UNIVERSITÀ DEGLI STUDI DI BERGAMO
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
E METODI MATEMATICI°

QUADERNI DEL DIPARTIMENTO

Department of Information Technology and Mathematical Methods

Working Paper

Series "*Mathematics and Statistics*"

n. 7/MS – 2008

Goodness of fit test for ergodic diffusions by discrete time observations: an innovation martingale approach

by

Hiroki Masuda, Ilia Negri, Yoichi Nishiyama

° Viale Marconi. 5, I – 24044 Dalmine (BG), ITALY, Tel. +39-035-2052339; Fax. +39-035-562779

[§] L'accesso alle *Series* è approvato dal Comitato di Redazione. I *Working Papers* della Collana dei Quaderni del Dipartimento di Ingegneria dell'Informazione e Metodi Matematici costituiscono un servizio atto a fornire la tempestiva divulgazione dei risultati dell'attività di ricerca, siano essi in forma provvisoria o definitiva.

Goodness of fit test for ergodic diffusions by discrete time observations: an innovation martingale approach

Hiroki Masuda

Graduate School of Mathematics, Kyushu Univeristy

6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

hiroki@math.kyushu-u.ac.jp

Ilia Negri

University of Bergamo

Viale Marconi, 5, 24044, Dalmine (BG), Italy

ilia.negri@unibg.it

and

Yoichi Nishiyama *

The Institute of Statistical Mathematics

4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan

nisiyama@ism.ac.jp

July 26, 2008

Running title: Goodness of fit test for ergodic diffusions

Abstract

We consider a nonparametric goodness of fit test problem for the drift coefficient of one-dimensional ergodic diffusions. Our test is based on discrete time observation of the processes, and the diffusion coefficient is a nuisance function which is estimated in our testing procedure. We prove that the limit distribution of our test is the supremum of the standard Brownian motion, and thus our test is asymptotically distribution free. We also show that our test is consistent under any fixed alternatives.

Keywords. Ergodic diffusion process, discrete time observation, invariance principle, asymptotically distribution free test.

*Corresponding author.

1 Introduction

Goodness of fit tests play an important role in theoretical and applied statistics, and the study for them has a long history. Such tests are really useful especially if they are *distribution free*, in the sense that their distributions do not depend on the underlying model. The origin goes back to the Kolmogorov-Smirnov and Crámer-von Mises tests in the i.i.d. case, established early in the 20th century, which are *asymptotically distribution free*.

This work deals with a goodness of fit test for diffusion processes. Despite the fact that in the last thirty years diffusion models have been proved to be immensely useful, not only in finance and more generally in economics science, but also in other fields such as biology, medicine, physics and engineering, the problem of goodness of fit tests for diffusion processes has still been a new issue in recent years.

The estimation theory, in both the parametric and the non parametric frameworks, has been studied by many authors in the last twenty years. Among the others we cite Kutoyants [13] and the references therein for the model based on continuous time observations. Regarding the model based on discrete time observations for different sampling schemes we recall the works of Yoshida [23], Kessler [10], Hoffmann [9], and Gobet *et al.* [8]. In the last paper an interesting historical overview on this topic is presented. The methods introduced for estimation of diffusion process have been successfully implemented and applied to financial data to study decision to optimally consume, save and invest, portfolio choice under many different constrains, contingent claim pricing. See e.g. Aıt-Sahalia [1] and reference therein.

Although their great importance in application, the theory of goodness of fit tests for diffusion has not received much attention from researcher as the theory of estimation has. Kutoyants [13] discusses some possibilities of the construction of such tests in his Section 5.4, where he considers the Kolmogorov-Smirnov statistics based on the continuous observation of a diffusion process. The goodness of fit test based on the Kolmogorov-Smirnov statistics is asymptotically consistent and the asymptotic distribution under the null hypothesis follows from the weak convergence of the empirical process to a suitable Gaussian process but these tests are not asymptotically distribution free. Note that the Kolmogorov-Smirnov statistics for ergodic diffusion process was studied in Fournie,[5], see also Fournie and Kutoyants [6] for more details, while the weak convergence of the empirical process was proved in Negri [16] (see van der Vaart and van Zanten, [22] for further developments). Dachian and Kutoyants [2] and Negri and Nishiyama [17] proposed some asymptotically distribution free tests but their results are based on *continuous time observation* of the diffusion processes. One of the interesting points of this paper is that the proposed test is based on *discrete time observation*, which is more realistic in applications.

As well as Negri and Nishiyama [17], [18], we take an approach based on a certain marked empirical process to construct an asymptotically distribution free test, where "empirical process" actually means an innovation martingale. In Lee and Wee [14] a similar approach based on the residual empirical process is proposed to study a goodness of fit test for diffusion process with the drift in a parametric form and a constant diffusion coefficient. Our approach based on the innovation martingale is motivated by the work of Koul and Stute [12] who considered a non-linear parametric time series model (see also Section 7.3 of Nishiyama [20] which is a reprint of his thesis in 1998). They studied the large sample behavior of the proposed test statistics under the null hypotheses and present a martingale transformation of the underlying process that makes tests based on it asymptotically distribution free. Some considerations on consistency have also been done. The approach is well expounded in Koul [11]. See Delgado and Stute [3] and references therein for more recent information. Our work is an attempt to develope the method in the continuous time model based on discrete time observation.

Now we turn to the description of the problem treated in this paper. Consider a one-dimensional stochastic differential equation (SDE)

$$X_t = X_0 + \int_0^t S(X_t)dt + \int_0^t \sigma(X_t)dW_t, \tag{1}$$

where the initial value X_0 is finite almost surely, S and σ are functions which satisfy some properties described in Section 2, and $t \rightsquigarrow W_t$ is a standard Wiener process defined on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$. We consider a case where a unique strong solution X to this SDE exists, and we shall assume that X is ergodic. We are interested in goodness of fit test for the drift coefficient S, while the diffusion coefficient σ^2 is an unknown nuisance function which we estimate in our testing procedure. That is, we consider the problem of testing hypothesis $H_0 : S = S_0$ versus $H_1 : S \neq S_0$ for a given S_0. The meaning of the alternatives "$S \neq S_0$" will be precisely stated in Section 4.

We denote $\mathrm{Log}\, m = \log(1 + m)$. We consider the following situation.

Sampling scheme. The process $X = \{X_t; t \in [0, \infty)\}$ is observed at times $0 = t_0^n < t_1^n < \cdots < t_n^n$ such that, as $n \to \infty$, $t_n^n \to \infty$ and $h_n = O(n^{-2/3}(\mathrm{Log}\, n)^{1/3})$ (which implies $nh_n^2 \to 0$) where $h_n = \max_{1\leq i\leq n} |t_i^n - t_{i-1}^n|$. $\diamond$

This condition implies $h_n \to 0$, so we may assume that $h_n \leq 1$ without loss of generality. We will propose an asymptotically distribution free test based on this sampling scheme, namely, *high frequency data*. We should mention that there is a huge literature on discrete time approximations of statistical estimators for diffusion processes; see e.g. the Introduction of Gobet *et al.* [8] for a review including not only high frequency cases but also low frequency cases. However, it seems difficult to obtain asymptotically distribution free results based on low frequency data.

The organization of the article is as follows. In Section 2, we state some conditions for (S, σ) which are assumed throughout this work. Section 3 gives the main result under the null hypothesis. In Section 4, we prove that our test is consistent under any fixed alternatives. A simulation study is given in Section 5. The proofs for some lemmas will be given in Section 6 and in the Appendix I. The Appendix II gives some simple sufficient conditions for some of our regularity conditions.

2 Preliminaries

Let us list some conditions for the pair of functions (S,σ).
A1. There exists a constant $K>0$ such that

$$|S(x)-S(y)| \le K|x-y|, \quad |\sigma(x)-\sigma(y)| \le K|x-y|.$$

$\diamond$

Under this condition, the SDE (1) has a unique strong solution X. Notice also that there exists a constant $K'>0$ such that

$$|S(x)| \le K^{'}(1+|x|), \quad |\sigma(x)| \le K^{'}(1+|x|).$$

To see this, just put $y=0$. The constant K' depends on the values $S(0)$ and $\sigma(0)$, however the constant K itself depends on the choice of the functions (S,σ). So it is convenient to introduce the notation

$$K_{S,\sigma} = \max\{K, K'\}.$$

This notation will be used throughout this article.

A2. The diffusion process X is regular. The speed measure $m_{S,\sigma}$ is finite and has the second moment. $\diamond$

In this case, the process X is ergodic. We denote by $f_{S,\sigma}$ the invariant density, and introduce the metric $\rho_{S,\sigma}$ on $[-\infty,\infty]$ given by

$$\rho_{S,\sigma}(x,y) = \sqrt{\int_{x\wedge y}^{x\vee y} (\sigma(z)^2 f_{S,\sigma}(z) + \phi(z))dz}. \tag{2}$$

where ϕ is the density of the standard Gaussian distribution. Without ϕ, the above $\rho_{S,\sigma}$ may not define a *metric* but just define a *semimetric*. The weak convergence theory which we review in the Appendix I requires that ρ is a metric, so we have added the Gaussian density ϕ. It is easy to see that the space $[-\infty,\infty]$ is compact and the metric entropy condition is satisfied:

$$\int_0^1 \sqrt{\log N([-\infty,\infty],\rho_{S,\sigma};\varepsilon)}d\varepsilon < \infty.$$

Here, when a metric space (T,ρ) is given, $N(T,\rho;\varepsilon)$ denotes the smallest number of closed balls with ρ-radius ε which cover T.

A3. The diffusion coefficient is bounded: $\sigma_*^2 := \sup_{x\in\mathbb{R}} \sigma(x)^2 < \infty$. The invariant density $f_{S,\sigma}$ is bounded. Furthermore,

$$\Sigma_{S,\sigma} := \sqrt{\int_{-\infty}^{\infty} \sigma(z)^2 f_{S,\sigma}(z)dz} > 0.$$

$\diamond$

A4. $\sup_{t\in[0,\infty)} E|X_t|^2 < \infty$. ◇

Now, we introduce an array of constants

$$-\infty = x_0^n < x_1^n < x_2^n < \cdots < x_{m(n)}^n < x_{m(n)+1}^n = \infty$$

such that, as $n \to \infty$,

$$\max_{2\le k\le m(n)} |x_k^n - x_{k-1}^n| \to 0, \quad x_1^n \downarrow -\infty, \quad x_{m(n)}^n \uparrow \infty.$$

For example, one may consider $x_k^n = -n + (k/n)$ with $k = 1, 2, ..., 2n^2$. Next we introduce a sequence of functions $z \rightsquigarrow \psi_k^n(z)$ on $(-\infty, \infty)$ which approximates the indicator function $1_{(-\infty, x_k^n]}$.

Definition 1 *Let a sequence of positive constants b_n be given. For every $k = 1, 2, ..., m(n)$, ψ_k^n is the continuous, piecewise linear function on $(-\infty, \infty)$ defined by*

$$\psi_k^n(z) = \begin{cases} 1, & z \in (-\infty, x_k^n], \\ \textit{line}, & z \in [x_k^n, x_k^n + b_n], \\ 0, & z \in [x_k^n + b_n, \infty). \end{cases}$$

Also we define $\psi_0^n \equiv 0$ and $\psi_{m(n)+1}^n \equiv 1$.

This function satisfies the following properties:

$$|\psi_k^n(z) - \psi_k^n(z')| \le b_n^{-1}|z - z'|;$$

$$|\psi_k^n(z) - 1_{(-\infty, x_k^n]}(z)| \le 1_{[x_k^n, x_k^n + b_n]}(z).$$

Now we make the following condition.

A5. In addition to $h_n = O(n^{-2/3}(\operatorname{Log} n)^{1/3})$, which implies $nh_n^2 \to 0$, we assume the following:
(i) $b_n^{-2} h_n \cdot \operatorname{Log} n \cdot \operatorname{Log} m(n) \to 0$;
(ii) $b_n \operatorname{Log} m(n) \to 0$. ◇

Typically, $\operatorname{Log} m(n) = O(\operatorname{Log} n^\alpha)$ for some $\alpha > 0$. In this case, the above (i) and (ii) are satisfied if we take $b_n = n^{-1/4}\operatorname{Log} n$.

Let us close this section with making some conventions. We denote by $C_\rho(T)$ the class of continuous functions defined on a metric space (T, ρ), and by $\ell^\infty(T)$ the class of bounded functions on T. We equip the uniform metric with the spaces. We denote by "$\to^p$" and "$\to^d$" the convergence in probability and in distribution as $n \to \infty$, respectively. The notation "$\to$" always means that we take the limit as $n \to \infty$. See e.g. van der Vaart and Wellner [21] for the weak convergence theory in the $\ell^\infty(T)$ space.

3 Asymptotically distribution free test

Throughout all this section, we shall suppose that **A1** – **A5** are satisfied for some (S_0, σ). We denote $\rho = \rho_{S_0,\sigma}$ which is given by (2).

Our test statistics is based on the random field $U^n = \{U^n(x); x \in [-\infty, \infty]\}$ defined by $U^n(-\infty) = 0$ and

$$U^n(x) = \frac{1}{\sqrt{t_n^n}} \sum_{i=1}^{n} \psi_k^n(X_{t_{i-1}^n})[X_{t_i^n} - X_{t_{i-1}^n} - S_0(X_{t_{i-1}^n})|t_i^n - t_{i-1}^n|]$$

for $x \in (x_{k-1}^n, x_k^n]$, $1 \leq k \leq m(n) + 1$. We call it the *smoothed score marked empirical process* based on discrete observation. This U^n is an approximation of the random field $V^n = \{V^n(x); x \in [-\infty, \infty]\}$ defined by

$$V^n(x) = \frac{1}{\sqrt{t_n^n}} \int_0^{t_n^n} 1_{(-\infty,x]}(X_t)[dX_t - S_0(X_t)dt],$$

which is the *score marked empirical process* based on continuous observation studied by Negri and Nishiyama [17].

We present some lemmas which will be proved in Section 6.

Lemma 2 *The random field U^n takes values in $\ell^\infty([-\infty, \infty])$, and the random field V^n takes values in $C_\rho([-\infty, \infty])$ almost surely.*

Lemma 3 $\sup_{x \in [-\infty,\infty]} |U^n(x) - V^n(x)| \to^p 0$.

So, once we establish a good asymptotic result for V^n, the same thing could hold also for U^n. Indeed, we have the following result for V^n which was essentially obtained by Negri and Nishiyama [17]. It is a fruit of the combination of the weak convergence theory for ℓ^∞-valued continuous martingales developed by Nishiyama [20] and a theorem for local time of ergodic diffusion processes given by van Zanten [24] (see also van der Vaart and van Zanten [22]).

Lemma 4 *$V^n \to^d G$ in $C_\rho([-\infty, \infty])$, where $G = \{G(x); x \in [-\infty, \infty]\}$ is a zero-mean Gaussian random field with co-variance given by*

$$EG(x)G(y) = \int_{-\infty}^{x \wedge y} \sigma(z)^2 f_{S_0,\sigma}(z)dz.$$

Almost all paths of G are uniformly ρ-continuous.

Combining Lemmas 3 and 4 we obtain the following.

Theorem 5 *$U^n \to^d G$ in $\ell^\infty([-\infty, \infty])$, where G is that in Lemma 4.*

By the continuous mapping theorem, we have the following.

Corollary 6 *$\sup_{x\in[-\infty,\infty]}|U^n(x)| \to^d \sup_{t\in[0,\Sigma^2]}|B_t| =^d \Sigma\sup_{t\in[0,1]}|B_t|$, where $t \rightsquigarrow B_t$ is a standard Brownian motion, $\Sigma = \Sigma_{S_0,\sigma}$, and where the notation "$=^d$" means that the distributions are the same.*

In order to obtain an asymptotically distribution free test, we need a consistent estimator for $\Sigma_{S_0,\sigma}$. The following lemma, which will be proved also in Section 6, gives us an answer.

Lemma 7 *The estimator*

$$\widehat{\Sigma}^n = \sqrt{\frac{1}{t_n^n}\sum_{i=1}^n |X_{t_i^n} - X_{t_{i-1}^n}|^2}$$

is consistent for $\Sigma_{S,\sigma}$.

We finally obtain our main result.

Theorem 8 *Under $H_0 : S = S_0$, it holds that*

$$\mathcal{T}^n = \frac{\sup_{x\in[-\infty,\infty]}|U^n(x)|}{\widehat{\Sigma}^n} \to^d \sup_{t\in[0,1]}|B_t|,$$

where $t \rightsquigarrow B_t$ is a standard Brownian motion.

One may think that it is more natural to consider the random field $\widetilde{U}^n = \{\widetilde{U}^n(x); x \in [-\infty,\infty]\}$ given by

$$\widetilde{U}^n(x) = \frac{1}{\sqrt{t_n^n}}\sum_{i=1}^n 1_{(-\infty,x]}(X_{t_{i-1}^n})[X_{t_i^n} - X_{t_{i-1}^n} - S_0(X_{t_{i-1}^n})|t_i^n - t_{i-1}^n|]$$

(that is, the case $b_n = 0$) instead of U^n. At least in our proof, the uniform approximation (Lemma 3) is due to the continuity of the function $z \rightsquigarrow \psi_k^n(z)$, so it does not seem easy to translate the result for V^n into that for $\widetilde{U}^n$. However, it is conjectured that the same result would hold also for $\widetilde{U}^n$; see a simulation study in Section 5.

4 Consistency of the test

In this section, we continue to assume all conditions in Section 3, including the properties of the given function S_0. We denote by $\mathcal{S}$ the class of functions S which satisfies **A1** – **A4** and

$$\int_{-\infty}^{x_S}(S(z) - S_0(z))f_{S,\sigma}(z)dz \neq 0 \quad \text{for some } x_S \in (-\infty,\infty]. \tag{3}$$

The precise description of our problem is testing the null hypothesis $H_0 : S = S_0$ versus the alternatives $H_1 : S \in \mathcal{S}$.

We will prove that our test is consistent. Fix $S \in \mathcal{S}$. We can write $U^n = U^n_S + U^n_\Delta$ where $U^n_S(-\infty) = U^n_\Delta(-\infty) = 0$ and, for $x \in (x^n_{k-1}, x_k]$, $1 \le k \le m(n)+1$,

$$U^n_S(x) = \frac{1}{\sqrt{t^n_n}} \sum_{i=1}^n \psi^n_k(X_{t^n_{i-1}})[X_{t^n_i} - X_{t^n_{i-1}} - S(X_{t^n_{i-1}})|t^n_i - t^n_{i-1}|]$$

and

$$U^n_\Delta(x) = \frac{1}{\sqrt{t^n_n}} \sum_{i=1}^n \psi^n_k(X_{t^n_{i-1}})(S(X_{t^n_{i-1}}) - S_0(X_{t^n_{i-1}}))|t^n_i - t^n_{i-1}|.$$

Now we have

$$\sup_{x\in[-\infty,\infty]} |U^n(x)| \ge \sup_{x\in[-\infty,\infty]} |U^n_\Delta(x)| - \sup_{x\in[-\infty,\infty]} |U^n_S(x)|.$$

Since S satisfies **A1** – **A4**, by the same argument as in Section 3, the random field U^n_S converges to the corresponding Gaussian random field with S_0 replaced by S. So the second term of the right hand side is $O_P(1)$, and thus it suffices to show

$$\sup_{x\in[-\infty,\infty]} |U^n_\Delta(x)| \ne O_P(1).$$

This is actually accomplished by the following lemma.

Lemma 9 *Choose $x_S \in (-\infty, \infty]$ as in (3). Then, $|U^n_\Delta(x_S)| \ne O_P(1)$.*

We therefore obtain the consistency of the test.

Theorem 10 *Under $H_1 : S \in \mathcal{S}$, it holds that*

$$\mathcal{T}^n = \frac{\sup_{x\in[-\infty,\infty]} |U^n(x)|}{\widehat{\Sigma}^n} \ne O_P(1).$$

5 Simulation study

In this section we observe finite-sample performance of our test statistics through numerical experiments. For true (data-generating) process we adopt the Ornstein-Uhlenbeck diffusion starting from the origin:

$$X_t = \int_0^t (-2X_s)ds + W_t. \tag{4}$$

For simplicity we here focus on the equidistant sampling case, that is, $h_n = t^n_i - t^n_{i-1}$ for every $i \le n$.

We are going to observe the following (a) and (b), in both of which we will take $x^n_k = -n + \frac{k}{n}$ for $k = 1, 2, \dots, 2n^2$, and $b_n = \frac{1}{100} n^{-1/4} \mathrm{Log} n$:

(a) asymptotic behavior of $\mathcal{T}^n_0 := \mathcal{T}^n$ with $S_0(x) = -2x$;

(b) asymptotic behavior of $\mathcal{T}^n_1 := \mathcal{T}^n$ with $S_0(x) = -4x$.

Concerning (a), from Theorem 8 it follows that the distribution of $\mathcal{T}_0^n$ asymptotically obeys

$$F(x) := P\left(\sup_{t\in[0,1]} |B_t| \leq x \right) = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left(-\frac{\pi^2(2k+1)^2}{8x^2} \right) \tag{5}$$

for $x \in \mathbb{R}$, where B is a standard Brownian motion. See e.g. 343 page of Feller [4] for this formula. Turning to (b), in view of Theorem 10 the variable $\mathcal{T}_1^n$ diverges in probability.

Throughout we take the significance level to be 0.05. Then we see that $F(x) = 0.95$ for $x \fallingdotseq 2.24$, hence the critical region is $\{x > 2.24\}$, and:

- $P(\mathcal{T}_0^n > 2.24)$ should tend to 0.05 in (a);
- $P(\mathcal{T}_1^n > 2.24)$ should tend to 1.0 in (b).

For several different terminal time t_n^n and sampling frequency h_n, we simulate 1000 independent copies of a discrete sample trajectory of (4) to obtain, say $\{(\mathcal{T}_0^{n,l}, \mathcal{T}_1^{n,l})\}_{l=1}^{1000}$. We then compute:

- the empirical size (e.s.) defined by $\sharp\{l : \mathcal{T}_0^{n,l} > 2.24\}/1000$, the sample proportion of making incorrect rejections of the null;
- the empirical power (e.p.) defined by $\sharp\{l : \mathcal{T}_1^{n,l} > 2.24\}/1000$, the sample proportion of making successful rejections of the null.

Table 1 summarizes the simulation results. We observe that: for a fixed h_n, empirical power gains along with increasing terminal time t_n^n; on the contrary, it is not the case when we make h_n smaller for a fixed t_n^n. This tells us that, in order to obtain high power of our test procedure, the large-time characteristic (i.e., the ergodicity) of the data sequence may be more important than the high frequency.

	$t_n^n = 10$		$t_n^n = 20$		$t_n^n = 50$	
h_n	e.s.	e.p.	e.s.	e.p.	e.s.	e.p.
0.1	0.043	0.438	0.059	0.633	0.067	0.943
	$(n = 100)$		$(n = 200)$		$(n = 500)$	
0.05	0.050	0.412	0.047	0.596	0.060	0.911
	$(n = 200)$		$(n = 400)$		$(n = 1000)$	

Table 1: Empirical sizes (e.s.) and empirical powers (e.p.) based on 1000 independent statistics. Here the significance level is 0.05, and $b_n = \frac{1}{100} n^{-1/4} \mathrm{Log} n$.

Figure 1 reports plots of the empirical distribution functions based on $(\mathcal{T}_0^{n,l})_{l=1}^{1000}$ in (a). This reveals that the distributional approximation of Theorem 8 is quite accurate even for small sample size (small observation window).

Finally, Table 2 shows results of experimental trials when $b_n = 0$. Though in this case our theory cannot confirm the same asymptotic behaviors of $\mathcal{T}_0^n$ and $\mathcal{T}_1^n$ as before, from the table we may expect that this is also the case.

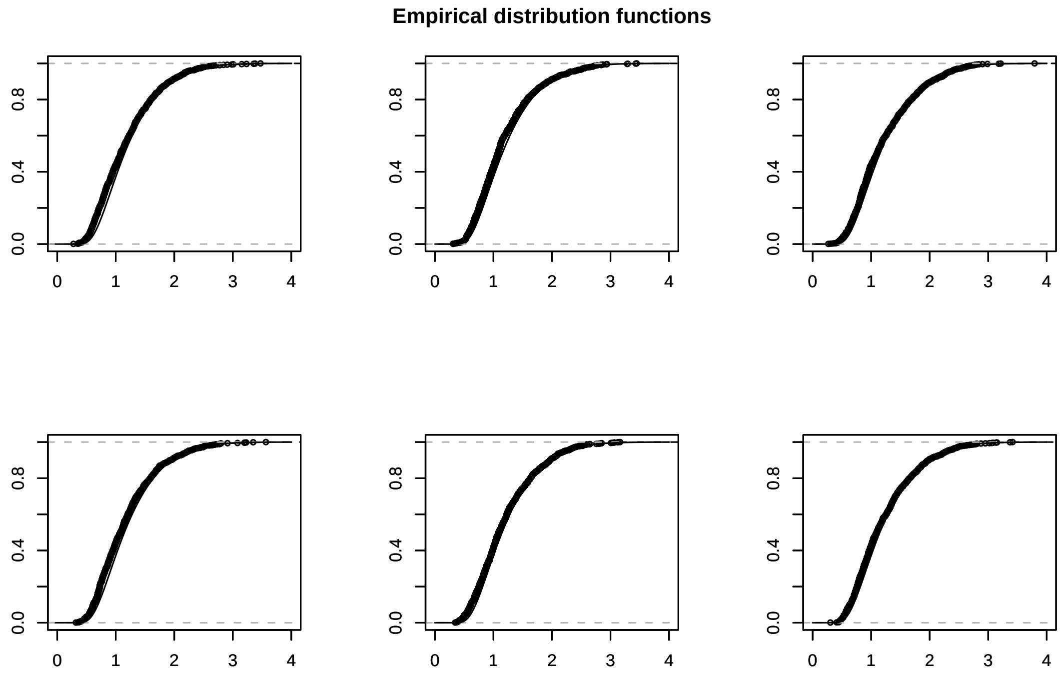

Figure 1: Plots of empirical distribution functions over $[0,4]$ based on $(\mathcal{T}_0^{n,l})_{l=1}^{1000}$ in (a), and the common straight lines in the six displays indicate the target distribution function (5). In each figure the two plots are almost piled up.

6 Proofs of Lemmas

Notation: For $x, y \geq 0$, the inequality $x \lesssim y$ means that there exists a constant $C > 0$, depending only on $K_{S,\sigma}$, σ_*^2 and the value $\sup_s E|X_s|^2$ for fixed (S, σ), such that $x \leq Cy$.

Proof of Lemmas 2 and 4. We apply Lemma 12 to the family $M^n = \{M^{n,x}; x \in [-\infty, \infty]\}$ of continuous martingales given by

$$M_t^{n,x} = \frac{1}{\sqrt{t_n^n}} \int_0^t 1_{(-\infty,x]}(X_s)\sigma(X_s)dW_s.$$

The metric entropy condition $\int_0^1 \sqrt{\log N([-\infty,\infty],\rho;\varepsilon)}d\varepsilon < \infty$ is trivially satisfied. The adapted quadratic modulus satisfies

$$\begin{aligned}\left|||M^n||_{\rho,t_n^n}^*\right|^2 &\leq \sup_{x<y} \frac{(t_n^n)^{-1}\int_{-\infty}^{\infty} 1_{(x,y]}(z)\sigma(z)^2 l_{t_n^n}(z) m_{S_0,\sigma}((-\infty,\infty)) f_{S_0,\sigma}(z)dz}{\int_{-\infty}^{\infty} 1_{(x,y]}(z)(\sigma(z)^2 f_{S_0,\sigma}(z) + \phi(z))dz} \\ &\leq \frac{\sup_{z\in(-\infty,\infty)} l_{t_n^n}(z)}{t_n^n} m_{S_0,\sigma}((-\infty,\infty)),\end{aligned}$$

	$t_n^n = 10$		$t_n^n = 20$		$t_n^n = 50$	
h_n	e.s.	e.p.	e.s.	e.p.	e.s.	e.p.
0.1	0.054	0.479	0.047	0.662	0.063	0.938
	($n = 100$)		($n = 200$)		($n = 500$)	
0.05	0.042	0.455	0.059	0.613	0.038	0.894
	($n = 200$)		($n = 400$)		($n = 1000$)	

Table 2: Empirical sizes (e.s.) and empirical powers (e.p.) based on 1000 independent statistics. Here the significance level is 0.05, and experimentally $b_n = 0$.

where $l_t(z)$ denotes the local time of the diffusion X with respect to the speed measure $m_{S_0,\sigma}$. Theorem 3.1 of van Zanten [24] says that $\sup_{z\in(-\infty,\infty)} l_t(z) = O_P(t)$ as $t \to \infty$. So Lemma 12 (i) implies that V^n takes values in $C_\rho([-\infty,\infty])$ almost surely, and Lemma 12 (ii) yields the assertion of Lemma 4. The assertion that U^n takes values in $\ell^\infty([-\infty,\infty])$ is trivial. □

In order to prove Lemma 3, we prepare a lemma.

Lemma 11 *For every $\varepsilon > 0$ there exists a constant $K > 0$ such that*

$$\limsup_n P\left(\frac{1}{\operatorname{Log} n}\max_{1\le i\le n}\frac{\sup_{t\in(t_{i-1}^n,t_i^n]}|X_t - X_{t_{i-1}^n}|^2}{t_i^n - t_{i-1}^n} \ge K\right) < \varepsilon.$$

Proof. It is sufficient to show that

$$E\left(\max_{1\le i\le n}\frac{\sup_{t\in(t_{i-1}^n,t_i^n]}|X_t - X_{t_{i-1}^n}|^2}{t_i^n - t_{i-1}^n}\right) = O(\operatorname{Log} n).$$

The random variable in the inside of the expectation on the left hand side is bounded by $4\{(I) + (II)\}$ where

$$\begin{aligned}(I) &= \max_{1\le i\le n}\frac{\left|\int_{t_{i-1}^n}^{t_i^n}|S_0(X_s)|ds\right|^2}{t_i^n - t_{i-1}^n},\\ (II) &= \max_{1\le i\le n}\frac{\sup_{t\in(t_{i-1}^n,t_i^n]}\left|\int_{t_{i-1}^n}^{t}\sigma(X_s)dW_s\right|^2}{t_i^n - t_{i-1}^n}.\end{aligned}$$

As for (I), it follows from Hölder's inequality that

$$(I) \le \max_{1\le i\le n}\frac{\int_{t_{i-1}^n}^{t_i^n}ds\int_{t_{i-1}^n}^{t_i^n}|S_0(X_s)|^2ds}{t_i^n - t_{i-1}^n} \lesssim h_n \sup_{s\in[0,t_n^n]}\{1 + |X_s|^2\}.$$

Using Hölder's inequality again, we have

$$
\begin{aligned}
E\sup_{s\in[0,t_n^n]}|X_s|^2 &\lesssim E|X_0|^2 + E\left(\left|\int_0^{t_n^n}|S_0(X_u)|du\right|^2\right) + E\left(\int_0^{t_n^n}\sigma(X_u)^2du\right)\\
&\leq E|X_0|^2 + t_n^n E\left(\int_0^{t_n^n}|S_0(X_u)|^2du\right) + \sigma_*^2 t_n^n\\
&\leq E|X_0|^2 + |t_n^n|^2\sup_u E|S_0(X_u)|^2 + \sigma_*^2 t_n^n\\
&\lesssim |t_n^n|^2.
\end{aligned}
$$

So $E(I) \lesssim h_n|t_n^n|^2 \leq n^2h_n^3 = O(\operatorname{Log} n)$. Here we used the assumption $h_n = O(n^{-2/3}(\operatorname{Log} n)^{1/3})$.

As for (II), it follows from Lemma 13 that

$$
\begin{aligned}
&P\left(\frac{\sup_{t\in(t_{i-1}^n,t_i^n]}|\int_{t_{i-1}^n}^t\sigma(X_s)dW_s|^2}{t_i^n - t_{i-1}^n} > x\right)\\
&= P\left(\sup_{t\in(t_{i-1}^n,t_i^n]}\left|\int_{t_{i-1}^n}^t\sigma(X_s)dW_s\right| > \sqrt{x}\sqrt{t_i^n - t_{i-1}^n}\right)\\
&\leq 2\exp\left(-\frac{x|t_i^n - t_{i-1}^n|}{2\int_{t_{i-1}^n}^{t_i^n}\sigma_*^2ds}\right)\\
&= 2\exp\left(-\frac{x}{2\sigma_*^2}\right).
\end{aligned}
$$

Apply Lemma 14 (i) to conclude that $E(II) \lesssim \operatorname{Log} n$. □

Proof of Lemma 3. Let us introduce the random fields Y_1^n, Y_2^n and Y_3^n given by $Y_1(-\infty) = Y_2(-\infty) = Y_3(-\infty) = 0$ and

$$
\begin{aligned}
Y_1^n(x) &= \frac{1}{\sqrt{t_n^n}}\sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n}\psi_k^n(X_{t_{i-1}^n})[dX_t - S_0(X_t)dt],\\
Y_2^n(x) &= \frac{1}{\sqrt{t_n^n}}\sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n}\psi_k^n(X_t)[dX_t - S_0(X_t)dt],\\
Y_3^n(x) &= \frac{1}{\sqrt{t_n^n}}\sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n}1_{(-\infty,x_k^n]}(X_t)[dX_t - S_0(X_t)dt],
\end{aligned}
$$

for $x \in (x_{k-1}^n, x_k^n]$, $1 \leq k \leq m(n)+1$. We will prove:

$$\sup_{x\in[-\infty,\infty]}|U^n(x) - Y_1^n(x)| \to^p 0; \tag{6}$$

$$\sup_{x\in[-\infty,\infty]}|Y_1^n(x) - Y_2^n(x)| \to^p 0; \tag{7}$$

$$\sup_{x\in[-\infty,\infty]} |Y_2^n(x) - Y_3^n(x)| \to^p 0; \tag{8}$$

$$\sup_{x\in[-\infty,\infty]} |Y_3^n(x) - V^n(x)| \to^p 0. \tag{9}$$

Proof of (6). Using Lemma 15, we have

$$\begin{aligned}
& E\sup_x |U^n(x) - Y_1^n(x)| \\
\leq\ & \frac{1}{\sqrt{t_n^n}} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} E|S_0(X_t) - S_0(X_{t_{i-1}^n})|dt \\
\leq\ & \frac{1}{\sqrt{t_n^n}} \sum_{i=1}^n K_{S_0,\sigma} \int_{t_{i-1}^n}^{t_i^n} E|X_t - X_{t_{i-1}^n}|dt \\
\lesssim\ & \frac{1}{\sqrt{t_n^n}} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} |t - t_{i-1}^n|^{1/2} dt \\
\leq\ & \frac{1}{\sqrt{t_n^n}} \cdot t_n^n \cdot h_n^{1/2} \\
\leq\ & \sqrt{n} h_n \to 0.
\end{aligned}$$

Proof of (7). We introduce the stopping time

$$\tau_n(K) = \inf\left\{ t \in [0, t_n^n] :\ \frac{\sup_{t\in(t_{i-1}^n, t_i^n]} |X_t - X_{t_{i-1}^n}|^2}{\operatorname{Log} n \cdot |t_i^n - t_{i-1}^n|} \geq K \right\}.$$

By Lemma 11, for every $\varepsilon > 0$ there exists a constant $K > 0$ such that $\limsup_n P(\tau_n(K) < t_n^n) < \varepsilon$. So it is enough to see that $\max_{1\leq k\leq m(n)+1} |\xi_k^n| \to^p 0$ where

$$\xi_k^n = \frac{1}{\sqrt{t_n^n}} \sum_{i=1}^n \int_{t_{i-1}^n \wedge \tau_n(K)}^{t_i^n \wedge \tau_n(K)} (\psi_k^n(X_{t_{i-1}^n}) - \psi_k^n(X_t))\sigma(X_t) dW_t.$$

Clearly $\xi_{m(n)+1}^n = 0$. For every $1 \leq k \leq m(n)$, note that ξ_k^n is a terminal variable of a continuous martingale. To apply the exponential inequality for continuous martingales, let us compute the predictable variation of ξ_k^n:

$$\begin{aligned}
& \frac{1}{t_n^n} \sum_{i=1}^n \int_{t_{i-1}^n \wedge \tau_n(K)}^{t_i^n \wedge \tau_n(K)} |\psi_k^n(X_{t_{i-1}^n}) - \psi_k^n(X_t)|^2 \sigma(X_t)^2 dt \\
\leq\ & \frac{1}{t_n^n} \sum_{i=1}^n \int_{t_{i-1}^n \wedge \tau_n(K)}^{t_i^n \wedge \tau_n(K)} b_n^{-2} |X_{t_{i-1}^n} - X_t|^2 \sigma_* dt \\
\leq\ & \frac{1}{t_n^n} \sum_{i=1}^n \int_{t_{i-1}^n \wedge \tau_n(K)}^{t_i^n \wedge \tau_n(K)} b_n^{-2} K \operatorname{Log} n |t_i^n - t_{i-1}^n| \sigma_*^2 dt \\
\leq\ & K\sigma_*^2 \cdot b_n^{-2} h_n \operatorname{Log} n.
\end{aligned}$$

Hence, by Lemmas 13 and 14, we have

$$E \max_{1\leq k\leq m(n)} |\xi_k^n| \lesssim \sqrt{K\sigma_*^2 \cdot b_n^{-2} h_n \operatorname{Log} n} \sqrt{\operatorname{Log} m(n)} \to 0.$$

Proof of (8). We introduce the stopping time

$$\tau_n(K) = \inf\left\{t \in [0, t_n^n] : \frac{\sup_{z\in(-\infty,\infty)} l_t(z)}{t_n^n} \geq K\right\},$$

where l_t denotes the local time of X with respect to the speed measure $m_{S_0,\sigma}$. By Theorem 3.1 of van Zanten [24], for every $\varepsilon > 0$ there exists a constant $K > 0$ such that $\limsup_n P(\tau_n(K) < t_n^n) < \varepsilon$. So it is enough to see that $\max_{1\leq k\leq m(n)+1} |\xi_k^n| \to^p 0$ where

$$\xi_k^n = \frac{1}{\sqrt{t_n^n}} \sum_{i=1}^n \int_{t_{i-1}^n\wedge\tau_n(K)}^{t_i^n\wedge\tau_n(K)} (\psi_k^n(X_t) - 1_{(-\infty,x_k^n]}(X_t))\sigma(X_t)dW_t.$$

Clearly $\xi_{m(n)+1}^n = 0$. To apply the exponential inequality for continuous martingales, let us compute the predictable variation of ξ_k^n:

$$\begin{aligned}
&\frac{1}{t_n^n} \sum_{i=1}^n \int_{t_{i-1}^n\wedge\tau_n(K)}^{t_i^n\wedge\tau_n(K)} |\psi_k^n(X_t) - 1_{(-\infty,x_k^n]}(X_t)|^2\sigma(X_t)^2 dt \\
\leq\ & \frac{1}{t_n^n} \sum_{i=1}^n \int_{t_{i-1}^n\wedge\tau_n(K)}^{t_i^n\wedge\tau_n(K)} 1_{[x_k^n, x_k^n+b_n]}(X_t)\sigma_*^2 dt \\
=\ & \sigma_*^2 \frac{1}{t_n^n} \int_{-\infty}^{\infty} 1_{[x_k^n,x_k^n+b_n]}(z) l_{\tau_n(K)}(z) m_{S_0,\sigma}((-\infty,\infty)) f_{S_0,\sigma}(z)dz \\
\leq\ & \sigma_*^2 K m_{S_0,\sigma}((-\infty,\infty)) \sup_z f_{S_0,\sigma}(z)\cdot b_n.
\end{aligned}$$

Hence, by Lemmas 13 and 14, we have

$$E \max_{1\leq k\leq m(n)} |\xi_k^n| \lesssim \sqrt{\sigma_*^2 K m_{S_0,\sigma}((-\infty,\infty)) \sup_z f_{S_0,\sigma}(z)\cdot b_n}\sqrt{\operatorname{Log} m(n)} \to 0.$$

Proof of (9). It is sufficient to show that

$$\max_{1\leq k\leq m(n)+1} \sup_{x\in(x_{k-1}^n, x_k^n]} |V^n(x_k^n) - V^n(x)| \to^p 0.$$

Notice that Lemma 12 implies also that for every $\varepsilon, \eta > 0$ there exists $\delta > 0$ such that

$$P\left(\sup_{\rho(x,y)<\delta} |V^n(x) - V^n(y)| > \varepsilon\right) < \eta.$$

Since $\rho(x, y) \leq (\sup_z(\sigma(z)^2 f_{S_0,\sigma}(z) + \phi(z)))\sqrt{|x-y|} \leq \text{constant}\sqrt{|x-y|}$, we have $\max_{2\leq k\leq m(n)} \rho(x_{k-1}^n, x_k^n) \to 0$. Also, it is clear that $\rho(-\infty, x_1^n) \to 0$ and $\rho(x_{m(n)}^n, \infty) \to 0$. Hence we have (9).

Now (6) – (9) have been proved, and the proof of Lemma 3 is finished. □

Proof of Lemma 7. By Itô's formula, we have

$$|X_{t_i^n}|^2 - |X_{t_{i-1}^n}|^2 = 2\int_{t_{i-1}^n}^{t_i^n} X_s dX_s + \int_{t_{i-1}^n}^{t_i^n} \sigma(X_s)^2 ds.$$

Since

$$|\widehat{\Sigma}^n|^2 = \frac{1}{t_n^n}\sum_{i=1}^n \left\{ |X_{t_i^n}|^2 - |X_{t_{i-1}^n}|^2 - 2X_{t_{i-1}^n}(X_{t_i^n} - X_{t_{i-1}^n}) \right\},$$

it is enough to show that

$$\frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} (X_s - X_{t_{i-1}^n}) dX_s \to^p 0$$

and

$$\frac{1}{t_n^n}\int_0^{t_n^n} \sigma(X_s)^2 ds \to^p \Sigma_{S,\sigma}^2.$$

The latter is nothing else than the ergodicity. As for the former, observe that

$$\begin{aligned}
&\frac{1}{t_n^n}\left|\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} (X_s - X_{t_{i-1}^n}) dX_s\right| \\
&\leq \frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} |X_s - X_{t_{i-1}^n}||S(X_s)|ds + \frac{1}{t_n^n}\left|\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} (X_s - X_{t_{i-1}^n})\sigma(X_s)dW_s\right|.
\end{aligned}$$

By Lemma 15, the expectation of the first term on the right hand side is

$$\begin{aligned}
&\frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} E(|X_s - X_{t_{i-1}^n}||S(X_s)|)ds \\
&\leq \frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \sqrt{E|X_s - X_{t_{i-1}^n}|^2}\sqrt{E|S(X_s)|^2}ds \\
&\lesssim \frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \sqrt{h_n}\sqrt{E|S(X_s)|^2}ds \\
&\leq \sqrt{h_n}\sup_s \sqrt{K_{S,\sigma}^2 E(1+|X_s|)^2} \\
&\to 0.
\end{aligned}$$

On the other hand, the expectation of the square of the second term on the right hand side is

$$\begin{aligned}
&\frac{1}{|t_n^n|^2}\sum_{i=1}^n E\int_{t_{i-1}^n}^{t_i^n} |X_s - X_{t_{i-1}^n}|^2\sigma(X_s)^2 ds \\
&\lesssim \frac{1}{|t_n^n|^2}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} |s - t_{i-1}^n|\sigma_*^2 ds \\
&\leq \frac{1}{t_n^n} h_n \sigma_*^2 \\
&\to 0.
\end{aligned}$$

This proves the consistency of our estimator. □

Proof of Lemma 9. We write $U^n_\Delta(x_S) = \sqrt{t^n_n} A^n_1$, where

$$A^n_1 = \frac{1}{t^n_n}\sum_{i=1}^n \psi^n_{k(n)}(X_{t^n_{i-1}})(S(X_{t^n_{i-1}}) - S_0(X_{t^n_{i-1}}))|t^n_i - t^n_{i-1}|$$

and $k(n)$ is the number k such that $x_S \in (x^n_{k-1}, x^n_k]$. It suffices to show that A^n_1 converges in probability to a constant which is not zero. Now, by the ergodicity we have

$$\begin{aligned} A^n_4 &= \frac{1}{t^n_n}\int_0^{t^n_n} 1_{(-\infty,x_S]}(X_t)(S(X_t) - S_0(X_t))dt \\ &\to \int_{-\infty}^{\infty} 1_{(-\infty,x_S]}(z)(S(z) - S_0(z))f_{S,\sigma}(z)dz \neq 0, \quad \text{almost surely.} \end{aligned}$$

Hence, introducing

$$A^n_2 = \frac{1}{t^n_n}\sum_{i=1}^n \int_{t^n_{i-1}}^{t^n_i} \psi^n_{k(n)}(X_{t^n_{i-1}})(S(X_t) - S_0(X_t))dt$$

and

$$A^n_3 = \frac{1}{t^n_n}\sum_{i=1}^n \int_{t^n_{i-1}}^{t^n_i} \psi^n_{k(n)}(X_t)(S(X_t) - S_0(X_t))dt,$$

we will show that $A^n_1 - A^n_2 \to^p 0$, $A^n_2 - A^n_3 \to^p 0$ and $A^n_3 - A^n_4 \to^p 0$.

To see $A^n_1 - A^n_2 \to^p 0$, notice that

$$\begin{aligned} E|A^n_1 - A^n_2| \;&\leq\; \frac{1}{t^n_n}\sum_{i=1}^n E\left(\int_{t^n_{i-1}}^{t^n_i} \psi^n_{k(n)}(X_{t^n_{i-1}})|S(X_{t^n_{i-1}}) - S(X_t)|dt\right) \\ &\quad + \frac{1}{t^n_n}\sum_{i=1}^n E\left(\int_{t^n_{i-1}}^{t^n_i} \psi^n_{k(n)}(X_{t^n_{i-1}})|S_0(X_{t^n_{i-1}}) - S_0(X_t)|dt\right). \end{aligned}$$

Using Lemma 15, the first term on the right hand side is bounded by

$$\begin{aligned} &\frac{1}{t^n_n}\sum_{i=1}^n \int_{t^n_{i-1}}^{t^n_i} E|S(X_{t^n_{i-1}}) - S(X_t)|dt \\ &\leq \frac{1}{t^n_n}\sum_{i=1}^n K_{S,\sigma}\int_{t^n_{i-1}}^{t^n_i} E|X_{t^n_{i-1}} - X_t|dt \\ &\leq \frac{1}{t^n_n}\sum_{i=1}^n K_{S,\sigma}\int_{t^n_{i-1}}^{t^n_i} \sqrt{E|X_{t^n_{i-1}} - X_t|^2}dt \\ &\lesssim \frac{1}{t^n_n}\sum_{i=1}^n \int_{t^n_{i-1}}^{t^n_i} |t - t^n_{i-1}|^{1/2}dt \\ &\leq \frac{1}{t^n_n}\cdot t^n_n \cdot h^{1/2}_n \to 0. \end{aligned}$$

The second term is also estimated in the same way.

To see $A_2^n - A_3^n \to 0$, notice that

$$\begin{aligned} E|A_2^n - A_3^n| &\le \frac{1}{t_n^n}\sum_{i=1}^n E\left(\int_{t_{i-1}^n}^{t_i^n} (\psi_{k(n)}^n(X_{t_{i-1}^n}) - \psi_{k(n)}^n(X_t))|S(X_t)|dt\right) \\ &\quad + \frac{1}{t_n^n}\sum_{i=1}^n E\left(\int_{t_{i-1}^n}^{t_i^n} (\psi_{k(n)}^n(X_{t_{i-1}^n}) - \psi_{k(n)}^n(X_t))|S_0(X_t)|dt\right). \end{aligned}$$

The first term on the right hand side is bounded by

$$\begin{aligned} & \frac{1}{t_n^n}\sum_{i=1}^n E\left(\int_{t_{i-1}^n}^{t_i^n} b_n^{-1}|X_{t_{i-1}^n} - X_t||S(X_t)|dt\right) \\ \le & \frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} b_n^{-1}\sqrt{E|X_{t_{i-1}^n} - X_t|^2}\sqrt{E|S(X_t)|^2}dt \\ \lesssim & \frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} b_n^{-1}\sqrt{|t - t_{i-1}^n|}dt \\ \le & \frac{1}{t_n^n}\cdot t_n^n \cdot b_n^{-1}\sqrt{h_n} \to 0. \end{aligned}$$

The second term is also estimated in the same way.

To see $A_3^n - A_4^n \to^p 0$, notice that

$$\begin{aligned} |A_3^n - A_4^n| &\le \frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} 1_{[x_{k(n)-1}^n, x_{k(n)}^n + b_n]}(X_t)|S(X_t)|dt \\ &\quad + \frac{1}{t_n^n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} 1_{[x_{k(n)-1}^n, x_{k(n)}^n + b_n]}(X_t)|S_0(X_t)|dt. \end{aligned}$$

Using the local time l_t, the first term on the right hand side is bounded by $m_{S,\sigma}((-\infty,\infty))$ times

$$\begin{aligned} & \int_{-\infty}^{\infty} 1_{[x_{k(n)-1}^n, x_{k(n)}^n + b_n]}(z)|S(z)|\frac{l_{t_n^n}(z)}{t_n^n} f_{S,\sigma}(z)dz \\ \le & \sqrt{\int_{-\infty}^{\infty} 1_{[x_{k(n)-1}^n, x_{k(n)}^n + b_n]}(z) f_{S,\sigma}(z)dz}\sqrt{\int_{-\infty}^{\infty} |S(z)|^2 \left|\frac{l_{t_n^n}(z)}{t_n^n}\right|^2 f_{S,\sigma}(z)dz}. \end{aligned}$$

Since $(t_n^n)^{-1}\sup_z l_{t_n^n}(z) = O_P(1)$, the right hand side is $o_P(1)$. The second term is also analyzed in the same way, and the lemma has been proved. □

Appendix I

First, let us review the theory of random fields generated by continuous martingales developed by Nishiyama [19], [20]. Let T be a set which is totally bounded with respect to a metric ρ. Let a family $M^n = \{M^{n,x}; x \in T\}$ of continuous martingales $t \rightsquigarrow M_t^{n,x}$ be given, where the underlying stochastic basis $(\Omega^n, \mathcal{F}^n, (\mathcal{F}_t^n)_{t\geq 0}, P^n)$ is common for all $x \in T$. Let a finite stopping time τ^n be given. We define the *quadratic modulus* $t \rightsquigarrow ||M^n||_{\rho,t}$ by

$$||M^n||_{\rho,t} = \sup_{x\neq y} \frac{\sqrt{\langle M^{n,x} - M^{n,y}\rangle_t}}{\rho(x,y)}.$$

Since T is not necessarily countable, $||M^n||_{\rho,t}$ may not have any measurability. So we introduce the *adapted quadratic modulus* $t \rightsquigarrow ||M^n||_{\rho,t}^*$ as any ($[0,\infty]$-valued) adapted process such that $||M^n||_\rho \leq ||M^n||_\rho^*$ almost surely. We denote by $N(T,\rho;\varepsilon)$ the smallest number of closed balls with ρ-radius ε which cover T. Then, we have the following result.

Lemma 12 *Suppose* $\int_0^1 \sqrt{\log N(T,\rho;\varepsilon)}d\varepsilon < \infty$.

(i) If $||M^n||_{\rho,\tau^n}^* < \infty$ *almost surely, then there exists a uniformly* ρ*-continuous version of the random field* $M_{\tau^n}^n = \{M_{\tau^n}^{n,x}; x \in T\}$.

(ii) If $||M^n||_{\rho,\tau^n}^* = O_P(1)$*, and if* $\langle M^{n,x}, M^{n,y}\rangle_{\tau^n} \rightarrow^p C^{(x,y)}$*, then the random field* $M_{\tau^n}^n$ *is asymptotically uniformly* ρ*-equicontinuous in probability and converges weakly to* G *in* $C_\rho(T)$*, where* $G = \{G(x); x \in T\}$ *is a zero-mean Gaussian random field with the co-variance* $EG(x)G(y) = C^{(x,y)}$*. Almost all paths of* G *are uniformly* ρ*-continuous.*

Proof. See Theorems 2.4.3 and 3.4.2 of Nishiyama [20]. The essence is the maximal inequality for continuous martingales; see Theorem 2.3 of Nishiyama [19]. □

Here, we state the *exponential inequality* for continuous martingales.

Lemma 13 *Let* M *be a continuous martingale, and let* τ *be a bounded stopping time. For every* $x, v > 0$ *it holds that*

$$P\left(\sup_{t\in[0,\tau]} |M_t| > x,\ \langle M\rangle_\tau \leq v\right) \leq 2\exp\left(-\frac{x^2}{2v}\right).$$

The following lemma, the *maximal inequality* for general random variables, is used in connection with Lemma 13 and plays a key role in our approach.

Lemma 14 *(i) Let* $X_1, ..., X_m$ *be arbitrary random variables which satisfy*

$$P(|X_i| > x) \leq 2\exp\left(-\frac{x}{a}\right)$$

for all x *and* i *and a fixed constant* $a > 0$*. Then there exists a universal constant* $C > 0$ *such that*

$$E\left(\max_{1\leq i\leq m} |X_i|\right) \leq Ca\log(1+m).$$

(ii) Let $X_1, ..., X_m$ be arbitrary random variables which satisfy

$$P(|X_i| > x) \leq 2\exp\left(-\frac{x^2}{b}\right)$$

for all x and i and a fixed constant $b > 0$. Then there exists a universal constant $C > 0$ such that

$$E\left(\max_{1\leq i\leq m} |X_i|\right) \leq C\sqrt{b}\sqrt{\log(1+m)}.$$

Proof. Use Lemmas 2.2.1 and 2.2.2 of van der Vaart and Wellner [21]. □

The following lemma is well known.

Lemma 15 *Let X be a solution to the SDE (1) for (S,σ) which satisfies* **A1**. *Let $k \geq 2$. Then, there exists a constant $C_k > 0$, depending only on k, such that for any $0 \leq t < t'$ with $|t'-t| \leq 1$*

$$E\sup_{u\in[t,t']} |X_u - X_t|^k \quad\leq\quad C_k K_{S,\sigma}^k \sup_s E(1+|X_s|)^k |t'-t|^{k/2}.$$

Proof. Use Hölder's inequality and Burkholder-Davis-Gundy's inequality. □

Appendix II

For convenience we here give two sets of simple sufficient conditions for $\int_{-\infty}^{\infty} x^2 f_{S,\sigma}(x)dx < \infty$ (the last part of **A2**) and **A4**, under the conditions **A1** and $\sigma_*^2 < \infty$ (a part of **A3**).

(a) There exist some constants $a, c, K > 0$ such that $E\exp(aX_0^2) < \infty$ and that $xS(x) \leq -cx^2$ for every $|x| \geq K$.

(b) S is bounded, and there exist some constants $a, c, K >$ such that $E\exp(a|X_0|) < \infty$ and that $xS(x) \leq -c|x|$ for every $|x| \geq K$.

To see these statements, we may apply the proofs for Propositions 1.1 and 5.1 in Gobet [7]. Another way is to apply the argument of the proof for Theorem 2.2 in Masuda [15] with Lyapunov-test functions such as $x \mapsto x^2$.

Acknowledgements. This work was supported by MIUR 2006 Grant and ISM Cooperative Research Program.

References

[1] Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusion. *J. Finance*, **54**, 1361-1395.

[2] Dachian, S. and Kutoyants, Yu.A. (2008). On the goodness-of-fit tests for some continuous time processes. In *Statistical Models and Methods for Biomedical and Technical Systems,* 395-413. Vonta, F., Nikulin, M., Limnios, N. and Huber-Carol C. (Eds), Birkhuser, Boston.

[3] Delgado, M. and Stute, W. (2008). Distribution-free specification tests of conditional models. *J. Econometrics* **143**, 37-55.

[4] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications. Volume II.* (2nd ed.) John Wiley & Sons, Inc., New York.

[5] Fournie, E., (1992). Un test de type Kolmogorov-Smirnov pour processus de diffusion ergodiques, *Rapporte de recherche*, **1696**, INRIA, Sophia-Antipolis.

[6] Fournie, E. and Kutoyants, Yu.A., (1993). Estimateur de la distance maximale pour le processus de diffusion ergodiques, *Rapporte de recherche*, **1952**, INRIA, Sophia-Antipolis.

[7] Gobet, E. (2002). LAN property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincare Probab. Statist.* **38**, 711-737.

[8] Gobet, E., Hoffmann, M. and Reiss, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32**, 2223-2253.

[9] Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Process. Appl.* **79**, 135-163.

[10] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 221-229.

[11] Koul, H. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models.* (2nd ed.) Springer, New York.

[12] Koul, H. and Stute. W. (1999). Nonparametric model checks in time series. *Ann. Statist.* **27**, 204-237.

[13] Kutoyants, Yu.A. (2004). *Statistical Inference for Ergodic Diffusion Processes.* Springer, London.

[14] Lee, S. and Wee, I. (2008). Residual empirical process for diffusion processes. *J. Korean Math. Soc.* **45** 683-693.

[15] Masuda, H. (2007). Ergodicity and exponential β-mixing bound for multidimensional diffusions with jumps. *Stochastic Process. Appl.* **117**, 35-56.

[16] Negri, I. (1998). Stationary distribution function estimation for ergodic diffusion process. *Statistical Inference for Stochastic Processes*, **1**, 61–84.

[17] Negri, I. and Nishiyama, Y. (2007a). Goodness of fit test for ergodic diffusion processes. *To appear in Ann. Inst. Statist. Math.*

[18] Negri, I. and Nishiyama, Y. (2007b). Goodness of fit test for small diffusions based on discrete observations. *Research Memorandum* **1054**, Inst. Statist. Math., Tokyo.

[19] Nishiyama, Y. (1999). A maximal inequality for continuous martingales and M-estimation in a Gaussian white noise model. *Ann. Statist.* **27**, 675-696.

[20] Nishiyama, Y. (2000). *Entropy Methods for Martingales.* CWI Tract **128**, Centrum voor Wiskunde en Informatica, Amsterdam.

[21] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer-Verlag, New York.

[22] van der Vaart, A.W. and van Zanten, H. (2005). Donsker theorems for diffusions: necessary and sufficient conditions. *Ann. Probab.* **33**, 1422-1451.

[23] Yoshida, N. (1992). Estimation for diffusion processes from discrete observations. *J. Multivariate Anal.* **41**, 220-242.

[24] van Zanten, H. (2003). On uniform laws of large numbers for ergodic diffusions and consistency of estimators. *Statist. Inference Stoch. Process.* **6**, 199-213.