Discussiones Mathematicae
Probability and Statistics 30 (2010) 117–122

SAMPLE PARTITIONING ESTIMATION FOR ERGODIC DIFFUSIONS: APPLICATION TO ORNSTEIN-UHLENBECK DIFFUSION

Luís Ramos

Department of Mathematics
New University of Lisbon, Portugal

e-mail: lpcr@fct.unl.pt

Abstract

When a diffusion is ergodic its transition density converges to its invariant density, see Durrett (1998). This convergence enabled us to introduce a sample partitioning technique that gives in each sub-sample, maximum likelihood estimators. The averages of these being a natural choice as estimators. To compare our estimators with the optimal we obtained from martingale estimating functions, see Sørensen (1998), we used the Ornstein-Uhlenbeck process for which exact simulations can be carried out.

Keywords: ergodic diffusions; martingale estimating functions; transition and invariant densities; maximum likelihood estimators.

2000 Mathematics Subject Classification: 60J60, 60H35, 62F10.

1. Introduction

The remainder of this article is organized as follows. In Section 2 we intruduce the definition of diffusions, followed by a definition of ergodic propriety and a few results for that kind of diffusions in Section 3. Our new results are described in Section 4. In Section 5 we present an application to the Ornstein-Uhlenbeck diffusion. Finally, Section 6 gives the conclusions.

2. Diffusions

A Diffusion is, see Iacus (2008, pg 33), a time-homogeneous stochastic differential equation

$$dX_t = b(X_t, \boldsymbol{\theta})dt + \sigma(X_t, \boldsymbol{\theta})dB_t, \tag{2.1}$$

where $\boldsymbol{\theta} \in \Theta \subset \Re^p$ is a multidimentional parameter and $\{B_t\}_{t\geq 0}$ is the Brownian motion or Wiener process. The functions $b : \Re \times \Theta \rightarrow \Re$ and $\sigma : \Re \times \Theta \rightarrow]0, +\infty[$ are known and such that the solution of (2.1) exist.

The function b is known as the drift coefficient of X, the function σ is known as the diffusion coefficient of X.

An example of a diffusion is the Ornstein-Uhlenbeck process given by the equation

$$dX_t = \theta X_t dt + \sigma dB_t.$$

3. Ergodicity

A diffusion is ergodic when there is a stochastic limit for the Hessian matrix of the score when the sample size tends to infinite, the observations being equally spaced, see Küchler & Sørensen (1997, pgs 123–127). We have

Theorem 3.1. *The transition densities of ergodic diffusion tend to the corresponding invariant densities when the time lag tends to infinite. See Durrett* (1996).

With $E =]l, r[$ the range of variation of X_t, $t > 0$, and $l < x_0 < r$, let the scale measure, be

$$s(x, \boldsymbol{\theta}) = \exp\left[-2\int_{x_0}^{x} \frac{b(y, \boldsymbol{\theta})}{\sigma^2(y, \boldsymbol{\theta})}dy\right] \quad ; \; x \in]l; r[\tag{3.1}$$

and speed measure,

$$m(x, \boldsymbol{\theta}) = \frac{1}{s(x, \boldsymbol{\theta})\sigma^2(x, \boldsymbol{\theta})} \quad ; \; x \in]l; r[. \tag{3.2}$$

Thus, we have

Theorem 3.2 (Sørensen (1998)). *The diffusion is ergodic whenever, for every $\boldsymbol{\theta} \in \Theta$,*

$$\int_{x_0}^{r} s(x,\boldsymbol{\theta})dx = \int_{l}^{x_0} s(x,\boldsymbol{\theta})dx = +\infty \tag{3.3}$$

and

$$M(\boldsymbol{\theta}) = \int_{l}^{r} m(x,\boldsymbol{\theta})dx < +\infty. \tag{3.4}$$

The invariant density being

$$f_{\boldsymbol{\theta}}(x) = \frac{m(x,\boldsymbol{\theta})}{M(\boldsymbol{\theta})} \quad ; \; x \in]l;r[. \tag{3.5}$$

4. LIMIT INDEPENDENCE AND SUB-SAMPLING

When the transition density converges to the invariant density there is $\overline{\triangle}$ such that for $t > \overline{\triangle}$, we can assume X_t to have the invariant density approximately. Thus, observations taken at times $t_1, \dots t_n$ with $t_1 > \overline{\triangle}$ and $t_j - t_{j-1} > \overline{\triangle}$, $j = 2, \dots k$ may be treated as *i.i.d.* with the invariant density. Since $\overline{\triangle}$ is not known we must obtain a lower bound for it. Given the observations $X_1, \dots X_n$ with $n = k \times m$, taken at times $t_1, \dots, t_n$, we can use Friedman test to check

$$H_0 : m\triangle > \overline{\triangle}.$$

So, when this hypothesis holds, the matched sub-samples corresponding of the lines of

$$\begin{array}{ccccc} X_1 & \cdots & X_j & \cdots & X_m \\ \vdots & & \vdots & & \vdots \\ X_{(i-1)m+1} & \cdots & X_{(i-1)m+j} & \cdots & X_{im} \\ \vdots & & \vdots & & \vdots \\ X_{(k-1)m+1} & \cdots & X_{(k-1)m+j} & \cdots & X_{km} \end{array}$$

will have the same distribution.

When H_0 is not rejected we can treat the observations in each column as being *i.i.d.* with the invariant density. Thus, from each column we can obtain a maximum likelihood estimator for $\boldsymbol{\theta}_j$, $j = 1, \dots, m$. Afterwards, we take the average.

5. Application to the Ornstein-Uhlenbeck diffusion

The invariant density for the Ornstein-Uhlenbeck diffusion is the normal density with mean 0 and variance $-\frac{\sigma^2}{2\theta}$, $\theta < 0$. Since this density has one parameter we take $\sigma = 1$ and estimate θ.

We, thus, obtained

$$\widehat{\theta} = \frac{1}{m} \sum_{j=1}^{m} \widehat{\theta}_j = \frac{1}{m} \sum_{j=1}^{m} \left[-\frac{2}{k} \sum_{i=1}^{k} X^2_{(i-1)m+j} \right]^{-1} . \tag{5.1}$$

We now compare our estimator with

$$\widetilde{\theta}_n = \frac{1}{\triangle} \ln \left(\frac{\sum_{i=1}^{n} X_{i-1} X_i}{\sum_{i=1}^{n} X^2_{i-1}} \right) \quad \textit{with} \quad \sum_{i=1}^{n} X_{i-1} X_i > 0$$

which, see Sørensen (1998), is derived using martingale estimating functions.

We used the transition density, which is known, to simulate trajectories of Ornstein-Uhlenbeck diffusion considering a few values of θ and we obtained the followed tables with mean, variance and the mean square error estimated values of two compared estimators. We have fixed $\sigma = 1$ and the time lag, $\triangle = 1$.

Table 1. $k = 10$, $m = 5$.

θ	Mean($\widehat{\theta}$)	Mean ($\widetilde{\theta}_n$)	Var($\widehat{\theta}$)	Var ($\widetilde{\theta}_n$)	MSE($\widehat{\theta}$)	MSE ($\widetilde{\theta}_n$)
− 0.001	− 0.1348	− 0.0317	0.0215	0.003	0.0392	0.0039
− 0.01	−0.1885	− 0.0488	0.0519	0.0044	0.0833	0.0058
− 0.1	− 0.389	− 0.1422	0.0898	0.0093	0.1724	0.011
− 1	− 3.8088	− 1.0775	15.5239	0.1641	23.258	0.1685
− 5	− 15.7905	− 2.5885	42.628	0.784	158.6364	6.5822
− 10	− 33.9045	− 2.5827	221.2223	1.2305	790.437	56.2212

Table 2. $k = 20$, $m = 20$.

θ	Mean($\widehat{\theta}$)	Mean ($\widetilde{\theta}_n$)	Var($\widehat{\theta}$)	Var ($\widetilde{\theta}_n$)	MSE($\widehat{\theta}$)	MSE ($\widetilde{\theta}_n$)
− 0.001	− 0.0078	− 0.0059	5e− 06	1e− 04	1e− 04	1e− 04
− 0.01	− 0.018	− 0.0163	2e− 04	2e− 04	2e− 04	2e− 04
− 0.1	− 0.1148	− 0.1067	8e− 04	7e− 04	0.001	7e− 04
− 1	− 1.1099	− 1.011	0.0084	0.0159	0.0204	0.0158
− 5	− 5.5128	− 3.3625	0.1963	0.9667	0.4573	3.6277
− 10	− 11.1713	− 3.9323	0.6569	1.482	2.0221	38.2718

As we expected the results for our estimator are better than Sørensen estimator when we consider a decomposition ($k = 20$, $m = 20$) which means we can assume the assumption of independence of the observations in each column and, consequently, observations having the approximately invariant distribution. The rate of convergence of transition density to invariant density for the Ornstein-Uhlenbeck diffusion depends of θ, and is faster for large absolute values of θ, so our results are better for these values of θ.

6. Final remarks

This paper points towards the use of the invariant density while converging out inference on ergodic diffusions. Another possible application of these will be the use of a Kolmogorov-Smirnov test to check the model. Thus, once the parameters are estimated we can apply such a test.

References

[1] B.M. Bibby and M. Sørensen, *Martingale estimation functions for discretely observed diffusion processes*, Bernoulli **1** (1995), 17–39.

[2] U. Küchler and M. Sørensen, *Exponential Families of Stochastic Processes*, Springer-Verlag 1997.

[3] S. Iacus, *Simulation and Inference for Stochastic Differential*, Equations with R Examples, Springer 2008.

[4] J.T. Mexia and G.C. Dias, Statistical Inference for Discretely Observed Diffusions of Diffusions, VII-Congresso da SPE, 14–16 Outubro, Osir 1999.

[5] B. Øksendal, *Stochastic Differential Equations*, An Introduction, Fifth Edition, Springer-Verlag 1998.

[6] M. Sørensen, Lecture Notes on "Statistical Inference for Discretely Observed Diffusions", Berlin Graduiertenkolleg 1998.

Received 18 January 2010