Goodness of fit test for diffusions by different sample schemes

(joint works with Y. Nishiyama and H. Masuda)

Ilia Negri

Department of Information Technology and Mathematical Methods

University of Bergamo (Italy)

ilia.negri@unibg.it

SAPS VII – Le Mans, 16-19 March 2009

Plan of the talk

- Introduction: goodness of fit test
- Goodness of fit tests for diffusion models
- Asymptotic distribution free test for ergodic diffusion
  - continuous time observation
  - discrete time observation
  - tick time sample scheme (not yet)
  - simulation studies

Introduction: goodness of fit test

Goodness of fit tests play an important role in theoretical and applied statistics.

Such test are useful if they are distribution free or asymptotically distribution free.

The origin goes back to the Kolmogorov-Smirnov and the Cramer-Von Mises tests in the i.i.d case.

Despite their importance in applications goodness of fit test for diffusion process has still been a new issue in recent years.

We present a test for the drift of a diffusion, when the diffusion coefficient is a nuisance function which is estimated in our test procedure.

Such tests are asymptotically distribution free and consistent. We consider the case based on continuos time observation and the case (more interesting for application) based on discrete time observation.

Goodness of fit tests for diffusion models

Given a general set-up $(\Omega, \{\mathcal{F}_t\} \subset \mathcal{F}, \mathbf{P})$, let $\{W_t : t \geq 0\}$ be a Brownian motion. Suppose that we observe the process $\{X_t : 0 \leq t \leq T\}$, solution of the stochastic differential equation

$$dX_t = S(X_t)dt + \sigma(X_t)dW_t, \quad X_0 \quad t \geq 0$$

To test the simple hypotheses

$$H_0 : S = S_0$$

against any alternative $S_1 \neq S_0$, where $S_1 \neq S_0$ means $\sup_x |S_1(x) - S_0(x)| > 0$, we can introduce the statistic

$$\Delta_T(X^T) = \sup_x \sqrt{T} |\widehat{F}_T(x) - F_{S_0}(x)| = \sup_x |\eta^T_{S_0}(x)|$$

where the empirical distribution function is defined as

$$\widehat{F}_T(x) = \frac{1}{T} \int_0^T \mathbf{1}_{\{X_t \leq x\}} dt.$$

and the empirical process as

$$\eta^T_S(x) = \sqrt{T}(\widehat{F}_T(x)) - F_S(x))$$

Goodness of fit test based on $\widehat{F}_T(x)$

The decision function for this test procedure is $\widehat{\phi}_T(X^T) = \mathbf{1}_{\{\Delta_T(x) > c_\alpha\}}$, where c_α is the solution of the equation

$$\mathbf{P}\left(\sup_x |\eta_{S_0}(x)| > c_\alpha\right) = \alpha$$

and $\eta_{S_0} = \{\eta_{S_0}(x) : x \in \mathbb{R}\}$ is a suitable Gaussian process.

Theorem (Kutoyants 2004)

Under some regularity conditions for S the test based on

$$\widehat{\phi}_T(X^T) = \mathbf{1}_{\{\Delta_T(X^T) > c_\alpha\}}$$

is asymptotically of size α and it is consistent.

The result follows from the weak convergence of the empirical process to the Gaussian process η_{S_0} studied in Negri (1998) and the continuous mapping theorem.

Unfortunately, due to the structure of the covariance function of η_{S_0} the Kolmogorov-Smirnov statistics is not asymptotically distribution free.

Goodness of fit test for diffusion: continuous time observation

Suppose we observe the ergodic diffusion processes solution of

$$dX_t = S(X_t)dt + \sigma(X_t)dW_t, \quad X_0 \quad t \geq 0$$

on the continuous time interval $[0, T]$

$$\{X_t : 0 \leq t \leq T\}$$

To test the simple hypothesis

$$H_0 : S = S_0$$

against any alternative $S_1 \neq S_0$, we propose a test statistics based on the *score marked empirical process*, defined as:

$$V_T(x) = \frac{1}{\sqrt{T}} \int_0^T \mathbf{1}_{(-\infty,x]}(X_t)(\mathrm{d}X_t - S_0(X_t)\mathrm{d}t)$$

Some other approaches in Dachian, Kutoyants (2007) and Kutoyants (2009)

Conditions on diffusion models

A1. There exists a constant $K > 0$ such that

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

Under this condition, the SDE has a unique strong solution X.

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

Under conditions **A1** and **A2** the diffusion process is ergodic and has an invariant density $f_{S,\sigma}$ given by

$$f_{S,\sigma}(y) = \frac{1}{G}\frac{1}{\sigma(y)^2} \exp\left\{2\int_0^y \frac{S(v)}{\sigma(v)^2}dv\right\}$$

where G is given by $G = m_{S,\sigma}(\mathbb{R})$.

Conditions on diffusion models

We introduce the metric $\rho_{S,\sigma}$ on $[-\infty, \infty]$ given by

$$\rho_{S,\sigma}(x,y) = \sqrt{\int_{x\wedge y}^{x\vee y} (\sigma(z)^2 f_{S,\sigma}(z) + \phi(z))dz}. \quad (1)$$

where ϕ is the density of the standard Gaussian distribution.

We introduce the following condition

A3. The invariant density $f_{S,\sigma}$ satisfies

$$\Sigma^2_{S,\sigma} := \int_{-\infty}^{+\infty} \sigma(z)^2 f_{S,\sigma}(z)dz \in (0, +\infty)$$

In order to obtain an asymptotically distribution free test we need a consistent estimator for $\Sigma^2_{S_0,\sigma}$, for a given S_0.

In continuos time observation we can suppose σ known, so we can compute $\Sigma^2_{S_0,\sigma}$ for a given S_0.

Some notations

We introduce the following notation. Let $(\mathbb{T}, \rho)$ be a metric space.

We denote by $C_\rho(\mathbb{T})$ the space of continuous functions on $\mathbb{T}$

We denote by $\ell_\rho^\infty(\mathbb{T})$ the space of bounded functions on $\mathbb{T}$

We equip both the spaces with the uniform metric

We denote by "$\rightarrow^p$" and "$\rightarrow^d$" the convergence in probability and in distribution respectively

The notation "$\rightarrow$" always means that we take the limit as $T \rightarrow \infty$ or $n \rightarrow \infty$

In our work we consider $\mathbb{T} = [-\infty, +\infty]$

Goodness of fit test for diffusion: continuous time observation

The test based on the score marked empirical process is asymptotically distribution free.

The main result is given by the following

Theorem. Assume **A1**, **A2** and **A3** for (S_0, σ). Under $H_0 : S = S_0$, suppose that a positive consistent estimator $\widehat{\Sigma}^T$ for $\Sigma_{S_0,\sigma}$ is given. Then it holds that

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

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

So the test based on the following statistical decision function

$$\phi_T^* = \mathbf{1}_{\left\{ \frac{\sup_{x\in[-\infty,\infty]} |V^T(x)|}{\widehat{\Sigma}^T} > c_\alpha \right\}}$$

where the *critical value* c_α is defined by

$$\mathbf{P}\left(\sup_{0\le t\le 1} |B_t| > c_\alpha \right) = \alpha,$$

is asymptotically distribution free (Negri and Nishiyama 2008).

It is well known that the distribution function of supremum over $t \in [0,1]$ of $|B_t|$ is given by

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

Proof of the result

The result is based on the following lemma

Lemma . $V^T \to^d G$ in $C_\rho[-\infty, +\infty]$, where $G = \{G(x) : x \in [-\infty, +\infty]\}$ is a zero-mean Gaussian process with co-variance

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

Almost all paths of G are uniformly ρ-continuous.

The proof is a fruit of the combination of the weak convergence theory for ℓ^∞-valued continuous martingales developed by Nishiyama (2000) and a theorem for local time of ergodic diffusion processes given by van Zanten (2003)

Moreover by the continuous mapping theorem, we have the following

Corollary. It holds that

$$\sup_{x \in [-\infty, +\infty]} |V^T(x)| \to^d \sup_{x \in [-\infty, +\infty]} |G(x)| =^d \sup_{t \in [0, \Sigma^2_{S_0,\sigma}]} |B_t| =^d \Sigma_{S_0,\sigma} \sup_{t \in [0,1]} |B_t|,$$

Consistency of the test

In order to have consistency we have to study the asymptotic behavior of $\mathcal{C}^T$ under a fixed alternative. Denote by $\mathcal{S}$ the class of function S which satisfy **A1**, **A2** and **A3** and

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

We have the following result

Theorem. Assume **A1**, **A2** and **A3** for (S, σ). Under $H_1 : S \in \mathcal{S}$, if $\widehat{\Sigma}^T$ is bounded in probability, then

$$\mathcal{C}^T = \frac{\sup_{x\in[-\infty,\infty]} |V^T(x)|}{\widehat{\Sigma}^T} \neq O_P(1)$$

Proof of the result

We can write

$$\sup_{x\in[-\infty,\infty]} |V_T(x)| \geq \sqrt{T} \sup_{x\in[-\infty,\infty]} |A_T(x)| - \sup_{x\in[-\infty,\infty]} |V_T^1(x)|,$$

where

$$V_T^1(x) = \frac{1}{\sqrt{T}} \int_0^T \mathbf{1}_{(-\infty,x]}(X_t)(\mathrm{d}X_t - S(X_t)\mathrm{d}t)$$

weakly converges to the corresponding Gaussian process so the limit process is tight.

Moreover, on the other hand, chosen x_S and defined

$$A_T(x) = \frac{1}{T} \int_0^T \mathbf{1}_{(-\infty,x]}(X_t)(S(X_t) - S_0(X_t))\mathrm{d}t$$

we have

$$\frac{1}{T} \int_0^T \mathbf{1}_{(-\infty,x_S]}(X_t)(S(X_t) - S_0(X_t))\mathrm{d}t \to^p \int_{-\infty}^{x_S} (S(z) - S_0(z)) f_{S,\sigma}(z)\ dz \neq 0$$

Goodness of fit test for diffusion: discrete time observations

The test procedure presented is based on the continuous observation of the process on $[0, T]$.

A new test procedure is based on discrete time observation which is more realistic in applications.

Precisely the sampling scheme is the following (we denote $\operatorname{Log} m = \log(1+m)$)

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

This condition implies $h_n \to 0$ so we can assume $h_n \le 1$ without loss of generality.

Goodness of fit test for diffusion: test statistic

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 time 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 time observation.

Some more conditions on diffusion model

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

Now, we introduce an array of constants in the state space

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

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

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

For example, one may consider $x_k^n = -n + (k/n)$ with $k = 1, 2, ..., 2n^2$.

The functions ψ_k^n on $(-\infty, \infty)$ are defined for every $k = 1, 2, ..., m(n)$, ψ_k^n as

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

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

Some more conditions on diffusion model

These functions satisfie the following properties:

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

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

We make the following condition on the sequence of positive constants b_n

A5. In addition to $h_n = O(n^{-2/3}(\mathrm{Log}\, n)^{1/3})$, which implies $nh_n^2 \to 0$, we assume the following:

(i) $b_n^{-2} h_n \cdot \mathrm{Log}\, n \cdot \mathrm{Log}\, m(n) \to 0$;

(ii) $b_n \mathrm{Log}\, m(n) \to 0$.

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

Asymptotically distribution free test

The main result is the following (see Masuda Negri and Nishiyama, 2008)

Theorem. Assume **A1 – A5** for (S_0, σ). 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 and $\widehat{\Sigma}^n$ is a consistent estimator for $\Sigma_{S_0,\sigma}$.

A consistent estimator for $\Sigma_{S_0,\sigma}$ is given by the following

Lemma. 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 $\Sigma_{S,\sigma}$.

Proof of the result

The result is based on the following lemmas

Lemma 1. 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 2. $\sup_{x\in[-\infty,\infty]} |U^n(x) - V^n(x)| \to^p 0$

Lemma 3. $V^n \to^d G \quad \text{in} \quad C_\rho([-\infty,\infty])$

Theorem. $U^n \to^d G \quad \text{in} \quad \ell^\infty([-\infty,\infty])$

Finally, by the continuous mapping theorem, we have the following.

$$\sup_{x\in[-\infty,\infty]} |U^n(x)| \to^d \sup_{t\in[0,\Sigma^2]} |B_t| =^d \Sigma \sup_{t\in[0,1]} |B_t|$$

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.

Proof of Lemmas

Proof of Lemma 1 and Lemma 3 follow from the theory of random fields generated by continuous martingales developed by Nishiyama (1999, 2000)

Proof of Lemma 2 follows by the *exponential inequality* for continuous martingales, from the *maximal inequality* for general random variables and from the following

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

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

To prove it, it is sufficient to show that

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

Consistency of the test

We denote by $\mathcal{S}$ the class of functions S which satisfies **A1 – A4** and

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

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 prove that our test is consistent.

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

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

Proof of the result

Fix $S \in \mathcal{S}$. We can write $U^n = U_S^n + U_\Delta^n$ where $U_S^n(-\infty) = U_\Delta^n(-\infty) = 0$ and, for $x \in (x_{k-1}^n, x_k]$, $1 \leq k \leq m(n) + 1$,

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

and

$$U_\Delta^n(x) = \frac{1}{\sqrt{t_n^n}} \sum_{i=1}^n \psi_k^n(X_{t_{i-1}^n})(S(X_{t_{i-1}^n}) - S_0(X_{t_{i-1}^n}))|t_i^n - t_{i-1}^n|.$$

Now we have

$$\sup_{x\in[-\infty,\infty]} |U^n(x)| \geq \sup_{x\in[-\infty,\infty]} |U_\Delta^n(x)| - \sup_{x\in[-\infty,\infty]} |U_S^n(x)|.$$

The random field U_S^n 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 for the first we have

Lemma. Choose $x_S \in (-\infty, \infty]$. Then, $|U_\Delta^n(x_S)| \neq O_P(1)$.

A simulation study for ergodic diffusion process

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:

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

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

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

(a) asymptotic behavior of test statistics with $S_0(x) = -2x$;

(b) asymptotic behavior of test statistics with $S_0(x) = -4x$.

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

if we denote our test statistics with

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

and

(a) $\mathcal{T}_0^n := \mathcal{T}^n$ with $S_0(x) = -2x$;

(b) $\mathcal{T}_1^n := \mathcal{T}^n$ with $S_0(x) = -4x$.

what we expect is

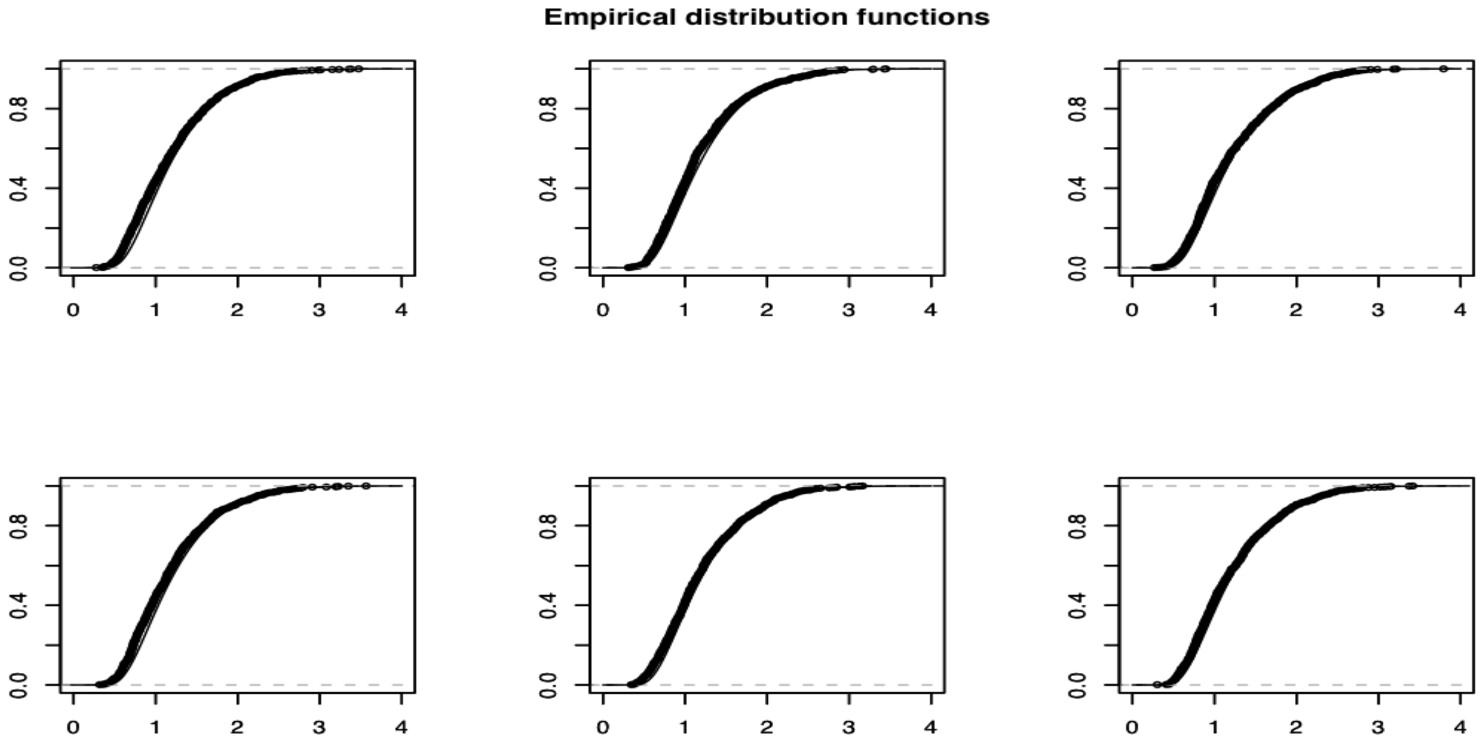
- $P(\mathcal{T}_0^n > 2.24)$ should tend to 0.05 in (a);
- $P(\mathcal{T}_1^n > 2.24)$ should tend to 1.0 in (b).

For several different terminal time t_n^n and sampling frequency h_n, we simulate 1000 independent copies of a discrete sample trajectory of (2) 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.

	$t_n^n = 10$		$t_n^n = 20$		$t_n^n = 50$	
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
	$(n = 200)$		$(n = 400)$		$(n = 1000)$	



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). In each figure the two plots are almost piled up.

Finally, it is conjectured that the same result would hold also for $\widetilde{U}^n$ given by

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

The next Table shows results of experimental trials for this test statistics. 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.

	$t_n^n = 10$		$t_n^n = 20$		$t_n^n = 50$	
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
	$(n = 200)$		$(n = 400)$		$(n = 1000)$	

It seems 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.

Tick time sample scheme

For every $T > 0$, the process $X_T = \{X_t : t \in [0, T]\}$ is observed at random times $0 = \tau_0^T < \tau_1^T < \cdots < \tau_{n(T)}^T < \tau_{n(T)+1}^T = T$ where $N(T) = \sup\{i : \tau_i^T < T\}$, $\tau_0^T = 0$

$$\tau_1^T = \inf\{t > 0 : X_t = a_p^T \quad \text{for some} \quad p\}$$

and

$$\tau_i^T = \inf\{t > \tau_{i-1}^T : X_t = a_{p-1}^T \text{ or } X_t = a_{p+1}^T \text{ if } X_{\tau_{i-1}^T} = a_p^T\} \quad i \geq 2$$

Here $\cup_p(a_p^T, a_{p+1}^T]$ is a countable partition of $(-\infty, +\infty)$ and we assume that $\inf_p |a_{p+1}^T - a_p^T| > 0$

We can prove that $N(T) < \infty$ almost surely.

We suppose that $h_T = o(T^{-1/2})$ as $T \to \infty$ where $h_T = \sup_p |a_{p+1}^T - a_p^T|$

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

$$U^T(x) = \frac{1}{\sqrt{T}} \sum_{i=1}^{n} \psi_k^T(X_{\tau_{i-1}^T})[X_{\tau_i^T} - X_{\tau_{i-1}^T} - S_0(X_{\tau_{i-1}^T})|\tau_i^T - \tau_{i-1}^T|]$$

for $x \in (x_{k-1}^T, x_k^T],\ 1 \leq k \leq m(T) + 1$.

It turns out that under $H_0 : S = S_0$ it holds that

$$\mathcal{D}^T := \frac{\sup_{x\in[-\infty,\infty]} |U^T(x)|}{\widehat{\Sigma}_2^T} \to^d \sup_{t\in[0,1]} |B_t|,$$

where

$$\widehat{\Sigma}_2^T = \sqrt{\frac{1}{T} \sum_{i=1}^{N(T)+1} |X_{\tau_i^T} - X_{\tau_{i-1}^T}|^2}$$

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

References

- Kutoyants, (2004). *Statistical Inference for Ergodic Diffusion Processes.* Springer.
- Kutoyants, (2009). Goodness of fit testing for ergodic Diffusion Processes. Preprint
- Masuda, Negri and Nishiyama (2008). Goodness of fit test for ergodic diffusions: an innovation martingale approach. (submitted).
- Negri, (1998). Stationary distribution function estimation for ergodic diffusion process. *Stat. Inf. for Stoch. Processes*.
- Nishiyama (1999). A maximal inequality for continuous martingales and M-estimation in a Gaussian white noise model *Ann. Statist.*
- Nishiyama, (2000). *Entropy Methods for Martingales*, CWI Tract **128**, Amsterdam.
- Negri and Nishiyama, (2008). Goodness of fit test for ergodic diffusion processes. *Ann. Inst. Statist. Math.*
- Negri and Nishiyama, (2009). Goodness of fit test for ergodic diffusion by tick time sampling scheme. (work in progress)
- Van Zanten, (2003). On empirical processes for ergodic diffusions and rates of convergence of M-estimators, *Scandinavian Journal of Statistics*.