Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 9(7): 19015–19030.
DOI: 10.3934/math.2024925
Received: 25 April 2024
Revised: 24 May 2024
Accepted: 28 May 2024
Published: 06 June 2024

Research article

Strong consistency of the nonparametric kernel estimator of the transition density for the second-order diffusion process

Li Yue and Wang Yunyan*

School of Science, Jiangxi University of Science and Technology, Ganzhou, China

* **Correspondence:** Email: yywang@jxust.edu.cn; Tel: +8615216127128.

Abstract: The integrals of diffusion processes are of significant importance in the field of finance, particularly in relation to stochastic volatility models, which are frequently employed to represent the temporal variability of stock prices. In this paper, we consider the strong consistency of the nonparametric kernel estimator of the transition density for second-order diffusion processes, using the moment inequalities of ρ-mixing sequences to demonstrate the strong consistency under some regularity conditions. Furthermore, the asymptotic mean square error is provided based on the deviation and variance of the transition density kernel estimator. The optimal bandwidth is found and thus the convergence rate of the kernel estimator is obtained. At the same time, our results are compared with the conclusions of the univariate density function.

Keywords: kernel estimator; transition density; strong consistency; second-order diffusion process; convergence rate; moment inequality
Mathematics Subject Classification: 62G05, 62G20

1. Introduction

In this study, we examine the second-order diffusion process, which is defined by a stochastic differential equation below:

$$\begin{cases} dY_t = X_t dt, \\ dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \end{cases} \tag{1.1}$$

where $\{B_t, t \geq 0\}$ is a standard Brownian motion, $\mu(\cdot)$ and $\sigma(\cdot)$ are the drift coefficient and diffusion coefficient, respectively. $\{X_t\}$ is often assumed to be a stationary, continuous-time, one-dimensional diffusion process as follows:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t. \tag{1.2}$$

Although numerous popular models in econometrics and finance consider the diffusion processes (1.2) to be solutions to stochastic differential equations, such as interest rates, stock prices, and exchange rates, this process has inherent limitations in its application.

In recent years, statistical research based on high-frequency data has attracted the attention of many scholars; see [1–3]. Most of the previous studies on diffusion models have assumed that the data are stationary. However, in practice, many stochastic processes in economics and finance exhibit the accumulation of all past perturbations and are regarded as non-stationary integral stochastic processes. To illustrate, in a discrete-time context, the unit-root model $\{y_t : t = 0, 1, 2, \cdots\}$ is a prototypical example, where $y_t = \alpha + y_{t-1} + \varepsilon_t(\varepsilon_t \sim i.i.d.N(0, 1))$. Notice that the y_t can be expressed as

$$y_t = y_0 + \sum_{k=1}^{t} x_k,$$

$x_t = \alpha + \varepsilon_t$. Obviously, the process $\{y_t : t = 0, 1, 2, \cdots\}$ is not stationary and thus requires modeling using a difference equation (e.g., $\Delta y_t = \alpha + \varepsilon_t$) to transform it into a stationary process. Just as Nicolau [4] says, all sample functions of a diffusion process, driven by a Brownian motion, are of unbounded variation and nowhere differentiable. Therefore, it is not feasible to model integrated and differentiated diffusion processes via the model (1.2).

Fortunately, in the context of the continuous case, the model (1.1) may be considered. Since the Y in the model is a differentiable process, defined in an integral form as

$$Y_t = Y_0 + \int_0^t X_u du.$$

Clearly, the second-order diffusion process is similar to the unit root model in the discrete setting, which can be made stationary by differencing, which addresses the limitation of the diffusion process in modeling differentiable stochastic processes.

Notably, unlike model (1.2), the estimation of model (1.1) raises new challenges. Firstly, current financial dynamic data typically shows the accumulation of all past disturbances. It is common to obtain the observations $\{Y_{i\Delta}, i = 1, 2, \cdots\}$ for model (1.1) rather than $\{X_{i\Delta}, i = 1, 2, \cdots\}$. Nevertheless, the conditional distribution of Y is typically unidentified, even in the context of a known distribution for X. So we cannot construct estimators based on $\{Y_{i\Delta}, i = 1, 2, \cdots\}$. Secondly, we cannot compute the value of X_{t_i} from $Y_{t_i} = Y_0 + \int_0^{t_i} X_u du$ for a fixed sampling interval. To this end, by discrete-time observations $\{Y_{i\Delta}, i = 1, 2, \cdots\}$ and given that

$$Y_{i\Delta} - Y_{(i-1)\Delta} = \int_0^{i\Delta} X_u du - \int_0^{(i-1)\Delta} X_u du = \int_{(i-1)\Delta}^{i\Delta} X_u du,$$

it is possible to obtain an approximation value for X at instants $\{t_i = i\Delta, i = 1, 2, \cdots\}$ (where $\Delta = t_i - t_{i-1}$) by the following formula:

$$\tilde{X}_{i\Delta} = \frac{Y_{i\Delta} - Y_{(i-1)\Delta}}{\Delta}, i = 1, 2, \cdots .$$

The degree of accuracy of $\tilde{X}_{i\Delta}$ is contingent upon the size of the Δ. For more details on model (1.1), please refer to [4–6].

The second-order diffusion process, or the integrated diffusion process, as it is also known, is the integral of the diffusion process. It plays a crucial role in finance, particularly in connection with stochastic volatility models; see [7–9]. Moreover, the second-order diffusion process is also used in physics, such as to model particle velocities on liquid surfaces [10] and the ice-core data [11]. The statistical inference is the premise of its applications, so many authors have studied the non-parametric statistical inference of the continuous-time second-order diffusion model based on high-frequency sampling. For the nonparametric estimation of the coefficients $\mu(\cdot)$ and $\sigma(\cdot)$, Nicolau [4] pioneered the development of the Nadaraya-Watson estimators. On this basis, Wang and Lin [12] constructed the corresponding local linear estimators, and Wang et al. [13] introduced of the innovative re-weighted estimators of the coefficients. Subsequently, Tang et al. [14] took into account the nonparametric deviation correction of the diffusion coefficient, and Yang et al. [15] proved that the nonparametric kernel estimators in [4] meet strong consistency.

The transition density function, as a crucial variable that reflects the distinctive characteristics of second-order diffusion models, has been a subject of investigation for scholars for a considerable period. It is useful to calculate the dynamic characteristics of fundamental variables and help us solve practical problems such as financial asset pricing and financial portfolio selection. Consequently, the estimation of the transition density function is a meaningful research topic. For the study of the strong consistency of the transition density estimator, Zhao and Liu [16] proposed a double kernel estimator of the conditional density function and studied its strong consistency. Then the strong convergence rate of this estimator was then investigated in [17] under certain mild conditions. Khardani and Semmar [18] investigated the transition density estimation in the context of response variables subject to censoring and derived the strong consistency of the estimator. Benkhaled et al. [19] studied local linear estimation of the transition density of a randomly censored scalar response variable in the context of a functional random covariate and established almost sure convergence under α-mixing dependence. The strong uniform consistency rate of the transition density estimator in the single functional index model was discussed in [20]. For other studies on strong consistency, see [21, 22].

Li et al. [23] initially developed the transition density estimation for the second-order diffusion process, proposing a kernel estimator that was subsequently shown to have weak consistency and be asymptotically normal. However, a corresponding demonstration of strong consistency has not yet been provided. In this paper, we wish to study the strong consistency of the nonparametric kernel estimator proposed in [23]. Strong consistent estimation has good convergence speed and stability; therefore, it has high reliability in practical applications. It is suitable for parameter estimation in large sample situations and can provide more accurate results.

The rest of this paper is organized into the following sections: In Section 2, we provide some assumptions for the main results. Section 3 gives the strong consistency theorem and finds the optimal bandwidth and convergence rate. Section 4 provides the proofs of the main results and the necessary lemmas. Section 5 offers a simulation example comparing the kernel estimator values with the exact transition density solutions for the second-order diffusion process. Finally, Section 6 concludes the paper.

2. Assumptions

Let $D = (l, r)\,(-\infty \le l < r \le +\infty)$ be the state space of stationary process X, and $p(y\,|x)$ be the transition density of $y = X_t$ at a given $x = X_s$ for model (1.1), $x, y \in D$. According to [23], for $\forall x, y \in D$, the kernel estimator of the transition density for the second-order diffusion process is defined as:

$$\hat{p}(y\,|x) = \frac{\sum_{i=1}^{n} K_h(x - \tilde{X}_{i\Delta}) K_h(y - \tilde{X}_{(i+1)\Delta})}{\sum_{i=1}^{n} K_h(x - \tilde{X}_{i\Delta})},$$

where $K(\cdot)$ is a symmetric density function on $\mathbb{R}$, h is defined as bandwidth, and $K_h(\cdot) = K(\cdot/h)/h$.

For $\forall x_0, x \in D$, let $s(x) = \exp\left\{-\int_{x_0}^{x} \frac{2\mu(u)}{\sigma^2(u)} du\right\}$ be the scale density function and $m(x) = (\sigma^2(x)s(x))^{-1}$ be the speed density function. The following assumptions will be used throughout the paper:

Assumption 1. *(i) The infinitesimal coefficients $\mu(\cdot)$ and $\sigma(\cdot)$ are time-aligned ß-measured functions on $D = (l, r)$ that have at least continuous second-order derivatives, where ß is the σ-fields generated by the Borel set on $D = (l, r)$. $\mu(\cdot)$ and $\sigma(\cdot)$ satisfy the local linear increasing condition and the local Lipschitz condition, namely, for an arbitrary compact subset $J \in (l, r)$, there are positive constants C_1, C_2 such that for $\forall x, y \in J$ have*

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \le C_1 \,|x - y|$$

and

$$|\mu(x)| + |\sigma(x)| \le C_2\,(1 + |x|)\,.$$

(ii) $\sigma^2\,(x) > 0$ for $x \in (l, r)$.

Remark 2.1. *By Theorem 5.5.15 of [24], Assumption 1 ensures that the stochastic differential equation has a solution that is uniquely strong.*

Assumption 2. $\int_l^x s(u)du = \int_x^r s(u)du = \infty$, $\int_l^r m(x)dx < \infty$.

Assumption 3. *$X_0 = x$ has an invariant distribution P^0.*

Assumption 4. $\limsup\limits_{x\to r} \left(\frac{\mu(x)}{\sigma(x)} - \frac{\sigma'(x)}{2}\right) < 0$, $\limsup\limits_{x\to l} \left(\frac{\mu(x)}{\sigma(x)} - \frac{\sigma'(x)}{2}\right) > 0$.

Remark 2.2. *The Assumptions 2–4 make sure that the $\left\{\tilde{X}_{i\Delta}\right\}$ is ergodic, stationary, and ρ-mixing [4].*

Assumption 5. *(i) The marginal density function $p(x)$ is a positive, continuous, and stationary function that possesses a continuous first derivative.*
(ii) The joint density function $p(x, y)$ is bounded by an independent constant.
(iii) The transition density function $p(y\,|x)$ may be considered a bounded, continuous function. It also has a continuous second-order partial derivative with respect to x and y, respectively.

Assumption 6. *The Kernel function, denoted $K(\cdot)$, is a bounded, symmetric, and continuously differentiable function on $\mathbb{R}$, and it satisfies the following properties: $\int K\,(u)\,du = 1$, $\int uK(u)du = 0$, $\int u^2K(u)du = K_1 < \infty$, $|u|\,K(u) \to 0(|u| \to \infty)$, $\int |K(u)|\,du < \infty$, $\int K^2(u)du = K_2 < \infty$.*

Assumption 7. $|K'(u)| \leq C$, $E\left|K'(\xi_{n,i})\right|^m = O(h)$, *where* $\xi_{n,i} = \theta((x - X_{i\Delta})/h) + (1-\theta)((x - \tilde{X}_{i\Delta})\big/h)$, $0 \leq \theta \leq 1$, *h is defined as a bandwidth and m is a positive integer.*

Assumption 8. *(i)* $\Delta \to 0$, $h \to 0$, $nh^2 \to \infty$ *as* $n \to \infty$.
(ii) $\Delta \log(1/\Delta)\big/h^4 \to 0$, $\Delta \log(1/\Delta)\big/h^2 \to 0$ *as* $n \to \infty$.

3. Main results

The asymptotic result can be expressed as follows:

Theorem 3.1. *Under Assumptions 1-8, for* $\forall x, y \in D$, *if there exists* $\tau > 0$ *such that*

$$\frac{\Delta \log(1/\Delta)}{nh^5} = O(n^{-\tau}),$$

then

$$\hat{p}(y|x) - p(y|x) = o_{a.s.}(1).$$

Remark 3.1. *It is well established that the estimator of the transition density is closely related to the choice of the kernel function and the bandwidth. A multitude of studies have demonstrated that the selection of kernel function is no longer a particularly significant factor; conversely, we must choose an appropriate bandwidth. Asymptotic analysis offers a straightforward approach for identifying the optimal bandwidth* h^*. *The most common method is to minimize the asymptotic mean square error (AMSE). Following, we take the estimator* $\hat{p}(y\,|x)$ *for example. [23] has given the asymptotic bias and variance of* $\hat{p}(y\,|x)$*:*

$$Bias\,(\hat{p}(y\,|x)) = \frac{1}{2}K_1h^2\left(\frac{\partial^2 p(y\,|x)}{\partial x^2} + 2\frac{p'(x)}{p(x)}\frac{\partial p(y\,|x)}{\partial x} + \frac{\partial^2 p(y\,|x)}{\partial y^2}\right),$$

$$Var(\hat{p}(y|x))=\frac{K_2^2 p(y|x)}{nh^2 p(x)}.$$

So we can easily write its AMSE:

$$\frac{1}{4}K_1^2h^4\left(\frac{\partial^2 p(y\,|x)}{\partial x^2} + 2\frac{p'(x)}{p(x)}\frac{\partial p(y\,|x)}{\partial x} + \frac{\partial^2 p(y\,|x)}{\partial y^2}\right)^2 + \frac{K_2^2 p(y|x)}{nh^2 p(x)}, \tag{3.1}$$

the derivative of the AMSE with respect to the parameter h is given by the expression

$$K_1^2h^3\left(\frac{\partial^2 p(y\,|x)}{\partial x^2} + 2\frac{p'(x)}{p(x)}\frac{\partial p(y\,|x)}{\partial x} + \frac{\partial^2 p(y\,|x)}{\partial y^2}\right)^2 - \frac{2K_2^2 p(y|x)}{nh^3 p(x)}. \tag{3.2}$$

By setting the formula to zero, the optimal bandwidth may then be determined

$$h^* = An^{-1/6}, \tag{3.3}$$

where $A = \left(\frac{2K_2^2 p(y|x)}{Bp(x)K_1^2}\right)^{\frac{1}{6}}$, $B = \left(\frac{\partial^2 p(y|x)}{\partial x^2} + 2\frac{p'(x)}{p(x)}\frac{\partial p(y|x)}{\partial x} + \frac{\partial^2 p(y|x)}{\partial y^2}\right)^2$. *In the (3.3), A is a function of unknown marginal density* $p(x)$ *and transition density* $p(y\,|x)$, *so* h^* *can not be obtained directly. Silverman [25]* *considered a rules of thumb method that replaces the unknown function in A with a consistent estimator to obtain a feasible, approximate optimal bandwidth $h_{\rm opt} = 1.06Sn^{-1/5}$, where S is the sample standard deviation. In this paper, the estimation uses the smoothing parameter h in both x and y directions, then the approximate optimal bandwidth*

$$h^*_{\rm opt} = 1.06Sn^{-1/6}.$$

Remark 3.2. *Substituting the h^* for h in (3.1), we can get*

$$\begin{aligned} AMSE^* &= \frac{1}{4}K_1^2\left(\frac{2K_2^2p(y|x)}{Bp(x)K_1^2n}\right)^{2/3}B + \frac{K_2^2p(y|x)}{n\left(\frac{2K_2^2p(y|x)}{Bp(x)K_1^2n}\right)^{1/3}p(x)} \\ &= \sqrt{2}\left(\frac{K_2^2K_1p(y|x)\sqrt{B}}{p(x)}\right)^{2/3}n^{-2/3} + \sqrt[3]{2}\left(\frac{K_2^2K_1p(y|x)\sqrt{B}}{p(x)}\right)^{2/3}n^{-2/3} \\ &= \left(\sqrt{2}+\sqrt[3]{2}\right)C^{2/3}n^{-2/3}, \end{aligned}$$

where $B = \left(\frac{\partial^2 p(y|x)}{\partial x^2} + 2\frac{p'(x)}{p(x)}\frac{\partial p(y|x)}{\partial x} + \frac{\partial^2 p(y|x)}{\partial y^2}\right)^2$, $C = \frac{K_2^2K_1p(y|x)\sqrt{B}}{p(x)}$. The convergence rate of the AMSE is of order $n^{-2/3}$, which is the same as the convergence rate of the kernel estimator obtained in [26].

Remark 3.3. *In comparison with the results obtained for a univariate kernel density estimator, the convergence rate of which is of order $n^{-4/5}$ (see [27]), it reveals that the convergence properties are superior in the univariate case, as a smaller sample size is required for estimation when the value of X is given.*

4. Lemmas and proofs

4.1. Lemmas

Lemma 4.1. *[28] Let $K(u)$ and $g(x)$ are the Borel-measured functions defined in $\mathbb{R}$ and satisfy the following conditions: (i) $\sup_{-\infty<u<\infty}|K(u)| < \infty$; (ii) $\int_{-\infty}^{\infty}|K(u)|\,du < \infty$; (iii) $\lim_{u\to\infty}|uK(u)| = 0$; (iv) $\int_{-\infty}^{\infty}|g(x)|\,dx < \infty$. Define*

$$g_n(x) = \frac{1}{h(n)}\int_{-\infty}^{\infty}K\left(\frac{u}{h(n)}\right)g(x-u)du,$$

where $h(n)$ is a sequence of positive constants that satisfy $\lim_{n\to\infty}h(n)\to 0$. Subsequently, for each point x of continuity of $g(x)$,

$$\lim_{n\to\infty}g_n(x) = g(x)\int_{-\infty}^{\infty}K(u)\,du.$$

Lemma 4.2. *[29] Suppose that $\{X_i : i > 1\}$ is a ρ-mixing sequence of random variables with the mixing coefficient $\rho(n) = O\left(n^{-\theta}\right)$ for some constant $\theta > 0$. If $EX_i = 0$ and $E|X_i|^r < \infty$ (where $r > 1$), then for any given $m \geq 1$, there exists a positive constant $C = C(m)$ such that*

$$E\left|\sum_{j=1}^{n}X_j\right|^r \leq Cn^{\delta(m)}\sum_{j=1}^{n}E\left|X_j\right|^r (1 < r \leq 2),$$

$$E\left|\sum_{j=1}^{n} X_j\right|^r \leq C n^{\delta(m)} \left\{ \sum_{j=1}^{n} E\left|X_j\right|^r + \left(\sum_{j=1}^{n} E\left(X_j^2\right) \right)^{r/2} \right\} (r > 2),$$

where $\delta(m) = (r-1)\alpha^m$ *and* $0 < \alpha \leq 1$.

Lemma 4.3. *[1, 24]*

$$P\left(\limsup_{n\to 0} \frac{k_n}{(\Delta \log(1/\Delta))^{\frac{1}{2}}} = k_0\right) = 1,$$

where k_0 *is a constant,*

$$k_n = \max_{1\leq i\leq n} \sup_{(i-1)\Delta\leq s\leq i\Delta} |X_s - X_{i\Delta}|.$$

Remark 4.1. *Let* $\lambda_n = (\Delta \log(1/\Delta))^{1/2}$. *According to Lemma 4.3, we have*

$$\max_{1\leq i\leq n} \left|\tilde{X}_{i\Delta} - X_{i\Delta}\right| = O_{a.s.}(\lambda_n).$$

Since

$$\max_{1\leq i\leq n} \left|\tilde{X}_{i\Delta} - X_{i\Delta}\right| = \max_{1\leq i\leq n} \left|\frac{1}{\Delta}\int_{(i-1)\Delta}^{i\Delta} X_s ds - X_{i\Delta}\right| \leq \frac{1}{\Delta}\max_{1\leq i\leq n} \int_{(i-1)\Delta}^{i\Delta} |X_s - X_{i\Delta}| ds.$$

Lemma 4.4. *For* $\forall x, y \in D$, *let*

$$\tilde{p}(x,y) = \frac{1}{nh^2}\sum_{i=1}^{n} K(\frac{x - X_{i\Delta}}{h})K(\frac{y - X_{(i+1)\Delta}}{h}),$$

$$\tilde{p}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_{i\Delta}}{h}),$$

$\tilde{p}(y|x) = \frac{\tilde{p}(x,y)}{\tilde{p}(x)}$ *is the kernel estimator of the transition density for model (1.2), where* $K(\cdot)$ *is a symmetric density function on* $\mathbb{R}$ *and* h *is bandwidth. Under Assumptions 1–6, Assumptions 8, if there exists* $\tau > 0$ *such that*

$$\frac{1}{nh_n^3} = O(n^{-\tau}),$$

then

$$\tilde{p}(y|x) - p(y|x) = o_{a.s.}(1).$$

Lemma 4.5. *[15] Suppose Assumptions 1–8 hold. If there exists* $\tau > 0$ *such that*

$$\frac{1}{nh_n} = O(n^{-\tau}),$$

then

$$\hat{p}(x) - p(x) = o_{a.s.}(1).$$

4.2. Proofs

The proof of Lemma 4.4. Since $\frac{1/(nh)}{1/(nh^3)} = h^2 \to 0$, by Lemma 5.5 of [15], we have

$$\tilde{p}(x) - p(x) = o_{a.s.}(1). \tag{4.1}$$

Furthermore, it is obvious that

$$\tilde{p}(y|x) - p(y|x) = \frac{\tilde{p}(x,y) - p(y|x)p(x) - [p(y|x)(\tilde{p}(x) - p(x))]}{\tilde{p}(x)}.$$

Thus, we need to prove $\tilde{p}(x,y) - p(y|x)p(x) = o_{a.s.}(1)$, which is equivalent to proving

$$\tilde{p}(x,y) - E\tilde{p}(x,y) = o_{a.s.}(1), \tag{4.2}$$

$$E\tilde{p}(x,y) \to p(x,y). \tag{4.3}$$

Firstly, we prove Eq (4.2). Let

$$G_{1,i} = K(\frac{x - X_{i\Delta}}{h})K(\frac{y - X_{(i+1)\Delta}}{h}) - E\left[K(\frac{x - X_{i\Delta}}{h})K(\frac{y - X_{(i+1)\Delta}}{h})\right],$$

clearly, $EG_{1,i} = 0$ and $\tilde{p}(x,y) - E\tilde{p}(x,y) = \frac{1}{nh^2}\sum_{i=1}^{n} G_{1,i}$. For $\forall s \geq 2$, from Lemma 4.1,

$$\frac{1}{h}E\left|K(\frac{x - X_{i\Delta}}{h})\right|^s = \frac{1}{h}\int_R \left|K\left(\frac{u}{h}\right)\right|^s p(x-u)du \to p(x)\int_R |K(u)|^s du < \infty.$$

So by Cr inequality and *Hölder* inequality,

$$\begin{aligned}
E\left|G_{1,i}\right|^s &\leq 2^s E\left|K(\frac{x - X_{i\Delta}}{h})K(\frac{y - X_{(i+1)\Delta}}{h})\right|^s \\
&\leq C\left[E\left|K\left(\frac{x - X_{i\Delta}}{h}\right)\right|^{2s}\right]^{1/2} \cdot \left[E\left|K\left(\frac{y - X_{(i+1)\Delta}}{h}\right)\right|^{2s}\right]^{1/2} \\
&\leq Ch \cdot \left[\frac{1}{h}E\left|K\left(\frac{x - X_{i\Delta}}{h}\right)\right|^{2s}\right]^{1/2} \cdot \left[\frac{1}{h}E\left|K\left(\frac{y - X_{(i+1)\Delta}}{h}\right)\right|^{2s}\right]^{1/2} \\
&\leq Ch.
\end{aligned}$$

By Markov inequality and Lemma 4.2, yields for $\varepsilon > 0$, $s > 2$, and $\delta > 0$,

$$\begin{aligned}
P(|\tilde{p}(x,y) - E\tilde{p}(x,y)| > \varepsilon) &\leq \frac{1}{(nh^2\varepsilon)^s}E\left|\sum_{i=1}^{n} G_{1,i}\right|^s \\
&\leq \frac{Cn^\delta}{(nh^2)^s}\left[\sum_{i=1}^{n} E\left|G_{1,i}\right|^s + \left(\sum_{i=1}^{n} E\left|G_{1,i}\right|^2\right)^{\frac{s}{2}}\right] \\
&\leq \frac{Cn^\delta}{(nh^2)^s}\left(nh + (nh)^{\frac{s}{2}}\right)
\end{aligned}$$

$$
\begin{aligned}
&\le Cn^{\delta}\frac{(nh)^{\frac{s}{2}}}{(nh^2)^s}\\
&= Cn^{\delta}\left(\frac{1}{nh^3}\right)^{\frac{s}{2}} = Cn^{\delta-\tau s/2} < \infty.
\end{aligned}
$$

Taking $s > \max(2, 2(1+\delta)/\tau)$ yields

$$
\sum_{i=1}^{n} P\left(|\tilde{p}(x,y) - E\tilde{p}(x,y)| > \varepsilon\right) < \infty.
$$

It is therefore known by the Borel-Cantelli lemma, $\tilde{p}(x,y) - E\tilde{p}(x,y) = o_{a.s.}(1)$. Following prove Eq (4.3).

According to Taylor's expansion and Assumption 6, as $h \to 0$,

$$
\begin{aligned}
E\tilde{p}(x,y) &= E\left[\frac{1}{nh^2}\sum_{i=1}^{n} K(\frac{x-X_{i\Delta}}{h})K(\frac{y-X_{(i+1)\Delta}}{h})\right]\\
&= \frac{1}{h^2}E\left[K(\frac{x-X_{i\Delta}}{h})K(\frac{y-X_{(i+1)\Delta}}{h})\right]\\
&= \frac{1}{h^2}\iint K(\frac{x-u}{h})K(\frac{y-v}{h})p(u,v)dudv\\
&= \frac{1}{h}\int K(\frac{y-v}{h})dv\cdot\frac{1}{h}\int K(\frac{x-u}{h})p(u,v)du\\
&= \frac{1}{h}\int K(\frac{y-v}{h})[p(x,y) + p'_y(x,y)(v-y)\\
&+ \frac{p_{yy}''(x,y)}{2}(v-y)^2 + O(h^2)]dv\\
&= p(x,y)\int K(z)dz + hp'_y(x,y)\int zK(z)dz\\
&+ \frac{h^2 p_{yy}''(x,y)}{2}\int z^2K(z)dz + O(h^2)\\
&\to p(x,y),
\end{aligned}
$$

where the penultimate equation is because of

$$
\begin{aligned}
\frac{1}{h}\int K(\frac{x-u}{h})p(u,v)du &= \frac{1}{h}\int K(\frac{x-u}{h})[p(x,y) + p'_x(x,y)(u-x)\\
&+ \frac{p_{xx}''(x,y)}{2}(u-x)^2 + p'_y(x,y)(v-y)\\
&+ \frac{p_{yy}''(x,y)}{2}(v-y)^2 + p_{xy}''(x,y)(u-x)(v-y)]du\\
&= (p(x,y) + p_y'(x,y)(v-y) + \frac{p_{yy}''(x,y)}{2}(v-y)^2)\int K(t)dt\\
&+ h\left(p'_x(x,y) + (v-y)p_{xy}''(x,y)\right)\int tK(t)dt
\end{aligned}
$$

$$+\frac{h^2 p_{xx}''(x,y)}{2}\int t^2K(t)dt$$
$$= p\,(x,y) + p_y'(x,y)(v-y) + \frac{p_{yy}''(x,y)}{2}(v-y)^2 + O(h^2).$$

The proof of Theorem 3.1. Notes

$$\hat{p}(y|x) - \tilde{p}(y|x) = \frac{\hat{p}(x,y)}{\hat{p}(x)} - \frac{\tilde{p}(x,y)}{\tilde{p}(x)}.$$

From Eq (4.1), Lemmas 4.4 and 4.5, we have $\tilde{p}(x) - p(x) = o_{a.s.}(1)$, $\tilde{p}(y|x) - p(y|x) = o_{a.s.}(1)$ and $\hat{p}(x) - p(x) = o_{a.s.}(1)$, respectively, so we only need to prove $\hat{p}(x,y) - \tilde{p}(x,y) = o_{a.s.}(1)$.

$$\begin{aligned}\hat{p}(x,y) - \tilde{p}(x,y) &= \frac{1}{nh^2}\sum_{i=1}^{n} K(\frac{x-\tilde{X}_{i\Delta}}{h})K(\frac{y-\tilde{X}_{(i+1)\Delta}}{h})\\ &- \frac{1}{nh^2}\sum_{i=1}^{n} K(\frac{x-X_{i\Delta}}{h})K(\frac{y-\tilde{X}_{(i+1)\Delta}}{h})\\ &+ \frac{1}{nh^2}\sum_{i=1}^{n} K(\frac{x-X_{i\Delta}}{h})K(\frac{y-\tilde{X}_{(i+1)\Delta}}{h})\\ &- \frac{1}{nh^2}\sum_{i=1}^{n} K(\frac{x-X_{i\Delta}}{h})K(\frac{y-X_{(i+1)\Delta}}{h})\\ &= A(x,y) + B(x,y).\end{aligned}$$

Denote

$$G_{2,i} = \left[K(\frac{x-\tilde{X}_{i\Delta}}{h}) - K(\frac{x-X_{i\Delta}}{h})\right]K(\frac{y-\tilde{X}_{(i+1)\Delta}}{h}),$$
$$G_{3,i} = K(\frac{x-X_{i\Delta}}{h})\left[K(\frac{y-\tilde{X}_{(i+1)\Delta}}{h}) - K(\frac{y-X_{(i+1)\Delta}}{h})\right],$$
$$\tilde{G}_{2,i} = G_{2,i} - EG_{2,i}, \tilde{G}_{3,i} = G_{3,i} - EG_{3,i}.$$

Then $A(x,y) = \frac{1}{nh^2}\sum_{i=1}^{n}\tilde{G}_{2,i} + \frac{1}{nh^2}\sum_{i=1}^{n}EG_{2,i}$, $B(x,y) = \frac{1}{nh^2}\sum_{i=1}^{n}\tilde{G}_{3,i} + \frac{1}{nh^2}\sum_{i=1}^{n}EG_{3,i}$. Taylor's expansion,

$$K\left(\frac{x-\tilde{X}_{i\Delta}}{h}\right) = K\left(\frac{x-X_{i\Delta}}{h}\right) + K'(\xi_{n,i})\frac{\tilde{X}_{i\Delta}-X_{i\Delta}}{h},$$

where $\xi_{n,i} = \theta\left(\frac{x-X_{i\Delta}}{h}\right) + (1-\theta)\left(\frac{x-\tilde{X}_{i\Delta}}{h}\right)$, $0 \le \theta \le 1$. Hence, by Assumption 6, Lemma 4.1, and Remark 4.1, there is

$$\frac{1}{h}E\left[K(\frac{x-X_{i\Delta}}{h})\right] = \frac{1}{h}\int K(\frac{u}{h})p(x-u)du \to p(x) < \infty,$$

$$\frac{1}{nh^2}\sum_{i=1}^{n}EG_{2,i} = \frac{1}{h^2}E\left[\left|K'(\xi_{n,i})(X_{i\Delta}-\tilde{X}_{i\Delta})\frac{1}{h}K(\frac{y-\tilde{X}_{(i+1)\Delta}}{h})\right|\right]$$

$$
\begin{aligned}
&\le h^{-3}(\Delta\log(1/\Delta))^{1/2}\cdot\left[E\left|K'(\xi_{n,i})\right|^{2}\right]^{1/2}\cdot\left[E\left|K(\frac{y-\tilde{X}_{(i+1)\Delta}}{h})\right|^{2}\right]^{1/2}\\
&= O\left(h^{-2}(\Delta\log(1/\Delta))^{1/2}\right)\to 0.
\end{aligned}
$$

For $\forall s\ge 2$, from Cr inequality, *Hölder* inequality, and Remark 4.1,

$$
\begin{aligned}
E\left|\tilde{G}_{2,i}\right|^{s} &\le 2^{s}E\left[\left|\frac{1}{h}K'(\xi_{n,i})(X_{i\Delta}-\tilde{X}_{i\Delta})K(\frac{y-\tilde{X}_{(i+1)\Delta}}{h})\right|^{s}\right]\\
&\le Ch^{-s}(\Delta\log(1/\Delta))^{s/2}\cdot\left[E\left|K'(\xi_{n,i})\right|^{2s}\right]^{1/2}\cdot\left[E\left|K(\frac{y-\tilde{X}_{(i+1)\Delta}}{h})\right|^{2s}\right]^{1/2}\\
&\le Ch^{-s+1}(\Delta\log(1/\Delta))^{s/2}.
\end{aligned}
$$

Similarly, suppose $\varepsilon>0$, $s>2$, and $\delta>0$, by Markov inequality and Lemma 4.2,

$$
\begin{aligned}
P\left(\left|\frac{1}{nh_n^2}\sum_{i=1}^{n}\tilde{G}_{2,i}\right|>\varepsilon\right) &\le \frac{1}{(nh^2\varepsilon)^{s}}E\left|\sum_{i=1}^{n}\tilde{G}_{2,i}\right|^{s}\\
&\le \frac{Cn^{\delta}}{(nh^2)^{s}}\left[nh^{-s+1}(\Delta\log(1/\Delta))^{\frac{s}{2}}+\left(nh^{-1}\Delta_n\log(1/\Delta)\right)^{\frac{s}{2}}\right]\\
&\le \frac{Cn^{\delta}}{(nh^2)^{s}}\left(nh^{-1}\Delta\log(1/\Delta)\right)^{\frac{s}{2}}\\
&\le Cn^{\delta}\left(\frac{\Delta\log(1/\Delta)}{nh^5}\right)^{\frac{s}{2}}\\
&\le Cn^{\delta-\tau s/2}.
\end{aligned}
$$

One of the last inequalities is given by

$$
\frac{\Delta\log(1/\Delta)}{nh^5}\Bigg/\frac{1}{nh^3}=\frac{\Delta\log(1/\Delta)}{h^2}\to 0.
$$

Taking $s>\max(2,2(1+\delta)/\tau)$ yields

$$
\sum_{i=1}^{n}P\left(\left|\frac{1}{nh^2}\sum_{i=1}^{n}\tilde{G}_{2,i}\right|>\varepsilon\right)<\infty.
$$

Based on the above results, from the Borel-Cantelli lemma, $A(x,y)=o_{a.s.}(1)$. In the same way, $B(x,y)=o_{a.s.}(1)$.

5. Simulation results

As pointed out in [4], the forms for the drift and diffusion coefficients in the second-order diffusion model can be found as $\mu(x)=\beta(\tau-x)$ and $\sigma(x)=\sqrt{\alpha^2+\lambda(x-\mu)^2}$, which implies that the specific model can be obtained by selecting the appropriate parameters. In this section, by choosing $\beta = 10$, $\tau = 0$, $\alpha^2 = 0.1$, $\lambda = 1$, and $\mu = 0.05$, we specify the following model:

$$\begin{cases} dY_t = X_t dt, \\ dX_t = -10X_t dt + \sqrt{0.1 + (X_t - 0.05)^2} dB_t, \end{cases}$$

where X is the ergodic process defined by the stochastic differential equation

$$dX_t = -10X_t dt + \sqrt{0.1 + (X_t - 0.05)^2} dB_t.$$

As computers are unable to simulate continuous orbits during the generation of sample orbital data, an Euler discretization scheme will be employed to discretize the sampling for simulating continuous processes. Namely, we shall take $t \in [0, T] = [0, 10]$, denote the observation time interval $\Delta = T/n$, where n is the sample size, and apply the discrete model

$$X_{t+1} = X_t - 10X_t\Delta + \sqrt{0.1 + (X_t - 0.05)^2} \left(\sqrt{\Delta} N(0, 1) \right)$$

to generate the sample paths for X, where the standard normal distribution $N(0, 1)$ is the random numbers produced by simulation. Recall that, the integral process $Y_t = Y_0 + \int_0^t X_u du$, and discretizing it, we get $Y_t - Y_{t-1} = \Delta X_t$. Figure 1 shows the simulated sample paths of X_t and Y_t with $n = 1000$ and $T = 10$. Obviously, the process X is consistent with the assumption of its stationarity, and the process Y conforms to its non-stationary assumption.

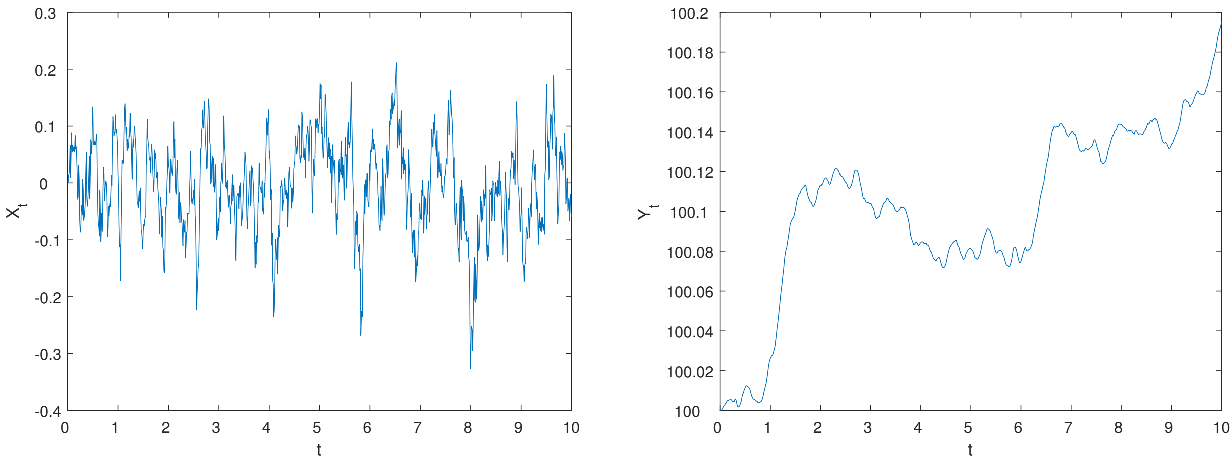

Figure 1. Simulated sample paths of processes X and Y.

In what follows, we shall evaluate the validity of the kernel estimator $\hat{p}(y|x)$. Setting a fixed observation time $T = 10$, selecting the Gaussian kernel $K(u) = \exp(-u^2/2)/\sqrt{2\pi}$, and taking the approximately optimal bandwidth $h = 1.06Sn^{-1/6}$ in (3.3), where S is the standard deviation of the sample. Figure 2 shows the comparison of the simulated curves for the exact and estimated values at different sampling intervals for $x = -0.1$ and $x = 0$, where the exact solutions are obtained by setting $\Delta = 0.0008$. Evidently, when the sample size n is larger, that is, the sampling interval Δ is smaller, the estimated simulation curve is closer to the exact value curve, which indicates that the kernel estimator of the transition density of the model (1.1) is consistent.

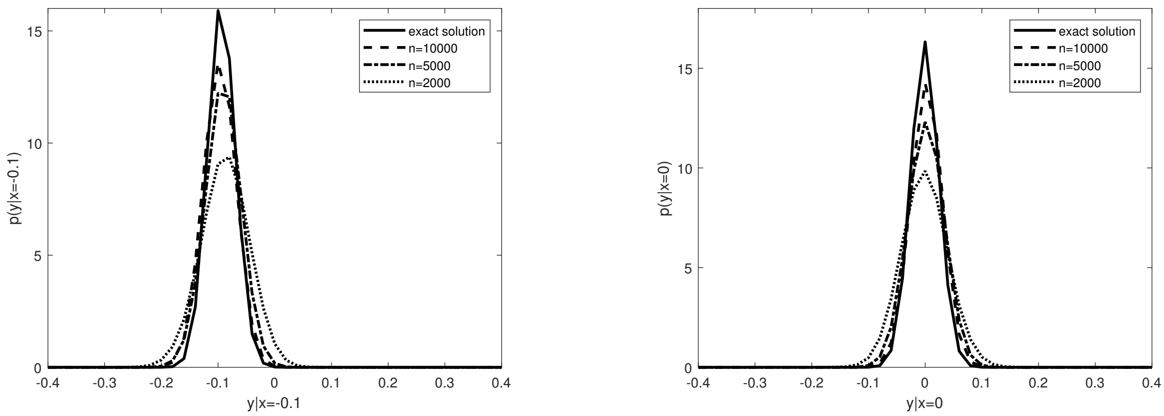

Figure 2. Kernel estimation simulation curves and exact value's curve.

6. Conclusions

Under relatively mild conditions, we use the moment inequalities of ρ-mixing sequences to prove that the nonparametric kernel estimator of the transition density of the second-order diffusion process satisfies strong consistency, thereby further improving upon the discussion presented in [23] regarding the estimator's weak convergence. The good theoretical property provides some guarantees for its application. In addition, we also found that the optimal bandwidth h^* of the transition density kernel estimator is of order $n^{-1/6}$ and the convergence rate is of order $n^{-2/3}$. Of course, h^* is not a practical bandwidth selection rule, because it contains unknown functions, but the rule may provide a guide for use in transition density estimation. It is also possible to obtain consistent estimators by using other nonparametric estimation techniques, including the local linear method and the re-weighted method. We leave it as a prospect for future research.

Author contributions

Yue Li: Conceptualization, Methodology, Writing-original draft, Writing-review & editing; Yunyan Wang: Conceptualization, Methodology, Supervision, Writing review & editing. All authors have read and approved the final version of the manuscript for publication.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors would like to thank the referees for their valuable suggestions, which greatly improved the structure and the presentation of the paper.

Additionally, this work was supported by the National Nature Science Foundation of China under Grant numbers 11461032 and 11401267, and Jiangxi University of Science and Technology under the Program of Qingjiang Excellent Young Talents.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. F. M. Bandi, P. C. B. Phillips, Fully nonparametric estimation of scalar diffusion models, *Econometrica*, **71** (2003), 241–283. https://doi.org/10.1111/1468-0262.00395
2. J. Fan, C. Zhang, A reexamination of diffusion estimators with applications to financial model validation, *J. Am. Stat. Assoc.*, **98** (2003), 118–134. https://doi.org/10.1198/016214503388619157
3. F. Comte, V. G. Catalot, Y. Rozenholc, Penalized nonparametric mean square estimation of the coefficients of diffusion processes, *Bernoulli*, **13** (2007), 514–543. https://doi.org/10.3150/07-BEJ5173
4. J. Nicolau, Nonparametric estimation of second-order stochastic differential equations, *Economet. Theor.*, **23** (2007), 880–898. https://doi.org/10.1017/S0266466607070375
5. A. Gloter, Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient, *ESAIM-Probab. Stat.*, **4** (2000), 205–227. https://doi.org/10.1051/ps:2000105
6. A. Gloter, Parameter estimation for a discretely observed integrated diffusion process, *Scand. J. Stat.*, **33** (2006), 83–104. https://doi.org/10.1111/j.1467-9469.2006.00465.x
7. J. Nicolau, Modeling financial time series through second-order stochastic differential equations, *Stat. Probabil. Lett.*, **78** (2008), 2700–2704. https://doi.org/10.1016/j.spl.2008.03.024
8. A. Carriero, T. E. Clark, M. Marcellino, Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors, *J. Econometrics*, **212** (2019), 137–154. https://doi.org/10.1016/j.jeconom.2019.04.024
9. X. J. He, W. Chen, A closed-form pricing formula for European options under a new stochastic volatility model with a stochastic long-term mean, *Math. Financ. Econ.*, **15** (2021), 381–396. https://doi.org/10.1007/s11579-020-00281-y
10. L. Rogers, W. David, *Diffusions, Markov processes and martingales: Volume 2, Itô calculus*, Cambridge University Press, **2** (2020). https://doi.org/10.1112/blms/21.1.106
11. P. D. Ditlevsen, S. Ditlevsen, K. K. Andersen, The fast climate fluctuations during the stadial and interstadial climate states, *Ann. Glaciol.*, **35** (2002), 457–462. https://doi.org/10.3189/172756402781816870

12. H. C. Wang, Z. Y. Lin, Local linear estimation of second-order diffusion models, *Commun. Stat.-Theor. M.*, **40** (2011), 394–407. https://doi.org/10.1080/03610920903391345
13. Y. Y. Wang, L. X. Zhang, M. T. Tang, Re-weighted functional estimation of second-order diffusion processes, *Metrika*, **75** (2012), 1129–1151. https://doi.org/10.1007/s00184-011-0372-6
14. M. T. Tang, Y. Y. Wang, Q. Q. Zhan, Non parametric bias reduction of diffusion coefficient in integrated diffusion processes, *Commun. Stat.-Theor. M.*, **51** (2022), 6435–6446. https://doi.org/10.1080/03610926.2020.1861298
15. S. C. Yang, S. Zhang, G. D. Xing, X. Yang, Strong consistency of nonparametric kernel estimators for integrated diffusion process, *Commun. Stat.-Theor. M.*, **53** (2024), 2792–2815. https://doi.org/10.1080/03610926.2022.2148540
16. L. C. Zhao, Z. J. Liu, Strong consistency of the kernel estimators of conditional density function, *Acta Math.*, **1** (1985), 314–318. https://doi.org/10.1007/BF02564838
17. L. G. Xue, M. Z. Fan, Strong convergence rates of the double kernel estimates of conditional density (in Chinese), *Chinese J. Eng. Math.*, **1** (2001), 89–93.
18. S. Khardani, S. Semmar, Nonparametric conditional density estimation for censored data based on a recursive kernel, *Electron. J. Stat.*, **8** (2015), 2541–2556. https://doi.org/10.1214/14-EJS960
19. A. Benkhaled, F. Madani, S. Khardani, Strong consistency of local linear estimation of a conditional density function under random censorship, *Arab. J. Math*, **9** (2020), 513–529. https://doi.org/10.1007/s40065-020-00282-1
20. N. Kadiri, M. Meghnafi, A. Rabhi, Strong uniform consistency rates of conditional density estimation in the single functional index model for functional data under random censorship, *REVSTAT-Stat. J.*, **20** (2022), 221–249. https://doi.org/10.57805/revstat.v20i2.374
21. H. Wang, W. X. Chen, B. J. Li, Large sample properties of maximum likelihood estimator using moving extremes ranked set sampling, *J. Korean. Stat. Soc.*, 2024. https://doi.org/10.1007/s42952-023-00251-2
22. B. Sayedana, M. Afshari, P. E. Caines, A. Mahajan, Strong consistency and rate of convergence of switched least squares system identification for autonomous markov jump linear systems, *IEEE. T. Automat. Contr.*, 2024. https://doi.org/10.1109/TAC.2024.3351806
23. Y. Li, Y. Y. Wang, M. T. Tang, Non parametric estimation of transition density for second-order diffusion processes, *Commun. Stat.-Theor. M.*, 2023, 1–13. https://doi.org/10.1080/03610926.2023.2234521
24. I. Karatzas, E. Shreve, *Brownian motion and stochastic calculus*, Springer, 2014. https://doi.org/10.1007/978-3-319-57511-7_2
25. B. W. Silverman, *Density estimation for statistics and data analysis*, London: Chapman and Hall, 1986. https://doi.org/10.2307/2348849
26. R. J. Hyndman, D. M. Bashtannyk, G. K. Grunwald, Estimating and visualizing conditional densities, *J. Comput. Graph. Stat.*, **5** (1996), 315–336. https://doi.org/10.1080/10618600.1996.10474715

27. D. W. Scott, Multivariate density estimation and visualization, *Handbook Comput. Stat.*, 2012, 549–569. https://doi.org/10.1007/978-3-642-21551-3_19

28. E. Parzen, On the estimation of a probability density function and mode, *Ann. Math. Stat.*, **33** (1962), 1065–1076. https://doi.org/10.2307/2237880

29. S. C. Yang, Moment inequality for mixing sequences and nonparametric estimation, *Acta Math. Sin.*, **40** (1997), 271–279. https://doi.org/10.12386/A1997sxxb0034