

# Two sample test for diffusion processes with a non-linear covariate

Yoichi Nishiyama

*The Institute of Statistical Mathematics*

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

nisiyama@ism.ac.jp

March 9, 2009

Running title: Two sample test for diffusion processes

## Abstract

Let two diffusion processes  $t \rightsquigarrow X_t^{(p)}$ ,  $p = 1, 2$ , with the drift  $S^{(p)}(Z_t)$  where the covariate  $t \rightsquigarrow Z_t$  is common for  $p = 1, 2$ , be observed. We test whether the functions  $S^{(p)}$  for  $p = 1, 2$  are the same or not. We present an asymptotically distribution free test and prove its consistency.

Keywords. Diffusion process, invariance principle, martingale, metric entropy, two sample test, weak convergence.

# 1 Introduction

Suppose that two diffusion processes  $t \rightsquigarrow X_t^{(p)}$  on  $\mathbb{R}$ ,  $p = 1, 2$ , given by

$$X_t^{(p)} = X_0^{(p)} + \int_0^t S^{(p)}(Z_s) ds + \int_0^t \sigma(Z_s) dW_s^{(p)}$$

are observed. Here,  $t \rightsquigarrow Z_t$  is an observable covariate process which is common for  $p = 1, 2$ , and  $W^{(1)}$  and  $W^{(2)}$  are independent standard Wiener processes. The covariate process  $Z$  is assumed to take values in a measurable space  $(I, \mathcal{I})$ , and it represents the effect of the underlying circumstance which is common for  $X^{(1)}$  and  $X^{(2)}$ , such as the social events, while  $S^{(p)}$ ,  $p = 1, 2$ , and  $\sigma$  are deterministic measurable functions on  $I$ . We assume that  $\sigma$  is known and common for  $p = 1, 2$ . We are interested in testing the hypothesis  $H_0 : S^{(1)} = S^{(2)}$  versus  $H_1 : S^{(1)} \neq S^{(2)}$ . The precise meaning of “ $S^{(1)} \neq S^{(2)}$ ” will be stated later. As for the process  $Z$ , we shall assume the ergodicity and the existence of a “good” local time. The latter is satisfied if the process  $Z$  is a one-dimensional regular diffusion with finite speed measure. Such a case already brings rich applications, for example in econometrics. Indeed, our model includes the so-called stochastic volatility models, although in that context the process  $Z$  is usually assumed to be unobservable.

This work is motivated by Negri and Nishiyama (2006) who considered a goodness of fit test problem for ergodic diffusions based on a marked empirical process; see Nishiyama (2009) for a review (in Japanese) including also discrete observation cases. We will present an asymptotically distribution free test and prove its consistency under any fixed alternative. Our main tool is a functional central limit theorem for continuous martingales indexed by an abstract set, which is actually a class of functions in the current case, established by Nishiyama (2000). The key idea for our two sample test statistic is found in Nishiyama (2008) who considered a counting process model with a non-linear covariate. By extending the idea we may be able to consider also  $k$  sample problems. However, our goal is to derive an *asymptotically distribution free* test, so we consider the case  $k = 2$  only, and pay special attention to the case where  $I$  is the one-dimensional Euclidean space.

We refer to van der Vaart and Wellner (1996) for the weak convergence theory in  $\ell^\infty(\Psi)$ -space, the space of bounded functions on a set  $\Psi$ . We denote by  $\rightarrow^p$  and  $\rightarrow^d$  the convergence in probability and the convergence in distribution, respectively. The limit notations always mean to take a limit as  $T \rightarrow \infty$ .

## 2 Result

We suppose the following two properties on the covariate  $Z$ :

**A1.** (Ergodicity.) There exists a probability measure  $\mu$  on  $(I, \mathcal{I})$  such that for every  $\mu$ -integrable function  $\psi$  on  $(I, \mathcal{I})$

$$\frac{1}{T} \int_0^T \psi(Z_t) dt \rightarrow^p \int_I \psi(z) \mu(dz), \quad \text{as } T \rightarrow \infty.$$

**A2.** (Existence of a good local time.) There exists a non-negative predictable processes  $t \rightsquigarrow l_t(z)$ ,  $z \in I$  and a measure  $m$  on  $(I, \mathcal{I})$  such that for any measurable function  $\psi$  on  $(I, \mathcal{I})$

$$\int_0^T \psi(Z_t) dt = \int_I \psi(z) l_T(z) m(dz), \quad \text{almost surely,}$$

provided the integrals on the both sides exist, and that

$$\sup_{z \in I} l_T(z) = O_P(T), \quad \text{as } T \rightarrow \infty.$$

We notice that the condition A2 is satisfied if  $I \subseteq \mathbb{R}$  and  $Z$  is a regular diffusion with finite speed measure  $m$ . See van Zanten (2003).

**Theorem 1** *Assume A1 and A2. Suppose that the class  $\Psi \subset L_2(m(dz)) \cap L_2(\mu(dz))$  satisfies the metric entropy condition:*

$$\int_0^1 \sqrt{\log N(\Psi, L_2(m(dz)), \varepsilon)} d\varepsilon < \infty,$$

where  $N(\Psi, L_2(m(dz)), \varepsilon)$  denotes the smallest number of  $\varepsilon$ -balls with respect to the  $L_2$ -metric  $\|\cdot\|_{L_2(m(dz))}$  which cover the class  $\Psi$ . Then, it holds that the random fields  $\xi_T = \{\xi_T(\psi); \psi \in \Psi\}$  defined by

$$\xi_T(\psi) = \frac{1}{\sqrt{2T}} \int_0^T \frac{\psi(Z_t)}{\sigma(Z_t)} \left[ (dX_t^{(1)} - S^{(1)}(Z_t)dt) - (dX_t^{(2)} - S^{(2)}(Z_t)dt) \right]$$

converges weakly in  $\ell^\infty(\Psi)$  to a zero-mean Gaussian random field  $G = \{G(\psi); \psi \in \Psi\}$  with the covariance

$$EG(\psi)G(\psi') = \int_I \psi(z)\psi'(z)\mu(dz).$$

Furthermore, almost all paths of  $\psi \rightsquigarrow G(\psi)$  are uniformly  $\rho$ -continuous with respect to the semimetric  $\rho(\psi, \psi') = \sqrt{E|G(\psi) - G(\psi')|^2}$ .

*Proof.* We apply Corollary 3.4.3 of Nishiyama (2000) to the terminal variables  $\{M_T^{T,\psi}; \psi \in \Psi\}$  of the continuous martingales  $t \rightsquigarrow M_t^{T,\psi}$  given by

$$M_t^{T,\psi} = \frac{1}{\sqrt{2T}} \int_0^t \psi(Z_s) [dW_s^{(1)} - dW_s^{(2)}].$$

The finite-dimensional convergence follows from the fact that

$$\langle M^{T,\psi}, M^{T,\psi'} \rangle_T = \frac{1}{T} \int_0^T \psi(Z_s)\psi'(Z_s) ds \xrightarrow{p} \int_I \psi(z)\psi'(z)\mu(dz).$$

As for [ME], the square of the *quadratic modulus* with respect to the metric  $\|\cdot\|_{L_2(m(dz))}$  is given by

$$\sup_{\substack{\psi, \psi' \in \Psi \\ \psi \neq \psi'}} \frac{T^{-1} \int_0^T |\psi(Z_s) - \psi'(Z_s)|^2 ds}{\|\psi - \psi'\|_{L_2(m(dz))}^2} \leq T^{-1} \sup_{z \in I} l_T(z) = O_P(1).$$

The proof is finished.  $\square$

When  $I = \mathbb{R}$ , by setting  $\bar{\Psi} = \{1_{(-\infty, u]}; u \in \mathbb{R}\}$ , the limit process in Theorem 1 becomes a time change of a standard Brownian motion  $\{B(v); v \in [0, 1]\}$ , namely,  $u \rightsquigarrow B(\mu(-\infty, u])$ . This suggests the following test statistic for the two sample problem.

**Theorem 2** *Let  $I = \mathbb{R}$ , and assume A1 and A2. Consider the test statistic*

$$\mathcal{S}_T = \sup_{u \in \mathbb{R}} \frac{1}{\sqrt{2T}} \left| \int_0^T \frac{1_{(-\infty, u]}(Z_t)}{\sigma(Z_t)} [dX_t^{(1)} - dX_t^{(2)}] \right|.$$

(i) *Assume  $u \rightsquigarrow \mu(-\infty, u]$  is continuous. Under the hypothesis  $H_0 : S^{(1)} = S^{(2)}$ , it holds that  $\mathcal{S}_T \xrightarrow{d} \sup_{v \in [0, 1]} |B(v)|$ , where  $v \rightsquigarrow B(v)$  is a standard Brownian motion.*

(ii) *Under the hypothesis  $H_1 : \int_{-\infty}^u \frac{1}{\sigma(z)} (S^{(1)}(z) - S^{(2)}(z)) \mu(dz) \neq 0$  for some  $u \in \mathbb{R}$ , it holds that  $P(\mathcal{S}_T \leq M) \rightarrow 0$  for any  $M$ .*

**Remark.** The limit distribution under the null hypothesis is given by

$$P\left(\sup_{v \in [0, 1]} |B(v)| \leq x\right) = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left(-\frac{\pi^2(2k+1)^2}{8x^2}\right).$$

See e.g. 343 page of Feller (1971) for this formula. Theorem 1 itself is true also for the case where  $I$  is a multi-dimensional space, but in such a case the limit is not distribution free.

*Proof.* The claim (i) follows directly from Theorem 1 with help of the continuous mapping theorem.

To show (ii), observe that

$$\begin{aligned} & \sup_{u \in \mathbb{R}} \frac{1}{\sqrt{2T}} \left| \int_0^T \frac{1_{(-\infty, u]}(Z_t)}{\sigma(Z_t)} [dX_t^{(1)} - dX_t^{(2)}] \right| \\ & \geq \sup_{u \in \mathbb{R}} \sqrt{T/2} \frac{1}{T} \left| \int_0^T \frac{1_{(-\infty, u]}(Z_t)}{\sigma(Z_t)} (S^{(1)}(Z_t) - S^{(2)}(Z_t)) dt \right| - \sup_{u \in \mathbb{R}} |\xi_T(1_{(-\infty, u]})|, \end{aligned}$$

where  $u \rightsquigarrow \xi_T(1_{(-\infty, u]})$  is the process in Theorem 1 in the case of  $\psi = 1_{(-\infty, u]}$ . Since the process  $u \rightsquigarrow \xi_T(1_{(-\infty, u]})$  converges weakly in  $\ell^\infty(\mathbb{R})$  to a tight law, the second term on the right hand side is tight. The first term on the right hand side tends to

$\infty$  in probability because there exists some  $u_0 \in \mathbb{R}$  such that  $\int_{\mathbb{R}} \frac{1_{(-\infty, u_0]}(z)}{\sigma(z)} (S^{(1)}(z) - S^{(2)}(z)) \mu(dz) \neq 0$  and for such a  $u_0$  it holds that

$$\frac{1}{T} \left| \int_0^T \frac{1_{(-\infty, u_0]}(Z_t)}{\sigma(Z_t)} (S^{(1)}(Z_t) - S^{(2)}(Z_t)) dt \right| \xrightarrow{p} \left| \int_{-\infty}^{u_0} \frac{1}{\sigma(z)} (S^{(1)}(z) - S^{(2)}(z)) \mu(dz) \right| > 0.$$

The proof is completed. □

## References

- [1] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications. Volume II.* (2nd ed.) John Wiley & Sons, Inc., New York.
- [2] Negri, I. and Nishiyama, Y. (2006). Goodness of fit test for ergodic diffusion processes. *To appear in Ann. Inst. Statist. Math.*
- [3] Nishiyama, Y. (2000). *Entropy Methods for Martingales.* CWI Tract **128**, Centrum voor Wiskunde en Informatica, Amsterdam.
- [4] Nishiyama, Y. (2008). Two sample test for counting processes with a non-linear covariate. *Research Memorandum 1071*, Inst. Statist. Math., Tokyo.
- [5] Nishiyama, Y. (2009). Nonparametric goodness of fit tests for diffusion processes. (In Japanese.) *To appear in Proc. Inst. Statist. Math.*
- [6] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer-Verlag, New York.
- [7] van Zanten, H. (2003). On uniform laws of large numbers for ergodic diffusions and consistency of estimators. *Statist. Inference Stoch. Process.* **6**, 199-213.