On second order minimax estimation of invariant density for ergodic diffusion

Arnak S. Dalalyan, Yury A. Kutoyants

Received: March 17, 2003; Accepted: March 3, 2004

Summary: There are many asymptotically first order efficient estimators in the problem of estimating the invariant density of an ergodic diffusion process nonparametrically. To distinguish between them, we consider the problem of asymptotically second order minimax estimation of this density based on a sample path observation up to the time $T$. It means that we have two problems. The first one is to find a lower bound on the second order risk of any estimator. The second one is to construct an estimator, which attains this lower bound. We carry out this program (bound+estimator) following Pinsker’s approach. If the parameter set is a subset of the Sobolev ball of smoothness $k > 1$ and radius $R > 0$, the second order minimax risk is shown to behave as $-T^{-2k/(2k-1)}\Pi(k, R)$ for large values of $T$. The constant $\Pi(k, R)$ is given explicitly.

1 Introduction

The Model. In this paper we deal with a diffusion process $X^T = \{X_t, 0 \leq t \leq T\}$ given by the stochastic differential equation

$$dX_t = S(X_t) \, dt + \sigma(X_t) \, dW_t, \quad X_0 = \xi, \quad 0 \leq t \leq T,$$

(1.1)

where $W^T = \{W_t, 0 \leq t \leq T\}$ is the standard one-dimensional Wiener process. We suppose that the trend coefficient $S(\cdot)$ and the diffusion coefficient $\sigma(\cdot)^2$ are such that the equation (1.1) has a unique weak solution and the conditions

$$\int_{-\infty}^{\infty} \exp \left\{ \int_0^x -2 \int_0^y \frac{S(z)}{\sigma(z)^2} \, dz \right\} \, dy \longrightarrow \pm \infty \quad \text{as} \quad x \to \pm \infty,$$

(1.2)

$$G(S) = \int_{-\infty}^{\infty} \sigma(y)^{-2} \exp \left\{ 2 \int_0^y \frac{S(z)}{\sigma(z)^2} \, dz \right\} \, dy < \infty$$

(1.3)

are fulfilled. We fix a diffusion coefficient $\sigma(\cdot)^2$ and denote by $\Sigma$ the class of all trend coefficients satisfying conditions (1.2) and (1.3). By these conditions the process $X =$

Key words and phrases: ergodic diffusion, invariant density estimation, second order minimax, lower bound, Pinsker’s constant
\{X_t, t \geq 0\} is recurrent positive with the density of invariant law

\[
f_S(x) = \frac{1}{G(S)\sigma(x)^2} \exp\left\{2 \int_0^x \frac{S(z)}{\sigma(z)^2} \, dz\right\}.
\] (1.4)

The aim of the present paper is to estimate the function \(f_S(\cdot)\) nonparametrically, given a sample path \(X^T = \{X_t, 0 \leq t \leq T\}\) of the diffusion process \(X\) defined by (1.1). We assume that \(\sigma(\cdot)\) is a known, sufficiently smooth, positive function such that \(\sigma + \sigma^{-1}\) increases at most polynomially. The trend coefficient is supposed to be unknown.

**First order efficiency.** During last twenty years the problem of stationary density estimation for continuous time processes has been studied by many authors. We mention here Nguyen [25], Prakasa Rao [29], Castellana and Leadbetter [4], Leblanc [23], Kutoyants [21], Bosq and Davydov [3], Bosq [2]. It is proven by Castellana and Leadbetter [4] that for processes satisfying some strong mixing condition, this density can be estimated with the “parametric” rate \(T^{-1/2}\). In most situations, the above mentioned mixing condition is not easy to verify since it involves the joint density of \(X_0\) and \(X_t\), which in general cannot be written explicitly.

In Kutoyants [21, 22], it is shown that for one dimensional ergodic diffusion process, one can obtain the same rate of convergence without verifying that mixing condition. The nonparametric analogue of the Fisher information in this problem is shown to be equal

\[
I(S, x) = \left[4f_S(x)^2 \mathbb{E}_S \left( \frac{X_{\{\xi \geq x\}} - F_S(\xi)}{\sigma(\xi) f_S(\xi)} \right)^2 \right]^{-1}.
\] (1.5)

For more detailed discussion on the Fisher information and its connection with bounding the minimax risk the interested reader is referred to Koshevnik and Levit [20], where the i.i.d. case is presented. Here and in the sequel, \(\mathbb{E}_S\) stands for the expectation with respect to the probability measure induced by the process (1.1) on the space of continuous functions from \([0, T]\) to \(\mathbb{R}\) equipped with the Borel \(\sigma\)-algebra defined by uniform convergence. Also, the random variable \(\xi\) is supposed to follow the invariant law.

In fact, the existence of \(\sqrt{T}\)-consistent estimators in the case of ergodic diffusion can also be explained by the fact that, due to (1.4), the invariant density is a smooth functional of the trend coefficient. Since the trend coefficient can be estimated with the standard nonparametric rate of convergence, it is natural that a smooth functional of trend is estimated with rate \(T^{-1/2}\). An example illustrating the connection between the minimax estimation of a functional parameter and the second order minimax estimation of a smooth functional of this parameter can be found in Golubev and Levit [18].

Furthermore, it is shown in [21] that under mild regularity conditions the local-time estimator

\[
f^\tau_T(x) = \frac{1}{T \sigma(x)^2} \int_0^T \text{sgn}(x - X_t) \, dX_t + \frac{|X_T - x| - |X_0 - x|}{T},
\]

kernel-type estimators \(\hat{f}_{K,T}(x)\) and a wide class of unbiased estimators \(\hat{f}_T(x)\) are consistent, asymptotically normal and asymptotically (first-order) efficient. Moreover, as it is proved in Kutoyants [22, Section 4.3] the corresponding lower bound on the minimax
risk and the efficiency of the estimators hold true for the \( L_2 \)-type risk as well. Particularly, if the bandwidth is of order \( T^{-1/2} \) or smaller, then the kernel estimator satisfies

\[
T \int_{\mathbb{R}} E_S[(\hat{f}_{K,T}(x) - f_S(x))^2] \, dx \xrightarrow{T \to \infty} \mathcal{R}(S) = \int_{\mathbb{R}} I(S,x)^{-1} \, dx.
\]

**Second order minimax approach.** The abundance of the first order efficient estimators motivates our interest in developing the second order minimax approach. The main concepts of this approach can be presented as follows. Any of above cited estimators attaining the lower bound \( (T \mathcal{I}(S,x))^{-1} \), is unbiased or has an asymptotically negligible bias with respect to the variance. Furthermore, the choice of the smoothing parameter (the bandwidth in the case of kernel estimator) does not make the main part of the variance smaller. However, it is clear that the oversmoothing (choosing the smoothing parameter larger) decreases the variance, but this effect is perceptible only on the second term of the asymptotic expansion of the variance. So the main idea of the second order minimax approach is to choose the smoothing parameter by minimizing the sum of the bias and the second term of the variance’s asymptotic expansion.

More precisely, we suppose that the function \( f_S(\cdot) \) is \( k \)-times differentiable and the trend coefficient \( S(\cdot) \) belongs to some set \( \Sigma_* \) defined below. The second order risk is then defined as follows:

\[
\mathcal{R}_T(\hat{f}_T, f_S) = \int_{\mathbb{R}} E_S(\hat{f}_T(x) - f_S(x))^2 \, dx - T^{-1} \mathcal{R}(S),
\]

where \( \hat{f}_T(x) \) is an arbitrary estimator of the density. It is evident that for asymptotically efficient estimators \( \hat{f}_T \), the quantity \( T \mathcal{R}_T(\hat{f}_T, f_S) \) tends to zero, as \( T \to \infty \). It can be shown that for some of these estimators there exists a non degenerate limit for \( T^{\frac{2k}{2k-1}} \mathcal{R}_T(\hat{f}_T, f_S) \) and for the others this limit is equal to infinity. Therefore we can compare the performance of these estimators according to the limits of this quantity.

It is noteworthy that for the local time estimator, the quantity \( \mathcal{R}_T(f^*_T, f_S) \) tends to zero with the speed \( T^{-3/2} \) (this assertion can be easily derived from the martingale representation of the local time estimator). It means that if the smoothness order \( k \) of \( f_S \) exceeds \( 3/2 \), we have \( T^{2k/(2k-1)} \mathcal{R}_T(\hat{f}_T, f_S) \to 0 \), and consequently the minimal limiting value of \( T^{2k/(2k-1)} \mathcal{R}_T(\hat{f}_T, f_S) \) is less than or equal to zero. We show in this paper that in the case of Sobolev regularity \( k \geq 2 \), this minimal value is strictly negative, which entails the inefficiency in second order of the local time estimator. Note that the condition that the function \( f_S \) belongs to a Sobolev ball of smoothness \( k \geq 2 \) means that the function \( S \) is absolutely continuous, which is not a strong restriction.

As it is typical for the results describing the asymptotic behaviour of the minimax risk, we divide the problem into two parts. The first one is to find a constant \( \Pi \) such that the following inequality holds

\[
\lim_{T \to \infty} \inf_{f_T} \sup_{S(\cdot) \in \Sigma_*} T^{\frac{2k}{2k-1}} \mathcal{R}_T(\hat{f}_T, f_S) \geq -\Pi,
\]

where the inf is taken over all estimators \( \hat{f}_T(\cdot) \). This inequality gives us the lower bound for the risks of all estimators. The second part is to construct an estimator \( \hat{f}_T(\cdot) \) which
Second order minimax density estimation

attains this bound, i.e., such that

\[
\lim_{T \to \infty} \sup_{S(.) \in \Sigma_*} T^{\frac{2k}{k-1}} \mathcal{R}_T(\hat{f}_T, f_S) = -\tilde{\Pi}.
\]

Such estimators will be called second order minimax over the parameter set \( \Sigma_* \). We show in this paper that the optimal constant \( \tilde{\Pi} \) is given by the expression

\[
\tilde{\Pi} = \tilde{\Pi}(k, R) = 2(2k - 1) \left( \frac{4k}{\pi(k - 1)(2k - 1)} \right)^{\frac{2k}{k-1}} R^{-\frac{2k}{k-1}}.
\]

(1.6)

The methodology. The approach adopted in this work is inspired by the paper Golubev and Levit [17], who have studied the second order minimax estimation of the distribution function (d. f.) in the i.i.d. case (see also Golubev and Levit [18] for the case of infinitely smooth d. f. estimation). In that paper, the authors prove that Pinsker’s phenomenon (see [28]), well known for classical problems of nonparametric curve estimation, appears in the problem of d. f. estimation when the second order minimax approach is considered. The list of articles concerning the Pinsker bound and related topics can be found in Nussbaum [26] and Efromovich [12]. For ergodic diffusion processes we have already obtained similar results in the problems of density derivative [8] and trend coefficient [9] estimation.

To obtain a lower bound, we construct a parametric family \{\( S(\vartheta, \cdot), \vartheta \in \Theta_T \)\}, which is almost entirely contained in the original nonparametric set \( \Sigma_* \). This allows us to evaluate the global minimax risk by the Bayesian risk in a finite dimensional parameter estimation problem. Then we choose the worst prior distribution and the least favorable parametric family and the risk of the latter problem is shown to be greater than \( T^{-2k/(2k-1)} \tilde{\Pi}(k, R) \).

To construct a second order minimax estimator we use the Hilbert space structure of \( L_2(\mathbb{R}, dx) \). Namely, we argue that the estimation of \( f_S \) in \( L_2 \)-norm is equivalent to the estimation in \( l_2 \)-norm of Fourier coefficients of \( f_S \) in a suitably chosen orthonormal basis of \( L_2^2(\mathbb{R}) \). Afterwards we proceed to a modification of the linear estimator of these coefficients permitting to remove some boundary effects. Then we prove that Pinsker’s filter leads to a second order minimax estimator.

Of course, the estimator proposed in this paper depends on the parameters \( k \) and \( R \) describing the parameter set. In the case when these parameters are not known, one should use an adaptive estimating procedure. Such a procedure can be provided using the unbiased risk minimization (see for instance Cavalier et al. [5]), the Lepski method (see Lepski and Spokoiny [24]) or blockwise adaptation (see Cavalier and Tsybakov [6]). The exact theoretical study of an adaptive procedure lies beyond the scope of this paper and will be done in forthcoming works. We believe that in our problem, combining the techniques developed in the papers Golubev and Levit [19], Golubev and Härdle [16] and Dalalyan [7], it is possible to construct an adaptive estimating procedure achieving the minimax bound obtained in this paper over a wide range of Sobolev balls.

It is worth noting that if the precise value of \( k \) is not known, but the function \( S \) is supposed to be absolutely continuous (therefore \( f'_S \) exists and is absolutely continuous), then the estimator proposed in Section 4 with \( k = 2 \) will have a second order risk of order \( T^{-4/3} \), that is significantly better than the second order risk of the local time estimator.
The parameter set. Let us now describe exactly the conditions imposed on the model. To simplify the computations we put \( \sigma(x) \equiv 1 \), so the diffusion process is given by
\[
dX_t = S(X_t) \, dt + dW_t, \quad X_0, \quad 0 \leq t \leq T,
\]
and we have to estimate the invariant density \( f_S(x) = G(S)^{-1} \exp \left\{ 2 \int_0^T S(v) \, dv \right\} \) based on the observations \( X_T = \{X_T, 0 \leq t \leq T\} \). Some comments on the possible generalizations to the case \( \sigma \neq 1 \) are given in the end of Section 4. For technical reasons, we replace conditions (1.2), (1.3) by a stronger one \( S \in \Sigma_{\gamma, (A_*, C_*, \nu_*)} \), where
\[
\Sigma_{\gamma, (A_*, C_*, \nu_*)} = \left\{ S \in \Sigma : \sgn(x)S(x) \leq -\gamma_* \quad \text{for} \quad |x| > A_* \right\}.
\]
Here \( \gamma_*, A_*, C_* \) and \( \nu_* \) are some positive constants. It is easy to see that for such functions \( S(\cdot) \) the conditions (1.2), (1.3) are fulfilled.

Fix some integer \( k > 1 \). The function \( S(\cdot) \) is supposed to be \( (k - 2) \)-times differentiable with absolutely continuous \((k - 2)\)th derivative and to belong to the set
\[
\Sigma(k, R) = \left\{ S(\cdot) \in \Sigma : \int_\mathbb{R} [f_S^{(k)}(x) - f_S^{(k)}(x)]^2 \, dx \leq R \right\},
\]
where \( R > 0 \) is some constant and \( f_S^{(k)}(x) \) is the \( k \)-th derivative (in the distributional sense) of the function \( f_S(x) \) w.r.t. \( x \). The set \( \Sigma(k, R) \) is a Sobolev ball of smoothness \( k \) and radius \( R \) centered at \( f_S = f_* \). The choice of the center is not arbitrary, it should be smoother than the other functions of the class. For simplicity we consider \( S_*(x) = -x \).

In this case the corresponding diffusion is an Ornstein-Uhlenbek process and the invariant density is infinitely differentiable. Finally we define the parameter set \( \Sigma_* = \Sigma_*(k, R) = \Sigma(k, R) \cap \Sigma_{\gamma, (A_*, C_*, \nu_*)} \).

The paper is organised as follows. The next section contains some heuristical arguments helping to understand the result. We state in Section 3 the lower bound result and prove it up to some technical lemmas. Section 4 is devoted to the construction of the second order minimax estimator. It contains also the proof of its minimaxity. The proofs of the technical results are gathered in Section 5.

2 Heuristical explanation

As we mentioned in the introduction, the local time estimator \( f^0_T(x) \) is asymptotically normal with mean \( f_S(x) \) and variance \( [TI(S, x)]^{-1} \). Since the local time is a sufficient statistic in the model of ergodic diffusion (see Kutoyants [22]), this latter is weakly asymptotically equivalent to the heteroscedastic Gaussian experiment
\[
Y_t = f_S(t) + \frac{2f_S(t)}{\sqrt{T}} \int_\mathbb{R} \frac{\chi_{\{u > t\}} - F_S(u)}{\sqrt{f_S(u)}} \, dB_u,
\]
where \( B = (B_u, u \in \mathbb{R}) \) is a Brownian motion. Let now \( v(\cdot) \) be a test function (infinitely differentiable with compact support) and \( V(\cdot) \) be its primitive. A formal use of
the integration by parts formula yields

\[
\int_{\mathbb{R}} v(t) Y_t \, dt = \int_{\mathbb{R}} v(t) f_S(t) \, dt + \frac{2}{\sqrt{T}} \int_{\mathbb{R}} v(t) f_S(t) \int_{\mathbb{R}} \frac{\chi_{(u>t)} - F_S(u)}{\sqrt{f_S(u)}} \, dB_u \, dt
\]

\[
= - \int_{\mathbb{R}} V(t) f'_S(t) \, dt - \frac{2}{\sqrt{T}} \int_{\mathbb{R}} V(t) \sqrt{f_S(t)} \, dB_t
\]

\[
- \frac{2}{\sqrt{T}} \int_{\mathbb{R}} V(t) f'_S(t) \int_{\mathbb{R}} \frac{\chi_{(u>t)} - F_S(u)}{\sqrt{f_S(u)}} \, dB_u \, dt.
\]

The third term of this sum can be dropped in the asymptotical results, since in contrast with the second one, it is uniformly bounded in $L^2$ (see [8]) in the following sense:

\[
\sup_{S \in \Sigma} \mathbb{E}_S \left[ \int_{\mathbb{R}} f'_S(t) \int_{\mathbb{R}} \frac{\chi_{(u>t)} - F_S(u)}{\sqrt{f_S(u)}} \, dB_u \right]^2 \, dt < \infty.
\]

Therefore, if we drop the third term we come to the Gaussian experiment

\[
dY_t = f'_S(t) \, dt + 2 \sqrt{f_S(t)} T^{-1} \, dB_t, \quad t \in \mathbb{R}.
\]

As it is noted by Golubev and Levit [18], second order minimax estimation of a function, which is a regular functional (invariant density in our case) of the unknown parameter, is closely related to the problem of first order minimax estimation of the derivative of this function. On the other hand, it is known (Golubev [15]) that in the Gaussian shift experiment $dY_t = \theta(t) \, dt + \epsilon \sqrt{I^{-1}(t)} \, dB_t$, the optimal constant depends on the Fisher information $I(t)$ only via the integral $\int I^{-1}(t) \, dt$. Due to the fact that $f_S$ is a probability density, the integral of the inverse of the Fisher information in our case is equal to 2 and does not depend on $S$.

This explains why we succeed to obtain the second order minimax (up to constant) bound over a global nonparametric class of the unknown parameter. These arguments show also that the restriction to the $L^2$-norm in the risk definition is essential for carrying out the optimal constants.

To end up with heuristic reasoning, let us remark that the reason, which made the obtention of the second order minimax constant possible in the problem of the distribution function $F(\cdot)$ estimation from i.i.d. observations, was the same. Indeed, according to Nussbaum [27], the statistical problem of $F'(\cdot)$ estimation is locally asymptotically equivalent to recovering the function $F'(\cdot)$ from the Gaussian shift experiment $dY_t = F'(t) \, dt + \sqrt{n^{-1} F'_0(t)} \, dB_t$, where $n$ is the size of the sample and $F_0$ is the center of localisation.

### 3 Lower bound

The first result of this work concerns the asymptotically exact lower bound on the $L_2$-risks of all estimators. This bound turns out to be almost the same as in the problem of the distribution function estimation of a sequence of i.i.d. random variables (Golubev and Levit [17]). The main reason of this similarity is that these constants appear as the solutions of two very similar optimisation problems.
Theorem 3.1 Let the integer $k > 1$, then
\[
\lim_{T \to \infty} \inf_{f_T} \sup_{S \in \Sigma} T^{-\frac{1}{2k}} \mathcal{R}_T(f_T, f_S) \geq -\hat{\Pi}(k, R),
\]  
where
\[
\hat{\Pi}(k, R) = 2(2k - 1) \left( \frac{4k}{\pi(k - 1)(2k - 1)} \right)^{\frac{2k}{2k - 1}} R^{-\frac{1}{2k - 1}}.
\]

Proof: The main steps of the proof are close to those of Theorem 1 in [9]. The minimax risk over the nonparametric set $\Sigma$ is evaluated by the Bayesian risk over a parametric set of increasing dimension (constructed like in Golubev [15] and Dalalyan, Kutoyants [9]), then the van Trees inequality is applied.

Let us introduce a parametric family of diffusion processes
\[
dX_t = S(\vartheta, X_t) \, dt + dW_t, \quad X_0, \quad 0 \leq t \leq T,
\]
where $S(\vartheta, \cdot)$ is chosen in the following way: We fix an increasing interval $[-A, A]$ with $A = A_T = \gamma^{-1} \log(T + 1)$ and a sequence of its sub-intervals $I_m = [a_m - AT^{-\beta}, a_m + AT^{-\beta}]$, where $\beta = (2k - 1)^{-1}$ and $a_m = 2mAT^{-\beta}$, $m = 0, \pm 1, \pm 2, \ldots, \pm M$. Here $M = M_T$ is the greatest integer such that the interval $I_M$ is entirely included in $[-A, A]$.

Let us denote
\[
S(\vartheta, x) = S_0(x) + \sum_{|m| \leq M} \sqrt{\frac{2A}{T^\beta f_0(a_m)}} \sum_{|i| \leq L} \vartheta_{i,m} g_{i,m}(x),
\]
where $\vartheta$ is the $(2M + 1) \times (2L + 1)$ matrix of coefficients $\vartheta_{i,m}$ and
\[
g_{i,m}(x) = \sqrt{T^\beta / A} e_i(T^\beta A^{-1}(x - a_m)) U(A - |x - a_m| T^\beta).
\]
Here $e_i(\cdot)$ is the trigonometric basis on $[-1, 1]$, that is
\[
e_i(x) = \begin{cases} \sin(\pi x), & \text{if } i > 0, \\ 1/\sqrt{2}, & \text{if } i = 0, \\ \cos(\pi x), & \text{if } i < 0, \end{cases}
\]
the function $U(\cdot)$ is $(k + 1)$-times differentiable, increasing, vanishing for $x \leq 0$ and equal to one for $x \geq 1$. The integer $L = L_T$ will be chosen later.

The parameter $\vartheta \in \Theta_T$ is of increasing dimension and
\[
|\vartheta_{i,m}| \leq K \sqrt{\sigma_l(\varepsilon)}, \quad \sigma_l(\varepsilon) = \frac{1}{2AT^{1-\beta}} \left( \left( \frac{A(1 - \varepsilon)}{l} \right)^k - 1 \right)_+, \quad (3.3)
\]
for $l \neq 0$, and $\sigma_0 = T^{-\beta}$, for $l = 0$. Above we used the standard notation $B_+ = \max(0, B)$; $\varepsilon$ is a positive number and
\[
\Lambda = \Lambda_T = A \left( \frac{R(k - 1)(2k - 1)}{4k\pi^{2k-2}} \right)^{\frac{1}{2k-1}}.
\]
The number \( L = L_T \) is now the integer part of \( \Lambda \). For the moment, the choice of the \( \sigma_I \) and \( \Lambda \) is uncomprehending, but it will be clarified a little later.

Note that for \( S(\cdot) \in \Sigma_{\gamma_1}(A_*, C_*, \nu_*) \) we have the estimate (see [9, Lemma 4 in Appendix])

\[
\sup_{\theta \in \Theta_T} \int_{|x| > A} f_\theta(x)^2 \mathbb{E}_\theta \left( \frac{X(x) - F_\theta(x)}{f_\theta(x)} \right)^2 dx \leq C e^{-\gamma_* A},
\]

which implies that the choice \( A = \gamma_*^{-1} \log T \) makes this term asymptotically negligible. Therefore it is sufficient to study the lower bound for the risk

\[
\hat{R}_T(\hat{f}_T, f_S) = \mathbb{E}_S \int_{-A}^A \left[ \hat{f}_T(x) - f_S(x) \right]^2 dx - \frac{1}{T} \int_{-A}^A 1^{-1}(S, x) dx
\]

because \( |\hat{R}_T(\hat{f}_T, f_S) - \hat{R}_T(\hat{f}_T, f_S)| \leq C T^{-2} \), and the order of the term \( \hat{R}_T \) will be shown to be less than \( T^{-2} \).

We consider now only the estimators belonging to the set \( \mathcal{W}_T = \{ \hat{f}_T \mid \hat{R}_T(\hat{f}_T, f_S) < 1 \} \). It is evident that in the proof of lower bound one can drop all the estimators do not belonging to \( \mathcal{W}_T \). Moreover, it is clear that for \( T \) large enough this set is not empty (since there exist consistent estimators). Suppose that the parameter \( \theta \) is a random matrix with a prior distribution \( Q(d\theta) \) on the set \( \Theta_T \). Then we have the obvious inequality

\[
\sup_{\theta \in \Theta_T} \hat{R}_T(\hat{f}_T, f_\theta) \geq \mathbb{E}_Q \hat{R}_T(\hat{f}_T, f_\theta) = \int_{\Theta_T} \hat{R}_T(\hat{f}_T, f_\theta) Q(d\theta),
\]

where the expectation \( \mathbb{E}_Q \) is defined by the last equality. We can write now the following sequence of inequalities:

\[
\begin{align*}
\inf_{\hat{f}_T \in \mathcal{W}_T} \sup_{S \in \Sigma_*} \hat{R}_T(\hat{f}_T, f_S) &\geq \inf_{\hat{f}_T \in \mathcal{W}_T} \sup_{S \in \Sigma_* \cap \Theta_T} \hat{R}_T(\hat{f}_T, f_S) \\
&\geq \inf_{\hat{f}_T \in \mathcal{W}_T} \int_{\Theta_T} \hat{R}_T(\hat{f}_T, f_\theta) Q(d\theta) - \sup_{\hat{f}_T \in \mathcal{W}_T} \int_{\Theta_T \setminus \Sigma_*} \hat{R}_T(\hat{f}_T, f_\theta) Q(d\theta) \\
&\geq \hat{R}_T(Q) - \sup_{\hat{f}_T \in \mathcal{W}_T} \int_{\Theta_T \setminus \Sigma_*} \hat{R}_T(\hat{f}_T, f_\theta) Q(d\theta).
\end{align*}
\]

Here \( \hat{R}_T(Q) \) denotes the second order Bayesian risk with respect to the prior \( Q \), that is

\[
\hat{R}_T(Q) = \mathbb{E}_Q [\hat{R}_T(\hat{f}_T, f_\theta)].
\]

The next step is to choose a prior distribution that maximises the second order Bayesian risk and, in the same time, is essentially concentrated on the set \( \Sigma_* \), that is the probability of the set \( \Theta_T \setminus \Sigma_* \) is sufficiently small. This distribution \( Q \) is usually called asymptotically least favorable prior distribution. In our case, it is defined in the following way:
let \( \eta_{l,m} \), \( l = 0, \pm 1, \pm 2, \ldots, \pm L \), \( m = 0, \pm 1, \ldots, \pm M \), be i.i.d. random variables with common density \( p(\cdot) \), such that
\[
|\eta_{l,m}| \leq K, \quad \mathbb{E}[\eta_{l,m}] = 0, \quad \mathbb{E}[\eta_{l,m}^2] = 1, \quad J = \int_{-K}^{K} \frac{\hat{p}(x)^2}{p(x)} \, dx = 1 + \varepsilon,
\]
where \( \varepsilon \to 0 \) as \( K \to \infty \) (for example, one can set \( K = \varepsilon^{-1} \)). Then \( Q \) is the distribution of the random array \( \vartheta_{l,m} = \sqrt{\sigma_l(\varepsilon)} \eta_{l,m} \), in other words, it is the product measure
\[
Q(d\vartheta) = \prod_{|m| \leq M} \prod_{|l| \leq L} \frac{1}{\sqrt{\sigma_l(\varepsilon)}} \frac{p(\vartheta_{l,m} \sqrt{\sigma_l(\varepsilon)})}{\sigma_l(\varepsilon)} \, d\vartheta_{l,m}.
\]
Therefore the Fisher information of one component of this prior distribution \( Q \) is
\[
J_{l,m} = J_l = (1 + \varepsilon) \sigma_l(\varepsilon)^{-1}.
\]

We shall evaluate now the Bayesian risk, which will give us the main part of the minimax risk. Remark firstly that the set of functions \( e_{l,m}(x) = \sqrt{A^{-1}T^\beta} e_l(A^{-1}T^\beta(x - a_m)) \) forms an orthonormal basis of the Hilbert space \( L^2[-A, A] \) and hence, using Parseval’s identity, we can write
\[
\mathbb{E}_Q(\hat{\mathcal{R}}_T(\tilde{f}_T, f_\vartheta)) = \mathbb{E} \int_{-A}^{A} [\tilde{f}_T(x) - f_\vartheta(x)]^2 \, dx - \frac{1}{T} \mathbb{E} \int_{-A}^{A} I_f^{-1}(S, x) \, dx
\]
\[
\geq \sum_{|m| \leq M} \sum_{|l| \leq L} \left[ \mathbb{E} |\psi_{l,m,T} - \psi_{l,m,\vartheta}|^2 - \frac{4}{T} \mathbb{E} |\Psi_{l,m,\vartheta}(\xi)|^2 \right]
\]
\[
- \sum_{|m| \leq M} \sum_{|l| > L} \frac{4}{T} \mathbb{E} |\Psi_{l,m,\vartheta}(\xi)|^2 := A_{1T} + A_{2T}.
\]

where \( \mathbb{E} \) is the expectation with respect to the probability measure \( \mathbb{P}_\vartheta^{(T)} \times Q(d\vartheta) \) and \( \psi_{l,m,T}, \psi_{l,m,\vartheta} \) and \( \Psi_{l,m,\vartheta}(y) \) are the Fourier coefficients
\[
\psi_{l,m,T} = \int_{I_m} \tilde{f}_T(x) e_{l,m}(x) \, dx,
\]
\[
\psi_{l,m,\vartheta} = \int_{I_m} f_\vartheta(x) e_{l,m}(x) \, dx,
\]
\[
\Psi_{l,m,\vartheta}(y) = \int_{I_m} f_\vartheta(x) \left( \frac{\chi_{\{y>x\}} - F_\vartheta(y)}{f_\vartheta(y)} \right) e_{l,m}(x) \, dx
\]
of the corresponding functions.

From the other nonparametric problems related to this model (see [8, 9]), it turned out that the quantity
\[
I_{l,m}(\vartheta) = \mathbb{E}_\vartheta \left( \frac{\partial S(\vartheta, \xi)}{\partial \vartheta_{l,m}} \right)^2 = \frac{2A}{T^\beta f_\vartheta(a_m)} \mathbb{E}_\vartheta \left[ \eta_{l,m}^2(\xi) \right] < \infty
\]
plays the role of Fisher information concerning the parameter \( \vartheta_{l,m} \) in the van Trees inequality (see Gill and Levit [14]). That is, the following inequality holds

\[
E[|\psi_{l,m,T} - \psi_{l,m,\vartheta}|^2] \geq \left( \frac{E_Q \partial \psi_{l,m,\vartheta}}{T E_Q I_{l,m}(\vartheta)} \right)^2 T E_Q I_{l,m}(\vartheta) + J_{l,m}.
\]

From this inequality one can deduce that

\[
E[|\psi_{l,m,T} - \psi_{l,m,\vartheta}|^2] \geq -\frac{4 J_l E_Q \psi_{l,m,\vartheta}(\xi)}{T(T E_Q I_{l,m}(\vartheta) + J_l)} + \mu_{l,m,T}, \quad (3.7)
\]

where

\[
\mu_{l,m,T} = \frac{(E_Q \partial \psi_{l,m,\vartheta})^2 - 4 E_Q \psi_{l,m,\vartheta}(\xi) E_Q |\psi_{l,m,\vartheta}(\xi)|^2}{T E_Q I_{l,m}(\vartheta) + J_l}.
\]

Due to Lemma 5.4, Section 5, the sum of the terms \( \mu_{l,m,T} \) is less than a power of \( A \) divided by \( T^{1+2\beta} \), which is clearly \( o_T(1)/T^{2\beta} \). This implies that the main part of the convergence of the second order Bayesian risk is given by the first terms of inequality (3.7).

Now we shall evaluate the sum of the first terms in the inequality (3.7). Since the function \( g_{l,m}(\cdot) \) is very close in supremum norm to the normalised function \( e_{l,m}(\cdot) \), we have

\[
E_Q[|l_{l,m}(\vartheta)|] \geq 2AT^{-\beta}(1 - \varepsilon).
\]

Remark that the sum over \( m \) of the first terms in the inequality (3.7) corresponding to the case \( l = 0 \) is smaller in order than \( T^{-\frac{2\beta}{1-\beta}} \), since the value \( J_0 \) is chosen significantly smaller than the other \( J_l \)'s. Further, using the result of Corollary 5.3, Section 5, together with the inequality of Minkowsky, we get

\[
A_{1T} \geq -\frac{4}{T} \sum_{m=-M}^{M} \sum_{l=\pm 1}^{L} J_l E_Q \psi_{l,m,\vartheta}(\xi)^2 + o_T(1) T \frac{1}{2^{2\beta} \pi^{2}}
\]

Further, direct calculations (for more details, see [9]) yield the estimate

\[
\int_{-A}^{A} [f^{(k)}(\vartheta, x) - f^{(k)}_{\star}(x)]^2 dx
\]

\[
= 4(1 + o_T(1)) \int_{-A}^{A} [S^{(k-1)}(\vartheta, x) - S^{(k-1)}_{\star}(x)]^2 f^2_{\star}(x) dx
\]

\[
= 4(1 + o_T(1)) \sum_{|m| \leq M} \frac{2A}{T^{\beta} f_{\star}(a_m)} \sum_{1 \leq |i| \leq L} \vartheta_{i,m}^2 f_{\star}^2(a_m) \left( \frac{\pi l A}{T^{\beta}} \right)^{2(k-1)}.
\]
If we evaluate now the mathematical expectation with respect to $Q$ of this expression, using the fact that the partial sums of the function $f_\ast(\cdot)$ tend to its integral (which equals 1), we obtain

$$
E \int_\mathbb{R} \left[ f^{(k)}(\vartheta, x) - f^{(k)}_\ast(x) \right]^2 dx \\
= 4(1 + o_T(1)) \sum_{|m| \leq M} \frac{2A f^{(k)}_\ast(a_m)}{T^\beta} \sum_{1 \leq |l| \leq L} \sigma_l(\varepsilon) \left( \frac{\pi l T^\beta}{A} \right)^{2(k-1)} \\
= 4(1 + o_T(1)) \sum_{1 \leq |l| \leq L} \sigma_l(\varepsilon) \left( \frac{\pi l T^\beta}{A} \right)^{2(k-1)}.
$$

(3.10)

We can now explain the choice (3.5) of the prior distribution. It is clear from (3.8) that the least favorable prior is the one, for which the Fisher informations $J_l$ are minimal under the constraint (3.10) involving only the variances $\{\sigma_l(\varepsilon), l = \pm 1, \ldots, \pm L\}$. It is well known that the minimum of the Fisher information of a random variable with zero mean and fixed variance $\sigma$ is attained by the Gaussian distribution and this minimum is equal to $\sigma^{-1}$. We can not use here directly the product of normal distributions as a prior distribution, since for H"{o}ffding’s inequality we need to have bounded random variables. That is why we use an approximation of Gaussian distribution and we pay a factor $1 + \varepsilon$ for the error of this approximation.

In order to determine the values of $\sigma_l$, we solved the maximisation problem derived from (3.9) and (4.3). Thus, we looked for the sequence $y = (y_l)$ maximising the functional

$$
\Phi(y) = \sum_{l=1}^L \frac{1}{\left[ 2(2T^{1-\beta}A y_l + 1) \right]^2},
$$

over the set $\mathcal{E}(k, R) = \{ y \in \mathbb{R}_+^L : 8 \sum_{l=1}^L y_l \left( \frac{\pi l T^\beta}{A} \right)^{2(k-1)} \leq R \}$. This can be done directly using the method of Lagrange multipliers and the result is given by $y'_l$ equal to $(2AT^{1-\beta})^{-1} \left( \left[ \Lambda l^{-1} \right]_k - 1 \right)_+$, which leads to the definition of $\sigma_l$ presented in the beginning of the proof.

Now we shall evaluate the sums in right hand sides of (3.8) and (3.10). For the second one, setting $\Lambda_\varepsilon = \Lambda(1 - \varepsilon)$ we have

$$
\sum_{1 \leq |l| \leq L} \sigma_l(\varepsilon) \left( \frac{\pi l T^\beta}{A} \right)^{2(k-1)} = \frac{1}{2AT^{1-\beta}} \sum_{1 \leq |l| \leq L} \left( \left[ \frac{\Lambda_\varepsilon}{l} \right]_k^{k-1} - 1 \right)_+ \left( \frac{\pi l T^\beta}{A} \right)^{2(k-1)}
$$

$$
= \frac{\pi^{2(k-1)} \Lambda_\varepsilon^{2k-1}}{2A^{2k-1}} \sum_{1 \leq |l| \leq L} \left( \left[ \frac{l}{\Lambda_\varepsilon} \right]_k^{k-2} - \left[ \frac{l}{\Lambda_\varepsilon} \right]^{2k-2} \right)_+ \frac{1}{\Lambda_\varepsilon}.
$$
Since $\Lambda_\varepsilon$ tends to infinity, when $T$ tends to infinity, the last sum is close to the integral over $[0, 1]$ of the function $x^{k-2} - x^{2k-2}$. This yields

$$\sum_{1 \leq |l| \leq L} \sigma_l(x) \left( \frac{\pi T \beta}{A} \right)^{2(k-1)} = (1 + o_T(1)) \frac{\pi^2 2^{k-1} \Lambda_\varepsilon^{k-1}}{2A^{2k-1}} \int_0^1 (x^{k-2} - x^{2k-2}) \, dx$$

$$= \frac{(1 + o_T(1)) \pi^2 2^{k-1} \Lambda_\varepsilon^{k-1}}{2A^{2k-1}} \frac{k}{(k-1)(2k-1)} \leq \frac{R(1 - \varepsilon)}{8}.$$  

This upper estimate permits us to use the Hoeffding inequality to obtain an exponential bound for $Q$-probability of the set $\Theta_T \setminus \Sigma_\ast$ (see [9] for details). Thus we get

$$Q(\Theta_T \setminus \Sigma_\ast) \leq CT^{-2}. \quad (3.11)$$

Evaluating the sum of (3.8) in the same way, we get

$$A_1 T \geq -8 \frac{(1 + \varepsilon)^2 A}{(1 - \varepsilon)^2 T^{1 + \beta} \pi^2} \sum_{l=1}^{L} \frac{1}{l^2} \left( \frac{l}{\Lambda_\varepsilon} \right) = -8 \frac{(1 + \varepsilon)^2 A(1 + o_T(1))}{(1 - \varepsilon)^2 T^{1 + \beta} \pi^2 A} \int_0^1 x^{k-2} \, dx$$

$$\geq -8 \frac{(1 + \varepsilon)^3 A}{(1 - \varepsilon)^2 T^{1 + \beta} \pi^2 A(k - 1)}.$$  

Similarly, due to Corollary 5.3, Section 5,

$$A_2 T = -8 \frac{A(1 + o_T(1))}{T^{1 + \beta} \pi^2} \sum_{l=L+1}^{\infty} l^{-2} = -8 \frac{A(1 + o_T(1))}{\Lambda_\varepsilon T^{1 + \beta} \pi^2} \int_1^\infty x^{-2} \, dx$$

$$\geq -8 A(T \Lambda_\varepsilon^{1 + \beta} \pi^2 \Lambda(k - 1)).$$

Putting together the last two estimates we get the following lower bound for the Bayesian risk with prior distribution $Q$:

$$E_Q R_T(\tilde{f}_T, f_\vartheta) \geq -8 \frac{(1 + \varepsilon)^3 A}{(1 - \varepsilon)^2 T^{1 + \beta} \pi^2 A(k - 1)} = -T^{-\frac{k-1}{2\pi^2 + 1}} \Pi(k, R) \frac{(1 + \varepsilon)^3}{(1 - \varepsilon)^2}.$$  

This inequality, (3.4) and (3.11) imply the desired result, since $\tilde{f}_T \in \mathcal{W}_T$ and the real number $\varepsilon$ can be chosen as small as we want.

\section{Estimator}

Our goal now is to construct an estimator attaining the lower bound obtained in the previous section. Note that in the problem of signal detection with Gaussian white noise, there is a linear estimator which is second order minimax. The linear filter defining this estimator utilises the so called Pinsker’s weights.

The same assertion holds true for the problem of distribution function estimation from i.i.d. observations. In this case the second order minimax estimator is linear with
respect to the (global) trigonometric basis on the interval \([-B, B]\), where \(B\) tends to infinity when the number of observations tends to infinity.

It turns out that an analogous result holds true in the model of ergodic diffusion, but one has to be careful in the definition of linear estimator. More precisely, the estimation of the function \(f\) in \(L^2\)-norm is equivalent to estimating in \(L^2\) its Fourier transform. Therefore, it would be natural to expect that the best linear estimator of the Fourier transform of \(f\) (which leads to a kernel estimator) is asymptotically optimal over the set of all possible estimators. But this assertion turned out to be wrong.

A very heuristical explanation of this fact is that when we consider the linear estimators of the Fourier transform, we disperse the information concerning each point over the whole real line and therefore, the estimator does not take into account the local structure of the model with respect to the space parameter. That is why, the localised bases of functions fit better to our model, in particular it appears that a slightly modified linear estimator with respect to a localised trigonometric basis is second order asymptotically minimax.

To construct our estimator, we introduce the localised orthonormal basis \(\{e_{l,m} : l, m \in \mathbb{Z}\}\) of \(L^2(\mathbb{R})\) defined in the following way

\[
e_{l,m}(x) = \frac{1}{\sqrt{2\delta_T}} \exp \left\{ \frac{i\pi l(x - a_m + \delta_T)}{\delta_T} \right\} \chi_{\{x \in I_m\}}.
\]

Above \(i = \sqrt{-1}\) and \(I_m\) is the interval of length \(2\delta_T\) with center \(a_m = 2m\delta_T\). We call this basis localised since \(\delta_T\) tends to zero when \(T \to \infty\). The Fourier coefficients of the local time and the invariant density with respect to this basis are

\[
\varphi^0_{T,l,m} = \int_{\mathbb{R}} \overline{\varphi}_{l,m}(x) f_T(x) \, dx = \frac{1}{T} \int_0^T \overline{\varphi}_{l,m}(X_t) \, dt,
\]

\[
\varphi_{S,l,m} = \int_{\mathbb{R}} \overline{\varphi}_{l,m}(x) f_S(x) \, dx = \mathbb{E}_S[\varphi_{l,m}(X_t)],
\]

where \(\overline{\varphi}_{l,m}\) states for the complex conjugate of \(e_{l,m}\). Of course, the estimation of the function \(f_S(\cdot)\) in \(L^2\) is equivalent to the estimation of the infinite matrix \(\{\varphi_{S,l,m} : l, m \in \mathbb{Z}\}\). Since \(\varphi^0_{T,l,m}\) is the empirical estimator of \(\varphi_{S,l,m}\), the linear estimator of \(\{\varphi_{S,l,m} : l, m \in \mathbb{Z}\}\) is defined as

\[
\bar{\varphi}_{T,l,m} = \varphi^0_{T,l,m} \varphi^0_{T,l,m}, \quad l, m \in \mathbb{Z},
\]

where \(\varphi^0_{T,l,m}\) are some numbers between 0 and 1. This leads us to the ensuing definition of the linear estimator of the invariant density: \(\bar{f}_T(x) = \sum_{l,m \in \mathbb{Z}} \varphi_{T,l,m} \varphi^0_{T,l,m} e_{l,m}(x)\).

In order to attain the lower bound of the previous section, this linear estimator should be modified. In some sense, we have to correct it at boundaries. Let us discuss it more in details. By condition the estimated function \(\hat{f}_S\) belongs to the Sobolev class \(\Sigma(k, R)\). We would like to derive from this condition an ellipsoid type inequality for the coefficients \(\varphi_{S,l,m}\). Using the integration by parts formula, for \(l \neq 0\), we get

\[
\varphi_{S,l,m} = \varphi f^{(k)}(l, m) \left( \frac{\delta_T}{i\pi l} \right)^k \left( \frac{\delta_T}{i\pi l} \right)^{k-1} \sum_{j=0}^{k-1} \left( f^{(j)}(a_m + \delta_T) - f^{(j)}(a_m - \delta_T) \right) \left( \frac{\delta_T}{i\pi l} \right)^{j+1}.
\]
This last sum will be denoted $q_{l,m}(f)$. Now the condition $\int_R \left( f_{S}^{(k)}(x) \right)^2 \, dx \leq R$ can be rewritten as
\[
\sum_{m \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} \left| \varphi_{S,l,m} + q_{l,m}(f_S) \right|^2 \left( \frac{\pi l}{\delta_T} \right)^{2k} \leq R. \tag{4.2}
\]

Recall that the regularity condition imposed on the estimated function determines the behaviour of the bias and does not influence the variance. That is why we want to modify the linear estimator defined by (4.1) so that the bias of the modified estimator be linear with respect to $\varphi_{S,l,m} + q_{l,m}(f_S)$ and its variance remains unchanged. It is easy to check that the sequence
\[
\varphi_{T,l,m}^* = \varphi_{T,l,m}(\varphi_{T,l,m} + q_{T,l,m}(f_S)) - q_{T,l,m}(f_S), \quad l, m \in \mathbb{Z},
\]
satisfies these conditions, but it is not an estimator since it involves the unknown function $f_S$ via the term $q_{T,l,m}(f_S)$. To overcome this problem, we substitute this term by a rate optimal estimator $\hat{q}_{T,l,m}$. Thus,
\[
\tilde{\varphi}_{T,l,m} = \varphi_{T,l,m}(\varphi_{T,l,m} + \hat{q}_{T,l,m}) - \hat{q}_{T,l,m}, \quad l, m \in \mathbb{Z},
\]
where
\[
\hat{q}_{T,l,m} = \frac{(-1)^l}{\sqrt{2\delta_T}} \sum_{j=0}^{k-1} \left( \hat{f}_{T}^{(j)}(a_m + \delta_T) - \hat{f}_{T}^{(j)}(a_m - \delta_T) \right) \left( \frac{\delta_T}{i\pi l} \right)^{j+1},
\]
and $Q$ is a kernel of order $k$ with support $[-1, 1]$. From now on, we set $h_T = T^{-1/(2k-1)}$ and $\delta_T = h_T^{1/2}$. Note that this trick with $q_{T,l,m}$ has been already exploited by Delattre and Hoffmann [10], it is particularly useful in the case of nonhomogeneous Fisher information.

We show in this section that there exists a modified linear estimator which is second order minimax. The weights $\varphi_{T,l,m}$ of this estimator are defined by $\hat{\nu}_{T,l,m} = \left( 1 - \frac{|l|}{\hat{\nu}^{(k+\mu_T)}} \right)_+$, where $\mu_T = 1/\sqrt{\log T}$ tends to zero as $T \to \infty$ and
\[
\hat{\nu}_{T} = \frac{8k\pi^{2(k-1)}}{TR(k-1)(2k-1)} \left( \frac{\pi}{\delta_T} \right)^{2k-1}.
\]
We denote by $\hat{f}_{T}^* \cdot$ the estimator of the invariant density defined via these weights. Note that $\hat{\nu}_{T}$ tends to infinity as $T$ tends to infinity.

**Theorem 4.1** Let $k > 1$, then the estimator $\hat{f}_{T}^*$ is second order asymptotically minimax, that is
\[
\lim_{T \to \infty} \sup_{S(\cdot) \in \Sigma} T^{\frac{2k}{2k-1}} \mathcal{R}(\hat{f}_{T}^*, f_S) = -\hat{\Pi}(k, R). \tag{4.3}
\]
Proof: Let $S(\cdot) \in \Sigma_s(k, R)$. Recall that the normalised first order minimax risk asymptotically equals

$$R(S) = 4 \int_{\mathbb{R}} f_S^2(x) E_S \left[ \frac{X(x) - F_S(x)}{f_S(x)} \right] dx = \sum_{l,m \in \mathbb{Z}} E_S \left[ \psi_{S,l,m}^2(\xi) \right]$$

where we applied the Parseval identity to the square integrable function $g_S(\cdot, x)$ having the following Fourier coefficients:

$$\psi_{S,l,m}(x) = 2 \int_{\mathbb{R}} \tau_{l,m}(y)f_S(y) \left[ \frac{1_{\{x>y\}} - F_S(x)}{f_S(x)} \right] dy.$$ 

Using this decomposition, one can rewrite the second order risk function as

$$\mathcal{R}(\hat{f}_T, f_S) = E_S \left[ \sum_{l,m \in \mathbb{Z}} \left( |\hat{\varphi}_{T,l,m} - \varphi_{S,l,m}|^2 - T^{-1}E_S[\psi_{S,l,m}^2(\xi)] \right) \right]$$

$$= \sum_{l,m \in \mathbb{Z}} E_S \left[ \hat{\varphi}_{T,l,m} \varphi_{T,l,m} - \varphi_{S,l,m} + (\hat{\varphi}_{T,l,m} - 1)\hat{q}_{T,l,m} \right]^2 - \frac{E_S[\psi_{S,l,m}^2(\xi)]}{T}$$

$$= \sum_{l,m \in \mathbb{Z}} E_S \left[ (\hat{\varphi}_{T,l,m} - \varphi_{T,l,m}^2) - (\varphi_{S,l,m} - \varphi_{T,l,m}^2) \right]^2 - \frac{1}{T} E_S[\psi_{S,l,m}^2(\xi)].$$

On the one hand, we shall show below that the term

$$\sum_{l,m \in \mathbb{Z}} E_S \left[ (\hat{\varphi}_{T,l,m} \varphi_{T,l,m} - \varphi_{S,l,m}) - (\varphi_{S,l,m} - \varphi_{T,l,m}^2) \right]^2$$

is of order $T^{-1}$. On the other hand, the second inequality of Lemma 5.5 implies that

$$\sum_{l,m \in \mathbb{Z}} E_S \left[ (\hat{\varphi}_{T,l,m} - 1)\hat{q}_{T,l,m} - \varphi_{S,l,m} \right]^2$$

is of order $T^{-(2k+1/2)/(2k-1)}$. So, a simple application of the H"older inequality yields

$$\mathcal{R}(\hat{f}_T, f_S) \leq \sum_{l,m \in \mathbb{Z}} \left( E_S \left[ \hat{\varphi}_{T,l,m} \varphi_{T,l,m} - \varphi_{S,l,m} \right] - (\varphi_{S,l,m} - \varphi_{T,l,m}^2) \right)^2$$

$$- T^{-1}E_S[\psi_{S,l,m}^2(\xi)] + CT^{-(2k+1/4)/(2k-1)}.$$ 

The stationarity of the process $X$ implies that the mathematical expectation of $\varphi_{T,l,m}$ is equal to $\varphi_{S,l,m}$, therefore

$$\mathcal{R}(\hat{f}_T, f_S) = \sum_{l,m \in \mathbb{Z}} \left( (\hat{\varphi}_{T,l,m} - 1)^2 |\varphi_{S,l,m} - q_{l,m}(f_S)|^2 + |\hat{\varphi}_{T,l,m}|^2 \text{Var}_S[\varphi_{T,l,m}] \right)$$

$$- T^{-1}E_S[\psi_{S,l,m}^2(\xi)] + CT^{-(2k+1/4)/(2k-1)}.$$
The first term of this sum is in the appropriate form and we show below that it is of order $T^{-\frac{2k}{2k+1}}$. To determine how behave the second and the third terms, we need several technical results.

**Lemma 4.2** The following relations hold:

\[
\sum_{l,m \in \mathbb{Z}} |\hat{\varphi}_{T,l,m}|^2 \left( \text{Var}_S[\varphi_{T,l,m}^o] - T^{-1} \mathbb{E}_S[\psi_{S,l,m}^2(\xi)] \right) \leq T^{-2/(2k+1)} o_T(1),
\]

\[
\sum_{l,m \in \mathbb{Z}} (1 - |\hat{\varphi}_{T,l,m}|^2) \mathbb{E}_S[\psi_{S,l,m}^2(\xi)] \geq (1 + o_T(1)) \sum_{l,m \in \mathbb{Z}^*} \frac{8f_S(a_m)(1 - |\hat{\varphi}_{T,l,m}|^2)}{\delta_T^{-2} \pi l^2}.
\]

The proof of this lemma is quite technical and relies essentially on the arguments used in the proofs of Lemmas 5.2–5.4, Section 5. That is why it will be omitted here.

Using this lemma, for any linear filter $\hat{\varphi} = \{\hat{\varphi}_{T,l,m}\}$ such that $|\hat{\varphi}_{T,l,m}| \leq 1$, the second order risk can be evaluated as follows:

\[
\mathcal{R}_T(\hat{f}_T, f_S) \leq \sum_{l,m \in \mathbb{Z}, l \neq 0} \left( |\hat{\varphi}_{T,l,m} - 1|^2 |\varphi_{S,l,m} - q_{l,m}(f_S)|^2 - \frac{8\delta_t^2 f_S(a_m)(1 - |\hat{\varphi}_{T,l,m}|^2)}{T \pi l^2} \right) + C T^{-\frac{2k+1/4}{2k+1}}.
\]

Recall that the function $f_S(\cdot)$ belongs to $\Sigma_*(k, R)$. Acting like in (4.2), we get

\[
(\pi \delta_T^{-1})^{2k} \sum_{l,m \in \mathbb{Z}} l^{2k} |\varphi_{S,l,m} - q_{l,m}(f_S) - \varphi_{*,l,m} + q_{l,m}(f_*)|^2 \leq R.
\]

Since the central function $f_*$ is smoother than $f_S$, we have

\[
\sum_{l,m \in \mathbb{Z}} l^{2k+2} |\varphi_{*,l,m} - q_{l,m}(f_*)|^2 < C < +\infty,
\]

where the constant $C$ does not depend on $T$. If we use this inequality and the convergence $\hat{\nu}_T \to \infty$, we obtain

\[
\sum_{l,m \in \mathbb{Z}} |1 - \hat{\varphi}_{T,l,m}|^2 |\varphi_{S,l,m} - q_{l,m}(f_S)|^2 \leq \hat{\nu}_T^{-2k} \sum_{l,m \in \mathbb{Z}} l^{2k} |\varphi_{S,l,m} - q_{l,m}(f_S) - \varphi_{*,l,m} + q_{l,m}(f_*)|^2 (1 + o_T(1)) \leq (\hat{\nu}_T^{-1})^{-2k} R (1 + o_T(1)).
\]

Since the coefficients $\hat{\varphi}_{T,l,m}$ do not depend on $m$, we shall omit the index $m$. We denote by $\mathbb{Z}_*$ the set of all integers different from zero. Using the fact that the integral of $f_S(\cdot)$
is equal to one and the convergence of partial sums, we get
\[ R_T(\hat{f}_T, f_S) \leq \frac{R_\delta^2 T}{(\pi^2)^{k/2}} \left( 1 + o_T(1) \right) - \sum_{l \in \mathbb{Z}} \frac{4\delta_T \left( 1 - |\hat{\varphi}_{T,l}|^2 \right)}{\pi^2 T l^2}. \]

It is easy to check that
\[ \sum_{l \in \mathbb{Z}} \frac{1 - |\hat{\varphi}_{T,l}|^2}{l^2} \sim \sum_{|l| \leq \nu} \frac{1}{l^2} \left( \frac{2}{\nu} \frac{|l|^k}{\nu^k} - \left| \frac{l}{\nu} \right|^{2k} \right) + \sum_{|l| > \nu} \frac{1}{l^2} \]
\[ \sim \frac{2}{\nu} \int_0^1 (2x^{k-2} - x^{2k-2}) \, dx + \frac{2}{\nu} \int_1^{\infty} x^{-2} \, dx \]
\[ = \frac{2}{\nu} \left( \frac{2}{k - 1} - \frac{1}{2k - 1} + 1 \right) = 4k^2 \frac{k^2}{\nu(k - 1)(2k - 1)}. \]

Above \( \alpha_T \sim b_T \) means \( \alpha_T = b_T \left( 1 + o_T(1) \right) \). We are able now to explain the choice of \( \hat{\nu} \); it is simply the value minimising the function
\[ G(\nu^{-1}) = \frac{R_\delta^2 T \nu^{-2k}}{\pi^{2k}} - \frac{16k^2 \delta_T \nu^{-1}}{\pi^2 T(k - 1)(2k - 1)}. \]

The substitution of \( \hat{\nu}_T \) by its value leads to the desired bound for the second order risk of our estimator:
\[ R_T(\hat{f}_T, f_S) \leq \left( 1 + o_T(1) \right) T^{-\frac{2k}{k-1}} \hat{\Pi}(k, R). \]

This completes the proof of the theorem. \( \square \)

**Remark 4.3** We supposed in all statements and proofs that the diffusion coefficient \( \sigma \) is identically 1, but this condition can be relaxed. There are two possible approaches for obtaining optimal constants in this case:

- Either one considers the weighted (second order) \( L^2 \)-risk

\[ E_S \int_R (\tilde{f}_T(x) - f_S(x))^2 \sigma^2(x) \, dx - \frac{4}{T} \int_R f_S^2(x) \sigma^2(x) \, dx \]
and defines the weighted Sobolev ball \( \left\{ \int (f_S^{(k)}(x) - f_\star^{(k)})^2 \sigma^2 \leq R \right\} \) as the parameter space. Then the optimal constant remains unchanged.

- Or one refuses to change the risk definition but accepts to work in local minimax settings, that is when the estimated function belongs to a narrowing vicinity of an unknown (but fixed) function \( f_\star \). In that case the optimal constant depends on the central function \( f_\star \) and has the following form:

\[ \hat{\Pi}(f_\star, k, R) = 2(2k - 1) \left( \int_R \frac{f_\star(x)}{\sigma^2(x)} \, dx \right)^{\frac{2k}{2k-1}} \left( \frac{4k}{\pi(k - 1)(2k - 1)} \right)^{\frac{2k}{2k-1}} R^{\frac{1}{2k-1}}. \]
Remark 4.4 The optimal constant $\Pi(k, R)$ we obtained in this work contains a factor 2 (therefore is twice smaller) which is absent in the second order optimal constant in the problem of distribution function estimation from i.i.d. observations (cf. Golubev and Levit [17]). In some sense, it reveals that the estimation problem in our settings is easier than in the settings of [17]. This phenomenon has a simple explanation.

In the definition of the parameter set, we required that the function $S$ satisfies the inequality $\text{sgn}(x)S(x) \leq \gamma_*$ for $x \notin [-A, A]$. This condition excludes automatically the density functions $f_S$ possessing heavy tails, moreover, it implies that these densities decrease exponentially fast at infinity. In particular, the densities close to those of uniform distributions over a large interval do not belong to our nonparametric class. But exactly these densities are the most difficult to estimate, because the observations distributed following these laws are quite dispersed on the whole real line and contain very few information about the local variation of the underlying density function.

Remark 4.5 We estimated the infinite matrix $\varphi_S = \{\varphi_{S,l,m}\}_{l,m \in \mathbb{Z}}$ of the Fourier coefficients of $f_S$ by the infinite matrix $\tilde{\varphi}_T = \{\tilde{\varphi}_{T,l,m}\}_{l,m \in \mathbb{Z}}$. In reality, this last matrix contains only a finite number of elements different from null. Indeed, for large values of $m$, the intervals $I_m$ contain no observation and corresponding coefficients $\tilde{\varphi}_{T,l,m}$ are therefore equal to zero. So we used here a kind of soft thresholding procedure.

5 Proofs of technical lemmas

Lemma 5.1 In the notation of Section 3, for any integer $l$ different from 0, we have

$$E[\Psi_{l,m,\vartheta}(\xi)\chi_{\{\xi \notin I_m\}}] \leq \frac{CA^3}{l^2T^3}. $$

Proof: We denote $\delta = A/T^3$ and divide the indicator into two parts

$$E[\Psi_{l,m,\vartheta}(\xi)\chi_{\{\xi \notin I_m\}}] = E[\Psi_{l,m,\vartheta}(\xi)\chi_{\{\xi > a_m + \delta\}}] + E[\Psi_{l,m,\vartheta}(\xi)\chi_{\{\xi < a_m - \delta\}}]. $$

The first term can be evaluated as follows:

$$E[\Psi_{l,m,\vartheta}(\xi)\chi_{\{\xi > a_m + \delta\}}] = E\left(\int_{I_m} f_{\vartheta}(x)e_{l,m}(x)\,dx\right)^2 \left(\frac{1 - F_{\vartheta}(\xi)}{f_{\vartheta}(\xi)}\chi_{\{\xi > a_m + \delta\}}\right)^2 \leq CEQ\left(\int_{I_m} f_{\vartheta}(x)e_{l,m}(x)\,dx\right)^2, $$

since one can prove exactly like in [9], that $f_{\vartheta}(x) = f_*(x)(1 + o_T(1))$ and

$$\int_{a_m + \delta}^{\infty} \left(\frac{1 - F_*(y)}{f_*(y)}\right)^2 f_*(y)\,dy < CAf_*^{-1}(a_m + \delta). $$
The first multiplier will be evaluated using the integration by parts formula:

\[
\int_{I_m} f_\vartheta(x) e_{l,m}(x) \, dx = \frac{\delta}{\pi l} \left[ (f_\vartheta(a_m + \delta) - f_\vartheta(a_m - \delta)) e_{-l,m}(a_m - \delta) \right. \\
\left. - \int_{I_m} f'_\vartheta(x) e_{-l,m}(x) \, dx \right]
\]

Using the fact that \(f_\vartheta(x)\) is differentiable, its derivative is \(2S_\vartheta(x) f_\vartheta(x)\) and that \(S_\vartheta(\cdot)\) is bounded by \(2A\) on the interval \([-A, A]\), for any \(\vartheta \in \Theta_T\), we get

\[
\left| \int_{I_m} f_\vartheta(x) e_{l,m}(x) \, dx \right| \leq \frac{CA\delta^2}{l^2} f_*^2(a_m).
\]

Combining these estimates we obtain the desired result for the first term. The term involving the indicator of the event \(\xi < a_m - \delta\) can be evaluated analogously. This completes the proof of the lemma. \(\square\)

**Lemma 5.2** If \(l \neq 0\) and \(y \in I_m\), then the following representation holds

\[
\Psi_{l,m,\vartheta}(y) = \frac{A(1 + \alpha_T(1))}{\pi l T^3} \left( e_{-l,m}(y) - e_{-l,m}(a_m - AT^{-\beta}) \right) + \frac{A\Phi_{l,m,\vartheta}(y)}{\pi l T^3},
\]

where the sequence \(\Phi_{l,m,\vartheta}\) is defined as follows

\[
\Phi_{l,m,\vartheta}(y) = \int_{I_m} f'_\vartheta(x) \left( \frac{\chi_{\{y > x\}} - F_\vartheta(y)}{f_\vartheta(y)} \right) e_{-l,m}(x) \, dx.
\]

**Proof:** The proof of this lemma is quite close to the proof of the preceding one. Using the integration by parts formula one gets

\[
\Psi_{l,m,\vartheta}(y) = \frac{\delta}{\pi lf_\vartheta(y)} \left( f_\vartheta(y) e_{-l,m}(y) - f_\vartheta(a_m - \delta) e_{-l,m}(a_m - \delta) \right) \\
- \frac{\delta F_\vartheta(y)}{\pi lf_\vartheta(y)} \left( f_\vartheta(a_m + \delta) - f_\vartheta(a_m) \right) e_{-l,m}(a_m - \delta) \\
- \frac{\delta}{\pi l} \int_{I_m} f'_\vartheta(x) \left( \frac{\chi_{\{y > x\}} - F_\vartheta(y)}{f_\vartheta(y)} \right) e_{-l,m}(x) \, dx.
\]

To finish the proof, it remains to remark that on the interval \(I_m\), one has \(|f_\vartheta(y) - f_\vartheta(a_m - \delta)| \leq C\delta|f'_\vartheta(y)| \leq CA\delta|f_\vartheta(y)|\), since \(f'_\vartheta(x) = 2S_\vartheta(x) f_\vartheta(x)\) and \(S_\vartheta(x)\) is bounded by \(1 + |x|\). \(\square\)

**Corollary 5.3** For any \(l\) different from 0, we define the number \(c_l\) to be equal to 1, if \(l < 0\), and to 3, if \(l > 0\). Then we have

\[
\mathbb{E} \left[ \Psi_{l,m,\vartheta}^2(\xi) \right] \leq \frac{A^2 f_*^2(a_m)(1 + \alpha_T(1))}{\pi^2 l^2 T^{2\beta}} \left( \sqrt{c_l} + \sqrt{\mathbb{E} \left[ \Phi_{l,m,\vartheta}^2(\xi) \right]} \right)^2 + \frac{CA^3}{l^2 T^{3\beta}},
\]
Second order minimax density estimation

and (see [9, Appendix], for the proof)

$$\sum_{l, m \in \mathbb{Z}} E \left[ \Phi^2_{l, m, \vartheta}(\xi) \right] \leq C.$$

**Lemma 5.4** The residual terms $\mu_{l,m,T}$ defined in (3.7) satisfy $|\mu_{l,m,T}| \leq C \delta^3 T^{-1}$ where $C$ is a constant.

**Proof:** The proof of this lemma is rather technical. We give below the main ideas of the proof and leave the technical details to the reader. First of all, due to Lemma 5.1, it is enough to prove this inequality for $\bar{\mu}_{l,m,T} = \left( E_{\mathcal{Q}} \frac{\partial \psi_{l,m,\vartheta}}{\partial \vartheta} \right)^2 - 4 E_{\mathcal{Q}} \left[ I_{l,m}(\vartheta) \right] E \left[ \Psi_{l,m,\vartheta}(\xi) \chi_{\{\xi \in I_m\}} \right]$

Next, it can be shown (see Kutoyants [22]) that

$$\frac{\partial \psi_{l,m,\vartheta}}{\partial \vartheta} = 2 \sqrt{\frac{2A}{f_s(a_m) T^2}} \int_{I_m} g_{l,m}(x) \Psi_{l,m,\vartheta}(x) f_\vartheta(x) \, dx.$$  (5.1)

Since all the functions under the integral are differentiable with bounded derivative and the length of the interval $I_m$ is $2 \delta_T$, we have

$$\frac{\partial \psi_{l,m,\vartheta}}{\partial \vartheta} = \sqrt{\frac{32 \delta^3_T}{f_s(a_m)}} g_{l,m}(a_m) \Psi_{l,m,\vartheta}(a_m) f_\vartheta(a_m) + \delta^3_T O(1).$$

Using the same arguments, one can check that

$$I_{l,m}(\vartheta) E_{\vartheta} \left[ \Psi_{l,m,\vartheta}(\xi) \chi_{\{\xi \in I_m\}} \right] = \frac{2 \delta_T}{f_s(a_m)} E_{\vartheta} \left[ g^2_{l,m}(\xi) \right] E_{\vartheta} \left[ \Psi^2_{l,m,\vartheta}(\xi) \chi_{\{\xi \in I_m\}} \right]$$

$$= \frac{8 \delta^3_T}{f_s(a_m)} g^2_{l,m}(a_m) \Psi^2_{l,m,\vartheta}(a_m) f_\vartheta(a_m)^2 + O_T(\delta^4_T).$$  (5.2)

It is clear now that the difference of the terms (5.1) and (5.2) is of order $\delta^4_T$. Since the denominator of $\bar{\mu}_{l,m,T}$ is larger than $C T \delta_T$, we get the desired inequality.

**Lemma 5.5** In the notation of Section 4, we have

$$E_S \left( \tilde{f}^{(j)}_T(a) - f^{(j)}_S(a) \right)^2 \leq C h^{-2j} (1 + |a|)^{\nu^*} f_S(a),$$  (5.3)

$$E_S \left( \tilde{q}_{T,l,m} - q_{l,m}(f_S) \right)^2 \leq C h^{-2j+1/2} \int_{I_m} (1 + |a|)^{\nu^*} f_S(a) \, da \sum_{j=0}^{k-1} \left( \frac{\hat{\nu}_T}{T} \right)^{2j+2} \hat{\nu}_T^{-1}. \quad (5.4)$$
We apply now the Taylor formula with the rest term in the integral form and the fact that $E$.

Let us introduce the functions $g$ since

\[ E[X] \text{ bounded via the Itô formula exactly like in Kutoyants (1998).} \]

The bias-variance decomposition yields

\[ E_S(f_T^{(j)}(a) - f_S^{(j)}(a))^2 \leq 2(E_S[f_T^{(j)}(a)] - f_S^{(j)}(a))^2 + 2E_S(f_T^{(j)}(a) - E_S[f_T^{(j)}(a)])^2. \]

In view of stationarity of $X$ and the equality $\int Q(u)du = 1$, we get

\[ E_S[f_T^{(j)}(a)] - f_S^{(j)}(a) = \int_{\mathbb{R}} Q(u)[f_S^{(j)}(a + uh_T) - f_S^{(j)}(a)] du. \]

We apply now the Taylor formula with the rest term in the integral form and the fact that $Q$ is a kernel of order $k$ (i.e. $\int u^l Q(u)du = 0$ for any integer $l \in [1, k]$):

\[ E_S[f_T^{(j)}(a)] - f_S^{(j)}(a) = \frac{1}{(k-j-1)!} \int_{\mathbb{R}} Q(u) \int_0^{uh_T} y^{k-j-1} f_S^{(k)}(a + uh_T - y) dy du. \]

Since $Q(u) = 0$ if $|u| > 1$, we obtain

\[ |E_S[f_T^{(j)}(a)] - f_S^{(j)}(a)| \leq \frac{h_T^{k-j-1}}{(k-j-1)!} \int_{-h_T}^{h_T} |f_S^{(k)}(a + y)| dy. \]

The Cauchy–Schwarz inequality yields

\[ |E_S[f_T^{(j)}(a)] - f_S^{(j)}(a)|^2 \leq \frac{h_T^{2k-2j-1}}{[(k-j-1)!]^2} \int_{-h_T}^{h_T} [f_S^{(k)}(a + y)]^2 dy. \hspace{1cm} (5.5) \]

We turn to the evaluation of the variance. We have

\[ f_T^{(j)}(a) - E_S[f_T^{(j)}(a)] = \frac{1}{Th_T^{k+j}} \int_0^T \left[ Q^{(j)}(X_t - a) h_T - E_S \left[ Q^{(j)} \left( \frac{X_t - a}{h_T} \right) \right] \right] dt. \]

Let us introduce the functions $g(u)$ and $H(u)$ as follows:

\[ g(u) = Q^{(j)} \left( \frac{u - a}{h_T} \right) - E_S \left[ Q^{(j)} \left( \frac{X_0 - a}{h_T} \right) \right], \]

\[ H(u) = \int_0^u \frac{2}{f_S(y)} \int_{-\infty}^{y} g(v) f_S(v) dv dy. \]

It can be easily checked that the Itô formula implies

\[ \int_0^T g(X_t) dt = H(X_T) - H(X_0) - \int_0^T H'(X_t) dW_t, \]

and therefore, due to stationarity,

\[ E_S \left( \int_0^T g(X_t) dt \right)^2 \leq 6E_S[H^2(X_0)] + 3TE_S[(H'(X_0))^2] \]
By integrating by parts, we get

\[ H'(u) = 2 \int_{-\infty}^{u} g(y) \, dy - \frac{2}{f_S(u)} \int_{-\infty}^{u} \int_{-\infty}^{y} g(v) \, dv \, f'_S(y) \, dy \]

\[ = 2 \int_{-\infty}^{u} Q^{(j)} \left( \frac{y - a}{h_T} \right) \, dy \]

\[ - \frac{2}{f_S(u)} \int_{\mathbb{R}} \int_{-\infty}^{y} Q^{(j)} \left( \frac{v - a}{h_T} \right) \, dv \, f'_S(y) \left( \chi_{\{ y < a \}} - F_S(u) \right) \, dy \]

\[ = 2h_T Q^{(j-1)} \left( \frac{u - a}{h_T} \right) \]

\[ - \frac{2h_T}{f_S(u)} \int_{\mathbb{R}} Q^{(j-1)} \left( \frac{y - a}{h_T} \right) f'_S(y) \left( \chi_{\{ y < a \}} - F_S(u) \right) \, dy \quad (5.6) \]

Using the inequality

\[ \frac{\chi_{\{ y < a \}} - F_S(u)}{f_S(u)} \leq C + \frac{C \chi_{\{ u \in [0, y] \}}}{\sqrt{f_S(u)}} \leq C + \frac{C \chi_{\{ u \in [0, y] \}}}{\sqrt{f_S(u)}}. \]

the equality \( f'_S(x) = S(x) f'_S(x) \) and the fact that \( S(\cdot) \) increases at most like a polynomial, one can prove that the second term in (5.6) is asymptotically negligible with respect to the first one. Therefore,

\[ E_S \left[ (H'(X_0))^2 \right] \leq 4(1 + o_T(1)) E_S \left[ \left( h_T Q^{(j-1)} \left( \frac{X_0 - a}{h_T} \right) \right)^2 \right] \]

\[ \leq 4(1 + o_T(1)) h_T^2 \int_{-1}^{1} \left[ Q^{(j-1)}(u) \right]^2 f_S(a + u h_T) \, du \]

\[ \leq C h_T^3 \int_{a-h_T}^{a+h_T} f_S(x) \, dx. \]

It is evident that the term involving the expectation of \( H^2(X_0) \) is asymptotically much smaller than the term \( T E_S [H'(X_0)^2] \). Consequently,

\[ E_S \left[ \tilde{f}_{ij}^j(a) - f_{ij}^j(a) \right]^2 \]

\[ \leq C h_T^{2k-2j-1} \int_{a-h_T}^{a+h_T} \left[ f_S^{(k)}(y) \right]^2 \, dy + \frac{C}{T h_T^{2j-1}} \int_{a-h_T}^{a+h_T} f_S(x) \, dx \]

\[ \leq C h_T^{2k-2j-1} \int_{a-h_T}^{a+h_T} f_S(x) \, dx \leq C h_T^{2k-2j} (1 + |a|)^{\nu} f_S(a). \]

Therefore, using the equality

\[ q_{T,l,n} - q_{l,n} (f_S) = \frac{(-1)^l}{\sqrt{2 \delta_T}} \sum_{j=0}^{k-1} \left( \frac{\delta_T}{\nu \pi l} \right)^{j+1} \int_{l}^{j+1} (\tilde{f}_{ij}^j(a) - f_{ij}^j(a)) \, dx \]
and the Cauchy–Schwarz inequality, we get

\[
E_S |\hat{q}_{T,l,m} - q_{l,m}(f_S)|^2 \\
\leq \frac{k}{2\delta_T} \sum_{j=0}^{k-1} \left( \frac{\delta_T}{\pi l} \right)^{2j+2} E_S \left( \int_{I_m} (f^{(j+1)}_T - f^{(j+1)}_S(a)) \, dx \right)^2 \\
\leq C \frac{k}{\delta_T} \sum_{j=0}^{k-1} \left( \frac{\delta_T}{T} \right)^{2j+2} h_T^{2j-2j(1 + |a_m| - \delta)} f_S(a_m - \delta) \\
\leq Ch_T^{2k+1} \delta_T^{-1} \int_{I_m} (1 + |a|)^{\nu_\ast} f_S(a) \, da \sum_{j=0}^{k-1} \left( \frac{\hat{\nu}_T}{T} \right)^{2j+2} \hat{\nu}_T^{-1}.
\]

In the last inequality we have used the fact that the step of localisation \( \delta_T \) is equal to the square root of \( h_T \) and that \( \hat{\nu}_T \) is of order \( \delta_T/h_T \).

\[\square\]

**References**


Second order minimax density estimation


Arnak S- Dalalyan  
Laboratoire de Probabilités  
Université Paris 6, Boîte courrier 188  
75252 Paris cedex 05, France  
dalalyan@ccr.jussieu.fr

Yu.A. Kutoyants  
Laboratoire de Statistique et Processus  
Université du Maine, Av. O. Messiaen  
72085 Le Mans cedex 9, France  
kutoyants@univ-lemans.fr