C. R. Acad. Sci. Paris, Ser. I 349 (2011) 897–900

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Probability Theory/Statistics

On the goodness-of-fit testing for a switching diffusion process

Test d'ajustement d'un processus de diffusion à changement de régime

Anis Gassem

Laboratoire de statistique et processus, Université du Maine, 72085 Le Mans cedex 9, France

ARTICLE INFO

Article history:
Received 18 April 2008
Accepted 8 June 2011
Available online 15 August 2011

Presented by Paul Deheuvels

ABSTRACT

We study the asymptotic behavior of the Cramér–von Mises type statistic in the goodness-of-fit hypotheses testing problem for ergodic diffusion processes. The basic (simple) hypothesis is defined by the stochastic differential equation with sign-type trend coefficient and known diffusion coefficient. It is shown that the limit distribution of the proposed test statistic (under hypothesis) is defined by the integral type functional of continuous Gaussian process. We provide the Karhunen–Loève expansion of the corresponding limiting process and show that the eigenfunctions in this expansion are expressed in terms of Bessel functions. This representation for the limit statistic allows us to approximate the threshold.

RÉSUMÉ

Nous considérons un test d'ajustement de Cramér–von Mises d'un processus de diffusion dont la dérive est de type signe. Il est montré que la distribution limite du test statistique proposé est définie par une fonctionnelle de type intégrale d'un processus gaussien continu. Nous obtenons la décomposition de Karhunen–Loève du processus limite correspondant et nous montrons que les fonctions propres de cette décompositon ont des expressions en termes de fonctions de Bessel. La représentation de la statistique limite nous permet de trouver le seuil.

1. Introduction

The problem of goodness-of-fit testing is one of the central themes of statistical theory and practice because it is always important to verify the correspondence between the mathematical model and the observed real data. The classical approaches to this hypothesis testing problem can be found in many books (see, e.g., Durbin [4], Greenwood and Nikulin [6] or Lehmann and Romano [10]). The most known tests are: Pearson's goodness-of-fit Chi-Squared test, Kolmogorov–Smirnov and Cramér–von Mises tests. We recall the construction of the Cramér–von Mises test. Suppose that we observe n independent identically distributed random variables $(X_1, \ldots, X_n) = X^n$ with continuous distribution function $F(x)$ and the basic hypothesis is simple: $F(x) = F_0(x)$. Then we can introduce the Cramér–von Mises statistic,

E-mail address: anisgassem@yahoo.fr.

1631-073X/$ – see front matter
doi:10.1016/j.crma.2011.06.010

$$W_n^2 = n\int_{-\infty}^{\infty}\left[\hat F_n(x) - F_0(x)\right]^2 \mathrm{d}F_0(x), \quad \hat F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i<x\}},$$

where $\hat F_n(x)$ is the empirical distribution function, and after the transformation $t = F_0(x)$, we get the standard form of the statistic W_n^2 and its limit distribution W^2 (under $\mathcal{H}_0$)

$$W_n^2 = n\int_0^1 \left(\tilde F_n(t) - t\right)^2 \mathrm{d}t = \int_0^1 \beta_n^2(t)\,\mathrm{d}t \quad\Longrightarrow\quad W^2 = \int_0^1 \beta^2(t)\,\mathrm{d}t.$$

Here $\beta(t)$ is Brownian bridge (Gaussian process with $\mathbf{E}\beta(t) = 0$ and $K(t,s) = \mathbf{E}\beta(t)\beta(s) = t\wedge s - ts$).

The hypothesis $\mathcal{H}_0$ is accepted, if $W_n^2 < c_{n,\alpha}$ and rejected, if $W_n^2 \geqslant c_{n,\alpha}$, where the constant $c_{n,\alpha}$ is solution of the equation $\mathbf{P}(W_n^2 \geqslant c_{n,\alpha}) = \alpha$. The constant $c_{n,\alpha}$ can be approximated by its limit ($n\to\infty$) value c_α, solution of the equation $\mathbf{P}(W^2 \geqslant c_\alpha) = \alpha$. It is possible to expand $\beta(\cdot)$ into the Karhunen–Loève series (see Ash and Gardner [1], pp. 38–43)

$$\beta(t) = \sum_{k=1}^{\infty}\frac{\sqrt{2}}{\pi k}\xi_k \sin(\pi k t), \quad t\in[0,1], \qquad W^2 \overset{d}{=} \sum_{k=1}^{\infty}\frac{\xi_k^2}{(\pi k)^2}, \quad \xi_k \quad \text{i.i.d.} \sim \mathcal{N}(0,1).$$

These relations are used to approximate the value c_α. More about Karhunen–Loève theory and the use of Karhunen–Loève expansion in the problems of goodness-of-fit testing can be found in Martynov [11] and Deheuvels and Martynov [3].

In this Note, we study the similar problem of goodness-of-fit testing, when the basic model is an ergodic diffusion process and the corresponding statistic is of Cramér–von Mises type.

Let $X^T = \{X_t, 0\leqslant t\leqslant T\}$ be an observed trajectory of continuous time ergodic diffusion process

$$\mathrm{d}X_t = S(X_t)\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0, \quad 0\leqslant t\leqslant T, \tag{1}$$

where X_0 is its initial value, independent of the Wiener process $\{W_t, t\geqslant 0\}$. The trend coefficient $S(\cdot)$ is supposed to be unknown to the observer and he has to verify if $S(x) = S_0(x)$, where $S_0(x)$ is some known function. The Cramér–von Mises type statistics, based on the observation X^T solution of (1) can be:

$$V_T^2 = T\int_{-\infty}^{\infty}\left(\hat F_T(x) - F_{S_0}(x)\right)^2 \mathrm{d}x, \qquad W_T^2 = T\int_{-\infty}^{\infty}\left(\hat f_T(x) - f_{S_0}(x)\right)^2 \mathrm{d}x,$$

where $\hat F_T(x)$ and $\hat f_T(x)$ are some estimates of the distribution function $F_{S_0}(x)$ of the invariant density $f_{S_0}(x)$ respectively. These statistics converge in distribution (under the null hypothesis) to quadratic functionals of Gaussian processes (see Kutoyants [9]). However, due to the structure of the covariance of these processes, the choice of the thresholds c_α for the tests $\varphi_T = 1_{\{V_T^2 > c_\alpha\}}, \phi_T = 1_{\{W_T^2 > c_\alpha\}}$ is much more complicate (see Dachian and Kutoyants [2]).

In the present work we are interested by the case when the trend coefficient $S_0(\cdot)$ of the process (1) under hypothesis takes just two values $+1$ and -1 (simple switching $S_0(x) = -\mathrm{sgn}(x)$), i.e., is discontinuous function (see Kutoyants [8], for the further details). We study the limit distribution of the corresponding Cramér–von Mises type statistic based on the empirical density estimator (local time estimator) of the invariant density. The main tool is the K-L expansion of the limit Gaussian process. Finally, using this expansion and numerical simulation we calculate the threshold of this test.

2. Simple switching diffusion

Let us introduce the simple switching diffusion process

$$\mathrm{d}X_t = -\mathrm{sgn}(X_t)\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0, \quad 0\leqslant t\leqslant T, \tag{2}$$

i.e., $S_0(x) = -\mathrm{sgn}(x)$. It is easy to see that the conditions of the existence of the solution and ergodicity (see, e.g., conditions $\mathcal{ES}$ and $\mathcal{A}_0$ in Kutoyants in [9]), are fulfilled and (2) is an ergodic diffusion process with the stationary density $f_{S_0}(x) = e^{-2|x|}$, for $x\in\mathbb{R}$.

We observe a trajectory X^T of (1) and we have to test the following hypotheses:

$$\mathcal{H}_0\colon \quad S(\cdot) = S_0(\cdot), \qquad \mathcal{H}_1\colon \quad S(\cdot) \neq S_0(\cdot).$$

To test this null hypothesis we use the following Cramér–von Mises type statistic:

$$W_T^2 = \int_{-\infty}^{\infty}\zeta_T^2(x)\,\mathrm{d}x = T\int_{-\infty}^{\infty}\left[f_T^\circ(x) - f_{S_0}(x)\right]^2 \mathrm{d}x, \quad f_T^\circ(x) = \frac{2\Lambda_T(x)}{T},$$

where $\Lambda_T(x)$ is the local time of the diffusion process and $f_T^\circ(x)$ is the local time estimator of the invariant density (see, Kutoyants in [9]).

Fix a number $\alpha \in (0,1)$ and define the class $\mathcal{K}'_\alpha$ of tests of asymptotic level $1-\alpha$ as follows:

$$\mathcal{K}'_\alpha = \Big\{\phi_T\colon \limsup_{T\to\infty} \mathbf{E}_{S_0}\phi_T\big(X^T\big) \leqslant \alpha\Big\}.$$

Let us introduce the Gaussian process $(\zeta_f(x),\ x\in\mathbb{R})$, with zero mean and covariance function

$$R_f(x,y) = 2\big(1_{\{x\vee y<0\}}e^{2(x\wedge y)} + 1_{\{x\wedge y>0\}}e^{-2(x\vee y)}\big) - \big(2\big(|x|+|y|\big)+\operatorname{sgn}(xy)\big)e^{-2(|x|+|y|)}.$$

Denote by d_α the value defined by the equation

$$\mathbf{P}\Bigg(\int_{-\infty}^{\infty} \zeta_f^2(x)\,\mathrm{d}x > d_\alpha\Bigg) = \mathbf{P}\big(W^2 > d_\alpha\big) = \alpha.$$

Theorem 2.1. *The Cramér–von Mises type test $\Phi_T(X^T) = 1_{\{W_T^2>d_\alpha\}}$ belongs to $\mathcal{K}'_\alpha$.*

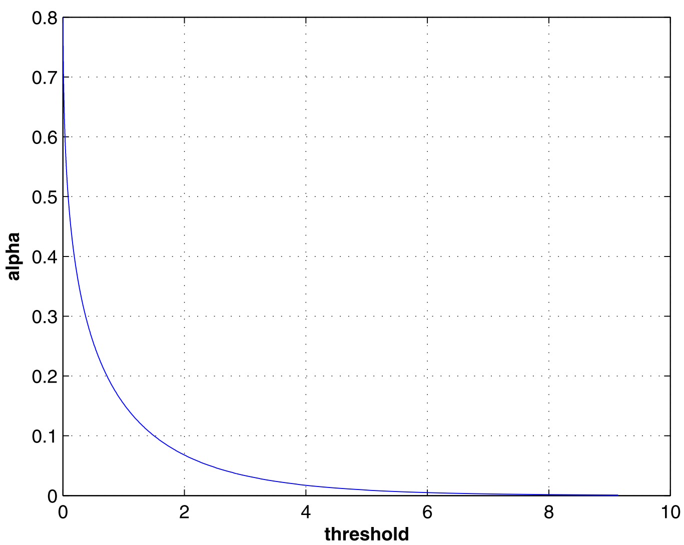
Proof. The main idea of the proof is that under $\mathcal{H}_0$ the distribution of W_T^2 converge to the distribution of W^2. In fact by using Theorem A22, in Ibragimov–Khasminskii [7], we show that for any constant $L\geqslant 0$, the distribution of $\int_{|x|\leqslant L}\zeta_T^2(x)\,\mathrm{d}x$ convergence to the distribution of $\int_{|x|\leqslant L}\zeta_f^2(x)\,\mathrm{d}x$. Then we show that $\int_{|x|>L}\zeta_T^2(x)\,\mathrm{d}x$ tends to zero as $L\to+\infty$. For more details see [5]. □

Let $J_n(\cdot)$ and $Y_n(\cdot)$ be the Bessel functions of first and second kind of order n, respectively, with $n=0,1,2,\ldots.$

Proposition 2.2. *The Gaussian process $(\zeta_f(x),\ x\in\mathbb{R})$ has a Karhunen–Loève expansion given by*

$$\zeta_f(x) = \sum_{n=1}^{\infty}\frac{2}{\nu_{2,n}}\xi_n\left\{\frac{\alpha(\nu_{2,n})e^{-|x|}J_1(\nu_{2,n}e^{-|x|}) - J_2(\nu_{2,n})(e^{-|x|}Y_1(\nu_{2,n}e^{-|x|}) + \frac{2}{\pi\nu_{2,n}})}{J_1(\nu_{2,n})\alpha(\nu_{2,n}) - Y_1(\nu_{2,n})J_2(\nu_{2,n})}\right\}$$
$$-\sum_{n=1}^{\infty}\frac{2}{\nu_{1,n}}\xi_n^*\operatorname{sgn}(x)e^{-|x|}\frac{J_1(\nu_{1,n}e^{-|x|})}{J_0(\nu_{1,n})},$$

where $\{\xi_n, n\geqslant 1\}$, $\{\xi_n^*, n\geqslant 1\}$ denote two independent sequences of i.i.d. $\mathcal{N}(0,1)$ random variables and $\nu_{1,n}$, $\nu_{2,n}$, $n=1,2,\ldots,$ are respectively the positive zeros of $J_1(\nu_{1,n})$ and $f(\nu_{2,n})$ defined by the equation $f(\nu_n) = J_2(\nu_n)\beta(\nu_n) - (J_0(\nu_n)-1)\alpha(\nu_n)$, with $\beta(\nu_n) = Y_0(\nu_n) - \frac{2}{\pi}(\ln(\nu_n/2)+\gamma)$, $\alpha(\nu_n) = Y_2(\nu_n) + \frac{4}{\pi\nu_n^2}$ and $\gamma = 0.577215\ldots$ the Euler's constant.



Values of some quantiles of the random variable W^2.

alpha	threshold
0.10	1.5008
0.05	2.4202
0.025	3.4332
0.010	4.8678
0.005	6.0036

Fig. 1. Choice of threshold.

Proof. The detailed proof is given in [5].

Note that the random variable W^2 is a weighted sum of independent χ_1^2 components. We make the truncation of this sum keeping 10^7 terms and using the numerical simulation of the corresponding Gaussian random variables we obtain values for the threshold d_α as shown in Fig. 1. $\square$

Acknowledgements

The author is very grateful to Professor Marina Kleptsyna for essential help in correcting the first version of this result, and to Professor Yury Kutoyants for suggesting this problem, and fruitful discussions.

References

[1] R.B. Ash, M.F. Gardner, Topics in Stochastic Processes, Academic Press, New York, 1975.
[2] S. Dachian, Yu.A. Kutoyants, On the goodness-of-fit testing for some continuous time processes, in: F. Vonta, et al. (Eds.), Statistical Models and Methods for Biomedical and Technical Systems, Birkhäuser, Boston, 2007, pp. 395–413.
[3] P. Deheuvels, G.V. Martynov, Cramér–von Mises-type tests with applications to tests of independence for multivariate extreme-value distributions, Commun. Statist. Theory Meth. 25 (4) (1996) 871–908.
[4] J. Durbin, Distribution Theory for Tests Based on the Sample Distribution Function, SIAM, Philadelphia, 1973.
[5] A. Gassem, Goodness-of-fit test for switching diffusion, Prépublication no. 08-7, Laboratoire de Statistique et Processus, Université du Maine, 2008, http://www.univ-lemans.fr/sciences/statist/download/Gassem/article_anis.pdf.
[6] P.E. Greenwood, M. Nikulin, A Guide to Chi-Squared Testing, John Wiley and Sons, New York, 1996.
[7] I.A. Ibragimov, R.Z. Khasminskii, Statistical Estimation: Asymptotic Theory, Springer-Verlag, New York, 1981.
[8] Yu.A. Kutoyants, On parameter estimation for switching ergodic diffusion processes, C. R. Acad. Sci. Paris, Ser. I 330 (2000) 925–930.
[9] Yu.A. Kutoyants, Statistical Inference for Ergodic Diffusion Processes, Springer-Verlag, London, 2004.
[10] E.L. Lehmann, J.P. Romano, Testing Statistical Hypotheses, 3rd edition, Springer-Verlag, New York, 2005.
[11] G.V. Martynov, Statistical tests based on empirical processes and related question, J. Soviet. Math. 61 (1992) 2195–2271.