On Brownian motion as a prior for nonparametric regression

J.H. van Zanten*

Eindhoven University of Technology
Department of Mathematics
P.O. Box 513
5600 MB Eindhoven
The Netherlands
j.h.v.zanten@tue.nl

Revised version, February 4, 2010

Abstract

In this paper we consider the use of Brownian motion as a prior in a nonparametric, univariate regression setting. Using change of measure theory for continuous semimartingales we derive an explicit stochastic differential equation characterization for the posterior. In combination with stochastic calculus tools this dynamical characterization of the posterior allows us to derive new asymptotic properties of the posterior.

1 Introduction

Consider a fixed design regression problem where we have data $X_1, \ldots, X_n$ satisfying

$$X_i = \theta(t_i) + \varepsilon_i,$$

with $t_1, \ldots, t_n \in [0,1]$ known design points, $\theta : [0,1] \to \mathbb{R}$ an unknown regression function belonging to some function space Θ and noise variables $\varepsilon_1, \ldots, \varepsilon_n$ that are independent and $N(0, \sigma^2)$ for $\sigma^2 > 0$. The Bayesian approach to making inference about θ begins with choosing a prior distribution Π on the space Θ. The prior and the data then give rise to the posterior distribution, which is in this case the random law on Θ defined by

$$\Pi(B \mid X_1, \ldots, X_n) = \frac{\int_B \exp\Big(\frac{1}{\sigma^2}\sum_{i=1}^n X_i\theta(t_i) - \frac{1}{2\sigma^2}\sum_{i=1}^n \theta^2(t_i)\Big)\Pi(d\theta)}{\int \exp\Big(\frac{1}{\sigma^2}\sum_{i=1}^n X_i\theta(t_i) - \frac{1}{2\sigma^2}\sum_{i=1}^n \theta^2(t_i)\Big)\Pi(d\theta)}$$

*Partially supported by the Netherlands Organization for Scientific Research NWO.

for a measurable $B \subseteq \Theta$. (We assume of course the necessary measurability to make this well defined.) Further inference is based on the posterior. The posterior mean $\hat{\theta}$, defined by

$$\hat{\theta}(t) = \int \theta(t)\, d\Pi(d\theta \,|\, X_1, \ldots, X_n),$$

can for instance serve as an estimator for the regression function.

Gaussian process priors are a popular choice in this setting. An important advantage is that they are conjugate, in the sense that the posterior is Gaussian as well. This make it relatively easy to implement the procedure. In addition, the class of Gaussian processes is very flexible in the sense that the sample paths of a Gaussian processes can have any regularity that is desired. See for instance the book Rasmussen and Williams (2006) or the website `www.gaussianprocess.org` for references to what is called *Gaussian process regression* in the machine learning community.

If the true regression function is thought to be highly irregular, Brownian motion can be a sensible choice for a prior model for θ. Recent asymptotic theory shows that if the true regression function θ_0 is Hölder continuous of order 1/2, then taking Brownian motion (with standard normal initial distribution) as a prior yields a posterior that contracts around θ_0 at the rate $n^{1/4}$, in the sense that for $M > 0$ large enough,

$$\Pi(\theta \in C[0,1] : \|\theta - \theta_0\|_n > Mn^{-1/4} \,|\, X_1, \ldots, X_n) \xrightarrow{P_{\theta_0}} 0 \tag{1.1}$$

as $n \to \infty$ (see Van der Vaart and Van Zanten (2008)). Here $\|f\|_n^2 = \sum f^2(t_i)$ and the convergence is in probability under the true model with regression function θ_0. The rate $n^{-1/4}$ is in fact the optimal rate if $\theta_0 \in C^{1/2}[0,1]$ and nothing else about θ_0 is known. Besides this consistency property, the Brownian motion prior is attractive because the Markov property of the Brownian motion makes the computation of the posterior even easier to implement. Computing the posterior mean for instance boils down to a filtering/smoothing problem that can be dealt with using Kalman smoothing techniques (see also the recursions given in Section 3).

In this paper we derive a number of new asymptotic properties of the posterior corresponding to a Brownian motion prior (in fact we consider a slightly larger class of priors, including for instance also the Ornstein-Uhlenbeck process). In particular, we study the asymptotic concentration of the posterior around the posterior mean. As a first step we consider the posterior covariance function

$$r_n(s,t) = \int (\theta(s) - \hat{\theta}(s))(\theta(t) - \hat{\theta}(t))\, d\Pi(d\theta \,|\, X_1, \ldots, X_n).$$

We prove, among other things, that for $s, t \in (0,1)$, it holds that $\sqrt{n} r_n(s,t) \to 0$ if $s \neq t$. In particular, evaluations of the regression function at different points are asymptotically independent under the posterior. Qualitatively, this means that even if n gets very large, the posterior does not "smooth out", and will continue to look rather "rough".

A consequence of the fact that $\sqrt{n} r_n(s,t) \to 0$ if $s \neq t$ is that $n^{1/4}\|\theta - \hat{\theta}\|_\infty \to \infty$ under the posterior. Hence, the width of uniform asymptotic credible bands (i.e. bands that contain a certain fraction of the posterior mass, asymptotically) must be of larger order than $n^{-1/4}$, if they exist. We prove in this paper that for L_p-norms with $p < \infty$, we do have credible regions with a radius of the order $n^{-1/4}$. For $p = 2$ we get the most precise result. For $0 < a < b < 1$, the posterior asymptotically gives mass $1 - \alpha$ to the set

$$\left\{\theta : \left| \|\theta - \hat{\theta}\|_{L^2[a,b]} - n^{-1/4}\sqrt{\frac{(b-a)\sigma}{2}} \right| \leq n^{-1/2}\xi_{\alpha/2}\frac{\sigma}{2}\right\},$$

where ξ_α is the upper α-quantile of the standard normal distribution. Note that this global credible region can be viewed as an annulus in $L^2[a,b]$ around the posterior mean $\hat{\theta}$ with a diameter of the order $n^{-1/4}$ and a width of the order $n^{-1/2}$. This result is related to a result of Cox (1993), who describes a similar phenomon in a more abstract setting where (multiply) integrated Brownian motion is used as prior. His approach is rather different than ours and leads to credible regions involving constants defined in terms of certain eigenvalues related to the prior. The priors we consider are not included in the setup of Cox (1993).

Technically our results rely on a dynamical, stochastic process characterization of the posterior. In the next section we use Girsanov's theorem to prove that the posterior is the law of a process solving a linear stochastic differential equation (SDE), and we give recurrent expressions for the coefficients of the SDE. To prove the asymptotic statistical results we study the asymptotic behaviour of the solutions of the recurrence relations and we employ martingale limit theorems.

The remainder of the paper is organized as follows. In the next section we give a precise definition of the class of SDE priors that we consider. We give a number of concrete examples and discuss the Markov property that we will need at a later stage. In section 3 we recall the nonparametric regression setting and we derive the SDE characterization of the posterior distribution corresponding to the SDE prior. In the final Section 4 we use the SDE representation to study the asymptotic behaviour of the posterior, yielding explicit asymptotic pointwise and L^p-credible regions.

2 Prior stochastic differential equations

2.1 Definition of the priors

The linear SDE characterization of the posterior can in fact be derived for a wider class of priors than the one we consider in the next section containing the asymptotic results. Since it does not require extra effort, we consider the broader class in the present section.

For the rigorous specification of the priors that we will employ, consider an auxiliary filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in[0,1]}, \mathbb{P})$. On this stochastic basis, let $W = (W(t))_{t\in[0,1]}$ be a standard Brownian motion and let ξ be a (possibly degenerate) Gaussian random variable with mean μ and variance ρ^2, which is $\mathscr{F}_0$-measurable, and hence independent of W. For measurable, locally bounded functions a, b and τ on $[0, 1)$, with τ is strictly positive, consider the SDE

$$dY(t) = (a(t)Y(t) + b(t))\, dt + \tau(t)\, dW(t), \qquad Y(0) = \xi. \tag{2.1}$$

Setting $A(t) = \int_0^t a(s)\, ds$ for $t \in [0, 1)$ and using the integration by parts formula to compute the differential $d(e^{-A(t)}Y(t))$ we see that (2.1) has a unique strong solution on $[0, 1)$, given by

$$Y(t) = e^{A(t)}\xi + \int_0^t e^{A(t)-A(s)}b(s)\, ds + \int_0^t e^{A(t)-A(s)}\tau(s)\, dW(s) \tag{2.2}$$

(see for instance also Karatzas and Shreve (1991), Section 5.6). Since the coefficients of the SDE (2.1) are deterministic functions and ξ and W are independent, the process Y is Gaussian. Moreover, it is easily verified that $Y(t)$ has an almost surely finite limit for $t \uparrow 1$, if and only if the limits

$$\lim_{t\uparrow 1} e^{A(t)}, \qquad \lim_{t\uparrow 1} \int_0^t e^{A(t)-A(s)}b(s)\, ds, \qquad \lim_{t\uparrow 1} \int_0^t e^{2(A(t)-A(s))}\tau^2(s)\, ds \tag{2.3}$$

exist and are finite. (This can for instance be seen by writing the stochastic integral in (2.2) as a time-changed Brownian motion.) We will assume that these limits exist, so that we can define $Y(1)$ as the almost sure limit $\lim_{t\uparrow 1} Y(t)$. The resulting process $Y = (Y(s))_{s\in[0,1]}$ indexed by the whole unit interval $[0, 1]$ then has a continuous modification and hence we can view its law as a distribution on the space $C[0, 1]$ of continuous functions on $[0, 1]$, endowed with its Borel σ-field. This law is denoted by $\Pi(\mu, \rho, a, b, \tau)$ and will serve as prior distribution in the nonparametric regression model that we consider in this paper.

2.2 Examples of prior SDE's

The limits in (2.3) clearly exist and are finite if a, b and τ are well defined and bounded on the whole unit interval. Several interesting priors are in fact already obtained if they are taken constant.

Example 2.1 (Brownian motion with drift)**.** If $a = 0$ and b and τ are constant, the process Y is simply a Brownian motion with linear drift. Indeed, we have

$$Y(t) = \xi + bt + \tau W(t)$$

in that case. This process is considered as a prior for nonparametric regression for instance in the recent paper Van der Vaart and Van Zanten (2008), as part of the much larger family of so-called Riemann-Liouville processes. In Bayesian regression the use of *integrated* Brownian motion priors is well established, and is in fact connected to using smoothing splines, see e.g. Wahba (1978).

Example 2.2 (Ornstein-Uhlenbeck process)**.** For a, b and τ constant and $a < 0$ the in SDE (2.1) reduces to the SDE of the stationary Ornstein-Uhlenbeck process. The SDE can in this case be written as

$$dY(t) = a\Big(Y(t) - \Big(-\frac{b}{a}\Big)\Big)\,dt + \tau\,dW(t).$$

This shows that $-b/a$ is the so-called mean-reversion level and $-a$ is the mean-reversion force. Indeed, it is clear from the SDE that at every time t there is a drift towards the level $-b/a$ which is proportional to $-a$ times the distance between Y_t and $-b/a$. If the distribution of the initial variable ξ is Gaussian with mean $m = -b/a$ and variance $s^2 = -\tau^2/(2a)$, the process Y is stationary (cf. Karatzas and Shreve (1991), Theorem, 5.6.7).

The stationary Ornstein-Uhlenbeck process is often mentioned as prior for regression in the machine learning literature, as particular example in the larger class of so-called Matérn processes. See for instance Rasmussen and Williams (2006) and the references therein.

More generally than in the Ornstein-Uhlenbeck example, it might be attractive to construct priors by employing SDE's of the form

$$dY(t) = -f(t)(Y(t) - l(t))\,dt + \tau(t)\,dW(t), \tag{2.4}$$

with f a function that takes nonnegative values. In that case the function l describes the "mean level" of the process, which might represent an overall prior idea about the regression function. The function f, which quantifies how strong the process Y is pulled toward the function l, can be used in conjunction with the diffusion function τ to specify the degree of uncertainty about the level l, which might vary from one location to another in the unit interval. This is reflected in the spread of the prior, which will then be dependent on the location as well.

In our setup, the pulling force $f(t)$ in (2.4) is allowed to become infinitely large for $t = 1$. This can be used to force the process Y to move to a fixed point at time 1. As a consequence, bridge-type processes are included in our framework.

Example 2.3 (Brownian bridge)**.** For real numbers l and r, consider the (2.1) with $\xi = l$, $a(t) = -1/(1-t)$, $b(t) = r/(1-t)$ and $\tau(t) = 1$, i.e.

$$dY(t) = -\frac{1}{1-t}(Y(t) - r)\,dt + dW(t), \qquad , Y_0 = l.$$

In this example the functions a and b explode at $t = 1$. However we have $\exp(A(t)) = 1 - t$, so that the limits in (2.3) exist and are finite (they are 0, r and 0, respectively). The resulting process Y is the Brownian bridge which starts at l at time 0 and ends up in r at time 1 (see Karatzas and Shreve (1991), Section 5.6.B).

The Brownian bridge may serve as a prior on periodic functions for instance. A discrete-time version of the process is used as prior in the paper Drignei (2006).

2.3 Markov property of the priors

For later reference we recall the fact that the process Y solving (2.1) is a Markov process with Gaussian transition densities. Indeed, it follows from (2.2) that for $0 \le s \le t \le 1$,

$$\begin{aligned} Y(t) = e^{A(t)-A(s)}Y(s) + \int_s^t e^{A(t)-A(u)}b(u)\,du \\ + \int_s^t e^{A(t)-A(u)}\tau(u)\,dW(u), \end{aligned} \tag{2.5}$$

with the proper interpretation for $t = 1$. Since the last term on the right is independent of $(Y(u))_{u\le s}$, and has a Gaussian distribution with mean zero and variance

$$\int_s^t e^{2(A(t)-A(u))}\tau^2(u)\,du,$$

it follows that the conditional distribution of $Y(t)$ given $(Y(u))_{u\le s}$ is Gaussian again, with mean

$$e^{A(t)-A(s)}Y(s) + \int_s^t e^{A(t)-A(u)}b(u)\,du$$

and variance

$$\int_s^t e^{2(A(t)-A(u))}\tau^2(u)\,du.$$

In other words, the process Y is Markov and its transition kernel $P_{s,t}$, defined by $P_{s,t}f(x) = \mathbb{E}(f(Y(t))\,|\,Y(s) = x)$ for a bounded measurable function f and $x \in \mathbb{R}$, is given by $P_{s,t}f(x) = \mathbb{E}f(p(s,t)x + q(s,t) + \sqrt{v(s,t)}N)$, where N is standard normal and

$$\begin{aligned} p(s,t) = e^{A(t)-A(s)}, \qquad q(s,t) = \int_s^t e^{A(t)-A(u)}b(u)\,du, \\ v(s,t) = \int_s^t e^{2(A(t)-A(u))}\tau^2(u)\,du. \end{aligned} \tag{2.6}$$

If a, b and τ are constant, the process Y is a time-homogenous Markov process. The quantities in the preceding display then only depend on s and t through the difference $t - s$. Specifically, if $a = 0$ we have $p(s,t) = 1$, $q(s,t) = b(t-s)$ and $v(s,t) = \tau^2(t-s)$, and if $a \ne 0$ it holds that

$$p(s,t) = e^{a(t-s)}, \quad q(s,t) = \frac{b}{a}\Big(e^{a(t-s)} - 1\Big), \quad v(s,t) = \frac{\tau^2}{2a}\Big(e^{2a(t-s)} - 1\Big).$$

3 Posterior stochastic differential equations

3.1 Statistical setting

We are now going to apply the priors introduced in the preceding section in our nonparametric regression problem. Recall that we have observations $X_1, \ldots, X_n$

satisfying

$$X_i = \theta(t_i) + \varepsilon_i, \tag{3.1}$$

with $t_1, \ldots, t_n \in [0,1]$ known design points, $\theta : [0,1] \to \mathbb{R}$ an unknown function, and noise variables $\varepsilon_1, \ldots, \varepsilon_n$ that are independent and $N(0, \sigma^2)$ for some $\sigma^2 > 0$. We will assume that the labels of the design points correspond to their order, so that $0 \leq t_1 < \cdots < t_n \leq 1$. As prior on the function θ we use the law $\Pi = \Pi(\mu, \rho, a, b, \tau)$ of the solution Y of the SDE (2.1) described in the preceding section, for some fixed choice of μ, ρ, a, b and τ satisfying the conditions of Section 2.1. The main result of this section is a characterization of the corresponding posterior as the law of a stochastic process satisfying an explicit SDE, depending on the data $X_1, \ldots, X_n$.

3.2 Characterization of the posterior

To describe the posterior SDE we introduce some notation. Recall the functions p, q and v associated to the prior that are defined in (2.6). Using these functions we recursively define two sequences α_k and β_k, depending on the data $X_1, \ldots, X_n$. We first set $\alpha_n = X_n/\sigma^2$, $\beta_n = 1/\sigma^2$, and $t_0 = 0$. Then for $k = 0, \ldots, n-1$, we define

$$\alpha_k = \frac{X_k}{\sigma^2} 1_{k \geq 1} + \frac{p(t_k, t_{k+1})(\alpha_{k+1} - \beta_{k+1} q(t_k, t_{k+1}))}{1 + \beta_{k+1} v(t_k, t_{k+1})}, \tag{3.2}$$

$$\beta_k = \frac{1}{\sigma^2} 1_{k \geq 1} + \frac{\beta_{k+1} p^2(t_k, t_{k+1})}{1 + \beta_{k+1} v(t_k, t_{k+1})}. \tag{3.3}$$

The numbers α_0 and β_0 are actually only needed if the first design point t_1 is strictly positive.

We can now formulate and prove the main SDE characterization of the posterior.

Theorem 3.1. *The posterior corresponding to the prior* $\Pi(\mu, \rho, a, b, \tau)$ *is given by* $\Pi(\bar{\mu}, \bar{\rho}, \bar{a}, \bar{b}, \tau)$, *where*

$$\bar{\mu} = \frac{\mu + \alpha_0 \rho^2}{1 + \beta_0 \rho^2}, \qquad \bar{\rho}^2 = \frac{\rho^2}{1 + \beta_0 \rho^2}$$

if $t_1 > 0$ *and*

$$\bar{\mu} = \frac{\mu + \alpha_1 \rho^2}{1 + \beta_1 \rho^2}, \qquad \bar{\rho}^2 = \frac{\rho^2}{1 + \beta_1 \rho^2}$$

if $t_1 = 0$. *Moreover, we have*

$$\bar{a}(t) = a(t) - \sum_{j=1}^{n} \frac{\beta_j p^2(t, t_j) \tau^2(t)}{1 + \beta_j v(t, t_j)} 1_{(t_{j-1}, t_j]}(t),$$

$$\bar{b}(t) = b(t) + \sum_{j=1}^{n} \frac{p(t, t_j)(\alpha_j - \beta_j q(t, t_j)) \tau^2(t)}{1 + \beta_j v(t, t_j)} 1_{(t_{j-1}, t_j]}(t),$$

and the diffusion function τ *remains unaltered.*

Proof. By construction, the law of θ under the prior $\Pi = \Pi(\mu, \rho, a, b, \tau)$ is the same as the law of the process Y defined in Section 2.1 under the auxiliary probability measure $\mathbb{P}$. Note that the posterior is the law of Y under the probability measure $\mathbb{Q}$ defined by

$$d\mathbb{Q} = \frac{\exp\left(\frac{1}{\sigma^2}\sum_{i=1}^n X_i Y(t_i) - \frac{1}{2\sigma^2}\sum_{i=1}^n Y^2(t_i)\right)}{\mathbb{E}\exp\left(\frac{1}{\sigma^2}\sum_{i=1}^n X_i Y(t_i) - \frac{1}{2\sigma^2}\sum_{i=1}^n Y^2(t_i)\right)}\, d\mathbb{P},$$

where the data $X_1, \ldots, X_n$ are considered as fixed. Let $Z = (Z(t))_{t\in[0,1]}$ be the density process of $\mathbb{Q}$ relative to $\mathbb{P}$, which is defined by

$$Z(t) = \mathbb{E}\Big(\frac{d\mathbb{Q}}{d\mathbb{P}} \mid \mathscr{F}_t\Big) = \frac{\mathbb{E}\Big(\exp\left(\frac{1}{\sigma^2}\sum_{i=1}^n X_i Y(t_i) - \frac{1}{2\sigma^2}\sum_{i=1}^n Y^2(t_i)\right) \mid \mathscr{F}_t\Big)}{\mathbb{E}\exp\left(\frac{1}{\sigma^2}\sum_{i=1}^n X_i Y(t_i) - \frac{1}{2\sigma^2}\sum_{i=1}^n Y^2(t_i)\right)}.$$

Let $M(t) = \int_0^t \tau(s)\, dW(s)$ be the continuous martingale part of Y. By Girsanov's theorem, see e.g. Jacod and Shiryaev (1987), Theorem III.3.11, it holds that the process N defined by

$$N(t) = M(t) - \int_0^t \frac{1}{Z(s)}\, d\langle M, Z\rangle(s) \tag{3.4}$$

is a $\mathbb{Q}$-local martingale, and $\langle N\rangle = \langle M\rangle$. It follows that we have the $\mathbb{Q}$-semimartingale decomposition

$$\begin{aligned} Y(t) = Y(0) &+ \int_0^t (a(s)Y(s) + b(s))\, ds \\ &+ \int_0^t \frac{1}{Z(s)}\, d\langle M, Z\rangle(s) + \int_0^t \tau(s)\, dU(s), \end{aligned} \tag{3.5}$$

where U is a standard Brownian motion under $\mathbb{Q}$.

To compute the middle integral on the right-hand side of the last display we first derive an expression for the conditional expectation

$$\mathbb{E}\Big(\exp\Big(\frac{1}{\sigma^2}\sum_{i=1}^n X_i Y(t_i) - \frac{1}{2\sigma^2}\sum_{i=1}^n Y^2(t_i)\Big) \mid \mathscr{F}_t\Big)$$

and for the density process Z. To simplify the notation somewhat we set $a_i = X_i/\sigma^2$ and $b = 1/\sigma^2$. If $t_n \le t$ there is nothing to compute. Now set $t_0 = 0$, fix $t \in (0, 1]$ and let $j \in \{1, \ldots, n\}$ be such that $t_{j-1} < t \le t_j$. Then the conditional expectation equals

$$\exp\Big(\sum_{i=1}^{j-1} a_i Y(t_i) - \frac{1}{2}b\sum_{i=1}^{j-1} Y^2(t_i)\Big)\mathbb{E}(\exp\Big(\sum_{i=j}^{n} a_i Y(t_i) - \frac{1}{2}b\sum_{i=j}^{n} Y^2(t_i)\Big) \mid \mathscr{F}_t).$$

By the Markov property of Y (see Section 2.3) we have

$$\begin{aligned}\mathbb{E}(\exp\Big(\sum_{i=j}^{n} a_i Y(t_i) - \frac{1}{2}b\sum_{i=j}^{n} Y^2(t_i)\Big) \mid \mathscr{F}_t)\\ = P_{t,t_j} h_{a_j,b} P_{t_j,t_{j+1}} h_{a_{j+1},b} \cdots P_{t_{n-1},t_n} h_{a_n,b}(Y(t)),\end{aligned}$$

where $h_{a,b}(x) = \exp(ax - bx^2/2)$. So with

$$\begin{aligned}c &= \Big(\mathbb{E}e^{\sum a_i Y(t_i) - \frac{1}{2}b\sum Y^2(t_i)}\Big)^{-1},\\ g_j(t,x) &= P_{t,t_j} h_{a_j,b} P_{t_j,t_{j+1}} h_{a_{j+1},b} \cdots P_{t_{n-1},t_n} h_{a_n,b}(x),\end{aligned}$$

we have

$$Z(t) = c g_j(t, Y(t)) \prod_{i=1}^{j-1} h_{a_i,b}(Y(t_i)), \qquad t_{j-1} < t \le t_j. \tag{3.6}$$

By Lemma 3.2 ahead and the identity (3.10) for the function h we have, for a constant C not depending on x,

$$h_{a_j,b} P_{t_j,t_{j+1}} h_{a_{j+1},b} \cdots P_{t_{n-1},t_n} h_{a_n,b} = C h_{\alpha_j,\beta_j}.$$

Applying P_{t,t_j} on the left and on the right and using the lemma once more we find that for a smooth function $C = C(t)$ not depending on x, we have

$$g_j(t,x) = C(t) h_{\frac{p(t,t_j)(\alpha_j - \beta_j q(t,t_j))}{1+\beta_j v(t,t_j)}, \frac{\beta_j p^2(t,t_j)}{1+\beta_j v(t,t_j)}}(x). \tag{3.7}$$

for $t_{j-1} < t \le t_j$. By (3.6), Itô's formula and the fact that Z is a martingale, it holds on the interval $(t_{j-1}, t_j]$ that

$$dZ(t) = c\Big(\prod_{i=1}^{j-1} h_{\alpha_i,\beta}(Y(t_i))\Big)\frac{\partial}{\partial x} g_j(t, Y(t))\, dM(t) = Z(t)\frac{\frac{\partial}{\partial x} g_j(t,Y(t))}{g_j(t,Y(t))}\, dM(t).$$

Hence, for $t_{j-1} < t \le t_j$, we have

$$\begin{aligned}d\langle Z, M\rangle(t) &= Z(t)\frac{\frac{\partial}{\partial x} g_j(t,Y(t))}{g_j(t,Y(t))}\, d\langle M\rangle(t)\\ &= Z(t)\frac{\frac{\partial}{\partial x} g_j(t,Y(t))}{g_j(t,Y(t))}\tau^2(t)\, dt.\end{aligned} \tag{3.8}$$

Using the fact that $(\partial/\partial x h_{\alpha,\beta}(x))/h_{\alpha,\beta}(x) = \alpha - \beta x$ we see that

$$\frac{\frac{\partial}{\partial x} g_j(t,x)}{g_j(t,x)} = \frac{p(t,t_j)(\alpha_j - \beta_j q(t,t_j))}{1+\beta_j v(t,t_j)} - \frac{\beta_j p^2(t,t_j)}{1+\beta_j v(t,t_j)} x.$$

Combining this with (3.8) and (3.5) shows that for $t_{j-1} < t \le t_j$,

$$\begin{aligned} dY(t) =& \Big(\Big(a(t) - \frac{\beta_j p^2(t,t_j)\tau^2(t)}{1+\beta_j v(t,t_j)}\Big)Y(t) \\ &+ \Big(b(t) + \frac{p(t,t_j)(\alpha_j - \beta_j q(t,t_j))\tau^2(t)}{1+\beta_j v(t,t_j)}\Big)\Big)\, dt \\ &+ \tau(t)\, dU_t. \end{aligned}$$

It remains to prove the assertion about the law of $Y(0)$ under $\mathbb{Q}$. The case $\rho^2 = 0$ is clear, so assume $\rho^2 > 0$. We have from (3.6) and (3.7) that if $t_1 > 0$, then

$$Z(0) = C \exp\Big(\frac{p(0,t_1)(\alpha_1 - \beta_1 q(0,t_1))}{1+\beta_1 v(0,t_1)} Y(0) - \frac{1}{2}\frac{\beta_1 p^2(0,t_1)}{1+\beta_1 v(0,t_1)} Y^2(0)\Big),$$

where $C > 0$ is a constant that ensures that $\mathbb{E}Z(0) = 1$. In this case it follows that under $\mathbb{Q}$, the random variable $Y(0)$ has a density proportional to

$$x \mapsto \exp\Big(\big(\frac{\mu}{\rho^2} + \frac{p(0,t_1)(\alpha_1 - \beta_1 q(0,t_1))}{1+\beta_1 v(0,t_1)}\big)x - \frac{1}{2}\big(\frac{1}{\rho^2} + \frac{\beta_1 p^2(0,t_1)}{1+\beta_1 v(0,t_1)}\big)x^2\Big).$$

The proof for the case $t_1 > 0$ is now easily completed. The case $t_1 = 0$ follows by an analogous argument. □

Lemma 3.2. *For $h_{a,b}(x) = \exp(ax - bx^2/2)$, with $a \in \mathbb{R}, b > 0$, and $0 \le s \le t \le 1$ we have*

$$P_{s,t}h_{a,b}(x) = \frac{h_{0,-\frac{1}{1+bv}}(a\sqrt{v})h_{\frac{a}{1+bv},\frac{b}{1+bv}}(q)}{\sqrt{1+bv}} h_{\frac{p(a-bq)}{1+bv},\frac{bp^2}{1+bv}}(x),$$

where $p = p(s,t)$, $q = q(s,t)$, $v = v(s,t)$ (see (2.6)).

Proof. In this proof we use the easily verifiable identities

$$h_{a,b}(cx+d) = h_{a,b}(d)h_{c(a-db),c^2b}(x) \tag{3.9}$$

and

$$h_{a_1,b_1}(x)h_{a_2,b_2}(x) = h_{a_1+a_2,b_1+b_2}(x). \tag{3.10}$$

As observed in Section 2 we have, with $m = px + q$ and N a standard Gaussian random variable,

$$P_{s,t}h_{a,b}(x) = \mathbb{E}h_{a,b}(m + \sqrt{v}N) = h_{a,b}(m)\mathbb{E}h_{(a-mb)\sqrt{v},bv}(N),$$

where the second equality follows from (3.9). For $\alpha \in \mathbb{R}$ and $\beta > 0$ we have

$$\mathbb{E}h_{\alpha,\beta}(N) = \mathbb{E}e^{\alpha N - \frac{1}{2}\beta N^2} = \frac{e^{\frac{1}{2}\frac{\alpha^2}{1+\beta}}}{\sqrt{1+\beta}} = \frac{1}{\sqrt{1+\beta}}h_{0,-(1+\beta)^{-1}}(\alpha).$$

Taking $\alpha = (a - bm)\sqrt{v}$, $\beta = bv$ and using the indentities (3.9)–(3.10) we get

$$\begin{aligned}
P_{s,t}h_{a,b}(x) &= \frac{1}{\sqrt{1+bv}} h_{a,b}(m) h_{0,-(1+bv)^{-1}}((a-bm)\sqrt{v}) \\
&= \frac{h_{0,-(1+bv)^{-1}}(a\sqrt{v})}{\sqrt{1+bv}} h_{a,b}(m) h_{-abv/(1+bv),-b^2v/(1+bv)}(m) \\
&= \frac{h_{0,-(1+bv)^{-1}}(a\sqrt{v})}{\sqrt{1+bv}} h_{a/(1+bv),b/(1+bv)}(m).
\end{aligned}$$

Now use the fact that $m = px + q$ and identity (3.9) to complete the proof. □

Note that the posterior SDE described by Theorem 3.1 is a linear SDE of the form (2.1) again, but with the functions a and b describing the prior drift replaced by their posterior versions $\bar{a}$ and $\bar{b}$. In the interval (t_{j-1}, t_j) between the two design points t_{j-1} and t_j, the drift of the posterior depends, through the numbers $\alpha_j, \ldots, \alpha_n$, on the data $X_j, X_{j+1}, \ldots, X_n$.

3.3 Exact sampling from the posterior

Given a concrete prior $\Pi(\mu, \rho, a, b, \tau)$ and data $X_1, \ldots, X_n$ it is straightforward to compute the sequences α_k and β_k and hence to evaluate the coefficients $\bar{a}$ and $\bar{b}$ of the posterior SDE and compute the initial law of the posterior process. Since an SDE of the form (2.1) admits the explicit solution (2.2), the restriction of the posterior process to a finite grid in $[0, 1]$ has a first order moving average representation. This can be used for exact sampling from the posterior on a finite grid. To formulate this result we use the posterior versions of the functions p, q and v defined in (2.6). We define

$$\begin{aligned}
\bar{p}(s,t) = e^{\bar{A}(t)-\bar{A}(s)}, \qquad \bar{q}(s,t) = \int_s^t e^{\bar{A}(t)-\bar{A}(u)} \bar{b}(u)\, du, \\
\bar{v}(s,t) = \int_s^t e^{2(\bar{A}(t)-\bar{A}(u))} \tau^2(u)\, du,
\end{aligned} \tag{3.11}$$

where $\bar{A}$ is the posterior version of A, i.e.

$$\bar{A}(t) = \int_0^t \bar{a}(s)\, ds.$$

The result then takes the following form.

Corollary 3.3. *Consider the setting of Theorem 3.1. For numbers $0 = s_0 < s_1 < \cdots < s_k \leq 1$, the posterior distribution of the vector $(\theta(s_0), \ldots, \theta(s_k))$ is the law of the random vector $(V_0, \ldots, V_k)$ defined by $V_0 \sim N(\bar{\mu}, \bar{\rho}^2)$ and for $j = 1, \ldots, k$,*

$$V_j = \bar{p}(s_{j-1}, s_j) V_{j-1} + \bar{q}(s_{j-1}, s_j) + \sqrt{\bar{v}(s_{j-1}, s_j)} Z_j, \tag{3.12}$$

$Z_1, \ldots, Z_k$ are independent, standard normal variables, independent of V_0.

Proof. This follows immediately from Theorem 3.1 and the fact that a process Y solving an SDE of the form (2.1) satisfies (2.5) for $s \leq t$. $\square$

4 Asymptotic results

4.1 asymptotic behaviour of the posterior covariance

We work in the setting and notation of Section 3 again. In this section we study the asymptotic behaviour of the posterior as the number of design points tends to infinity. Specifically, we are interested in the asymptotic spread of the posterior around the posterior mean if the number of design points gets large. For simplicity we consider the regular design case, so the design points t_j are equally spaced in $[0, 1]$, i.e. $t_j = j/n$. Moreover, we assume that the functions a, b and τ defining the prior are constant. As observed in Section 2, the Brownian motion with drift and the Ornstein-Uhlenbeck process are included in this setup.

By Theorem 3.1 and the representation (2.2) of the solution of an SDE of the form (2.1), the posterior corresponding to the SDE prior $\Pi = \Pi(\mu, \rho, a, b, \tau)$ is Gaussian. The mean function of the posterior is the posterior mean $\hat{\theta}$, defined by

$$\hat{\theta}(t) = \int \theta(t)\, d\Pi(d\theta \,|\, X_1, \ldots, X_n),$$

and the posterior covariance function r, defined by

$$r(s,t) = \int (\theta(s) - \hat{\theta}(s))(\theta(t) - \hat{\theta}(t))\, d\Pi(d\theta \,|\, X_1, \ldots, X_n)$$

for $s, t \in [0, 1]$, is given by

$$r(s,t) = e^{\bar{A}(s)+\bar{A}(t)}\bar{\rho}^2 + \tau^2 \int_0^{s\wedge t} e^{\bar{A}(s)+\bar{A}(t)-2\bar{A}(u)}\, du. \tag{4.1}$$

Note that, as is well known, the posterior covariance does not depend on the data.

The following theorem describes the asymptotic behaviour of the posterior covariance function r.

Theorem 4.1. *If $a < \tau^2/\sigma^2$, then for $s, t \in (0, 1)$ it holds that*

$$\sqrt{n} r(s,t) \to \begin{cases} \dfrac{\sigma\tau}{2} & \text{if } s = t, \\ 0 & \text{if } s \neq t, \end{cases}$$

as $n \to \infty$.

Proof. We have $\beta_0/\sqrt{n} \to 1/(\sigma\tau)$ by Lemma 4.2. Hence, by Theorem 3.1,

$$\sqrt{n}\bar{\rho}^2 \to 1_{\rho>0}\sigma\tau$$

as $n \to \infty$. For $t \in (0,1)$, Lemma 4.3 shows that

$$\frac{\bar{A}(t)}{\sqrt{n}} = \int_0^t \frac{\bar{a}(u)}{\sqrt{n}}\, du \to -t\frac{\tau}{\sigma}$$

and it follows that $\bar{A}(t) \to -\infty$. Combining these facts we see that as $n \to \infty$,

$$\sqrt{n} e^{\bar{A}(s)+\bar{A}(t)} \bar{\rho}^2 \to 0$$

as $n \to \infty$. Next, observe that for $s \le t < 1$ we have

$$\begin{aligned}
&\sqrt{n} \int_0^s e^{\bar{A}(s)+\bar{A}(t)-2\bar{A}(u)}\, du \\
&\quad = \frac{\sigma}{2\tau}\int_0^s (-2\bar{a}(u)) e^{\bar{A}(s)+\bar{A}(t)-2\bar{A}(u)}\, du \\
&\quad\quad - \frac{1}{2}\int_0^s \Big(\frac{\sqrt{n}}{\bar{a}(u)} + \frac{\sigma}{\tau}\Big)(-2\bar{a}(u)) e^{\bar{A}(s)+\bar{A}(t)-2\bar{A}(u)}\, du.
\end{aligned}$$

The first term on the right equals

$$\frac{\sigma}{2\tau}\Big(e^{\bar{A}(t)-\bar{A}(s)} - e^{\bar{A}(t)+\bar{A}(s)}\Big),$$

which converges to 1 if $s = t$ and to 0 otherwise. Lemma 4.3 implies that the second term vanishes asymptotically. □

The following lemma, used in the proof of the preceding theorem, describes the asymptotic behaviour of the sequence β_k defined by (3.3) in the case that a, b and τ are constant. Then $p(s,t) = \exp(a(t-s))$ and $v(s,t) = \tau^2(\exp(2a(t-s)) - 1)/(2a)$ (which is understood as $v(s,t) = \tau^2(t-s)$ if $a = 0$), so that in the present case $t_j = j/n$, the sequence is given by $\beta_n = 1/\sigma^2$ and for $k = 0, \dots, n-1$,

$$\beta_k = \frac{1}{\sigma^2} 1_{k \ge 1} + \frac{p_n \beta_{k+1}}{1 + \beta_{k+1}\tau^2 w_n/n},$$

with $p_n = \exp(2a/n)$ and $w_n = n(\exp(2a/n) - 1)/(2a)$.

Lemma 4.2. *Suppose that $n \to \infty$ and j_n is such that $j_n/n \le C < 1$ for some constant C. Then if $a < \tau^2/\sigma^2$,*

$$\max_{0 \le j \le j_n} \Big|\frac{\beta_j}{\sqrt{n}} - \frac{1}{\sigma\tau}\Big| = O\Big(\frac{1}{\sqrt{n}}\Big).$$

Proof. Fix $c > 0$ and for $n \in \mathbb{N}$, define the functions f_n by setting

$$f_n(x) = c + \frac{x p_n}{1 + x w_n/n}.$$

Consider the sequence x_k^n defined recursively by $x_0^n = c$, and for $k \ge 1$, $x_k^n = f_n(x_{k-1}^n)$. Observe that with $c = \tau^2/\sigma^2$, we have that $\tau^2\beta_k = x_{n-k}^n$ for $k =$

$1, \dots, n$, and $\tau^2\beta_0 = x_n^n - c$. For the function f_n we have $f_n(0) = c$, $f_n(x) \to np_n/w_n$ as $x \to \infty$, and

$$f_n'(x) = \frac{p_n}{(1 + xw_n/n)^2}, \qquad f_n''(x) = -\frac{2p_n w_n}{n(1 + xw_n/n)^3}.$$

Hence, in particular, f_n is a nonnegative, increasing, convex function on the positive half-line. The properties of f_n imply it has a unique positive fixed point x_n^*. Solving the quadratic equation $x = f_n(x)$ shows that

$$x_n^* = \frac{cw_n/n + p_n - 1 + \sqrt{(1 - p_n - cw_n/n)^2 + 4cw_n/n}}{2w_n/n}.$$

Since $p_n - 1 \sim 2a/n$ and $w_n - 1 \sim a/n$, it follows that

$$\frac{x_n^*}{\sqrt{n}} = \sqrt{c} + O\Big(\frac{1}{\sqrt{n}}\Big) \tag{4.2}$$

as $n \to \infty$. Next we show that if k/n is bounded away from 0, then for n large enough, $x_k^n/\sqrt{n}$ is bounded away from 0 and infinity. To see this, note that in view of the computations above we have

$$|x_n^* - x_k^n| = |f_n(x_n^*) - f_n(x_{k-1}^n)| = \int_{x_{k-1}^n}^{x_n^*} \frac{p_n}{(1 + xw_n/n)^2}\,dx \le \frac{p_n|x_n^* - x_{k-1}^n|}{(1 + cw_n/n)^2}.$$

Iterating this inequality we get $|x_n^* - x_k^n| \le p_n^k(1 + cw_n/n)^{-2k}|x_n^* - c|$, and hence

$$\Big|\frac{x_k^n}{\sqrt{n}} - \frac{x_n^*}{\sqrt{n}}\Big| \le \Big(\frac{p_n}{(1 + cw_n/n)^2}\Big)^k \Big|\frac{x_n^*}{\sqrt{n}} - \frac{c}{\sqrt{n}}\Big|. \tag{4.3}$$

We have

$$\log\Big(\frac{p_n}{(1 + cw_n/n)^2}\Big)^k = \frac{k}{n}\Big(2a - 2n\log\Big(1 + \frac{cw_n}{n}\Big)\Big).$$

The second factor on the right converges to $2(a - c)$ as $n \to \infty$ and hence, by the assumptions of the lemma, is negative and bounded away from 0 for n large enough. It follows that if k/n is bounded away from 0, then the right hand side of (4.3) is bounded by $K\sqrt{c}$ for some $K \in (0, 1)$. Using the triangle inequality it follows that for k/n bounded away from 0 and n large enough,

$$\Big|\frac{x_k^n}{\sqrt{n}} - \sqrt{c}\Big| \le K'\sqrt{c}$$

for a constant $K' \in (0, 1)$, and hence $x_k^n/\sqrt{n} \ge (1 - K')\sqrt{c}$. The fact that the sequence is bounded also follows easily from (4.3).

Now suppose that $k_n, n \to \infty$ such that $k_n \le n$ and k_n/n is bounded away from 0. Arguing as above we have, for $k \ge k_n$,

$$|x_n^* - x_k^n| = |f_n(x_n^*) - f_n(x_{k-1}^n)| = \int_{x_{k-1}^n}^{x_n^*} \frac{p_n}{(1 + xw_n/n)^2}\,dx \le \frac{p_n|x_n^* - x_{k-1}^n|}{(1 + x_{k-1}^n w_n/n)^2}.$$

By the preceding paragraph we have, for n large enough, $x^n_{k-1} \geq K\sqrt{n}$ for some constant $K > 0$. It follows that

$$|x^*_n - x^n_k| \leq \frac{e^{2a/n}}{\left(1 + Kw_n/\sqrt{n}\right)^2} |x^*_n - x^n_{k-1}|$$

for $k \geq k_n$ and hence, by iteration,

$$\max_{k_n \leq k \leq n} |x^*_n - x^n_k| \leq \frac{e^{2a}}{\left(1 + Kw_n/\sqrt{n}\right)^{2k_n}} |x^*_n - c|.$$

If k_n/n is bounded away from 0, then for n large enough

$$\log \left(1 + Kw_n/\sqrt{n}\right)^{2k_n} \sim 2Kw_n \frac{k_n}{\sqrt{n}} \geq K'\sqrt{n}$$

for some constant $K' > 0$. We see that if $k_n, n \to \infty$ such that k_n/n is bounded away from 0, then for n large enough,

$$\max_{k_n \leq k \leq n} \left| \frac{x^n_k}{\sqrt{n}} - \frac{x^*_n}{\sqrt{n}} \right| \leq K_1 e^{-K_2\sqrt{n}}$$

for some $K_1, K_2 > 0$. Combined with (4.2) this shows that

$$\max_{k_n \leq k \leq n} \left| \frac{x^n_k}{\sqrt{n}} - \sqrt{c} \right| = O\Big(\frac{1}{\sqrt{n}}\Big),$$

by the triangle inequality. Now recall that $c = \tau^2/\sigma^2$ and that $\tau^2\beta_k = x^n_{n-k}$ for $k = 1, \ldots, n$ and $\tau^2\beta_0 = x^n_n - c$, to complete the proof. □

The following auxiliary result concerns the function $\bar{a}$ appearing in the drift of the posterior SDE and is also used in the proof of Theorem 4.1. Again, it deals with the case that a, b and τ are constant and $t_j = j/n$.

Lemma 4.3. *If $a < \tau^2/\sigma^2$ then*

$$\sup_{s \in [0,t]} \left| \frac{\bar{a}(s)}{\sqrt{n}} + \frac{\tau}{\sigma} \right| = O\Big(\frac{1}{\sqrt{n}}\Big),$$

and

$$\sup_{s \in [0,t]} \left| \frac{\sqrt{n}}{\bar{a}(s)} + \frac{\sigma}{\tau} \right| = O\Big(\frac{1}{\sqrt{n}}\Big)$$

for all $t \in (0, 1)$.

Proof. Fix $t \in (0, 1)$. For $t_{j-1} < s \leq t_j$ it holds that

$$\frac{\bar{a}(s)}{\sqrt{n}} = \frac{a}{\sqrt{n}} - \frac{\tau^2 p^2(s, t_j)\beta_j/\sqrt{n}}{1 + \beta_j v(s, t_j)},$$

hence with j_n such that $t_{j_n-1} < t \le t_{j_n}$, we have

$$\sup_{s\in[0,t]}\Big|\frac{\bar a(s)}{\sqrt{n}} + \frac{\tau}{\sigma}\Big| \le \frac{|a|}{\sqrt{n}} + \tau^2 \max_{j\le j_n} \sup_{t_{j-1}<s\le t_j}\Big|\frac{1}{\sigma\tau} - \frac{\beta_j p^2(s,t_j)/\sqrt{n}}{1+\beta_j v(s,t_j)}\Big|$$
$$\le \frac{|a|}{\sqrt{n}} + \tau^2 \max_{j\le j_n}\frac{\beta_j v(t_{j-1},t_j)}{\sigma\tau} + \tau^2 \max_{j\le j_n} \sup_{t_{j-1}<s\le t_j}\Big|\frac{\beta_j p^2(s,t_j)}{\sqrt{n}} - \frac{1}{\sigma\tau}\Big|.$$

The first term on the right is clearly $O(n^{-1/2})$. For the second one this is true as well, in view of Lemma 4.2 and the fact that for n large enough, $v(t_{j-1},t_j)$ is bounded by a multiple of $1/n$. For the last term, observe that

$$\sup_{t_{j-1}<s\le t_j}\Big|\frac{\beta_j p^2(s,t_j)}{\sqrt{n}} - \frac{1}{\sigma\tau}\Big| = \sup_{t_{j-1}<s\le t_j} p^2(s,t_j)\Big|\frac{\beta_j}{\sqrt{n}} - \frac{1}{\sigma\tau p^2(s,t_j)}\Big|$$
$$\le e^{2a/n}\Big|\frac{\beta_j}{\sqrt{n}} - \frac{1}{\sigma\tau}\Big| + O\Big(\frac{1}{n}\Big).$$

Hence, since $j_n \sim tn$, we can apply Lemma 4.2 to complete the proof of the first statement of the lemma.

The first statement implies that for n large enough, for n large enough, $\bar a/\sqrt{n}$ is bounded away from 0 on $[0,t]$. Therefore, the second statement follows from the first one. □

Theorem 4.1 only deals with the posterior covariance function on the interior $(0,1)\times(0,1)$ of the unit square. Due to boundary effects, the behaviour on the boundary of the square is different and not completely tractable. Another remark is that perhaps somewhat surprsingly, it is not necessary for the asymptotic results that $a \le 0$. The parameter a may be positive, but is restricted by the requirement that $a < \tau^2/\sigma^2$. It seems however that priors that are of practical interest always satisfy $a \le 0$. For $a > 0$ one gets a non-stationary Ornstein-Uhlenbeck-type process which moves away from 0 very quickly.

The following corollary follows immediately from Theorem 4.1 and the fact that the posterior is Gaussian. We denote weak convergence by the symbol "$\Rightarrow$".

Corollary 4.4. *If $a < \tau^2/\sigma^2$, then for $t \in (0,1)$ we have that under the posterior,*

$$n^{1/4}(\theta(t) - \hat\theta(t)) \Rightarrow N\Big(0, \frac{\sigma\tau}{2}\Big)$$

as $n \to \infty$.

The corollary implies that the width of pointwise Bayesian confidence intervals, or credible intervals, for θ is asymptotically of the order $n^{-1/4}$. More precisely, we see that for $t \in (0,1)$ the marginal posterior distribution for $\theta(t)$ asymptotically gives mass $1-\alpha$ to the interval

$$\Big[\hat\theta(t) - n^{-1/4}\sqrt{\sigma\tau/2}\,\xi_{\alpha/2}, \hat\theta(t) + n^{-1/4}\sqrt{\sigma\tau/2}\,\xi_{\alpha/2}\Big],$$

where ξ_α is the upper α-quantile of the standard normal distribution.

Observe that Theorem 4.1 implies that for $s \neq t$, the variables $\theta(s)$ and $\theta(t)$ are asymptotically independent under the posterior. This implies that the pointwise weak convergence in (4.4) does extend to functional weak convergence in the space $C[s,t]$ of continuous functions on an interval $[s,t] \subset (0,1)$ for instance. In fact, we will see below that under the posterior, we do not even have weak convergence of the uniform norm $\|\theta - \hat{\theta}\|_{C[s,t]}$ at the the rate $n^{1/4}$. For L^p-norms with $p < \infty$ we have however positive convergence results.

4.2 Asymptotic L^p-confidence regions

We have the following asymptotic results for the L^p-distance between θ and $\hat{\theta}$ under the posterior.

Theorem 4.5. *If $a < \tau^2/\sigma^2$, then for $0 < s \leq t < 1$ and $p = 2, 4, \dots$ we have*

$$n^{p/4}\mathbb{E}\Big(\int_s^t |\theta(u) - \hat{\theta}(u)|^p\, du\Big) \to c_p(t-s)\Big(\frac{\sigma\tau}{2}\Big)^{p/2}$$

under the posterior, and

$$n^{1/4}\|\theta - \hat{\theta}\|_{L^p[s,t]} \xrightarrow{\mathbb{P}} c_p^{1/p}(t-s)^{1/p}\sqrt{\frac{\sigma\tau}{2}},$$

where $c_p = 1 \cdot 3 \cdots (p-3)(p-1)$.

Proof. Let R_n denote the process $\theta - \hat{\theta}$ under the posterior, i.e. R_n is the centered Gaussian process with covariance function (4.1). In this notation, the first statement of the theorem reads

$$n^{p/4}\mathbb{E}\int_s^t R_n^p(u)\, du \to c_p(t-s)\Big(\frac{\sigma\tau}{2}\Big)^{p/2} \tag{4.4}$$

for $p = 2, 4, \dots$. We are going to prove this by induction on p.

Recall that an SDE of the form (2.1) has a unique solution given by (2.2). Hence, the process R_n satisfies the SDE

$$dR_n(t) = \bar{a}(t)R_n(t)\, dt + \tau dW(t),$$

where W is a Brownian motion. By Itô's formula, the process R_n^p then satisfies

$$\begin{aligned} R_n^p(t) - R_n^p(s) = &\int_s^t \Big(p\bar{a}(u)R_n^p(u) + \frac{1}{2}\tau^2 p(p-1)R_n^{p-2}(u)\Big)\, du \\ &+ \tau p \int_s^t R_n^{p-1}(u)\, dW(u). \end{aligned} \tag{4.5}$$

Since R_n is a continuous Gaussian process it is also a measurable random element of $L^q[s,t]$ for all $q \geq 1$, and hence it holds that

$$\mathbb{E}\int_s^t R_n^{2(p-1)}(u)\, du < \infty$$

(cf. e.g. Ledoux and Talagrand (1991)). It follows that the expectation of the stochastic integral term is 0 and we get

$$
\begin{aligned}
n^{(p-2)/4}\mathbb{E}(R_n^p(t) - R_n^p(s)) = pn^{(p-2)/4}\mathbb{E}\int_s^t \bar{a}(u)R_n^p(u)\,du \\
+ \frac{1}{2}\tau^2 p(p-1)n^{(p-2)/4}\mathbb{E}\int_s^t R_n^{p-2}(u)\,du.
\end{aligned}
$$

By equivalence of Gaussian norms and Theorem 3.1 the left-hand side is $O(n^{-1/2})$, so

$$
\begin{aligned}
&-n^{(p-2)/4}\mathbb{E}\int_s^t \bar{a}(u)R_n^p(u)\,du \\
&\qquad = \frac{1}{2}\tau^2(p-1)n^{(p-2)/4}\mathbb{E}\int_s^t R_n^{p-2}(u)\,du + O(n^{-1/2}).
\end{aligned}
\tag{4.6}
$$

By Lemma 4.3, we have

$$
\begin{aligned}
\sqrt{n}\int_s^t R_n^p(u)\,du &= -\frac{\sigma}{\tau}\int_s^t \bar{a}(u)R_n^p(u)\,du \\
&\quad + \int_s^t \bar{a}(u)\Big(\frac{\sqrt{n}}{\bar{a}(u)} + \frac{\sigma}{\tau}\Big)R_n^p(u)\,du \\
&= \Big(-\frac{\sigma}{\tau} + o(1)\Big)\int_s^t \bar{a}(u)R_n^p(u)\,du.
\end{aligned}
\tag{4.7}
$$

Combined with (4.6), we get

$$
n^{p/4}\mathbb{E}\int_s^t R_n^p(u)\,du = \Big(\frac{\sigma}{\tau}+o(1)\Big)\frac{(p-1)\tau^2}{2}n^{(p-2)/4}\mathbb{E}\int_s^t R_n^{p-2}(u)\,du + O(n^{-1/2}).
$$

By induction it now easily follows that (4.4) holds for $p = 2, 4, \ldots$.

The second statement of the theorem asserts that

$$
n^{p/4}\int_s^t R_n^p(u)\,du \xrightarrow{\mathbb{P}} c_p(t-s)\Big(\frac{\sigma\tau}{2}\Big)^{p/2}
$$

for $p = 2, 4, \ldots$, where the arrow denotes convergence in probability. In view of (4.7) it suffices to show that

$$
-n^{(p-2)/4}\int_s^t \bar{a}(u)R_n^p(u)\,du \xrightarrow{\mathbb{P}} c_p(t-s)\frac{\tau}{\sigma}\Big(\frac{\sigma\tau}{2}\Big)^{p/2}, \tag{4.8}
$$

for $p = 2, 4, \ldots$. We are going to prove this by induction on p again. By the relation (4.5) obtained from Itô's formula,

$$
\begin{aligned}
-n^{(p-2)/4}\int_s^t \bar{a}(u)R_n^p(u)\,du &= -\frac{n^{(p-2)/4}}{p}(R_n^p(t) - R_n^p(s)) \\
&\quad + \frac{1}{2}\tau^2(p-1)n^{(p-2)/4}\int_s^t R_n^{p-2}(u)\,du \\
&\quad + \tau n^{(p-2)/4}\int_s^t R_n^{p-1}(u)\,dW(u).
\end{aligned}
\tag{4.9}
$$

Theorem 4.1 and equivalence of Gaussian norms imply that the first term on the right converges to 0 in probability. The last term can be written as $M_n(t)$, where M_n is a contintuous local martingale with $M_n(s) = 0$ and bracket

$$\langle M_n \rangle(t) = \tau^2 n^{(p-2)/2} \int_s^t R_n^{2p-2}(u)\, du.$$

By the first part of the theorem, $\langle M_n \rangle(t) \to 0$ in probability. Hence, by standard results in martingale limit theory, see for instance the exponential inequality on p. 153 of Revuz and Yor (1991), $M_n(t) \to 0$ is probability as well. Statement (4.8) now easily follows by induction. □

For $p = 2, 4, \ldots$ and $0 < s < t < 1$, the theorem shows that for all $\varepsilon > 0$, the posterior mass is asymptotically concentrated on the L^p-ring

$$\Big\{\theta : \frac{c_p^{1/p}(t-s)^{1/p}\sqrt{\sigma\tau/2} - \varepsilon}{n^{1/4}} \le \|\theta - \hat{\theta}\|_{L^p[s,t]} \le \frac{c_p^{1/p}(t-s)^{1/p}\sqrt{\sigma\tau/2} + \varepsilon}{n^{1/4}}\Big\}$$

around the posterior mean $\hat{\theta}$. In particular, we have for every even p a Bayesian asymptotic confidence region around the posterior mean in the L^p-norm, with radius of the order $n^{1/4}$. Since every L^p-norm $\|\theta - \hat{\theta}\|_{L^p[s,t]}$ is bounded by the uniform norm

$$\|\theta - \hat{\theta}\|_{C[s,t]} = \sup_{s \le u \le t} |\theta(u) - \hat{\theta}(u)|,$$

the theorem also implies that for all $\varepsilon > 0$ and $p = 2, 4, \ldots$,

$$\Pi\Big(n^{1/4}\|\theta - \hat{\theta}\|_{C[s,t]} \ge c_p^{1/p}(t-s)^{1/p}\sqrt{\sigma\tau/2} - \varepsilon \,|\, X_1, \ldots, X_n\Big) \to 1.$$

Since by Stirlings approximation, $c_p^{1/p} \ge ((p-1)!)^{1/(2p)} \to \infty$ as $p \to \infty$, we have the following corollary of Theorem 4.5.

Corollary 4.6. *If $a < \tau^2/\sigma^2$, then for $0 < s \le t < 1$ we have*

$$n^{1/4}\|\theta - \hat{\theta}\|_{C[s,t]} \xrightarrow{\mathbb{P}} \infty$$

under the posterior.

This shows again that under the posterior, $n^{1/4}(\theta - \hat{\theta})$ does not converge weakly in the space of continuous functions and that the width of uniform Bayesian confidence bands around the posterior mean, if they exist, must be of strictly larger order than $n^{-1/4}$.

It is natural to ask the "higher order question" of how fast we can let the ε appearing in the Bayesian L^p-confidence ring above tend to zero, while still capturing almost all posterior mass. The following theorem answers this question for $p = 2$, showing that we can take ε of the order $n^{-1/4}$.

Theorem 4.7. *If $a < \tau^2/\sigma^2$, then for $0 < s \le t < 1$ we have the weak convergence*

$$n^{1/4}\Big(n^{1/2}\|\theta - \hat{\theta}\|^2_{L^2[s,t]} - (t-s)\frac{\sigma\tau}{2}\Big) \Rightarrow N\Big(0, (t-s)\frac{\sigma^3\tau}{2}\Big)$$

under the posterior.

Proof. Using the notation of the proof of Theorem 4.5 again, we have to show that we have the weak convergence

$$n^{1/4}\Big(n^{1/2}\int_s^t R_n^2(u)\,du - (t-s)\frac{\sigma\tau}{2}\Big) \Rightarrow N\Big(0, (t-s)\frac{\sigma^3\tau}{2}\Big).$$

By Lemma 4.3 we have

$$\begin{aligned} n^{1/4}\Big|n^{1/2}\int_s^t R_n^2(u)\,du + \frac{\sigma}{\tau}\int_s^t \bar{a}(u)R_n^2(u)\,du\Big| \\ = n^{1/4}\Big|\int_s^t \bar{a}(u)\Big(\frac{\sqrt{n}}{\bar{a}(u)} + \frac{\sigma}{\tau}\Big)R_n^2(u)\,du\Big| \\ = O(n^{-1/4})\Big|\int_s^t \bar{a}(u)R_n^2(u)\,du\Big|. \end{aligned}$$

In the proof of Theorem 4.5 it was shown that the last integral converges in probability (cf. (4.8)), hence

$$n^{1/4}\Big|n^{1/2}\int_s^t R_n^2(u)\,du + \frac{\sigma}{\tau}\int_s^t \bar{a}(u)R_n^2(u)\,du\Big| \xrightarrow{\mathbb{P}} 0.$$

For the proof of the theorem it therefore suffices to show that the random variable Z_n defined by

$$Z_n = n^{1/4}\Big(-\frac{\sigma}{\tau}\int_s^t \bar{a}(u)R_n^2(u)\,du - (t-s)\frac{\sigma\tau}{2}\Big)$$

converges weakly to a $N(0, (t-s)\sigma^3\tau/2)$-distribution.

Relation (4.9) in the proof of Theorem 4.5 shows that

$$Z_n = -\frac{\sigma}{2\tau}n^{1/4}(R_n^2(t) - R_n^2(s)) + \sigma n^{1/4}\int_s^t R_n(u)\,dW(u).$$

The stochastic integral on the right can be written as $M_n(t)$, where M_n is is continuous local martingale with $M_n(s) = 0$ and

$$\langle M_n\rangle(t) = \sigma^2 n^{1/2}\int_s^t R_n^2(u)\,du.$$

By Theorem 4.5,

$$\langle M_n\rangle(t) \xrightarrow{\mathbb{P}} (t-s)\frac{\sigma^3\tau}{2}.$$

Hence, by the central limit theorem for continuous martingales (cf. e.g. Jacod and Shiryaev (1987)), we have the weak convergence $M_n \Rightarrow N(0, (t-s)\sigma^3\tau/2)$. This completes the proof since by Theorem 4.1, $n^{1/4}R_n(s) \to 0$ and $n^{1/4}R_n(t) \to 0$ in probability. □

For the norm of $\theta - \hat{\theta}$ itself, rather than its square, we obtain the following.

Corollary 4.8. *If $a < \tau^2/\sigma^2$, then for $0 < s < t < 1$ we have the weak convergence*

$$n^{1/4}\Big(n^{1/4}\|\theta - \hat{\theta}\|_{L^2[s,t]} - \sqrt{\frac{(t-s)\sigma\tau}{2}}\Big) \Rightarrow N\Big(0, \frac{\sigma^2}{4}\Big)$$

under the posterior.

Proof. This follows from Theorem 4.7 by the delta method, see for instance Van der Vaart (1998), Chapter 3. □

The result shows that asymptotically, the posterior is concentrated on an "L^2-annulus" around the posterior mean, with diameter of the order $n^{-1/4}$ and width of the order $n^{-1/2}$. For $\alpha \in (0,1)$, the annulus

$$\Big\{\theta : \Big|\|\theta - \hat{\theta}\|_{L^2[a,b]} - n^{-1/4}\sqrt{\frac{(b-a)\sigma\tau}{2}}\Big| \leq n^{-1/2}\xi_{\alpha/2}\frac{\sigma}{2}\Big\},$$

is a $(1-\alpha)$-credible region for θ, i.e. it contains a fraction $1-\alpha$ of the posterior mass.

References

Cox, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.* **21**(2), 903–923.

Drignei, D. (2006). Empirical Bayesian analysis for high-dimensional computer output. *Technometrics* **48**(2), 230–240.

Jacod, J. and Shiryaev, A. N. (1987). *Limit theorems for stochastic processes*. Springer-Verlag, Berlin.

Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*. Springer-Verlag, New York.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach spaces*. Springer-Verlag, Berlin.

Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, Massachusetts.

Revuz, D. and Yor, M. (1991). *Continuous martingales and Brownian motion*. Springer-Verlag, Berlin.

Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press, Cambridge.

Van der Vaart, A. W. and Van Zanten, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Annals of Statistics* **36**, 1435–1463.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40**(3), 364–372.