

# Penalized nonparametric drift estimation in a continuous time one-dimensional diffusion process

Eva LÖCHERBACH\*, Dasha LOUKIANOVA†, Oleg LOUKIANOV‡

March 11, 2009

## Abstract

Let  $X$  be a one dimensional positive recurrent diffusion observed in continuous time. Without assuming strict stationarity of the process, we propose a nonparametric estimator of the drift function obtained by penalization. Our estimators belong to a finite-dimensional function space whose dimension is chosen according to the data. Our risk-bounds for the estimator are non-asymptotic and hold in a non-stationary regime.

*Key words* : diffusion process, adaptive estimation, regeneration method, continuous time Markov processes, model selection, deviation inequalities.

*MSC 2000* : 60 F 99, 60 J 35, 60 J 55, 60 J 60, 62 G 99, 62 M 05

## 1 Introduction

Let  $X_t$  be a one-dimensional ergodic diffusion process given by

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t, \quad X_0 = x, \quad (1.1)$$

where  $W$  is a standard Brownian motion. Assuming that the process is positive recurrent but not in the stationary regime, i.e. not starting from the invariant measure, we want to estimate the unknown drift function  $b$  on a fixed interval  $K$  from observations of  $X$  during the time interval  $[0, t]$ , for fixed  $t$ . More precisely we aim at studying nonparametric adaptive estimators for the unknown drift  $b$ .

---

\*Centre de Mathématiques, Faculté de Sciences et Technologie, Université Paris-Est Val-de-Marne, 61 avenue du Général de Gaulle, 94010 Créteil, France. E-mail: locherbach@univ-paris12.fr

†Département de Mathématiques, Université d'Evry-Val d'Essonne, Bd François Mitterrand, 91025 Evry, France. E-mail: dasha.loukianova@univ-evry.fr

‡Département Informatique, IUT de Fontainebleau, Université Paris-Est, route Hurtaut, 77300 Fontainebleau, France. E-mail: oleg@iut-fbleau.fr

Nonparametric estimation for continuous-time diffusion models has been widely studied over the last decades. Let us cite for example the extensive work of Kutoyants (2004) on this subject. The adaptive estimation for the drift at a fixed point has been studied by Spokoiny (2000) who uses Lepskii’s method in order to construct an adaptive procedure. Dalalyan (2005) uses kernel-type estimators and considers a weighted  $L^2$ -risk, where the weight is given by the invariant density. He has to work under quite strong ergodicity assumptions.

Our aim in this paper is twofold. Firstly, we aim at introducing a simple nonparametric estimation procedure based on model selection. Our estimator is obtained by minimizing a contrast function within a fixed finite-dimensional sub-space of  $L^2(K, dx)$  – quite in the spirit of mean square estimation and following ideas presented by Comte et al. (2007) for discretely observed diffusions. These finite-dimensional sub-spaces include spaces such as piecewise polynomials or compactly supported wavelets. The risk we consider for a given estimator  $\hat{b}$  of  $b$  is the expectation of an empirical  $L^2$ -norm defined by

$$E_x \|\hat{b} - b\|_t^2, \text{ where } \|\hat{b} - b\|_t^2 = \frac{1}{t} \int_0^t (\hat{b} - b)^2(X_s) ds.$$

The dimension of the space is chosen by a data-driven method using a penalization.

Secondly, we aim at working under the less restrictive assumptions on the ergodicity properties of the process that seem to be possible. We do not impose the condition of a spectral gap and do not work under the condition of existence of exponential moments for the invariant measure, though we do have to impose the existence of a certain number of moments. Finally, note that we do not work in the stationary regime: the process starts from a fixed point  $x \in K$ , and is not yet in equilibrium. Note also that our approach is non-asymptotic in time. But we have to suppose that  $t \geq t_0$  for some fixed time horizon  $t_0$  that is needed for theoretical reasons and defined precisely later in the text (see theorem 5.2 and in corollary 7.1).

Some results that we use in this paper are interesting as well from a probabilistic point of view. A main ingredient of the proofs is an exponential inequality ensuring that empirical norm and theoretical  $L^2$ -norm are not too far away. A whole probabilistic literature exists on deviation inequalities. We use a deviation inequality given in Loukianova et al. (2009) – see also the references therein for a background on this subject. The important point in the present text is that we can use the so-called regeneration method in order to obtain the inequalities – which is a method that could be extended as well to the case of multi-dimensional diffusions by using Nummelin splitting. Moreover, it is important to have the precise form of all constants appearing in the inequality. This needs some cumbersome work. Some results are needed comparing empirical norms to theoretical ones in a more direct way using local time (see proposition 5.1). These results are purely one-dimensional in spirit.

The paper is organized as follows. In section 2, we describe our framework and give some examples for possible models. Section 3 deals with non-adaptive and section 4 with adaptive drift estimation. In section 5, we collect the necessary probabilistic results : comparison of theoretical and empirical norms and deviation inequality. Section 6, 7 and 8 contain proofs. Finally, section 9 presents some numerical simulations.

**Acknowledgments.** The subject of this paper has been proposed to the authors by Fabienne Comte and Valentine Genon-Catalot during their research period at the University Paris-

Descartes in spring 2008. The authors thank both of them for their kindness, their patience, and last, but not least for all the time spent on discussions and explanations on the paper.

Eva Löcherbach has been partially supported by an ANR projet : Ce travail a bénéficié d'une aide de l'Agence Nationale de la Recherche portant la référence ANR-08-BLAN-0220-01.

## 2 Model assumptions

Let  $X_t$  be a solution of (1.1). We would like to estimate the drift function  $b$  on a fixed interval  $K$ , say  $K = [0, 1]$ . We shall say that  $b$  belongs to the class of models  $\mathcal{M}(M_0, b_0, \gamma)$ , if the following assumptions are fulfilled.

**Assumption 2.1** 1.  $b$  and  $\sigma$  are locally Lipschitz and there is some  $C < 0$ , such that  $|b(x)| \leq C(1 + |x|)$ .

2. There exist  $0 < \sigma_0^2 \leq \sigma_1^2 < \infty$  such that for all  $x$ ,  $\sigma_0^2 \leq \sigma^2(x) \leq \sigma_1^2$ .

3. There exist positive constants  $M_0$  and  $b_0$  such that  $K \subset [-M_0, M_0]$  and, for all  $x$  with  $|x| \leq M_0$ ,  $|b(x)| \leq b_0$ .

4. We suppose that there is a positive constant  $\gamma$  satisfying  $2\gamma > 31\sigma_1^2$  such that for all  $x$  with  $|x| \geq M_0$ ,

$$xb(x) \leq -\gamma.$$

Under this assumption, the diffusion admits a unique invariant probability measure which is given as follows. Define the scale density

$$s(x) = \exp\left(-2 \int_0^x \frac{b(u)}{\sigma^2(u)} du\right),$$

let  $m(x) = 1/(s(x)\sigma^2(x))dx$  and  $M = \int_{-\infty}^{+\infty} m(x)dx$ . Then

$$\mu(dx) = p(x)dx, \text{ where } p(x) = M^{-1}m(x).$$

It follows from the above assumptions that the invariant density  $p$  is bounded from above and below on any compact interval. So we have

$$0 < p_0 \leq p(x) \leq p_1 < \infty \text{ for all } x \in [0, 1]. \quad (2.2)$$

Note that the assumption 2.1.4,  $2\gamma > 31\sigma_1^2$ , implies that hitting times for the diffusion admit a moment of order  $p = 16$ , which will be needed in the sequel (see theorem 6.1).

The following condition will be needed for results concerning adaptive estimators.

**Assumption 2.2** 1. *Nesting condition.*  $(\mathcal{S}_m)_{m \in \mathcal{M}_t}$  is a collection of models  $\mathcal{S}_m \subset L^2(K, dx)$  such that there exists a space denoted by  $\mathcal{S}_t$ , belonging to the collection, such that  $\mathcal{S}_m \subset \mathcal{S}_t$  for all  $m \in \mathcal{M}_t$ . We denote by  $D_m$  the dimension of  $\mathcal{S}_m$ , and by  $D_t$  the dimension of  $\mathcal{S}_t$ . Moreover, we suppose that  $|\mathcal{M}_t| \leq D_t$ .

2. We suppose moreover that there exists  $\Phi_0 > 0$  such that for all  $m \in \mathcal{M}_t$ , for all  $h \in \mathcal{S}_m$ ,

$$\|h\|_\infty \leq \Phi_0 D_m^{1/2} \|h\|.$$

Here,  $\|h\|^2 = \int_K h^2(x) dx$  is the usual  $L^2(K, dx)$ -norm.

3. We suppose that

$$\sum_{m \in \mathcal{M}_t} e^{-D_m} \leq C,$$

where the constant  $C$  does not depend on  $t$ .

4. Dimension condition.

$$D_t \leq t.$$

5. Let  $\{\varphi_1, \dots, \varphi_{D_t}\}$  be an orthonormal basis of  $\mathcal{S}_t \subset L^2(K, dx)$ . We assume that there exists a positive constant  $\Phi_1$  such that for all  $i$ ,

$$\text{card}\{j : \|\varphi_i \varphi_j\|_\infty \neq 0\} \leq \Phi_1.$$

## 2.1 Example for models

We present a collection of models that can be used for estimation. We consider the space of piecewise polynomials, as introduced for example in Baraud et al. (2001) and Comte et al. (2007). Fix an integer  $r \geq 0$ . For  $p \in \mathbb{N}$ , consider the dyadic subintervals  $I_{j,p} = [(j-1)2^{-p}, j2^{-p}]$ , for any  $1 \leq j \leq 2^p$ . On each subinterval  $I_{j,p}$ , we consider polynomials of degree less or equal to  $r$ , so we have polynomials  $\varphi_{j,l}, 0 \leq l \leq r$  of degree  $l$ , such that  $\varphi_{j,l}$  is zero outside  $I_{j,p}$ . Then the space  $\mathcal{S}_m$ , for  $m = (r, p)$ , is defined as the space of all functions that can be written as

$$t(x) = \sum_{j=1}^{2^p} \sum_{l=0}^r t_{j,l} \varphi_{j,l}(x).$$

Hence,  $D_m = (r+1)2^p$ . Then the collection of spaces  $(\mathcal{S}_m, m \in \mathcal{M}_t)$  is such that

$$\mathcal{M}_t = \{m = (r, p), p \geq 1, r \in \{0, \dots, r_{max}\}, 2^p(r_{max} + 1) \leq D_t\}.$$

It is well known, see for instance Comte et al. (2007), that for this model the assumption of norm connection 2.2 2. is satisfied. Note moreover that for a fixed  $\varphi_{j,l} \in \mathcal{S}_t$ ,

$$\text{card}\{(j', l') : \varphi_{j',l'} \varphi_{j,l} \neq 0\} = \text{card}\{(j', l') : \varphi_{j',l'} \neq 0\} \leq r_{max} + 1,$$

which does not depend on  $t$ . Hence assumption 2.2 5. is satisfied. Finally, it is easy to check that also assumption 2.2 3. holds:

$$\begin{aligned} \sum_{m \in \mathcal{M}_t} e^{-D_m} &= \sum_{r=0}^{r_{max}} \sum_{p: 2^p(r_{max}+1) \leq D_t} e^{-(r+1)2^p} \\ &\leq \sum_{r=0}^{r_{max}} \sum_{p: 2^p(r_{max}+1) \leq D_t} e^{-2^p} \\ &\leq (r_{max} + 1) \sum_{k \geq 0} e^{-k} < +\infty, \end{aligned}$$

where the last quantity does not depend on  $t$ .

We could also consider spaces generated by compactly supported wavelets, as considered in Baraud et al. (2001) or in Hoffmann (1999). The most restrictive assumption that we need to guaranty is assumption 2.2 5. It is for that reason that we cannot consider trigonometric spaces for instance.

### 3 Non-adaptive drift estimation

Recall that  $K = [0, 1]$ . We want to estimate the unknown drift function  $b$  on the interval  $K$ . Fix a subspace  $\mathcal{S}_m \subset L^2(K, dx)$  of dimension  $D_m$  with orthonormal basis  $\{\varphi_1, \dots, \varphi_{D_m}\}$ . Denote by  $\|h\|$  the  $L^2(K, dx)$ -norm, and  $\|h\|_\mu^2 = \int_K h^2(x)\mu(dx)$ . We shall write shortly  $b_K(x) := b(x)1_K(x)$  for the restriction of the function  $b$  to the interval  $K$ . In what follows, all functions  $\varphi_i, \psi_i$  will be restricted to  $K$ , i.e. zero outside  $K$ .

The estimator  $\hat{b}_m$  of  $b$  on  $K$  is a minimizer on  $\mathcal{S}_m$  of the following contrast function:

$$\gamma_t(h) = \frac{1}{t} \int_0^t h^2(X_s) ds - \frac{2}{t} \int_0^t h(X_s) dX_s. \quad (3.3)$$

Put

$$\|h\|_t^2 = \frac{1}{t} \int_0^t h^2(X_s) ds \quad (3.4)$$

and denote the corresponding quadratic form on  $\mathcal{S}_m$  by

$$T_m(h, f) = \frac{1}{t} \int_0^t h(X_s)f(X_s) ds \text{ for all } f, h \in \mathcal{S}_m.$$

To insure the existence of  $\hat{b}_m$  we need some condition under which  $T_m$  is positive-definite. Let

$$A_t = \{\forall m \in \mathcal{M}_t; \quad \rho_m \geq t^{-1/2}\}, \quad (3.5)$$

where  $\rho_m = \inf_{h \in \mathcal{S}_m; \|h\|=1} T_m(h, h)$ . We finally define

$$\hat{b}_m = \arg \min_{h \in \mathcal{S}_m} \gamma_t(h) \text{ on } A_t \text{ and } \hat{b}_m = 0 \text{ on } A_t^c. \quad (3.6)$$

Since  $T_m$  is not degenerated on  $A_t$ , the minimizer exists and is unique on  $A_t$ .

We define the risk of the estimator  $\hat{b}_m$  as

$$E_x \|\hat{b}_m - b_K\|_t^2 = E_x \left( \frac{1}{t} \int_0^t (\hat{b}_m - b_K)^2(X_s) ds \right).$$

Thus, our risk is the expectation of an empirical norm.

Let  $b_m$  be the  $L^2(K, dx)$ -projection of  $b_K$  onto  $\mathcal{S}_m$ . We have the following first result.

**Proposition 3.1** *Suppose that  $t \geq t_0$  where  $t_0$  is given by corollary 7.1. Under our assumptions 2.1 and 2.2, for all  $b \in \mathcal{M}(M_0, b_0, \gamma)$ ,*

$$\begin{aligned} E_x \|\hat{b}_m - b_K\|_t^2 &\leq 3\kappa \|b_m - b_K\|^2 + 16\sigma_1^2 \frac{\kappa}{p_0} \frac{D_m}{t} \\ &\quad + C (b_0^2 + \sigma_1^2 \Phi_0^2) t^{-1}. \end{aligned} \quad (3.7)$$

Here,  $\kappa = \kappa(t) = \frac{2}{\sigma_0^2} (\frac{2}{t} + \frac{2\sigma_1}{\sqrt{t}} + 2b_0 + \frac{\sigma_1^2}{2})$  (see proposition 5.1 below), and  $C$  is a positive constant.

**Proof** The proof follows ideas of Comte et al. (2007). From the definition of  $\gamma_t$  it follows that on  $A_t$ ,

$$\hat{b}_m = \sum_{i=1}^{D_m} \hat{\alpha}_i \varphi_i,$$

with  $\hat{\alpha}$  satisfying

$$T\hat{\alpha} = \frac{1}{t} \int_0^t \varphi(X_s) dX_s, \quad T_{ij} = \frac{1}{t} \int_0^t \varphi_i(X_s) \varphi_j(X_s) ds. \quad (3.8)$$

Here,

$$\int_0^t \varphi(X_s) dX_s = \begin{pmatrix} \int_0^t \varphi_1(X_s) dX_s \\ \vdots \\ \int_0^t \varphi_{D_m}(X_s) dX_s \end{pmatrix}.$$

Define

$$\Omega_t = \left\{ h \in S_t, \quad \frac{1}{2}\mu(h^2) \leq \|h\|_t^2 \leq \frac{3}{2}\mu(h^2) \right\}. \quad (3.9)$$

Let  $b_m = \sum_{i=1}^{D_m} \alpha_i \varphi_i$  be the  $L^2(K, dx)$ -projection of  $b_K$  onto  $\mathcal{S}_m$ . In what follows it will be useful to pass to an orthonormal basis  $\{\psi_1, \dots, \psi_{D_m}\}$  of  $\mathcal{S}_m$  viewed as a subspace of  $L^2(K, d\mu)$ . Hence, our estimator can be rewritten as

$$\hat{b}_m = \sum_{i=1}^{D_m} \hat{\beta}_i \psi_i, \quad \text{and} \quad b_m = \sum_{i=1}^{D_m} \beta_i \psi_i.$$

Then for any two functions  $h$  and  $g$ ,

$$\gamma_t(h) - \gamma_t(g) = \|h - b_K\|_t^2 - \|g - b_K\|_t^2 - \frac{2}{t} \int_0^t (h - g)(X_s) \sigma(X_s) dW_s,$$

whence

$$\|\hat{b}_m - b_K\|_t^2 \mathbf{1}_{A_t} \leq \|b_m - b_K\|_t^2 \mathbf{1}_{A_t} + 2 \sum_{i=1}^{D_m} (\hat{\beta}_i - \beta_i) \left( \frac{1}{t} \int_0^t \psi_i(X_s) dW_s \right) \mathbf{1}_{A_t}. \quad (3.10)$$

We will investigate this equation on  $A_t \cap \Omega_t$  on the one hand and on  $A_t \cap \Omega_t^c$  on the other hand. We start by treating (3.10) on  $A_t \cap \Omega_t$ . Using Cauchy-Schwartz's inequality, and noting that  $\sum_{i=1}^{D_m} (\hat{\beta}_i - \beta_i)^2 = \|\hat{b}_m - b_m\|_\mu^2$ , we have

$$\begin{aligned} \|\hat{b}_m - b_K\|_t^2 \mathbf{1}_{\Omega_t \cap A_t} &\leq \|b_m - b_K\|_t^2 \mathbf{1}_{\Omega_t \cap A_t} + 2 \sum_{i=1}^{D_m} (\hat{\beta}_i - \beta_i) \left( \frac{1}{t} \int_0^t \psi_i(X_s) \sigma(X_s) dW_s \right) \mathbf{1}_{\Omega_t \cap A_t} \\ &\leq \|b_m - b_K\|_t^2 + \frac{1}{8} \|\hat{b}_m - b_m\|_\mu^2 \mathbf{1}_{\Omega_t \cap A_t} \\ &\quad + 8 \sum_{i=1}^{D_m} \left( \frac{1}{t} \int_0^t \psi_i(X_s) \sigma(X_s) dW_s \right)^2. \end{aligned} \quad (3.11)$$

Then on  $\Omega_t \cap A_t$ ,

$$\frac{1}{8} \|\hat{b}_m - b_m\|_\mu^2 \mathbf{1}_{\Omega_t \cap A_t} \leq \frac{1}{2} (\|\hat{b}_m - b_K\|_t^2 + \|b_m - b_K\|_t^2) \mathbf{1}_{\Omega_t \cap A_t}.$$

Plugging this into (3.11) gives

$$\|\hat{b}_m - b_K\|_t^2 \mathbf{1}_{\Omega_t \cap A_t} \leq 3 \|b_m - b_K\|_t^2 + 16 \sum_{i=1}^{D_m} \left( \frac{1}{t} \int_0^t \psi_i(X_s) \sigma(X_s) dW_s \right)^2.$$

We have

$$E_x \|\hat{b}_m - b_K\|_t^2 \mathbf{1}_{\Omega_t \cap A_t} \leq \frac{3}{t} E_x \int_0^t (b_m - b_K)^2(X_s) ds + \frac{16\sigma_1^2}{t^2} \sum_{i=1}^{D_m} E_x \int_0^t \psi_i^2(X_s) ds.$$

Using proposition 5.1, we can write for any positive function  $f$  having support on  $K$ ,

$$E_x \int_0^t f(X_s) ds \leq \kappa t \int_K f d\lambda,$$

where the constant  $\kappa$  is explicitly given in proposition 5.1 and does only depend on the model constants  $b_0, \sigma_0, \sigma_1$ . Using this estimation, we obtain the following bound for the integrated risk restricted on  $\Omega_t \cap A_t$ :

$$E_x \|\hat{b}_m - b_K\|_t^2 \mathbf{1}_{\Omega_t \cap A_t} \leq 3\kappa \|b_m - b_K\|^2 + 16\sigma_1^2 \frac{\kappa}{p_0} \frac{D_m}{t}. \quad (3.12)$$

We now consider the risk restricted on  $\Omega_t^c \cap A_t$ . Let  $\tilde{b}_m$  be the almost surely defined orthogonal projection of  $b_K$  onto  $\mathcal{S}_m$  w.r.t.  $\|\cdot\|_t$ . We have

$$\begin{aligned} \|b_K - \hat{b}_m\|_t^2 \mathbf{1}_{\Omega_t^c \cap A_t} &= \|b_K - \tilde{b}_m\|_t^2 \mathbf{1}_{\Omega_t^c \cap A_t} + \|\tilde{b}_m - \hat{b}_m\|_t^2 \mathbf{1}_{\Omega_t^c \cap A_t} \\ &\leq \|b_K\|_t^2 \mathbf{1}_{\Omega_t^c} + \|\tilde{b}_m - \hat{b}_m\|_t^2 \mathbf{1}_{\Omega_t^c \cap A_t}. \end{aligned} \quad (3.13)$$

From the definition of  $\tilde{b}_m$  it follows that  $\tilde{b}_m = \sum_{i=1}^{D_m} \tilde{\alpha}_i \varphi_i$ , with  $\tilde{\alpha}$  satisfying

$$T\tilde{\alpha} = \frac{1}{t} \int_0^t \varphi(X_s) b(X_s) ds. \quad (3.14)$$

Recall that on  $A_t$ ,  $\hat{b}_m = \sum_{i=1}^{D_m} \hat{\alpha}_i \varphi_i$ , with  $\hat{\alpha}$  given by (3.8). Hence on  $A_t$ , we can write  $\hat{\alpha} - \tilde{\alpha} = T^{-1}M_t$ , where

$$M_t = \frac{1}{t} \int_0^t \varphi(X_s) \sigma(X_s) dW_s = \begin{pmatrix} \frac{1}{t} \int_0^t \varphi_1(X_s) \sigma(X_s) dW_s \\ \vdots \\ \frac{1}{t} \int_0^t \varphi_{D_m}(X_s) \sigma(X_s) dW_s \end{pmatrix}.$$

So on  $A_t$  we have  $\hat{b}_m - \tilde{b}_m = \varphi^*(\hat{\alpha} - \tilde{\alpha}) = \varphi^* T^{-1} M_t$ , where  $\varphi^* = (\varphi_1, \dots, \varphi_{D_m})$ , and (we denote by  $*$  the matrix-transposition operation),

$$(\hat{b}_m - \tilde{b}_m)^2(X_s) = M_t^* (T^*)^{-1} \varphi \varphi^*(X_s) T^{-1} M_t.$$

So,

$$\|\tilde{b}_m - \hat{b}_m\|_t^2 = \frac{1}{t} \int_0^t (\tilde{b}_m - \hat{b}_m)^2(X_s) ds = M_t^* (T^*)^{-1} T T^{-1} M_t = M_t^* T^{-1} M_t = \langle T^{-1} M_t, M_t \rangle,$$

which gives

$$\|\tilde{b}_m - \hat{b}_m\|_t^2 \mathbf{1}_{\Omega_t^c \cap A_t} \leq \frac{1}{t^{-1/2}} \|M_t\|_t^2 \mathbf{1}_{\Omega_t^c} = t^{1/2} \sum_{i=1}^{D_m} \left( \frac{1}{t} \int_0^t \varphi_i(X_s) \sigma(X_s) dW_s \right)^2 \mathbf{1}_{\Omega_t^c}. \quad (3.15)$$

Using our assumptions on  $b$  we have

$$E_x \|b_K\|_t^2 \mathbf{1}_{\Omega_t^c} \leq b_0^2 P_x(\Omega_t^c).$$

Using Burkholder-Davis-Gundy inequalities and the hypothesis  $\|\varphi_i^2\|_\infty \leq \Phi_0^2 D_m$ , it follows from (3.15),

$$\begin{aligned} E_x \|\tilde{b}_m - \hat{b}_m\|_t^2 \mathbf{1}_{\Omega_t^c \cap A_t} &\leq \frac{t^{1/2}}{t^2} \sum_{i=1}^{D_m} E_x \left( \left( \int_0^t \varphi_i(X_s) \sigma(X_s) dW_s \right)^2 \mathbf{1}_{\Omega_t^c} \right) \\ &\leq t^{-3/2} \sum_{i=1}^{D_m} \sqrt{E_x \left( \int_0^t \varphi_i(X_s) \sigma(X_s) dW_s \right)^4 P_x(\Omega_t^c)} \\ &\leq t^{-3/2} \sum_{i=1}^{D_m} \sqrt{C(4) E_x \left( \int_0^t \varphi_i^2(X_s) \sigma^2(X_s) ds \right)^2 P_x(\Omega_t^c)}. \end{aligned}$$

Here,  $C(4)$  is a Burkholder-Davis-Gundy constant. But

$$\int_0^t \varphi_i^2(X_s) \sigma^2(X_s) ds \leq \Phi_0^2 D_m \sigma_1^2 t,$$

hence

$$\begin{aligned} E_x \|\tilde{b}_m - \hat{b}_m\|_t^2 \mathbf{1}_{\Omega_t^c \cap A_t} &\leq \sqrt{C(4)} t^{-3/2} \sum_{i=1}^{D_m} \sqrt{\Phi_0^4 D_m^2 \sigma_1^4 t^2 P_x(\Omega_t^c)} \\ &\leq \sqrt{C(4)} \sigma_1^2 \Phi_0^2 t^{-1/2} D_m^2 \sqrt{P_x(\Omega_t^c)}. \end{aligned} \quad (3.16)$$

From (3.13), the integrated risk on  $\Omega_t^c \cap A_t$  satisfies

$$E_x \|b_K - \hat{b}_m\|_t^2 \mathbf{1}_{\Omega_t^c \cap A_t} \leq (b_0^2 + C \sigma_1^2 \Phi_0^2 t^{-1/2} D_m^2) \sqrt{P_x(\Omega_t^c)} \quad (3.17)$$

$$\leq (b_0^2 + C \sigma_1^2 \Phi_0^2) t^{-1/2} D_m^2 \sqrt{P_x(\Omega_t^c)}. \quad (3.18)$$

As a consequence, the full integrated risk satisfies, on  $A_t$ , since  $D_m^2 \leq t^2$ ,

$$\begin{aligned} E_x \|\hat{b}_m - b_K\|_t^2 \mathbf{1}_{A_t} &\leq 3\kappa \|b_m - b_K\|^2 + 16\sigma_1^2 \frac{\kappa}{p_0} \frac{D_m}{t} \\ &\quad + (b_0^2 + C \sigma_1^2 \Phi_0^2) t^{3/2} \sqrt{P_x(\Omega_t^c)}. \end{aligned} \quad (3.19)$$

This - together with corollary 7.1 and with theorem 6.1, applied with  $p = 12$  - finishes our proof. •

**Remark 3.2** *In the case when  $X$  is in the stationary regime, i.e. starting from the invariant measure  $\mu$ , the above estimation can be improved to*

$$\begin{aligned} E_\mu \|\hat{b}_m - b_K\|_t^2 &\leq 3p_1 \|b_m - b_K\|^2 + 16\sigma_1^2 \frac{D_m}{t} \\ &\quad + C (b_0^2 + \sigma_1^2 \Phi_0^2) t^{-1}. \end{aligned} \quad (3.20)$$

Let us finish this section with some comments on (3.7). It is natural to choose the dimension  $D_m$  that balances the bias term  $\|b_m - b_K\|^2$  and the variance term which is of order  $D_m/t$ . If one assumes that  $b_K$  belongs to some Besov space  $B_{2,\infty}^\alpha([0,1])$ , then it can be shown that  $\|b_m - b_K\|^2 \leq C D_m^{-2\alpha}$ , see for example Hoffmann (1999). Thus the best choice of  $D_m$  is to take

$$D_m = t^{\frac{1}{2\alpha+1}}$$

and then we obtain

$$E_x (\|\hat{b}_m - b_K\|_t^2) \leq C t^{-\frac{2\alpha}{2\alpha+1}} + C (b_0^2 + \sigma_1^2 \Phi_0^2) t^{-1},$$

and this yields exactly the classical non-parametric rate  $t^{-\frac{2\alpha}{2\alpha+1}}$ . This choice however supposes the knowledge of the regularity  $\alpha$  of the unknown drift function, and that is why we have to introduce an adaptive estimation scheme, in order to choose automatically the best dimension  $D_m$  in the case when we do not know the regularity  $\alpha$ .

## 4 Adaptive drift estimation

We now introduce an adaptive scheme of estimation that ensures an automatic choice of the best linear subspace  $\mathcal{S}_m$ , i.e. the best dimension  $D_m$  adapted to the estimation procedure. We do this by adding a penalization term  $pen(m)$  that will be chosen later. So put

$$\hat{m} := \arg \min_{m \in \mathcal{M}_t} \left[ \gamma_t(\hat{b}_m) + pen(m) \right]. \quad (4.21)$$

Then the estimator that we propose is the following adaptive estimator

$$\hat{b}_{\hat{m}} := \begin{cases} \sum_n 1_{\{\hat{m}=n\}} \hat{b}_n & \text{on } A_t \\ 0 & \text{on } A_t^c \end{cases}. \quad (4.22)$$

We have the following theorem.

**Theorem 4.1** *Suppose that assumptions 2.1 and 2.2 are satisfied. Suppose that  $t \geq t_0$ , where  $t_0$  is defined in corollary 7.1. Let*

$$pen(m) \geq \chi \sigma_1^2 \frac{D_m}{t},$$

where  $\chi$  is a universal constant. Then we have for all  $b \in \mathcal{M}(M_0, b_0, \gamma)$ ,

$$E_x \|\hat{b}_{\hat{m}} - b_K\|_t^2 \leq 3\kappa \inf_{m \in \mathcal{M}_t} (\|b_m - b_K\|^2 + pen(m)) + \frac{K'}{t}, \quad (4.23)$$

where  $\kappa = \kappa(t) = \frac{2}{\sigma_0^2} \left( \frac{2}{t} + \frac{2\sigma_1}{\sqrt{t}} + 2b_0 + \frac{\sigma_1^2}{2} \right)$  (compare to proposition 5.1).

**Remark 4.2** *Compare (4.23) to (3.7). (4.23) means that the adaptive estimator  $\hat{b}_{\hat{m}}$  achieves automatically the bias-variance equilibrium (not exactly, of course, but almost).*

**Proof** Put

$$\nu_t(f) := \frac{1}{t} \int_0^t f(X_s) \sigma(X_s) dW_s.$$

The same argument that yields (3.10) in the non-adaptive case gives for any  $m \in \mathcal{M}_t$ ,

$$\|\hat{b}_{\hat{m}} - b_K\|_t^2 1_{A_t} \leq \|b_m - b_K\|_t^2 1_{A_t} + 2\nu_t(\hat{b}_{\hat{m}} - b_m) 1_{A_t} + (pen(m) - pen(\hat{m})) 1_{A_t}. \quad (4.24)$$

Here, a special attention has to be paid to the term  $\nu_t(\hat{b}_{\hat{m}} - b_m)$ , since it is not a priori clear that this stochastic integral is well-defined. On  $\hat{m} = n$ ,  $\hat{b}_{\hat{m}} - b_m$  is an element of  $S_n + \mathcal{S}_m$  viewed as linear subspace of  $L^2(K, \mu)$ . Put  $k = \dim(S_n + \mathcal{S}_m)$  and let  $\{\psi_1, \dots, \psi_k\}$  be an orthonormal basis of this subspace. Then  $1_{\{\hat{m}=n\}}(\hat{b}_{\hat{m}} - b_m) = 1_{\{\hat{m}=n\}} \sum_{i=1}^k \hat{\beta}_i \psi_i$ , and we define on  $\hat{m} = n$ ,

$$\nu_t(\hat{b}_{\hat{m}} - b_m) := \sum_{i=1}^k \hat{\beta}_i \nu_t(\psi_i).$$

Hence,  $\nu_t(\hat{b}_{\hat{m}} - b_m)$  is well-defined and linear. Thus we may write

$$\nu_t(\hat{b}_{\hat{m}} - b_m) \leq \|\hat{b}_{\hat{m}} - b_m\|_\mu \cdot \nu_t \left( \frac{\hat{b}_{\hat{m}} - b_m}{\|\hat{b}_{\hat{m}} - b_m\|_\mu} \right) \leq \|\hat{b}_{\hat{m}} - b_m\|_\mu \cdot \sup_{h \in \mathcal{S}_m + \mathcal{S}_{\hat{m}}, \|h\|_\mu = 1} |\nu_t(h)|.$$

Write for short

$$G_m(m') := \sup_{h \in \mathcal{S}_m + \mathcal{S}_{m'}, \|h\|_\mu = 1} |\nu_t(h)|.$$

We now investigate (4.24). First, on  $A_t \cap \Omega_t$ , using that  $2ab \leq \frac{1}{8}a^2 + 8b^2$ ,

$$\begin{aligned} \|\hat{b}_{\hat{m}} - b_K\|_t^2 &\leq \|b_m - b_K\|_t^2 + 2\|\hat{b}_{\hat{m}} - b_m\|_\mu G_m(\hat{m}) + [\text{pen}(m) - \text{pen}(\hat{m})] \\ &\leq \|b_m - b_K\|_t^2 + \frac{1}{8}\|\hat{b}_{\hat{m}} - b_m\|_\mu^2 + 8 G_m^2(\hat{m}) + [\text{pen}(m) - \text{pen}(\hat{m})] \\ &\leq \|b_m - b_K\|_t^2 + \frac{1}{2} \left( \|\hat{b}_{\hat{m}} - b_K\|_t^2 + \|b_K - b_m\|_t^2 \right) \\ &\quad + 8 G_m^2(\hat{m}) + [\text{pen}(m) - \text{pen}(\hat{m})] \\ &\leq \frac{3}{2}\|b_m - b_K\|_t^2 + \frac{1}{2}\|\hat{b}_{\hat{m}} - b_K\|_t^2 + 8 G_m^2(\hat{m}) + [\text{pen}(m) - \text{pen}(\hat{m})]. \end{aligned} \quad (4.25)$$

This yields finally, on  $A_t \cap \Omega_t$ ,

$$\|\hat{b}_{\hat{m}} - b_K\|_t^2 \leq 3\|b_m - b_K\|_t^2 + 16 G_m^2(\hat{m}) + 2 [\text{pen}(m) - \text{pen}(\hat{m})]. \quad (4.26)$$

Now, as in Comte et al. (2007), put  $p(m, m') := p(m) + p(m')$ , where

$$p(m) := \chi \sigma_1^2 \frac{D_m}{t} \quad (4.27)$$

and where  $\chi$  is a universal constant. Then

$$\begin{aligned} G_m^2(\hat{m})1_{\Omega_t} &\leq [(G_m^2(\hat{m}) - p(m, \hat{m}))1_{\Omega_t}]_+ + p(m, \hat{m}) \\ &\leq \sum_{n \in \mathcal{M}_t} [(G_m^2(n) - p(m, n))1_{\Omega_t}]_+ + p(m, \hat{m}). \end{aligned}$$

Now we choose  $\text{pen}(m)$  such that  $8p(m, m') \leq \text{pen}(m) + \text{pen}(m')$ , i.e.

$$\text{pen}(m) := 8\chi \sigma_1^2 \frac{D_m}{t}.$$

We have:

**Proposition 4.3** *Under the assumptions of theorem 4.1,*

$$E_x \left( (G_m^2(m') - p(m, m'))1_{\Omega_t} \right)_+ \leq 1, 6\chi \sigma_1^2 \frac{1}{t} e^{-D_{m'}}. \quad (4.28)$$

The proof of proposition 4.3 will be given in section 8 below.

For any  $n$ , let  $\{\varphi_1^n, \dots, \varphi_{D_n}^n\}$  be an orthonormal basis of  $S_n$ . On  $A_t \cap \Omega_t^c$ , using (3.13) and (3.15), we have

$$\begin{aligned}
\|\hat{b}_{\hat{m}} - b_K\|_t^2 \mathbf{1}_{\{A_t \cap \Omega_t^c\}} &= \sum_{n \in \mathcal{M}_t} \mathbf{1}_{\{\hat{m}=n; A_t \cap \Omega_t^c\}} \|\hat{b}_n - b_K\|_t^2 \\
&\leq \|b_K\|_t^2 \mathbf{1}_{\Omega_t^c} + \sum_{n \in \mathcal{M}_t} \mathbf{1}_{\{\hat{m}=n\}} \|\tilde{b}_n - \hat{b}_n\|_t^2 \mathbf{1}_{\Omega_t^c \cap A_t} \\
&\leq \|b_K\|_t^2 \mathbf{1}_{\Omega_t^c} + \sum_{n \in \mathcal{M}_t} \mathbf{1}_{\{\hat{m}=n\}} t^{1/2} \sum_{i=1}^{D_n} \left( \frac{1}{t} \int_0^t \varphi_i^n(X_s) \sigma(X_s) dW_s \right)^2 \mathbf{1}_{\Omega_t^c} \\
&\leq \|b_K\|_t^2 \mathbf{1}_{\Omega_t^c} + \sum_{n \in \mathcal{M}_t} t^{1/2} \sum_{i=1}^{D_n} \left( \frac{1}{t} \int_0^t \varphi_i^n(X_s) \sigma(X_s) dW_s \right)^2 \mathbf{1}_{\Omega_t^c},
\end{aligned}$$

The same calculus that yields (3.17) now gives

$$E_x \|\hat{b}_{\hat{m}} - b_K\|_t^2 \mathbf{1}_{\{A_t \cap \Omega_t^c\}} \leq C (b_0^2 + \sigma_1^2 \Phi_0^2) t^{-1/2} D_t^2 |\mathcal{M}_t| \sqrt{P_x(\Omega_t^c)}. \quad (4.29)$$

(4.26), (4.28) and (4.29) yield finally, for any  $m$ , using assumption 2.2,

$$\begin{aligned}
E_x \|\hat{b}_{\hat{m}} - b_K\|_t^2 \mathbf{1}_{A_t} &\leq 3E_x \|b_m - b_K\|_t^2 + 4pen(m) + 16 \sum_{n \in \mathcal{M}_t} 1, 6\chi \sigma_1^2 \frac{1}{t} e^{-D_n} \\
&\quad + C (b_0^2 + \sigma_1^2 \Phi_0^2) t^{-1/2} D_t^2 |\mathcal{M}_t| \sqrt{P_x(\Omega_t^c)} \\
&\leq 3\kappa \|b_m - b\|^2 + 4pen(m) + C\chi \sigma_1^2 \frac{1}{t} \\
&\quad + C(b_0^2, \sigma_1^2) t^{-1/2} D_t^3 P_x(\Omega_t^c)^{1/2}.
\end{aligned}$$

Now, recall that by theorem 6.1, since  $D_t^3 \leq t^3$ , taking  $p = 16$ ,

$$P_x(\Omega_t^c)^{1/2} \leq Ct^{-7/2}$$

and by corollary 7.1,

$$P_x(A_t^c) \leq Ct^{-1}.$$

As a consequence,

$$E_x \|\hat{b}_{\hat{m}} - b_K\|_t^2 \leq 3\kappa \inf_{m \in \mathcal{M}_t} (\|b_m - b\|^2 + pen(m)) + C\chi \sigma_1^2 \frac{1}{t} + C(b_0^2, \sigma_1^2) t^{-1}.$$

This finishes the proof. •

## 5 Probabilistic tools

In this section, we give some probabilistic results that are needed for the proofs.

## 5.1 Comparing empirical and theoretical norms

It is important to be able to compare empirical and theoretical norms. One way of doing this is given by the next proposition.

**Proposition 5.1** *For any positive function  $f$  having support on  $K$  we have*

$$\frac{1}{t} E_x \int_0^t f(X_s) ds \leq \kappa(t) \int_0^1 f(x) dx,$$

where  $\kappa(t) = \frac{2}{\sigma_0^2} \left( \frac{2}{t} + \frac{2\sigma_1}{\sqrt{t}} + 2b_0 + \frac{\sigma_1^2}{2} \right)$ .

**Proof** By the occupation time formula and since  $f$  has support in  $K = [0, 1]$ ,

$$E_x \int_0^t f(X_s) ds = \int_0^1 f(y) \frac{2}{\sigma^2(y)} E_x L_t^y dy.$$

We will derive a bound on  $E_x L_t^y$  for  $y \in [0, 1]$ . We have

$$E_x L_t^0 - E_x |L_t^y - L_t^0| \leq E_x L_t^y \leq E_x L_t^0 + E_x |L_t^y - L_t^0| \quad (5.30)$$

and

$$|L_t^y - L_t^0| \leq |y| + \left| \int_0^t \mathbf{1}_{\{0 < X_s < y\}} \sigma(X_s) dW_s \right| + \int_0^t \mathbf{1}_{\{0 < X_s < 1\}} |b(X_s)| ds. \quad (5.31)$$

Taking expectation we obtain

$$E_x \int_0^t \mathbf{1}_{\{0 < X_s < 1\}} |b(X_s)| ds \leq t b_0, \quad (5.32)$$

and by norm inclusion and isometry,

$$\begin{aligned} E_x \left| \int_0^t \mathbf{1}_{\{0 < X_s < y\}} \sigma(X_s) dW_s \right| &\leq \left( E_x \left( \int_0^t \mathbf{1}_{\{0 < X_s < y\}} \sigma(X_s) dW_s \right)^2 \right)^{1/2} \\ &\leq \left( E_x \left( \int_0^t \mathbf{1}_{\{0 < X_s < 1\}} \sigma^2(X_s) ds \right) \right)^{1/2} \leq \sigma_1 \sqrt{t}. \end{aligned} \quad (5.33)$$

In conclusion,

$$E_x L_t^y \leq E_x L_t^0 + 1 + \sigma_1 \sqrt{t} + t b_0 = C_0 + L, \quad (5.34)$$

where  $L := 1 + \sigma_1 \sqrt{t} + t b_0$  and  $C_0 = E_x L_t^0$ . We also have  $C_0 - L \leq E_x L_t^y \leq C_0 + L$ , so

$$t \geq E_x \int_0^t \mathbf{1}_{[0,1]}(X_s) ds = \int_0^1 \frac{2 E_x L_t^y}{\sigma^2(y)} dy \geq \frac{2(C_0 - L)}{\sigma_1^2},$$

whence

$$C_0 \leq L + \sigma_1^2 t / 2,$$

and thus finally,

$$E_x L_t^y \leq 2L + \sigma_1^2 t / 2 = 2(1 + \sigma_1 \sqrt{t} + t b_0) + \sigma_1^2 t / 2.$$

This concludes the proof. •

## 5.2 Deviation inequality

In this section, we give a useful deviation inequality for the one-dimensional ergodic diffusion process  $X$ . It is a consequence of deviation inequalities obtained by Loukianova et al. (2009), paying in particular attention to all constants, which – in view of statistical applications – is crucial.

**Theorem 5.2** *Let  $f$  be measurable bounded function with compact support  $K = [0, 1]$  such that  $\mu(f) > 0$ . Suppose that there exists  $p > 1$  such that*

$$2\gamma > (2p - 1)\sigma_1^2. \quad (5.35)$$

*Then for any  $0 < \varepsilon \leq 1$ , we have the following polynomial bound.*

$$P_x \left( \left| \frac{1}{t} \int_0^t f(X_s) ds - \mu(f) \right| > \varepsilon \right) \leq K(p)t^{-p/2}\varepsilon^{-p}\mu(|f|)^p, \quad (5.36)$$

*where  $K(p)$  is positive and finite, depending on the coefficients of the diffusion.*

**Remark 5.3** *The proof of the above theorem relies on the regeneration method, that consists in cutting the trajectory of the process into independent excursions out of a fixed point. More precisely, let  $0 \leq a < b \leq 1$  and define a sequence of stopping times  $(S_n)_n, (R_n)_n$  as follows.  $S_0 = R_0 = 0$ ,*

$$S_1 = \inf\{t \geq 0 : X_t = b\}, \quad R_1 = \inf\{t \geq S_1 : X_t = a\}, \quad (5.37)$$

*and for any  $n \geq 1$ ,  $S_{n+1} = R_n + S_1 \circ \theta_{R_n}$ ,  $R_{n+1} = R_n + R_1 \circ \theta_{R_n}$ . Then the sequence of stopping times  $(R_n)_n$  defines the sequence of successive excursions of the process out of the point  $a$ . The method applies if we dispose of a moment of order  $p$  for the regeneration time  $R_{n+1} - R_n$ , and this moment exists under the condition (5.35).*

## 6 Control of $\Omega_t$

Recall the definition of  $\Omega_t$  in (3.9). We use the above deviation theorem 5.2 in order to control the decay of  $\Omega_t^c$ .

**Theorem 6.1** *Grant assumption 2.2 and suppose that*

$$2\gamma > (2p - 1)\sigma_1^2.$$

*Then we have that*

$$P_x(\Omega_t^c) \leq Ct^{-\frac{1}{2}(p-2)},$$

*where  $C$  denotes a universal constant not depending on the dimension.*

**Proof** Recall that  $\|f\|$  denotes the usual  $L^2(K, dx)$ -norm. For any function  $f$ , write

$$Z_t(f) := \frac{1}{t} \int_0^t f(X_s) ds - \mu(f).$$

Since  $\|f\|_\mu^2 = 1$  implies that  $\|f\|^2 \leq p_0^{-1}$ , we have that

$$P_x(\Omega_t^c) \leq P_x\left(\sup_{f \in \mathcal{S}_t, \|f\| \leq 1} |Z_t(f^2)| > 0, 5p_0\right).$$

Let  $\{\varphi_1, \dots, \varphi_{D_t}\}$  be an orthonormal basis of  $\mathcal{S}_t \subset L^2(K, dx)$  and note that any function  $f$  with  $\|f\| \leq 1$  can be written as

$$f = \sum_{i=1}^{D_t} a_i \varphi_i \text{ with } \sum a_i^2 \leq 1.$$

Therefore,

$$\begin{aligned} P_x(\Omega_t^c) &\leq P_x\left(\sup_{\|f\| \leq 1} |Z_t(f^2)| > 0, 5p_0\right) \\ &= P_x\left(\sup_{\sum a_i^2 \leq 1} \sum_{i,j} a_i a_j |Z_t(\varphi_i \varphi_j)| > 0, 5p_0\right). \end{aligned}$$

Write

$$C_{ij} := \mu(|\varphi_i \varphi_j|)$$

and fix some positive number  $\varepsilon$ . On the set

$$\{|Z_t(\varphi_i \varphi_j)| \leq C_{ij} \varepsilon, \forall i, j\},$$

we have that

$$\sup_{\sum a_i^2 \leq 1} a_i a_j |Z_t(\varphi_i \varphi_j)| \leq \varepsilon \varrho(C),$$

where  $\varrho(C)$  is the biggest eigenvalue of the matrix  $C$ . Then choosing  $\varepsilon := p_0/(4\varrho(C))$ , we conclude that

$$P_x(\Omega_t^c) \leq P_x(\exists i, j : |Z_t(\varphi_i \varphi_j)| > C_{ij} \varepsilon).$$

By theorem 5.2, we have the upper bound

$$P_x(|Z_t(\varphi_i \varphi_j)| > C_{ij} \varepsilon) \leq K(p) \varrho(C)^p t^{-p/2}.$$

Note that due to assumption 2.2, 5. and since  $\mu(|\varphi_i \varphi_j|) \leq p_1$ , we have that

$$\varrho(C) \leq \Phi_1 p_1$$

where the upper bound does not depend on  $t$ . Indeed, using that  $2u_i u_j \leq u_i^2 + u_j^2$ , we have that

$$\begin{aligned} \varrho(C) &= \sup_{u \in \mathbb{R}^{D_t}, \|u\| \leq 1} \langle Cu, u \rangle = \sup_{u \in \mathbb{R}^{D_t}, \|u\| \leq 1} \sum_{i,j} C_{ij} u_i u_j \\ &\leq \sup_{u \in \mathbb{R}^{D_t}, \|u\| \leq 1} \sum_{i,j} C_{ij} u_i^2 \\ &= \sup_{u \in \mathbb{R}^{D_t}, \|u\| \leq 1} \sum_i u_i^2 \sum_{j: \varphi_i \varphi_j \neq 0} \mu(|\varphi_i \varphi_j|) \\ &\leq \sup_{u \in \mathbb{R}^{D_t}, \|u\| \leq 1} \sum_i u_i^2 \Phi_1 p_1 \leq \Phi_1 p_1. \end{aligned}$$

Using once more that

$$\sum_i \sum_j 1_{\{\varphi_i \varphi_j \neq 0\}} \leq D_t \cdot \Phi_1,$$

due to assumption 2.2, 5., we conclude that

$$P_x(\Omega_t^c) \leq C D_t t^{-p/2} \leq C t^{-(p/2-1)},$$

where  $C$  denotes a universal constant not depending on the dimension. •

## 7 Control of $A_t$

It still remains to control the probability of  $A_t^c$ . Note that by definition of  $A_t$  and of  $\Omega_t$ , if we take  $t$  sufficiently large such that

$$p_0 \geq 2t^{-1/2} \text{ or equivalently } t \geq \frac{4}{p_0^2},$$

obviously  $A_t^c \subset \Omega_t^c$ . But we have to be able to put hands on the lower bound  $p_0$  of the invariant density on  $[0, 1]$ .

Our approach relies on fine estimations of the scale density under our model assumptions 2.1. Suppose w.l.o.g. that the constant  $M_0$  of assumption 2.1.3 is greater or equal to 1. Then it is easy to see that for any  $x$  such that  $|x| \leq M_0$ ,

$$s^{-1}(x) \leq e^{\frac{2M_0 b_0}{\sigma_0^2}}.$$

and for  $|x| \geq M_0$ ,

$$s^{-1}(x) \leq e^{\frac{2M_0 b_0}{\sigma_0^2}} \left( \frac{M_0}{|x|} \right)^{\frac{2\gamma}{\sigma_1^2}}.$$

Then we get immediately that

$$M = \int_{-\infty}^{+\infty} (s(x)\sigma^2(x))^{-1} dx \leq \frac{2M_0}{\sigma_0^2} e^{\frac{2M_0 b_0}{\sigma_0^2}} \left[ \frac{2\gamma}{2\gamma - \sigma_1^2} \right] =: M_+.$$

This yields the following lower bound for all  $x \in [0, 1]$ ,

$$p(x) \geq \frac{1}{M_+} \frac{1}{\sigma_1^2} e^{-2b_0/\sigma_0^2}.$$

Hence a possible choice is

$$p_0 = \frac{1}{M_+} \frac{1}{\sigma_1^2} e^{-2b_0/\sigma_0^2}. \tag{7.38}$$

This yields the following corollary.

**Corollary 7.1** *Suppose that*

$$2\gamma > 5\sigma_1^2.$$

*Under our model assumptions 2.1, let  $p_0$  be as in (7.38). Then for all  $t \geq t_0 := \frac{4}{p_0}$ , we have that*

$$P_x(A_t^c) \leq Ct^{-1}.$$

**Proof** This follows immediately from the above considerations, noting that  $P_x(A_t^c) \leq P_x(\Omega_t^c) \leq Ct^{-1}$ , applying theorem 6.1 with  $p = 3$ . •

## 8 Proof of proposition 4.3

We finally give the proof of proposition 4.3. In what follows, we shall also need a Bernstein inequality for martingales. Write

$$\nu_t(f) := \frac{1}{t} \int_0^t f(X_s) \sigma(X_s) dW_s.$$

Using the classical Bernstein inequality for continuous martingales (see for instance Dzhaparidze and van Zanten (2001)), we recall that

$$P_x(\nu_t(f) \geq a, \|f\|_t^2 \leq v^2) \leq \exp\left(-\frac{ta^2}{2\sigma_1^2 v^2}\right). \quad (8.39)$$

Recall that  $\|f\|_t^2 = \frac{1}{t} \int_0^t f^2(X_s) ds$  and  $\|f\|_\mu^2 = \mu(f^2)$ .

The proof of proposition 4.3 follows a chaining argument as developed in Baraud et al. (2001), pages 45–47. By Lorentz et al. (1996), for any linear subspace  $S$  of  $L^2([0, 1], \mu)$  of dimension  $d$  one can find a set  $T_\delta \subset B$ , where  $B$  is the unit ball of  $S \subset L^2([0, 1], \mu)$ , such that

$$\text{card}(T_\delta) \leq \left(\frac{3}{\delta}\right)^d, \text{ and } \forall f \in B \exists f_\delta \in T_\delta : \|f - f_\delta\|_\mu \leq \delta.$$

Apply this to the linear space  $\mathcal{S}_m + \mathcal{S}_{m'}$  of dimension  $d(m') \leq D_m + D_{m'}$ . Consider  $\delta_k$ -sets  $T_k = T_{\delta_k}$  where  $\delta_k = \delta_0 2^{-k}$ , where  $\delta_0 < 1$  is to be chosen later. Set  $H_k := \log \text{card}(T_k)$ . Write  $B_{m'} := \{f \in \mathcal{S}_m + \mathcal{S}_{m'} : \|f\|_\mu \leq 1\}$ . Then for any  $f \in B_{m'}$ , one can find a sequence  $(f_k)_k$  with  $f_k \in T_k$  such that  $\|f - f_k\|_\mu \leq \delta_k$ . Hence we get

$$f = f_0 + \sum_{k \geq 1} (f_k - f_{k-1}).$$

Then as in Baraud et al. (2001),

$$\|f_0\|_\mu \leq 1, \|f_k - f_{k-1}\|_\mu^2 \leq 5\delta_{k-1}^2/2.$$

In the following, we shall work in restriction to the set  $\Omega_t$ . Write  $P_t$  for the measure  $P_x(\cdot \cap \Omega_t)$ . Put as in Baraud et al. (2001),

$$\Delta := \sqrt{3}\sigma_1 \left( \sqrt{x_0} + \sum_{k \geq 1} \delta_{k-1} \sqrt{5x_k/2} \right),$$

then we have that

$$\begin{aligned} P_t \left( \sup_{f \in B_{m'}} \nu_t(f) \geq \Delta \right) &= P_t \left( \exists (f_k)_k, f_k \in T_k : \nu_t(f_0) + \sum_{k \geq 1} \nu_t(f_k - f_{k-1}) \geq \Delta \right) \\ &\leq P_1 + P_2, \end{aligned}$$

where

$$P_1 = \sum_{f_0 \in T_0} P_t \left( \nu_t(f_0) \geq \sqrt{3x_0}\sigma_1 \right),$$

and

$$P_2 = \sum_{k \geq 1} \sum_{f_k \in T_k, f_{k-1} \in T_{k-1}} P_t \left( \nu_t(f_k - f_{k-1}) \geq \sigma_1 \delta_{k-1} \sqrt{15x_k/2} \right).$$

Recall (8.39) : Since on  $\Omega_t$ ,  $\|f\|_t^2 \leq \frac{3}{2}\|f\|_\mu^2$ , we have for all  $x > 0$ ,

$$P_t \left( \nu_t(f) \geq \sqrt{3}\sigma_1 \sqrt{x} \|f\|_\mu \right) \leq \exp(-tx).$$

We apply this inequality, remarking that  $\|f_0\|_\mu \leq 1$ , hence

$$P_1 \leq \text{card}(T_0) \exp(-tx_0) = \exp(H_0 - tx_0)$$

and, since  $\|f_k - f_{k-1}\|_\mu^2 \leq 5\delta_{k-1}^2/2$ ,

$$P_2 \leq \sum_{k \geq 1} \exp(H_{k-1} + H_k - tx_k).$$

Now, choose  $x_0$  such that

$$tx_0 = H_0 + D_{m'} + \tau$$

and  $x_k$  such that

$$tx_k = H_{k-1} + H_k + D_{m'} + kd(m') + \tau.$$

Then, if  $d(m') \geq 1$ , we obtain as in Baraud et al. (2001),

$$P_t \left( \sup_{f \in B_{m'}} \nu_t(f) \geq \Delta \right) \leq 1, 6e^{-\tau} e^{-D_{m'}}. \quad (8.40)$$

Else,  $d(m') = 0$ , hence  $\mathcal{S}_m + \mathcal{S}_{m'} = \{0\}$ , and (8.40) holds trivially. Exactly as in Baraud et al. (2001), it can be shown that

$$t\Delta^2 \leq \chi\sigma_1^2(D_{m'} + D_m + \tau),$$

and then we conclude as there

$$E_x \left[ \left( G_m^2(m') - \chi\sigma_1^2 \frac{D_{m'} + D_m}{t} \right)_+ 1_{\Omega_t} \right] \leq 1, 6\chi\sigma_1^2 \frac{1}{t} e^{-D_{m'}}.$$

## 9 Simulations

We have not made intensive numeric simulations for our estimator, since they would be redundant with respect to existing estimation schemes. Indeed, though it is possible to simulate “exactly” a sampled diffusion’s trajectory (see Beskos et al [3]), the discretization of stochastic integrals, needed to construct our estimator, produces the estimator studied in Comte et al [4].

Nevertheless, our results apply to a larger than in [4] class of diffusions. To give an example, we take

$$dX_t = -\frac{\gamma X_t}{1 + X_t^2} dt + dW_t, \quad X_0 = x, \quad \gamma > \frac{1}{2}.$$

This diffusion is positive recurrent with stationary distribution

$$\mu(dx) \sim \frac{dx}{(1 + x^2)^\gamma}.$$

Note that we do not assume the existence of exponential moments, neither the  $\beta$ -mixing of  $X_t$  as in [4]. Moreover, the initial distribution has not to be stationary.

To produce a realization at time  $t$  of  $X_t$  started at  $x$ , the algorithm EA1 of [3] requires (in our example) that

$$\alpha(u) = -\frac{\gamma X_t}{1 + X_t^2}$$

be  $C^1$ , and that

$$\alpha^2(u) + \alpha'(u) = \frac{(\gamma^2 + \gamma)x^2 - \gamma}{(1 + x^2)^2}$$

be bounded, which is the case here. Moreover, setting  $A(y) = \int^y \alpha(u) du$ , the function

$$h(y) = \exp\left(A(y) - \frac{(y-x)^2}{2t}\right) = \frac{C}{(1+x^2)^{\gamma/2}} \exp\left(-\frac{(y-x)^2}{2t}\right)$$

must be integrable, and an exact realization of a r.v. with a density proportional to  $h$  must be possible. All these conditions being fulfilled here, we can generate a trajectory of  $X_t$  sampled at discrete moments  $k\Delta$ ,  $k = 0, 1, \dots, n$ . The integrals involved in the construction of the estimator are then approximated by Stieltjes sums.

To fulfill the assumption 2.1.4, we take  $\gamma = 16$ . A sampled trajectory of  $X_t$  with  $X_0 = 1$  was generated using 100000 equidistant time moments, with a step  $\Delta = 0.01$ , hence  $t = 1000$ . The estimation has been made on an interval covering 98% of the generated data points.

We have chosen the family of Legendre dyadic polynomials, with  $r_{\max} = 4$ , to construct an estimator. The dimension  $D_m = 2^p(r+1)$  for  $m = (p, r)$  was bounded by 20 to ensure that the sampled process has visited the support of every polynomial. The penalty  $pen(m)$  was taken equal to  $4D_m/t$ .

**Remark 9.1** *The quality of estimation is quite satisfactory on the intervals  $K$  with the lower bound  $p_0$  of the invariant density  $p(\cdot)$  large enough. Interestingly enough, we have never succeeded to get a good approximation on intervals with nearly vanishing  $p_0$ . This is quite in accordance with the statement of the proposition 3.1.*

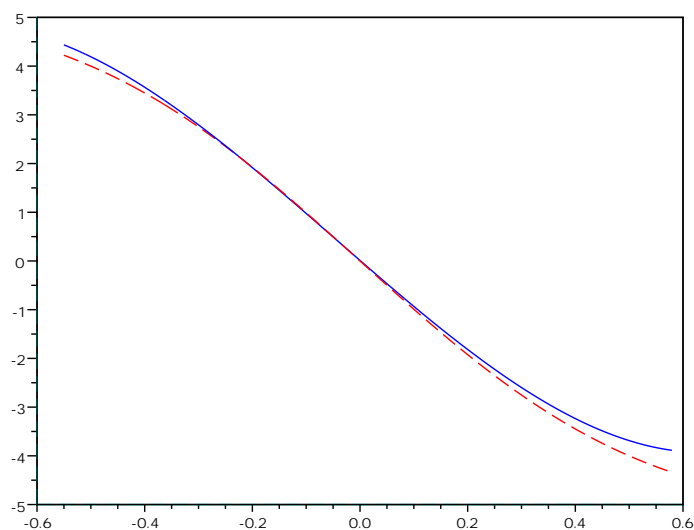


Figure 1: Solid line: estimated drift; dashed line: true drift. The adaptive estimation algorithm on the interval  $[-0.56, 0.58]$  selects  $\hat{m} = (0, 3)$ .

## References

- [1] Baraud, Y.; Comte, F.; Viennet, G. Adaptive estimation in autoregression or  $\beta$ -mixing regression via model selection. *Ann. Stat.* 29, No.3, 839-875 (2001).
- [2] Baraud, Y.; Comte, F.; Viennet, G. Model selection for (auto-)regression with dependent data. *ESAIM, Probab. Stat.* 5, 33-49 (2001).
- [3] Beskos, A.; Papaspiliopoulos, O.; Roberts, G.O. Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli* 12, 1077–1098 (2006).
- [4] Comte, F.; Genon-Catalot, V.; Rozenholc, Y. Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* 13, No. 2, 514-543 (2007).
- [5] Dalalyan, A. Sharp adaptive estimation of the drift function for ergodic diffusions. *Ann. Stat.* 33, No. 6, 2507-2528 (2005).
- [6] Dzhaparidze, K.; van Zanten, J.H. On Bernstein-type inequalities for martingales. *Stochastic Processes Appl.* 93, No.1, 109-117 (2001).
- [7] Hoffmann, M. Adaptive estimation in diffusion processes. *Stochastic Processes Appl.* 79, No. 1, 135-163 (1999).
- [8] Kutoyants, Yury A. *Statistical inference for ergodic diffusion processes*. Springer Series in Statistics. London: Springer (2004).
- [9] Lorentz, G.G.; von Golitschek, M.; Makovoz, Y. *Constructive approximation: advanced problems*. Grundlehren der Mathematischen Wissenschaften. 304. Berlin: Springer (1996).

- [10] Loukianova, D.; Loukianov, O.; Löcherbach, E. Deviation bounds in ergodic theorem for positive recurrent one-dimensional diffusions and integrability of hitting times. Manuscript, 2009.
- [11] Spokoiny, V. G. Adaptive drift estimation for nonparametric diffusion model. *Ann. Stat.* 28, No. 3, 815-836 (2000).