}-x|\leq\delta_n} a_i^n\big(f(x_{i,n})-f(x)\big)\right| \\
&= \left|\sum_{|x_{i,n}-x|\leq\delta_n} a_i^n\big(f'(\xi_i)-f'(x)\big)(x_{i,n}-x) + f'(x)\sum_{|x_{i,n}-x|\leq\delta_n} a_i^n(x_{i,n}-x)\right| \\
&\lesssim \sum_{|x_{i,n}-x|\leq\delta_n} a_i^n|\xi_i-x|^\delta|x_{i,n}-x| + \frac{1}{n} \\
&\leq \sum_{|x_{i,n}-x|\leq\delta_n} a_i^n|x_{i,n}-x|^{1+\delta} + \frac{1}{n} \lesssim k^{\left(\frac{1}{2}-\frac{1}{1-\beta}\right)(1+\delta)} + \frac{1}{n}.
\end{aligned}$$

If $\beta \leq 0$, then this is smaller than $k^{\left(\frac{1}{2}-\frac{1}{1-\beta}\right)(1+\delta)}$ for all $\delta \in (0,1)$. Note that the result also holds for $\delta = 0$ and $\delta = 1$ (for $\delta = 1$ we can repeat the application of the mean value theorem above and obtain a second derivative). We see that the result also follows for $\alpha \in [1,2]$. □

In order to derive results on the coverage, we will also need results on the posterior variance and the variance of the posterior mean.

**Lemma 1.8.** *The variance of the posterior mean*

$$t_n^2 := \mathrm{var}_f \sum_{i=1}^n a_i^n Y_{i,n} = \sum_{i=1}^n (a_i^n)^2$$

*satisfies* $t_n^2 \sim \frac{1}{4\sqrt{k}}$.

*Proof.* We have

$$t_n^2 = A^2 \sum_{i=1}^{i_n} \lambda_+^{2i} + \tilde{A}^2 \sum_{i=1}^{n-i_n} \lambda_+^{2i} + O\big(e^{-k^\gamma}\big).$$

Now considering $m = \lfloor sn \rfloor$ for some $s \in (0,1)$, we see that

$$\sum_{i=1}^m \lambda_+^{2i} = \frac{\lambda_+^{2m+2}-\lambda_+^2}{\lambda_+^2-1} \sim \frac{\sqrt{k}}{2}\lambda_+^{2m}, \tag{1.7}$$

since $\lambda_+^2 = 1 + \frac{1}{\sqrt{k}}\sqrt{4+\frac{1}{k}} + O\big(\frac{1}{k}\big)$ and the other term is negligible by Lemma 1.2. We conclude that

$$t_n^2 \sim \frac{1}{4k}\lambda_+^{-2i_n}\frac{\sqrt{k}}{2}\lambda_+^{2i_n} + \frac{1}{4k}\lambda_+^{-2(n-i_n)}\frac{\sqrt{k}}{2}\lambda_+^{2(n-i_n)} \sim \frac{1}{4\sqrt{k}},$$

as desired. □

The analysis of the posterior variance $\sigma_n^2$ is more involved. Recall that

$$\sigma_n^2 := \mathrm{E}\big[\big(W_x - \hat{f}_n(x)\big)^2 \mid Y\big].$$

Note that $W_x - \hat{f}_n(x)$ is $L^2$-orthogonal to $Y$, since $\hat{f}_n(x)$ is the projection of $W_x$ onto $Y$. Since all quantities involved have multivariate mean-zero normal distributions, it follows that $W_x - \hat{f}_n(x)$ is independent of $Y$. We conclude

$$\begin{aligned}\sigma_n^2 = \mathrm{E}\big[W_x - \hat{f}_n(x)\big]^2 &= \mathrm{E}\left[W_x - \sum_{i=1}^n a_i^n(W_{x_{i,n}} + \varepsilon_{i,n})\right]^2 \\ &= \mathrm{E}\left[W_x - \sum_{i=1}^n a_i^n W_{x_{i,n}}\right]^2 + \mathrm{E}\left[\sum_{i=1}^n a_i^n \varepsilon_{i,n}\right]^2 = s_n^2 + t_n^2,\end{aligned}$$

where the last equality defines $s_n^2$. To determine the behaviour of $s_n^2$, we require the following result:

**Proposition 1.9.** *Let $m \in \mathbb{N}$ and $s \geq x_{m,n}$ such that $|s - x_{m,n}| = o(1/\sqrt{k})$ and*

$$\Lambda(s,m) = \sum_{i=1}^m \lambda_+^i (W_s - W_{x_{i,n}}).$$

*Then $E\Lambda(s,m)^2 \sim \frac{1}{2}\sqrt{k}\lambda_+^{2m}$.*

*Proof.* We may write $W_s - W_{x_{i,n}} = (W_s - W_{x_{m,n}}) + \sum_{j=i+1}^m V_j$, where $(W_s - W_{x_{m,n}})$ and $V_j = W_{x_{j,n}} - W_{x_{j-1,n}} \sim \mathcal{N}\big(0, \frac{1}{k}\big)$ are independent random variables. Then

$$\begin{aligned}\mathrm{E}\Lambda(s,m)^2 &= \mathrm{var}\left[(W_s - W_{x_{m,n}})\sum_{i=1}^m \lambda_+^i\right] + \mathrm{var}\sum_{i=1}^m \lambda_+^i \sum_{j=i+1}^m V_j \\ &= \mathrm{var}\sum_{j=1}^m V_j \sum_{i=1}^{j-1}\lambda_+^i + o\big(\sqrt{k}\lambda_+^{2m}\big) \\ &= \frac{1}{k}\sum_{j=1}^m \left(\sum_{i=1}^{j-1}\lambda_+^i\right)^2 + o\big(\sqrt{k}\lambda_+^{2m}\big),\end{aligned}$$

where we use that $\sum_{i=1}^m \lambda_+^i \sim \sqrt{k}\lambda_+^m$ by a similar argument as applied in (1.7). We have

$$\left(\sum_{i=1}^{j-1}\lambda_+^i\right)^2 = \left(\frac{\lambda_+^j - \lambda_+}{\lambda_+ - 1}\right)^2 = \frac{\lambda_+^{2j} - 2\lambda_+^{j+1} + \lambda_+^2}{(\lambda_+ - 1)^2}$$

Summing over $j$, we obtain

$$\begin{aligned} \mathrm{E}\Lambda(s,m)^2 &= \frac{1}{k(\lambda_+ - 1)^2}\left(\frac{\lambda_+^{2m+2} - \lambda_+^2}{\lambda_+^2 - 1} - 2\frac{\lambda_+^{m+2} - \lambda_+^2}{\lambda_+ - 1}\right) + o\big(\sqrt{k}\lambda_+^{2m}\big) \\ &= \frac{\lambda_+^{2m+2}}{k(\lambda_+ - 1)^2(\lambda_+^2 - 1)} + o\big(\sqrt{k}\lambda_+^{2m}\big) \sim \tfrac{1}{2}\sqrt{k}\lambda_+^{2m} \end{aligned}$$

as desired. □

Note that the condition $|s - x_{m,n}| = o(1/\sqrt{k})$ is certainly satisfied if $|s - x_{m,n}| = O(1/n)$. We are now able to conclude our analysis of $s_n^2$.

**Corollary 1.10.** *We have*

$$s_n^2 = E\left(W_x - \sum_{i=1}^n a_i^n W_{x_{i,n}}\right)^2 \sim \frac{1}{4\sqrt{k}}.$$

*Proof.* Note that

$$s_n^2 = \mathrm{E}\left(\sum_{i=1}^n a_i^n (W_x - W_{x_{i,n}})\right)^2 + O\big(e^{-k^\gamma}\big),$$

since $\sum_{i=1}^n a_i^n - 1$ is exponentially small by Corollary 1.4. Now we use the fact that Brownian motion has independent increments to write

$$\begin{aligned} \mathrm{E}\left(\sum_{i=1}^n a_i^n (W_x - W_{x_{i,n}})\right)^2 &= \mathrm{E}\left(\sum_{i=1}^{i_n} a_i^n (W_x - W_{x_{i,n}})\right)^2 \\ &\quad + \mathrm{E}\left(\sum_{i=i_n+1}^{n} a_i^n (W_x - W_{x_{i,n}})\right)^2. \end{aligned}$$

Consider the first term: using Proposition 1.9 we obtain

$$\begin{aligned} \mathrm{E}\left(\sum_{i=1}^{i_n} a_i^n (W_x - W_{x_{i,n}})\right)^2 &= A^2\mathrm{E}\left(\sum_{i=1}^{i_n} \lambda_+^i (W_x - W_{x_{i,n}})\right)^2 + O\big(e^{-k^\gamma}\big) \\ &= A^2\mathrm{E}\Lambda(x,i_n)^2 + O\big(e^{-k^\gamma}\big) \sim \frac{1}{8\sqrt{k}}. \end{aligned}$$

For the second term, we have

$$\begin{aligned}
\mathrm{E}&\left(\sum_{i=i_n+1}^{n} a_i^n (W_x - W_{x_{i,n}})\right)^2 \\
&= \tilde{A}^2 \mathrm{E}\left(\sum_{i=1}^{n-i_n} \lambda_+^i (W_x - W_{x_{n-i+1,n}})\right)^2 + O\big(e^{-k^\gamma}\big) \\
&= \tilde{A}^2 \mathrm{E}\left(\sum_{i=1}^{n-i_n} \lambda_+^i (W_{1-x+\frac{1}{2n_+}} - W_{x_{i,n}})\right)^2 + O\big(e^{-k^\gamma}\big) \\
&= \tilde{A}^2 \mathrm{E}\Lambda\big(1-x+\tfrac{1}{2n_+}, n-i_n\big)^2 + O\big(e^{-k^\gamma}\big) \sim \frac{1}{8\sqrt{k}},
\end{aligned}$$

where in the second equality we use the fact that the process $(W_{1+\frac{1}{2n_+}} - W_{1+\frac{1}{2n_+}-t})_{t\in[0,1]}$ is again a Brownian motion on $[0,1]$. The result follows. □

Combining the above results, we understand the posterior variance.

**Corollary 1.11.** *The posterior variance $\sigma_n^2 = s_n^2 + t_n^2$ satisfies $\sigma_n^2 \sim \frac{1}{2\sqrt{k}}$.*

Finally, we arrive at the proof of our main result.

*Proof of Theorem 1.1.* We have

$$P_f\big(f(x) \in C_\eta\big) = P_f\left(|\hat{f}_n(x) - f(x)| < \zeta_\eta \sigma_n\right) = P(|V_n| < \zeta_\eta),$$

where $V_n \sim \mathcal{N}\left(\frac{\mu_n}{\sigma_n}, \frac{t_n^2}{\sigma_n^2}\right)$. We have $t_n^2/\sigma_n^2 \to \frac{1}{2}$, while by Corollary 1.7 we have

$$\frac{\mu_n}{\sigma_n} \lesssim k^{\left(\frac{1}{2}-\frac{1}{1-\beta}\right)\alpha+\frac{1}{4}}.$$

This exponent is negative if $\alpha > \xi_\beta$, hence in this case we have $V_n \rightsquigarrow \mathcal{N}(0,1/2)$. For $\alpha = \xi_\beta$ and $f(t) = C|x-t|^\alpha$, we have $\mu_n/\sigma_n \to \sqrt{2}C\Gamma(\alpha+1)$. On the other hand, if $\alpha < \xi_\beta$ and we choose $f$ such that the bias is at least $k^{\left(\frac{1}{2}-\frac{1}{1-\beta}\right)\alpha}$ (e.g. $f(t) = |x-t|^\alpha$), we have $V_n \rightsquigarrow \infty$. □

As we noted in the introduction, we see that as $\beta \to 1$, the range of $\alpha$ for which we obtain a favourable coverage increases. Furthermore, the rate at which the bias tends to zero is also increasing in $\beta$. On the other hand, the rate $\sigma_n$ at which the credible set contracts decreases as $\beta$ grows. More precisely, we have

$$\mu_n \lesssim n^{-\frac{1}{2}(\beta+1)\alpha}, \qquad \sigma_n \asymp n^{\frac{1}{4}(\beta-1)}.$$

Equating these exponents, we obtain the optimal choice $\beta = \frac{1-2\alpha}{1+2\alpha}$. We see that for $\alpha = \frac{1}{2}$, the optimal choice is $\beta = 0$. As $\alpha$ decreases, larger values of $\beta$ (and hence a less smooth prior) are optimal. For $\alpha > \frac{1}{2}$, we should use negative values of $\beta$ (and hence a smoother prior). Note however that the optimal choice of $\beta$ is not convenient in practice. Indeed, in this case we have $\alpha = \xi_\beta$ in Theorem 1.1, which does not guarantee a useful coverage. Hence it is preferable to choose $\beta$ slightly larger, so that we are in the case $\alpha > \xi_\beta$.

## 1.4 Discussion

The results in this chapter concern a Gaussian process prior with a fixed scaling, given by a parameter $\beta$. If the smoothness level $\alpha$ of the true regression function is known, then $\beta$ can be chosen to obtain a conservative confidence level of minimal width. In real-world applications the smoothness level will typically not be known in advance. Within the Bayesian setting one might again take it to be distributed according to a prior. Alternatively, one might replace it by an estimator. In view of results in [Szabó et al., 2013] and [Szabó et al., 2015] it is to be expected that such an adaptive procedure will destroy coverage for some functions in the Hölder classes. We will study this problem in more detail in the following two chapters.

Another possible extension is to consider functions that are smoother than $C^2$. In this case, we might replace the prior $W$ by an integrated form of Brownian motion. Here we obtain a system of equations similar to (1.2), but significantly more complex. In the simplified form for the case of standard integrated Brownian motion, there are now five rows with non-zero right hand side, rather than two. The method used in this chapter cannot be applied to solve this system; other techniques are required.