Modeling an estimate for a parameter of a random diffusion process

T V Avdeeva[1,2] and L M Illicheva[2]

[1] Department of Mathematical Physics, National Technical University of Ukraine "Igor Sikorskiy Kiyv Politechnical Institute", 37 Peremogu ave., Kyiv 03056, Ukraine
[2] Department of Applied Mathematics, National Aviation University, 1 Liubomyra Huzara ave., Kyiv 03058, Ukraine

[3] E-mail: avdeeva.tetyana@gmail.com

Abstract. Various special cases of estimation of unknown parameters of diffusion processes are considered. Three different martingal estimating functions based on discrete-time observations of a diffusion process are considered. One is the discretized continuous-time score function adjusted by its compensator. The other two emerge naturally if optimality properties of the first are considered. In addition to the general normal correlation theorem in the discrete case, an overview of research on this topic is given. The following approaches are considered: martingale parameter estimation for discretly observed diffusion processes; explicit expression for the parameter of the diffusion process; optimal linear non-stationary filtering (Kalman-Bucy method)

1. Introduction

In the study of physical phenomena, one often deals with case-dependent quantities that change over time. A probabilistic process is used to describe such phenomena. If we have a main probability space; T - the set of real numbers, then the probabilistic process is called a random variable $x_t(\omega)$, depending on t $\in T$. Diffusion processes began to be considered as early as the 19th century.

Historically, standard Brownian motion was first studied - the random movement of pollen particles suspended in water. The exact definition of Brownian motion involves the use of a measure in the space of trajectories, and only then the Brownian motion received a solid mathematical foundation. In 1923, N. Wiener introduced the concept of a process that now bears his name.

The Wiener process describes the chaotic motion of a microscopic particle in a medium (liquid or gas) under the influence of collisions with molecules of this medium. There is also a model of thermal noise in the conductor due to the chaotic movement of electric charge carriers.

The parameters of diffusion processes have a significant effect on their properties. The work is devoted to the consideration of some examples of estimating the parameters of diffusion processes.

2. Martingale parameter estimation for discretly observed diffusion processes

Considered one-dimensional diffusion processes defined by the following class of stochastic differential equations:

$$dX_t = b(X_t;\Theta)dt + \sigma(X_t;\Theta)dW_t \,, X_0 = x_0 \qquad (1)$$

Here, the drift and the diffusion coefficient do not depend on the time t; the function σ is assumed to be positive; the functions b and σ are supposed to be known and twice continuously differentiable with respect to both arguments; assume that (1) has a unique solution for all Θ in some open subset of the real time. The parameter Θ is to be estimated from discrete equidistant observations of $\{X_t\}$: $X_{\Delta}, X_{2\Delta}, \dots X_{n\Delta}$. Filtration is defined as $\Im_i = \sigma\ (X_{\Delta}, \dots, X_{i\Delta}), i = 1,2, \dots$
Parameter estimation, concerning diffusion process, based on the likelihood function L. The likelihood function for discrete observations is a product of transition densities, so we are looking for good approximation for such function. Approximation of continuous-time likelihood function is considered by replacing Lebesque intergrals and Ito integrals by Rieman-Ito sums. The estimator $\overline{L}$ obtained in this way can be strongly biased, therefore consider martingal estimating function from $\overline{L}$. Two functions are considered as such an assessment for Θ [1]:

$$\widetilde{G_n}(\Theta) = \sum_{i=1}^{n} \frac{\dot{b}(X_{(i-1)\Delta};\Theta)}{\sigma^2(X_{(i-1)\Delta};\Theta)} \{X_{i\Delta} - F(X_{(i-1)\Delta};\Theta) \tag{2}$$

$$G_n^*(\Theta) = \sum_{i=1}^{n} \frac{\dot{F}(X_{(i-1)\Delta};\Theta)}{\phi(X_{(i-1)\Delta};\Theta)} \{X_{i\Delta} - F\left(X_{(i-1)\Delta};\Theta\right)\} \tag{3}$$

where

$$F(x;\Theta) = E_{\Theta}(X_{\Delta}|X_0 = x);\ \phi\left(X_{(i-1)\Delta};\Theta\right) = E_{\Theta}\{(X_{i\Delta} - F\left(X_{(i-1)\Delta};\Theta\right))^2|X_{(i-1)\Delta}\}$$

A dot denotes differentiation with respect to the parameter Θ.

$G_n^*(\Theta)$ is optimal, that is $G_n^*(\Theta)$ is in some sense closest to the score function based on the usually unknown exact likelihood function within the class functions:

$$G_n(\Theta) = \sum_{i=1}^{n} g_{i-1}(\Theta)\ \{X_{i\Delta} - F\left(X_{(i-1)\Delta};\Theta\right)\}\ , \tag{4}$$

where g_{i-1} is $\Im_{i-1}$ - measurable and continuously differentiable function of Θ, $i = 1, \dots n$. The optimal estimating function is found within the class given by (4) in the sense of giving the smallest asymptotic confidence interval around Θ and smallest asymptotic dispersion.

That two martingale estimating functions (2,3) result in consistent and asymptotically normally distributed estimators when the underlying diffusion is ergodic.

A third estimating function which is denoted by $G^+(\Theta)$, given by

$$\begin{aligned} G^+(\Theta) = \sum_{i=1}^{n} \Big(\dot{b}(X_{(i-1)\Delta};\Theta)\Delta + \tfrac{1}{2}\Delta^2\big[\dot{b}(X_{(i-1)\Delta};\Theta) b'\left(X_{(i-1)\Delta};\Theta\right) + \\ b\left(X_{(i-1)\Delta};\Theta\right) b'\left(X_{(i-1)\Delta};\Theta\right) + \tfrac{1}{2}\{\dot{\sigma}^2 b''\left(X_{(i-1)\Delta};\Theta\right) + \\ \sigma^2 b''\left(X_{(i-1)\Delta};\Theta\right)\}\big]\Big) \frac{X_{i\Delta} - F\left(X_{(i-1)\Delta};\Theta\right)}{\phi(X_{(i-1)\Delta};\Theta)}\ , \end{aligned} \tag{5}$$

where the prime denotes differentiation with respect to x.

$G^+(\Theta)$ is based on the expansion for small Δ:

$$F(x;\Theta) = x + \Delta b(x;\Theta) + \frac{1}{2}\Delta^2\left\{b(x;\Theta)b'(x;\Theta) + \sigma^2(x;\Theta)b''(x;\Theta)\right\} + O(\Delta^3)$$

So, there are three different zero-mean P_{Θ} - martingale estimating functions: one of them, G^*, is optimal in the class (4) and other two, $\tilde{G}$ and G^+, are the first- and second-order approximations in Δ of G^*.

To find the parameter Θ estimate, it is necessary to solve the equations: $G^*(\Theta) = 0$.

Since in many cases it is difficult to find the roots of G^*, good approximations to G^* are important.

So, using this approach, an estimate of the autocorrelation parameter Θ was obtained for reciprocal diffusion process [2]:

$$dX_t = -\Theta\left(X_t - \frac{\alpha}{\beta-1}\right)dt + \sqrt{\frac{2\Theta}{\beta-1}\ {X_t}^2}\ dW_t,\ t \geq 0, \tag{6}$$

where $\Theta > 0, \alpha > 0, \beta > 1, W_t$- Wiener process. It is denoted :

$$\mu(x) = -\Theta\left(x - \frac{\alpha}{\beta - 1}\right), \qquad \sigma(x) = \sqrt{\frac{2\Theta}{\beta - 1}x^2}$$

Differential equation of this type is one of the most popular short term interest rate models. The diffusion process $X = \{X_t, t \geq 0\}$ that solves this equation is ergodic with unvariant reciprocal gamma probability density function $rg(x)$:

$$rg(x) = \begin{cases} 0, & x \leq 0 \\ \dfrac{\alpha^\beta}{\Gamma(\beta)} x^{\beta-1} e^{-\alpha x}, & x > 0 \end{cases}$$

Then the estimate for Θ takes the form : $\bar{\bar{G}}_n(\Theta) = \sum_{i=1}^{n} \frac{\dot{\mu}(X_{(i-1)\Delta};\Theta)}{\sigma^2(X_{(i-1)\Delta};\Theta)}\left[X_i - F\left(X_{(i-1)\Delta};\Theta\right)\right]$

$$F(x;\Theta) = E[X_1|X_0 = x] = xe^{-\Theta} + \mu(1 - e^{-\Theta})$$

Estimation function is a zero mean martingal with respect to filtration $\Im_t$.

The corresponding estimation equation : $\bar{\bar{G}}_n(\Theta) = 0$ provided an estimator $\tilde{\Theta}$ of an unknown autocorrelation parameter Θ that is under some specific conditions P-consistent and asymptotically normal.

Estimator $\tilde{\Theta}$ of the autocorrelation parameter Θ is derived:

$$\tilde{\Theta} = ln\frac{\sum_{i=1}^{n}\frac{\tilde{B}_1(X_{(i-1)})}{(X_{i-1})^2}(X_{i-1} - \mu)}{\sum_{i=1}^{n}\frac{\tilde{B}_1(X_{(i-1)})}{(X_i)^2}(X_i - \mu)}$$

where $\tilde{B}_1(x)$ is Bessel polynomial[3-5].

3. Explicit expression for the parameter of the diffusion process

In some cases, it is possible to obtain an explicit expression for the parameter of the diffusion process. Consider two cases of exact estimation of the parameter of the diffusion process [4].

Let be $\xi = (\xi_t, \Im_t), 0 \leq t \leq T$ has a differential:

$$d\xi_t = \Theta dt + dW_t\ , \quad \xi_0 = 0 \qquad (7)$$

where Θ - $\Im_0$ is measurable normal random value, $N(m, \gamma)$, which does not depend from Wiener process $W = (W_t, \Im_t)$. Then $E\left(\Theta\middle|\Im_t^\xi\right) = \frac{m+\gamma\xi_t}{1+\gamma t}$ and process is diffusion process with differential:

$$d\xi_t = \frac{m+\gamma\xi_t}{1+\gamma t}dt + d\overline{W}_t\ , \quad \Im_t^\xi = \Im_t^{\overline{W}}, 0 \leq t \leq T$$

This representation serves as a special case of solving the optimal filtration equation, whose solution is based on the following statement.

Let be $\Theta = \Theta(\omega)$ - random value, $E\Theta^4 < \infty$ and the observed process $\xi = (\xi_t), 0 \leq t \leq T$ allows differential:

$$d\xi_t = [A_0(t,\xi) + A_1(t,\xi)\Theta]dt + B(t,\xi)dW_2(t),$$

where coefficients satisfy some conditions and conditional distribution $P(\Theta \leq \xi_0) = F_{\xi_0}(a)$ is Gaussian. Then

$$m_t = \frac{m_0 + \gamma_0 \int_0^t \frac{A_1(s,\xi)}{B^2(s,\xi)}[d\xi_s - A_0(s,\xi)ds]}{1 + \gamma_0 \int_0^t (\frac{A_1(s,\xi)}{B(s,\xi)})^2\, ds}$$

$$\gamma_t = \frac{\gamma_0}{1 + \gamma_0 \int_0^t (\frac{A_1(s,\xi)}{B(s,\xi)})^2\, ds}$$

In this case, an explicit expression $\frac{m+\gamma\xi_t}{1+\gamma t}$ was obtained for the unknown parameter Θ of the diffusion process (7).

Let's consider another case. Let there be a pair of random processes $(\Theta,\xi) = (\Theta_t,\xi_t)$, $0 \le t \le T$, where the unobservable component Θ is a Markov process with a finite or countable set of states, and the observed process is represented as a stochastic differential:

$$d\xi_t = A_t(\Theta_t,\xi)dt + B_t(\xi)dW_t$$

where W_t is Wiener process[6-8].

If a $\Theta = (\Theta_t, \Im_t), 0 \le t \le T$ - real Markov process with values in a countable set $K = \{\alpha,\beta,\dots\}$, continuous on the right; W_t - standard Wiener process, independent of Θ, ξ_0- $\Im_0$ - measurable random variable, independent of Θ.

The implementation $\xi_0^t = \{\xi_s, s \le t\}$ of the observed process ξ is known up to a point in time $0 \le t \le T$. The construction of estimates for the quantity Θ_t on ξ_0^t is the problem of filtering the unobservable process Θ. A convenient estimation characteristic for Θ is the posterior probability[9-11]:

$$\pi_\beta(t) = P\left(\Theta_t = \beta \middle| \Im_t^\xi\right), \beta \in K$$

With the help $\pi_\beta(t), \beta \in K$ can be obtained a variety of estimates of the value Θ_t. In particular, the conditional mathematical expectation $\left(\Theta_t \middle| \Im_t^\xi\right) = \sum_{\beta\in K} \beta\, \pi_\beta(t)$, which is an estimate that is optimal in the mean square sense. The estimate $\beta_t(\xi)$ obtained from the condition:

$$max_\beta\, P\left(\Theta_t = \beta \middle| \Im_t^\xi\right) = \pi_{\beta_t(\xi)}(t)$$

- the estimate that maximizes the posterior probability $\pi_\beta(t)$.

If a random variable $\Theta = \Theta(\omega)$ takes two values β and α with probabilities p and $(1-p)$, respectively. A special case of (7) is random process $\xi_t, t \ge 0$ with

$$d\xi_t = \Theta dt + dW_t, \quad \xi_0 = 0$$

Then the posterior probability $\pi(t) = P\left(\Theta = \beta \middle| \Im_t^\xi\right)$ satisfies the equation:

$$d\pi(t) = (\beta-\alpha)\pi(t)\big(1-\pi(t)\big)\big[d\xi_t - \big(\alpha + \pi(t)(\beta-\alpha)\big)dt\big],$$
$$\pi(0) = p$$

In particular, if $\beta = 1, \alpha = 0,$ then

$$d\pi(t) = \pi(t)\big(1-\pi(t)\big)[d\xi_t - \pi(t)dt], \qquad \pi(0) = p$$

If is the Radon-Nikodym density: $\lambda(t) = \frac{d\mu_1}{d\mu_0}(t, \xi)$ of the measure μ_1 corresponding to the process ξ with $\Theta = 1$ in the measure μ_0 corresponding to the process ξ with $\Theta = 0$, then from the Bayes formula at $p < 1$ it follows:

$$\pi(t) = \frac{p}{1-p} \frac{\lambda(t)}{(1 + \frac{p}{1-p}\lambda(t))}$$

In this case $\lambda(t) = \exp\{\xi_t - \frac{1}{2}\}$.

The posterior probability $\pi(t)$ is a sufficient statistic in the problem of distinguishing between two simple hypotheses - H_0: $\Theta = 0$ and H_1: $\Theta = 1$.

4. Optimal linear non-stationary filtering (Kalman-Bucy method)

On a probability space, a two-dimensional Gaussian random process $(\Theta_t, \xi_t), 0 \leq t \leq T$ is considered that satisfies the stochastic differential equations[12]:

$$d\Theta_t = a(t)\Theta_t dt + b(t)dW_1(\mathrm{t}) \tag{8}$$

$$d\xi_t = A(t)\Theta_t dt + B(t)dW_2(\mathrm{t}), \tag{9}$$

where W_1, W_2 are two independent Wiener processes, Θ_0, ξ_0 - $\mathfrak{J}_0$ are measurable.

The process $\Theta_t, 0 \leq t \leq T$ is inaccessible to observation, only values $\xi_t, 0 \leq t \leq T$ are observed that carry incomplete (due to the presence of a multiplier $A(t)$ and interference $B(t)dW_2(\mathrm{t})$) information about the values Θ_t. It is necessary at every moment in time to evaluate (filter) the values Θ_t by implementation $\xi_0^t = \{\xi_s, 0 \leq s \leq t\}$.

The optimal estimator for Θ_t, that is, the best in the root-mean-square sense, coincides with the conditional mathematical expectation $m_t = \pi_t(\Theta)$. The estimation (filtering) error is denoted as:

$$\gamma_t = E(\Theta_t - m_t)^2$$

The method used by Kalman and Bucy to find m_t and γ_t made it possible to obtain for these quantities a closed system of recurrent equations.

The process $(\Theta_t, \xi_t), 0 \leq t \leq T$ considered by Kalman and Bucy is Gaussian. As a consequence, the optimal estimate m_t turns out to be linear. However, in the conditionally Gaussian case, for m_t can also be obtained a closed system of equations, although the estimate m_t will be, in general, nonlinear.

That is, under certain conditions on the functions $a(t), b(t), A(t), B(t)$ in (8,9), the conditional mathematical expectation m_t and root-mean-square filtering error γ_t satisfy the following system of equations:

$$dm_t = a(t)m_t dt + \frac{\gamma_t A(t)}{B^2(t)}(d\xi_t - A(t)m_t dt)$$

$$\dot{\gamma}_t = 2a(t)\gamma_t - \frac{A^2(t){\gamma_t}^2}{B^2(t)} + b^2(t)$$

The system of equations has a unique continuous solution (for γ_t - in the class of non-negative functions).

5. Normal correlation theorem in the discrete case

The main results on estimating the parameters of conditionally Gaussian processes are based on the normal correlation theorem and the concept of pseudoinverse matrices [4].

An matrix A^+(order: n x m) is called pseudoinverse to the non-degenerated matrix $A = A_{nxm}$ if the following conditions are satisfied: $AA^+A = A, A^+ = UA^* = A^*V$, where U and V – some matrices. Matrix A^+rows and columns serve as linear combinations of matrix A^*rows and columns.

The normal correlation theorem states that for a Gaussian vector $(\Theta,\xi) = ([\Theta_1,\dots,\Theta_k],[\xi_1,\dots,\xi_l])$ with $m_\Theta = E\Theta, m_\xi = E\xi$, $D_{\Theta\Theta} = cov(\Theta,\Theta), D_{\Theta\xi} = cov(\Theta,\xi), D_{\xi\xi} = cov(\xi,\xi)$, then conditional mathematical expectation $E(\Theta|\xi)$ and conditional covariance:

$$cov(\Theta,\Theta|\xi) = E\{[\Theta - E(\Theta|\xi)][\Theta - E(\Theta|\xi)]^*|\xi\}$$

have a form:

$$E(\Theta|\xi) = m_\Theta + D_{\Theta\xi} {D_{\xi\xi}}^+ (D_{\Theta\xi})^* \quad (10)$$

$$cov(\Theta,\Theta|\xi) = D_{\Theta\Theta} - D_{\Theta\xi} {D_{\xi\xi}}^+ (D_{\Theta\xi})^* \quad (11)$$

As a consequence, we obtain, in general, the well-known statement about the optimal estimation of the parameter Θ. If $k = l = 1, D\xi > 0$, then

$$E(\Theta|\xi) = E\Theta + \frac{cov(\Theta,\xi)}{D\xi}(\xi - E\xi)$$

$$D(\Theta|\xi) = D\Theta - \frac{cov^2(\Theta,\xi)}{D\xi},$$

where $D(\Theta|\xi) = E\{[\Theta - E(\Theta|\xi)]^2|\xi\}$.
Introducing the correlation coefficient: $\rho = \frac{cov(\Theta,\xi)}{\sigma_\Theta \sigma_\xi}$, $\sigma_\Theta = +\sqrt{D\Theta}, \sigma_\xi = +\sqrt{D\xi}$,
formulas get the known form:

$$E(\Theta|\xi) = E\Theta + \rho(\xi - E\xi)$$

$$D(\Theta|\xi) = \sigma^2(1 - \rho^2)$$

Also, this fundamental result (10,11) leads to an interesting consequence:
Let be $\Theta = b_1\varepsilon_1 + b_2\varepsilon_2$, $\xi = B_1\varepsilon_1 + B_2\varepsilon_2$, where $\varepsilon_1, \varepsilon_2$ - independent Gaussian values, $E\varepsilon_i = 0$, $D\varepsilon_i = 0, i = 1,2,\ B_1^2 + B_2^2 > 0$, then

$$E(\Theta|\xi) = \frac{b_1B_1 + b_2B_2}{B_1^2 + B_2^2}\xi$$

$$D(\Theta|\xi) = \frac{(B_1b_2 - b_1B_2)^2}{B_1^2 + B_2^2}$$

These statements, based on the theorem of normal correlation in the discrete case, have become widespread in statistical studies of diffusion processes.

Conclusions

Diffusion processes are widely used in technical applications, economics, and finance. Estimation of the unknown parameters of diffusion processes is associated with a number of difficulties, which are not an easy task to overcome. The paper discusses the main approaches to solving this issue. For this, statistical estimates, equations of optimal linear and nonlinear filtering are used. These approaches are illustrated with examples.

References

[1] B M Bibby and M Sorensen 1995Martingal estimation functions for discretely observed diffusion processes, Bernoulli 1(1/2) pp 017-039

[2] A N Shiryayev (ed.) 1992 Selected Works of A N Kolmogorov, On Analytic Methods in Probability Theory, Vol. II, Kluwer Academic Publisher, Dordrecht pp 62-108

[3] E Wong 1964 The construction of a class of stationary Markov processes, in Sixteen Symposium in Applied Mathematics – Stochastic Processes in Mathematical Physics and Engineering, Vol. 16, R. Bellman, ed., American Mathematical Society, Providence, RI pp 264-276

[4] F Avram, N N Leonenko, and N Šuvak 2010 On spectral properties of Fisher–Snedecor diffusion, submitted for publication

[5] J L Forman and M Sørensen 2008 The Pearson diffusions: A class of statistically tractable diffusion processes, Scand. J. Statist. 35 pp 438-465

[6] N N Leonenko and N Šuvak 2010 Statistical inference for reciprocal gamma diffusion process, J Statist. Plann. Inference 140, pp 30-51

[7] N N Leonenko and N Šuvak 2010 Statistical inference for Student diffusion process, Stoch. Anal. Appl. 28, pp 972-1002.

[8] F Avram, N N Leonenko, and L Rabehasaina 2009 Series expansions for the first passage distribution of Wong-Pearson jump-diffusion, Stoch. Anal. Appl. 27, pp 770–796

[9] Y.A. Kutoyants 2004 Statistical Inference for Ergodic Diffusion Processes, Springer, NewYork, Heidelberg, Berlin

[10] Y A Kutoyants and N Yoshida 2007 Moment estimation for ergodic diffusion processes, Bernoulli 13 pp 933-951

[11] K Itô and H McKean 1974 *Diffusion Processes and their Sample Paths*, Springer, NewYork, Heidelberg, Berlin

[12] R S Liptser and A N Shiryayev 1977 Statistics of RandomPprocesses I,II, Springer, New York