ISSN 0005-1179, Automation and Remote Control, 2015, Vol. 76, No. 10, pp. 1771–1783. © Pleiades Publishing, Ltd., 2015.
Original Russian Text © V.D. Konakov, A.R. Markova, 2015, published in Avtomatika i Telemekhanika, 2015, No. 10, pp. 74–89.

Linear Trend Exclusion for Models Defined with Stochastic Differential and Difference Equations

V. D. Konakov and A. R. Markova

National Research University Higher School of Economics, Moscow, Russia
e-mail: VKonakov@hse.ru, anna.r.markova@gmail.com

Received October 30, 2014

Abstract—We consider a sequence of Markov chains that weakly converge to a diffusion process. We assume that the trend contains a linearly growing component. The usual parametrix method does not apply since the trend is unbounded. We show how to modify the parametrix method in order to get local limit theorems in this case.

DOI: 10.1134/S0005117915100057

1. INTRODUCTION

In this introduction, we show where the problem we consider has arisen from. Note that the introduction is not intended to provide a comprehensive survey.

The importance of studying discretizations of stochastic differential equations (SDE) follows already from the fact that all known approximate modeling techniques for SDEs are based on some kind of discretization. It is well known that SDE solutions can be obtained in a closed form only for a small number of subclasses of stochastic equations, so various discretization methods (the Euler–Maruyama scheme, Milstein's scheme, schemes based on higher order stochastic Taylor decompositions) attracts a lot of attention from researchers. A natural question arises: do discretization schemes reproduce asymptotic properties of the original process (such as the mixing rate, principle of large deviations and so on). Another important question is how "close" the resulting approximation is, in some sense, to a solution of the original SDE. Our study deals with that second question. Historically, the first considered scheme was a sequence of Markov chains defined on a grid with grid step tending to zero, and weak convergence of measures (transition probabilities) to the transition probability of some limit diffusion process was usually studied. The first general results on weak convergence in this scheme were obtained by A.V. Skorokhod in 1961 [1]. Skorokhod's results held for a quite general class of Markov chains and processes that might have jumps. For continuous diffusion, the most general results on weak convergence were obtained in the book of D. Stroock and S. Varadhan [2]. They developed an approach based on solving the so-called "martingale problem." We emphasize once again that these first results dealt with weak convergence of measures. Modern theory of weak convergence for probability measures is a theory with decades of history; it is primarily related to the names of A.N. Kolmogorov, J. Doub, M. Donsker, Yu.V. Prokhorov, Skorokhod, L. Le Cam and Varadhan.

We now assume that transition probabilities of both Markov chains and the limit diffusion process are absolutely continuous with respect to the Lebesgue measure. Then it is natural to ask when these transition densities also converge, i.e., when the corresponding local limit theorem holds. To answer that question, V. Konakov and S. Molchanov [3] proposed a discrete version of the parametrix method that was further refined and generalized in a cycle of works by Konakov and E. Mammen [4–7]. The parametrix method has been known for a long time in the theory of differential equations; it was proposed by E. Levy back in 1907 [8, 9] and was then developed in the works of A. Friedman [10], A. Il'in, A. Kalashnikov, and O. Oleinik [11] and others. However, for our purposes this version of the parametrix method does not apply. In 1967, H. McKean and I. Singer [12] proposed a modification of the parametrix method that turned out to admit a discrete version and let them develop a new method for proving local limit theorems for transition densities in a sequence of Markov chains that weakly converge to a limit diffusion process. One significant requirement in the proof of these results was the condition that shift and diffusion coefficients are bounded. Boundedness was needed for the series constructed in the parametrix method to converge. This requirement narrowed down the applicability of these results and did not let the authors consider a number of important specific models. The purpose of this work is to define a procedure that lets us exclude a linearly growing trend component and reduce the problem to the already studied problem with bounded shift and diffusion coefficients. This procedure was applied both to diffusion and to Markov chains. For SDEs, a similar trend exclusion procedure was previously applied in the work of F. Delarue and S. Menozzi [13] to get two-sided bounds on the transition density for some degenerate SDEs of Kolmogorov type. To the best of our knowledge, for Markov chains this procedure is novel. The essence of our proposed procedure is simple: we compensate a growing trend by returning along the trajectories of the system of ordinary differential equations that results by discarding the "Brownian" component of the SDE. For this compensated process, with Ito's formula we can write its stochastic differential and SDE that it satisfies. This SDE now has a bounded shift coefficient. For Markov chains we perform a similar procedure, only instead of a differential equation we use a difference equation, and return back not along the trajectories of differential equations but along its Euler polylines. Then, applying known results for the bounded case, we can with a simple transformation return to the original problem and obtain local limit theorems in the original model. In further work, we also plan to consider a more general case of a growing trend with a bounded gradient and the case of an unbounded diffusion coefficient.

2. THE ESSENCE OF THE PARAMETRIX METHOD AND NECESSARY PRELIMINARIES FROM THE THEORY OF DIFFERENTIAL AND DIFFERENCE EQUATIONS

Consider the class of stochastic problems that can be solved with the parametrix method. Suppose that on the interval $[0,1]$ there is a sequence of partitions $\mathit{\Gamma}_n = \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$, $n = 1, 2, \ldots$, and a sequence of Markov chains $X_t^{(n)}$ with discrete time and continuous state space. Chains $X_t^{(n)}$ are defined on the lattice $\mathit{\Gamma}_n$, have initial distribution $\delta_{x_0}(\cdot)$, and the one-step transition probability has density

$$p^{(n)}\left(\frac{1}{n}, x, A\right) = P\left(X^{(n)}_{\frac{i+1}{n}} \in A | X^{(n)}_{\frac{i}{n}} = x\right) = \int\limits_A p^{(n)}_{\frac{i}{n},x}\left(\frac{1}{n}, x, z\right) dz.$$

Under this condition, transition probability over n steps is also absolutely continuous with respect to the Lebesgue measure and has density $p^{(n)}(1, x_0, z)$. The problem is to find conditions under which the local limit theorem holds, i.e., conditions under which the density $p^{(n)}(1, x_0, z)$ can be approximated with a density of some diffusion process $p(1, x_0, z)$. For these purposes, we use the analytical method based on a special version of the parametrix method (McKean and Singer [12]) whose classical version was proposed by Levy in 1907 [8, 9]. Next we give a brief description of the parametrix method in the form of McKean and Singer.

The Parametrix Method in the Form of McKean and Singer

Consider a diffuse process Y_t which is a solution for the SDE

$$dY = b(t, Y)dt + \sigma(t, Y)dB(t), \quad Y(0) = x \in \mathrm{R}^{\mathrm{d}}, \quad t \in [0, 1],$$

where $B(t)$ is the standard Wiener process, $\sigma(z)$ is a symmetric matrix such that matrix $a(z) = \sigma(z)\sigma^{\mathrm{T}}(z)$ satisfies the uniform ellipticity condition, functions $b(z)$ and $a(z)$ are bounded and satisfy Hölder's condition, and, moreover, there exist bounded and continuous derivatives $\frac{\partial a_{ij}}{\partial z_j}$, $\frac{\partial^2 a_{ij}}{\partial z_i \partial z_j}$, and $\frac{\partial b_i}{\partial z_i}$ that satisfy Hölder's condition.

Consider the direct and reverse Kolmogorov equations:

$$-\frac{\partial p(t-s,x,y)}{\partial s} = L_x p = \frac{1}{2}\sum_i a_{ij}(x)\frac{\partial^2 p(t-s,x,y)}{\partial x_i \partial x_j} + \sum_i b_i(x)\frac{\partial p(t-s,x,y)}{\partial x_i} \tag{1}$$

and

$$\frac{\partial p(t-s,x,y)}{\partial t} = L_x^{\mathrm{T}} p = \frac{1}{2}\sum_i \frac{\partial^2 \left[a_{ij}(y)p(t-s,x,y)\right]}{\partial y_i \partial y_j} - \sum_i \frac{\partial \left[b_i(y)p(t-s,x,y)\right]}{\partial y_i}. \tag{2}$$

Consider another diffuse process $\tilde{Y} = \tilde{Y}_{s,x,y}$ defined on the interval $s \leqslant t \leqslant 1$, which represents a solution of the following SDE:

$$d\tilde{Y}(t) = b(y)dt + \sigma(y)dB(t), \quad \tilde{Y}(s) = x \in \mathrm{R}^{\mathrm{d}}, \quad t \in [s,1].$$

Processes of the form $\tilde{Y} = \tilde{Y}_{s,x,y}$ are called diffusions frozen at point y. The transition density of such a diffusion is Gaussian

$$\tilde{p}^{(y)}(t-s,x,y) = (2\pi)^{-\frac{d}{2}}(t-s)^{-\frac{d}{2}}\left(\det a(y)\right)^{-\frac{1}{2}}$$
$$\times \exp\left(-\frac{1}{2(t-s)}\left\{y-x-b(y)(t-s)\right\}^{\mathrm{T}} a^{-1}(y)\left\{y-x-b(y)(t-s)\right\}\right).$$

We introduce the necessary notation and operations. We define the singular kernel $H(t-s,x,y)$ and binary operation $\otimes$ of convolution type as follows:

$$H(t-s,x,y) = \frac{1}{2}\sum_{i,j}\left(a_{ij}(x) - a_{ij}(y)\right)\frac{\partial^2 \tilde{p}(t-s,x,y)}{\partial x_i \partial x_j} + \sum_i \left(b_i(x) - b_i(y)\right)\frac{\partial \tilde{p}(t-s,x,y)}{\partial x_i},$$

$$(f \otimes g)(s,t,x,y) = \int_s^t d\tau \int_{\mathrm{R}^{\mathrm{d}}} f(s,\tau,x,z)g(\tau,t,z,y)dz.$$

The fundamental solution of Eqs. (1) and (2) can be represented as

$$p(t-s,x,y) = \sum_{r=0}^{\infty}\left(\tilde{p} \otimes H^{(r)}\right)(t-s,x,y), \tag{3}$$

where $H^{(r)} = H^{(r-1)} \otimes H$.

It is important to represent the transition density in the form of McKean and Singer because this representation can be applied in the discrete case. Consider a sequence of partitions $\mathit{\Gamma}_n = \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$, $n = 1, 2, \ldots$, and a sequence of Markov chains $X_t^{(n)}$ with discrete time and continuous state space. Chains $X_t^{(n)}$ are defined on the lattice $\mathit{\Gamma}_n$, have initial distribution $\delta_{x_0}(\cdot)$, and the chain's dynamics is defined by recurrent relation

$$X_{\frac{i+1}{n}}^{(n)} = X_{\frac{i}{n}}^{(n)} + \frac{1}{n}b\left(X_{\frac{i}{n}}^{(n)}\right) + \frac{1}{\sqrt{n}}\varepsilon_{\frac{i+1}{n}}^{(n)}, \quad 0 \leqslant i \leqslant n-1, \quad X_0^{(n)} = x. \tag{4}$$

For each $0 < s = \frac{j}{n} < 1$ and $x,\, y \in \mathrm{R^d}$ we define a Markov chain $\tilde{X}_t^{(n)} = \tilde{X}_{s,x,y}^{(n)}$. This chain is defined on the lattice $\left\{\frac{j}{n}, \frac{j+1}{n}, \ldots, 1\right\}$ with recurrent relation

$$\tilde{X}_{\frac{i+1}{n}}^{(n)} = \tilde{X}_{\frac{i}{n}}^{(n)} + \frac{1}{n} b\,(y) + \frac{1}{\sqrt{n}} \tilde{\varepsilon}_{\frac{i+1}{n}}^{(n)}, \quad j \leqslant i \leqslant n-1, \quad \tilde{X}_{\frac{j}{n}}^{(n)} = x. \tag{5}$$

We introduce discrete counterparts $H_n(t-s,x,y)$ and $\otimes_n$ of the singular kernel $H(t-s,x,y)$ and binary operation $\otimes$:

$$H_n\left(\frac{j'}{n} - \frac{j}{n}, x, y\right) = \left(L^n - \tilde{L}^{n,y}\right) \tilde{p}_n^y \left(\frac{j'}{n} - \frac{j+1}{n}, x, y\right),$$

$$(f \otimes_n g)\left(\frac{j'}{n} - \frac{j}{n}, x, y\right) = \sum_{i=j}^{j'-1} \frac{1}{n} \int\limits_{\mathrm{R^d}} f\left(\frac{i}{n} - \frac{j}{n}, x, z\right) g\left(\frac{j'}{n} - \frac{i}{n}, z, y\right) dz,$$

where L^n and $\tilde{L}^{n,y}$ are infinitesimal operators of chains (4) and (5), and $\tilde{p}_n^y\left(\frac{i}{n}, x, y\right)$ is the transition density of chain (5).

The transition density of the Markov chain (4) can be represented as

$$p_n\left(\frac{j'}{n} - \frac{j}{n}, x, y\right) = \sum_{r=0}^{j'-j} \left(\tilde{p}_n \otimes_n H_n^{(r)}\right)\left(\frac{j'}{n} - \frac{j}{n}, x, y\right). \tag{6}$$

The proximity of left-hand sides of representations (3) and (6) can be established with a detailed analysis of the series. One can show that

$$\tilde{p} \approx \tilde{p}_n, \quad H^{(r)} \approx H_n^{(r)}, \quad \otimes \approx \otimes_n.$$

Some Known Facts from the Theory of Differential and Difference Equations

Consider a linear uniform differential equation (LUDE):

$$x' = A(t)x, \quad x \in \mathrm{R^d}, \quad A(t) : \mathrm{R^d} \mapsto \mathrm{R^d}, \quad t \in [0,1], \tag{7}$$

where $A(t)$ is a continuous matrix function.

Consider a set X of all solutions for Eqs. (7) defined on the interval $[0,1]$. The set X is a vector space consisting of functions $\phi : [0,1] \mapsto \mathrm{R^d}$.

Together with the vector differential Eq. (7) we will consider the matrix differential equation

$$X' = A(t)X. \tag{8}$$

We will say that the matrix function $\Phi(t)$ is a solution of Eq. (8) on the interval $[0,1]$ if $\Phi(t)$ is continuously differentiable for $t \in [0,1]$ and $\Phi'(t) = A(t)\Phi(t)$, $t \in [0,1]$.

The next theorem establishes a relation between solutions for Eqs. (7) and (8).

Theorem 1 [14, p. 33]. *Let $A(t)$ be a continuous $n \times n$ matrix-valued function on the interval $[0,T]$, and let $\Phi(t)$ be an $n \times n$ matrix-valued function with columns $\phi_1(t), \phi_2(t), \ldots, \phi_n(t)$:*

$$\Phi(t) = [\phi_1(t), \phi_2(t), \ldots, \phi_n(t)], \quad t \in [0,T].$$

Then Φ is a solution of the matrix differential Eq. (8) on $[0,T]$ if and only if each column ϕ_i is a solution of the vector differential Eq. (7) on $[0,T]$, $i = 1,2,\ldots,n$. Moreover, if Φ is a solution of matrix Eq. (8) then

$$x(t) = \Phi(t)c$$

is a solution of the vector differential Eq. (7) for any $n \times 1$ vector of constants c.

We introduce the following definitions.

Definition 1. A fundamental system of solutions for Eq. (7) is the basis of the vector space X.

Definition 2. A matrix whose columns form a fundamental system of solutions is called a fundamental matrix of differential Eq. (7).

Consider a linear uniform difference equation of order k:

$$x(s+k) + a_1(s)x(s+k-1) + \ldots + a_k(s)x(s) = 0. \tag{9}$$

The set of solutions for difference Eq. (9) is a vector space as well. Similar to the definition for differential equations, the *fundamental matrix of a difference equation* is a matrix whose columns form the *fundamental system of solutions for Eq.* (9), i.e., the basis of the vector space of all solutions for this difference equation.

Note that as a fundamental matrix for a differential or difference equation at the initial time moment $t = 0$ it is convenient to take the unit matrix of the corresponding dimension. In this case $x(t) = \Phi(t)c$ is a solution vector for Eq. (7) that satisfies initial conditions $x(0) = \Phi(0)c = c$.

3. EXCLUDING THE LINEAR TREND COMPONENT FOR A DIFFUSION AND A MARKOV CHAIN

Consider the following diffusion model:

$$dY = \{b(t)Y + m(t,Y)\}dt + \sigma(t,Y)dB(t), \quad Y(0) = x \in \mathrm{R^d}, \quad t \in [0,1], \tag{10}$$

where $B(t)$ is the standard Wiener process. Interval $[0,1]$ is taken purely for convenience and can be replaced with any interval.

Consider also a sequence of Markov chains with the same initial conditions as in the diffusion model (10):

$$\begin{gathered} X_n\left(\frac{k+1}{n}\right) = X_n\left(\frac{k}{n}\right) + \frac{1}{n}\left\{b_n\left(\frac{k}{n}\right)X_n\left(\frac{k}{n}\right) + m_n\left(\frac{k}{n}, X_n\left(\frac{k}{n}\right)\right)\right\} + \frac{1}{\sqrt{n}}\varepsilon_n\left(\frac{k+1}{n}\right), \\ X_n(0) = x \in \mathrm{R^d}, \quad k = 0,1,2,\ldots,n. \end{gathered} \tag{11}$$

As for the innovations ε_n, we make a standard Markov assumption, namely assume that the random value $\varepsilon_n\left(\frac{k+1}{n}\right)$ for fixed "past" $X_n\left(\frac{i}{n}\right) = x(i)$, $i = 0,1,2,\ldots,k$, depends only on the value of the process $x(k)$ at the last time moment $\frac{k}{n}$ and has conditional distribution density $q_{n,\frac{k}{n},x(k)}(\cdot)$ which is an element of the family of densities $q_{n,t,x}(\cdot)$ parameterized with three parameters $(n,t,x) \in N \times [0,1] \times \mathrm{R^d}$. We make the following assumptions regarding the family of densities $q_{n,t,x}(\cdot)$ and coefficients of Eq. (10).

1. Symmetrix matrix $a(t,x) = \sigma(t,x)\sigma^{\mathrm{T}}(t,x)$ is bounded and positive definite: there exists $C > 1$ such that $C^{-1} \leqslant \theta^{\mathrm{T}}a(t,x)\theta \leqslant C$ for all θ such that $\mid \theta \mid = 1$, $x \in \mathrm{R^d}$, $t \in [0,1]$.

2. Matrix functions $b_n(t)$ and $b(t)$ are continuous on $[0,1]$. Functions $a(t,x)$ and $m(t,x)$ together with their first derivatives are uniformly continuous and bounded with respect to (t,x), $x \in \mathrm{R^d}$, $t \in [0,1]$ and Lipschitz with respect to variable x with Lipschitz constant independent of t. Moreover, second derivatives $\frac{\partial^2 a(t,x)}{\partial x_i \partial x_j}$, $1 \leqslant i,\ j \leqslant d$, exist and satisfy Hölder's condition with respect to variable x with a constant independent of t.

3. $\int q_{n,t,x}(z)zdz = 0$, $\int q_{n,t,x}(z)zz^{\mathrm{T}}dz \triangleq a_n(t,x)$, $x \in \mathrm{R^d}, t \in [0,1]$.

4. There exists a positive integer S' and function $\psi(x) : \mathrm{R^d} \mapsto R$, $\sup_{x\in\mathrm{R^d}} \mid \psi(x) \mid < \infty$ such that $\int \mid x \mid^S \psi(x)dx < \infty$, where $S = 2dS' + 4$, and for all sufficiently large n and $z \in \mathrm{R^d}$ it holds

that

$$| D_z^v q_{n,t,x}(x) | \leqslant \psi(z), \quad | v | = 0,1,2,3,4,$$
$$| D_x^v q_{n,t,x}(x) | \leqslant \psi(z), \quad | v | = 0,1,2,$$

where $x \in \mathrm{R}^{\mathrm{d}}$, $t \in [0,1]$.

The existence of a transition density for model (11) immediately follows from the model's assumptions. The existence of a transition density for the diffusion model (10) may be obtained from Hermander's theory [15], but under stronger assumptions on the coefficients. However, the existence of a transition density in this model can also be proven with the parametrix method under weaker conditions on the coefficients that Hermander's theory requires. We first consider the diffuse model (10). If the model's coefficients are bounded, the existence of a transition density follows from the work of Il'in, Kalashnikov, and Oleinik [11] or from the work of Konakov and Mammen [4], which constructs the parametrix for such an equation. But in model (10), the trend is unbounded and increases linearly. To apply the parametrix method for this model, we first get rid of the trend and consider a new model with bounded coefficients. Then we apply the parametrix method for this new model with bounded coefficients and return to the original model with an unbounded linearly increasing trend.

Consider a linear system of ordinary differential equations (ODE):

$$y'(t) = b(t)y(t), \quad y(0) = x \in \mathrm{R}^{\mathrm{d}}, \quad t \in [0,1]. \tag{12}$$

Let $\Phi(t)$ be a fundamental matrix corresponding to this system. We remind that for system (12) this means that all solutions can be continued to the entire segment $[0,1]$, and the fundamental matrix is a matrix whose columns are independent solutions of this system that satisfy initial conditions. As the initial fundamental matrix we take the unit matrix $\Phi(0) = I$. The fundamental matrix is a solution of equation $\Phi'(t) = b(t)\Phi(t)$ and a nondegenerate matrix on the interval $[0,1]$. The inverse matrix $\Phi^{-1}(t)$ satisfies equation $[\Phi^{-1}(t)]' = -\Phi^{-1}(t)b(t)$ and initial condition $\Phi^{-1}(0) = I$. To remove the trend's linear component, we consider the process $\tilde{Y}(t) = f(t, Y(t))$, where $f(t,y) = \Phi^{-1}(t)y$. Function $f(t,y)$ is continuous on $[0,1] \times \mathrm{R}^{\mathrm{d}}$ and has continuous partial derivatives $\frac{\partial f}{\partial t}$ and $\frac{\partial f}{\partial y_i}$, so to compute the stochastic differential of process $\tilde{Y}(t)$ we can use Ito's formula. We get that

$$\begin{aligned} d\tilde{Y}(t) &= d\left[\Phi^{-1}(t)Y(t)\right] = \Phi^{-1}(t)dY(t) + d\Phi^{-1}(t)Y(t) \\ &= \Phi^{-1}(t)\left(\{b(t)Y(t) + m(t,Y(t))\}dt + \sigma(t,Y(t))\,dB(t)\right) + [\Phi^{-1}(t)]'dt\,Y(t) \\ &= \Phi^{-1}(t)b(t)Y(t)dt + \Phi^{-1}(t)m(t,Y(t))\,dt + \Phi^{-1}(t)\sigma(t,Y(t))\,dB(t) \\ &\quad -\Phi^{-1}(t)b(t)Y(t) = \Phi^{-1}(t)m(t,Y(t))\,dt + \Phi^{-1}(t)\sigma(t,Y(t))\,dB(t) \\ &= \tilde{m}\left(t,\tilde{Y}(t)\right)dt + \tilde{\sigma}\left(t,\tilde{Y}(t)\right)dB(t), \end{aligned}$$

where $\tilde{m}\left(t,\tilde{Y}(t)\right) = \Phi^{-1}(t)m\left(t,\Phi(t)\tilde{Y}(t)\right)$, $\tilde{\sigma}\left(t,\tilde{Y}(t)\right) = \Phi^{-1}(t)\sigma\left(t,\Phi(t)\tilde{Y}(t)\right)$.

It is clear that process $\tilde{Y}(t)$ that we have introduced is a diffusion process satisfying the SDE with bounded trend $\tilde{m}\left(t,\tilde{Y}(t)\right)$ and positive definite diffusion matrix $\tilde{\sigma}\left(t,\tilde{Y}(t)\right)$:

$$d\tilde{Y}(t) = \tilde{m}\left(t,\tilde{Y}(t)\right)dt + \tilde{\sigma}\left(t,\tilde{Y}(t)\right)dB(t).$$

Indeed,

$$\begin{aligned} \tilde{a}(t,y) &= \tilde{\sigma}\left(t,\tilde{Y}(t)\right)\left[\tilde{\sigma}\left(t,\tilde{Y}(t)\right)\right]^{\mathrm{T}} = \Phi^{-1}(t)\sigma(t,\Phi(t)y)\left[\Phi^{-1}(t)\sigma(t,\Phi(t)y)\right]^{\mathrm{T}} \\ &= \Phi^{-1}(t)\sigma(t,\Phi(t)y)\left[\sigma(t,\Phi(t)y)\right]^{\mathrm{T}}[\Phi^{-1}(t)]^{\mathrm{T}} = \Phi^{-1}(t)a(t,\Phi(t)y)\left[\Phi^{-1}(t)\right]^{\mathrm{T}} \end{aligned}$$

and, consequently,

$$\theta^{\mathrm{T}}\tilde{\sigma}(t,y)[\tilde{\sigma}(t,y)]^{\mathrm{T}}\theta = 0 \Leftrightarrow \vartheta^{\mathrm{T}}\sigma(t,y)[\sigma(t,y)]^{\mathrm{T}}\vartheta = 0, \quad \vartheta = [\Phi^{-1}(t)]^{\mathrm{T}}\vartheta.$$

It remains use the fact that matrix $a(t,\Phi(t)y) = \sigma(t,\Phi(t)y)[\sigma(t,\Phi(t)y)]^{\mathrm{T}}$ is positive definite.

The existence of a transition density $\rho_{\tilde{Y}}(t)$ for process $\tilde{Y}(t)$ has been proven with the parametrix method in [4]. With known transformation formulas, the transition density of the process $Y(t) = \Phi(t)\tilde{Y}(t)$ is

$$\rho_Y(s,t,x,y) = \det[\Phi^{-1}(t)]\rho_{\tilde{Y}}(s,t,\Phi^{-1}(s)x,\Phi^{-1}(t)y). \tag{13}$$

For model (11), we consider the trend removal procedure which is a discrete counterpart of the above procedure for the diffusion equation. Consider a difference equation without trend:

$$\frac{x_n\left((k+1)h\right) - x_n(kh)}{h} = b_n(kh)X_n(kh), \quad x_n(0) = x$$

on the grid $\Gamma = \{0, h, 2h, \ldots, nh = 1\}$, $h = \frac{1}{n}$.

In matrix notation,

$$x_n\left((k+1)h\right) = \left(I + h\, b_n(kh)\right) x_n(kh), \quad x_n(0) = x.$$

Iterating the latter equality, we get

$$\begin{aligned}
&x_n(h) \ \ = \left(I + h\, b_n(0)\right) x,\\
&x_n(2h) = \left(I + h\, b_n(h)\right) x_n(h) = \left(I + h\, b_n(h)\right)\left(I + h\, b_n(0)\right) x,\\
&\ldots\\
&x_n(kh) = \Phi_n(kh)x,
\end{aligned}$$

where $\Phi_n(kh) = \left(I + h\, b_n\left((k-1)h\right)\right)\Phi_n\left((k-1)h\right)\Phi_n(kh)$ is the fundamental matrix from the theory of difference equations [16], the discrete counterpart of the fundamental matrix $\Phi(t)$ defined on the grid $\{0, h, 2h, \ldots, nh = 1\}$ with initial condition $\Phi_n(0) = I$. We define a new Markov chain:

$$\tilde{X}_n(kh) = \Phi_n^{-1}(kh)X_n(kh), \quad \tilde{X}_n(0) = x.$$

By known transformation formulas, the transition density of the Markov chain $X_n(kh) = \Phi_n(kh)\,\tilde{X}_n(kh)$ is

$$\rho_{X_n}(ih, jh, x, y) = \det\left[\Phi^{-1}(jh)\right]\rho_{\tilde{X}_n}\left(ih, jh, \Phi^{-1}(ih)x, \Phi^{-1}(jh)y\right). \tag{14}$$

Thus, according to model (11) we have that

$$\begin{aligned}
\tilde{X}_n\left((k+1)h\right) &= \Phi_n^{-1}\left((k+1)h\right)\ X_n\left((k+1)h\right)\\
&= \Phi_n^{-1}(kh)\left(I + h\, b_n(kh)\right)^{-1}\left\{\left(I + h\, b_n(kh)\right)X_n(kh)\right.\\
&\quad \left.+ h\, m_n\left(kh,\ X_n(kh)\right) + \sqrt{h}\varepsilon_n\left((k+1)h\right)\right\}\\
&= \Phi_n^{-1}(kh)X_n(kh) + h\,\Phi_n^{-1}(kh)\ \left(I + h\, b_n(kh)\right)^{-1}\ m_n\left(kh, X_n(kh)\right)\\
&\quad + \sqrt{h}\,\Phi_n^{-1}(kh)\ \left(I + h\, b_n(kh)\right)^{-1}\ \varepsilon_n\left((k+1)h\right)\\
&= \tilde{X}_n(kh) + h\,\tilde{m}_n\left(kh,\ \tilde{X}_n(kh)\right) + \sqrt{h}\,\tilde{\varepsilon}_n\left((k+1)h\right),
\end{aligned}$$

where

$$\tilde{m}_n\left(kh,\ \tilde{X}_n(kh)\right) = \Phi_n^{-1}(kh)\ (I + h\, b_n(kh))^{-1}\ m_n\left(kh,\ \Phi_n(kh)\ \tilde{X}_n(kh)\right),$$
$$\tilde{\varepsilon}_n\left((k+1)h\right) = \Phi_n^{-1}(kh)\ (I + h\, b_n(kh))^{-1}\ \varepsilon_n\left((k+1)h\right).$$

Now the definition of $\tilde{\varepsilon}_n$ implies that the random value $\tilde{\varepsilon}_n((k+1)h)$ for fixed "past" $\tilde{X}_n(ih) = \tilde{x}(i)$, $i = 0, 1, 2, \ldots, k$, depends only on the value of the process $\tilde{x}(k)$ at the last time moment kh and has conditional density

$$\tilde{q}_{n,kh,\tilde{x}(k)}(z) = \det\ \Phi_n\left((k+1)h\right)\ q_{n,kh,\Phi_n(kh)\tilde{x}(k)}\left(\Phi_n\left((k+1)h\right) z\right) \tag{15}$$

from the family of densities

$$\det\ \Phi_n\left(([tn]+1)h\right) \tilde{q}_{n,t,\Phi_n(([tn]+1)h)x(k)}\left(\Phi_n\left(([tn]+1)h\right) z\right)$$

that depends on three parameters $(n, t, x) \in N \times [0, 1] \times \mathrm{R}^{\mathrm{d}}$. Densities $\tilde{q}_{n,kh,\tilde{x}(k)}(z)$ in (15) satisfy conditions 3 and 4 formulated in Section 3, with the difference that $\phi(x)$ from condition 4 is replaced with $C\,\phi(x)$, where C is a constant. Making a change of variables $v = \Phi\left(\frac{[tn]+1}{n}\right) z$, for $t = kh$ we have

$$\begin{gathered}
\int \tilde{q}_{n,t,\tilde{x}}(z)dz = \det \Phi_n\left(\frac{[tn]+1}{n}\right) \int q_{n,t,\Phi_n\left(\frac{[tn]}{n}\right)\tilde{x}}\left(\Phi_n\left(\frac{[tn]+1}{n}\right) z\right) zdz \\
= \det \Phi_n\left(\frac{[tn]+1}{n}\right) \det \Phi_n^{-1}\left(\frac{[tn]+1}{n}\right) \int q_{n,t,\Phi_n\left(\frac{[tn]}{n}\right)\tilde{x}}(v) \Phi_n^{-1}\left(\frac{[tn]+1}{n}\right) vdv \\
= \Phi_n^{-1}\left(\frac{[tn]+1}{n}\right) \int q_{n,t,\Phi_n\left(\frac{[tn]}{n}\right)\tilde{x}}(v)vdv = 0, \\
\int \tilde{q}_{n,t,\tilde{x}}(z) z_i z_j dz = \det \Phi_n\left(\frac{[tn]+1}{n}\right) \int q_{n,t,\Phi_n\left(\frac{[tn]}{n}\right)\tilde{x}}\left(\Phi_n\left(\frac{[tn]+1}{n}\right) z\right) z_i z_j dz \\
= \int q_{n,t,\Phi_n\left(\frac{[tn]}{n}\right)\tilde{x}}(v) \left[\Phi_n^{-1}\left(\frac{[tn]+1}{n}\right) v\right]_i \left[\Phi_n^{-1}\left(\frac{[tn]+1}{n}\right) v\right]_j dv \\
= \left\{\Phi_n^{-1}\left(\frac{[tn]+1}{n}\right) \int q_{n,t,\Phi_n\left(\frac{[tn]}{n}\right)\tilde{x}}(v) vv^{\mathrm{T}} dv \left[\Phi_n^{-1}\left(\frac{[tn]+1}{n}\right)\right]^{\mathrm{T}}\right\}_{ij} \\
= \left\{\Phi_n^{-1}\left(\frac{[tn]+1}{n}\right) a_n\left(t, \Phi_n\left(\frac{[tn]}{n}\right)\tilde{x}\right) \left[\Phi_n^{-1}\left(\frac{[tn]+1}{n}\right)\right]^{\mathrm{T}}\right\}_{ij} \triangleq \tilde{a} + n(t, \tilde{x}).
\end{gathered} \tag{16}$$

The vector function $\Phi_n(t)x$ at points $t = kh$, $k = 0, 1, 2, \ldots, n$, coincides with the Euler polyline for equations $y'(t) = b_n(t)y(t)$, $y(0) = x \in \mathrm{R}^{\mathrm{d}}$. Thus, in the case of a diffusion process the growing trend is compensated by returning along the trajectories of differential equations $y'(t) = b_n(t)y(t)$, $y(0) = x \in \mathrm{R}^{\mathrm{d}}$, and in case of a Markov chain, by returning along the Euler polylines of this equation. According to well-known properties of Euler polylines [17], $\Phi_n\left(\left[\frac{t}{h}\right]\right) x \to \Phi(t)$ uniformly on the interval $[0, 1]$, and, taking the above properties into account, we see that (12) and (16) imply that

$$\tilde{a}_n(t, \tilde{x}) = \int \tilde{q}_{n,t,x}(z) zz^{\mathrm{T}} \to \tilde{a}(t, x) = \tilde{\sigma}(t, x)[\tilde{\sigma}(t, x)]^{\mathrm{T}}, \quad n \to \infty.$$

Let us now show how statements obtained for models with a bounded trend can be transformed into the corresponding statements for models containing a linear component in the trend. For simplicity we consider the case of a family of densities $q_{n,t,x}(\cdot)$ independent of the n parameter, i.e.,

$q_{n,t,x}(\cdot) = q_{t,x}(\cdot)$. Suppose that the family $q_{t,x}(\cdot)$ and coefficients of Eqs. (10) satisfy conditions 1–4 of Section 3. Then for $\tilde{m}(t,x)$ and $\tilde{\sigma}(t,x)$ satisfy the conditions of Theorem 1.1 from [4], and, consequently, it holds that

$$\sup_{x,y\in \mathrm{R^d}} \left(1 + \|y - x\|^{2(S'-1)}\right) |p_{\tilde{X}_n}(0,1,x,y) - p_{\tilde{Y}}(0,1,x,y)| = O\left(\frac{1}{\sqrt{n}}\right). \tag{17}$$

Using relations (13) and (14), we formulate the following result which is a corollary of (17).

Theorem 2. *Suppose that conditions 1–4 of Section 3 are satisfied. Then*

$$\sup_{x,y\in \mathrm{R^d}} \left(1 + \|\Phi^{-1}(1)y - x\|^{2(S'-1)}\right) |\det \Phi_n(1)\, p_{X_n}(0,1,x,\Phi_n(1)\Phi^{-1}(1)y)$$
$$- \det \Phi(1)\, p_Y(0,1,x,y)| = O\left(\frac{1}{\sqrt{n}}\right).$$

If $b(t) \equiv b$ *then*

$$\sup_{x,y\in \mathrm{R^d}} \left(1 + \|\Phi^{-1}(1)y - x\|^{2(S'-1)}\right) |p_{X_n}(0,1,x,y) - p_Y(0,1,x,y)| = O\left(\frac{1}{\sqrt{n}}\right).$$

Proof of Theorem 2 is given in the Appendix.

Remark. Since

$$C_1\|y - \Phi(1)x\| \leqslant \|\Phi^{-1}(1)y - x\| \leqslant C_2\|y - \Phi(1)x\|,$$

statements of Theorem 2 can also be written with a factor $\left(1 + \|y - \Phi(1)x\|^{2(S'-1)}\right)$ instead of $\left(1 + \|\Phi^{-1}(1)y - x\|^{2(S'-1)}\right)$, i.e., a non-uniform estimate of the convergence rate in this theorem results either by shifting the terminal point y back (pull back) or by shifting the initial point x forward (push forward).

4. SAMPLE EXCLUSION OF THE LINEAR TREND COMPONENT FOR A DIFFUSION MODEL

Consider a model presented in [18]:

$$dX_t = \{\beta(t)(a(t) - X_t)\}dt + \sigma(t, X_t)dB_t, \quad X_0 = x \in \mathrm{R^d}, \tag{18}$$

where B_t, $t \geqslant 0$, is the standard Wiener process. We assume that function $\sigma(t, X_t)$ is bounded.

Consider a system of LUDE: $x'(t) = -\beta(t)x(t)$, $y(0) = x \in \mathrm{R^d}$ and its fundamental matrix $\Phi(t)$: $\Phi'(t) = -\beta(t)\Phi(t)$, $\Phi(0) = I$, where I is the unit matrix. We introduce the process $\tilde{X}_t = \Phi^{-1}(t)X_t$, and by Ito's lemma [19] we get the following stochastic differential for this process:

$$d\tilde{X}_t = d\left[\Phi^{-1}(t)X_t\right] = \Phi^{-1}(t)dX_t + d\Phi^{-1}(t)X_t$$
$$= \Phi^{-1}(t)\beta(t)a(t)dt + \Phi^{-1}(t)\sigma(t, X_t)dB_t = \tilde{m}(t, \tilde{X}_t)dt + \tilde{\sigma}(t, \tilde{X}_t)dB_t,$$

where $\tilde{m}(t, \tilde{X}_t) = \Phi^{-1}(t)\beta(t)a(t)$, $\tilde{\sigma}(t, \tilde{X}_t) = \Phi^{-1}(t)\sigma(t, \Phi(t)X_t)$.

Thus, process $\tilde{X}_t$ can be represented as

$$\tilde{X}_t = \tilde{X}_0 + \int_0^{\mathrm{T}} \tilde{m}(s, \tilde{X}_s)ds + \int_0^{\mathrm{T}} \tilde{\sigma}(s, \tilde{X}_s)dB_s$$
$$= x + \int_0^{\mathrm{T}} \Phi^{-1}(s)\beta(s)a(s)ds + \int_0^{\mathrm{T}} \Phi^{-1}(s)\sigma(s, \Phi(s)\tilde{X}_s)dB_s.$$

Consider the one-dimensional case with constant coefficients: $a(t) \equiv a$, $\beta(t) \equiv \beta$, $\sigma(t) \equiv \sigma$. Then model (18) corresponds to Vasicek's model of interest rate evolution [20]:

$$dX_t = \{\alpha\beta - \beta X_t\}dt + \sigma dB_t, \quad X_0 = x \in \mathrm{R}. \tag{19}$$

For model (19), SDE solutions can be found explicitly, so the trend exclusion procedure for the diffusion equations is not especially interesting, but it allows us to check the correctness of this method.

A solution of SDE (19) can be represented as [20]

$$X_t = x\ \exp(-\beta t) + a(1 - \exp(-\beta t)) + \sigma\ \exp(-\beta t) \int\limits_0^{\mathrm{T}} \exp(\beta s)dB_s. \tag{20}$$

Let us now show how to solve SDE (19) with the linear trend component exclusion procedure.

The fundamental matrix for SDE (19) has the form $\Phi(t) = \exp(-\beta t)$. Then process $\tilde{X}_t = \Phi^{-1}(t)X_t = \exp(\beta t)X_t$ can be represented as

$$\tilde{X}_t = x + a(\exp(\beta t) - 1) + \sigma \int\limits_0^{\mathrm{T}} \exp(\beta s)dB_s.$$

Solution found by linear trend component exclusion now takes the form

$$\begin{aligned} X_t &= \Phi(t)\left(x + a(\exp(\beta t) - 1) + \sigma \int\limits_0^{\mathrm{T}} \exp(\beta s)dB_s\right) \\ &= x\ \exp(-\beta t) + a(1 - \exp(-\beta t)) + \sigma\ \exp(-\beta t) \int\limits_0^{\mathrm{T}} \exp(\beta s)dB_s, \end{aligned}$$

which coincides with (20).

The linear trend component exclusion method is not applicable to more complex methods such as the Cox–Ingersoll–Ross modified interest rate evolution model [21] or its extension, the Hull–White model [22], since a modification of the parametrix method for unbounded diffusion has not yet been studied. The Cox–Ingersoll–Ross model differs from Vasicek's model above in the form of its volatility functions $\sigma(t, X_t) \equiv \sigma\sqrt{X_t}$, and the trend removal procedure for this model is a separate problem. The Hull–White model also contains unbounded diffusion $\sigma(t, X_t) \equiv \sigma(t)\sqrt{X_t}$, so the linear trend exclusion procedure for this model requires further study.

However, apart from Vasicek's interest rate model, the trend exclusion method can be applied to Heston's stochastic volatility model [23] in case when the function of volatility occurring in the return on investment equation is bounded. Consider a two-dimensional case with the example of Heston's stochastic volatility model [23]:

$$\begin{cases} dS_t = \mu S_t dt + f(v_t, S_t)dB_t^1 \\ dv_t = k(\theta - v_t)dt + \xi g(v_t)dB_t^2, \end{cases} \tag{21}$$

where $B_t = (B_t^1, B_t^2)$ is the standard Wiener process, $f(v_t, S_t)$ and $g(v_t)$ are bounded functions.

We represent the system of SDEs [21] in matrix form:

$$\begin{pmatrix} dS_t \\ dv_t \end{pmatrix} = \left[\begin{pmatrix} \mu & 0 \\ 0 & -k \end{pmatrix}\begin{pmatrix} S_t \\ v_t \end{pmatrix} + \begin{pmatrix} 0 \\ k\theta \end{pmatrix}\right]dt + \begin{pmatrix} f(v_t, S_t) & 0 \\ 0 & \xi g(v_t) \end{pmatrix}\begin{pmatrix} dB_t^1 \\ dB_t^2 \end{pmatrix}.$$

Consider a system of LUDEs:

$$x'(t) = \begin{pmatrix} \mu & 0 \\ 0 & -k \end{pmatrix} x(t).$$

The fundamental matrix for this system has the form

$$\Phi(t) = \begin{pmatrix} \exp(\mu t) & 0 \\ 0 & \exp(-kt) \end{pmatrix}.$$

Then process

$$\begin{pmatrix} \tilde{S}_t \\ \tilde{v}_t \end{pmatrix} = \begin{pmatrix} \exp(-\mu t) & 0 \\ 0 & \exp(kt) \end{pmatrix} \begin{pmatrix} S_t \\ v_t \end{pmatrix}$$

can be represented as

$$\begin{pmatrix} \tilde{S}_t \\ \tilde{v}_t \end{pmatrix} = \begin{pmatrix} S_0 \\ v_0 \end{pmatrix} + \begin{pmatrix} 0 \\ \theta(\exp(kt) - 1) \end{pmatrix} + \int_0^{\mathrm{T}} \begin{pmatrix} \exp(-\mu s) & 0 \\ 0 & \exp(ks) \end{pmatrix} \begin{pmatrix} f(v_s, S_s) & 0 \\ 0 & \xi g(v_s) \end{pmatrix} dB_s.$$

By excluding the linear trend component, we get the following representation for the solutions of this system of SDEs:

$$\begin{pmatrix} S_t \\ v_t \end{pmatrix} = \begin{pmatrix} \exp(\mu t) & 0 \\ 0 & \exp(-kt) \end{pmatrix} \begin{pmatrix} S_0 \\ v_0 \end{pmatrix} + \begin{pmatrix} 0 \\ \theta(1 - \exp(-kt)) \end{pmatrix}$$
$$+ \begin{pmatrix} \exp(\mu t) & 0 \\ 0 & \exp(-kt) \end{pmatrix} \int_0^{\mathrm{T}} \begin{pmatrix} \exp(-\mu s) & 0 \\ 0 & \exp(ks) \end{pmatrix} \begin{pmatrix} f(v_s, S_s) & 0 \\ 0 & \xi g(v_s) \end{pmatrix} dB_s.$$

5. CONCLUSION

It is known [11, 24] that the parametrix method and its discrete counterpart [4, 5] assume that the shift and diffusion coefficients are bounded. At the same time, many important models have unbounded shift coefficients, including models corresponding to stochastic recurrent estimation procedures that have a linearly growing shift coefficient.

Namely, Markov chains and limit diffusion processes with a linear trend component arise in recurrent estimation procedures based on the Robbins–Monroe method. The work [25] proves a number of results on the weak convergence of recurrent estimation procedures to finite-dimensional distributions of some limit diffusion process (Theorem 6.3, Chap. 6; Theorems 3.1 and 5.2, Chap. 8). These theorems assume the existence of densities, so a natural question arises: do, in addition to weak convergence, the densities also converge, i.e., does the corresponding local limit theorem hold? The parametrix method combined with the trend exclusion method let us answer this question positively. This application of our approach will be the subject of a separate publication; the purpose of this work was to propose a procedure that lets one exclude linearly growing trend components and reduce the problem to a previously studied problem with bounded trend. Our method is also applicable to more general models with trends that have bounded gradients, but the formulas are less explicit and Euler polylines are constructed locally, so in this case we can speak of the local limit theorem over small time. This will also be a subject of further study.

ACKNOWLEDGMENTS

This work was supported by the Scientific Foundation of the National Research University Higher School of Economics.

APPENDIX

Proof of Theorem 2. Applying (17) to points x and $\Phi^{-1}(1)y$, we get

$$\sup_{x,y\in\mathrm{R}^{\mathrm{d}}}\left(1+\|\Phi^{-1}(1)y-x\|^{2(S'-1)}\right)|p_{\tilde{X}_n}(0,1,x,\Phi^{-1}(1)y)-p_{\tilde{Y}}(0,1,x,\Phi^{-1}(1)y)|=O\left(\frac{1}{\sqrt{n}}\right). \quad (A.1)$$

From (13) and (14) we have that

$$p_{\tilde{Y}}(s,t,x,z)=\det\Phi(t)p_Y(s,t,\Phi(s)x,\Phi(t)z), \quad (A.2)$$

$$p_{\tilde{X}_n}(ih,jh,x,v)=\det\Phi_n(jh)p_{X_n}(ih,jh,\Phi_n(ih)x,\Phi_n(jh)v). \quad (A.3)$$

Substituting into (A.2) and (A.3) $s=0,\ t=1,\ i=0,\ j=n,\ x,$ and $z=\Phi^{-1}(1)y,\ v=\Phi^{-1}(y)$, we get from (A.1) that

$$\sup_{x,y\in\mathrm{R}^{\mathrm{d}}}\left(1+\|\Phi^{-1}(1)y-x\|^{2(S'-1)}\right)|\det\Phi_n(1)\ p_{X_n}(0,1,x,\Phi_n(1)\Phi^{-1}(1)y)$$
$$-\det\Phi(1)\ p_Y(0,1,x,y)|=O\left(\frac{1}{\sqrt{n}}\right). \quad (A.4)$$

Let $b(t)\equiv b$. Then $\Phi(1)=e^b$ is the matrix exponent, $\Phi_n(1)=\left(I+\frac{b}{n}\right)^n$ for sufficiently large n and

$$\|\Phi(1)-\Phi_n(1)\|\leqslant e^a-\left(1+\frac{a}{n}\right)^n\leqslant\frac{a^2e^a}{n}, \quad (A.5)$$

where $a=\|b\|$. The first inequality in (A.5) has been proven in [17, p. 98]. To show the second inequality, it suffices to study the sign of the derivative near the point $x=0$ for the function $f(x)\triangleq a^2e^ax-e^a+(1+ax)^{1/x}$. Besides, Lemmas 3.1 and 3.2 from [4] imply the estimate

$$p_{\tilde{Y}}(s,t,x,y)\leqslant Ce^{-C\|y-x\|^2},$$

so for the transition density $p_Y(0,1,x,y)$ we have

$$p_Y(0,1,x,y)=\det\Phi^{-1}p_{\tilde{Y}}(0,1,x,\Phi^{-1}(1)y)\leqslant Ce^{-C\|\Phi^{-1}(1)y-x\|^2}. \quad (A.6)$$

The second statement of Theorem 2 now follows from (A.4)–(A.6).

REFERENCES

1. Skorokhod, A.V., *Issledovaniya po teorii sluchainykh protsessov* (Studies in the Theory of Random Processes), Kiev: Kiev. Univ., 1961.

2. Stroock, D.W. and Varadhan, S.R.S., *Multidimensional Diffusion Processes*, Grundlehren der Mathematischen Wissenschaft, vol. 233, Berlin: Springer, 1979.

3. Konakov, V. and Molchanov, S., On the Convergence of Markov Chains to Diffusion Processes, *Theory Probab. Math. Stat.*, 1984, vol. 31, pp. 51–64.

4. Konakov, V. and Mammen, E., Local Limit Theorems for Transition Densities of Markov Chains Converging to Diffusions, *Probab. Theory Relat. Fields*, 2000, no. 117, pp. 551–587.

5. Konakov, V. and Mammen, E., Local Approximations of Markov Random Walks by Diffusions, *Stoch. Proc. Appl.*, 2001, no. 96 (1), pp. 73–98.

6. Konakov, V. and Mammen, E., Edgeworth Type Expansions for Transition Densities of Markov Chains Converging to Diffusions, *Bernoulli: J. Math. Statist. Probab.*, 2005, vol. 11, no. 4, pp. 591–641.

7. Konakov, V. and Mammen, E., Small Time Edgeworth-Type Expansions for Weakly Convergent Nonhomogeneous Markov Chains, *Probab. Theory Relat. Fields*, 2009, vol. 143, no. 1, pp. 137–176.

8. Levy, E.E., Sulle equazioni lineari alle derivate parziali totalmente ellittiche, *Rendiconti della Reale Accademia dei Lincei, Classe di Scienze Fisiche, Matematiche, Naturali*, 1907, vol. 16, no. 12, pp. 932–938.

9. Levy, E.E., Sulle equazioni lineari totalmente ellittiche alle derivate parzial, *Rendiconti del Circolo Matematico di Palermo*, 1907, vol. 24, no. 1, pp. 275–317.

10. Friedman, A., *Partial Differential Equations of Parabolic Type*, Englewood Cliffs: Prentice-Hall, 1964.

11. Il'in, A.M., Kalashnikov, A.S., and Oleinik, O.A., Second Order Linear Equations of Parabolic Type, *Usp. Mat. Nauk*, 1962, vol. XVII, no. 3 (105), pp. 3–146.

12. McKean, H.P. and Singer, I.M., Curvature and the Eigenvalues of the Laplacian, *J. Differ. Geometry*, 1967, no. 1, pp. 43–69.

13. Delarue, F. and Menozzi, S., Density Estimation for a Random Noise Propagating Through a Chain of Differential Equations, *J. Func. Anal.*, 2010, vol. 259, no. 6, pp. 1577–1630.

14. Kelley, W. and Peterson, A., *The Theory of Differential Equations: Classical and Qualitative*, Englewood Cliffs: Prentice Hall, 2004.

15. Nualart, D., *The Malliavin Calculus and Related Topics*, New York: Springer, 2006.

16. Elaydi, S., *An Introduction to Difference Equations*, New York: Springer Science+Business Media, 2005.

17. Arnol'd, V.I., *Obyknovennye differentsial'nye uranveniya* (Ordinary Differential Equations), Moscow: Nauka, 1971.

18. Koo, B. and Linton, O., *Semiparametric Estimation of Locally Stationary Diffusion Models*, Discussion paper no. EM/2010/551, August 2010.

19. Øksendal, B., *Stochastic Differential Equations. An Introduction with Applications*, New York: Springer, 2003. Translated under the title *Stokhasticheskie differentsial'nye uravneniya. Vvedenie v teoriyu i prilozheniya*, Moscow: Mir, 2003.

20. Vasicek, O., An Equilibrium Characterization of the Term Structure, *J. Finan. Econom.*, 1977, no. 5 (2), pp. 177–188.

21. Cox, J.C., Ingersoll, J.E., and Ross, S.A., A Theory of the Term Structure of Interest Rates, *Econometrica*, 1985, no. 53, pp. 385–407.

22. Hull, J. and White, A., Pricing Interest-Rate-Derivative Securities, *Rev. Finan. Stud.*, 1990, vol. 3, no. 4, pp. 573–592.

23. Heston, S.L., A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Rev. Finan. Stud.*, 1993, vol. 6, no. 2, pp. 327–343.

24. Konakov, V.D., The Parametrix Method for Diffusions and Markov Chains, *Preprint of Math. Mech. Dept., MSU*, 2012, no. 2012, 2012.

25. Nevel'son, M.B. and Khas'minskii, R.Z., *Stokhasticheskaya approksimatsiya i rekurrentnoe otsenivanie* (Stochastic Approximation and Recurrent Estimation), Moscow: Nauka, 1972.

This paper was recommended for publication by A.V. Nazin, a member of the Editorial Board