Statistical Inference for Stochastic Processes **6:** 89–107, 2003.

Asymptotically Efficient Estimation of the Derivative of the Invariant Density

ARNAK S. DALALYAN and YURY A. KUTOYANTS
Laboratoire de Statistique et Processus, Université du Maine, Avenue Messiaen, 72085 Le Mans, Cedex 9, France. e-mails: arnak.dalalyan.etu@univ-lemans.fr, kutoyants@univ-lemans.fr

(Received in final form 17 April 2001; Accepted 30 May 2002)

Abstract. The problem of estimation of the derivative of the invariant density is considered for a one-dimensional ergodic diffusion process. The lower minimax bound on the L^2-type risk of all estimators is proposed and an asymptotically efficient (up to the constant) in the sense of this bound kernel-type estimator is constructed.

AMS 1999 Subject Classification: 62M05, 62G07, 62G20.

Key words: ergodic diffusion, invariant density, derivative estimation, asymptotical efficiency, Pinsker's constant.

1. Introduction

In this work we consider a diffusion process X given by the stochastic differential equation

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

where W_t denotes a standard Wiener process defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and the initial value X_0 is a random variable independent of the Wiener process. We suppose that the following condition is fulfiled:

$$G(S) = \int_{-\infty}^{\infty} \exp\left\{2\int_0^x S(v)\,\mathrm{d}v\right\}\mathrm{d}x < \infty. \tag{2}$$

By this condition the process (1) is recurrent positive, that is, there exists an invariant probability distribution with the density function

$$f_S(y) = \frac{1}{G(S)}\exp\left\{2\int_0^y S(v)\,\mathrm{d}v\right\}$$

and the process $\{X_t,\ t \geqslant 0\}$ has ergodic properties [3, 5].

The trend coefficient $S(\cdot)$ is supposed to be unknown, so the invariant density $f_S(\cdot)$ is unknown function, too. The problem of estimation of this density by observations $X^T = \{X_t, 0 \leqslant t \leqslant T\}$ as $T \to \infty$ has been considered by many authors (see [1, 2, 9, 10, 12] and references therein). Particularly, in [9] a lower bound on the minimax risk of all estimators is proposed and several estimators (local-time, kernel-type, unbiased) attaining this bound in the limit are constructed. These estimators are called asymptotically efficient when the time of observation tends to infinity. It is also shown that they are $\sqrt{T}$ consistent and asymptotically normal.

There are several open problems closely related with the results described above. First, to find an estimator of the density which is asymptotically efficient in the second order, say, as it is done in distribution function estimation problem in i.i.d. case by Golubev and Levit [8]. Second, to construct an asymptotically efficient estimator of the trend coefficient $S(\cdot)$. Third, to find an asymptotically efficient estimator of the derivative of the invariant density. One of the possible approaches in these problems is the so-called minimax approach that we develop in this paper.

Remark that there exists a close relation between the problems cited above. In fact, the trend coefficient can be written as

$$S(x) = \frac{f_S'(x)}{2f_S(x)}.$$

It can be easily shown that this function can be estimated with a rate depending on the smoothness of $S(\cdot)$. Say, if the function $S(\cdot)$ is k-times continuously differentiable, then the best possible rate has to be $T^{k/(2k+1)}$. Therefore, if one considers the estimator of $S(x)$ constructed via some efficient estimators of $f_S(x)$ and $f_S'(x)$, then, roughly speaking, the error of estimation of the density $f_S(x)$ is asymptotically smaller than the error of $f_S'(x)$ estimation. Thus, the main problem is to construct an asymptotically efficient estimator of the derivative $f_S'(x)$.

The present work is devoted to the problem of this derivative $f_S'(\cdot)$ estimation. A similar problem was discussed by Lucas [11] in the case where X^T is stationary and continuous but not necessarily a Markov process. The author obtained some conditions under which a $\sqrt{T}$-convergent kernel-type estimator exists, but pointed out that the class of processes satisfying these conditions is quite narrow.

We present below two results concerning the same problem. The first one is the lower minimax bound on the L_2-type risk of all estimators (Section 2). The second provides a kernel-type estimator, which is asymptotically efficient in the sense of this bound (Section 3). We finish by concluding remarks where we propose some possible generalizations of these results (Section 4).

Note that the statement of the problem and the results obtained in this paper lie in the framework of *Pinsker's bound approach* [14] (see [13] for the discussion and references) and the schemes of the proofs follow the work by Golubev and Levit [8] (see as well the paper by Schipper [15]). The main difficulty in our case, which does not exist in the case of density estimation by i.i.d. data, is the following. When we take sup of the risk on the class of functions $S(\cdot)$ satisfying condition (2), we touch the case $G(S) = \infty$, which corresponds to the null recurrent process, and in this case we have no more the law of large numbers. Therefore to prove the asymptotic efficiency of the estimators we need to control the divergence of this integral. It is done using three conditions (3)–(5), given below.

To formulate the main results we need some notation. Let $\mathcal{C}_T$ be the space (equipped with the topology of the uniform convergence) of all continuous functions defined on $[0, T]$ and let $\mathcal{B}_T$ be the σ-algebra of its Borelian subsets. The process $X^T = \{X_t,\ 0 \leqslant t \leqslant T\}$ induces a probability measure on the space $(\mathcal{C}_T, \mathcal{B}_T)$. This measure and the mathematical expectation with respect to this measure will be denoted by $\mathbf{P}_S$ and $\mathbf{E}_S$, respectively. In the cases where the function $S(\cdot)$ depends on a parameter θ, we will write $\mathbf{P}_\theta$ and $\mathbf{E}_\theta$ instead of $\mathbf{P}_{S_\theta}$ and $\mathbf{E}_{S_\theta}$.

The set of all trend coefficients $S(\cdot)$, which provide the (weak) existence, uniqueness and ergodicity of X_t (see [3]) will be denoted by Σ_0. We suppose that the initial value X_0 has the density $f_S(\cdot)$. Under this condition $\{X_t,\ t \geqslant 0\}$ is a stationary process.

We consider the nonparametric problem of $f'_S(\cdot)$ derivative function estimation from continuous path observations and give the exact asymptotics of the quadratic minimax risk. That is, we find a constant $C(\Sigma)$ such that

$$\lim_{T\to\infty} \inf_{\bar{\vartheta}_T} \sup_{S\in\Sigma} T^{2k/(2k+1)} \mathbf{E}_S \int_{\mathbb{R}} \left(\bar{\vartheta}_T(x) - f'_S(x)\right)^2 \mathrm{d}x = C(\Sigma),$$

where the infimum is taken over the class of all possible estimators.

To define the set Σ we need some additional notation. Let χ_A denote the indicator of an event $A \subset \Omega$ and $F_S(\cdot)$ be the probability distribution of the random variable ξ with density function $f_S(\cdot)$, that is,

$$F_S(x) = \mathbf{P}(\xi \leqslant x).$$

Throughout this paper the constants (do not depending on T and S) will be denoted by C.

The parameter space Σ is defined in the following way. Let us fix a real number $R > 0$ and an integer $k \geqslant 2$ and define the ellipsoid

$$\Sigma(k, R) = \left\{ S(\cdot) \in \Sigma_0 \,\middle|\, \int_{\mathbb{R}} \left[f_S^{(k+1)}(x) \right]^2 \mathrm{d}x \leqslant R \right\}.$$

We can define now the infinite-dimensional parameter space Σ as the set of all $S \in \Sigma(k, R)$, such that for a constant D (does not depending on S) the following conditions are fulfiled:

1. There exists a positive constant B_0 (does not depending on S) such that

$$\sup_{B>B_0} B^2 \int_{|x|>B} \left[f'_S(x)\right]^2 \mathrm{d}x < D, \tag{3}$$

$$\sup_{B>B_0} B^{-2} \int_{|x|<B} \left[f'_S(x)\right]^2 \mathbf{E}_S \left[\int_0^\xi \frac{\chi_{\{y>x\}} - F_S(y)}{f_S(y)} \mathrm{d}y \right]^2 \mathrm{d}x < D. \tag{4}$$

2. The following estimate holds:

$$\int_{\mathbb{R}} \left[f_S'(x)\right]^2 \mathbf{E}_S\left[\frac{\chi_{\{\xi>x\}} - F_S(\xi)}{f_S(\xi)}\right]^2 \,\mathrm{d}x < D. \tag{5}$$

Of course, the set Σ depends on the constants k, R, D and B_0. In the sequel this set will be denoted by Σ_D.

2. Lower Bound

In this section we will establish a lower bound on the minimax risk

$$\inf_{\bar\vartheta_T} \sup_{S\in\Sigma} R_T(\bar\vartheta_T, f_S') = \inf_{\bar\vartheta_T} \sup_{S\in\Sigma} \mathbf{E}_S \int_{\mathbb{R}} \left(\bar\vartheta_T(x) - f_S'(x)\right)^2 \,\mathrm{d}x, \tag{6}$$

where the inf is taken over all estimators of the derivative.

THEOREM 1. *The following inequality is true*

$$\lim_{D\to\infty} \liminf_{T\to\infty} \inf_{\bar\vartheta_T} \sup_{S\in\Sigma_D} T^{2k/(2k+1)} R_T(\bar\vartheta_T, f_S') \geqslant P(k, R), \tag{7}$$

where the Pinsker constant $P(k, R)$ is defined by

$$P(k, R) = (2k+1)\left(\frac{\pi(k+1)(2k+1)}{4k}\right)^{-2k/(2k+1)} R^{1/(2k+1)}. \tag{8}$$

Proof. To prove this inequality we minorate the minimax risk (6) by the minimax risk over a properly chosen parametric family. The dimension of this family depends on T and tends to infinity when $T \to \infty$.

Thus, let us introduce the following parameterizations

$$S_\theta(x) = \sum_{|i|\leqslant L} \theta_i e_i(x) U(A - |x|), \quad x \in [-A, A], \tag{9}$$

where $A = \ln(1+T)$ and $U(x)$ is $k+1$ times differentiable increasing function vanishing for $x \leqslant 0$ and equal to 1 for $x \geqslant 1$. The functions $\{e_i\}_{i\in\mathbb{Z}}$ are the elements of the trigonometric basis of $L^2[-A, A]$, that is,

$$e_i(x) = \frac{1}{\sqrt{A}} \begin{cases} \sin\left(\dfrac{\pi i x}{A}\right) & \text{if } i > 0, \\ \dfrac{1}{\sqrt{2}} & \text{if } i = 0, \\ \cos\left(\dfrac{\pi i x}{A}\right) & \text{if } i < 0. \end{cases}$$

The positive number $L = L_T$ will be chosen later.

In the outside of the interval $[-A, A]$ we put

$$S_\theta(x) = S_0(x) = -\operatorname{sgn}(x)(k+2)\left(|x| - A\right)^{k+1}.$$

It is evident that the function $S_\theta(\cdot)$ defined in this way is k times differentiable and the kth derivative is continuous. Consequently, the density $f_\theta(\cdot)$ is $k+1$ times continuously differentiable.

Since for any fixed θ, the function $S_\theta(\cdot)$ is locally Lipshitz and

$$\sup_{x\in\mathbb{R}} \frac{xS_\theta(x)+1}{1+x^2} < \infty,$$

there exists a unique solution of (1) (see [3]) corresponding to the case $S = S_\theta$.

In the sequel we consider the parametric space Γ_T of all sequences $\{\theta_i\}_{|i|\leqslant L}$ such that $|\theta_i| \leqslant G\sqrt{\sigma_i}$ for all $i \in \mathbb{Z}$ such that $|i| \leqslant L$. Here

$$\sigma_i = \sigma_{i,T} = \frac{2A}{T}\left(\left|\frac{\alpha}{i}\right|^k - 1\right)_+, \quad i \neq 0 \tag{10}$$

with

$$\alpha = \alpha_T = A\left(\frac{TR(k+1)(2k+1)}{4k\pi^{2k}}\right)^{1/(2k+1)} \tag{11}$$

and $\sigma_0 = 0$. The integer L is chosen to be equal $[\alpha]$.

LEMMA 1. *For any $m = 0, 1, \ldots, k$, we have the following estimates*

$$\sup_{\theta\in\Gamma_T} \sup_{x\in[-A,A]} \left| \left(\int_0^x S_\theta(v)\,\mathrm{d}v\right)^{(m)} \right| \leqslant CA^{k+2}T^{-1/(4k+2)}. \tag{12}$$

Proof. For $m = 0$, using the inequality $|\theta_i| \leqslant G\sqrt{\sigma_i}$ and the definitions (10) and (11), we obtain

$$\begin{aligned}\left|\int_0^x S_\theta(v)\,\mathrm{d}v\right| &\leqslant A \sup_{|x|\leqslant A} \left|S_\theta(x)\right| \leqslant A \sup_{|x|\leqslant A} \sum_{i\neq 0} \left|\theta_i e_i(x)\right| \leqslant \sqrt{A} \sum_{i\neq 0} |\theta_i| \\ &= \frac{4GA}{\sqrt{T}} \sum_{i=1}^{L} \left(\left|\frac{\alpha}{i}\right|^k - 1\right)_+^{1/2} \leqslant \frac{4GA\alpha^{(k/2)+1}}{\sqrt{T}} \leqslant \frac{CA^{k+2}}{T^{1/(4k+2)}}\,.\end{aligned}$$

Remark that this chain of inequalities proves the bound (12) for $m = 1$ as well. Suppose that this bound holds for $1, 2, \ldots, m-1$ and prove it for m. Since all the derivatives of the function $U(\cdot)$ up to the order m are bounded, we have

$$\begin{aligned}\left|S_\theta^{(m)}(x)\right| &\leqslant \frac{1}{\sqrt{A}} \sum_{i\neq 0} |\theta_i| \left(\frac{\pi \mathrm{i}}{A}\right)^m + \frac{CA^{k+2}}{T^{1/(4k+2)}} \\ &\leqslant \frac{C}{A^m} \sum_{i\neq 0} i^m \sqrt{\sigma_i} + \frac{CA^{k+2}}{T^{1/(4k+2)}} \\ &\leqslant \frac{C}{A^{m-1}\sqrt{T}} \sum_{i=1}^{\alpha} i^{k-1} \left(\frac{\alpha}{i}\right)^{k/2} + \frac{CA^{k+2}}{T^{1/(4k+2)}} \\ &\leqslant \frac{C\alpha^k}{A^{m-1}\sqrt{T}} + \frac{CA^{k+2}}{T^{1/(4k+2)}} \leqslant \frac{CA^{k+2}}{T^{1/(4k+2)}}.\end{aligned}$$

□

COROLLARY 1. *The following relation holds*

$$f_\theta(x) = \frac{1+o_T(1)}{2A}\begin{cases}\exp\left\{-2(|x|-A)^{k+2}\right\} & if \quad |x|>A,\\ 1 & if \quad |x|\leqslant A.\end{cases}$$

The proof is easy and immediately follows from Lemma 1.

Let $\{\xi_i\}_{i\in\mathbb{Z}}$ be i.i.d. random variables with common probability density $p(x)$ such that

$$|\xi_i|<G,\qquad \mathbf{E}(\xi_i)=0,\qquad \mathbf{E}(\xi_i^2)=1,\qquad I=\int\frac{[p'(x)]^2}{p(x)}\,\mathrm{d}x=1+\varepsilon,$$

where $\varepsilon\to 0$ when $G\to\infty$.

We introduce a prior distribution Λ on Γ_T putting

$$\theta_i=\sqrt{\sigma_i(\varepsilon)}\,\xi_i$$

with

$$\sigma_i(\varepsilon)=\frac{2A(1+\varepsilon)}{T}\left(\left|\frac{\alpha(1-\varepsilon)}{i}\right|^k-1\right)_+$$

for each i different from 0. The coefficient θ_0 will be deterministic and equal to 0. The Fisher information I_i of this prior distribution with respect to θ_i (for $i\neq 0$) is then equal to $(1+\varepsilon)/\sigma_i(\varepsilon)$.

Since the minimax risk can be evaluated below by the Bayesian one with a correction term (which is shown to be small in order), we seek now a lower bound of the Bayesian risk corresponding to a prior distribution Λ, defined by

$$\mathcal{R}_T(\Lambda)=\inf_{\bar\vartheta_T}\mathbb{E}\int_{\mathbb{R}}\left(\bar\vartheta_T(x)-f_\theta'(x)\right)^2\,\mathrm{d}x,$$

where $\mathbb{E}$ is the mathematical expectation with respect to the probability measure $\mathbf{P}_\theta(\mathrm{d}X^T)\times\Lambda(\mathrm{d}\theta)$.

Let $\psi_{i,\theta}$ and $\psi_{i,T}$ be the Fourier coefficients on $[-A,A]$ of $f_\theta'(\cdot)$ and $\bar\vartheta_T(\cdot)$, respectively, that is,

$$\psi_{i,\theta}=\int_{-A}^{A}f_\theta'(x)e_i(x)\,\mathrm{d}x,\qquad \psi_{i,T}=\int_{-A}^{A}\bar\vartheta_T(x)e_i(x)\,\mathrm{d}x.$$

Since $\{e_i(\cdot)\}$ is an orthonormal sequence and the function $f_\theta'(\cdot)$ belongs to the Hilbert space generated by this sequence, using the Parseval's identity one has

$$\mathcal{R}_T(\Lambda)\geqslant\inf_{\bar\vartheta_T}\mathbb{E}\int_{-A}^{A}\left(\bar\vartheta_T(x)-f_\theta'(x)\right)^2\,\mathrm{d}x\geqslant\inf_{\bar\vartheta_T}\sum_{|i|<L}\mathbb{E}\left(\psi_{i,T}-\psi_{i,\theta}\right)^2.$$

By two-dimensional van Trees inequality (see [4])

$$\mathcal{R}_T(\Lambda) \geqslant \big(1+o_T(1)\big) \sum_{0<i\leqslant L} \frac{\big(\mathbb{E}\big[\partial\psi_{i,\theta}/\partial\theta_i + \partial\psi_{-i,\theta}/\partial\theta_{-i}\big]\big)^2}{T\mathbf{E}\big(I_i(\theta)+I_{-i}(\theta)\big)+I_i+I_{-i}}, \tag{13}$$

where $I_i(\theta)$ is defined by the formula

$$I_i(\theta) = \mathbf{E}_\theta\bigg[\frac{\partial S_\theta}{\partial\theta_i}(\xi)\bigg]^2.$$

The elementary calculations show that

$$\frac{\partial f_\theta'}{\partial\theta_i}(x) \;=\; 2f_\theta(x)\frac{\partial S_\theta}{\partial\theta_i}(x) + 2f_\theta'(x)\,\mathbf{E}_\theta\bigg[\int_\xi^x \frac{\partial S_\theta}{\partial\theta_i}(v)\,\mathrm{d}v\bigg].$$

Using Lemma 1, one can easily prove that

$$\sup_{|x|\leqslant A} f_\theta'(x)\,\mathbf{E}_\theta\bigg[\int_\xi^x \frac{\partial S_\theta}{\partial\theta_i}(v)\,\mathrm{d}v\bigg] = A^{-2}o_T(1), \tag{14}$$

which gives us

$$\frac{\partial f_\theta'}{\partial\theta_i}(x) = 2f_\theta(x)e_i(x)U(A-|x|) + A^{-2}o_T(1). \tag{15}$$

So the partial derivative of $\psi_{\theta,i}$ with respect to θ_i can be evaluated like follows

$$\begin{aligned}\frac{\partial\psi_{i,\theta}}{\partial\theta_i} &= \int_{-A}^{A} e_i(x)\,\frac{\partial f_\theta'}{\partial\theta_i}(x)\,\mathrm{d}x\\ &= 2\int_{-A}^{A} e_i^2(x)U(A-|x|)f_\theta(x)\,\mathrm{d}x + A^{-1}o_T(1)\\ &= 2\int_{-A}^{A} e_i^2(x)f_\theta(x)\,\mathrm{d}x + A^{-1}o_T(1).\end{aligned}$$

This equality and the elementary identity $e_i^2(x)+e_{-i}^2(x) = 1/A$ imply

$$\frac{\partial\psi_{i,\theta}}{\partial\theta_i} + \frac{\partial\psi_{-i,\theta}}{\partial\theta_{-i}} = \frac{2}{A}(1+o_T(1)). \tag{16}$$

It is not difficult to show that

$$I_i(\theta)+I_{-i}(\theta) = \int_{-A}^{A}\big(e_i^2(x)+e_{-i}^2(x)\big)f_\theta(x)\,\mathrm{d}x + \frac{O_T(1)}{A^2} = \frac{1+o_T(1)}{A}\,.$$

Now the inequality (13) can be rewritten as

$$\mathcal{R}_T(\Lambda) \geqslant 4A^{-1}\big(1+o_T(1)\big)\sum_{0<i\leqslant L}\frac{\sigma_i(\varepsilon)}{T\sigma_i(\varepsilon)+2A(1+\varepsilon)}\,. \tag{17}$$

Simple calculations show that the last sum is equal to

$$\frac{4k\alpha(1-\varepsilon)}{TA(k+1)}\,(1+o_T(1)).$$

Replacing α by expression (11) we get the inequality

$$\mathcal{R}_T(\Lambda)\geqslant T^{-2k/(2k+1)}P(k,R)(1+o_T(1))(1-\varepsilon)$$

with

$$P(k,R)=(2k+1)\left(\frac{\pi(k+1)(2k+1)}{4k}\right)^{-2k/(2k+1)}R^{1/(2k+1)}. \tag{18}$$

Remark 1. It is important to emphasize that the values (10) and (11) are the solutions of the maximization problem related with the functional

$$\Psi(y)=\sum_{i>0}\frac{y_i}{Ty_i+2A}$$

over the set

$$\mathcal{E}(k,R)=\left\{y=(y_i)_{i>0}\,\middle|\,2A^{-2}\sum_{i>0}y_i\left(\frac{\pi\mathrm{i}}{A}\right)^{2k}\leqslant R\right\}.$$

This choice is based on relation (17) and Lemma 2.

Now we prove some lemmas which will permit us to complete the proof of the lower bound.

LEMMA 2. *There exists a constant* $D>0$ *such that the functions* $\{S_\theta(\cdot)\}_{\theta\in\Gamma_T}$ *satisfy conditions* (3)–(5), *if* T *is sufficiently large.*

Proof. We show firstly that there exists a constant D such that

$$\sup_{\theta\in\Gamma_T}\int_{\mathbb{R}}\left[f_\theta'(u)\right]^2\mathbf{E}_\theta\left[\frac{\chi_{\{\xi>u\}}-F_\theta(\xi)}{f_\theta(\xi)}\right]^2\mathrm{d}u<D \tag{19}$$

for all $T>T_0$. Remark that for $z>y>A$ the following inequality is true

$$(z-A)^{k+2}-(y-A)^{k+2}\geqslant(z-y)(y-A)^{k+1},$$

it follows that

$$\frac{1-F_\theta(y)}{f_\theta(y)}\leqslant\int_y^\infty\exp\left\{-(z-y)(y-A)^{k+1}\right\}\mathrm{d}z=\frac{1}{(y-A)^{k+1}}\,.$$

This leads us to the inequality

$$\int_A^\infty f_\theta'^2(u)\mathbf{E}_\theta\left[\left(\frac{1-F_\theta(\xi)}{f_\theta(\xi)}\right)^2\chi_{\{\xi>u\}}\right]\mathrm{d}u\leqslant\int_A^\infty f_\theta'^2(u)\,\mathbf{E}_\theta\left[\frac{\chi_{\{\xi>u\}}}{(\xi-A)^{2k+2}}\right]\mathrm{d}u$$

$$\leqslant\int_A^\infty f_\theta'^2(u)\,(u-A)^{-2k}\,\mathrm{d}u<1 \tag{20}$$

for T large enough. In the same way, the inequality

$$(y-A)^{k+2} \leqslant (x-A)^{k+1}(y-A) \quad \text{for } A \leqslant y \leqslant x,$$

implies that

$$\begin{aligned} f_\theta(u)\int_A^u \frac{F_\theta^2(y)}{f_\theta(y)}\,\mathrm{d}y &\leqslant \int_A^u \frac{f_\theta(u)}{f_\theta(y)}\,\mathrm{d}y \leqslant \int_A^u \mathrm{e}^{(u-A)^{k+1}(y-u)}\,\mathrm{d}y \\ &\leqslant \frac{1}{(u-A)^{k+1}}\,. \end{aligned}$$

Hence

$$\begin{aligned} &\int_A^\infty f_\theta'^2(u)\mathbf{E}_\theta\left[\left(\frac{F_\theta(\xi)}{f_\theta(\xi)}\right)^2 \chi_{\{A<\xi<u\}}\right]\mathrm{d}u \leqslant C\int_A^\infty f_\theta'^2(u)\,|f_\theta'(u)|^{-1}\,\mathrm{d}u \\ &\quad = C\int_A^\infty |f_\theta'(u)|\,\mathrm{d}u = Cf_\theta(A) \;<\; 1 \end{aligned} \tag{21}$$

if T is large enough. Then, we have

$$\int_{-A-1}^{A} \frac{F_\theta^2(y)}{f_\theta(y)}\,\mathrm{d}y \leqslant \int_{-A-1}^{A} \frac{1}{f_\theta(y)}\,\mathrm{d}y = 4A^2\big(1+o_T(1)\big)$$

and

$$\int_A^\infty f_\theta'^2(u)\mathbf{E}_\theta\left[\left(\frac{F_\theta(\xi)}{f_\theta(\xi)}\right)^2 \chi_{\{-A-1<\xi<A\}}\right]\mathrm{d}u \leqslant 8A^2\int_A^\infty f_\theta'^2(u)\,\mathrm{d}u \;<\; C. \tag{22}$$

Proceeding like in the proof of (20), one can show that

$$\int_{-\infty}^{-A-1} \frac{F_\theta^2(y)}{f_\theta(y)}\,\mathrm{d}y \leqslant \sup_{y\leqslant -A-1} \frac{F_\theta^2(y)}{f_\theta^2(y)} \leqslant \frac{1}{(A+1-A)^{2k+2}} = 1.$$

Thus,

$$\int_A^\infty f_\theta'^2(u)\mathbf{E}_\theta\left[\left(\frac{F_\theta(\xi)}{f_\theta(\xi)}\right)^2 \chi_{\{\xi<-A\}}\right]\mathrm{d}u \leqslant 1 \tag{23}$$

for T large enough. Combining the inequalities (20)–(23), we obtain

$$\int_A^\infty f_\theta'^2(u)\mathbf{E}_\theta\left[\left(\frac{\chi_{\{\xi>u\}}-F_\theta(\xi)}{f_\theta(\xi)}\right)^2\right]\mathrm{d}u \leqslant C. \tag{24}$$

In the same way it can be shown that

$$\int_{-\infty}^{-A} f_\theta'^2(u)\mathbf{E}_\theta\left[\left(\frac{\chi_{\{\xi>u\}}-F_\theta(\xi)}{f_\theta(\xi)}\right)^2\right]\mathrm{d}u \leqslant C. \tag{25}$$

For $u \in [-A, A]$, using Lemma 1 we have

$$f'_\theta(u) = \frac{o_T(1)}{2A}$$

and

$$\begin{aligned}\mathbf{E}_\theta\left[\left(\frac{\chi_{\{\xi>u\}} - F_\theta(\xi)}{f_\theta(\xi)}\right)^2\right] \leqslant\ & \mathbf{E}_\theta\left[\left(\frac{1 - F_\theta(\xi)}{f_\theta(\xi)}\right)^2 \chi_{\{\xi>A+1\}}\right] + \\ & + \mathbf{E}_\theta\left[\left(\frac{1}{f_\theta(\xi)}\right)^2 \chi_{\{-A-1\leqslant\xi\leqslant A+1\}}\right] + \\ & + \mathbf{E}_\theta\left[\left(\frac{F_\theta(\xi)}{f_\theta(\xi)}\right)^2 \chi_{\{\xi<-A-1\}}\right] \leqslant C.\end{aligned}$$

Hence

$$\int_{-A}^{A} f'^2_\theta(u) \mathbf{E}_\theta\left[\left(\frac{\chi_{\{\xi>u\}} - F_\theta(\xi)}{f_\theta(\xi)}\right)^2\right] \mathrm{d}u \leqslant C$$

and the inequality (19) is proved.

We pass now to the proof of the following estimate

$$\int_{A}^{\infty} \left[f'_\theta(u)\right]^2 \mathbf{E}_\theta\left[\int_0^\xi \frac{\chi_{\{y>u\}} - F_\theta(y)}{f_\theta(y)}\,\mathrm{d}y\right]^2 \mathrm{d}u < DA^2. \tag{26}$$

To prove it, one has to consider the cases $\xi < -A$, $\xi \in [-A, A]$, $\xi \in [A, u]$ and $\xi > u$. In the first, third and fourth cases we obtain (26) proceeding like in the proof of (19). For the second one, we have

$$\begin{aligned}&\int_{A}^{\infty} \left[f'_\theta(u)\right]^2 \mathbf{E}_\theta\left[\int_0^\xi \frac{\chi_{\{y>u\}} - F_\theta(y)}{f_\theta(y)}\,\mathrm{d}y\ \chi_{\{-A\leqslant\xi\leqslant A\}}\right]^2 \mathrm{d}u \\ &\quad \leqslant \int_{A}^{\infty} \left[f'_\theta(u)\right]^2 \left[\int_{-A}^{A} \frac{1}{f_\theta(y)}\,\mathrm{d}y\right]^2 \mathrm{d}u \\ &\quad \leqslant CA^4 \int_{A}^{\infty} \left[f'_\theta(u)\right]^2 \mathrm{d}u = DA^2.\end{aligned} \tag{27}$$

So we proved (26) which implies the estimate (4).

It remains to check the condition (3). We suppose that $A > 2$ and $B > 2$. If $A > B/2$, then

$$\int_{B}^{\infty} \left[f'_\theta(x)\right]^2 \mathrm{d}x \leqslant \int_{0}^{\infty} \left[f'_\theta(x)\right]^2 \mathrm{d}x = \frac{C}{A^2} \leqslant \frac{D}{B^2}\,.$$

For $A \leqslant B/2$, we have

$$\begin{aligned}\int_B^\infty \left[f'_\theta(x)\right]^2 \,\mathrm{d}x &\leqslant C \int_B^\infty (x-A)^{2k+2}\,\mathrm{e}^{-2(x-A)^{k+2}}\,\mathrm{d}x \\ &\leqslant C \int_B^\infty (x-A)^{2k+3}\,\mathrm{e}^{-2(x-A)^{k+2}}\,\mathrm{d}x \\ &= D \int_{(B-A)^{k+2}}^\infty y\,\mathrm{e}^{-2y}\,\mathrm{d}y \leqslant D \int_{(B-A)}^\infty \mathrm{e}^{-2y}\,\mathrm{d}y \\ &\leqslant D\,\mathrm{e}^{-2(B-A)} < D\,\mathrm{e}^{-B} < \frac{D}{B^2}. \end{aligned} \tag{28}$$

This inequality completes the proof of Lemma 2. □

LEMMA 3. *The following relation holds*

$$\mathcal{C}_T = \left\{\theta \in \Gamma_T \;\middle|\; \frac{1}{A^2}\sum_{i\in\mathbb{Z}}\left(\frac{\pi\mathrm{i}}{A}\right)^{2k}\theta_i^2 < R(1-\varepsilon)\right\} \subseteq \left\{\theta \mid S_\theta \in \Sigma_D\right\} \tag{29}$$

if T is large enough.

Proof. Since $\mathcal{C}_T$ is a subset of Γ_T, the conditions (3)–(5) are satisfied for any $\theta \in \mathcal{C}_T$. So, only the condition

$$\sup_{\theta\in\mathcal{C}_T}\int_{\mathbb{R}}\left[f_\theta^{(k+1)}(x)\right]^2\mathrm{d}x \leqslant R$$

needs to be checked. Note that

$$f_\theta^{(k+1)}(x) = \left[2S_\theta^{(k)}(x) + P\left(S_\theta^{(k-1)}(x),\ldots,S_\theta(x)\right)\right]f_\theta(x),$$

where $P(z_1,\ldots,z_k)$ is a polynomial. By Lemma 1

$$S_\theta^{(m)}(x) = o_T(1), \quad m = 0,1,\ldots,k-1.$$

So, on the one hand,

$$\left[P\left(S_\theta^{(k-1)}(x),\ldots,S_\theta(x)\right)\right]^2 = o_T(1) \quad \text{for } x\in[-A,A].$$

On the other hand, using the orthonormality of the trigonometric basis e_i and the definition of $\mathcal{C}_T$, we obtain

$$\begin{aligned} 4\int_{-A}^A\left[S_\theta^{(k)}(x)f_\theta(x)\right]^2\mathrm{d}x &= \frac{1+o_T(1)}{A^2}\int_{-A}^A\left[S_\theta^{(k)}(x)\right]^2\mathrm{d}x \\ &= \frac{1+o_T(1)}{A^2}\sum_{i\in\mathbb{Z}}\left(\frac{\pi\mathrm{i}}{A}\right)^{2k}\theta_i^2 \\ &< R(1-\varepsilon)\left(1+o_T(1)\right) \end{aligned}$$

for any $\theta \in \mathcal{C}_T$. Consequently,

$$\int_{-A}^{A} \left[f_\theta^{(k+1)}(x)\right]^2 \mathrm{d}x = 4\int_{-A}^{A} \left[S_\theta^{(k)}(x) f_0(x)\right]^2 \mathrm{d}x (1 + o_T(1)) \leqslant R(1-\varepsilon)(1+o_T(1)). \tag{30}$$

It can be easily checked that

$$\int_{|x|>A} \left[f_\theta^{(k+1)}(x)\right]^2 \mathrm{d}x \leqslant CA^{-2}. \tag{31}$$

Combining (30) and (31) we obtain

$$\int_{\mathbb{R}} \left[f_\theta^{(k+1)}(x)\right]^2 \mathrm{d}x \leqslant R(1-\varepsilon) + o_T(1).$$

This completes the proof of Lemma 3. □

LEMMA 4. *The probability of the event $S_\theta \notin \Sigma_D$ is exponentially small and consequently*

$$\Lambda(S_\theta \notin \Sigma_D) = o(T^{-1}).$$

Proof. The relation (29) implies

$$\Lambda(S_\theta \notin \Sigma_D) \leqslant \Lambda(\theta \notin \mathcal{C}_T).$$

Remark now that

$$\frac{1}{A^2}\mathbf{E}\left[\sum_{i\in\mathbb{Z}} \left(\frac{\pi \mathrm{i}}{A}\right)^{2k} \theta_i^2\right] = \frac{1}{A^2}\sum_{i\in\mathbb{Z}} \left(\frac{\pi \mathrm{i}}{A}\right)^{2k} \sigma_i(\varepsilon) = R(1-\varepsilon)^{2k+1}\big(1+o_T(1)\big)(1+\varepsilon).$$

Hence, for T sufficiently large

$$\frac{1}{A^2}\mathbf{E}\left[\sum_{i\in\mathbb{Z}} \left(\frac{\pi \mathrm{i}}{A}\right)^{2k} \theta_i^2\right] \leqslant R(1-\varepsilon)^2.$$

The Hoeffding's inequality gives us the following upper bound

$$\begin{aligned}\Lambda(\theta \notin \mathcal{C}_T) &= \Lambda\left\{\frac{1}{A^2}\sum_{i\in\mathbb{Z}} \left(\frac{\pi \mathrm{i}}{A}\right)^{2k} \theta_i^2 \geqslant R(1-\varepsilon)\right\} \\ &\leqslant \Lambda\left\{\frac{1}{A^2}\sum_{i\in\mathbb{Z}} \left(\frac{\pi \mathrm{i}}{A}\right)^{2k} (\theta_i^2 - \mathbf{E}\theta_i^2) \geqslant R\varepsilon(1-\varepsilon)\right\} \\ &\leqslant \exp\left\{-\frac{R^2\varepsilon^2(1-\varepsilon)^2}{2Q}\right\},\end{aligned}$$

where

$$Q = \frac{G^4}{A^4}\sum_{i\in\mathbb{Z}}\left(\frac{\pi \mathrm{i}}{A}\right)^{4k}\sigma_i^2 \leqslant \frac{C}{T^2}\sum_{|i|<\alpha} i^{4k}\left(\left|\frac{\alpha}{i}\right|^k - 1\right)^2 \leqslant \frac{C}{T^2}\sum_{|i|<\alpha} i^{2k}\alpha^{2k}$$
$$\leqslant CT^{-2}\alpha^{4k+1} = CT^{-1/(2k+1)}.$$

So we have

$$\Lambda(S_\theta \notin \Sigma_D) \leqslant \Lambda(\theta \notin \mathcal{C}_T) \leqslant \exp\{-CT^{1/(2k+1)}\}.$$

The Lemma 4 is proved. □

Now everything is ready to finish the proof of Theorem 1.

Note first that we can consider only those estimators $\bar{\vartheta}_T$ for which

$$R_T(\bar{\vartheta}_T, f_0') = \mathbf{E}_0 \int_{\mathbb{R}} \left(\bar{\vartheta}_T(x) - f_0'(x)\right)^2 \mathrm{d}x < 1$$

with $f_0(\cdot) = f_{S_0}(\cdot)$ and S_0 is the function S_θ corresponding to the value $\theta = 0$. The set of all estimators satisfying this inequality will be denoted by $\mathcal{W}_T$. For the other estimators the result is evident.

We have the following obvious inequalities:

$$\begin{aligned}\inf_{\bar{\vartheta}_T\in\mathcal{W}_T}\sup_{S\in\Sigma_D} R_T(\bar{\vartheta}_T, f_S') &\geqslant \inf_{\bar{\vartheta}_T\in\mathcal{W}_T}\sup_{S\in\Sigma_D\cap\Gamma_T} R_T(\bar{\vartheta}_T, f_S')\\ &\geqslant \inf_{\bar{\vartheta}_T\in\mathcal{W}_T}\int_{\Gamma_T} R_T(\bar{\vartheta}_T, \theta)\Lambda(\mathrm{d}\theta) -\\ &\quad - \sup_{\bar{\vartheta}_T\in\mathcal{W}_T}\int_{\Gamma_T\setminus\Sigma_D} R_T(\bar{\vartheta}_T, \theta)\Lambda(\mathrm{d}\theta)\\ &\geqslant \mathcal{R}_T(\Lambda) - \sup_{\bar{\vartheta}_T\in\mathcal{W}_T}\int_{\Gamma_T\setminus\Sigma_D} R_T(\bar{\vartheta}_T, \theta)\Lambda(\mathrm{d}\theta).\end{aligned}$$

We have already found a lower bound for the first term. The second term can be bounded as follows

$$\sup_{\bar{\vartheta}_T\in\mathcal{W}_T}\int_{\Gamma_T\setminus\Sigma_D} R_T(\bar{\vartheta}_T, \theta)\Lambda(\mathrm{d}\theta) \leqslant (8\sup_{\theta\in\Gamma_T}\|f_\theta'\|_2^2 + 2)\Lambda(S_\theta \notin \Sigma_D).$$

It follows from Lemma 1 that the L^2 norm of f_θ' is bounded uniformly on $\theta \in \Gamma_T$. Consequently, using the Lemma 4 we obtain

$$\inf_{\bar{\vartheta}_T\in\mathcal{W}_T}\sup_{S\in\Sigma_D} R_T(\bar{\vartheta}_T, f_S') \geqslant \mathcal{R}_T(\Lambda) - o(T^{-1}).$$

Therefore

$$\lim_{D\to\infty}\liminf_{T\to\infty}\inf_{\bar{\vartheta}_T}\sup_{S\in\Sigma_D} T^{2k/(2k+1)} R_T(\bar{\vartheta}_T, f_S') \geqslant P(k, R)(1-\varepsilon).$$

This completes the proof of inequality (7), since ε can be taken as small as we want. □

3. Efficiency of a Kernel-type Estimator

Now we have a lower bound for the minimax risk, to prove its optimality we have to find an estimator achieving it. So, we investigate in this section the behavior of the following estimator

$$\vartheta_{K,T}(x) = \frac{2}{T}\int_0^T K_T(x - X_t)\,\chi_{\{|X_t|<B_T\}}\,\mathrm{d}X_t,$$

where $K_T(\cdot)$ is a kernel-type function and B_T is a positive number. Let us denote

$$K_T^*(x) = \alpha_{0,T} K^*(x\alpha_{0,T}) \tag{32}$$

with

$$K^*(x) = \frac{1}{\pi}\int_0^1 (1 - u^k)\,\cos(ux)\,\mathrm{d}u$$

and

$$\alpha_{0,T} = \left(\frac{\pi RT(k+1)(2k+1)}{4k}\right)^{1/(2k+1)}.$$

We show in this section that, for $B_T = \sqrt{T}$, the estimator $\vartheta_{K^*,T}$ is asymptotically efficient, that is, this estimator achieves the lower bound obtained in the previous section. This means that the formula (32) gives the optimal kernel in the problem of the first derivative of the invariant density estimation.

Particularly, if $k = 2$, then the optimal kernel has the following form

$$K^*(x) = \frac{2(\sin x - x\cos x)}{\pi x^3}.$$

For $x = 0$ this function is equal to 2/3.

THEOREM 2. *We have*

$$\lim_{D\to\infty}\limsup_{T\to\infty}\inf_{\bar\vartheta_T}\sup_{S\in\Sigma_D} T^{2k/(2k+1)} R_T(\bar\vartheta_T, f_S') \leqslant P(k, R).$$

Proof. It is evident that

$$\inf_{\bar\vartheta_T}\sup_{S\in\Sigma_D} R_T(\bar\vartheta_T, f_S') \leqslant \inf_K \sup_{S\in\Sigma_D} R_T(\vartheta_{K,T}, f_S').$$

Hence, it is sufficient to evaluate the risk

$$R_T(K, f_S') = \mathbf{E}_S\int_{\mathbb{R}} (\vartheta_{K,T}(x) - f_S'(x))^2\,\mathrm{d}x,$$

where $S(\cdot) \in \Sigma_D$ is the unknown trend coefficient.

To evaluate this risk we will use the Fourier transformations. Let us denote

$$\begin{aligned}\varphi_S(\lambda) &= \int_{\mathbb{R}} e^{i\lambda x}\, f_S'(x)\,dx, \qquad \varphi_{K,T}(\lambda) = \int_{\mathbb{R}} e^{i\lambda x}\, \vartheta_{K,T}(x)\,dx,\\ \varphi_K(\lambda) &= \int_{\mathbb{R}} e^{i\lambda x}\, K_T(x)\,dx, \qquad \varphi_T(\lambda) = \frac{1}{T}\int_0^T e^{i\lambda X_t}\, \chi_{\{|X_t|<B_T\}}\,dX_t.\end{aligned}$$

By Parseval's identity

$$R_T(K, f_S') = \frac{1}{2\pi}\,\mathbf{E}_S \int_{\mathbb{R}} \left|\varphi_{K,T}(\lambda) - \varphi_S(\lambda)\right|^2 d\lambda.$$

As the estimator $\vartheta_{K,T}$ is a convolution, its Fourier transform is product of two Fourier transforms. Indeed

$$\begin{aligned}\varphi_{K,T}(\lambda) &= \frac{2}{T}\int_{\mathbb{R}} e^{i\lambda x}\int_0^T K_T(x - X_t)\,\chi_{\{|X_t|<B_T\}}\,dX_t\,dx\\ &= \frac{2}{T}\int_0^T \int_{\mathbb{R}} e^{i\lambda x}\, K_T(x - X_t)\,dx\ \chi_{\{|X_t|<B_T\}}\,dX_t\\ &= \frac{2}{T}\int_0^T e^{i\lambda X_t}\,\varphi_K(\lambda)\chi_{\{|X_t|<B_T\}}\,dX_t = 2\varphi_K(\lambda)\,\varphi_T(\lambda).\end{aligned}$$

So the quadratic risk can be rewritten as

$$\begin{aligned}R_T(K, f_S') &= \frac{1}{2\pi}\,\mathbf{E}_S \int_{\mathbb{R}} |2\varphi_K(\lambda)\varphi_T(\lambda) - \varphi_S(\lambda)|^2\,d\lambda\\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \mathbf{E}_S\left|2\varphi_K(\lambda)\varphi_T(\lambda) - \varphi_S(\lambda)\right|^2 d\lambda\\ &= \frac{2}{\pi}\int_{\mathbb{R}} |\varphi_K(\lambda)|^2\ \mathbf{Var}_S[\varphi_T(\lambda)]\,d\lambda\ +\\ &\quad + \frac{1}{2\pi}\int_{\mathbb{R}} \left|2\varphi_K(\lambda)\mathbf{E}_S\varphi_T(\lambda) - \varphi_S(\lambda)\right|^2 d\lambda.\end{aligned}$$

For the mathematical expectation of $\varphi_T(\lambda)$, we have

$$\begin{aligned}\mathbf{E}_S\varphi_T(\lambda) &= \frac{1}{T}\,\mathbf{E}_S\left[\int_0^T e^{i\lambda X_t}\, S(X_t)\chi_{\{|X_t|<B_T\}}\,dt\right]\\ &= \mathbf{E}_S\left[e^{i\lambda\xi}\, S(\xi)\chi_{\{|\xi|<B_T\}}\right]\\ &= \frac{1}{2}\,\varphi_S(\lambda) - \frac{1}{2}\int_{|u|>B_T} e^{i\lambda u}\, f_S'(u)\,du.\end{aligned}$$

The following lemma describes the behavior of the bias and variance terms and tells us how to choose B_T.

LEMMA 5. *If the function $K_T(\cdot)$ is such that $|\varphi_K(\lambda)| \leqslant 1$ for each λ, then there exists a constant $C = C_D$ such that, for any $S \in \Sigma_D$,*

$$\int_{\mathbb{R}} |\varphi_K(\lambda)|^2 \,\mathbf{Var}_S[\varphi_T(\lambda)]\,\mathrm{d}\lambda \leqslant \left[\frac{\|\varphi_K\|_2}{\sqrt{T}} + \frac{C}{\sqrt{T}} + \frac{CB_T}{T}\right]^2,$$

$$\int_{\mathbb{R}} \left|\, 2\varphi_K(\lambda)\mathbf{E}_S[\varphi_T(\lambda)] - \varphi_S(\lambda)\right|^2 \mathrm{d}\lambda \leqslant \left[\,\left\|(\varphi_K - 1)\varphi_S\right\|_2 + \frac{C}{B_T}\,\right]^2,$$

where $\|\cdot\|_2$ denotes the L^2-norm over $\mathbb{R}$.

Proof. We prove here only the first inequality, the second one can be proved in the same way. Note that Itô formula gives us the following representation

$$\begin{aligned}\varphi_T(\lambda) - \mathbf{E}_S[\varphi_T(\lambda)] \;&=\; H_S(\lambda, X_T) - H_S(\lambda, X_0) + \\ &\quad + \int_0^T \left[\mathrm{e}^{\mathrm{i}\lambda X_t} - g_S(\lambda, X_t)\right] \mathrm{d}W_t\end{aligned}$$

with

$$g_S(\lambda, y) = 2\int_{-B_T}^{B_T} \mathrm{e}^{\mathrm{i}\lambda u} f_S'(u)\, \frac{\chi_{\{u<y\}} - F_S(y)}{f_S(y)}\,\mathrm{d}u$$

and $H_S(\lambda, x) = \int_0^x g_S(\lambda, y)\,\mathrm{d}y$. Consequently,

$$\begin{aligned}\mathbf{Var}_S[\varphi_T(\lambda)] \;&=\; T^{-2}\mathbf{E}_S \Big| H_S(\lambda, X_T) - H_S(\lambda, X_0) + \\ &\quad + \int_0^T \left(\mathrm{e}^{\mathrm{i}\lambda X_t}\, \chi_{\{|X_t|<B_T\}} - g_S(\lambda, X_t)\right) \mathrm{d}W_t \Big|^2.\end{aligned}$$

Using the triangle inequality we get

$$\int_{\mathbb{R}} |\varphi_K(\lambda)|^2 \,\mathbf{Var}_S[\varphi_T(\lambda)]\,\mathrm{d}\lambda \leqslant \left(\sqrt{A_1} + 2\sqrt{A_2} + \sqrt{A_3}\right)^2 \tag{33}$$

with

$$\begin{aligned}A_1 \;&=\; \frac{1}{T^2}\int_{\mathbb{R}} |\varphi_K(\lambda)|^2 \mathbf{E}_S \Big| \int_0^T \mathrm{e}^{\mathrm{i}\lambda X_t}\, \chi_{\{|X_t|<B_T\}}\,\mathrm{d}W_t \Big|^2 \mathrm{d}\lambda \\ &\leqslant\; \frac{1}{T}\int_{\mathbb{R}} |\varphi_K(\lambda)|^2\,\mathrm{d}\lambda, \\ A_2 \;&=\; \frac{1}{T^2}\int_{\mathbb{R}} |\varphi_K(\lambda)|^2 \mathbf{E}_S \big| H_S(\lambda, \xi)\big|^2\,\mathrm{d}\lambda, \\ A_3 \;&=\; \frac{1}{T^2}\int_{\mathbb{R}} |\varphi_K(\lambda)|^2 \mathbf{E}_S \Big| \int_0^T g_S(\lambda, X_t)\,\mathrm{d}W_t \Big|^2 \mathrm{d}\lambda.\end{aligned}$$

To evaluate the term A_2 we use the fact that $|\varphi_K(\lambda)|$ is less than 1. Thus, according Parseval's identity and condition (4) we have

$$\begin{aligned} A_2 &\leqslant \frac{4}{T^2}\,\mathbf{E}_S \int_{\mathbb{R}} \left| \int_{-B_T}^{B_T} \mathrm{e}^{\mathrm{i}\lambda u}\, f_S'(u) \int_0^{\xi} \frac{\chi_{\{y>u\}} - F_S(y)}{f_S(y)}\,\mathrm{d}y\,\mathrm{d}u \right|^2 \mathrm{d}\lambda \\ &= \frac{8\pi}{T^2} \int_{-B_T}^{B_T} \left[f_S'(u)\right]^2 \mathbf{E}_S \left[\int_0^{\xi} \frac{\chi_{\{y>u\}} - F_S(y)}{f_S(y)}\,\mathrm{d}y \right]^2 \mathrm{d}u \leqslant \frac{8\pi D B_T^2}{T^2}\,. \end{aligned}$$

Repeating exactly the same arguments one can check that

$$A_3 \leqslant \frac{8\pi D}{T}.$$

Lemma 5 is proved. □

We choose the kernel-type function $K_T(\cdot)$ in the following way

$$\varphi_K(\lambda) = \varphi_\alpha(\lambda) = \left(1 - \left|\frac{\lambda}{\alpha}\right|^k\right)_+,$$

where $\alpha = \alpha_T$ is a positive number. As we will see below, the integrals of the right-hand sides in Lemma 5 converge both to zero with the rate $T^{-k/(2k+1)}$. Therefore the choice $B_T = \sqrt{T}$ gives us the following upper estimate for quadratic risk

$$R_T(K, f_S') \leqslant L_T(\varphi_\alpha, \varphi_S)\big(1 + o_T(1)\big),$$

where $o_T(1)$ tends to zero uniformly on $S \in \Sigma$ and

$$L_T(\varphi_\alpha, \varphi_S) = \frac{1}{2\pi T} \int_{\mathbb{R}} \big(4|\varphi_\alpha(\lambda)|^2 + T|\varphi_\alpha(\lambda) - 1|^2 |\varphi_S(\lambda)|^2\big)\,\mathrm{d}\lambda.$$

Since the function $S(\cdot)$ is in the ellipsoid $\Sigma(k, R)$, its Fourier transform should belong to the following set

$$\Phi = \left\{ \varphi \;\Big|\; \frac{1}{2\pi} \int_{\mathbb{R}} \lambda^{2k} |\varphi(\lambda)|^2\,\mathrm{d}\lambda \leqslant R \right\}.$$

Replacing in L_T the function φ_α by its explicit expression, we obtain

$$L_T(\varphi_\alpha, \varphi_S) = \frac{2}{\pi T} \int_{-\alpha}^{\alpha} \left(1 - \left|\frac{\lambda}{\alpha}\right|^k\right)^2 \mathrm{d}\lambda + \frac{1}{2\pi\alpha^{2k}} \int_{-\alpha}^{\alpha} |\lambda|^{2k} |\varphi_S(\lambda)|^2\,\mathrm{d}\lambda.$$

Since φ_S belongs to Φ, the second term of the right-hand side is less than R/α^{2k} and the first term can be calculated explicitly:

$$\int_{-\alpha}^{\alpha} \left(1 - \left|\frac{\lambda}{\alpha}\right|^k\right)^2 \mathrm{d}\lambda = \frac{4\alpha k^2}{(k+1)(2k+1)}.$$

It leads us to the following inequality

$$\inf_{\alpha>0} \sup_{S\in\Sigma} L_T(\varphi_\alpha, \varphi_S) \leqslant \inf_{\alpha>0} \left\{ \frac{8k^2\alpha}{\pi T(k+1)(2k+1)} + \frac{R}{\alpha^{2k}} \right\} = \inf_{\alpha>0} G(\alpha).$$

The function $G(\alpha)$ is continuously differentiable and strictly convex, consequently it attains the minimum at the point α_0 satisfying the following equation

$$\frac{8k^2}{\pi T(k+1)(2k+1)} = \frac{2kR}{\alpha_0^{2k+1}},$$

which leads to

$$\alpha_0 = \left(\frac{R\pi T(k+1)(2k+1)}{4k} \right)^{1/(2k+1)}$$

and

$$\inf_{\alpha>0} G(\alpha) = G(\alpha_0) = \frac{(2k+1)R}{\alpha_0^{2k}} = P(k,R)\, T^{-2k/(2k+1)} .$$

This completes the proof of Theorem 2. □

4. Remarks

1. One can use exactly the same arguments to find the Pinsker's constant in the problem of $f_S^{(l)}(\cdot)$ estimation when $S \in \Sigma(k+l-1, R)$. The optimal rate of convergence φ_T and the Pinsker's constant are then

   $\varphi_T = T^{\beta}$,

   $$P_l(k,R) = (1+2\beta)^{-1} \left(\frac{\pi(k+2l-1)(2k+2l-1)}{4k} \right)^{2\beta} R^{1+2\beta}$$

   with $\beta = -k/(2k+2l-1)$.
2. The condition (5) can be replaced by

   $$\sup_{S\in\Sigma} B^{-\tau} \int_{-B}^{B} f_S'^2(x)\, \mathbf{E}_S \left[\frac{\chi_{\{\xi>x\}} - F_S(\xi)}{f_S(\xi)} \right]^2 \mathrm{d}x < \infty,$$

   where τ is positive and less than $2/(2k+1)$. This condition is a little weaker than (5), but it does not enlarge significantly the class of diffusion processes.

References

1. Bosq, D. and Davydov, Y.: Local time and density estimation in continuous time, *Math. Meth. Statist.* **8**(1) (1999), 22–45.
2. Castellana, J. V. and Leadbetter, M. R.: On smoothed density estimation for stationary processes, *Stoch. Proc. Appl.* **21** (1986), 179–193.
3. Durett, R.: *Stochastic Calculus: A Practical Introduction*, CRC Press, Boca Raton, FL, 1996.

4. Gill, R. D. and Levit, B. Ya.: Application of the van Trees inequality: a Bayesian Cramer-Rao bound, *Bernoulli* **1** (1995), 59–79.
5. Gikhman, I. I. and Skorokhod, A. V.: *Introduction to Theory of Random Processes*, Saunders, Philadelphia, PA, 1969.
6. Golubev, G. K.: LAN in problems of non-parametric estimation of functions and lower bounds for quadratic risks, *Theory Probab. Appl.* **36**(1) (1991), 152–157.
7. Golubev, G. K.: Non-parametric estimation of smooth densities in L^2, *Problems Inform. Transmission* **28**(1) (1992), 52–62.
8. Golubev, G. K. and Levit, B. Ya.: On the second order minimax estimation of distribution functions, *Math. Meth. Statist.* **5**(1) (1996), 1–31.
9. Kutoyants, Yu. A.: Efficient density estimation for ergodic diffusion processes, *Statist. Inf. Stoch. Proc.* **1**(2) (1998), 131–155.
10. Leblanc, F.: Density estimation for a class of continuous time process, *Math. Meth. Statist.* **6**(2) (1997), 171–199.
11. Lucas, A.: Can we estimate the density's derivative with suroptimal rate? *Statist. Inf. Stoch. Proc.* **1**(1) (1998), 29–41.
12. Nguyen, H. T.: Density estimation in a continuous-time Markov processes, *Ann. Statist.* **7** (1979), 341–348.
13. Nussbaum, M.: Minimax risk: Pinsker's bound. In: S. Kotz (ed.), *Encyclopedia of Statistical Sciences*, Vol. 3, Wiley, New York, 1999, pp. 451–460.
14. Pinsker, M. S.: Optimal filtering of square integrable signals in Gaussian white noise, *Problems Inform. Transmission* **16** (1980), 120–133.
15. Schipper, M.: Optimal rates and constants in L^2-minimax estimation of probability density functions, *Math. Meth. Statist.* **5**(3) (1996), 253–274.