ASYMPTOTICALLY EFFICIENT TREND COEFFICIENT ESTIMATION FOR ERGODIC DIFFUSION

A. S. DALALYAN AND YU. A. KUTOYANTS

ABSTRACT. The problems of nonparametric estimation of the derivative of the invariant density function and of the trend coefficient are considered for a one dimensional ergodic diffusion process in the asymptotics of large samples. In both problems the lower minimax (local) bounds on the L^2-type risks of all estimators are proposed and the asymptotically efficient (up to the Pinsker's constant) in the sense of these bounds estimators are constructed.

1. INTRODUCTION

We consider a diffusion process $X^T = \left\{X_t,\ 0 \le t \le T\right\}$ given by the stochastic differential equation

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

where $W^T = \left\{W_t,\ 0 \le t \le T\right\}$ is the standard Wiener process and the initial value ξ is independent of W^T. We suppose that the function $S(\cdot)$ is unknown and we have to estimate it by the observations X^T in the asymptotics of large samples, i.e.; as $T \to \infty$. We suppose as well that this function is k-times differentiable and satisfies the condition

$$\overline{\lim_{|x|\to\infty}}\ \mathrm{sgn}(x)\, S(x) < 0. \tag{2}$$

Note that as the function $S(\cdot)$ is locally Lipschitz and satisfies the condition

$$xS(x) \le B$$

with some $B \ge 0$, the stochastic differential equation (1) has a unique strong solution (see [6], p. 190). By condition (2), the process (1) has a unique invariant distribution with the density function

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

2000 *Mathematics Subject Classification.* 62M05, 62G07, 62G20.

Key words and phrases. ergodic diffusion, invariant density, derivative estimation, trend estimation, asymptotic efficiency, Pinsker's constant.

where

$$G(S) = \int_{\mathbb{R}} \exp\left\{2\int_0^x S(v)\,\mathrm{d}v\right\}\mathrm{d}x$$

is the normalizing constant (see [11]). Throughout this paper we suppose that the initial value ξ follows the invariant law, implying thus the stationarity of the process X^T.

Note that the problem of trend estimation is closely related to the problem of the estimation of the derivative $f'_S(x)$ of the invariant density because we have the elementary equality

$$S(x) = \frac{f'_S(x)}{2f_S(x)}. \tag{3}$$

Therefore, to estimate $S(x)$ we have to estimate at first the functions $f'_S(x)$ and $f_S(x)$. Remind that the density $f_S(x)$ can be estimated with the rate $\sqrt{T}$ [16] and the derivative $f'_S(x)$ can be estimated with the rate $T^{\frac{k}{2k+1}}$ [4],[30], which depends on the smoothness of the function $S(\cdot)$. Therefore the main contribution to the limit variance of the estimator of $S(x)$ is due to the derivative estimator.

The problems of the invariant density and trend estimation were considered by many authors (see [1], [2], [16], [19], [25], [26] for density estimation and [1], [8], [22], [25], [28] for trend estimation). The problem of asymptotically efficient estimation of the derivative (lower bound with Pinsker's constant and efficient estimator) was treated in [4] but the type of the $\mathcal{L}_2$ risk (global) does not fit well to the problem of trend estimation. That is why in the present work we establish first the lower local minimax bound on the risk of all estimators of the derivative and construct an asymptotically efficient estimator in the sense of this bound. Then we use these results in the problem of trend estimation. The possible generalizations and some heuristic arguments are discussed in Section 4. Some of the technical results are stated in Appendix.

There is an essential difference between the problem of trend coefficient estimation and the estimation of the derivative of the invariant density. In the problem of derivative estimation, as it is explained in Subsection 4.2, the Fisher information $I(S, x)$ at a point x is equal to $f_S^{-1}(x)$. This implies, in particular, that the function $I^{-1}(S, \cdot)$ is integrable and its integral over $\mathbb{R}$ is equal to one. This property permitted us (see [4]) to obtain the Pinsker's constant for the estimation of derivative in a global setting. To understand better the relation between the optimal constant and the inverse of the Fisher information, the reader is referred to the work [13].

In contrast with the problem of derivative estimation, the Fisher information of the problem of trend coefficient estimation at the point x is $I_{trend}(S,x) = f_S(x)$, the inverse of which is evidently not integrable over $\mathbb{R}$. Moreover, the supremum over $S \in \Sigma$ of the integral of $I^{-1}_{trend}(S,\cdot)$ even over a bounded interval is equal to infinity for reasonable nonparametric classes Σ. This explains why we consider in this paper the local minimax risk. Furthermore, in the problem of trend estimation we define the risk as the mean integrated squared error weighted by the function $f^2_S(\cdot)$, in order to obtain a finite optimal constant, independent of the function S.

This work lies in the stream of the studies initiated by M. S. Pinsker [27] in 1980 (see [24] for more details and references) and is inspired by the works [13], [14] and [15] (see as well [29], where a similar method is used). According to this approach, the $\mathcal{L}_2$ risk of nonparametric estimation problem is minorated by the risk of a sequence of parametric estimation problems with a vector parameter of growing dimension. This parameter is supposed to be a random vector and we seek the least favorable a priori distribution of this parameter. The solution of this problem provides the lower minimax bound in the original nonparametric estimation problem.

The next step is to construct an estimator which is asymptotically efficient in the sense of this bound, that is an estimator which attains asymptotically this bound. This program was already realized for i. i. d. model of observations (see [7]) as well as for some others (signal in Gaussian white noise [27], nonhomogeneous Poisson process [17], regression models [23]) and the present work adds the one-dimensional ergodic diffusion process on this list.

Note that there is also another possibility for obtaining lower bounds in this type of nonparametric estimation problems. It concerns the asymptotic equivalence in Le Cam's sense of the initial experiment to another experiment with known lower bound. Actually there are some results of asymptotic equivalence for statistical models related to diffusion processes (see [5], [9], [10] and [21]), but none of them permit to derive a lower bound in our setup.

2. Estimation of the Derivative of the Invariant Density

We suppose that the trend coefficient $S(\cdot)$ of the diffusion process (1) is k-times differentiable, satisfies the condition (2) and we have to estimate the derivative $f'_S(x) = 2S(x)\, f_S(x)$ by the observations X^T.

Fix a trend coefficient $S_0(\cdot)$ and introduce its δ-neighborhood in the uniform metrics

$$V_\delta(S_0) = \Big\{ S : \mathbb{R} \to \mathbb{R} \;\Big|\; \sup_x \big|S(x) - S_0(x)\big| \le \delta \Big\}.$$

Then define the set

$$\Sigma_\delta(k, R, S_0) = \Big\{ S \in V_\delta(S_0) \;\Big|\; \int_{\mathbb{R}} \big[f_S^{(k+1)}(x) - f_{S_0}^{(k+1)}(x) \big]^2 \mathrm{d}x \le 4R \Big\},$$

which is the parameter space in our problem. To simplify the notation we will write $\Sigma_\delta(k, R, S_0)$ as Σ_δ and $f_{S_0}(\cdot)$ as $f_0(\cdot)$.

We measure the quality of estimation of the function $f_S'(\cdot)$ by an estimator $\bar\vartheta_T(\cdot)$ with the help of the $\mathcal{L}_2$-type risk function

$$R_T(\bar\vartheta_T, f_S') = \mathbf{E}_S \int_{\mathbb{R}} \big(\bar\vartheta_T(x) - f_S'(x) \big)^2 \mathrm{d}x.$$

We define the local minimax risk as

$$\tilde r_\delta(T) = \inf_{\bar\vartheta_T} \sup_{S \in \Sigma_\delta} R_T(\bar\vartheta_T, f_S'),$$

where the infimum is taken over all possible estimators $\bar\vartheta_T(\cdot)$. This risk is called local because the supremum is taken over all $S(\cdot)$ which belong to a shrinking neighborhood of a central function $S_0(\cdot)$.

The goal of this section is to find the lower bound on this risk function and then to construct an estimator which attains this bound.

2.1. Conditions.

Naturally we suppose that the function S_0 is k-times continuously differentiable and satisfies the condition

$$\overline{\lim_{|x| \to \infty}} \operatorname{sgn}(x) S_0(x) < 0. \tag{4}$$

We need two additional conditions on the "central" function $S_0(\cdot)$.

Suppose that the function $S_0(\cdot)$ has a polynomial majorant, that is there exist two positive constants C and ν such that

$$\big|S_0(x)\big| \le C\big(1 + |x|^\nu\big) \quad \text{for any } x \in \mathbb{R}. \tag{5}$$

Moreover, for some $\tau > 0$, the integral

$$\int_{\mathbb{R}} \lambda^{2k+\tau} \big|\varphi_0(\lambda)\big|^2 \mathrm{d}\lambda \tag{6}$$

is finite, where $\varphi_0(\cdot)$ is the Fourier transform of $f_0'(\cdot) = f_{S_0}'(\cdot)$. This condition means that the function $S_0(\cdot)$ is a little bit smoother than the other functions $S(\cdot)$ of Σ_δ. For example, if S_0 is $(k+1)$-times continuously differentiable and $S_0^{(k+1)}$ has a polynomial majorant, then this condition is satisfied with $\tau = 2$.

2.2. **Lower Bound.** To establish a lower bound we use some ideas developed in [13]. Firstly we restrict ourselves to the problem of the estimation of $f'_S(\cdot)$, for $S(\cdot)$ belonging to a properly chosen parametric family which is essentially concentrated on Σ_δ, and where minimax estimation is nearly as difficult as for the vicinity Σ_δ. The dimension of this family increases to infinity when T tends to infinity. Then we use a minoration of the minimax risk by the Bayesian one with respect to any prior distribution, plus a term which is asymptotically negligible. The last step is to find the worst prior and to evaluate the corresponding Bayesian risk. We obtain so the following result.

Theorem 1. *Suppose that the central function $S_0(\cdot)$ satisfies the conditions* (4)–(6)*, then*

$$\varliminf_{\delta\to 0}\ \varliminf_{T\to\infty} T^{\frac{2k}{2k+1}}\, \tilde r_\delta(T) \geq 4P(k,R),$$

where

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

Proof. The proof of this theorem is quite close to the proof of the lower bound in [4]. The main difference is in the construction of the parametric family (inspired by [13]).

We fix an increasing interval $[-A,A]$ with $A = A_T = \log(T+1)$ and a sequence of its sub-intervals $I_m = \big[a_m - AT^{-\beta}, a_m + AT^{-\beta}\big]$, where $\beta = (2k+1)^{-1}$ and

$$a_m = 2mAT^{-\beta}, \qquad m = 0, \pm 1, \pm 2, \dots, \pm M.$$

Here $M = M_T$ is the greatest integer such that $I_M \subseteq [-A,A]$. Let us introduce now the parameterization

$$S(\vartheta,x) = S_0(x) + \sum_{|m|<M} \sqrt{\frac{2A}{T^\beta f_0(a_m)}} \sum_{1\leq|i|\leq L} \vartheta_{i,m}\phi_{i,m}(x),$$

where

$$\phi_{i,m}(x) = \sqrt{T^\beta/A}\; e_i\big(T^\beta A^{-1}(x-a_m)\big)\, U(A - |x-a_m|T^\beta).$$

Here $e_i(\cdot)$ is the trigonometric basis on $[-1,1]$, that is

$$e_i(x) = \begin{cases} \sin(\pi i x)\ , & \text{if } i > 0, \\ 1/\sqrt{2} \ , & \text{if } i = 0, \\ \cos(\pi i x)\ , & \text{if } i < 0, \end{cases}$$

the function $U(x)$ is $(k+1)$-times differentiable, increasing, vanishing for $x \le 0$ and equal to one for $x \ge 1$. The integer $L = L_T$ will be chosen later.

It is easy to show that the functions $S(\vartheta,\cdot)$ are k-times continuously differentiable and coincide with S_0 outside of the interval $[-A,A]$. Note also that the function $U(A - |x-a_m|T^\beta)$ is a smooth approximation of the indicator function $\chi_{I_m}(x)$.

In order to avoid complicated indices we use the following notations: $f(\vartheta,x)$ denotes $f_{S(\vartheta,\cdot)}(x)$ and $f'(\vartheta,x)$ is the derivative of $f(\vartheta,x)$ with respect to x.

The parametric space Γ_T that we consider is the set of all finite sequences $\{\vartheta_{i,m}\}_{1\le|i|\le L, |m|\le M}$ such that $|\vartheta_{i,m}| \le G\sqrt{\sigma_i(\varepsilon)}$ for all i and m. Here G and ε are two positive numbers to be defined later and

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

with

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

and $\sigma_0 = 0$, where $a_+ = \max(a,0)$. The integer L is chosen to be equal $[\alpha]$ (integer part of α). Note that the dimension of the space Γ_T increases like $T^\beta A_T$, as $T \to \infty$.

We define now the set $\mathcal{W}_T = \{\bar\vartheta_T \mid R_T(\bar\vartheta_T, f_0') < 1\}$. It is evident that in the proof of lower bound one can drop all the estimators do not belonging to $\mathcal{W}_T$. Thus, for any probability distribution Λ on Γ_T, the following chain of inequalities holds

$$\begin{aligned} \tilde r_\delta(T) &\ge \inf_{\bar\vartheta_T\in\mathcal{W}_T} \sup_{S\in\Sigma_\delta} R_T(\bar\vartheta_T, f_S') \ge \inf_{\bar\vartheta_T\in\mathcal{W}_T} \sup_{S\in\Sigma_\delta\cap\Gamma_T} R_T(\bar\vartheta_T, f_S') \\ &\ge \inf_{\bar\vartheta_T\in\mathcal{W}_T} \int_{\Gamma_T} R_T\big(\bar\vartheta_T, f'(\vartheta,\cdot)\big)\,\mathrm{d}\Lambda(\vartheta) \\ &\qquad - \sup_{\bar\vartheta_T\in\mathcal{W}_T} \int_{\Gamma_T\setminus\Sigma_\delta} R_T\big(\bar\vartheta_T, f'(\vartheta,\cdot)\big)\,\mathrm{d}\Lambda(\vartheta) \\ &\ge \mathcal{R}_T(\Lambda) - \sup_{\bar\vartheta_T\in\mathcal{W}_T} \int_{\Gamma_T\setminus\Sigma_\delta} R_T\big(\bar\vartheta_T, f'(\vartheta,\cdot)\big)\,\mathrm{d}\Lambda(\vartheta). \end{aligned} \tag{7}$$

Here $\mathcal{R}_T(\Lambda)$ denotes the Bayesian risk with respect to the prior distribution Λ in the problem of $f'(\vartheta,\cdot)$ estimation over the parameter set Γ_T, that is

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

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

The next step of the proof is to choose a prior distribution Λ which maximizes asymptotically the Bayesian risk and which is essentially concentrated on the set $\left\{\vartheta \in \Gamma_T \,\middle|\, S(\vartheta,\cdot) \in \Sigma_\delta\right\}$. This Λ is called asymptotically least favorable prior distribution.

Some heuristic ideas presented in [15] allow us to choose the prior distribution in the following way. We fix a continuously differentiable probability density $p(x)$ with support in $[-G,\, G]$ and such that

$$\int_{\mathbb{R}} x\, p(x)\, \mathrm{d}x = 0, \qquad \int_{\mathbb{R}} x^2 p(x)\, \mathrm{d}x = 1, \qquad J = \int_{\mathbb{R}} \frac{[p'(x)]^2}{p(x)}\, \mathrm{d}x = 1+\varepsilon.$$

Note that these conditions imply that ε and G are connected and ε decreases to zero when G increases to infinity. The prior distribution Λ can be defined now as the product measure

$$\mathrm{d}\Lambda(\vartheta) = \prod_{1\le|i|\le L} \prod_{|m|\le M} \lambda_{i,m}(\vartheta_{i,m})\, \mathrm{d}\vartheta_{i,m}$$

with

$$\lambda_{i,m}(u) = \lambda_i(u) = \frac{1}{\sqrt{\sigma_i(\varepsilon)}}\, p\left(\frac{u}{\sqrt{\sigma_i(\varepsilon)}}\right)$$

for all integers i different from 0. The Fisher information $J_{i,m}$ of this prior distribution Λ with respect to the coefficient $\vartheta_{i,m}$ is then equal to $(1+\varepsilon)/\sigma_i(\varepsilon)$.

In order to evaluate the Bayesian risk $\mathcal{R}_T(\Lambda)$, we denote by $\psi_{i,m,\vartheta}$ and $\psi_{i,m,T}$ the Fourier coefficients on $[-A,A]$ of $f'(\vartheta,\cdot)$ and $\bar\vartheta_T(\cdot)$ with respect to the orthonormal sequence

$$e_{i,m}(x) = \sqrt{T^\beta/A}\; e_i\left(T^\beta A^{-1}(x-a_m)\right)\chi_{\{x\in I_m\}}.$$

This means that

$$\psi_{i,m,\vartheta} = \int_{-A}^{A} f'(\vartheta,x) e_{i,m}(x)\mathrm{d}x,$$

$$\psi_{i,m,T} = \int_{-A}^{A} \bar\vartheta_T(x) e_{i,m}(x)\mathrm{d}x.$$

Since the $L^2[-A,A]$-norm of the function $\bar\vartheta_T(\cdot)-f'(\vartheta,\cdot)$ is greater than the norm of its projection on the subspace of $L^2[-A,A]$ generated by the orthonormal sequence $\{e_{i,m}\}_{1\le|i|\le L,|m|\le M}$, we have

$$\begin{aligned}\mathcal{R}_T(\Lambda) &\ge \inf_{\bar\vartheta_T}\mathbb{E}\int_{-A}^{A}\big(\bar\vartheta_T(x)-f'(\vartheta,x)\big)^2\mathrm{d}x \\ &\ge \inf_{\bar\vartheta_T}\sum_{|m|\le M}\sum_{1\le|i|\le L}\mathbb{E}\big(\psi_{i,m,T}-\psi_{i,m,\vartheta}\big)^2.\end{aligned} \tag{8}$$

According to van Trees inequality (see [12]), we have

$$\mathcal{R}_T(\Lambda) \ge \sum_{|m|\le M}\sum_{1\le|i|\le L}\frac{\big(\mathbb{E}\big[\partial\psi_{i,m,\vartheta}/\partial\vartheta_{i,m}\big]\big)^2}{\int I_{i,m}(\vartheta)\,\mathrm{d}\Lambda(\vartheta)+J_{i,m}}\ , \tag{9}$$

where $I_{i,m}(\vartheta)$ is defined by the formula

$$\begin{aligned}I_{i,m}(\vartheta) &= \mathbf{E}_\vartheta\Bigg[\frac{\partial\log f(\vartheta,X_0)}{\partial\vartheta_{i,m}}+\int_0^T\frac{\partial S(\vartheta,X_t)}{\partial\vartheta_{i,m}}\,\mathrm{d}X_t \\ &\qquad -\frac{1}{2}\int_0^T\frac{\partial S^2(\vartheta,X_t)}{\partial\vartheta_{i,m}}\,\mathrm{d}t\Bigg]^2 \\ &= \mathbf{E}_\vartheta\Bigg[\frac{\partial\log f(\vartheta,X_0)}{\partial\vartheta_{i,m}}+\int_0^T\frac{\partial S(\vartheta,X_t)}{\partial\vartheta_{i,m}}\,\mathrm{d}W_t\Bigg]^2.\end{aligned}$$

This expression of the Fisher information can be obtained immediately from the explicit form of the likelihood ratio (see, for example, [20] Theorem 7.7). Note that the stochastic integral and the differential operator $\partial/\partial\vartheta_{i,m}$ are interchangeable in our case, since the dependence of $S(\vartheta,\cdot)$ on ϑ is linear.

Since W is independent of X_0 and the expectation of a stochastic integral with respect to a local martingale is zero, the cross term in $I_{i,m}(\vartheta)$ is equal to zero. Therefore, we have

$$I_{i,m}(\vartheta)=\mathbf{E}_\vartheta\left[\frac{\partial\log f(\vartheta,X_0)}{\partial\vartheta_{i,m}}\right]^2+\mathbf{E}_\vartheta\left[\int_0^T\frac{\partial S(\vartheta,X_t)}{\partial\vartheta_{i,m}}\,\mathrm{d}W_t\right]^2.$$

By a simple differentiation of the invariant density we obtain

$$\frac{\partial\log f(\vartheta,x)}{\partial\vartheta_{i,m}}=2\int_0^x\frac{\partial S(\vartheta,v)}{\partial\vartheta_{i,m}}\,\mathrm{d}v-2\mathbf{E}_\vartheta\left[\int_0^\xi\frac{\partial S(\vartheta,v)}{\partial\vartheta_{i,m}}\,\mathrm{d}v\right].$$

Using the fact that only $CT^\beta A_T$ components of the vector $\vartheta\in\Gamma_T$ are non-zero and each one is less than $CT^{-k\beta}A_T^k$, one can easily show that

$$f(\vartheta,x)=f_0(x)\big(1+o_T(1)\big),$$

where $o_T(1)$ is uniform on $x \in \mathbb{R}$ and $\vartheta \in \Gamma_T$. Further, since $|a_m| \leq A$ and S_0 has a polynomial majorant, using Taylor formula one can easily check that

$$f_0(x) = f_0(a_m)\big(1 + o_T(1)\big),$$

for any $x \in I_m$. On the other hand, due to the property (4), there exists an interval $[-b, b]$ such that the function $f_0(\cdot)$ is bounded above and below on this interval by some strictly positive constants and satisfies the property $\operatorname{sgn}(x)S(x) < -\gamma < 0$ outside of this interval. Consequently, $\sup_{x \geq a_m} f_0(x) \leq Ce^{-\gamma(x-a_m)} f_0(a_m)$ for any positive a_m and a similar estimate is true for $a_m < 0$. Therefore, for any $a_m > 0$, we have

$$\mathbf{E}_\vartheta \left(\frac{\partial \log f(\vartheta, X_0)}{\partial \vartheta_{i,m}} \right)^2 \leq C \mathbf{E}_\vartheta \left[\chi_{\{\xi > a_m - AT^{-\beta}\}} f_0(a_m)^{-1} \right] \leq C. \tag{10}$$

At the same time, due to the stationarity of the process X^T and the form of the parameterization, we have

$$\begin{aligned} \mathbf{E}_\vartheta \left[\int_0^T \frac{\partial S(\vartheta, X_t)}{\partial \vartheta_{i,m}} \, \mathrm{d}W_t \right]^2 &= \frac{2AT}{T^\beta f_0(a_m)} \, \mathbf{E}_\vartheta \left[\phi_{i,m}(\xi) \right]^2 \\ &= 2AT^{1-\beta}\big(1 + o_T(1)\big). \end{aligned} \tag{11}$$

Combining (10) and (11) for T sufficiently large we get

$$I_{i,m} \leq 2AT^{1-\beta}(1 + \varepsilon).$$

Using the same kind of arguments (see also the proof of relation (17) in [4]), we obtain

$$\begin{aligned} \frac{\partial \psi_{i,m,\vartheta}}{\partial \vartheta_{i,m}} &= 2 \int_{I_m} \frac{\partial S(\vartheta, x)}{\partial \vartheta_{i,m}} \, f(\vartheta, x) e_{i,m}(x) \, \mathrm{d}x \big(1 + o_T(1)\big) \\ &= 2 \int_{I_m} \sqrt{\frac{2A}{T^\beta f_0(a_m)}} \, f_0(x) e_{i,m}^2(x) \, \mathrm{d}x \big(1 + o_T(1)\big) \\ &= 2\sqrt{2A f_0(a_m) T^{-\beta}} \, \big(1 + o_T(1)\big), \end{aligned}$$

where $o_T(1)$ tends to zero uniformly on i, m and $\vartheta \in \Gamma_T$.

Now the inequality (9) can be rewritten like follows:

$$\mathcal{R}_T(\Lambda) \geq \big(1 + o_T(1)\big) \sum_{|m| \leq M} \sum_{|i| \leq L} \frac{8AT^{-\beta} f_0(a_m)}{2AT^{1-\beta}(1 + \varepsilon) + J_{i,m}} \,. \tag{12}$$

Using the convergence of Riemann's sums of a continuous function to its integral, it is easy to show that

$$\sum_{|m| \leq M} 2AT^{-\beta} f_0(a_m) = \int_{-A}^{A} f_0(x) \, \mathrm{d}x \big(1 + o_T(1)\big) = 1 + o_T(1). \tag{13}$$

Therefore, using once more the convergence of Riemann's sums, we get

$$\begin{aligned}\mathcal{R}_T(\Lambda) &\geq \frac{1+o_T(1)}{1+\varepsilon}\sum_{|i|\leq L}\frac{4\sigma_i(\varepsilon)}{2AT^{2k\beta}\sigma_i(\varepsilon)+1}\\ &= \frac{4\big(1+o_T(1)\big)}{AT^{2k\beta}(1+\varepsilon)}\sum_{i=1}^{L}\left(1-\left|\frac{i}{\alpha(1-\varepsilon)}\right|^k\right)\\ &= \frac{4\alpha(1-\varepsilon)\big(1+o_T(1)\big)}{AT^{2k\beta}(1+\varepsilon)}\int_0^1(1-x^k)\,\mathrm{d}x\\ &\geq 4(1-\varepsilon)^2T^{-2k/(2k+1)}P(k,R)\big(1+o_T(1)\big).\end{aligned}$$

To finish the proof of theorem, we have to show that the second term in (7) tends to zero faster than the first term. Proceeding exactly like in Lemma 3 of [4], one can check that

$$\begin{aligned}\int_{\mathbb{R}}&\big[\big(S(\vartheta,x)f(\vartheta,x)-S_0(x)f_0(x)\big)^{(k)}\big]^2\mathrm{d}x\\ &= \int_{\mathbb{R}}\big[S^{(k)}(\vartheta,x)-S_0^{(k)}(x)\big]^2f_0^2(x)\mathrm{d}x\,\big(1+o_T(1)\big)\\ &= \frac{2A}{T^\beta}\sum_{|m|\leq M}\sum_{|i|\leq L}f_0(a_m)\vartheta_{i,m}^2\left(\frac{\pi i T^\beta}{A}\right)^{2k}\big(1+o_T(1)\big).\end{aligned}$$

Let us denote the last sum by $V_T(\vartheta)$ and introduce a subset of $\mathbb{R}^{\mathbb{Z}\times\mathbb{Z}}$ as

$$\mathcal{C}_T=\Big\{\vartheta=\{\vartheta_{i,m}\}\;\Big|\;V_T(\vartheta)\leq R(1-\varepsilon)\Big\}.$$

For T sufficiently large, the set $\{S(\vartheta,\cdot)\,|\,\vartheta\in\mathcal{C}_T\}$ is included in the set Σ_δ. Moreover, using (13), we have

$$\begin{aligned}\mathbf{E}\big[V_T(\vartheta)\big] &= 2AT^{(2k-1)\beta}\sum_{|m|\leq M}f_0(a_m)\sum_{|i|\leq L}\sigma_i(\varepsilon)\Big(\frac{\pi i}{A}\Big)^{2k}\big(1+o_T(1)\big)\\ &= \sum_{|i|\leq L}T^{2k\beta}\sigma_i(\varepsilon)\Big(\frac{\pi i}{A}\Big)^{2k}\big(1+o_T(1)\big)\leq R(1-\varepsilon)^{2k}\big(1+o_T(1)\big).\end{aligned}$$

We can apply now the Hoeffding's inequality to show that the Λ-measure of the complement set $\mathcal{C}_T^c$ decreases like an exponent, consequently

$$\Lambda\big(S(\vartheta,\cdot)\notin\Sigma_\delta\big)\leq\Lambda(\mathcal{C}_T^c)=o(T^{-1}).$$

Now the second term of (7) can be evaluated as follows

$$\sup_{\bar\vartheta_T\in\mathcal{W}_T}\int_{\Gamma_T\setminus\Sigma_\delta}R_T(\bar\vartheta_T,\vartheta)\,\mathrm{d}\Lambda(\vartheta)\leq\big(8\sup_{\vartheta\in\Gamma_T}\|f'(\vartheta,\cdot)\|_2^2+2\big)\,\Lambda\big(S(\vartheta,\cdot)\notin\Sigma_\delta\big).$$

It can be checked that the L^2-norm of $f'(\vartheta,\cdot)$ is bounded uniformly on $\vartheta \in \Gamma_T$. Consequently,

$$\tilde{r}_\delta(T) \geq \mathcal{R}_T(\Lambda) - o(T^{-1}) \geq 4(1-\varepsilon)^2 T^{-2k\beta} P(k,R)\big(1+o_T(1)\big).$$

Therefore

$$\lim_{\delta\to 0}\lim_{T\to\infty} T^{\frac{2k}{2k+1}}\,\tilde{r}_\delta(T) \geq 4P(k,R)(1-\varepsilon)^2.$$

This completes the proof of the theorem, since ε can be taken as small as we want. □

2.3. **Asymptotically Efficient estimator.** Now we have an asymptotically lower bound for the minimax risk. To show that this bound is sharp we have to find an estimator attaining it. In the most models of nonparametric curve estimation where the Pinsker's constant is obtained, the lower bound is attained by the estimators minimizing the linear risk. In our problem the analog of the linear risk is the risk of kernel-type estimators. That is

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

where $\vartheta_{K,T}$ is any kernel-type estimator defined as

$$\vartheta_{K,T}(x) = \frac{2}{T}\int_0^T K_T(x - X_t)\,\mathrm{d}X_t. \tag{14}$$

The function $K_T(\cdot)$ is called kernel and is, for the moment, any squared integrable function. The corresponding risk is said to be linear because of the linearity of the Fourier transform of $\vartheta_{K,T}$ with respect to the kernel-type function's Fourier transform (see the relation (16) below).

Let us denote

$$K_T^*(x) = \frac{\alpha_{0,T}}{\pi}\int_0^1 (1 - u^{k+\rho_T})\cos(u\alpha_{0,T}x)\,du \tag{15}$$

with

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

and $\rho_T = 1/\log\log(T+1)$. We show in this section that the estimator $\vartheta_{K^*,T}$ is asymptotically efficient, in the sense that this estimator achieves the lower bound obtained in the previous section. This means that the formula (15) gives an asymptotically optimal kernel in the problem of the first derivative of the invariant density estimation.

Definition 1. An estimator $\hat{\vartheta}_T(\cdot)$ is said to be asymptotically efficient in the problem of $f_S'(\cdot)$ estimation, if

$$\lim_{\delta\to 0}\lim_{T\to\infty}\sup_{S\in\Sigma_\delta} T^{\frac{2k}{2k+1}}\,\mathbf{E}_S\int_{\mathbb{R}}\left(\hat{\vartheta}_T(x)-f_S'(x)\right)^2\mathrm{d}x = 4P(k,R).$$

Theorem 2. *Let the conditions* (4)–(6) *be satisfied. Then the kernel-type estimator* $\vartheta_{K^*,T}(\cdot)$ *is asymptotically efficient.*

Proof. To prove this result we will proceed in the following way. We consider first the linear minimax risk over the class of all kernel-type functions which are square integrable. Then we seek the best kernel among all functions whose Fourier transform has the form

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

that is we minimize the risk of the corresponding estimator with respect to $\alpha>0$, tending to infinity not very slowly. We prove that this minimum is asymptotically equal to $4T^{\frac{2k}{2k+1}}P(k,R)$ and is attained at the point $\alpha_0=\alpha_{0,T}$. Finally we apply the inverse Fourier transform to obtain

$$\begin{aligned}
K_T^*(x)&=\frac{1}{2\pi}\int_{\mathbb{R}}e^{i\lambda x}\left(1-\left|\frac{\lambda}{\alpha_0}\right|^{k+\rho}\right)_+\mathrm{d}\lambda\\
&=\frac{1}{\pi}\int_0^{\alpha_0}\cos(\lambda x)\left(1-\left|\frac{\lambda}{\alpha_0}\right|^{k+\rho}\right)\mathrm{d}\lambda\\
&=\frac{\alpha_0}{\pi}\int_0^1\cos(\alpha_0 ux)\left(1-u^{k+\rho}\right)\mathrm{d}\lambda.
\end{aligned}$$

Let us present now the proof in details. As we said, our aim is to find an upper bound for the linear risk $R_T^L(K,f_S')$, where $S(\cdot)\in\Sigma_\delta$ is the unknown trend coefficient and K is a squared integrable function. To evaluate this risk we will use the Fourier transformations. Let us denote

$$\varphi_S(\lambda)=\int_{\mathbb{R}}e^{i\lambda x}f_S'(x)\,\mathrm{d}x,\qquad \varphi_{K,T}(\lambda)=\int_{\mathbb{R}}e^{i\lambda x}\vartheta_{K,T}(x)\,\mathrm{d}x,$$
$$\varphi_K(\lambda)=\int_{\mathbb{R}}e^{i\lambda x}K_T(x)\,\mathrm{d}x,\qquad \varphi_T(\lambda)=\frac{1}{T}\int_0^T e^{i\lambda X_t}\mathrm{d}X_t.$$

By Parseval's identity

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

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

$$\varphi_{K,T}(\lambda) = \frac{2}{T}\int_{-\infty}^{\infty} e^{i\lambda x}\int_0^T K_T(x-X_t)\,\mathrm{d}X_t\,\mathrm{d}x \tag{16}$$
$$= \frac{2}{T}\int_0^T\int_{-\infty}^{\infty} e^{i\lambda x}K_T(x-X_t)\,\mathrm{d}x\,\mathrm{d}X_t$$
$$= \frac{2}{T}\int_0^T e^{i\lambda X_t}\varphi_K(\lambda)\,\mathrm{d}X_t = 2\varphi_K(\lambda)\,\varphi_T(\lambda).$$

It is easy to verify that the mathematical expectation of $2\varphi_T(\lambda)$ is equal to $\varphi_S(\lambda)$. Therefore the quadratic risk can be rewritten as

$$R_T^L(K,f_S') = \frac{1}{2\pi}\,\mathbf{E}_S\int_{\mathbb{R}}\left|2\varphi_K(\lambda)\varphi_T(\lambda)-\varphi_S(\lambda)\right|^2\mathrm{d}\lambda$$
$$= \frac{1}{2\pi}\int_{\mathbb{R}}\left(4\left|\varphi_K(\lambda)\right|^2\mathbf{Var}_S\,\varphi_T(\lambda) + |\varphi_K(\lambda)-1|^2|\varphi_S(\lambda)|^2\right)\mathrm{d}\lambda.$$

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

$$\varphi_K(\lambda) = \varphi_\alpha(\lambda) = \left(1-\left|\frac{\lambda}{\alpha}\right|^{k+\rho}\right)_+, \tag{17}$$

where $\alpha=\alpha_T$ is a positive number tending to infinity and $\rho=\rho_T>0$ tends to zero such that

$$\lim_{T\to\infty}\rho_T\log\alpha_T = \infty.$$

The key result in the proof of this theorem is the following lemma, which describes the asymptotic behavior of the variance of the empirical characteristic function.

Lemma 1. *For any $S\in\Sigma_\delta$, we have*

$$\int_{\mathbb{R}}\left|\varphi_\alpha(\lambda)\right|^2\mathbf{Var}_S\,\varphi_T(\lambda)\,\mathrm{d}\lambda \le \frac{1}{T}\int_{\mathbb{R}}\left|\varphi_\alpha(\lambda)\right|^2\mathrm{d}\lambda\big(1+o_T(1)\big)\ .$$

Proof. By definition of φ_T we have

$$\varphi_T(\lambda) = \frac{1}{T}\int_0^T e^{i\lambda X_t}S(X_t)\,\mathrm{d}t + \frac{1}{T}\int_0^T e^{i\lambda X_t}\,\mathrm{d}W_t.$$

The variance of the last term of the RHS is equal to one over T. Thus, if we show that the quantity

$$Q_T = \int_{\mathbb{R}}\left|\varphi_\alpha(\lambda)\right|^2\mathbf{Var}_S\left[\frac{1}{T}\int_0^T e^{i\lambda X_t}S(X_t)\,\mathrm{d}t\right]\mathrm{d}\lambda$$

converges to zero faster than

$$\frac{1}{T}\int_{\mathbb{R}} \big|\varphi_\alpha(\lambda)\big|^2 \mathrm{d}\lambda = \frac{C\alpha}{T},$$

the lemma will be proven. Let us prove that

$$Q_T = \int_{\mathbb{R}} \big|\varphi_\alpha(\lambda)\big|^2 \mathbf{Var}_S\left[\frac{1}{T}\int_0^T e^{i\lambda X_t} S(X_t)\mathrm{d}t\right]\mathrm{d}\lambda \le \frac{C}{T}.$$

Indeed, using the well known occupation times formula and martingale representation of local time (see Lemma 5), we get

$$\int_0^T e^{i\lambda X_t} S(X_t)\,\mathrm{d}t - T\,\mathbf{E}_S\big[e^{i\lambda\xi}S(\xi)\big] = \hat{H}(\lambda, X_T) - \hat{H}(\lambda, X_0) - \int_0^T \hat{g}(\lambda, X_t)\,\mathrm{d}W_t,$$

where $\hat{H}(\cdot, X_t)$ and $\hat{g}(\cdot, X_t)$ are respectively the Fourier transforms of the functions $S(\cdot)H(\cdot, X_t)$ and $S(\cdot)g(\cdot, X_t)$ defined by (31) and (32). Using once more the Parseval's identity one can show that

$$\begin{aligned} Q_T &\le \int_{\mathbb{R}} \mathbf{Var}_S\left[\frac{1}{T}\int_0^T e^{i\lambda X_t} S(X_t)\mathrm{d}t\right]\mathrm{d}\lambda \\ &\le \frac{8}{T^2}\int_{\mathbb{R}} \mathbf{E}_S\big|\hat{H}(\lambda,\xi)\big|^2\mathrm{d}\lambda + \frac{2}{T}\int_{\mathbb{R}} \mathbf{E}_S\big|\hat{g}(\lambda,\xi)\big|^2\mathrm{d}\lambda \\ &= \frac{8}{T^2}\int_{\mathbb{R}} S^2(x)\,\mathbf{E}_S\big[H^2(x,\xi)\big]\mathrm{d}x + \frac{2}{T}\int_{\mathbb{R}} S^2(x)\,\mathbf{E}_S\big[g^2(x,\xi)\big]\mathrm{d}x. \end{aligned}$$

where the first inequality is due to the fact that $|\varphi_\alpha(\cdot)| \le 1$. The last two integrals are bounded uniformly on $S \in \Sigma_\delta$ because of Lemma 4 and the fact that the trend coefficient S has a polynomial majorant. The proof of Lemma 1 is finished. □

We return to the proof of Theorem 2. By virtue of the last lemma, we have the following upper estimate for the quadratic risk

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

where $o_T(1)$ tends to zero uniformly on $S \in \Sigma_\delta$ 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 set $\Sigma_\delta(k, R)$, the Fourier transform φ_S should belong to the set

$$\Phi = \Big\{\varphi \;\Big|\; \frac{1}{2\pi}\int_{-\infty}^{\infty} \lambda^{2k}|\varphi(\lambda) - \varphi_0(\lambda)|^2\mathrm{d}\lambda \le 4R\Big\},$$

where $\varphi_0 = \varphi_{S_0}$. Replacing in L_T the function φ_α by its explicit expression, we obtain

$$L_T(\varphi_\alpha, \varphi_S) = \frac{2}{\pi T} \int_{-\alpha}^{\alpha} \left(1 - \left|\frac{\lambda}{\alpha}\right|^{k+\rho}\right)^2 \mathrm{d}\lambda + \frac{1}{2\pi} \int_{-\alpha}^{\alpha} \left|\frac{\lambda}{\alpha}\right|^{2k+2\rho} |\varphi_S(\lambda)|^2 \mathrm{d}\lambda + \frac{1}{2\pi} \int_{|\lambda|>\alpha} |\varphi_S(\lambda)|^2 \mathrm{d}\lambda. \tag{18}$$

Since φ_S belongs to the set Φ, and S_0 satisfies condition (6), we have

$$\int_{-\alpha}^{\alpha} \left|\frac{\lambda}{\alpha}\right|^{2k+2\rho} |\varphi_S(\lambda) - \varphi_0(\lambda)|^2 \mathrm{d}\lambda + \int_{|\lambda|>\alpha} |\varphi_S(\lambda) - \varphi_0(\lambda)|^2 \mathrm{d}\lambda \le \frac{8\pi R}{\alpha^{2k}}$$

and, for T large enough,

$$\begin{aligned}\int_{-\alpha}^{\alpha} \left|\frac{\lambda}{\alpha}\right|^{2k+2\rho} |\varphi_0(\lambda)|^2 \mathrm{d}\lambda + \int_{|\lambda|>\alpha} |\varphi_0(\lambda)|^2 \mathrm{d}\lambda &\le \int_{\mathbb{R}} \left|\frac{\lambda}{\alpha}\right|^{2k+2\rho} |\varphi_0(\lambda)|^2 \mathrm{d}\lambda \\ &\le \int_{\mathbb{R}} \frac{|\lambda|^{2k} + |\lambda|^{2k+\tau}}{\alpha^{2k+2\rho}} |\varphi_0(\lambda)|^2 \mathrm{d}\lambda \le \frac{C}{\alpha^{2k+2\rho}} .\end{aligned}$$

Thus the second and the third terms in (18) can be bounded by

$$\frac{4R}{\alpha^{2k}} + \frac{C}{\alpha^{2k+\rho}} = \frac{4R}{\alpha^{2k}} \big(1 + o_T(1)\big),$$

and the first term can be calculated explicitly:

$$\frac{2}{\pi T} \int_{-\alpha}^{\alpha} \left(1 - \left|\frac{\lambda}{\alpha}\right|^{k+\rho}\right)^2 \mathrm{d}\lambda = \frac{8\alpha k^2 \big(1 + o_T(1)\big)}{\pi T (k+1)(2k+1)} .$$

It is clear now that α_T should be of order $T^{1/(2k+1)}$ in order to have the best rate of convergence for the risk of the estimator corresponding to (17). It leads us to the following inequality

$$\begin{aligned}\inf_{\alpha>0} \sup_{S\in\Sigma_\delta} L_T(\varphi_\alpha, \varphi_S) &\le \big(1 + o_T(1)\big) \inf_{\alpha>0} \left[\frac{8k^2\alpha}{\pi T(k+1)(2k+1)} + \frac{4R}{\alpha^{2k}}\right] \\ &= \big(1 + o_T(1)\big) \inf_{\alpha>0} G(\alpha).\end{aligned}$$

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

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

which leads to

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

and

$$\inf_{\alpha>0} G(\alpha) = G(\alpha_0) = \frac{4(2k+1)R}{\alpha_0^{2k}} = 4P(k,R)\, T^{-\frac{2k}{2k+1}} \ .$$

By an elementary verification it can be shown that the function φ_{α_0} is the Fourier transform of the kernel-type function K_T^* defined by (15). We proved so that

$$\sup_{S\in\Sigma_\delta} R_T^L(K_T^*, f_S') \le 4P(k,R) T^{-\frac{2k}{2k+1}} \big(1 + o_T(1)\big).$$

This inequality, associated with the lower bound of Theorem 1, completes the proof of Theorem 2. □

3. Trend Coefficient Estimation

3.1. **Lower Bound.** We establish now a lower bound for the minimax risk in the problem of the trend coefficient estimation. In order to do it, we use the lower bound established in the previous section and some results concerning the behavior of the local time estimator. Let us denote

$$r_T(\Sigma_\delta) = \inf_{\bar S_T} \sup_{S\in\Sigma_\delta} \mathbf{E}_S \int_{\mathbb{R}} \big(\bar S_T(x) - S(x)\big)^2 f_S^2(x)\, \mathrm{d}x.$$

We denote by $f_T^\circ(\cdot)$ the local time estimator of the invariant density

$$f_T^\circ(x) = \frac{1}{T}\int_0^T \operatorname{sgn}(x - X_t)\, \mathrm{d}X_t + \frac{|X_T - x| - |X_0 - x|}{T} \ . \tag{19}$$

Using the occupation times formula one can check that f_T° is the derivative (in the sense of distributions) of the empirical distribution function. We will need in the sequel some exponential inequalities concerning the risk of the estimator $f_T^\circ(x)$ at the point x in order to ensure the integrability of some functions. We state now a result providing such an estimate.

Lemma 2. *For any $p \ge 1$ there exist two constants C and γ such that*

$$\mathbf{E}_S\big[f_T^\circ(x) - f_S(x)\big]^{2p} \le C T^{-p} e^{-\gamma|x|}$$

for any $x \in \mathbb{R}$ and $S \in \Sigma_\delta$.

Proof. The proof is an obvious consequence of Lemmas 4 and 5 stated in the Appendix. □

Theorem 3. *Let the central function* $S_0(\cdot)$ *satisfy the conditions* (4)–(6), *then*

$$\underline{\lim}_{\delta\to 0}\ \underline{\lim}_{T\to\infty}\ T^{\frac{2k}{2k+1}} r_T(\Sigma_\delta) \geq P(k,R),$$

where the Pinsker's constant $P(k,R)$ *is defined in the previous section.*

Proof. If we denote $\bar\vartheta_T(\cdot) = 2\bar S_T(\cdot) f_T^\circ(\cdot)$, then using triangular inequality we obtain

$$\mathbf{E}_S \int_{\mathbb{R}} \big(\bar S_T(x) - S(x)\big)^2 f_S^2(x)\,\mathrm{d}x \geq (\sqrt{A_2} - \sqrt{A_1})^2,$$

where we used the following notations

$$A_1 = \mathbf{E}_S \int_{\mathbb{R}} \bar S_T^2(x)\big(f_S(x) - f_T^\circ(x)\big)^2\,\mathrm{d}x,$$
$$A_2 = \frac{1}{4}\,\mathbf{E}_S \int_{\mathbb{R}} \big(\bar\vartheta_T(x) - f_S'(x)\big)^2\,\mathrm{d}x.$$

Since $\bar\vartheta_T(\cdot)$ is an estimator of $f_S'(\cdot)$, the term A_2 can be evaluated using the results of previous section. In order to evaluate the first term, note that we can consider only those estimators $\bar S_T(\cdot)$ which satisfy the condition

$$\big|\bar S_T(x)\big| \leq b_T\, e^{x/\log b_T}. \tag{20}$$

where $b_T = \log(T+1)$. Indeed, for T large enough, we have

$$\sup_{S\in V_\delta} \big|S(x)\big| \leq C(1+|x|^\nu) \leq b_T\, e^{x/\log b_T}.$$

Consequently the risk of the estimator $\tilde S_T(x) = \min\big(\bar S_T(x), b_T\, e^{x/\log b_T}\big)$ is less than the risk of $\bar S_T(\cdot)$. This means that in the proof of the lower bound we can consider only the estimators satisfying the condition (20).

The inequality (20) associated with Lemma 2 gives the estimate

$$A_1 \leq C b_T^2 T^{-1}$$

for any $S \in \Sigma_\delta$ and T large enough. Then, using the result of the previous section, we have

$$\begin{aligned}
\underline{\lim}_{\delta\to 0}\ \underline{\lim}_{T\to\infty}\ T^{\frac{k}{2k+1}} \sqrt{r_T(\Sigma_\delta)} &\geq \underline{\lim}_{\delta\to 0}\ \underline{\lim}_{T\to\infty}\ \inf_{\bar S_T}\ \sup_{S\in\Sigma_\delta} T^{\frac{k}{2k+1}}\big(\sqrt{A_2} - \sqrt{A_1}\big)\\
&\geq \underline{\lim}_{\delta\to 0}\ \underline{\lim}_{T\to\infty}\ \inf_{\bar S_T}\ \sup_{S\in\Sigma_\delta} T^{\frac{k}{2k+1}}\big(\sqrt{A_2} - C b_T T^{-1/2}\big)\\
&\geq \sqrt{P(k,R)}.
\end{aligned}$$

This completes the proof of the theorem. □

3.2. **Asymptotically Efficient Estimator.** We now come to the construction of an estimator of the trend coefficient which is asymptotically efficient in the sense of the lower bound obtained in Theorem 3.

Definition 2. An estimator $\tilde{S}_T(\cdot)$ is said to be asymptotically efficient in the problem of trend coefficient estimation, if

$$\lim_{\delta\to 0}\lim_{T\to\infty}\sup_{S\in\Sigma_\delta} T^{\frac{2k}{2k+1}}\,\mathbf{E}_S\left[\int_{\mathbb{R}}\big(\tilde{S}_T(x)-S(x)\big)^2 f_S^2(x)\,\mathrm{d}x\right]=P(k,R). \tag{21}$$

Remind that

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

We have already found an efficient estimator $\vartheta_{K^*,T}(\cdot)$ of $f_S'(\cdot)$. From now on we will write $\vartheta_T^*(\cdot)$ instead of $\vartheta_{K^*,T}(\cdot)$. We know also (see [16]) that the invariant density $f_S(\cdot)$ is "well" estimated by the local time estimator $f_T^\circ(\cdot)$ defined by (19). It is proved in [16] that this estimator is asymptotically efficient and converges to the invariant density $f_S(\cdot)$ with the rate $\sqrt{T}$. Thus, it is very natural to estimate $S(x)$ by

$$\bar{S}_T(x)=\frac{\vartheta_T^*(x)}{2f_T^\circ(x)}\ .$$

The great disadvantage of this estimator is that the denominator part is equal to zero if there is no observation at the point x. Therefore this estimator is equal to infinity with a strictly positive probability. That is why the corresponding risk is always equal to infinity.

Fortunately we can get round this difficulty using the following idea. The main property of the estimator f_T° that we need is the convergence to f_S with a rate faster than $T^{k/(2k+1)}$. So we can replace f_T° by another estimator which converges slower than $\sqrt{T}$ but faster than $T^{k/(2k+1)}$, and which is bounded away from 0.

These heuristics lead us to investigate the behavior of the estimator

$$\hat{S}_T(x)=\frac{\vartheta_T^*(x)}{2f_T^\circ(x)+\varepsilon_T e^{-l_T|x|}} \tag{22}$$

where $\vartheta_T^*(\cdot)$ is the asymptotically efficient estimator of the derivative of the invariant density studied in the previous section, $f_T^\circ(\cdot)$ is the local time estimator, $\varepsilon_T=T^{-(1-\kappa)/2}$ and $l_T=\big[\log(T+1)\big]^{-1}$. The positive constant κ is chosen to be strictly smaller than $\beta=1/(2k+1)$.

To prove the asymptotic efficiency we need also the following estimate.

Lemma 3. *There exists a constant C such that $\mathbf{E}_S\big[{\vartheta^*_T}^4(x)\big] \le CT^{4\beta}$ for any $x \in \mathbb{R}$ and $S \in \Sigma_\delta$.*

Proof. Remark firstly that using $(a+b)^4 \le 8a^4 + 8b^4$ and Hölder's inequalities, one can show that

$$\begin{aligned}\mathbf{E}_S\big[{\vartheta^*_T}^4(x)\big] &\le \frac{C}{T^4}\,\mathbf{E}_S\bigg[\int_0^T K^*_T(x-X_t)S(X_t)\mathrm{d}t\bigg]^4 \\ &\qquad\qquad + \frac{C}{T^4}\,\mathbf{E}_S\bigg[\int_0^T K^*_T(x-X_t)\,\mathrm{d}W_t\bigg]^4 \\ &\le C\mathbf{E}_S\big[K^{*4}_T(x-\xi)S^4(\xi)\big] + C\mathbf{E}_S\big[K^{*4}_T(x-\xi)\big].\end{aligned}$$

It follows from (15) that the supremum norm of the optimal kernel K^*_T is bounded by CT^β, that is

$$\sup_{x\in\mathbb{R}}\big|K^*_T(x)\big| \le CT^\beta.$$

Since S has a polynomial majorant, its moments are uniformly bounded on Σ_δ. These estimates imply the inequality of Lemma 3. □

Theorem 4. *If the conditions* (4)–(6) *are satisfied, then the estimator $\hat S_T$ is asymptotically efficient in the problem of trend coefficient estimation.*

Proof. Let us denote $\bar f_T(x) = f^\circ_T(x) + \varepsilon_T e^{-l_T|x|}$ and

$$\mathbb{B}_T(x) = \Big\{\ \omega\ \Big|\ \ f_S(x) - f^\circ_T(x) < \varepsilon_T\ e^{-l_T|x|}\Big\}.$$

Remark that according to Lemma 2 and the inequality of Chebyshev, we have

$$\mathbf{P}_S\big[\mathbb{B}^c_T(x)\big] \le CT^{-\kappa p}e^{-\gamma_*|x|},$$

where p can be chosen as large as we want and $\gamma_* < \gamma$ (see (28)). Using Cauchy-Schwarz inequality and Lemma 3, together with estimate (30), one can show that

$$\mathbf{E}_S\Big[\big(\hat S_T(x) - S(x)\big)^2 f^2_S(x)\chi_{\mathbb{B}^c_T(x)}\Big] \le \frac{C\varepsilon_T^{-2}\,e^{-2\gamma|x|}}{T^{\kappa p - 2\beta}} \le CT^{2-\kappa p}e^{-2\gamma|x|}.$$

We choose p such that $\kappa p > 3$, then we have

$$\int_{\mathbb{R}}\mathbf{E}_S\Big[\big(\hat S_T(x) - S(x)\big)^2 f^2_S(x)\chi_{\mathbb{B}^c_T(x)}\Big]\mathrm{d}x \le CT^{-1}.$$

To evaluate the risk over the set $\mathbb{B}_T(x)$, we use the triangular inequality

$$\int_{\mathbb{R}}\mathbf{E}_S\Big[\big(\hat S_T(x) - S(x)\big)^2 f^2_S(x)\chi_{\mathbb{B}_T(x)}\Big]\mathrm{d}x \le (\sqrt{A}_1 + \sqrt{A}_2)^2,$$

where

$$A_1 = \int_{\mathbb{R}} \mathbf{E}_S\Big[\big(\hat{S}_T(x) - S(x) f_S(x)/\bar{f}_T(x)\big)^2 f_S^2(x)\chi_{\mathbb{B}_T(x)}\Big]\mathrm{d}x,$$

$$A_2 = \int_{\mathbb{R}} \mathbf{E}_S\Big[\big(S(x) f_S(x)/\bar{f}_T(x) - S(x)\big)^2 f_S^2(x)\chi_{\mathbb{B}_T(x)}\Big]\mathrm{d}x.$$

Since $f_S(x) < \bar{f}_T(x)$ for any $\omega \in \mathbb{B}_T(x)$, and $2S(x)f_S(x) = f_S'(x)$, we have the following obvious inequalities

$$4A_1 \le \int_{\mathbb{R}} \mathbf{E}_S\big(\vartheta_T^*(x) - f_S'(x)\big)^2\mathrm{d}x,$$

$$4A_2 \le \int_{\mathbb{R}} S^2(x)\, \mathbf{E}_S\big(\bar{f}_T(x) - f_S(x)\big)^2\mathrm{d}x.$$

The condition (5) and Lemma 2 imply that the term A_2 is of order $\varepsilon_T^2 l_T^{-2\nu}$, which is smaller than $T^{-2k/(2k+1)}$ since $\kappa < 1/(2k+1)$. Finally, using the result of Theorem 2, we obtain (21). □

4. Concluding Remarks

4.1. **Weighted Sobolev balls.** Note that the parameter space Σ_δ we considered up to now is defined via the k^{th}-derivative of the difference $f_S' - f_0'$. This is a quite natural and classic definition in the case where the parameter of interest is the derivative function $f_S'(\cdot)$. But we used the same parameter space in the problem of estimating the trend coefficient as well. We show below that the results of the Section 3 hold true if we replace the parameter space $\Sigma_\delta(k, R, S_0)$ by a weighted Sobolev ball $\widetilde{\Sigma}_\delta(k, R, S_0)$, defined via k^{th}-derivative of the difference $S - S_0$ and the weight function f_S^2. This new setup seems to be more natural for the problem of trend estimation.

We give now the exact definitions and statement of the results. Let $\widetilde{V}_\delta(S_0)$ be the set of all k-times continuously differentiable functions $S(\cdot)$ satisfying the condition

$$\sup_{x\in\mathbb{R}} \big|S^{(i)}(x) - S_0^{(i)}(x)\big| \le \delta, \qquad i = 0, 1, \ldots, k-1. \tag{23}$$

Then we define the weighted Sobolev space by

$$\widetilde{\Sigma}_\delta(k, R, S_0) = \Big\{S \in \widetilde{V}_\delta(S_0) \,\Big|\, \int_{\mathbb{R}} \big[\big(S - S_0\big)^{(k)}(x)\big]^2 f_S^2(x)\mathrm{d}x \le R\Big\}.$$

In order to prove that the minimax risk over this space $\widetilde{\Sigma}_\delta$ has the same asymptotics as the minimax risk over Σ_δ, we need the following condition:

$$|S_0^{(k)}(x)| \leq C(1+|x|^\nu), \qquad \forall x \in \mathbb{R}, \tag{24}$$

for some positive constants C and ν.

Theorem 5. *Let conditions* (4), (6) *and* (24) *be fulfilled, then*

$$\lim_{\delta\to 0}\lim_{T\to\infty} r_T(\widetilde{\Sigma}_\delta) T^{2k/(2k+1)} = P(k,R),$$

and the estimator $\hat{S}_T(\cdot)$ is asymptotically efficient.

Proof. The lower bound part of this theorem can be derived from the proof of Theorem 1 using the obvious fact that $\mathcal{C}_T \subseteq \widetilde{\Sigma}_\delta$.

The efficiency of the estimator $\hat{S}_T(\cdot)$ follows from the inclusion

$$\widetilde{\Sigma}_\delta(k,R,S_0) \subseteq \Sigma_\delta(k,R+o_\delta(1),S_0)$$

where the term $o_\delta(1)$ tends to zero when $\delta \to 0$. Let us prove now this inclusion. In the sequel $\|h\|_2$ will denote the L^2-norm of a function $h:\mathbb{R}\to\mathbb{R}$.

Assume that the function S belongs to the set $\tilde{\Sigma}_\delta(k,R,S_0)$ and let $P(x_0,\ldots,x_{k-1})$ be a polynomial such that

$$f_S^{(k+1)} = \big(2S^{(k)} + P(S,S',\ldots,S^{(k-1)})\big)\, f_S.$$

Then, by virtue of conditions (23) and (24), we have

$$\begin{aligned}\big\|f_S^{(k+1)}(x) - f_0^{(k+1)}\big\|_2 &\leq 2\big\|S^{(k)} f_S - S_0^{(k)} f_0\big\|_2 \\ &\quad + \big\|P(S,\ldots,S^{(k-1)})f_S(x) - P(S_0,\ldots,S_0^{(k-1)})f_0\big\|_2 \\ &\leq 2\big\|(S^{(k)}-S_0^{(k)})f_S\big\|_2 + 2\big\|S_0^{(k)}(f_S-f_0)\big\|_2 \\ &\quad + o_\delta(1) + C\left[\int_{\mathbb{R}} (1+|x|^{\nu_1})^2 \big(f_S(x)-f_0(x)\big)^2 \mathrm{d}x\right]^{1/2}.\end{aligned} \tag{25}$$

It remains to verify that for any positive number n, we have

$$\lim_{\delta\to 0}\sup_{S\in\tilde{\Sigma}_\delta}\int_{\mathbb{R}} x^{2n}\big(f_S(x)-f_0(x)\big)^2\mathrm{d}x = 0. \tag{26}$$

Using the evident inequality $(a+b)^2 \leq 2a^2 + 2b^2$, we get

$$\begin{aligned}\int_{\mathbb{R}} x^{2n}\big(f_S(x)-f_0(x)\big)^2\mathrm{d}x &\leq 2\int_{\mathbb{R}} x^{2n}\left(\frac{1}{G(S)} - \frac{1}{G(S_0)}\right)^2 e^{4\int_0^x S_0(v)\mathrm{d}v}\mathrm{d}x \\ &\quad + \int_{\mathbb{R}} \frac{2x^{2n}}{G^2(S)}\left(e^{4\int_0^x S(v)\mathrm{d}v} - e^{4\int_0^x S_0(v)\mathrm{d}v}\right)^2\mathrm{d}x.\end{aligned}$$

Since the function $S(\cdot)$ belongs to the δ-neighborhood of $S_0(\cdot)$ in uniform metrics and satisfies the condition (4), the difference $G(S)-G(S_0)$

tends to zero as $\delta \to 0$. Therefore the first integral of the last inequality tends to zero. To estimate the second integral, one can remark that due to condition (4), there exist two positive constants b and γ such that

$$e^{4\int_0^x S(v)\mathrm{d}v} - e^{4\int_0^x S_0(v)\mathrm{d}v} \leq \chi_{\{|x|<2b\}} C(e^{4\delta b} - 1) + \chi_{\{|x|\geq 2b\}} C e^{-2\gamma|x|}.$$

This inequality implies immediately (26). So the second and the forth terms in (25) are $o_\delta(1)$, and the first term is less than $\sqrt{R}$ since $S \in \widetilde{\Sigma}_\delta$. Consequently

$$\big\| f_S^{(k+1)} - f_0^{(k+1)} \big\|_2 \leq 2\sqrt{R} + o_\delta(1).$$

This completes the proof. □

4.2. **Analogy with Gaussian white noise model.** As we said in the Introduction, there is no rigorous result providing a link between the model of ergodic diffusion and the other models of nonparametric statistics. But the results obtained in the previous sections reveal an evident similarity with the corresponding results in the Gaussian white noise model (see [27]). Therefore, for better understanding of the model of ergodic diffusion, it is helpful to write down the Gaussian white noise experiment having the same asymptotic properties.

Let $\{B_u,\, u \in \mathbb{R}\}$ be a two sided standard Wiener process. Some heuristic considerations suggest to consider the model

$$\mathrm{d}Y_u = S(u) f_S(u)\, \mathrm{d}u + \sqrt{\frac{f_0(u)}{T}}\, \mathrm{d}B_u, \quad u \in \mathbb{R}, \tag{27}$$

which seems to be the best candidate for the role of a locally asymptotically equivalent experiment for ergodic diffusion model (1). In order to explain these heuristic arguments, let us denote $h_0(u) = S_0(u) f_0(u)$ and let $v(\cdot)$ be a continuous function with support in $[-1, 1]$. Set

$$h(z, u) = h_0(u) + T^{-k\beta} v\big(T^\beta(u - z)\big),$$

with $\beta = 1/(2k + 1)$ and suppose that the function $v(\cdot)$ is such that $h(z, u) = S(u) f_S(u)$ for a $S \in \Sigma_\delta$. In this case, the log-likelihood corresponding to the ergodic diffusion model has essentially the following form (see, for example, [20] Theorem 7.7):

$$\log\left[\frac{\mathrm{d}\mathbf{P}_S}{\mathrm{d}\mathbf{P}_{S_0}}\,(X^T)\right] = \int_0^T \big(S(X_t) - S_0(X_t)\big)\, \mathrm{d}W_t \\ -\frac{1}{2}\int_0^T \big(S(X_t) - S_0(X_t)\big)^2 \mathrm{d}t.$$

Taking into account the relation $h(z,u) = S(u)f_S(u)$, we obtain the following obvious equality:

$$S(u) - S_0(u) = \frac{h(z,u) - h_0(u)}{f_0(u)} + \frac{2h(z,u)\big(f_0(u) - f_S(u)\big)}{f_S(u)f_0(u)} .$$

Since $f_S(\cdot)$ is the integral of $2h(z,\cdot)$, the difference $f_S(u) - f_0(u)$ tends to zero faster than $h(z,u) - h_0(u)$, as $T \to \infty$. This implies that

$$\int_0^T \big(S(X_t) - S_0(X_t)\big)^2 \mathrm{d}t \sim \int_0^T \frac{T^{-2k\beta} v^2\big(T^\beta(X_t - z)\big)}{f_0^2(X_t)}\, \mathrm{d}t.$$

Using the low of large numbers we obtain:

$$\begin{aligned}\int_0^T \big(S(X_t) - S_0(X_t)\big)^2 \mathrm{d}t &\sim T^\beta \int_{\mathbb{R}} \frac{v^2\big(T^\beta(x - z)\big)}{f_0^2(x)}\, f_0(x)\,\mathrm{d}x \\ &= \int_{-1}^1 \frac{v^2(w)\,\mathrm{d}w}{f_0(z + wT^{-\beta})} \sim \frac{\|v\|_2^2}{f_0(z)}.\end{aligned}$$

Roughly speaking, this relation means that the Fisher information (see [13] for precise definition) in the model of ergodic diffusion is $I(S,z) = 1/f_S(z)$, if the law of X^T is parameterized by the function $h(x) = S(x)f_S(x)$. At the same time, the small parameter is $1/\sqrt{T}$. That is why we guess that the models (27) and (1) are asymptotically equivalent.

It is worthy to note that the integral of $I^{-1}(S,\cdot)$ over $\mathbb{R}$ is equal to one, which explains the fact that the optimal constant obtained in Section 2 does not depend on the central function S_0 (see [13] for the expression of the optimal constant in general case).

The Gaussian white noise experiment described by (27) has heteroscedastic form, since the noise component contains an unknown function. Heuristically, the homoscedastic form of the equivalent experiment should be

$$\mathrm{d}Y_u = S(u)\sqrt{f_S(u)}\,\mathrm{d}u + \frac{1}{\sqrt{T}}\,\mathrm{d}B_u, \quad u \in \mathbb{R}.$$

One can deduce from this representation that the Fisher information in the problem of trend estimation is equal to $f_S(z)$. This fact can be also verified directly, like we have done it above for $h = Sf_S$.

4.3. **Kernel-type estimators of the invariant density.** Let $K(\cdot)$ be a symmetric positive function of real argument supported by the interval $[-1,1]$. It is well known (see [2], [16]) that the kernel-type estimator

$$f_{K,T}(x) = \frac{1}{h_T T} \int_0^T K\left(\frac{x - X_t}{h_T}\right) \mathrm{d}t$$

converges with the rate $\sqrt{T}$ to the invariant density if the kernel $K(\cdot)$ is continuous,

$$\int_{-1}^{1} K(u)\,\mathrm{d}u = 1$$

and h_T tends to zero as $T \to \infty$.

Using the occupation times formula and the inequality of Hölder, we obtain the following sequence of inequalities:

$$\begin{aligned}
\mathbf{E}_S \big[f_{K,T}(x) - f_S(x)\big]^{2p} & \\
&= \mathbf{E}_S \left[\frac{1}{h_T} \int_{\mathbb{R}} K\left(\frac{x-y}{h_T}\right) \big(f_T^\circ(y) - f_S(x)\big)\,\mathrm{d}x\right]^{2p} \\
&= \mathbf{E}_S \left[\int_{-1}^{1} K(u) \big(f_T^\circ(x + uh_T) - f_S(x)\big)\,\mathrm{d}x\right]^{2p} \\
&\leq C\mathbf{E}_S \int_{-1}^{1} \big[f_T^\circ(x + uh_T) - f_S(x)\big]^{2p}\,\mathrm{d}u.
\end{aligned}$$

It is evident now that for the quadratic risk of $f_{K,T}$ we have an upper estimate analogous to the one of Lemma 2. This implies that in the definition (22) of asymptotically efficient estimator one can replace the local time estimator by $f_{K,T}(\cdot)$. The estimator of the trend coefficient obtained in this way is still asymptotically efficient and its calculation can be easier in some cases.

4.4. **Trend estimation via local polynomial smoothers.** There exists a number of other methods of trend coefficient estimation. They use in general various extensions of the maximum likelihood method in nonparametric setup. One of these methods is based on the local approximation of the log-likelihood by polynomials (see [28]).

The precise formulation of the method is the following. Let $K_1(\cdot)$ and $K_2(\cdot)$ be two kernel functions and define

$$s^* = \arg\max_{s \in \mathbb{R}^n} l_T(s, x, X^T),$$

where l_T is a local approximation of the log-likelihood by a polynomial of order $n-1$, that is

$$l_T(s,x,X^T) = 2\int_0^T \left[\sum_{r=0}^{n-1} s_r(x-X_t)^r\right] K_1\left(\frac{x-X_t}{h_T}\right) \mathrm{d}X_t - \int_0^T \left[\sum_{r=0}^{n-1} s_r(x-X_t)^r\right]^2 K_2\left(\frac{x-X_t}{h_T}\right) \mathrm{d}t,$$

where (the bandwidth) h_T is a positive number tending to zero as $T \to \infty$. The estimator of the function $S(\cdot)$ at the point x is then $S_T^*(x) = s_0^*$ defined as the first coordinate of s^*.

In order to avoid very complicated expressions, for any $r \in \mathbb{N}$ we denote

$$\mu_r(T) = \frac{1}{T}\int_0^T \left(\frac{x-X_t}{h_T}\right)^r K_2\left(\frac{x-X_t}{h_T}\right) \mathrm{d}t,$$

$$\nu_r(T) = \frac{1}{T}\int_0^T \left(\frac{x-X_t}{h_T}\right)^r K_1\left(\frac{x-X_t}{h_T}\right) \mathrm{d}X_t.$$

Then it can be easily shown that for $n = 1$ this method gives the estimator

$$S_T^*(x) = \frac{\nu_0(T)}{\mu_0(T)} ,$$

which is essentially the same as the estimator defined in previous remark. As for $n = 2$, the estimator is

$$\tilde{S}_T^*(x) = \frac{\nu_0(T)\mu_2(T) - \nu_1(T)\mu_1(T)}{\mu_0(T)\mu_2(T) - \mu_1(T)^2}.$$

Each one of the terms $\mu_r(T)$ is asymptotically of the order h_T, but if the kernel K_2 is symmetric, then the term $\mu_1(T)$ is of order h_T^2. This means that the estimator S_T^* has asymptotically the same properties as $\tilde{S}_T^*$. In conclusion, we can note that the estimator $\hat{S}_T$ considered in subsection 3.2 can be derived from the method of local polynomial smoothing as well.

5. Appendix

In this section we state some technical lemmas from [18] and since this book is not yet published we present a sketch of the proofs. We suppose that the function $S_0(\cdot)$ satisfies the conditions (2) and (5). To formulate the first lemma, we need the following notations:

$$\Phi_S(x) = f_S^{2p}(x)\,\mathbf{E}_S\left[\frac{\chi_{\{\xi>x\}} - F_S(\xi)}{f_S(\xi)}\right]^{2p},$$
$$\Psi_S(x) = f_S^{2p}(x)\,\mathbf{E}_S\left[\int_0^\xi \frac{\chi_{\{y>x\}} - F_S(y)}{f_S(y)}\,\mathrm{d}y\right]^{2p},$$

where p is a real number greater than 1.

Lemma 4. *There exist two positive constants C and γ (depending only on p and S_0) such that, for δ small enough, the following inequalities are fulfilled:*

$$\Phi_S(x) \le Ce^{-\gamma|x|}, \quad \Psi_S(x) \le Ce^{-\gamma|x|}, \tag{28}$$

for any $S \in V_\delta(S_0)$.

Proof. We prove only the first inequality, the proof of the second one is similar. Note that since the function S_0 satisfies condition (2), one can find two positive constants b and γ such that

$$S_0(x)\,\mathrm{sgn}\,x < -2\gamma \tag{29}$$

when $|x| > b$. Therefore, on the one hand, for $\delta < \gamma$, we have

$$S(x)\,\mathrm{sgn}\,x < -\gamma$$

for any $x \notin [-b, b]$ and $S \in V_\delta(S_0)$. On the other hand, since S_0 is continuous and $S \in V_\delta(S_0)$, there is a constant D such that

$$\sup_{S\in V_\delta(S_0)}\ \sup_{|x|\le b} \big|S(x)\big| < D.$$

Thus, for any $S \in V_\delta(S_0)$, the normalizing constant $G(S)$ can be bounded like follows

$$\begin{aligned} G(S) &= \int_{\mathbb{R}} \exp\left\{2\int_0^x S(v)\,\mathrm{d}v\right\}\mathrm{d}x \\ &= \int_{|x|\le b} \exp\left\{2\int_0^x S(v)\,\mathrm{d}v\right\}\mathrm{d}x + \int_{|x|>b} \exp\left\{2\int_0^x S(v)\,\mathrm{d}v\right\}\mathrm{d}x \\ &\le \int_{|x|\le b} e^{2D|x|}\mathrm{d}x + \int_{|x|>b} e^{2Db-2\gamma(|x|-b)}\mathrm{d}x < \infty. \end{aligned}$$

In the same way, one can show that

$$\inf_{S\in V_\delta(S_0)} G(S) \ge \int_{|x|\le b} \exp\big\{-2D|x|\big\}\mathrm{d}x > 0.$$

These two estimates and the explicit form of the density $f_S(x)$ imply that

$$m_b = \inf_{S\in V_\delta} \inf_{|x|\le b} f_S(x) > 0 \quad \text{and} \quad \sup_{S\in V_\delta} f_S(x) \le C e^{-\gamma|x|}. \tag{30}$$

So if we show that the quantity $\Phi_S(x)/f_S(x)$ is bounded for any $S \in V_\delta(S_0)$ and $x \in \mathbb{R}$, then the first inequality of (28) will be proved.

Using inequalities (30), for $x \in [-b, b]$, we have

$$\begin{aligned}\frac{\Phi_S(x)}{f_S(x)} &= \int_{\mathbb{R}} \left(\frac{f_S(x)}{f_S(y)}\right)^{2p-1} \left(\chi_{\{y>x\}} - F_S(y)\right)^{2p} \mathrm{d}y \\ &\le C + \int_b^\infty \left(\frac{f_S(x)}{f_S(y)}\right)^{2p-1} \left(1 - F_S(y)\right)^{2p} \mathrm{d}y \\ &\quad + \int_{-\infty}^{-b} \left(\frac{f_S(x)}{f_S(y)}\right)^{2p-1} F_S^{2p}(y)\,\mathrm{d}y.\end{aligned}$$

Note that for any $y > b$ the following estimate is true

$$\frac{1 - F_S(y)}{f_S(y)} = \int_y^\infty \exp\left\{2\int_y^z S(v)\,\mathrm{d}v\right\}\mathrm{d}z \le \int_y^\infty e^{-2\gamma(z-y)}\,\mathrm{d}z = \frac{1}{2\gamma}\ .$$

It is evident that the same estimate is true for $F_S(y)/f_S(y)$ when y is less than $-b$. Consequently

$$\frac{\Phi_S(x)}{f_S(x)} \le C + \frac{C}{(2\gamma)^{2p}} \int_{|y|>b} f_S(y)\,\mathrm{d}y.$$

Since the last integral is less than 1 for any S, the inequality (28) is proved for $x \in [-b, b]$.
For x less than $-b$, we can estimate the integrals over $]-\infty, x[$, $[-b, b]$ and $]b, +\infty[$ as above. Concerning the integral over the interval $[x, -b]$, it is bounded by

$$\int_x^{-b} \left(\frac{f_S(x)}{f_S(y)}\right)^{2p-1} \mathrm{d}y \le \int_x^{-b} e^{-2(2p-1)\gamma(y-x)}\,\mathrm{d}y < \frac{1}{2(2p-1)\gamma}\ .$$

The case $x > b$ can be proved in the same way. □

Now we state a result concerning the local time estimator. This result is proved in [16] and gives a representation of the local time estimator via a local martingale.

Lemma 5. *Let X_t be an ergodic diffusion with invariant density $f_S(x)$ and $f_T^\circ(x)$ be the local time estimator defined by* (19). *Then*

$$f_T^\circ(x) - f_S(x) = \frac{H(x, X_T) - H(x, X_0)}{T} - \frac{1}{T}\int_0^T g(x, X_t)\,\mathrm{d}W_t,$$

for any $x \in \mathbb{R}$, where

$$H(x,u) = 2f_S(x)\int_0^u \left(\frac{\chi_{\{y>x\}} - F_S(y)}{f_S(y)}\right) \mathrm{d}y, \tag{31}$$

$$g(x,u) = 2f_S(x)\left(\frac{\chi_{\{u>x\}} - F_S(u)}{f_S(u)}\right). \tag{32}$$

Acknowledgments. The authors are grateful to the anonymous referees for careful reading of this paper and valuable comments.

References

[1] G. Banon, *Nonparametric identification for diffusion processes*, SIAM J. Control and Optim., 16 (1978), 3, pp. 380-395.

[2] J. V. Castellana and M. R. Leadbetter, *On smoothed density estimation for stationary processes*, Stoch. Proc. Appl., 21 (1986), pp. 179-193.

[3] A. S. Dalalyan, *Estimation non paramétrique asymptotiquement efficace pour des processus de diffusion ergodiques*, Thèse de doctorat de l'Université du Maine (2001). `http://www.univ-lemans.fr/~s972601/links/publis.html`

[4] A. S. Dalalyan and Yu. A. Kutoyants, *Asymptotically efficient estimation of the derivative of the invariant density*, Prepublication 00-4 (2000), Université du Maine, Le Mans, submitted.

[5] S. Delattre, and M. Hoffmann, *Asymptotic equivalence for a null recurent diffusion*, Prepublication n° 567 (2000), Universités Paris 6 et 7.

[6] R. Durrett, *Stochastic Calculus: a practical introduction.* CRC Press, Boca Raton, 1996.

[7] S. Yu. Efroimovich and M. S. Pinsker, *Estimating a square integrable probability density of a random variable.* Problems Inform. Transmission, 18 (1982), 3, pp. 19–38.

[8] L. Galtchouk and S. Pergamenshchikov, *Sequential nonparametric adaptive estimation of the drift coefficient in diffusion processes*, Math. Methods Statist., 10 (2001), 3, pp. 316–330.

[9] V. Genon-Catalot, C. Laredo and M. Nussbaum, *A diffusion representation of the nonparametric IID experiment on an interval*, WIAS preprint 413 (1998).

[10] V. Genon-Catalot, C. Laredo and M. Nussbaum, *Asymptotic equivalence of estimating a Poisson intensity and a positive diffusion drift*, Ann. Statistics, 30 (2002), pp. 731-753.

[11] I. I. Gikhman and A. V. Skorokhod, *Introduction to Theory of Random Processes.* W.B. Saunders, Philadelphia, 1969.

[12] R. D. Gill and B. Ya. Levit, *Application of the van Trees inequality: a Bayesian Cramer-Rao bound*, Bernoulli, 1 (1995), pp. 59–79.

[13] G. K. Golubev, *LAN in problems of non-parametric estimation of functions and lower bounds for quadratic risks*, Theory Probab. Appl., 36 (1991), 1, 152–157.

[14] G. K. Golubev, *Non-parametric estimation of smooth densities in L^2* (in Russian), Problems Inform. Transmission, 28 (1992), 1, pp. 52–62.

[15] G. K. Golubev and B. Ya. Levit, *On the second order minimax estimation of distribution functions*, Math. Methods Statist., 5 (1996), 1, pp. 1–31.

[16] Yu. A. Kutoyants, *Efficient Density Estimation for Ergodic Diffusion*, Statistical Inference for Stochastic Processes, 1 (1998), 2, pp. 131–155.

[17] Yu. A. Kutoyants, *Statistical Inference for Spatial Poisson Processes,* Lecture Notes in Statistics 134, Springer-Verlag, New York, 1998.

[18] Yu. A. Kutoyants, *Statistical Inference for Ergodic Diffusion Processes,* in preparation.

[19] F. Leblanc, *Density estimation for a class of continuous time process.* Math. Methods of Statist., 6 (1997), 2, pp. 171–199.

[20] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes.* I,II, Springer, New York, 1977.

[21] G. Milstein and M. Nussbaum, *Diffusion approximation for nonparametric autoregression*, Probab. Theory Relat. Fields, 112 (1998), pp. 535–543.

[22] H. T. Nguyen, *Density estimation in a continuous-time Markov processes*, Ann. Statist., 7 (1979), pp. 341–348.

[23] M. Nussbaum, *Spline smoothing in regression models and asymptotic efficiency in* L_2, Ann. Statist., 13 (1985), 3, pp. 984–998.

[24] M. Nussbaum, *Minimax risk: Pinsker's bound*, In Encyclopedia of Statistical Sciences, Update Volume 3, 451-460 (S.Kotz, Ed.), Wiley, New York, 1999.

[25] T. D. Pham, *Nonparametric estimation of the drift coefficient in the diffusion equation*, Math. Operationsforsch. Statist., Ser. Statistics, 1 (1981), pp. 61–73.

[26] B. L. S. Prakasa Rao, *Nonparametric density estimation for stochastic processes from sampled data*, Publ. ISUP, XXXV (1990), pp. 51–84.

[27] M. S. Pinsker, *Optimal filtering of square integrable signals in Gaussian white noise.* Problems Inform. Transmission, 16 (1980), pp. 120–133.

[28] V. G. Spokoiny, *Adaptive drift estimation for nonparametric diffusion model*, Ann. Statistics, 2 (2000), pp. 815–836.

[29] M. Schipper, *Optimal rates and constants in* L^2*-minimax estimation of probability density functions*, Math. Methods Statist, 5 (1996), 3, pp. 253–274.

[30] J. H. Van Zanten, *Rates of convergence and asymptotic normality of kernel estimators for ergodic diffusion processes*, Journal of Nonparametric Statistics, 13 (2001), pp. 833-850.

Laboratoire de Statistique et Processus, Université du Maine, Avenue O. Messiaen, 72085 Le Mans, Cedex 9, France

E-mail address: arnak.dalalyan@univ-lemans.fr

E-mail address: kutoyants@univ-lemans.fr