Thèse de Doctorat
de l'Université du Maine

Spécialité : **Mathématiques**
Option : **Statistique**

Sujet de la thèse :

Estimation non-paramétrique asymptotiquement efficace pour des processus de diffusion ergodiques

Présentée par

Arnak Dalalyan

pour obtenir le grade de Docteur de l'Université du Maine

Soutenue le 19 décembre 2001

La composition du jury :

Président du jury	:	J. Mémin (Université de Rennes I)
Directeur de thèse	:	Y. A. Koutoyants (Université du Maine)
Rapporteur	:	G. K. Golubev (Université d'Aix-Marseille I)
Rapporteur	:	C. Larédo (INRA)
Examinateur	:	D. Bosq (Université Paris VI)
Examinateur	:	P. Pilibossian (Université Paris VI)

Remerciements

Mes remerciements s'adressent tout d'abord à Youri A. Koutoyants, Professeur à l'Université du Maine, qui est à l'origine de ce travail. Il m'est difficile de résumer en quelques mots tout ce qu'il m'a apporté par son implication tout au long de ma thèse, tant sur le plan scientifique que relationnel. Ses précieux conseils et son sens de l'humour m'ont toujours aidé dans les situations difficiles.

Je souhaite témoigner ma profonde gratitude à Boris S. Nahapetyan, Directeur de Recherche à l'Académie Nationale des Sciences d'Arménie, qui très tôt a su me faire partager son intérêt pour la théorie des probabilités. Je lui suis reconnaissant d'avoir cru en moi et de m'avoir transmis sa passion pour la recherche.

Je tiens à exprimer ma reconnaissance à Georgii Golubev, Directeur de Recherche au CNRS, et Catherine Larédo, Directeur de Recherche à l'INRA, de l'intérêt qu'ils ont manifesté à l'égard de ma thèse en acceptant d'être rapporteurs.
J'adresse aussi mes vifs remerciements aux autres membres du jury : Jean Mémin, Professeur à L'Université Rennes 1, Denis Bosq, Professeur à l'Université Paris 6 et Philippe Pilibossian, Maître de conférences à l'Université Paris 6.

Je remercie également l'ensemble du personnel du département de Mathématiques de l'Université du Maine, de m'avoir accueilli pendant ces trois années, aidé dans les tâches administratives et guidé sur le plan pédagogique.

Un grand merci à mes amis Hayk, Grigor, Lucinée, Serguey, Eduard et Karina qui ont su être là aux bons moments et m'apporter un soutien moral. Leur amitié m'est très précieuse et que ces quelques lignes leur expriment mon sincère et profond attachement.

Enfin, c'est du fond du cœur que je remercie mes frères et mes cousins : ils ont été, sont et resteront les piliers qui me soutiennent sans limite. Je réserve le plus grand des mercis à mes parents. C'est le plus chaleureusement possible que je les remercie d'être tout simplement ce qu'ils sont.

Contents

Introduction

Les processus de diffusion font actuellement partie de l'ensemble des objets les plus étudiés par les mathématiciens. Cela s'explique principalement par leur application dans différentes disciplines. En effet, ce sont les outils utilisés en physique pour décrire les systèmes dynamiques perturbés par un bruit aléatoire; ils revêtent également une importance particulière dans le champs théorique des mathématiques financières.

En ce qui concerne les mathématiques, les processus de diffusion suscitent un grand intérêt théorique car ils sont à la fois des semi-martingales et des processus de Markov. De fait, la propriété de Markov permet de classifier les diffusions en trois catégories : les transientes, les nul-récurrentes et les récurrentes positives. Ces dernières sont dotées de la propriété de revenir "suffisamment vite" dans le voisinage du point de départ, et ce quel que soit le point de départ et le voisinage en question. Ce comportement des trajectoires des processus de diffusion est important d'un point de vue statistique, car il implique que le nombre de passages du processus X par un point (arbitraire) a est grand, si le temps d'observation T est suffisamment long. Par conséquent, en augmentant le temps T, nous pourrons estimer la valeur du paramètre inconnu (par exemple, de la fonction de répartition invariante ou de sa densité) au point a de manière de plus en plus précise. Soulignons que pour les processus de diffusion la propriété de la récurrence positive est équivalente à l'ergodicité.

Afin de décrire de façon plus précise le sujet de notre étude, nous tenons à rappeler que les processus de diffusion peuvent être donnés en tant que solutions d'Équations Différentielles Stochastiques (EDS) homogènes

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

où W est un mouvement Brownien standard et ξ une valeur initiale éventuellement aléatoire, indépendante de W. Les fonctions $S(\cdot)$ et $\sigma^2(\cdot)$ sont appelées dérive (ou coefficient de dérive) et coefficient de diffusion. En mathématiques financières, ces fonctions sont interprétées comme étant la tendance et la volatilité de l'évolution du cours d'une option financière, ce cours étant modélisé par le processus X.

Cette thèse traite de certains problèmes de statistique non-paramétrique liés aux processus de diffusion ergodiques. Aussi, afin que nous puissions parler d'un processus de diffusion associé à la dérive S et au coefficient de diffusion σ^2, il faut que l'EDS correspondant à ces fonctions ait une solution. C'est pourquoi, dans le Chapitre 1 nous présentons les résultats principaux concernant l'existence, l'unicité et certaines propriétés probabilistes des solutions des EDS. Les deux premières sections de ce chapitre ont pour but d'introduire l'intégration stochastique au sens d'Itô, qui représente un outil indispensable pour travailler avec les processus de diffusion.

Notons ici que les résultats obtenus dans ce mémoire de thèse s'inscrivent dans le cadre des recherches entreprises par Pinsker (cf. [**50**]). Nous commentons dans le Chapitre 2 les différents aspects du théorème de Pinsker et présentons une liste non-exhaustive des modèles où des résultats similaires ont été obtenus. Dans nos démonstrations, nous suivons essentiellement les travaux de Golubev [**23**], [**24**] et de Golubev et Levit [**28**].

L'un des problèmes classiques de la statistique est celui de l'identification du modèle. Dans le cadre des processus de diffusion, ce problème peut être énoncé de la façon suivante : on observe une trajectoire d'un processus de diffusion sur l'intervalle du temps $[0, T]$ et on souhaite déterminer la loi de ce processus, ou, plus exactement, en trouver une loi approchée. Ce problème nous amène directement à l'étude des estimateurs de la dérive et du coefficient de diffusion, car ce sont les seuls paramètres régissant le processus X. Étudier le comportement des estimateurs de la dérive, est l'objectif principal de ce travail de recherche. Pour cela, on suppose que le coefficient de diffusion σ^2 est connu, ce qui n'est pas une condition très contraignante car, en utilisant la variation quadratique de X, la fonction $\sigma^2(\cdot)$ peut être estimée sans erreur.

La méthode que nous utilisons pour estimer la dérive d'une diffusion ergodique repose sur la formule $f_S'(x) = 2S(x)\, f_S(x)$, où f_S et f_S' désignent la densité invariante et sa dérivée. Cette formule suggère d'utiliser les estimateurs de f_S et de f_S' dans la construction de l'estimateur de S. C'est ainsi que l'on arrive au second problème non-paramétrique considéré dans ce travail de recherche. Il consiste à estimer la dérivée f_S' de la densité invariante, en utilisant une trajectoire de X observée jusqu'à l'instant T. Notons que ce problème a déjà été considéré par d'autres auteurs (Pham, Prakasa Rao, van Zenten) qui ont étudié le comportement asymptotique des estimateurs à noyau. Ils ont obtenu, sous réserve de certaines conditions de régularité, la vitesse de convergence de ces estimateurs sans prouver son optimalité.

Dans les Chapitres 3 et 4, nous nous intéressons à une propriété beaucoup plus fine des estimateurs de la dérivée que celles démontrées par les auteurs précités. En effet, nous étudions notamment l'efficacité asymptotique à la constante près dans la classe de tous les estimateurs possibles. L'asymptotique considérée est celle des grands échantillons, c'est-à-dire lorsque le temps d'observation tend vers l'infini. Nous établissons, dans un premier temps, une borne inférieure du risque minimax pour des espaces de paramètres convenablement choisis (essentiellement les boules de l'espace de Sobolev). Dans un deuxième temps, nous prouvons que le bon choix du noyau et de bandwidth conduit à un estimateur qui atteint asymptotiquement cette borne inférieure. Les estimateurs possédant cette propriété sont appelés asymptotiquement efficaces ou asymptotiquement minimax. En d'autres termes, nous construisons un estimateur de la dérivée de la densité invariante qui est presque aussi performant que l'estimateur idéal (minimax), dans le cas où le temps d'observation est suffisamment long.
De plus, nous montrons que la même méthode peut être appliquée afin d'estimer des dérivées d'ordre supérieur de la densité invariante. Nous prouvons également que ces résultats restent vrais pour des espaces de paramètres plus généraux, analogues aux ellipsoïdes de l^2. Enfin, dans le Chapitre 4 nous développons l'approche localement minimax et démontrons qu'une légère modification de l'estimateur considéré au Chapitre 3 nous permet de construire un estimateur localement asymptotiquement efficace.

Les méthodes exposées dans le Chapitre 4, nous donnent enfin la possibilité – au Chapitre 5 – de réaliser notre objectif initial : estimer la dérive de façon efficace. C'est la notion d'efficacité locale qui s'avère être la mieux adaptée à ce problème. En effet, en utilisant l'estimateur de la dérivée étudié au Chapitre 4 et un estimateur de la densité invariante introduit par Kutoyants (cf. [**36**]), nous construisons un estimateur asymptotiquement efficace de la dérive.

Les résultats des Chapitres 3, 4 et 5 font l'objet de deux articles actuellement soumis à la publication dans les revues *Statistical Inference for Stochastic Processes* [**5**] et *Mathematical Methods of Statistics* [**6**].

Notations

a_+	The positive part of $a \in \mathbb{R}$, $a_+ = \max(0, a)$
$[a]$	The integer part of $a \in \mathbb{R}$.
$a \vee b$	The maximum of real numbers a and b.
$a \wedge b$	The minimum of real numbers a and b.
$b_\varepsilon = o_\varepsilon(a_\varepsilon)$, $\varepsilon \to \varepsilon_0$	The sequence b_ε is asymptotically smaller than a_ε, *i. e.*, $b_\varepsilon / a_\varepsilon \to 0$ as $\varepsilon \to \varepsilon_0$.
$b_\varepsilon = O_\varepsilon(a_\varepsilon)$, $\varepsilon \to \varepsilon_0$	The sequence b_ε is asymptotically bounded by the sequence a_ε, *i. e.*, $|b_\varepsilon| \leq C|a_\varepsilon|$, $\forall \varepsilon$.
$b_\varepsilon \sim a_\varepsilon$, $\varepsilon \to \varepsilon_0$	The sequences b_ε and a_ε are asymptotically equivalent, *i. e.*, $b_\varepsilon / a_\varepsilon \to 1$ as $\varepsilon \to \varepsilon_0$.
$\mathrm{C}[a, b]$, $\mathrm{C}(a, b)$	The set of all continuous functions from $[a, b]$ (resp. (a, b)) to $\mathbb{R}$.
$L^2[a, b]$	The set of all square integrable functions on $[a, b]$.
$\mathbb{1}_A(x)$	The indicator (or characteristic) function of the set $A \subset \mathbb{R}$, *i. e.*, $\mathbb{1}_A(x) = 1$ if $x \in A$, otherwise $\mathbb{1}_A(x) = 0$.
$\mathbf{P}$, $\mathbf{Q}$	Probability measures.
$\mathbf{E}$	The Lebesgue integral w.r.t. $\mathbf{P}$.
$\xi \sim \mathbf{Q}$	The random variable ξ follows the law $\mathbf{Q}$, *i. e.*, the probability of the event $\xi \in A$ is equal to $\mathbf{Q}(A)$.

CHAPTER 1

Preliminaries

In this chapter we introduce the main concepts of stochastic calculus necessary to deal with diffusion processes. The proofs of almost all theorems of this chapter can be found in the book [**51**] of Revuz and Yor. We also set up in this chapter the notations used in this work.

1.1. Brownian Motion and Local Martingales

Throughout this work we are given a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ and $\mathscr{B}(\mathbb{R})$ denotes the σ-algebra of Borel subsets of $\mathbb{R}$. Any collection $X = (X_t)_{t \in A}$ of random variables indexed by a subset A of $\mathbb{R}$ is called *random process*. For each $\omega \in \Omega$ the function $t \mapsto X_t(\omega)$ is called *realization*, *path* or *trajectory* of random process X. In the sequel we consider only the case where $A \subseteq \mathbb{R}_+ = [0, \infty[$ and use the notation $X^T = (X_t)_{t \in [0,T]}$ for each T strictly positive.

Definition 1. *A random process $W = (W_t)_{t \in \mathbb{R}_+}$ is called Brownian Motion or standard Wiener process if it fulfills the following three conditions:*

i) $W_0 = 0$ *a.s.*
ii) *The increments of W are independent, that is for any partition $0 < t_1 < t_2 < \ldots < t_n$ the random variables $W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent.*
iii) *For any couple $t, s \in \mathbb{R}_+$ the random variable $W_t - W_s$ is Gaussian with zero mean and variance $|t - s|$.*

The existence of such a process is ensured by Kolmogorov's theorem. Moreover, using Kolmogorov's continuity criterion (see, for example, [**51**] page 25) one can prove that there exists a modification of this process which is a.s. continuous. From now on, we deal only with this continuous modification. In fact, this modification can be even chosen to be Hölder of order α for any $\alpha \in [0, 1/2[$.

Any nondecreasing family $(\mathscr{F}_t)_{t\geq 0}$ of sub-σ-algebras of $\mathscr{F}$ is called *filtration* and the quadruple $\big(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P}\big)$ is said to be a *filtered probability space*. We say that a random process X is $\mathscr{F}_t$*-adapted* if for any positive t the random variable X_t is $\mathscr{F}_t$-measurable. These definitions permit us to introduce another (more general) definition of Brownian Motion:

Definition 2. *A random process $W = (W_t)_{t\in\mathbb{R}_+}$ with continuous paths is called $\mathscr{F}_t$-Brownian Motion if it is $\mathscr{F}_t$-adapted and fulfills the following two conditions:*

i) $W_0 = 0$ *a.s.*
ii) *For any couple $t, s \in \mathbb{R}_+$, $t > s$ the random variable $W_t - W_s$ is a centered Gaussian with variance equal to $t - s$ and independent of $\mathscr{F}_s$.*

A measurable set A is called negligible, if $\mathbf{P}(A) = 0$. A σ-algebra $\mathscr{G}$ is called complete if all the subsets of a negligible set belong to $\mathscr{G}$, *i. e.*,

$$B \subseteq A \in \mathscr{G}, \quad \mathbf{P}(A) = 0 \qquad \Rightarrow \qquad B \in \mathscr{G}.$$

From now on, we suppose that the σ-algebras $\mathscr{F}$ and $(\mathscr{F}_t)_{t\geq 0}$ are complete. Let X be a random process. We denote by $\mathscr{F}_T^X$ the minimal complete σ-algebra generated by the random variables $\{X_t\}_{t\in[0,T]}$. It is evident that the family $(\mathscr{F}_t^X)_{t\geq 0}$ is a filtration. It is called *natural filtration* of X. One can easily show that a continuous path random process W is a Brownian Motion in the sense of Definition 1, if and only if it is $\mathscr{F}_t^W$ -Brownian Motion.

Definition 3. *A random process $M = (M_t)_{t\geq 0}$ is called $\mathscr{F}_t$-martingale if it is $\mathscr{F}_t$-adapted, $M_t \in L^1$ for any $t \geq 0$, and*

$$\mathbf{E}(M_t \,|\, \mathscr{F}_s) = M_s$$

for any $t > s \geq 0$.

In the sequel we consider only the case of continuous martingales, that is the martingales whose paths are continuous for almost all ω's. If $\mathscr{F}_t = \mathscr{F}_t^M$ then we say simply that M is martingale. Note that for an $\mathscr{F}_t$-Brownian Motion W we have the following obvious equalities:

$$\mathbf{E}(W_t \,|\, \mathscr{F}_s) = W_s + \mathbf{E}(W_t - W_s \,|\, \mathscr{F}_s) = W_s + \mathbf{E}(W_t - W_s) = W_s.$$

Consequently W is an $\mathscr{F}_t$-martingale. A random process M is called *local martingale* if, for any $n \in \mathbb{N}$, the stopped process $M^{(n)} = (M_{t\wedge\tau_n})_{t\geq 0}$ is a martingale, where τ_n is the stopping time defined as $\tau_n = \inf\big\{t \geq 0 \,\big|\, |M_t| \geq n\big\}$.

1.2. Stochastic Integrals. Itô Formula

We introduce in this section the notion of stochastic integration with respect to a continuous martingale. Namely, for any random process H satisfying some measurability and integrability conditions and any martingale M, we define the process

$$\left(\int_0^t H_s\,dM_s\right)_{t\geq 0}. \tag{1.1}$$

Remark that this integral can not be defined by a path by path procedure as a Stiltjes integral, because in general a martingale is not almost surely of finite variation. Moreover, a martingale is a.s. of finite variation if and only if it is a.s. constant. Nevertheless, there is a global method which is used in order to define (1.1). For a detailed exposition of the theory of stochastic integration the reader is referred to the book of Revuz and Yor [**51**].

Let us give now some definitions. A random process A is called increasing (resp. of finite variation) if it is $\mathscr{F}_t$-adapted, $A_0 = 0$ a.s., and the paths $A_.(\omega)$ are continuous and increasing (resp. of finite variation) for almost all $\omega \in \Omega$. A random process H is *progressively measurable*, if for any positive t, the mapping $X : [0,t]\times\Omega \longrightarrow \mathbb{R}$ is $\mathscr{B}([0,t])\otimes\mathscr{F}_t$-measurable. It is clear that for any random process H, bounded on every interval $[0,t]$ for almost all ω's, and for any process of bounded variation A, one can define

$$\int_0^t H_s(\omega)\,dA_s(\omega) \tag{1.2}$$

as a Stiltjes integral. If ω is in the set where $A_.(\omega)$ is not of finite variation or $H_.(\omega)$ is not locally integrable with respect to $dA_s(\omega)$, we set $\int_0^t H_s(\omega)\,dA_s(\omega) = 0$.

THEOREM 1.1. *For any continuous local martingale $M = (M_t)_{t\geq 0}$ there exists a unique increasing process denoted by $(\langle M\rangle_t)_{t\geq 0}$, such that the process $\big(M_t^2 - \langle M\rangle_t\big)_{t\geq 0}$ is a continuous local martingale. Moreover, for any positive number T, if $0 = t_0^n < t_1^n < \ldots < t_{k_n}^n = T$ is a sequence of embedded subdivisions of $[0,T]$ such that $\sup_{i=1,\ldots,k_n}(t_i^n - t_{i-1}^n) \to 0$ as $n\to\infty$, then*

$$\langle M\rangle_T = \lim_{n\to\infty}\sum_{i=1}^{k_n}(M_{t_i^n} - M_{t_{i-1}^n})^2$$

in the sense of the convergence in probability. The random process $\langle M\rangle$ is called quadratic variation of M.

If the martingale M vanishes at 0 and, for a $T>0$, the random variable M_T is squared integrable, then

$$\mathbf{E}[M_t^2] = \mathbf{E}[\langle M\rangle_t], \qquad \text{for any} \quad t \in [0,T].$$

In the sequel we will need the following generalization of this property:

THEOREM 1.2 (BDG inequality). *For every $p>0$, there exists a constant C_p such that, for any continuous martingale M vanishing at zero,*

$$\mathbf{E}[M_t^{2p}] \le C_p \mathbf{E}[\langle M\rangle_t^p], \qquad \textit{for any} \quad t \ge 0. \tag{1.3}$$

The presentation of the construction of the stochastic integral is out of scope of this section. We want only to precise here that for any local martingale M and a progressively measurable process H such that, for almost all ω's,

$$\sup_{s\le t} H_s < \infty, \qquad \forall t \ge 0, \tag{1.4}$$

one can define an $\mathscr{F}_t$-adapted process denoted by $\Big(\int_0^t H_s\,dM_s\Big)_{t\ge0}$ and satisfying the following properties:

1. The application $(H,M)\mapsto \int_0^{\cdot} H_s\,dM_s$ is bilinear.
2. The process $\int_0^{\cdot} H_s\,dM_s$ is a local martingale and its quadratic variation is $\langle \int_0^{\cdot} H_s\,dM_s\rangle_t = \int_0^t H_s^2\,d\langle M\rangle_s$.
3. If H and K are two locally bounded processes, then, for any positive T, $\int_0^T H_t\,d\Big(\int_0^t K_s\,dM_s\Big) = \int_0^T H_tK_t\,dM_t$.
4. If $\mathbf{E}\Big[\int_0^T H_s^2\,d\langle M\rangle_s\Big] < \infty$, then for any $t\in[0,T]$, we have

$$\mathbf{E}\Big[\int_0^t H_s\,dM_s\Big] = 0, \quad \mathbf{E}\Big[\Big(\int_0^t H_s\,dM_s\Big)^2\Big] = \mathbf{E}\Big[\int_0^t H_s^2\,d\langle M\rangle_s\Big].$$

The processes H satisfying the condition (1.4) are called *locally bounded*. Further, a random process X is called *semimartingale*, if it can be represented in the form

$$X_t = X_0 + M_t + A_t,$$

where M is a local martingale vanishing at zero and A is a process of finite variation. This representation is obviously unique and the quadratic variation of X is then equal to $\langle M\rangle$. If H is a locally bounded process, then we define its integral with respect to X as the sum of integrals $\int_0^{\cdot} H_s\,dM_s$ and $\int_0^{\cdot} H_s\,dA_s$, where the first integral is understood in the Itô sense and the second one is a Stiltjes integral.

The result that we formulate below is one of the main tools of the stochastic calculus.

THEOREM 1.3 (Itô formula). *Let X be a continuous semimartingale and let F be a two times continuously differentiable function from $\mathbb{R}$ to itself, then for any positive T,*

$$F(X_T) = F(X_0) + \int_0^T F'(X_t)\,dX_t + \frac{1}{2}\int_0^T F''(X_t)d\langle X\rangle_t.$$

This implies, in particular, that the random process $\big(F(X_t)\big)_{t\geq 0}$ is a continuous semimartingale.

1.3. Stochastic Differential Equations

Stochastic Differential Equations (SDE) has been introduced in order to provide a mathematical model for a differential equation perturbed by a random noise. As in the theory of ordinary differential equations, here also the main problems are the existence, uniqueness and stability (in some probabilistic sense) of solutions. In our framework we are only interested in the one-dimensional case, but all the notions and results concerning the existence and the uniqueness of solution that we give below hold for the multidimensional case as well.

The exact definition is the following:

DEFINITION 4. *Let $S(\cdot)$ and $\sigma(\cdot)$ be locally bounded measurable functions from $\mathbb{R}$ to itself. A solution of SDE*

$$dX_t = S(X_t)\,dt + \sigma(X_t)\,dW_t, \quad X_0 = x, \quad t \geq 0, \tag{1.5}$$

is the giving of

- *a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$, where the filtration $\mathscr{F}_t$ is right-continuous and complete,*
- *an $\mathscr{F}_t$-Brownian Motion W defined on this space,*
- *an $\mathscr{F}_t$ adapted continuous process X defined on the same probability space such that for almost every ω, we have $X_0 = x$ and*

$$X_T - X_0 = \int_0^T S(X_t)\,dt + \int_0^T \sigma(X_t)\,dW_t, \quad \forall T > 0. \tag{1.6}$$

We say that this solution is strong, if X is $\mathscr{F}_t^W$-adapted.
We say that there is strong uniqueness of solution, if for any fixed filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ and $\mathscr{F}_t$-Brownian Motion W, any two solutions X and $\tilde{X}$ of (1.5) are indistinguishable, i. e.,

$$\mathbf{P}(\omega : \ X_t(\omega) = \tilde{X}_t(\omega), \, \forall t \geq 0) = 1.$$

The function $S(\cdot)$ is called *drift* or *trend coefficient*, and $\sigma(\cdot)$ is called *diffusion coefficient* or *volatility*. There are different sufficient conditions on these functions ensuring the strong existence and uniqueness of solution of (1.5). The most convenient result for our purposes is the following:

THEOREM 1.4 (cf. [**14**, page 190]). *If the functions $S(\cdot)$ and $\sigma(\cdot)$ are locally Lipschitz, i. e., for any positive K, there exists a constant L_K such that*

$$|S(x) - S(y)| + |\sigma(x) - \sigma(y)| < L_K|x - y|, \quad \forall x, y \in [0, K],$$

and for some constant B we have

$$2xS(x) + \sigma^2(x) \leq B(1 + |x|^2), \quad \forall x \in \mathbb{R},$$

then the equation (1.5) has a unique strong solution for any initial value $x \in \mathbb{R}$.

Under the conditions of this theorem, one can prove that for any filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ and an $\mathscr{F}_t$-Brownian Motion W there exists, for any random variable X_0 independent of W, a unique strong solution of (1.5) with initial value X_0. Throughout this work we suppose that the SDE (1.5) has a unique strong solution. However, for statistical purposes it is enough to require the existance of a weak solution.

DEFINITION 5. *A continuous random process X is called diffusion process, if there exist two functions S and σ from $\mathbb{R}$ to itself, such that the equality (1.6) is satisfied for any positive T.*

It is evident that any diffusion process is a semimartingale. Moreover, any solution of equation (1.6) possess the strong Markov property: *for any stopping time T and for any borelian function $\Phi : C(\mathbb{R}_+) \longrightarrow \mathbb{R}_+$, we have*

$$\mathbf{E}[\Phi(X_{T+t}, t \geq 0)|\mathscr{F}_T] = \mathbf{E}[\Phi(X_{T+t}, t \geq 0)|X_T].$$

So the solution of (1.6) is a strong Markov process. We say that a Markov process is *ergodic or possess ergodic properties* if there is a random variable X_0, such that the process X_t having X_0 as initial value is stationary. The law (distribution) of this random variable is called *invariant law (distribution)* of the process X, the density of this distribution with respect to the Lebesgue measure (if it exists) is called *invariant density* of the process X. The next result gives a sufficient condition on the coefficients of the equation (1.5) ensuring the ergodicity of the solution.

THEOREM 1.5. *Let the diffusion coefficient $\sigma(\cdot)$ be strictly positive and the following conditions be fulfilled*

$$V(S,\sigma,x) = \int_0^x \exp\left\{-2\int_0^y \frac{S(v)}{\sigma^2(v)}\,dv\right\}dy \to \pm\infty, \quad \text{as} \quad x \to \pm\infty$$

and

$$G(S,\sigma) = \int_{-\infty}^{\infty} \frac{1}{\sigma^2(x)} \exp\left\{2\int_0^x \frac{S(v)}{\sigma^2(v)}\,dv\right\}dx < \infty.$$

Then the process X defined by (1.6) possess ergodic properties and the invariant density is given by formula

$$f_{S,\sigma}(y) = \frac{1}{G(S,\sigma)\sigma^2(y)} \exp\left\{2\int_0^y \frac{S(v)}{\sigma^2(v)}\,dv\right\}. \tag{1.7}$$

The proof of this theorem can be found in [**21**]. A good exposition of results concerning ergodic properties of the solution of (1.6) can be found in [**31**], for more details see [**15**]. It is proved there that under the conditions of last theorem, the process X is recurrent positive (in the Harris sense). In fact, the first condition of this theorem guarantees the recurrence of the process X and the second one the positiveness. The recurrence means that for any initial value x and for any open ball $B \in \mathbb{R}$, the process starting at x visits B for almost every ω's. A recurrent process is called positive, if the mathematical expectation of the visiting time of an open ball B is finite for any B. It is well known that for a Markov process the positiveness does not depend on the choice of the initial value x. Moreover, if the process X issuing from x_0 is recurrent and positive with respect to an open ball $B_0 \subseteq \mathbb{R} \setminus \{x_0\}$, then it is so for any initial value x and any open ball $B \subseteq \mathbb{R}$. This feature can be interpreted as a kind of stability of solution.

An important consequence of ergodicity is the so called *strong law of large numbers*. It can be formulated in the following way.

THEOREM 1.6 (see [**31**]). *Let the conditions of previous theorem be satisfied and let us denote the invariant low by $m_{S,\sigma}$. Then, for any initial value x and for any*

$m_{S,\sigma}$-integrable function f, we have

$$\mathbf{P}\Big\{\frac{1}{T}\int_0^T f(X_t)\,dt \underset{T\to\infty}{\longrightarrow} \int_{\mathbb{R}} f(x)\,m_{S,\sigma}(dx)\Big\} = 1,$$

where X is the solution of (1.5).

Theorems 1.4 and 1.5 tell us that under suitable conditions on the trend coefficient S and diffusion coefficient σ there exists a unique strong solution of equation (1.5) and this solution is an ergodic Markov process. But the conditions of Theorem 1.5 have a form which is a little complicated. That is why we formulate below a result which gives a very simple condition ensuring the existence, uniqueness and ergodicity of diffusion process.

PROPOSITION 1. *Let the real functions $S(\cdot)$ and $\sigma(\cdot)$ be continuously differentiable and satisfy the condition*

$$\overline{\lim_{|x|\to\infty}}\ \mathrm{sgn}(x)\,\frac{S(x)}{\sigma^2(x)} < 0. \tag{1.8}$$

Let the diffusion coefficient $\sigma(\cdot)$ be bounded away from zero and has at most linear growth (i. e., $|\sigma(x)| \le C(1+|x|)$ for any x). Then the conditions of Theorem 1.4 and Theorem 1.5 are satisfied.

PROOF. Let us rewrite condition (1.8) in the following form: there exist two positive constants A and γ such that

$$\mathrm{sgn}(x)\,\frac{S(x)}{\sigma^2(x)} < -\gamma, \tag{1.9}$$

for any x satisfying the inequality $|x| > A$. We check firstly the conditions of Theorem 1.4. Since both S and σ are continuously differentiable, they are locally Lipschitz. Once more, using the continuity, we have

$$C_0 = \sup_{|x|\le A}\big[2xS(x) + \sigma^2(x)\big] < \infty.$$

Due to (1.9) the function $xS(x)$ is negative outside of the interval $[-A, A]$. Since $\sigma(\cdot)$ has at most linear growth, for a positive constant C_1, we have

$$\sigma^2(x) \le C_1(1+|x|)^2 \le 2C_1(1+x^2).$$

Combining these inequalities, we obtain

$$\sup_{x\in\mathbb{R}}\big[2xS(x) + \sigma^2(x)\big] \le C_0 + 2C_1(1+x^2) \le B(1+x^2).$$

To check the conditions of Theorem 1.5, we use (1.9). For any positive $x > A$, we have

$$\begin{aligned} V(S,\sigma,x) &= V(S,\sigma,A) + \int_A^x \exp\left\{-2\int_0^y \frac{S(v)}{\sigma^2(v)}\,dv\right\} dy \\ &= V(S,\sigma,A) + C\int_A^x \exp\left\{-2\int_A^y \frac{S(v)}{\sigma^2(v)}\,dv\right\} dy \\ &\geq V(S,\sigma,A) + C\int_A^x \exp\left\{2\gamma(y-A)\right\} dy \\ &= C_1 + C_2 e^{2\gamma(x-A)} \longrightarrow +\infty, \end{aligned}$$

when $x \to +\infty$. The conditions $V(S,\sigma,x) \to -\infty$ as $x \to -\infty$ and $G(S) < \infty$ are proved in the same way. □

1.4. Local Time and It's Martingale Representation

From now on, we fix a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ and a random process X – strong solution of SDE whose coefficients satisfy the conditions of Theorem 1.4. A very important feature of the diffusion processes from statistical point of view is the fact that the empirical distribution function $\hat{F}_T(x) = \frac{1}{T}\int_0^T \mathbb{1}_{\{X_t\leq x\}}dt$ has a Radon-Nikodym derivative with respect to the Lebesgue measure for any positive T. To formulate the corresponding result we need the following:

Tanaka formula. *For any real number a, there exists an increasing continuous process $(L_t^a)_{t\geq 0}$ called the local time of X at the point a such that,*

$$|X_t - a| = |X_0 - a| + \int_0^t \operatorname{sgn}(X_s - a)\,dX_s + L_t^a,$$

in particular, $|X_t - a|$ is a semimartingale.

Note that this formula is something like Itô formula for the function $F(x) = |x-a|$, which is not two times continuously differentiable. More generally, using the local times one can prove the extension of Itô formula to any function F which can be represented as a difference of two convex functions (Itô-Tanaka formula). In this case we have

$$F(X_T) - F(X_0) = \int_0^T F'_-(X_s)\,dX_s + \frac{1}{2}\int_0^T L_s^a\,F''(da),$$

where F'_- is the left side derivative of F and F'' is the second order derivative of F in the sense of distributions. Remind that the second order derivative of any convex function is a Radon measure.

The next result is one of the most interesting and useful properties of the local time.

THEOREM 1.7 (Occupation times formula). *There is a* $\mathbf{P}$*-negligible set outside of which*

$$\int_0^T \Phi(X_s)\, d\langle X\rangle_s = \int_{\mathbb{R}} \Phi(a) L_T^a\, da$$

for every positive number T *and every positive Borel function* Φ.

This result is true for any continuous semimartingale X. In the particular case of a diffusion process given by equation (1.6), this formula can be rewritten as

$$\int_0^T \Phi(X_s)\sigma^2(X_s)\, ds = \int_{\mathbb{R}} \Phi(a) L_T^a\, da.$$

Thus, if we denote

$$f_T^\circ(a) = \frac{L_T^a}{T\sigma^2(a)}, \tag{1.10}$$

then the empirical distribution function is

$$\hat{F}_T(x) = \frac{1}{T}\int_0^T \mathbb{1}_{\{X_s \le x\}} ds = \int_{\mathbb{R}} \frac{\mathbb{1}_{\{a\le x\}}}{T\sigma^2(a)}\, L_T^a\, da = \int_{-\infty}^x f_T^\circ(a)\, da.$$

It is proved (see [**51**], page 209) that there is a modification of the process $L = (L_T^a,\ T \ge 0,\, a \in \mathbb{R})$ which is a.s. continuous in T and cadlag in a. That is why we will always suppose that the function $f_T^\circ(\cdot)$ is a.s. cadlag. The last equality means that the empirical distribution function $\hat{F}_T$ of a diffusion process has a cadlag derivative (in the distributions sense). Remark that it is not the case for the empirical distribution function of a sequence of i.i.d. random variables.

The function f_T° is called *local time estimator* of invariant density. It has been introduced and studied by Kutoyants in [**38**]. In the same paper, the following representation of the local time estimator is proved:

$$f_T^\circ(x) - f(x) = \frac{2f(x)}{T}\int_{X_0}^{X_T} \left(\frac{\mathbb{1}_{\{v>x\}} - F(v)}{\sigma^2(v)f(v)}\right) dv - \frac{2f(x)}{T}\int_0^T \left(\frac{\mathbb{1}_{\{X_t>x\}} - F(X_t)}{\sigma^2(X_t)f(X_t)}\right) dW_t, \tag{1.11}$$

where $F(\cdot)$ and $f(\cdot)$ are respectively the invariant distribution function and the invariant density of an ergodic diffusion process X. This equality is called martingale representation of local time estimator, since the second term of the right hand side is a local martingale and the first one is in the most of cases negligible with respect to it.

Using this representation and the occupation times formula, we obtain for any integrable function $h(\cdot)$ such that $\mathbf{E}[h(\xi)] = \int_{\mathbb{R}} h(x)f(x)\,dx = 0$,

$$\begin{aligned} \frac{1}{T}\int_0^T h(X_t)\,dt &= \int_{\mathbb{R}} h(x)f_T^\circ(x)\,dx = \int_{\mathbb{R}} h(x)\big[f_T^\circ(x) - f(x)\big]\,dx \\ &= \frac{H(X_T) - H(X_0)}{T} - \frac{1}{T}\int_0^T g(X_t)\,dW_t, \end{aligned} \tag{1.12}$$

where we used the following auxiliary notations:

$$\begin{aligned} H(y) &= \int_0^y \int_{-\infty}^{\infty} 2h(x)f(x)\left(\frac{1\!\!1_{\{v>x\}} - F(v)}{\sigma^2(v)f(v)}\right)dx\,dv \\ &= \int_0^y \frac{2}{\sigma^2(v)f(v)}\int_{-\infty}^{v} h(x)f(x)\,dx\,dv \end{aligned} \tag{1.13}$$

$$g(y) = \frac{2}{\sigma^2(y)f(y)}\int_{-\infty}^{y} h(x)f(x)\,dx \tag{1.14}$$

The equality (1.12) plays a very important role in this work and is one of the main technical tools to prove the statistical properties for the model of ergodic diffusion.

1.5. Measures Corresponding to Diffusion Processes

Let $X = (X_t,\ t \geq 0)$ be a diffusion process defined on the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$ and given by equation (1.5), with initial condition $X_0 = x$. As we already said, this process is a.s. continuous. Consequently it induces a probability measure $\mathbf{P}_{x,\sigma,S}^{(T)}$ on the measurable space $(E^T, \mathscr{E}^T) = \big(C[0,T], \mathscr{B}(C[0,T])\big)$ defined by

$$\mathbf{P}_{x,\sigma,S}^{(T)}(A) = \mathbf{P}(\omega\,:\,X_{t_1}(\omega) \in A_1, \dots, X_{t_n}(\omega) \in A_n) \tag{1.15}$$

for any cylindric set $A = \big\{g \in C[0,T]\,\big|\,g(t_1) \in A_1, \dots, g(t_n) \in A_n\big\}$. Here $A_1, \dots, A_n$ are some Borel subsets of $\mathbb{R}$ and $\{t_1, \dots, t_n\}$ is an arbitrary subdivision of $[0,T]$. It is well known that the family of all cylindric sets generates the Borelian σ-algebra $\mathscr{E}^T$, consequently there exists a unique measure on $(E^T, \mathscr{E}^T)$

satisfying (1.15). We are interested in the behavior of the probability measures $\mathbf{P}^{(T)}_{x,\sigma,S}$ for different trend coefficients S when the diffusion coefficient σ is fixed. That is why we will often write $\mathbf{P}^{(T)}_{x,S}$ instead of $\mathbf{P}^{(T)}_{x,\sigma,S}$.

From statistical point of view, it is important to know whether two probability measures $\mathbf{P}^{(T)}_{x,S_1}$ and $\mathbf{P}^{(T)}_{x,S_2}$ are mutually absolutely continuous or not. The answer to this question is positive under fairly mild conditions and can be formulated as follows.

THEOREM 1.8. *Let the functions S_1, S_2 and σ be such that there exists a unique (weak) solution of each one of equations*

$$dX_t = S_1(X_t)\,dt + \sigma(X_t)\,dW_t, \quad X_0 = x,$$
$$dY_t = S_2(Y_t)\,dt + \sigma(Y_t)\,dW_t, \quad Y_0 = x,$$

where the initial value $x \in \mathbb{R}$ is the same for two equations. Let $\mathbf{P}^{(T)}_{x,S_1}$ and $\mathbf{P}^{(T)}_{x,S_2}$ be the probability measures corresponding to the solutions of these equations. Assume that the functions S_1, S_2, σ are locally bounded and

$$\mathbf{P}\left\{\omega : \int_0^T \left(\frac{S_1(X_t) - S_2(X_t)}{\sigma(X_t)}\right)^2 dt < +\infty\right\} = 1.$$

Then $\mathbf{P}^{(T)}_{x,S_1}$ is absolutely continuous with respect to $\mathbf{P}^{(T)}_{x,S_2}$ and the Radon-Nikodym derivative is given by formula

$$\frac{d\mathbf{P}^{(T)}_{x,S_1}}{d\mathbf{P}^{(T)}_{x,S_2}}(Y^T) = \exp\left\{\int_0^T \left(\frac{S_1(Y_t) - S_2(Y_t)}{\sigma(Y_t)}\right) dW_t - \frac{1}{2}\int_0^T \left(\frac{S_1(Y_t) - S_2(Y_t)}{\sigma(Y_t)}\right)^2 dt\right\}. \tag{1.16}$$

As a consequence of this theorem one can prove that if σ is bounded away from zero, and the functions S_1 and S_2 are bounded on any finite interval, then the measures $\mathbf{P}^{(T)}_{x,S_1}$ and $\mathbf{P}^{(T)}_{x,S_2}$ are mutually absolutely continuous.

Now let us suppose that the conditions of Theorem 1.5 are fulfilled for both couples (S_1, σ) and (S_2, σ). In this case, there exist invariant densities $f_1 = f_{S_1,\sigma}$ and $f_2 = f_{S_2,\sigma}$. Assume that the processes X and Y are as in Theorem 1.8 with only difference that the initial values of these processes are random: the density of X_0 w.r.t. the Lebesgue measure is f_1 and the density of Y_0 is f_2. In this case the law induced by X (resp. by Y) on the measurable space $(E, \mathscr{E}) = \big(C(\mathbb{R}_+); \mathscr{B}(C(\mathbb{R}_+))\big)$ will be denoted by $\mathbf{P}_{S_1}$ (resp. $\mathbf{P}_{S_2}$) and its restriction to the subspace $(E^T, \mathscr{E}^T) = \big(C[0,T]; \mathscr{B}(C[0,T])\big)$ by $\mathbf{P}_{S_1}^{(T)}$ (resp. by $\mathbf{P}_{S_2}^{(T)}$). As an immediate consequence of previous theorem one can prove that the measure $\mathbf{P}_{S_1}^{(T)}$ is absolutely continuous w.r.t. the measure $\mathbf{P}_{S_2}^{(T)}$ and the Radon-Nikodym density is

$$\frac{d\mathbf{P}_{S_1}^{(T)}}{d\mathbf{P}_{S_2}^{(T)}}(Y^T) = \frac{f_1(Y_0)}{f_2(Y_0)} \exp\left\{ \int_0^T \left(\frac{S_1(Y_t) - S_2(Y_t)}{\sigma(Y_t)} \right) dW_t - \frac{1}{2} \int_0^T \left(\frac{S_1(Y_t) - S_2(Y_t)}{\sigma(Y_t)} \right)^2 dt \right\}. \tag{1.17}$$

This formula needs perhaps some explanations. The stochastic integral figuring in the exponent of the right hand side is taken w.r.t. the Brownian Motion corresponding to the diffusion process Y. Therefore, the Radon-Nikodym density is defined by (1.17) on a set $E_1^T \subseteq E^T$ whose $\mathbf{P}_{S_2}^{(T)}$-measure is equal to one. Indeed, let us denote by $(\Omega, \mathscr{F}, \mathbf{P})$ the probability space on which the process Y is defined. Then the set E_1^T is the image of Ω by application $Y^T : \Omega \to E^T$ such that $Y^T(\omega) = (Y_t(\omega))_{t\in[0,T]}$. In the outside of E_1^T this density can be supposed to be equal to zero.

Note also that the stochastic integral with respect to the Brownian Motion in (1.17) can be replaced by the stochastic integral with respect to the semimartingale Y. This substitution leads us to the following formula:

$$\frac{d\mathbf{P}_{S_1}^{(T)}}{d\mathbf{P}_{S_2}^{(T)}}(Y^T) = \frac{f_1(Y_0)}{f_2(Y_0)} \exp\left\{ \int_0^T \left(\frac{S_1(Y_t) - S_2(Y_t)}{\sigma^2(Y_t)} \right) dY_t - \frac{1}{2} \int_0^T \left(\frac{S_1^2(Y_t) - S_2^2(Y_t)}{\sigma^2(Y_t)} \right) dt \right\}. \tag{1.18}$$

1.6. Examples

1. A simple example of a diffusion process is the so called Ornstein-Uhlenbek process, which is given by stochastic differential equation

$$dX_t = (aX_t + b)\, dt + \sigma\, dW_t, \quad X_0 = x, \quad t > 0.$$

Here a, b and $\sigma \neq 0$ are some real constants. A simple verification of the conditions of Theorem 1.4 shows that this equation has always a unique strong solution. Moreover, if $a < 0$ then this solution is ergodic.

2. Suppose that X is a random process solution of the SDE

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

where the random variable ξ is independent of W and the coefficients $S(\cdot)$ and $\sigma(\cdot)$ satisfy the conditions of Theorem 1.4. If, in addition, the function $\sigma(\cdot)$ is bounded away from 0, has at most linear growth and

$$\varlimsup_{|x|\to\infty} \frac{\operatorname{sgn}(x)\, S(x)}{\sigma^2(x)} < 0,$$

then, according to Proposition 1, X is an ergodic diffusion process. In this case the invariant density decreases exponentially implying the finiteness of the moments of any (algebraic) order.

3. There is another class of functions $S(\cdot)$ and $\sigma(\cdot)$ ensuring the strong existence, uniqueness and ergodicity of the corresponding diffusion process. Suppose that

$\sigma(\cdot)$ is bounded away from zero,
$\sigma(\cdot)$ has at most linear growth,
$S(\cdot)$ and $\sigma(\cdot)$ are locally Lipschitz and

$$\varlimsup_{|x|\to\infty} \frac{2xS(x)}{\sigma^2(x)} < -1.$$

In this case also there exists a unique solution of the corresponding SDE and this solution possess ergodic properties. But this class has a disadvantage: in general even the second order moment does not exist.

Remark also that the moments of any order exist if we replace the last condition by following one:

$$\varlimsup_{|x|\to\infty} \frac{2xS(x)}{\sigma^2(x)} = -\infty.$$

Note that the classes of functions described in this example contain the class of the previous example.

CHAPTER 2

Nonparametric Estimation

In this chapter we introduce the main notions of statistics essentially following the work of Nemirovski (cf. [**43**]) and the monograph of Johnstone (cf. [**34**]). Then we present the theorem of Pinsker and other similar results (for detailed discussion about it, see Nussbaum [**47**]). Finally we describe the model we are dealing with and formulate an important inequality for this model.

2.1. Generalities

We start this chapter by a brief outline of the main concepts of nonparametric estimation and statistics in general. A typical problem of statistics can be described as follows:

> We are given a Polish (*i. e.,* metric, separable and complete) *space of observations* E and a family of probability measures $(\mathbf{P}_h)_{h\in\mathcal{H}}$ on the measurable space $(E, \mathscr{E})$. Here $\mathscr{E}$ is the σ-algebra of Borel sets of E and the family of probability measures is parameterized by a *parameter* h varying in a subset $\mathcal{H}$ of a metric space.
>
> The goal is, given an *observation* – a realization y of random variable $Y \sim \mathbf{P}_h$ associated with an *unknown* $h \in \mathcal{H}$, to make conclusions about h, e.g.,
>
> · [Evaluation of a functional] To estimate the value $\Phi(h)$ at h of a given functional Φ from $\mathcal{H}$ to $\mathbb{R}$, or, more generally, to any metric space M,
> · [Hypotheses testing] Given a partition $\mathcal{H} = \bigcup_{i=1}^{N} \mathcal{H}_i$ of $\mathcal{H}$, to decide to which element $\mathcal{H}_i$ of the partition h belongs.

In this work we consider only some problems of evaluation of several functionals. An interesting case of evaluation problem is the so called identification problem: it corresponds to the case $\Phi(h) = h$.

In all statistical problems, the "candidate solution" is an estimate – an application $\bar{\Phi} : E \longrightarrow M$ measurable w.r.t. the Borelian σ-algebra of M. The application

$\bar{\Phi}$ is called *estimator* of Φ, and $\bar{\Phi}(y)$ is the *estimated value* of $\Phi(h_*)$, where h_* is the true value of parameter (*i. e.,* $Y \sim \mathbf{P}_{h_*}$). Further, we denote by $\mathbf{E}_h$ the mathematical expectation w.r.t. $Y \sim \mathbf{P}_h$.

Normally, it is impossible to recover the true value h_* exactly, and we should be satisfied with estimators $\bar{\Phi}$ which are with "high" probability "close" to the true value. In order to quantify the quality of estimators, and to be able to compare two estimators, the notions of *loss function* and *risk of estimators* are introduced. Let W be a nondecreasing positive function defined on $\mathbb{R}_+$ and vanishing at the origin: $W(0) = 0$. Then W is called loss function and the corresponding risk is defined as

$$R(\bar{\Phi}, \Phi(h)) = \mathbf{E}_h \Big[W\Big(d\big(\bar{\Phi}(Y), \Phi(h)\big)\Big)\Big],$$

where $d(\cdot, \cdot)$ is a metric on M. A typical example of loss function is the quadratic loss $W(x) = x^2$. The associated risk is often called *quadratic risk* or *mean squared error*. It is very natural to say that an estimator $\bar{\Phi}$ of $\Phi(h)$ is good if the risk $R(\bar{\Phi}, \Phi(h))$ is small. It is evident, that an estimator can be good for a value h and in the same time very bad for another value h'. This implies that, in order to be able to compare two estimators, one has to eliminate the dependence on the "true value" of parameter from the risk. There are two standard ways to do it:

[Bayesian approach] To take average of $R(\bar{\Phi}, \Phi(h))$ over a given a priori distribution π of $h \in \mathcal{H}$. This leads to the *Bayesian risk*

$$B(\bar{\Phi}, \pi) = \int_{\mathcal{H}} R(\bar{\Phi}, \Phi(h))\, \pi(dh).$$

[Minimax approach] To take the supremum of $R(\bar{\Phi}, \Phi(h))$ over $h \in \mathcal{H}$, thus coming to the worst-case risk

$$R(\bar{\Phi}, \mathcal{H}) = \sup_{h \in \mathcal{H}} R(\bar{\Phi}, \Phi(h))$$

of an estimator $\bar{\Phi}$ on the parameter set $\mathcal{H}$. Thus, the best estimator is the one minimizing the risk $R(\bar{\Phi}, \mathcal{H})$ and the quality of the "ideal" estimator becomes the *minimax risk*

$$R(\mathcal{H}) = \inf_{\bar{\Phi}} R(\bar{\Phi}, \mathcal{H}) = \inf_{\bar{\Phi}} \sup_{h \in \mathcal{H}} R(\bar{\Phi}, \Phi(h))$$

In this work, we always use the minimax approach. The major reason for this choice is that we intend to work with infinite-dimensional parameter sets, and these sets usually do not admit "natural" a priori distributions. In general, if the parameter space $\mathcal{H}$ can not be parameterized by a finite number of real parameters, the corresponding statistical problem is called *nonparametric*.

DEFINITION 6. *An estimator $\hat{\Phi}$ is called efficient (or minimax) if it satisfies the equality $R(\hat{\Phi}, \mathcal{H}) = \inf_{\bar{\Phi}} R(\bar{\Phi}, \mathcal{H}) = R(\mathcal{H})$.*

It is not difficult to show that the minimax and Bayesian approaches are connected by formula

$$R(\mathcal{H}) = \inf_{\bar{\Phi}} \sup_{\pi} B(\bar{\Phi}, \pi), \tag{2.1}$$

where the supremum is taken over all possible probability distributions π on $\mathcal{H}$. Indeed, on the one hand, we have

$$B(\bar{\Phi}, \pi) = \int_{\mathcal{H}} R(\bar{\Phi}, \Phi(h))\, \pi(dh) \leq \sup_{h \in \mathcal{H}} R(\bar{\Phi}, \Phi(h)).$$

This inequality implies that $\inf_{\bar{\Phi}} \sup_{\pi} B(\bar{\Phi}, \pi) \leq R(\mathcal{H})$. On the other hand, if we denote by δ_h the Dirac measure at the point $h \in \mathcal{H}$, then

$$\sup_{\pi} B(\bar{\Phi}, \pi) \geq B(\bar{\Phi}, \delta_h) = R(\bar{\Phi}, \Phi(h)), \text{ for any } h \in \mathcal{H}.$$

Consequently $\sup_{\pi} B(\bar{\Phi}, \pi) \geq \sup_{h \in \mathcal{H}} R(\bar{\Phi}, \Phi(h))$ and

$$\inf_{\bar{\Phi}} \sup_{\pi} B(\bar{\Phi}, \pi) \geq \inf_{\bar{\Phi}} \sup_{h \in \mathcal{H}} R(\bar{\Phi}, \Phi(h)) = R(\mathcal{H}).$$

This proves the equality (2.1) which is often used in order to obtain lower bounds for minimax risk, since

$$R(\mathcal{H}) = \inf_{\bar{\Phi}} \sup_{\pi} B(\bar{\Phi}, \pi) \geq \sup_{\pi} \inf_{\bar{\Phi}} B(\bar{\Phi}, \pi).$$

The prior distribution π_0 satisfying the equality

$$\inf_{\bar{\Phi}} B(\bar{\Phi}, \pi_0) = R(\mathcal{H})$$

is called *least favorable prior*. It is the structure of the minimax estimator and the least favorable prior, as well as their effect on the minimax risk, that is of chief statistical interest.

Remark that if the set $\mathcal{H}$ is too large, the minimax risk $R(\mathcal{H})$ can be equal to infinity. In this case the comparison of different estimators using minimax approach becomes impossible. That is why, in almost all nonparametric problems where the minimax approach is used, the set $\mathcal{H}$ is supposed to be a compact subset of a metric space, which guarantees the finiteness of $R(\mathcal{H})$.

2.2. Asymptotic Approach

As we have already said, the "ideal goal" of minimax approach is to find the minimax estimator and to evaluate the minimax risk. As a matter of fact, even in the very simple problems, this goal can not be achieved. That is why we are enforced to simplify our goal, and the standard simplification is to let the volume of data Y to increase to infinity. More precisely, we suppose that the dimension of the space $E = E^T$ is equal to T. For example,

if $Y = Y^n = (Y_1, \ldots, Y_n)$ is a random vector, then $T = n \in \mathbb{N}$,

if $Y = Y^T = (Y_t)_{t\in[0,T]}$ is a random process, then $T \in \mathbb{R}$.

When the dimension T is treated as a varying parameter, the minimax risk (as well as the Bayesian risk and the other notions introduced in previous section) becomes a function of T:

$$R_T(\mathcal{H}) = \inf_{\bar{\Phi}_T} \sup_{h\in\mathcal{H}} R_T(\bar{\Phi}_T, \Phi(h)) = \inf_{\bar{\Phi}_T} \sup_{h\in\mathcal{H}} \mathbf{E}_{h,T}\Big[W_T\Big(d\big(\bar{\Phi}_T(Y), \Phi(h)\big)\Big)\Big],$$

where the infimum is taken over the set of all possible estimates $\bar{\Phi}_T(y)$ of $\Phi(h)$ via observations $y = y^T$. In the case of asymptotic approach we are interested in the properties of estimators when the sample size is large, *i. e.,* as $T \to \infty$. The main asymptotic properties are the consistency, optimality of the rate of convergence and asymptotic efficiency of estimators.

An estimator $\hat{\Phi}_T$ is said to be *consistent*, if, for any value $h \in \mathcal{H}$, the risk $R_T\big(\hat{\Phi}_T, \Phi(h)\big)$ tends to zero when T goes to infinity. We say that it is *uniformly consistent*, if

$$\lim_{T\to\infty} R_T(\hat{\Phi}_T, \mathcal{H}) = 0.$$

From now on, we suppose that the loss function has the form $W_T(u) = W(\varphi_T u)$ where φ_T is a positive function of T. We say that φ_T is the *optimal rate of convergence* in the problem of estimation of $\Phi(h)$ with loss function W, if there exist two strictly positive constants c_* and c^*, and a family of estimators $(\hat{\Phi}_T)_{T\geq 0}$, such that

$$\varliminf_{T\to\infty} \inf_{\bar{\Phi}_T} \sup_{h\in\mathcal{H}} R_T(\bar{\Phi}_T, \Phi(h)) \geq c_*$$

and

$$\varlimsup_{T\to\infty} \sup_{h\in\mathcal{H}} R_T(\hat{\Phi}_T, \Phi(h)) \leq c^*.$$

Sometimes the estimators attaining the optimal rate of convergence are called *weakly asymptotically efficient.*

An estimator $\hat{\Phi}_T$ will be called *(strongly) asymptotically efficient* or *asymptotically minimax* if the constants c_* and c^* coincide: $c_* = c^* = c$. This definition is equivalent to the equality

$$R_T(\hat{\Phi}_T, \mathcal{H}) = \big(1 + o_T(1)\big)\, R_T(\mathcal{H}) = \big(1 + o_T(1)\big)c, \tag{2.2}$$

where $o_T(1)$ tends to zero as $T \to \infty$. Our goal becomes either to find an *asymptotically efficient estimator* $\hat{\Phi}_T(y)$. In this work we will distinguish two kinds of asymptotic efficiency. The estimators satisfying (2.2) will be called asymptotically efficient in the global minimax sense. In contrast with this, an estimator $\hat{\Phi}_T$ is called asymptotically efficient in the local minimax sense if

$$\overline{\lim_{\delta\to 0}}\ \overline{\lim_{T\to\infty}}\ R_T\big(\hat{\Phi}_T, V_\delta(h_0)\big) = \underline{\lim_{\delta\to 0}}\ \underline{\lim_{T\to\infty}}\ R_T\big(V_\delta(h_0)\big) \in \,]\,0, \infty\,[.$$

Here h_0 is an arbitrary element of $\mathcal{H}$ and the set $V_\delta(h_0) \subseteq \mathcal{H}$ is a δ-neighborhood of h_0. This element h_0 is commonly called center of localization.

Very often the asymptotically efficient estimator depends on some characteristics of the parameter space $\mathcal{H}$. These characteristics can be the smoothness of elements of $\mathcal{H}$ (if it is a functional space), the diameter of $\mathcal{H}$ in some metrics, etc. The dependence of an estimator on these characteristics is not convenient for practical applications. That is the reason of developing some adaptive methods in statistics which permit to estimate the unknown parameters without using their global characteristics.

2.3. Pinsker's Theorem

Let us consider the model of signal recovery in Gaussian white noise: given a realization $y = (y_t,\, t \in [0,1])$ of the random process $Y = Y^\varepsilon$ satisfying the equality

$$dY_t = \vartheta(t)\, dt + \varepsilon\, dW_t, \quad t \in [0,1], \tag{2.3}$$

one has to estimate the unknown function $\vartheta(\cdot) \in \Theta$. Here W_t is a Brownian Motion on the interval $[0,1]$. The function $\vartheta(\cdot)$ is called *signal* and $\varepsilon\, dW_t$ is called *Gaussian white noise.*

Integrating the equality (2.3) with respect to an orthonormal basis $\big(\varphi_k(\cdot), k \in \mathbb{Z}\big)$ of $L^2[0,1]$, we obtain the following sequence of equations:

$$y_k = \theta_k + \varepsilon \xi_k, \quad k \in \mathbb{Z}, \tag{2.4}$$

where $y_k = \int_0^1 \varphi_k(t)\, dY_t$ and $\xi_k = \int_0^1 \varphi_k(t)\, dW_t$ are the corresponding Fourier coefficients. It is easy to verify that the random variables $(\xi_k)_{k\in\mathbb{Z}}$ are independent and $\xi_k \sim N(0,1)$ for any $k \in \mathbb{Z}$. The model (2.4) is usually called *Gaussian sequence model*. The statistical problem is to estimate the sequence $\boldsymbol{\theta} = (\theta_k)_{k\in\mathbb{Z}}$ using the observations $(y_k)_{k\in\mathbb{Z}}$. In this model Pinsker (1980) made a significant discovery (cf. [**50**]). He studied the asymptotic behavior of the minimax risk

$$R_\varepsilon(\Theta) = \inf_{\bar{\boldsymbol{\theta}}_\varepsilon} \sup_{\boldsymbol{\theta}\in\Theta} \sum_{k\in\mathbb{Z}} \mathbf{E}_{\boldsymbol{\theta}} (\bar{\theta}_{k,\varepsilon} - \theta_k)^2$$

as $\varepsilon \to 0$, and found its exact asymptotics for a general class of parameter spaces Θ. To discuss Pinsker's result we need some definitions. An estimator $\bar{\boldsymbol{\theta}}_\varepsilon$ of $\boldsymbol{\theta}$ is called *linear*, if $\bar{\boldsymbol{\theta}}_\varepsilon = \bar{\boldsymbol{\theta}}_\varepsilon(\boldsymbol{\lambda}_\varepsilon) = (\lambda_{k,\varepsilon} y_k)_{k\in\mathbb{Z}}$, for a real sequence $\boldsymbol{\lambda}_\varepsilon = (\lambda_{k,\varepsilon})_{k\in\mathbb{Z}}$. The minimax linear risk is then defined as

$$R_\varepsilon^L(\Theta) = \inf_{\boldsymbol{\lambda}_\varepsilon} \sup_{\boldsymbol{\theta}\in\Theta} \sum_{k\in\mathbb{Z}} \mathbf{E}_{\boldsymbol{\theta}} (\lambda_{k,\varepsilon} y_k - \theta_k)^2.$$

The theorem of Pinsker contains several assertions. First, it gives an exact evaluation of the linear minimax risk and writes down the linear (nonasymptotically) minimax estimator in the problem of signal recovery in Gaussian noise (not necessarily white). It is proved when the parameter space

$$\Theta = \Theta(\boldsymbol{a}, R) = \Big\{\boldsymbol{\theta} \;\Big|\; \sum_{k\in\mathbb{Z}} a_k \theta_k^2 \le R\Big\} \tag{2.5}$$

is an ellipsoid in l^2. Secondly, it shows that in the case of white noise the general minimax risk is asymptotically equivalent to the minimax linear risk, *i. e.,*

$$R_\varepsilon(\Theta(\boldsymbol{a}, R)) \sim R_\varepsilon^L(\Theta(\boldsymbol{a}, R)), \qquad \varepsilon \to 0.$$

This means that there exist linear estimators which are asymptotically efficient. In fact, the result of Pinsker is even more general, because he considered the case $\xi_k \sim N(0, \sigma_k^2)$, where the variances σ_k^2 satisfy the condition

$$\sup_{k\colon a_k<d} \sigma_k^2 = o\bigg(\sum_{k\colon a_k<d} \sigma_k^2\bigg), \qquad d \to \infty.$$

It is evident that this condition is fulfilled in the case of Gaussian white noise, which corresponds to $\sigma_k = \text{const}$.

It was previously proved by Ibragimov and Khas'minskii ([**32**], for further results see [**33**, chapter VII, §§ 4-5]) that in the model "signal in Gaussian white noise" the optimal rate of convergence depends on the smoothness of underlying signal and is equal to $\varepsilon^{\frac{2m}{2m+1}}$ if the function $\vartheta(\cdot)$ is m times differentiable. In statistical estimation theory the problems where the optimal rate of convergence is less than ε is often called ill-posed (note that the quantity ε^{-2} plays in the Gaussian sequence model the same role as the sample size in nonparametric curve estimation problems). Pinsker's theorem was the first result providing an asymptotically efficient estimator (in the strong sense) in an ill-posed problem. Pinsker's argument proceeds by showing that there are Gaussian priors that are asymptotically least favorable, and that these priors are essentially concentrated on the ellipsoid $\Theta(\boldsymbol{a}, R)$. The Bayes estimators corresponding to these priors are linear, and are essentially the linear minimax estimators. This leads to asymptotic efficiency, since the Bayes risk for Gaussian priors coincides with the minimax linear risk (for more details see [**34**]).

An important corollary of Pinsker's theorem can be obtained considering a particular case of ellipsoid: Sobolev balls. For signal in Gaussian noise model the Sobolev ball $\mathcal{F}(m, R)$ is the set of all m-times differentiable (in the distributions sense) functions $\vartheta(\cdot) \in L^2(0,1)$, satisfying the inequality

$$\int_0^1 \left[\vartheta^{(m)}(x)\right]^2 dx \leq R. \tag{2.6}$$

If one choose $(\varphi_k)_{k\in\mathbb{Z}}$ to be the trigonometric basis on $[0,1]$, the condition corresponding to (2.6) in the sequence model is obtained by setting $a_k = |2\pi k|^{2m}$. In this case, for $\sigma_k \equiv 1$, Pinsker wrote down the explicit form of the first term in the asymptotic expansion of the minimax risk with respect to ε:

$$R_\varepsilon\big(\mathcal{F}(m, R)\big) = P(m, R)\,(\varepsilon^2)^{\frac{2m}{2m+1}}\big(1 + o_\varepsilon(1)\big),$$

where $P(m, R)$ is now called Pinsker's constant and is defined by

$$P(m, R) = (2m+1)\left(\frac{m}{\pi(m+1)(2m+1)}\right)^{\frac{2m}{2m+1}} R^{\frac{1}{2m+1}}. \tag{2.7}$$

An asymptotically efficient linear estimator is then

$$\theta^*_{i,\varepsilon} = \big(1 - |i/\alpha|^m\big)_+ y_i\ , \quad i \in \mathbb{Z},$$

with $\alpha^{2m+1} = R(m+1)(2m+1)/(m\varepsilon^2)$.

This result of Pinsker inspired a considerable literature. Investigations have been done in order to simplify the proof of this result and to generalize it, to prove analogous results in other ill-posed problems of nonparametric statistics, to study higher order efficiency for well-posed problems. We give below a list of results stimulated by Pinsker's theorem.

First, we mention the papers where the authors considered the same model as Pinsker and proved more general results. Thus Tsybakov (1997) proved that the result of Pinsker holds, if instead of quadratic loss one considers more general (possibly bounded) loss functions. Remind that Le Cam's notion of asymptotic equivalence of experiments permits to transfer the minimax lower bounds obtained for one model under bounded loss to another model, which is asymptotically equivalent to the first one. There are several results proving the asymptotic equivalence of some models with Gaussian sequence model (see, for example, [**34**, page 19]). So, the lower bound part of Tsybakov's result under bounded loss is true not only for Gaussian sequence model, but also for any equivalent one. Tsybakov calculated also the rate of convergence and the optimal constant in the problem of signal's derivatives estimation.

Bellitser and Levit (cf. [**2**]) gave a quite elementary proof of Pinsker's theorem. They were the first to derive the asymptotically exact lower bound using van Trees inequality. The expansion of the minimax risk over ellipsoids proposed by the authors, contains in some cases not only the first order term, but the second order term as well.

Golubev and Khas'minskii [**27**] showed that some inverse problems for partial differential equations are equivalent to the Gaussian sequence model with exponentially increasing variances. In the case of parabolic equation, they proved that the linear minimax estimator is asymptotically efficient of any (algebraic) order.

Note that all these results concern the estimation problem over l^2-ellipsoids. Donoho, Liu and MacGibbon (cf. [**9**]) considered the cases where the parameter space Θ has a different shape. They established that if Θ is quadratically convex, the minimax linear risk is within a factor 1.25 of the minimax risk nonasymptotically. They showed also, that if parameter space is an l^p-body with $p < 2$ (which is not quadratically convex), the minimax linear risk does not tend to 0 at the same rate as the minimax risk. In this setting, nonlinear estimators improve dramatically on linear estimators. The situation corresponding to the

l^q-loss is discussed in [**10**]. If the parameter space is an ellipsoid in l^p, the authors proved that the Pinsker's phenomenon occurs if and only if $q = p = 2$. Further, in the paper [**12**] of Donoho and Johnstone, the authors studied the estimation problem over Besov balls and proved that scalar nonlinearities applied to wavelet coefficients give estimators which are asymptotically efficient.

2.4. Analogous Results for Other Models

1. Spectral Density Estimation. In 1981 Efromovich and Pinsker [**16**] proved a similar result in the problem of spectral density estimation for discrete data. The problem is the following: given a realization of a Gaussian stationary sequence $\xi_1, \xi_2, \ldots, \xi_n$ having $f(\cdot)$ as spectral density, one has to estimate the function $f(\cdot) \in \mathcal{F}$, where $\mathcal{F}$ is a subset of $L^2([-1/2, 1/2])$. The ellipsoids in this model are defined by formula (2.5), setting $\theta_k = \int_{-1/2}^{1/2} f(\lambda)\cos(2\pi k\lambda)\, d\lambda$. The quality of estimation is measured by the minimax risk, using the following loss function:

$$W_n(\bar{f}_n, f) = \int_{-1/2}^{1/2} \big(\bar{f}_n(\lambda) - f(\lambda)\big)^2 d\lambda \Big/ \int_{-1/2}^{1/2} f^2(\lambda)\, d\lambda.$$

Authors constructed an estimator, which is showed to be asymptotically efficient if a_k increases to infinity as $k \to \infty$ faster than $\log k$.

2. Density Estimation. The efficiency results in the problem of density estimation based on i. i. d. data has been firstly obtained by Efromovich and Pinsker [**17**]. They treated the case where the observed i. i. d. random variables $y_1, \ldots, y_n$ take values in $[0, 1]$, which implies that their common density $f(\cdot)$ with respect to the Lebesgue measure (if it exists) vanishes outside of the interval $[0, 1]$. This density is assumed to exist and to be in a set $\mathcal{F}_+(\boldsymbol{a}, R)$ containing all probability density functions $f(\cdot) = \sum_{k=0}^{\infty} \theta_k \varphi_k(\cdot)$ such that $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots) \in \Theta(\boldsymbol{a}, R)$. Here $(\varphi_k,\ k = 0, 1, \ldots)$ is an orthonormal basis in $L^2(0, 1)$ satisfying some conditions. For any estimator $\bar{f}_n(\cdot)$, the quality of estimation is measured by the mean integrated squared error

$$R_n(\bar{f}_n, f) = \mathbf{E}_f \left[\int_0^1 \big(\bar{f}_n(x) - f(x)\big)^2 dx \right].$$

They constructed an estimator which turned out to be asymptotically efficient if $a_k / \log k \to +\infty$ as $k \to +\infty$. In the main case of periodic Sobolev class $\mathcal{F}_+(m, R)$ (which corresponds to $a_{2j} = a_{2j-1} = (2\pi j)^{2m}$, $j \geq 1$, $a_0 = 0$, and (φ_k)

is the trigonometric basis), they proved that the asymptotics of the minimax risk is the same as in the Gaussian sequence model (with $\varepsilon^2 = n^{-1}$), *i. e.*,

$$R_n(\mathcal{F}(m,R)) = n^{-\frac{2m}{2m+1}} P(m,R)\,\big(1+o_n(1)\big).$$

The proof of Efromovich and Pinsker relied essentially on a kind of uniform LAN property, generalized later by Golubev [**23**]. Schipper [**52**] proved the same result for densities having possibly unbounded support. The integrals in the expressions of the risk function and in the ellipsoid condition are then taken over whole $\mathbb{R}$. The author considered also the analytic class $\mathcal{A}(\mu,L)$, where $\mathcal{A}(\mu,L)$ is the set of all densities admitting a bounded analytic continuation in the strip $\big\{z\in\mathbb{C}\,\big|\,|\operatorname{Im} z|<\mu\big\}$ and such that $\|g_f\|^2 \le L$ with $g_f(x) = \lim_{y\to\pm\mu}\operatorname{Re} f(x+iy)$. In this case the first order term in the asymptotic expansion of the minimax risk does not depend on L, more precisely

$$R_n(\mathcal{A}(\mu,L)) = \frac{\log n}{2\pi\mu n}\,\big(1+o_n(1)\big).$$

In all these situations, the asymptotically efficient estimator is provided by a kernel-type estimating procedure. Using the lower bound obtained in [**23**], Golubev (cf. [**24**]) proved that a slight modification of the (globally) asymptotically minimax kernel-type estimator obtained in these works leads to an asymptotically efficient estimator in the local minimax sense. He proposed also an adaptive estimator and proved its asymptotic efficiency.

3. Intensity Function Estimation for Poisson Processes. Let us consider a Poisson process $Y^n = (Y_t,\, 0\le t\le n)$ with intensity measure Λ. Suppose that this measure is absolutely continuous with respect to the Lebesgue measure and $S(\cdot)$ is the corresponding Radon-Nikodym density. Kutoyants [**37**] considered the estimation problem of this function S over the class $\mathcal{F}_m(S_*,R)$ of all periodic functions with period one such that $\int_0^1 S(x)\,dx \le S_*$ and

$$\int_0^1 \big[S^{(m)}(x)\big]^2 dx \le R.$$

In this problem the observation is a trajectory of process Y^n. It is proved that the minimax risk has the following asymptotics:

$$R_n\big(\mathcal{F}_m(S_*,R)\big) \sim n^{-\frac{2m}{2m+1}} P(m,R)\, S_*^{\frac{2m}{2m+1}},$$

as $n\to\infty$. In fact, this problem is equivalent to the problem of intensity function estimation based on n independent trajectories of a Poisson process on $[0,1]$.

4. Signal Recovery in Mixed Gaussian White Noise. Let us consider the random process $Y^n = (Y_t^n,\ t \in \mathbb{R})$ given by

$$dY_t^n = \vartheta(t)\,dt + \sqrt{\frac{V_t}{n}}\,dW_t, \quad t \in \mathbb{R}.$$

Here W is a two sided Brownian Motion and $V = (V_t,\ t \in \mathbb{R})$ is a positive random process. Delattre and Hoffmann studied in [**7**] the following problem: given a realization y^n of the random process Y^n over a random interval $[A, B]$, to estimate the unknown function $\vartheta(\cdot)$ having a priori information that $\vartheta \in \Theta$. The random interval $[A, B]$ is supposed to be in a bounded interval $[a, b]$ and the triplet (V, A, B) is assumed to be independent of W. The law of this triplet is supposed to be known. The authors obtained an asymptotic lower bound of the weighted minimax risk

$$R_n(\bar\vartheta_n, \vartheta) = \mathbf{E}_\vartheta^n\Big[\int_a^b \big(\bar\vartheta_n(t) - \vartheta(t)\big)^2 \Gamma_t\,dt\Big]$$

when Θ is a weighted Sobolev class

$$W(m, \rho) = \Big\{\vartheta : \mathbb{R} \to \mathbb{R}\ \Big|\ \int_a^b \big|\vartheta^{(m)}(x)\big|^2 \rho(x)\,dx \le 1\Big\}.$$

Here $\Gamma_t = \Gamma_t(V, A, B)\mathbb{1}_{\{t\in[A,B]\}}$ is a continuous positive random process and $\rho(\cdot)$ is a continuous positive function. Under suitable conditions on Γ, V and ρ, the authors proved that the optimal rate of convergence is $n^{-\frac{m}{2m+1}}$ and computed the optimal constant. It turned out that the asymptotically efficient estimator for this model depends not only on the realization of V, but also on its law.

5. Regression Function Estimation. Consider the nonparametric regression model

$$Y_i^n = f(x_i^n) + \xi_i, \quad i = 1, \ldots, n,$$

where observations Y_i^n are taken at distinct points x_i^n of a finite interval $[a, b]$. The assumptions on the random errors ξ_i are $\mathbf{E}[\xi_i] = 0$ and $\mathbf{E}[\xi_i \xi_j] = \sigma^2 \delta_{ij}$ for any $i, j = 1, \ldots, n$ (δ_{ij} is the Kronecker symbol, *i. e.*, δ_{ij} is 1 if $i = j$ and 0 if $i \neq j$). The nonrandom design points x_i^n are assumed to be generated by a density $g(\cdot)$ on $[a, b]$ such that

$$\int_a^{x_i^n} g(t)\,dt = \frac{2i-1}{2n}, \quad \text{for any}\ \ i = 1, \ldots, n,$$

where $g(\cdot)$ is assumed to be continuous and strictly positive on $[a, b]$. Speckman [**53**] studied the problem of estimation of $f(\cdot)$ under the risk

$$\tilde{R}(\bar{f}_n, f) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{E}_f \big[\big(\bar{f}_n(x_i^n) - f(x_i^n) \big)^2 \big].$$

Using spline smoothing, he showed that

$$\tilde{R}_n\big(\mathcal{F}(m, R)\big) \sim n^{-\frac{2m}{2m+1}} P(m, R) \, [\sigma^2 c_m(g)]^{\frac{2m}{2m+1}}, \tag{2.8}$$

where

$$c_m(g) = \Big(\int_a^b g^{\frac{1}{2m}}(x) \, dx \Big)^{-1}.$$

In the paper [**46**] of Nussbaum the case of normal ξ_i and uniform design ($g \equiv 1$ with $[a, b] = [0, 1]$) was studied, taking as risk function the mean integrated squared error. He proved that the asymptotics (2.8) holds for this risk as well. The constant $c_m(g)$ is then evidently equal to one. An analogous result for Besov balls is proved by Donoho and Johnstone in [**13**]. Golubev and Nussbaum (cf. [**30**]) discussed the case where $\xi_1, \dots, \xi_n$ are independent but not necessarily Gaussian. They obtained a lower bound for weighted L^2-type risk and claimed its attainability by minimax linear smoothing spline estimator.

6. Distribution Function Estimation. Let $\xi_1, \dots, \xi_n$ be a sequence of i. i. d. random variables with common distribution function $F(x) = \mathbf{P}(\xi_1 \le x)$. In the problem of estimation of the distribution function $F(\cdot)$ using i. i. d. data, Dvoretzky, Kiefer and Wolfowitz (1956) proved the asymptotic minimaxity of empirical distribution function

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{\xi_i \le x\}}.$$

The rate of convergence is then equal to $n^{-1/2}$ and the asymptotic variance is $R_0(F) = \int_{\mathbb{R}} F(x)\big(1 - F(x)\big) \, dx$. Further, Golubev and Levit (cf. [**28**]) studied the asymptotic behavior of the second order minimax risk

$$R_n^{(2)}(\mathcal{F}) = \inf_{\bar{F}_n} \sup_{F \in \mathcal{F}} \Big\{ \mathbf{E}_F \int_{\mathbb{R}} [\bar{F}_n(x) - F(x)]^2 dx - n^{-1} R_0(F) \Big\}$$

as n tends to infinity. In the case where the parameter space is a Sobolev class they established the following result:

$$R_n^{(2)}\big(\mathcal{F}(m+1, R)\big) \sim -n^{-\frac{2m+2}{2m+1}} (2m+1) \Big(\frac{m+1}{\pi m (2m+1)} \Big)^{\frac{2m+2}{2m+1}} R^{-\frac{1}{2m+1}} \ ,$$

as $n \to \infty$. In the same paper an analogous result is obtained for analytic class $\mathcal{A}(\mu, L)$. In the article [**29**], the authors investigated the second order asymptotic efficiency in the local sense for analytic class of distributions. They described also (first order) asymptotically efficient estimators of derivatives of distribution function in the local minimax sense.

Other results similar to Pinsker's theorem are proved by Golubev [**25**], Golubev and Härdle [**26**], Donoho and Johnstone [**11**], etc.

2.5. The Model of Ergodic Diffusion

From now on, we assume that the random process X is the solution of SDE

$$dX_t = S(X_t)\,dt + \sigma(X_t)\,dW_t, \quad t \geq 0, \tag{2.9}$$

where the coefficients S and σ satisfy the conditions of Theorem 1.4 and Theorem 1.5. In addition, we suppose that the initial value $X_0 = \xi$, where ξ is a random variable which follows the invariant law and is independent of W. Thus the process X is stationary and possess ergodic properties.

The statistical model we consider is the following: the observation is a continuous trajectory x^T of the solution of (2.9) on the finite time interval $[0, T]$, the unknown parameter is the trend coefficient $S(\cdot)$, the diffusion coefficient σ is supposed to be known and the properties of estimators are studied in the asymptotics of large samples, *i. e.,* as $T \to \infty$. We do not precise here the type of the loss function since it is specific to each problem, but generally it will be something like mean integrated squared error.

In this setting, the nonparametric statistical problems which are traditionally of the major interest are the estimation of the invariant distribution function F_S, the estimation of the invariant density f_S and finally the estimation of the trend coefficient S. The first two problems has been thoroughly studied by several authors. A detailed exposition of all these results can be found in [**38**] (see also [**35**], [**36**], [**42**]). Throughout this work the diffusion coefficient is supposed known, so the problem of its estimation does not arise.

It appeared in the previous works, that there was an analogy between the problem of stationary distribution function estimation for ergodic diffusion and the classical problem of distribution function estimation for i. i. d. observations. In particular, it is proved that in both problems the optimal rate of convergence is $1/\sqrt{T}$ and that the empirical distribution function is asymptotically efficient.

However, the analogy between the model of ergodic diffusion and i. i. d. model breaks as soon as we consider the problem of invariant density estimation. It is well known that in the case of density estimation based on i. i. d. data the optimal rate of convergence depends on the smoothness of the estimating density function and, in particular, is always worse than $1/\sqrt{n}$ (here n is the sample size, *i. e.,* the number of observations). In contrast with this, in the problem of invariant density estimation for ergodic diffusion processes the rate $1/\sqrt{T}$ is attained by several estimators. Some exact results concerning the $\sqrt{T}$-efficiency in the problem of invariant density estimation are recalled in Chapter 5.

Our principal aim in this work is to investigate the behaviour of the estimators of the trend coefficient. We show that a natural way to achieve this aim pass by invariant density's derivative estimation. This is the other problem considered in this paper.

2.6. Van Trees Inequality

A very important tool in the proof of lower bounds for the risk of all possible estimators is van Trees inequality. It permits to minorate the minimax risk by something does not depending on estimators. This inequality can be interpreted as a Bayesian version of Cramér-Rao lower bound and is due to van Trees (see [**55**]). Further generalizations of this inequality and its applications for obtaining optimal rates of convergence can be found in [**22**].

Van Trees inequality is applicable in the following setting:

- the trend coefficient $S(x) = S_{\boldsymbol{\theta}}(x)$ is a function of $\boldsymbol{\theta}$ and x, with $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d) \in \Theta$ and $\Theta = \prod_{i=1}^{d} [a_i, b_i] \subseteq \mathbb{R}^d$,
- for any value of parameter $\boldsymbol{\theta}$, the function $S_{\boldsymbol{\theta}}$ fulfills the conditions of Theorem 1.4 and Theorem 1.5,
- we are given a probability density $p(\boldsymbol{\theta})$ (with respect to the Lebesgue measure) on Θ,
- $p(\boldsymbol{\theta})$ is supposed to be differentiable, positive on the interior of Θ and zero on its boundary.

We use below the following notations:

$$\partial_{\theta_i} u(\boldsymbol{\theta}) = \frac{\partial u}{\partial \theta_i}(\boldsymbol{\theta}), \qquad I_i = \int_\Theta \big[\partial_{\theta_i} \log p(\boldsymbol{\theta})\big]^2 p(\boldsymbol{\theta})\, d\boldsymbol{\theta},$$

where $u(\boldsymbol{\theta})$ is an arbitrary differentiable function. In the theory of parameter estimation the quantity I_i is well known and is called Fischer information corresponding to the coefficient θ_i. It can be rewritten in an equivalent form:

$$I_i = \int_\Theta \frac{\big[\partial_{\theta_i} p(\boldsymbol{\theta})\big]^2}{p(\boldsymbol{\theta})}\, d\boldsymbol{\theta}.$$

We denote by $\mathbf{P}_{\boldsymbol{\theta}}$ the measure $\mathbf{P}_{S_{\boldsymbol{\theta}}}$ and by $\Lambda_T(\boldsymbol{\theta}, x^T)$ the restriction on $\mathscr{F}_T$ of the Radon-Nikodym density of $\mathbf{P}_{\boldsymbol{\theta}}$ w.r.t. the measure $\mathbf{P}_0 = \mathbf{P}_{\boldsymbol{\theta}_0}$, where $\boldsymbol{\theta}_0$ is an arbitrary (fixed) value of parameter.

THEOREM 2.1. *Let the random process X be a solution of the SDE (2.9) on $[0,T]$ such that the initial value X_0 follows the invariant law. Assume that the trend coefficient $S = S_{\boldsymbol{\theta}}$ depends on a d-dimensional parameter $\boldsymbol{\theta} \in \Theta$. Let $\big(\Phi_i(\boldsymbol{\theta}),\ i = 1, \ldots, d\big)$ be a sequence of differentiable functions of $\boldsymbol{\theta}$. Then, for any estimators $\bar{\Phi}_{i,T} = \bar{\Phi}_{i,T}(x^T)$ of $\Phi_i(\boldsymbol{\theta})$, the Bayes risk can be bounded as follows:*

$$\sum_{i=1}^d \mathbb{E}\left[(\hat{\Phi}_{i,T} - \Phi_i(\boldsymbol{\theta}))^2\right] \geq \frac{\Big(\sum_{i=1}^d \int_\Theta \left[\partial_{\theta_i} \Phi_i(\boldsymbol{\theta})\right] p(\boldsymbol{\theta})\, d\boldsymbol{\theta}\Big)^2}{\sum_{i=1}^d \int_\Theta I_i(\boldsymbol{\theta}) p(\boldsymbol{\theta})\, d\boldsymbol{\theta} + \sum_{i=1}^d I_i},$$

where $\mathbb{E}$ denotes the integration with respect to the probability measure $\mathbf{P}_{\boldsymbol{\theta}}(dx^T) \times p(\boldsymbol{\theta})\, d\boldsymbol{\theta}$ and the term $I_i(\boldsymbol{\theta})$ is defined by

$$I_i(\boldsymbol{\theta}) = \mathbf{E}_{\boldsymbol{\theta}}\big[\partial_{\theta_i} \log \Lambda(\boldsymbol{\theta}, X^T)\big]^2.$$

PROOF. The proof of this theorem is based on the application of Cauchy-Schwarz inequality and integration by parts formula. It can be found in [**22**]. □

Van Trees inequality, as it is stated above, is true for any statistical model. In our particular case we can do some further developpements. In fact, using the formula (1.18), one can easily check the equality

$$I_i(\boldsymbol{\theta}) = \mathbf{E}_{\boldsymbol{\theta}}\left[\partial_{\theta_i} \log f_{\boldsymbol{\theta}}(X_0) + \partial_{\theta_i}\left(\int_0^T \frac{S_{\boldsymbol{\theta}}(X_t)}{\sigma^2(X_t)}\, dX_t\right) - \int_0^T \frac{\partial_{\theta_i} S^2_{\boldsymbol{\theta}}(X_t)}{2\sigma^2(X_t)}\, dt\right]^2.$$

In the case where the differential operator ∂_{θ_i} and the stochastic integral w.r.t. X are interchangeable, we have

$$\begin{aligned}\partial_{\theta_i}\left(\int_0^T \frac{S_{\boldsymbol{\theta}}(X_t)}{\sigma^2(X_t)}\,dX_t\right) - \int_0^T \frac{\partial_{\theta_i} S^2_{\boldsymbol{\theta}}(X_t)}{2\sigma^2(X_t)}\,dt \\ &= \int_0^T \frac{\partial_{\theta_i} S_{\boldsymbol{\theta}}(X_t)}{\sigma^2(X_t)}\,dX_t - \int_0^T \frac{\partial_{\theta_i} S_{\boldsymbol{\theta}}(X_t)}{\sigma^2(X_t)}\,S_{\boldsymbol{\theta}}(X_t)\,dt \\ &= \int_0^T \frac{\partial_{\theta_i} S_{\boldsymbol{\theta}}(X_t)}{\sigma^2(X_t)}\left[dX_t - S_{\boldsymbol{\theta}}(X_t)\,dt\right] \\ &= \int_0^T \frac{\partial_{\theta_i} S_{\boldsymbol{\theta}}(X_t)}{\sigma(X_t)}\,dW_t.\end{aligned}$$

Consequently, we obtain

$$I_i(\boldsymbol{\theta}) = \mathbf{E}_{\boldsymbol{\theta}}\left[\partial_{\theta_i}\log f_{\boldsymbol{\theta}}(X_0) + \int_0^T \frac{\partial_{\theta_i} S_{\boldsymbol{\theta}}(X_t)}{\sigma(X_t)}\,dW_t\right]^2. \tag{2.10}$$

As we mentioned, the last equality is true if the differential operator and the stochastic integral are interchangeable. Throughout this work we consider only the case where the dependence of S on θ_i is linear. In this case the condition of interchangeability is obviously fulfilled.

CHAPTER 3

Estimation of the Derivative of Invariant Density

3.1. The Problem

In this chapter we study the problem of invariant density's derivative estimation given a trajectory $x^T = \big(x(t),\, t \in [0,T]\big)$ of a diffusion process X. We assume that X is defined by the stochastic differential equation

$$dX_t = S(X_t)\,dt + dW_t, \quad X_0 = \xi, \quad t \in \mathbb{R}. \tag{3.1}$$

As before, $W = (W_t,\, t \in \mathbb{R})$ denotes a Brownian Motion defined on a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ and ξ is a random variable independent of W. We suppose that the trend coefficient S belongs to the set Σ_0 of all real functions satisfying the conditions of Theorem 1.4 and Theorem 1.5 with $\sigma \equiv 1$. By these conditions the process X is ergodic, *i. e.,* there exists an invariant measure with the density function (w. r. t. the Lebesgue measure)

$$f_S(y) = \frac{1}{G(S)}\,\exp\Big\{2\int_0^y S(v)\,dv\Big\},$$

where $G(S)$ is the normalizing constant:

$$G(S) = \int_{\mathbb{R}} \exp\Big\{2\int_0^y S(v)\,dv\Big\}\,dy.$$

We assume that the initial value ξ follows the invariant law ensuring thus the stationarity of the process X. The trend coefficient $S(\cdot)$ is supposed to be unknown, so the invariant density $f_S(\cdot)$ and its derivatives are unknown functions too. We are interested in the estimation of the first order derivative $f_S'(\cdot)$ using integrated quadratic risk

$$R_T(\bar\vartheta_T, f_S') = \int_{\mathbb{R}} \big(\bar\vartheta_T(x) - f_S'(x)\big)^2 dx$$

in order to measure the error of estimation. Here $\bar\vartheta_T$ is an arbitrary estimator of unknown function f_S', *i. e.,* a function on $\mathbb{R}\times\Omega$ which is $\mathscr{B}(\mathbb{R})\otimes\mathscr{F}_T$-measurable.

Our aim is to determine the exact asymptotics of the minimax risk as T tends to infinity. The parameter space considered in this chapter is a subset of $\Sigma_0 \cap \Sigma(m, R)$, where $\Sigma(m, R)$ is the set of all functions S such that $f'_S \in \mathcal{F}(m, R)$, *i. e.*, $f'_S \in L^2(\mathbb{R})$ is m-times differentiable in the distributions sense and

$$\|f_S^{(m+1)}\|_2^2 = \int_{\mathbb{R}} \left[f_S^{(m+1)}(x)\right]^2 dx \leq R.$$

From now on, we denote by $\|\cdot\|_2$ the norm of the Hilbert space $L^2(\mathbb{R})$. It appears that there are some similarities between the problem of density's derivative estimation and the signal recovery problem in white noise. Namely, the optimal rate of convergence and the constant that we obtain are almost the same as those of Pinsker's theorem. More precisely, we prove that

$$\inf_{\bar{\vartheta}_T} \sup_{S \in \Sigma} R_T(\bar{\vartheta}_T, f'_S) \sim (4T^{-1})^{\frac{2m}{2m+1}} P(m, R).$$

For the exact definition of the parameter space we need several notations. Let $\big(\mathrm{C}(\mathbb{R}_+), \mathscr{B}(C(\mathbb{R}_+))\big)$ be the space (equipped with the topology of the uniform convergence on the compacts) of all continuous functions from $\mathbb{R}_+$ to $\mathbb{R}$ and the σ-algebra of its Borelian subsets. The process $X = \{X_t,\ t \geq 0\}$ induces a probability measure on the measurable space $\big(\mathrm{C}(\mathbb{R}_+), \mathscr{B}(C(\mathbb{R}_+))\big)$. This measure and the Lebesgue integral with respect to this measure will be denoted by $\mathbf{P}_S$ and $\mathbf{E}_S$ respectively. In the cases where the function $S(\cdot)$ depends on a parameter θ, we will write $\mathbf{P}_\theta$ and $\mathbf{E}_\theta$ instead of $\mathbf{P}_{S_\theta}$ and $\mathbf{E}_{S_\theta}$.

Let $\mathbb{1}_A$ denote the indicator of the set $A \subset \mathbb{R}$ and $F_S(\cdot)$ be the distribution function associated with density $f_S(\cdot)$, *i. e.*,

$$F_S(x) = \mathbf{P}(\xi \leq x) = \int_{-\infty}^{x} f_S(y)\, dy.$$

The constants do not depending on the sample size T and trend coefficient $S(\cdot)$ will be denoted by C and D.

We can define now the infinite dimensional parameter space Σ as the set of all $S \in \Sigma(m, R) \cap \Sigma_0$, such that for a constant D the following conditions are fulfilled:
1. There exist positive constants B_1 and B_2 such that

$$\sup_{B > B_1} B^2 \int_{|x|>B} \left[f'_S(x)\right]^2 dx < D, \tag{3.2}$$

$$\sup_{B > B_2} B^{-2} \int_{|x|<B} \left[f'_S(x)\right]^2 \mathbf{E}_S \left[\int_0^{\xi} \frac{\mathbb{1}_{\{y>x\}} - F_S(y)}{f_S(y)}\, dy\right]^2 dx < D. \tag{3.3}$$

2. The following estimate holds:

$$\int_{\mathbb{R}} \left[f'{}_S(x)\right]^2 \mathbf{E}_S \left[\frac{\mathbb{1}_{\{\xi>x\}} - F_S(\xi)}{f_S(\xi)}\right]^2 dx < D. \tag{3.4}$$

So the parameter space $\Sigma = \Sigma(m, R, B_1, B_2, D)$ depends on B_i, $i = 1, 2$, and D as well. However, these constants have no effect on the asymptotics of the minimax risk. Moreover, it will be clear from our proof, that the quantities of the left hand sides of (3.2), (3.3) and (3.4) needn't be bounded by the same constant, these constants can be different in each case.

3.2. Some Motivations

The most interesting and the mostly studied nonparametric problems in the statistics of diffusion processes are the problems of the trend coefficient, diffusion coefficient, invariant density and invariant distribution function estimation. For us, the estimation of the derivative of density is mainly interesting because of its implications for trend coefficient estimation. To explain the link between these two problems, note that the following equality is true:

$$S(x) = f'_S(x)/2f_S(x).$$

So an intuitive approach to the problem of trend coefficient estimation could be to estimate the derivative function f'_S and the density f_S, then to take the half of their quotient as an estimator of S. The only missing link to realize this schema is the derivative function's estimator, since some asymptotically efficient estimators of f_S are given in [**36**].

The second motivation for studying this problem is its possible applications in the invariant density estimation. Denote by M^* the maximum of the observed path x^T. The local time estimator f_T°, which is an asymptotically efficient estimator of the invariant density, is equal to 0 on $[M^*, \infty[$. In this case, for the values $x \in [M^*, M^* + \varepsilon]$ with small ε, the estimator

$$\bar{f}_S(x) = f_T^\circ(M^*) + \bar{f}'_T(M^*)(x - M^*)$$

constructed with the help of an estimator $\bar{f}'_T$ of the derivative f'_S can be better than the local time estimator.

Finally, the results of this chapter permit to understand better the model of ergodic diffusion and reveal its similarity with some other models of nonparametric curve estimation. We show that the Pinsker's phenomenon occurs in our model just from the first order derivative estimation. Roughly speaking, this means that there is a kind of shift between the estimation problems for ergodic diffusion and i. i. d. observations. In other words, the model of ergodic diffusion is one order smoother than the i. i. d. model.

The problem of the invariant density's derivative estimation has been previously considered by Lucas in [**40**], who proved in more general setup that the rate of convergence in this problem for the kernel estimators is (strictly) slower than $1/\sqrt{T}$.

3.3. Lower Bound

In this section, using the method presented in [**28**], we establish a lower bound on the minimax risk

$$\inf_{\bar\vartheta_T} \sup_{S\in\Sigma} R_T(\bar\vartheta_T, f_S') = \inf_{\bar\vartheta_T} \sup_{S\in\Sigma} \mathbf{E}_S \left\| \bar\vartheta_T - f_S' \right\|_2^2 , \tag{3.5}$$

where the inf is taken over all estimators of density's derivative.

The sketch of the proof is the following. First, we define a parametric family of trend coefficients and find a lower bound in the problem of density's derivative estimation under the condition that S belongs to this parametric family. To do it, we minorate the integral on $\mathbb{R}$ by the integral over a bounded interval $[-A_T, A_T]$ and apply the Parseval's identity using the trigonometric basis $\{e_i\}_{i\in\mathbb{Z}}$ on $[-A_T, A_T]$. Secondly, we consider a prior distribution on this parametric family and applying the two-dimensional van Trees inequality we obtain a lower bound. The advantage of this bound is that it depends only on the prior distribution and is the same for all estimators. Then we prove that the appropriate choice of the prior distribution leads to the desirable lower bound. The final and perhaps the most difficult step is to show that this (asymptotically least favorable) prior distribution is essentially concentrated on the parameter space Σ.

THEOREM 3.1. *If $m > 1$, then the following inequality is true*

$$\varliminf_{T\to\infty} \inf_{\bar\vartheta_T} \sup_{S\in\Sigma} T^{\frac{2m}{2m+1}} R_T(\bar\vartheta_T, S) \geq 4^{\frac{2m}{2m+1}} P(m,R), \tag{3.6}$$

where $P(m,R)$ is the Pinsker's constant defined by (2.7).

PROOF. The first step of the proof, as it has been already said, is the construction of a parametric family. The dimension of this family has to be finite for any T and tend to infinity when $T \to \infty$.
Thus, following Golubev and Levit [**28**], we introduce the parameterization

$$S_{\boldsymbol{\theta}}(x) = \sum_{|i|\leq L} \theta_i e_i(x)\, U(A - |x|), \quad x \in [-A, A], \quad \boldsymbol{\theta} \in \mathbb{R}^{2L+1}, \tag{3.7}$$

where $A = A_T = \ln(1+T)$ and $U(x)$ is $m+1$ times differentiable increasing function vanishing for $x \leq 0$ and equal to one for $x \geq 1$. The functions $\{e_i\}_{i\in\mathbb{Z}}$ are the elements of the trigonometric basis on $[-A, A]$, *i. e.*,

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

The positive number $L = L_T$ will be chosen later. Remark that the function $U(A - |x|)$ vanishes for $|x| \geq A$ and is equal to one for $|x| \leq A - 1$. Roughly speaking, this is a kind of smooth approximation of the indicator function of the interval $[-A, A]$.

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

$$S_{\boldsymbol{\theta}}(x) = S_0(x) = -\operatorname{sgn}(x)(m+2)\big(|x| - A\big)^{m+1}.$$

The main goal of this choice is to allow the tails of the invariant density $f_{\boldsymbol{\theta}}$ and its derivative to be exponentially small.
It is evident that the function $S_{\boldsymbol{\theta}}(\cdot)$ defined in this way is m times differentiable and the m^{th} derivative is continuous. Consequently the density $f_{\boldsymbol{\theta}}(\cdot)$ is $m+1$ times continuously differentiable. So we can hope that at least for some values of parameter $\boldsymbol{\theta}$ the function $S_{\boldsymbol{\theta}}$ will belong to Σ.

Since for any fixed $\boldsymbol{\theta}$, the function $S_{\boldsymbol{\theta}}(\cdot)$ is locally Lipschitz and

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

there exists a unique strong solution of (3.1) corresponding to the trend coefficient $S = S_{\boldsymbol{\theta}}$ and diffusion coefficient $\sigma \equiv 1$. This solution possess ergodic properties, since the conditions of Proposition 1 (section 1.3) are satisfied.

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

$$\sigma_i = \sigma_{i,T} = \frac{2A}{T}\left(\left|\frac{\alpha}{i}\right|^m - 1\right)_+, \qquad i \neq 0, \tag{3.8}$$

with

$$\alpha = \alpha_T = A\left(\frac{TR(m+1)(2m+1)}{4m\pi^{2m}}\right)^{1/(2m+1)} \tag{3.9}$$

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

Let $\{\xi_i\}_{i\in\mathbb{Z}}$ be i. i. d. random variables such that

$$|\xi_i| < G, \quad \mathbf{E}\xi_i = 0, \quad \mathbf{E}\xi_i^2 = 1,$$

The law of the random variables ξ_i is supposed to be absolutely continuous w.r.t. the Lebesgue measure and their common probability density $p(x)$ is assumed to be differentiable. Moreover, we assume that the Fisher information is

$$I = \int \frac{[p'(x)]^2}{p(x)}\, dx = 1 + \varepsilon,$$

where $\varepsilon \to 0$ when $G \to \infty$. Since the Fisher information of the Gaussian distribution is equal to 1, one can choose $p(\cdot)$ as a smooth approximation with bounded support of the Gaussian density.

We introduce a prior distribution $\Lambda = \Lambda_T$ on Γ_T putting

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

with

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

for each i different from 0. The coefficient θ_0 will be deterministic and equal to 0. The density of random variable θ_i is then $\sigma_i^{-1/2} p\left(x\sigma_i^{-1/2}\right)$ and the corresponding Fischer information I_i is equal to $(1+\varepsilon)/\sigma_i(\varepsilon)$ (for $i \neq 0$).

Since the minimax risk is bounded below by the Bayesian one, we will try now to find a lower bound of the Bayesian risk defined by

$$B_T(\Lambda) = \inf_{\bar\vartheta_T} \mathbb{E} \int_{-\infty}^{\infty} \left(\bar\vartheta_T(x) - f'_{\boldsymbol{\theta}}(x)\right)^2 dx,$$

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

Let $\psi_{i,\boldsymbol{\theta}}$ and $\psi_{i,T}$ be the Fourier coefficients on $[-A,A]$ of the derivative $f'_{\boldsymbol{\theta}}$ and its estimator $\bar{\vartheta}_T$ respectively, *i. e.*,

$$\psi_{i,\boldsymbol{\theta}} = \int_{-A}^{A} f'_{\boldsymbol{\theta}}(x) e_i(x) dx,$$
$$\psi_{i,T} = \int_{-A}^{A} \bar{\vartheta}_T(x) e_i(x) dx.$$

Since $\{e_i(\cdot)\}$ is an orthonormal sequence and $f'_{\boldsymbol{\theta}}$ belongs to the Hilbert space generated by this sequence, using the Parseval's identity we obtain

$$\begin{aligned} B_T(\Lambda) &\geq \inf_{\bar{\vartheta}_T} \mathbb{E} \int_{-A}^{A} \left(\bar{\vartheta}_T(x) - f'_{\boldsymbol{\theta}}(x)\right)^2 dx \\ &\geq \inf_{\bar{\vartheta}_T} \sum_{0<|i|\leq L} \mathbb{E}\left(\psi_{i,T} - \psi_{i,\boldsymbol{\theta}}\right)^2 . \end{aligned}$$

Now we apply the van Trees inequality (see Theorem 2.1) with $d = 2$ to the coefficients θ_i and θ_{-i}. This leads to the inequality

$$B_T(\Lambda) \geq \sum_{0<i\leq L} \frac{\left(\mathbb{E}\left[\partial_{\theta_i}\psi_{i,\boldsymbol{\theta}} + \partial_{\theta_{-i}}\psi_{-i,\boldsymbol{\theta}}\right]\right)^2}{\mathbb{E}\left(I_i(\boldsymbol{\theta}) + I_{-i}(\boldsymbol{\theta})\right) + I_i + I_{-i}} \tag{3.10}$$

where, according to (2.10), the Fisher information $I_i(\boldsymbol{\theta})$ is

$$I_i(\boldsymbol{\theta}) = \mathbf{E}_{\boldsymbol{\theta}}\left[\partial_{\theta_i} \log f_{\boldsymbol{\theta}}(X_0) + \int_0^T \left[\partial_{\theta_i} S_{\boldsymbol{\theta}}(X_t)\right] dW_t\right]^2.$$

LEMMA 3.1. *For any $k = 0, 1, \ldots, m$, we have the following estimate:*

$$\sup_{\boldsymbol{\theta}\in\Gamma_T} \sup_{x\in[-A,A]} \left| \left(\int_0^x S_{\boldsymbol{\theta}}(v) dv \right)^{(k)} \right| \leq C A^{m+2} T^{-1/(4m+2)}. \tag{3.11}$$

PROOF. For $m = 0$, using the inequality $|\theta_i| \leq G\sqrt{\sigma}_i$ and the definitions (3.8), (3.9), we obtain

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

Remark that this chain of inequalities proves the bound (3.11) for $k = 1$ as well. Suppose that this bound holds for $1, 2, \ldots, k$ and prove it for $k+1$. Since all the derivatives of the function $U(\cdot)$ up to the order m are bounded, we have

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

The lemma is proved. □

COROLLARY 1. *The following relation holds*

$$f_{\boldsymbol{\theta}}(x) = \frac{1 + o_T(1)}{2A} \begin{cases} \exp\big\{ -2(|x| - A)^{m+2} \big\} & , \ \textit{if} \quad |x| > A, \\ 1 & , \ \textit{if} \quad |x| \le A. \end{cases}$$

PROOF. The proof is easy and follows immediately from Lemma 3.1. Indeed, for any $x \in [-A, A]$, we have

$$\exp\Big\{2\int_0^x S_{\boldsymbol{\theta}}(v)\,dv\Big\} = e^{o_T(1)} = 1 + o_T(1).$$

At the same time, for $x > A$, the following estimate is true:

$$\begin{aligned}\exp\Big\{2\int_0^x S_{\boldsymbol{\theta}}(v)\,dv\Big\} &= \exp\Big\{2\int_0^A S_{\boldsymbol{\theta}}(v)\,dv + 2\int_A^x S_{\boldsymbol{\theta}}(v)\,dv\Big\} \\ &= \big(1 + o_T(1)\big)\ \exp\Big\{ -2(x-A)^{m+2}\Big\}.\end{aligned}$$

In the same way, for $x < -A$, we obtain

$$\exp\Big\{2\int_0^x S_{\boldsymbol{\theta}}(v)\,dv\Big\} = \big(1 + o_T(1)\big)\ \exp\Big\{ -2(-x-A)^{m+2}\Big\}.$$

Integrating these equalities, one can check that

$$G(S_{\boldsymbol{\theta}}) = \frac{1 + o_T(1)}{2A}.$$

Corollary 1 is proved. □

COROLLARY 2. *For T large enough, we have*

$$I_i(\boldsymbol{\theta}) \leq \frac{T}{2A}\ (1+\varepsilon).$$

PROOF. First, we use the triangular inequality:

$$\sqrt{I_i(\boldsymbol{\theta})} \leq \left(\mathbf{E}_{\boldsymbol{\theta}}\left[\partial_{\theta_i}\log f_{\boldsymbol{\theta}}(X_0)\right]^2\right)^{\frac{1}{2}} + \left(\mathbf{E}_{\boldsymbol{\theta}}\left[\int_0^T \partial_{\theta_i} S_{\boldsymbol{\theta}}(X_t)\, dW_t\right]^2\right)^{\frac{1}{2}}.$$

Next, on the one hand, using the properties of stochastic integral, the Foubini theorem and the stationarity of the process X, we obtain

$$\mathbf{E}_{\boldsymbol{\theta}}\left[\int_0^T \partial_{\theta_i} S_{\boldsymbol{\theta}}(X_t)\, dW_t\right]^2 = \mathbf{E}_{\boldsymbol{\theta}}\left[\int_0^T \left(\partial_{\theta_i} S_{\boldsymbol{\theta}}(X_t)\right)^2 dt\right] = T\mathbf{E}_{\boldsymbol{\theta}}\left[\partial_{\theta_i} S_{\boldsymbol{\theta}}(\xi)\right]^2.$$

Since $S_{\boldsymbol{\theta}}$ depends on $\boldsymbol{\theta}$ only on the interval $[-A, A]$ where $f_{\boldsymbol{\theta}}$ is almost constant (Corollary 1) and the function U is always bounded by one, we have

$$\mathbf{E}_{\boldsymbol{\theta}}\left[\partial_{\theta_i} S_{\boldsymbol{\theta}}(\xi)\right]^2 = \int_{-A}^{A} e_i^2(x) U(A-|x|)\, f_{\boldsymbol{\theta}}(x) dx \leq \frac{1}{2A}\ \left(1+o_T(1)\right).$$

On the other hand, we have

$$\partial_{\theta_i} G(\boldsymbol{\theta}) = \int_{\mathbb{R}} \partial_{\theta_i} e^{2\int_0^x S_\theta(v)\, dv} dx = 2\int_{\mathbb{R}}\left[\int_0^x \partial_{\theta_i} S_{\boldsymbol{\theta}}(v)\, dv\right] e^{2\int_0^x S_\theta(v)\, dv} dx$$

and

$$\begin{aligned}
\partial_{\theta_i} f_{\boldsymbol{\theta}}(y) &= \frac{\partial_{\theta_i}\exp\left\{2\int_0^y S_{\boldsymbol{\theta}}(v)\, dv\right\}}{G(\boldsymbol{\theta})} - \frac{\partial_{\theta_i} G(\boldsymbol{\theta})\exp\left\{2\int_0^y S_{\boldsymbol{\theta}}(v)\, dv\right\}}{G^2(\boldsymbol{\theta})} \\
&= 2f_{\boldsymbol{\theta}}(y)\int_0^y \partial_{\theta_i} S_{\boldsymbol{\theta}}(v)\, dv - \frac{\partial_{\theta_i} G(\boldsymbol{\theta})}{G(\boldsymbol{\theta})}\ f_{\boldsymbol{\theta}}(y) \\
&= 2f_{\boldsymbol{\theta}}(y)\int_{\mathbb{R}}\left[\int_x^y \partial_{\theta_i} S_{\boldsymbol{\theta}}(v)\, dv\right] f_{\boldsymbol{\theta}}(x)\, dx. \qquad (3.12)
\end{aligned}$$

Using the fact that $\partial_{\theta_i} S_{\boldsymbol{\theta}}$ is bounded by one on the interval $[-A, A]$ and equals zero elsewhere, it can be easily checked that

$$\left|\partial_{\theta_i}\log[f_{\boldsymbol{\theta}}(y)]\right| = \left|\frac{\partial_{\theta_i} f_{\boldsymbol{\theta}}(y)}{f_{\boldsymbol{\theta}}(y)}\right| \leq 2A.$$

If we put all these estimates together, we obtain

$$\sqrt{I_i(\boldsymbol{\theta})} \leq 2A + \sqrt{\frac{T}{2A}\ \left(1+o_T(1)\right)}.$$

It is clear now that the desired inequality is satisfied for T large enough. □

We continue now the proof of Theorem 3.1. By virtue of the identity $f'_{\boldsymbol{\theta}} = 2S_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}$ and the equality (3.12), we have

$$\frac{\partial f'_{\boldsymbol{\theta}}}{\partial \theta_i}(x) = 2 f_{\boldsymbol{\theta}}(x) \frac{\partial S_{\boldsymbol{\theta}}}{\partial \theta_i}(x) + 2 f'_{\boldsymbol{\theta}}(x)\, \mathbf{E}_{\boldsymbol{\theta}} \int_{\xi}^{x} \frac{\partial S_{\boldsymbol{\theta}}}{\partial \theta_i}(v)\, dv.$$

Using Lemma 3.1, one can easily prove that

$$\sup_{|x|\le A} f'_{\boldsymbol{\theta}}(x)\, \mathbf{E}_{\boldsymbol{\theta}} \int_{\xi}^{x} \frac{\partial S_{\boldsymbol{\theta}}}{\partial \theta_i}(v)\, dv = o_T(A^{-2}), \tag{3.13}$$

which gives us

$$\frac{\partial f'_{\boldsymbol{\theta}}}{\partial \theta_i}(x) = 2 f_{\boldsymbol{\theta}}(x) e_i(x) U(A - |x|) + o_T(A^{-2}). \tag{3.14}$$

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

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

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

$$\frac{\partial \psi_{i,\boldsymbol{\theta}}}{\partial \theta_i} + \frac{\partial \psi_{-i,\boldsymbol{\theta}}}{\partial \theta_{-i}} = \frac{2}{A}\,(1 + o_T(1)). \tag{3.15}$$

So, the inequality (3.10) can be rewritten now as

$$B_T(\Lambda) \ge \frac{4\big(1 + o_T(1)\big)}{A(1+\varepsilon)} \sum_{0<i\le L} \frac{\sigma_i(\varepsilon)}{T\sigma_i(\varepsilon) + 2A}\ . \tag{3.16}$$

It is well known that the Riemann's sums of a continuous function on a finite interval converge to the integral of this function, consequently

$$\begin{aligned}\sum_{0<i\le L} \frac{\sigma_i(\varepsilon)}{T\sigma_i(\varepsilon) + 2A} &= \frac{1}{T} \sum_{0<i\le L} \left(\frac{i}{\alpha(1-\varepsilon)}\right)^m \left(\left(\frac{\alpha(1-\varepsilon)}{i}\right)^m - 1\right) \\ &= \alpha(1-\varepsilon) T^{-1}\big(1 + o_T(1)\big) \int_0^1 (1 - x^m)\, dx \\ &= \frac{\alpha(1-\varepsilon)m}{T(m+1)}\ \big(1 + o_T(1)\big).\end{aligned}$$

Replacing α by its expression (3.9) and using the evident inequality $(1+\varepsilon)^{-1} \geq 1-\varepsilon$, we obtain

$$B_T(\Lambda) \geq \frac{4\alpha m(1-\varepsilon)^2}{AT(m+1)} \big(1+o_T(1)\big)$$
$$= (4T^{-1})^{\frac{2m}{2m+1}} P(m,R) \big(1+o_T(1)\big)(1-\varepsilon)^2.$$

REMARK. It is important to emphasize that the values (3.8) and (3.9) are the solutions of the maximization problem related to the functional

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

over the set

$$\mathcal{E}(m,R) = \left\{ y = (y_i)_{i>0} \;\Big|\; 2A^{-2} \sum_{i>0} y_i \left(\frac{\pi i}{A}\right)^{2m} \leq R \right\}.$$

In the article [**28**] the authors explain heuristically how this maximization problem arise.

Thus we proved that the desired lower bound holds for the Bayesian risk (since ε can be chosen as small as we want). We have to show now that the chosen prior distribution is essentially concentrated on Σ, *i. e.,* that the Λ-probability of the set $\big\{\boldsymbol{\theta} \in \Gamma_T \,\big|\, S_{\boldsymbol{\theta}} \notin \Sigma\big\}$ tends to zero sufficiently fast. To do it, we need several auxiliary results.

LEMMA 3.2. *The functions* $\big\{S_{\boldsymbol{\theta}}(\cdot)\big\}_{\boldsymbol{\theta}\in\Gamma_T}$ *satisfy the conditions* (3.2)–(3.4)*, if* T *is sufficiently large.*

The proof of this lemma is very technical and quite long, that is why it is delayed to the end of this section.

LEMMA 3.3. *The following relation holds*

$$\mathcal{C}_T = \left\{ \boldsymbol{\theta} \in \Gamma_T \;\Big|\; \frac{1}{A^2} \sum_{|i|\leq L} \left(\frac{\pi i}{A}\right)^{2m} \theta_i^2 < R(1-\varepsilon) \right\} \subseteq \left\{ \boldsymbol{\theta} \;\Big|\; S_{\boldsymbol{\theta}} \in \Sigma \right\}, \qquad (3.17)$$

if T *is large enough.*

PROOF. Since $\mathcal{C}_T$ is a subset of Γ_T, the conditions (3.2) – (3.4) are satisfied for any $\boldsymbol{\theta} \in \mathcal{C}_T$. So, only the condition

$$\sup_{\boldsymbol{\theta}\in\mathcal{C}_T} \int_{\mathbb{R}} \big[f_{\boldsymbol{\theta}}^{(m+1)}(x)\big]^2 dx \leq R$$

needs to be checked. Note that

$$f_{\boldsymbol\theta}^{(m+1)}(x) = \big[2S_{\boldsymbol\theta}^{(m)}(x) + P\big(S_{\boldsymbol\theta}^{(m-1)}(x),\dots,S_{\boldsymbol\theta}(x)\big)\big]f_{\boldsymbol\theta}(x),$$

where $P(z_1,\dots,z_k)$ is a polynomial. By Lemma 3.1

$$\sup_{x\in[-A,A]} S_{\boldsymbol\theta}^{(k)}(x) = o_T(1), \qquad k=0,1,\dots,m-1. \tag{3.18}$$

So, on the one hand,

$$\sup_{x\in[-A,A]} \big[P\big(S_{\boldsymbol\theta}^{(m-1)}(x),\dots,S_{\boldsymbol\theta}(x)\big)\big]^2 = o_T(1).$$

On the other hand, using the fact that all the derivatives of U up to the order m are bounded and vanish for $x\notin[0,1]$, we obtain

$$\begin{aligned}\int_{-A}^{A} 4\big[S_{\boldsymbol\theta}^{(m)}(x) f_{\boldsymbol\theta}(x)\big]^2 dx &= \frac{1+o_T(1)}{A^2}\int_{-A}^{A}\big[S_{\boldsymbol\theta}^{(m)}(x)\big]^2dx \\ &= \frac{1+o_T(1)}{A^2}\sum_{|i|\le L}\Big(\frac{\pi i}{A}\Big)^{2m}\theta_i^2 \\ &\le R(1-\varepsilon)\big(1+o_T(1)\big)\end{aligned}$$

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

$$\begin{aligned}\int_{-A}^{A}\big[f_{\boldsymbol\theta}^{(m+1)}(x)\big]^2dx &= 4\int_{-A}^{A}\big[S_{\boldsymbol\theta}^{(m)}(x)f_0(x)\big]^2dx\big(1+o_T(1)\big) \\ &\le R(1-\varepsilon)\big(1+o_T(1)\big).\end{aligned} \tag{3.19}$$

It can be easily checked that

$$\int_{|x|>A}\big[f_{\boldsymbol\theta}^{(m+1)}(x)\big]^2dx \le CA^{-2}. \tag{3.20}$$

Combining (3.19) and (3.20) we obtain

$$\int_{-\infty}^{\infty}\big[f_{\boldsymbol\theta}^{(m+1)}(x)\big]^2dx \le R(1-\varepsilon)+o_T(1).$$

This completes the proof of Lemma 3.3. □

LEMMA 3.4. *The probability of the event $S_{\boldsymbol\theta}\notin\Sigma$ decreases exponentially to zero. Consequently*

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

Proof. The relation (3.17) implies that

$$\Lambda(S_{\boldsymbol\theta} \notin \Sigma) \le \Lambda(\boldsymbol\theta \notin \mathcal{C}_T).$$

We use now the shortened notation $\alpha_\varepsilon = \alpha(1-\varepsilon)$. Then we have

$$\begin{aligned}
\frac{1}{A^2}\,\mathbf{E}\Big[\sum_{|i|\le L}\Big(\frac{\pi i}{A}\Big)^{2m}\theta_i^2\Big] &= \frac{1}{A^2}\sum_{|i|\le L}\Big(\frac{\pi i}{A}\Big)^{2m}\sigma_i(\varepsilon)\\
&= \frac{2}{A^2}\sum_{i=1}^{L}\Big(\frac{\pi i}{A}\Big)^{2m}\frac{2A}{T}\bigg(\Big|\frac{\alpha(1-\varepsilon)}{i}\Big|^m - 1\bigg)\\
&= \frac{4\pi^{2m}\alpha_\varepsilon^{2m+1}}{TA^{2m+1}}\sum_{i=1}^{L}\frac{1}{\alpha_\varepsilon}\Big(\frac{i}{\alpha_\varepsilon}\Big)^{2m}\bigg(\Big|\frac{\alpha_\varepsilon}{i}\Big|^m - 1\bigg)\\
&= \frac{4\pi^{2m}\alpha_\varepsilon^{2m+1}}{TA^{2m+1}}\int_0^1 (x^m - x^{2m})\,dx\,\big(1+o_T(1)\big)\\
&= R(1-\varepsilon)^{2m+1}\big(1+o_T(1)\big).
\end{aligned}$$

Hence, for T sufficiently large,

$$\frac{1}{A^2}\,\mathbf{E}\Big[\sum_{|i|\le L}\Big(\frac{\pi i}{A}\Big)^{2m}\theta_i^2\Big] \le R(1-\varepsilon)^2.$$

Furthermore, according to Hoeffding's inequality, we obtain the following upper estimate:

$$\begin{aligned}
\Lambda(\boldsymbol\theta \notin \mathcal{C}_T) &= \Lambda\Big\{\frac{1}{A^2}\sum_{|i|\le L}\Big(\frac{\pi i}{A}\Big)^{2m}\theta_i^2 \ge R(1-\varepsilon)\Big\}\\
&\le \Lambda\Big\{\frac{1}{A^2}\sum_{|i|\le L}\Big(\frac{\pi i}{A}\Big)^{2m}(\theta_i^2 - \mathbf{E}\theta_i^2) \ge R\varepsilon(1-\varepsilon)\Big\}\\
&\le \exp\Big\{-\frac{R^2\varepsilon^2(1-\varepsilon)^2}{2Q}\Big\},
\end{aligned}$$

where

$$\begin{aligned}
Q &= \frac{G^4}{A^4}\sum_{|i|\le L}\Big(\frac{\pi i}{A}\Big)^{4m}\sigma_i^2 \le \frac{C}{T^2}\sum_{|i|\le \alpha} i^{4m}\bigg(\Big|\frac{\alpha}{i}\Big|^m - 1\bigg)^2 \le \frac{C}{T^2}\sum_{|i|\le\alpha} i^{2m}\alpha^{2m}\\
&\le CT^{-2}\alpha^{4m+1} = CT^{-1/(2m+1)}.
\end{aligned}$$

So, we have

$$\Lambda(S_{\boldsymbol\theta} \notin \Sigma) \le \Lambda(\boldsymbol\theta \notin \mathcal{C}_T) \le \exp\{-CT^{1/(2m+1)}\}.$$

Lemma 3.4 is proved. □

Now everything is ready to finish the proof of Theorem 3.1.

Note firstly that we can consider only the estimators $\bar\vartheta_T$ satisfying the inequality

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

with $f_0(\cdot) = f_{S_0}(\cdot)$ and S_0 is the function $S_{\boldsymbol\theta}$ corresponding to the value $\boldsymbol\theta = 0$. The set of all estimators satisfying this inequality will be denoted by $\mathcal{W}_T$. For estimators do not belonging to $\mathcal{W}_T$ the result is evident.

We have the following obvious inequalities:

$$\begin{aligned}\inf_{\bar\vartheta_T\in\mathcal{W}_T} \sup_{S\in\Sigma} R_T(\bar\vartheta_T, f_S') &\geq \inf_{\bar\vartheta_T\in\mathcal{W}_T} \sup_{\boldsymbol\theta\in\Theta} R_T(\bar\vartheta_T, f_{\boldsymbol\theta}') \\ &\geq \inf_{\bar\vartheta_T\in\mathcal{W}_T} \int_{\Gamma_T} R_T(\bar\vartheta_T, \boldsymbol\theta)\Lambda(d\boldsymbol\theta) - \sup_{\bar\vartheta_T\in\mathcal{W}_T} \int_{\Gamma_T\setminus\Theta} R_T(\bar\vartheta_T, \boldsymbol\theta)\Lambda(d\boldsymbol\theta) \\ &\geq B_T(\Lambda) - \sup_{\bar\vartheta_T\in\mathcal{W}_T} \int_{\Gamma_T\setminus\Theta} R_T(\bar\vartheta_T, \boldsymbol\theta)\Lambda(d\boldsymbol\theta)\end{aligned}$$

where $\Theta = \{\boldsymbol\theta \in \Gamma_T \,|\, S_{\boldsymbol\theta} \in \Sigma\}$. We have already found a lower bound for the first term. The second term can be bounded as follows

$$\sup_{\bar\vartheta_T\in\mathcal{W}_T} \int_{\Gamma_T\setminus\Theta} R_T(\bar\vartheta_T, \boldsymbol\theta)\Lambda(d\boldsymbol\theta) \leq \big(8 \sup_{\boldsymbol\theta\in\Gamma_T} \|f_{\boldsymbol\theta}'\|_2^2 + 2\big)\Lambda(S_{\boldsymbol\theta} \notin \Sigma).$$

It follows from Lemma 3.1 that the L^2 norm of $f_{\boldsymbol\theta}'$ is bounded uniformly on $\boldsymbol\theta \in \Gamma_T$. Consequently, using Lemma 3.4 we obtain

$$\inf_{\bar\vartheta_T\in\mathcal{W}_T} \sup_{S\in\Sigma} R_T(\bar\vartheta_T, f_S') \geq B_T(\Lambda) - o(T^{-1}).$$

Therefore

$$\varliminf_{T\to\infty} \inf_{\bar\vartheta_T} \sup_{S\in\Sigma} T^{\frac{2m}{2m+1}} R_T(\bar\vartheta_T, S) \geq 4^{\frac{2m}{2m+1}} P(m,R)(1-\varepsilon).$$

This proves the inequality (3.6), since ε can be taken as small as we want.

Proof of Lemma 3.2. We show first that there exists a constant D such that

$$\sup_{\boldsymbol\theta\in\Gamma_T} \int_{\mathbb{R}} \left[f_{\boldsymbol\theta}'(u)\right]^2 \mathbf{E}_{\boldsymbol\theta} \left[\frac{\mathbb{1}_{\{\xi>u\}} - F_{\boldsymbol\theta}(\xi)}{f_{\boldsymbol\theta}(\xi)}\right]^2 du < D \tag{3.21}$$

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

$$(z-A)^{m+2} - (y-A)^{m+2} \geq (z-y)(y-A)^{m+1},$$

consequently we have the estimate

$$\frac{1-F_{\boldsymbol{\theta}}(y)}{f_{\boldsymbol{\theta}}(y)} \leq \int_y^\infty \exp\left\{-(z-y)(y-A)^{m+1}\right\}\, dz = \frac{1}{(y-A)^{m+1}}\ . \tag{3.22}$$

This leads us to the inequality

$$\begin{aligned}\int_A^\infty {f'_{\boldsymbol{\theta}}}^2(u)\mathbf{E}_{\boldsymbol{\theta}}\left[\left(\frac{1-F_{\boldsymbol{\theta}}(\xi)}{f_{\boldsymbol{\theta}}(\xi)}\right)^2 \mathbb{1}_{\{\xi>u\}}\right]du &\leq \int_A^\infty {f'_{\boldsymbol{\theta}}}^2(u)\,\mathbf{E}_{\boldsymbol{\theta}}\left[\frac{\mathbb{1}_{\{\xi>u\}}}{(\xi-A)^{2m+2}}\right]du \\ &\leq \int_A^\infty {f'_{\boldsymbol{\theta}}}^2(u)\,(u-A)^{-2m}\,du < 1 \end{aligned} \tag{3.23}$$

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

$$(y-A)^{m+2} \leq (x-A)^{m+1}(y-A), \qquad \text{for}\ \ A\leq y\leq x,$$

implies that, for any $u>A$,

$$f_{\boldsymbol{\theta}}(u)\int_A^u \frac{F^2_{\boldsymbol{\theta}}(y)}{f_{\boldsymbol{\theta}}(y)}\,dy \leq \int_A^u \frac{f_{\boldsymbol{\theta}}(u)}{f_{\boldsymbol{\theta}}(y)}\,dy \leq \int_A^u e^{(u-A)^{m+1}(y-u)}dy \leq \frac{1}{(u-A)^{m+1}}.$$

Hence

$$\begin{aligned}\int_A^\infty {f'_{\boldsymbol{\theta}}}^2(u)\mathbf{E}_{\boldsymbol{\theta}}\left[\left(\frac{F_{\boldsymbol{\theta}}(\xi)}{f_{\boldsymbol{\theta}}(\xi)}\right)^2 \mathbb{1}_{\{A<\xi<u\}}\right]du &\leq C\int_A^\infty {f'_{\boldsymbol{\theta}}}^2(u)\,|f'_{\boldsymbol{\theta}}(u)|^{-1}du \\ &= C\int_A^\infty |f'_{\boldsymbol{\theta}}(u)|\,du = Cf_{\boldsymbol{\theta}}(A) < 1\end{aligned} \tag{3.24}$$

if T is large enough. Then, we have

$$\int_{-A-1}^A \frac{F^2_{\boldsymbol{\theta}}(y)}{f_{\boldsymbol{\theta}}(y)}\,dy \leq \int_{-A-1}^A \frac{1}{f_{\boldsymbol{\theta}}(y)}\,dy = 4A^2\big(1+o_T(1)\big)$$

and

$$\int_A^\infty {f'_{\boldsymbol{\theta}}}^2(u)\ \mathbf{E}_{\boldsymbol{\theta}}\left[\left(\frac{F_{\boldsymbol{\theta}}(\xi)}{f_{\boldsymbol{\theta}}(\xi)}\right)^2 \mathbb{1}_{\{-A-1<\xi<A\}}\right]du \leq 8A^2\int_A^\infty {f'_{\boldsymbol{\theta}}}^2(u)du < C. \tag{3.25}$$

Proceeding like in the proof of (3.22), one can show that

$$\int_{-\infty}^{-A-1} \frac{F^2_{\boldsymbol{\theta}}(y)}{f_{\boldsymbol{\theta}}(y)}\,dy \leq \sup_{y\leq -A-1}\frac{F^2_{\boldsymbol{\theta}}(y)}{f^2_{\boldsymbol{\theta}}(y)} \leq \frac{1}{(A+1-A)^{2m+2}} = 1. \tag{3.26}$$

Thus,

$$\int_A^\infty {f'_{\boldsymbol{\theta}}}^2(u)\,\mathbf{E}_{\boldsymbol{\theta}}\left[\left(\frac{F_{\boldsymbol{\theta}}(\xi)}{f_{\boldsymbol{\theta}}(\xi)}\right)^2 \mathbb{1}_{\{\xi<A\}}\right]du \leq 1, \tag{3.27}$$

for T large enough. Combining the inequalities (3.23), (3.24) and (3.27), we obtain

$$\int_A^\infty {f'_{\boldsymbol{\theta}}}^2(u)\,\mathbf{E}_{\boldsymbol{\theta}}\left[\left(\frac{\mathbb{1}_{\{\xi>u\}}-F_{\boldsymbol{\theta}}(\xi)}{f_{\boldsymbol{\theta}}(\xi)}\right)^2\right]du \leq C.$$

In the same way it can be shown that

$$\int_{-\infty}^{-A} f_{\boldsymbol{\theta}}'^{2}(u)\,\mathbf{E}_{\boldsymbol{\theta}}\left[\left(\frac{\mathbb{1}_{\{\xi>u\}}-F_{\boldsymbol{\theta}}(\xi)}{f_{\boldsymbol{\theta}}(\xi)}\right)^2\right]du \le C.$$

For $u \in [-A, A]$, using Lemma 3.1, we have

$$f_{\boldsymbol{\theta}}'(u) = \frac{o_T(1)}{2A}$$

and, by virtue of (3.22) and (3.26),

$$\mathbf{E}_{\boldsymbol{\theta}}\left[\left(\frac{\mathbb{1}_{\{\xi>u\}}-F_{\boldsymbol{\theta}}(\xi)}{f_{\boldsymbol{\theta}}(\xi)}\right)^2\right] \le \mathbf{E}_{\boldsymbol{\theta}}\left[\left(\frac{1-F_{\boldsymbol{\theta}}(\xi)}{f_{\boldsymbol{\theta}}(\xi)}\right)^2\mathbb{1}_{\{\xi>A+1\}}\right] + \mathbf{E}_{\boldsymbol{\theta}}\left[\frac{\mathbb{1}_{\{|\xi|\le A+1\}}}{f_{\boldsymbol{\theta}}^2(\xi)}\right]$$
$$+\mathbf{E}_{\boldsymbol{\theta}}\left[\left(\frac{F_{\boldsymbol{\theta}}(\xi)}{f_{\boldsymbol{\theta}}(\xi)}\right)^2\mathbb{1}_{\{\xi<-A-1\}}\right] \le CA^2.$$

It follows from these estimates that

$$\int_{-A}^{A} f_{\boldsymbol{\theta}}'^{2}(u)\,\mathbf{E}_{\boldsymbol{\theta}}\left[\left(\frac{\mathbb{1}_{\{\xi>u\}}-F_{\boldsymbol{\theta}}(\xi)}{f_{\boldsymbol{\theta}}(\xi)}\right)^2\right]du \le C$$

and the inequality (3.21) is proved.

We pass now to the proof of the inequality (3.3). For this, we need the following estimate:

$$\int_{A}^{\infty}\left[f'_{\boldsymbol{\theta}}(u)\right]^2\mathbf{E}_{\boldsymbol{\theta}}\left[\int_0^{\xi}\frac{\mathbb{1}_{\{y>u\}}-F_{\boldsymbol{\theta}}(y)}{f_{\boldsymbol{\theta}}(y)}\,dy\right]^2du < DA^2. \tag{3.28}$$

To prove it, one has to consider the cases $\xi < -A$, $\xi \in [-A, A]$, $\xi \in [A, u]$ and $\xi > u$. In the first, third and forth cases (3.28) can be obtained proceeding like in the proof of (3.21). In the second case, we have

$$\begin{aligned}\int_{A}^{\infty}\left[f'_{\boldsymbol{\theta}}(u)\right]^2\,\mathbf{E}_{\boldsymbol{\theta}}&\left[\int_0^{\xi}\frac{\mathbb{1}_{\{y>u\}}-F_{\boldsymbol{\theta}}(y)}{f_{\boldsymbol{\theta}}(y)}\,dy\;\mathbb{1}_{\{-A\le\xi\le A\}}\right]^2du\\ &\le \int_A^{\infty}\left[f'_{\boldsymbol{\theta}}(u)\right]^2\left[\int_{-A}^{A}\frac{1}{f_{\boldsymbol{\theta}}(y)}\,dy\right]^2du\\ &\le CA^4\int_A^{\infty}\left[f'_{\boldsymbol{\theta}}(u)\right]^2du = DA^2.\end{aligned} \tag{3.29}$$

For $B \le A$, the condition (3.3) is obviously fulfilled, since Lemma 3.1 implies that

$$\begin{aligned}\int_{-B}^{B}\left[f'_{\boldsymbol{\theta}}(u)\right]^2\mathbf{E}_{\boldsymbol{\theta}}\left[\int_0^{\xi}\frac{\mathbb{1}_{\{y>u\}}-F_{\boldsymbol{\theta}}(y)}{f_{\boldsymbol{\theta}}(y)}\,dy\right]^2du &\le 16A^2\int_{-B}^{B}\left[f'_{\boldsymbol{\theta}}(u)\right]^2\mathbf{E}_{\boldsymbol{\theta}}|\xi|^2\,du\\ &\le o_T(1)B.\end{aligned}$$

Here we used the fact that on $[-A, A]$ the derivative $f'_{\boldsymbol{\theta}}$ decreases to zero faster than any negative power of A. For $B > A$, the condition (3.3) follows from two previous estimates, indeed

$$\int_{-B}^{B} \left[f'_{\boldsymbol{\theta}}(u)\right]^2 \mathbf{E}_{\boldsymbol{\theta}} \left[\int_0^{\xi} \frac{\mathbb{1}_{\{y>u\}} - F_{\boldsymbol{\theta}}(y)}{f_{\boldsymbol{\theta}}(y)}\, dy\right]^2 du \leq \int_{-A}^{A} + \int_{|x|>A} \leq DA^2 \leq DB^2.$$

So we proved the estimate (3.3).

It remains to check the condition (3.2). We suppose that $A > 2$ and $B_2 > 2$. If $A > B/2$, then

$$\int_B^{\infty} [f'_{\boldsymbol{\theta}}(x)]^2\, dx \leq \int_0^{\infty} [f'_{\boldsymbol{\theta}}(x)]^2\, dx = \frac{C}{A^2} \leq \frac{D}{B^2}\ .$$

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

$$\begin{aligned} \int_B^{\infty} [f'_{\boldsymbol{\theta}}(x)]^2\, dx &\leq C \int_B^{\infty} (x-A)^{2m+2} e^{-2(x-A)^{m+2}} dx \\ &\leq C \int_B^{\infty} (x-A)^{2m+3} e^{-2(x-A)^{m+2}} dx \\ &= D \int_{(B-A)^{m+2}}^{\infty} y e^{-2y} dy \leq D \int_{(B-A)}^{\infty} e^{-2y} dy \\ &\leq De^{-2(B-A)} < De^{-B} < \frac{D}{B^2}. \end{aligned} \tag{3.30}$$

This inequality completes the proof of Lemma 3.2 and Theorem 3.1. □

3.4. Asymptotically Efficient Estimator

Now we have a lower bound for the minimax risk. To prove its optimality we have to find an estimator attaining this bound. Recall that a traditional estimator of the derivative is the so called kernel-type estimator:

$$\bar{\vartheta}_{K,T}(x) = \frac{2}{T} \int_0^T K_T(x - X_t)\, dX_t.$$

One of the simplest "good" properties of this estimator is the following: if the kernel K is equal to the delta function at 0, then the estimator is unbiased. Indeed,

$$\mathbf{E}_S\big[\bar{\vartheta}_{K,T}(x)\big] = 2\mathbf{E}_S\big[K_T(x-\xi)S(\xi)\big] = 2S(x)f_S(x) = f'_S(x).$$

Of course, the kernel can not be a distribution, that is why the kernel function is usually chosen to be an approximation of delta function. However, in our setting, we need to modify slightly the kernel-type estimator in order that it would asymptotically efficient. So, we investigate in this section the behavior of the estimator

$$\vartheta_{K,T}(x) = \frac{2}{T}\int_0^T K_T(x - X_t)\, 1\!\!1_{\{|X_t|<B_T\}} dX_t,$$

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

$$K_T^*(x) = b_T^* K^*(x b_T^*) \tag{3.31}$$

with

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

and

$$b_T^* = \left(\frac{\pi R T (m+1)(2m+1)}{4m}\right)^{\frac{1}{2m+1}}.$$

We show in this section that, for $B_T = \sqrt{T}$, the estimator $\vartheta_{K^*,T}$ is asymptotically efficient, *i. e.*, this estimator achieves the lower bound obtained in the previous section. This means that the formula (3.31) gives the optimal kernel in the problem of the invariant density's derivative estimation. In the proof of the upper bound we do not use directly the form of the optimal bandwidth b_T^*, in order to show how it can be obtained as a solution of a minimization problem.

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

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

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

THEOREM 3.2. *We have*

$$\overline{\lim_{T\to\infty}}\ \inf_{\bar\vartheta_T}\sup_{S\in\Sigma}\ T^{\frac{2m}{2m+1}}\, R_T(\bar\vartheta_T, f_S') \le 4^{\frac{2m}{2m+1}} P(m,R).$$

PROOF. It is evident that

$$\inf_{\bar\vartheta_T}\sup_{S\in\Sigma}\ R_T(\bar\vartheta_T, f_S') \le \inf_{K\in L^2}\ \sup_{S\in\Sigma}\ R_T(\vartheta_{K,T}, f_S').$$

Hence, it is sufficient to find an upper bound of the (linear) risk

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

where $S(x) \in \Sigma$ is the unknown trend coefficient.

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

$$\varphi_S(\lambda) = \int_{\mathbb{R}} e^{i\lambda x} f'_S(x)\,dx, \quad \varphi_{K,T}(\lambda) = \int_{\mathbb{R}} e^{i\lambda x} \vartheta_{K,T}(x)\,dx,$$
$$\varphi_K(\lambda) = \int_{\mathbb{R}} e^{i\lambda x} K_T(x)\,dx, \quad \varphi_T(\lambda) = \frac{1}{T}\int_0^T e^{i\lambda X_t} 1\!\!1_{\{|X_t|<B_T\}} dX_t.$$

The condition $K \in L^2$ implies that $\theta_{K,T}$ is a.s. square integrable. That is why the Parseval's identity can be used:

$$R_T(K,S) = \frac{1}{2\pi}\,\mathbf{E}_S \int_{\mathbb{R}} |\varphi_{K,T}(\lambda) - \varphi_S(\lambda)|^2\,d\lambda.$$

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

$$\begin{aligned}
\varphi_{K,T}(\lambda) &= \frac{2}{T}\int_{\mathbb{R}} e^{i\lambda x}\int_0^T K_T(x - X_t)\,1\!\!1_{\{|X_t|<B_T\}}\,dX_t\;dx \\
&= \frac{2}{T}\int_0^T \int_{\mathbb{R}} e^{i\lambda x} K_T(x - X_t)\,dx\;1\!\!1_{\{|X_t|<B_T\}}\,dX_t \\
&= \frac{2}{T}\int_0^T e^{i\lambda X_t}\varphi_K(\lambda)\,1\!\!1_{\{|X_t|<B_T\}}\,dX_t = 2\varphi_K(\lambda)\,\varphi_T(\lambda).
\end{aligned}$$

By the way, this is the reason for which the kernel-type estimators substitute the linear estimators in the problems of nonparametric curve estimation.
So the linear quadratic risk can be rewritten as

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

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

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

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

LEMMA 3.5. *If the function $K_T(\cdot)$ is such that $|\varphi_K(\lambda)| \le 1$ for each λ, then there exists a constant C such that*

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

$$\big\|\, 2\varphi_K \mathbf{E}_S[\varphi_T] - \varphi_S\big\|_2 \le \big\|\big(\varphi_K - 1\big)\varphi_S\big\|_2 + \frac{C}{B_T},$$

for any $S \in \Sigma$.

PROOF. We prove here only the first inequality, the second one can be proved in the same way. Note that using the equalities (1.12)–(1.14) one can check that the following decomposition is true:

$$\begin{aligned} T\big(\varphi_T(\lambda) - \mathbf{E}_S\big[\varphi_T(\lambda)\big]\big) &= H_S(\lambda, X_T) - H_S(\lambda, X_0) \\ &\quad + \int_0^T \big[e^{i\lambda X_t} 1\!\!1_{\{|X_t|<B\}} - g_S(\lambda, X_t)\big]\, dW_t \end{aligned}$$

with

$$g_S(\lambda, y) = 2\int_{-B_T}^{B_T} e^{i\lambda u} f_S'(u)\, \frac{1\!\!1_{\{u<y\}} - F_S(y)}{f_S(y)}\, du,$$

and $H_S(\lambda, x) = \int_0^x g_S(\lambda, y)\, dy$. Consequently

$$\begin{aligned} \mathbf{Var}_S\big[\varphi_T(\lambda)\big] &= T^{-2}\mathbf{E}_S\, \Big| H_S(\lambda, X_T) - H_S(\lambda, X_0) \\ &\qquad + \int_0^T \big(e^{i\lambda X_t} 1\!\!1_{\{|X_t|<B_T\}} - g_S(\lambda, X_t)\big)\, dW_t\Big|^2. \end{aligned}$$

Using the triangular inequality we get

$$\int_{\mathbb{R}} |\varphi_K(\lambda)|^2 \mathbf{Var}_S\big[\varphi_T(\lambda)\big] d\lambda \le \big(\sqrt{A_1} + 2\sqrt{A_2} + \sqrt{A_3}\big)^2, \tag{3.32}$$

with

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

Using the properties of Itô integral and the evident inequality

$$\left| \, e^{i\lambda X_t} \mathbb{1}_{\{|X_t|<B_T\}} \right| \le 1,$$

we obtain $A_1 \le T^{-1} \|\varphi_K\|_2^2$. To estimate the term A_2 we use the fact that $|\varphi_K(\lambda)|$ is less than 1. Thus, according to Parseval's identity and condition (3.3) we have

$$\begin{aligned} A_2 &\le \frac{4}{T^2} \, \mathbf{E}_S \int_{\mathbb{R}} \left| \int_{-B_T}^{B_T} e^{i\lambda u} f_S'(u) \int_0^{\xi} \frac{\mathbb{1}_{\{y>u\}} - F_S(y)}{f_S(y)} \, dy \, du \right|^2 d\lambda \\ &= \frac{8\pi}{T^2} \int_{-B_T}^{B_T} \left[f_S'(u) \right]^2 \mathbf{E}_S \left[\int_0^{\xi} \frac{\mathbb{1}_{\{y>u\}} - F_S(y)}{f_S(y)} \, dy \right]^2 du \le \frac{C B_T^2}{T^2} \; . \end{aligned}$$

Repeating exactly the same arguments one can check that

$$A_3 \le C/T.$$

Lemma 3.5 is proved. □

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

$$\varphi_K(\lambda) = \varphi_b(\lambda) = \left(1 - \left| \frac{\lambda}{b} \right|^m \right)_+ ,$$

where $b = b_T$ is a positive number. As we will see below, the L^2-norms figuring in the right hand sides of the inequalities of Lemma 3.5 converge both to zero with the rate $T^{-m/(2m+1)}$. Therefore the choice $B_T = \sqrt{T}$ guarantees the following upper estimate for the quadratic risk:

$$R_T(K, f_S') \le L_T(\varphi_b, \varphi_S) \big(1 + o_T(1) \big),$$

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

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

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

$$\Phi = \left\{ \varphi \;\Big|\; \frac{1}{2\pi} \int_{-\infty}^{\infty} \lambda^{2m} |\varphi(\lambda)|^2 d\lambda \le R \right\}.$$

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

$$L_T(\varphi_b, \varphi_S) = \frac{2}{\pi T} \int_{-b}^{b} \left(1 - \left| \frac{\lambda}{b} \right|^m \right)^2 d\lambda + \frac{b^{-2m}}{2\pi} \int_{-b}^{b} |\lambda|^{2m} |\varphi_S(\lambda)|^2 d\lambda.$$

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

$$\int_{-b}^{b}\left(1-\left|\frac{\lambda}{b}\right|^m\right)^2 d\lambda = \frac{4bm^2}{(m+1)(2m+1)}.$$

It leads to the inequality

$$\inf_{b>0}\sup_{S\in\Sigma} L_T(\varphi_b,\varphi_S) \le \inf_{b>0}\left\{\frac{8m^2 b}{\pi T(m+1)(2m+1)} + Rb^{-2m}\right\} = \inf_{b>0} G(b).$$

The function $G(b)$ is continuously differentiable and strictly convex, consequently it attains the minimum at the point b_T^* satisfying the equation

$$\frac{8m^2}{\pi T(m+1)(2m+1)} = \frac{2mR}{(b_T^*)^{2m+1}},$$

which leads to

$$b_T^* = \left(\frac{R\pi T(m+1)(2m+1)}{4m}\right)^{\frac{1}{2m+1}}$$

and

$$\inf_{b>0} G(b) = G(b_T^*) = \frac{(2m+1)R}{(b_T^*)^{2m}} = P(m,R)\left(4T^{-1}\right)^{\frac{2m}{2m+1}} .$$

Theorem 3.2 is proved. □

3.5. Concluding Remarks

1. One can use exactly the same arguments to find the asymptotics of the minimax risk in the problem of higher order derivative estimation under the condition that it belongs to Σ. For the derivative of order l, one can check that the rate of convergence is $\varphi_T = T^{-\beta}$, and the optimal constant is

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

with $\beta = m/(2m+2l-1)$. The only modification to do in the proof of lower bound is to set

$$\alpha = \alpha_T = A\left(\frac{TR(m+2l-1)(2m+2l-1)}{4m\pi^{2m+2l-2}}\right)^{\frac{1}{2m+2l-1}}.$$

The rest follows from the relation

$$\partial_{\theta_i}\psi_{i,\boldsymbol{\theta}} \sim \frac{1}{A}\left(\frac{\pi i}{A}\right)^{l-1}.$$

Remark that for $l \geq 2$, we do not require anymore that m is greater than 1. The obtained asymptotics holds for any positive m. To prove that this lower bound is attainable, one can use the kernel type estimator $\vartheta_{K_*,T}$, where the Fourier transform of the optimal kernel is defined by

$$\varphi_{K_*}(\lambda) = (i\lambda)^{l-1}\left(1 - \left|\frac{\lambda}{b_*}\right|^m\right)_+ .$$

The optimal bandwidth is

$$b_* = \left(\frac{TR\pi(m+2l-1)(2m+2l-1)}{4m}\right)^{-\frac{1}{2m+2l-1}} .$$

Taking the inverse Fourier transform, we get

$$K_{*,T}(x) = b_*^l K_l(b_* x), \quad \text{with} \quad K_l(x) = \frac{1}{\pi}\int_0^1 u^{l-1}(1-u^m)\phi_l(ux)\,du,$$

where $\phi_l(x)$ is the $(l-1)^{th}$ derivative of $\cos x$.

2. Some words about the condition (3.4). We need it in order that the right hand side of the first inequality in Lemma 3.5 be asymptotically bounded by $T^{-1}\|\varphi_K\|_2^2$. But for this goal, the following weaker condition is enough:

$$\sup_{S\in\Sigma} B^{-\tau}\int_{-B}^{B} f'^2_S(x)\,\mathbf{E}_S\left[\frac{1\!\!1_{\{\xi>x\}} - F_S(\xi)}{f_S(\xi)}\right]^2 dx < \infty,$$

where τ is positive and less than $1/(2m+1)$. Theoretically, if we replace the condition (3.4) by this one, the parameter space becomes larger. But from the practical viewpoint, the parameter space Σ defined with the help of (3.4) includes all the interesting cases.

3. As Pinsker's theorem, our result also can be extended to more general ellipsoids. In the problem of density estimation based on i. i. d. data such a generalization is contained in the paper [**24**] of Golubev.

In our case, the exact formulation of the result is the following. Let $\rho(\cdot)$ be a positive, continuous and symmetric function on $\mathbb{R}$ satisfying the property $\rho(\lambda) \sim |\lambda|^\gamma/\sqrt{2\pi}$ as $\lambda \to \infty$. We define the ellipsoid $\mathcal{F}(\rho, R)$ as the set of all functions $g \in L^2(\mathbb{R})$ such that

$$\int_{\mathbb{R}} \rho^2(\lambda)\big|\,\varphi(g,\lambda)\big|^2 d\lambda \leq R,$$

where $\varphi(g,\cdot)$ denotes the Fourier transform of the function g. Let us denote by Σ_ρ the set of all trend coefficients $S \in \Sigma_0$ such that $f'_S \in \mathcal{F}(\rho, R)$ and assume that the conditions (3.2)–(3.4) are satisfied. Then, for $\gamma > 1$, we have

$$\inf_{\bar\vartheta_T} \sup_{S\in\Sigma_\rho} R_T(\bar\vartheta_T, f_S') \sim (4T^{-1})^{\frac{2\gamma}{2\gamma+1}} P(\gamma, R).$$

We give now some ideas of the proof. To obtain the lower bound, only the proof of Lemma 3.3 should be modified. One has to prove that

$$\sup_{\boldsymbol{\theta}\in\mathcal{C}_T} \int_{\mathbb{R}} \rho^2(\lambda) \big| \varphi(f_{\boldsymbol{\theta}}', \lambda)\big|^2 d\lambda \le R.$$

The sketch of the proof of this inequality is the following:

$$\begin{aligned}\int_{\mathbb{R}} \rho^2(\lambda) \big| \varphi(f_{\boldsymbol{\theta}}', \lambda)\big|^2 d\lambda &= \int_{|\lambda|\le \pi\alpha/A} \rho^2(\lambda) \big| \varphi(f_{\boldsymbol{\theta}}', \lambda)\big|^2 d\lambda + o_T(1) \\ &= \sum_{|k|\le\alpha} \frac{\pi}{A}\, \rho^2\Big(\frac{\pi k}{A}\Big) \Big| \varphi\Big(f_{\boldsymbol{\theta}}', \frac{\pi k}{A}\Big)\Big|^2 + o_T(1) \\ &\sim \sum_{0\le k\le\alpha} \frac{1}{A}\, \Big(\frac{\pi k}{A}\Big)^{2\gamma} \Big| \varphi\Big(f_{\boldsymbol{\theta}}', \frac{\pi k}{A}\Big)\Big|^2 + o_T(1).\end{aligned}$$

Using the integration by parts formula, Lemma 3.1 and the orthonormality of trigonometric basis, one can show that

$$\begin{aligned}\Big| \varphi\Big(f_{\boldsymbol{\theta}}', \frac{\pi k}{A}\Big)\Big| &\le 2 \Big| \int_{-A}^{A} \exp\Big\{\frac{i\pi k x}{A}\Big\}\, S_{\boldsymbol{\theta}}(x) f_{\boldsymbol{\theta}}(x)\, dx \Big| + \frac{o_T(1)}{k^{\gamma+1}} \\ &= \frac{2\sqrt{A}\, |\theta_k + i\theta_{-k}| \big(1 + o_T(1)\big)}{2A} + \frac{o_T(1)}{k^{\gamma+1}} \\ &\sim A^{-1/2} |\theta_k + i\theta_{-k}|.\end{aligned}$$

This estimate implies immediately the result of Lemma 3.3.

To prove the attainability of this lower bound, one should consider the kernel-type estimator $\vartheta_{K,T}$ such that

$$\varphi_K(\lambda) = \Big(1 - \frac{\rho(\lambda)}{|b|^\gamma}\Big)_+$$

and use the following inequality:

$$\begin{aligned}\int_{\mathbb{R}} \Big(1 - \frac{\rho(\lambda)}{|b|^\gamma}\Big)_+^2 d\lambda &\le 2b\varepsilon + b \int_{\varepsilon\le|u|\le 2} \Big(1 - \frac{\rho(bu)}{|b|^\gamma}\Big)_+^2 d\lambda \\ &\sim b\varepsilon + b\int_\varepsilon^2 (1 - u^\gamma)_+^2 du.\end{aligned}$$

Here we used the fact that the optimal bandwidth has to tend to infinity as $T \to \infty$. To complete the proof, one should argue that ε can be chosen as small as we want.

4. From the beginning of this chapter, we assume that the diffusion coefficient is identically equal to one. But one can obtain the same results for the process

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

where $\sigma(\cdot)$ is a known continuous function, bounded away from zero (that is $\inf_{x\in\mathbb{R}} |\sigma(x)| > 0$) and $m+1$ times differentiable. We assume also that it has at most linear growth which guarantees the existence and uniqueness of solution. In this setting, the quantity

$$\frac{\mathbb{1}_{\{y>x\}} - F_S(y)}{f_S(y)}$$

figuring in the conditions (3.3) and (3.4) should be replaced by

$$\frac{\mathbb{1}_{\{y>x\}} - F_S(y)}{\sigma^2(y) f_S(y)} .$$

The only thing that needs to be changed in the proof of lower bound is the definition of parameterization: one has to consider the family of trend coefficients $S^\sigma_{\boldsymbol{\theta}} = \sigma^2 S_{\boldsymbol{\theta}}$. The asymptotically efficient estimator is then

$$\vartheta_{K,T}(x) = \frac{2}{T} \int_0^T K_T(x - X_t)\, \frac{\mathbb{1}_{\{|X_t|<B_T\}}}{\sigma^2(X_t)}\, dX_t.$$

The rest of the proof does not change.

5. As a final remark we would like to justify the conditions (3.2)–(3.4). Of course, our arguments are only of heuristic character. Remind that Σ_0 was the set of all trend coefficients providing strong existence, uniqueness and ergodicity of the diffusion process (with $\sigma = 1$). Denote by Σ^{nr} the set of all trend coefficients such that the diffusion process exists and is null recurrent (for example, $S \equiv 0 \in \Sigma^{nr}$). It is clear that the sets Σ^{nr} and Σ_0 are disjoint, nevertheless when we take sup over $S \in \Sigma_0$ we touch also the boundary values which belong to Σ^{nr}. But for null recurrent processes we are not able to obtain asymptotically exact upper bounds, since the law of large numbers does not work. Thus we impose the conditions (3.2)–(3.4), in order to forbid the trend coefficient S to go near of some boundary elements of Σ^{nr}. It is worthy to emphasize that these conditions are not very restrictive. For example, all the trend coefficients S satisfying

$$\overline{\lim_{|x|\to\infty}}\; xS(x) < -2$$

fulfill these conditions (with some constants B_1, B_2, D).

CHAPTER 4

Estimation of the Derivative of Invariant Density: Local Minimax Approach

4.1. The Setting

In this chapter we continue the study of invariant density's derivative estimation developing the local minimax approach. As we have already said, our final goal is to use the derivative's estimator in the problem of trend coefficient estimation. It turns out that the asymptotically efficient estimator in the global minimax sense does not fit well to the problem of trend estimation. That is why in the present chapter we investigate the asymptotic behavior of the local minimax risk. We find a lower bound for it and construct an asymptotically efficient estimator in the local minimax sense. This estimator is used in Chapter 5 in order to estimate the trend coefficient.

The results of this chapter are proved in the following setting. As before, we observe a continuous path x^T of the random process X on the finite interval $[0, T]$. The underlying diffusion process X is defined by SDE

$$dX_t = S(X_t)\,dt + dW_t,\ X_0 = \xi,\ t \geq 0,$$

where the random initial value ξ follows the invariant law and is independent of the Brownian Motion W. The unknown parameter is the trend coefficient $S(\cdot)$ and we want to estimate the first order derivative function $f'_S(x) = 2S(x)f_S(x)$, where $f_S(x)$ is the invariant density defined by (1.7). We suppose that the trend coefficient S belongs to the δ-neighborhood $\Sigma_\delta(m, R, S_0)$ of a function S_0 defined in the following way:

1°. for any real x, we have $\left|S(x) - S_0(x)\right| \leq \delta$,

2°. the function $f'_S(x) - f'_{S_0}(x)$ belongs to the set $\mathcal{F}(m, R)$.

The "central" function S_0 is unknown for statistician. To simplify the notations we will write Σ_δ instead of $\Sigma_\delta(m, R, S_0)$ and $f_0(\cdot)$ instead of $f_{S_0}(\cdot)$.

The quality of estimation is measured with the help of the L_2-type risk

$$R_T(\bar{\vartheta}_T, f'_S) = \mathbf{E}_S \int_{\mathbb{R}} \left(\bar{\vartheta}_T(x) - f'_S(x)\right)^2 dx,$$

where $\bar{\vartheta}_T$ is an estimator of f'_S. We define the local minimax risk as

$$R_T(\Sigma_\delta) = \inf_{\bar{\vartheta}_T} \sup_{S\in\Sigma_\delta} R_T(\bar{\vartheta}_T, f'_S),$$

where the infimum is taken over all possible estimators $\bar{\vartheta}_T$ (functions on $\mathscr{C}_T \times \mathbb{R}$ which are $\mathscr{B}_T \otimes \mathscr{B}(\mathbb{R})$-measurable). This risk is called local because the supremum is taken over all trend coefficients $S(\cdot)$ which belong to a shrinking neighborhood of a central function $S_0(\cdot)$.

Assume that the trend coefficient S_0 satisfy the conditions of Proposition 1 of Section 1.3, *i. e.,*

$$\delta_0 = \overline{\lim_{|x|\to\infty}} S_0(x)\,\mathrm{sgn}(x) < 0. \tag{4.1}$$

Thus, for δ sufficiently small (more precisely, $\delta < -\delta_0$), it can be shown that $\Sigma_\delta(m, R, S_0)$ is a subset of Σ_0. We assume also that S_0 has a polynomial majorant, which means that there exist two positive numbers C and ν, such that

$$\left|S_0(x)\right| \leq C(1 + |x|^\nu) \qquad \forall x \in \mathbb{R}. \tag{4.2}$$

We prove in the next section that in this setting the local minimax risk has asymptotically the same lower bound as the global minimax risk. However, to prove the attainability of this bound, we need an additional condition on the central function $S_0(\cdot)$.

We assume that for some $\tau > 0$,

$$\int_{\mathbb{R}} |\lambda|^{2m+\tau} \left|\varphi_0(\lambda)\right|^2 d\lambda < \infty, \tag{4.3}$$

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 $(m+1)$-times continuously differentiable and the derivatives $S_0^{(k)}$, $k = 0, \ldots, m+1$ increase like polynomials, then this condition is satisfied with $\tau = 2$. Indeed, since the derivative $f_0^{(m+1)}$ is the product of a polynomial depending on $S_0, \ldots, S_0^{(m+1)}$ and f_0, it has exponentially decreasing tails. Consequently

$$\int_{\mathbb{R}} \lambda^{2m+2} \left|\varphi_0(\lambda)\right|^2 d\lambda = 2\pi \|f_0^{(m+1)}\|_2^2 < \infty.$$

4.2. Lower Bound

To establish a lower bound we follow [**23**]. 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 whole space Σ_δ. The dimension of this family is finite, but depends on T and increases to infinity when T tends to infinity. Then we use the fact that the minimax risk is bounded below by the Bayesian one, with respect to any prior distribution. The last step is to find the worst prior and to evaluate the corresponding Bayesian risk. We obtain the following result.

THEOREM 4.1. *Suppose that the central function $S_0(\cdot)$ is m-times differentiable and satisfies the conditions (4.1) and (4.2), then*

$$\varliminf_{\delta\to 0}\varliminf_{T\to\infty} T^{\frac{2m}{2m+1}}\, R_T(\Sigma_\delta) \geq 4^{\frac{2m}{2m+1}}\ P(m,R),$$

where $P(m,R)$ is the constant of Pinsker, that is

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

PROOF. The proof of this theorem is essentially the same as the proof of the lower bound obtained in the previous chapter. The main difference is in the definition of the parametric family. The parameterization which we use is inspired by [**23**].

We fix a positive number A supposed to be large and define the sequence of subintervals $J_k = \big[a_k - AT^{-\beta}, a_k + AT^{-\beta}\big[$ of interval $[-A, A[$, where $\beta = (2m+1)^{-1}$ and

$$a_k = 2kAT^{-\beta}, \qquad k = 0, \pm 1, \pm 2, \dots, \pm M.$$

Here $M = M_T$ is the greatest integer such that $J_M \subseteq [-A, A]$. It is evident that these intervals J_k are disjoint and cover "almost" whole interval $[-A, A[$. Let us introduce now the parameterization

$$S(\boldsymbol{\theta}, x) = S_0(x) + \sum_{|k|<M} \sqrt{\frac{2A}{T^\beta f_0(a_k)}}\ \sum_{|i|<L} \theta_{i,k}\phi_{i,k}(x),$$

where

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

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 $(m+1)$-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.

It is easy to show that the functions $S(\boldsymbol{\theta}, \cdot)$ are m-times continuously differentiable and coincide with S_0 outside of the interval $[-A, A]$. Note that the function $U(A - |x - a_k| T^\beta)$ is a smooth approximation of the indicator function $\mathbb{1}_{J_k}(x)$. Roughly speaking, on each interval J_k, the trend $S(\boldsymbol{\theta}, \cdot)$ is obtained from the trend coefficient used in the proof of lower bound for global minimax risk by scaling and shifting by properly chosen numbers.

The parametric space Γ_T that we consider is the set of all finite sequences $\left\{\theta_{i,k}\right\}_{|i| \leq L, |k| \leq M}$ such that $|\theta_{i,k}| \leq G\sqrt{\sigma_i}$ for all i and m. Here

$$\sigma_i = \sigma_{i,T} = \frac{1}{2AT^{2m\beta}} \left(\left| \frac{\alpha}{i} \right|^m - 1 \right)_+ , \qquad i \neq 0,$$

with

$$\alpha = A \left(\frac{R(m+1)(2m+1)}{4m\pi^{2m}} \right)^{1/(2m+1)}$$

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 with the speed T^β, as $T \to \infty$.

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

$$|\xi_{i,k}| < G, \quad \mathbf{E}\xi_{i,k} = 0, \quad \mathbf{E}\xi_{i,k}^2 = 1, \quad I = \int \frac{[p'(x)]^2}{p(x)} \, dx = 1 + \varepsilon,$$

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

We introduce a prior distribution Λ on Γ_T setting

$$\theta_{i,k} = \sqrt{\sigma_i(\varepsilon)} \ \xi_{i,k}$$

with

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

for all integers i different from 0. The coefficients $\theta_{0,m}$ will be deterministic and equal to 0. The Fisher information $I_{i,k}$ of this prior distribution with respect to $\theta_{i,k}$ (for $i \neq 0$) is then equal to $(1+\varepsilon)/\sigma_i(\varepsilon)$.

Since the minimax risk is bounded below by the Bayesian one, we are looking for a lower bound of the Bayesian risk defined by

$$B_T(\Lambda) = \inf_{\bar{\vartheta}_T} \mathbb{E} \int_{\mathbb{R}} \big(\bar{\vartheta}_T(x) - f'(\boldsymbol{\theta}, x)\big)^2 dx,$$

where $\mathbb{E}$ is the mathematical expectation with respect to the probability measure $\mathbf{P}_{\boldsymbol{\theta}}(dx^T) \times \Lambda(d\boldsymbol{\theta})$. Here $f(\boldsymbol{\theta}, x) = f_{S(\boldsymbol{\theta},\cdot)}(x)$, and $f'(\boldsymbol{\theta}, x)$ is the derivative of $f(\boldsymbol{\theta}, x)$ with respect to x.

Let $\psi_{i,k,\boldsymbol{\theta}}$ and $\psi_{i,k,T}$ be the Fourier coefficients on $[-A, A]$ of $f'(\boldsymbol{\theta}, \cdot)$ and $\bar{\vartheta}_T$ with respect to the orthonormal sequence

$$e_{i,k}(x) = \sqrt{T^\beta/A}\; e_i\big(T^\beta A^{-1}(x - a_k)\big)\,\mathbb{1}_{\{x\in J_k\}}, \quad k \in \mathbb{Z}.$$

This means that

$$\begin{aligned}
\psi_{i,k,\boldsymbol{\theta}} &= \int_{-A}^{A} f'(\boldsymbol{\theta}, x)\, e_{i,k}\,(x) dx, \\
\psi_{i,k,T} &= \int_{-A}^{A} \bar{\vartheta}_T(x)\, e_{i,k}(x)\, dx.
\end{aligned}$$

Let E be the linear subspace of $L^2[-A, A]$ generated by the family $\{e_{i,k}(\cdot)\}$. It is well known that the L^2-norm of any element h of $L^2[-A, A]$ is greater than the L^2-norm of its projection $pr_E(h)$ on the subspace E. Applying this fact to the function $\bar{\vartheta}_T(\cdot) - f'(\boldsymbol{\theta}, \cdot)$ and using the orthonormality of the sequence $\{e_{i,k}(\cdot)\}$, we get

$$\begin{aligned}
B_T(\Lambda) &\geq \inf_{\bar{\vartheta}_T} \mathbb{E} \int_{-A}^{A} \big(\bar{\vartheta}_T(x) - f'(\boldsymbol{\theta}, x)\big)^2 dx \\
&\geq \inf_{\bar{\vartheta}_T} \mathbb{E} \left\| pr_E\big(\bar{\vartheta}_T(\cdot) - f'(\boldsymbol{\theta}, \cdot)\big)\right\|^2_{L^2[-A,A]} \\
&= \inf_{\bar{\vartheta}_T} \sum_{|k|\leq M} \sum_{|i|\leq L} \mathbb{E}(\psi_{i,k,T} - \psi_{i,k,\boldsymbol{\theta}})^2.
\end{aligned}$$

We use now van Trees inequality to eliminate the inf over all possible estimators. So we get

$$B_T(\Lambda) \geq \sum_{|k|\leq M} \sum_{|i|\leq L} \frac{\big(\mathbb{E}\big[\partial_{\theta_{i,k}} \psi_{i,k,\boldsymbol{\theta}}\big]\big)^2}{\mathbf{E}\big[I_{i,k}(\boldsymbol{\theta})\big] + I_{i,k}}\;, \tag{4.4}$$

where $I_{i,k}(\boldsymbol{\theta})$ is defined by the formula

$$I_{i,k}(\boldsymbol{\theta}) = \mathbf{E}_{\boldsymbol{\theta}}\left[\partial_{\theta_{i,k}} \log f(\boldsymbol{\theta}, X_0) + \int_0^T \left[\partial_{\theta_{i,k}} S(\boldsymbol{\theta}, X_t)\right] dW_t\right]^2 .$$

Since $|a_k| \leq A$ and S_0 is a continuous function (consequently bounded on the interval $[-A, A]$), the following estimate holds

$$f_0(x) = f_0(a_k)\, \exp\left\{2 \int_{a_k}^x S_0(v)\, dv\right\} = f_0(a_k)\left(1 + o_T(1)\right),$$

for any $x \in J_k$. Repeating word by word the proof of Corollary 2, one can easily show that

$$I_{i,k}(\boldsymbol{\theta}) = T\big(1 + o_T(1)\big)\, \mathbf{E}_{\boldsymbol{\theta}}\left[\partial_{\theta_{i,k}} S(\boldsymbol{\theta}, \xi)\right]^2 \leq \frac{2AT\big(1+\varepsilon\big)}{T^\beta} .$$

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

$$f(\boldsymbol{\theta}, x) = f_0(x)\big(1 + o_T(1)\big),$$

where $o_T(1)$ is uniform on $x \in \mathbb{R}$ and $\boldsymbol{\theta} \in \Gamma_T$. Hence, proceeding like in the proof of global lower bound, we get

$$\begin{aligned}
\partial_{\theta_{i,k}} \psi_{i,k,\boldsymbol{\theta}} &= 2 \int_{J_k} \partial_{\theta_{i,k}} S(\boldsymbol{\theta}, x)\, f(\boldsymbol{\theta}, x)\, e_{i,k}(x)\, dx \big(1 + o_T(1)\big) \\
&= 2 \int_{J_k} \sqrt{\frac{2A}{T^\beta f_0(a_k)}}\; f_0(x)\, e_{i,k}^2(x)\, dx \big(1 + o_T(1)\big) \\
&= 2 \int_{J_k} \sqrt{\frac{2A f_0(a_k)}{T^\beta}}\; e_{i,k}^2(x)\, dx \big(1 + o_T(1)\big).
\end{aligned}$$

Using the obvious equality $\int_{J_k} e_{i,k}^2(x)\, dx = 1$, we obtain

$$\partial_{\theta_{i,k}} \psi_{i,k,\boldsymbol{\theta}} = 2\sqrt{2Af_0(a_k) T^{-\beta}}\; \big(1 + o_T(1)\big).$$

Now, the inequality (4.4) can be rewritten like follows:

$$B_T(\Lambda) \geq \big(1 + o_T(1)\big) \sum_{|k| \leq M} \sum_{|i| \leq L} \frac{8AT^{-\beta} f_0(a_k)}{2AT^{2m\beta}(1+\varepsilon) + I_{i,k}} .$$

Using the convergence of Riemann's sums, it is easy to show that

$$\sum_{|k| \leq M} 2AT^{-\beta} f_0(a_k) = \int_{-A}^{A} f_0(x)\, dx \big(1 + o_T(1)\big). \tag{4.5}$$

Hence, by virtue of the inequality $(1+\varepsilon)^{-1} \geq 1-\varepsilon$, we have

$$\begin{aligned} B_T(\Lambda) &\geq \big(1+o_T(1)\big)(1-\varepsilon)\int_{-A}^{A} f_0(x)\,dx \sum_{|i|\leq L} \frac{4\sigma_i}{2AT^{2m\beta}\sigma_i+1} \\ &\geq \frac{4(1-\varepsilon)^2}{AT^{2m\beta}}\int_{-A}^{A} f_0(x)\,dx \sum_{i=1}^{L}\left(1-\frac{i^m}{\alpha^m}\right) \\ &\geq \frac{4(1-\varepsilon)^2}{AT^{2m\beta}}\int_{-A}^{A} f_0(x)\,dx \int_1^{L+1}\left(1-\frac{x^m}{\alpha^m}\right)_+ dx. \end{aligned}$$

Now we choose the positive number A such that the $\int_{-A}^{A} f_0(x)\,dx > 1-\varepsilon$ and

$$\int_1^{L+1}\left(1-\frac{x^m}{\alpha^m}\right)_+ dx \geq \alpha\int_{\frac{1}{\alpha}}^{1}(1-u^m)du \geq \frac{(1-\varepsilon)\alpha m}{m+1}\ .$$

Replacing α by its value, after simple calculations we obtain

$$B_T(\Lambda) \geq (1-\varepsilon)^4(4T^{-1})^{\frac{2m}{2m+1}}P(m,R).$$

Using condition (4.2) and proceeding exactly like in the proof of Theorem 3.1, it can be shown that

$$\begin{aligned} &\int_{\mathbb{R}}\left[\big(2S(\boldsymbol{\theta},x)f(\boldsymbol{\theta},x)\ -\ 2S_0(x)f_0(x)\big)^{(m)}\right]^2 dx \\ &\qquad = \int_{\mathbb{R}} 4\big[S^{(m)}(\boldsymbol{\theta},x)-S_0^{(m)}(x)\big]^2 f_0^2(x)dx\,\big(1+o_T(1)\big) \\ &\qquad = \frac{4A}{T^\beta}\sum_{|k|\leq M}\sum_{|i|\leq L} f_0(a_k)\theta_{i,k}^2\left(\frac{\pi i T^\beta}{A}\right)^{2m}\big(1+o_T(1)\big). \end{aligned}$$

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

$$\mathcal{C}_T = \Big\{\boldsymbol{\theta} = \{\theta_{i,k}\}\ \Big|\ V_T(\boldsymbol{\theta}) \leq R(1-\varepsilon)\Big\}.$$

For T sufficiently large, the set $\{S(\boldsymbol{\theta},\cdot)\,|\,\boldsymbol{\theta}\in\mathcal{C}_T\}$ is included in the set $\Sigma_\delta(S_0)$. Moreover, using (4.5), we have

$$\begin{aligned} \mathbf{E}\big[V_T(\boldsymbol{\theta})\big] &= 4AT^{(2m-1)\beta}\sum_{|k|\leq M} f_0(a_k)\sum_{|i|\leq L}\sigma_i(\varepsilon)\left(\frac{\pi i}{A}\right)^{2m}\big(1+o_T(1)\big) \\ &\leq \sum_{|i|\leq L} 2T^{2m\beta}\sigma_i(\varepsilon)\left(\frac{\pi i}{A}\right)^{2m}\big(1+o_T(1)\big) \leq R(1-\varepsilon)^{2m}\big(1+o_T(1)\big). \end{aligned}$$

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

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

Thus, if we denote $\mathcal{W}_T=\big\{\bar\vartheta_T\,\big|\,R_T(\bar\vartheta_T,f_0')<1\big\}$ and $\Theta=\big\{\boldsymbol{\theta}\in\Gamma_T\,\big|\,S(\boldsymbol{\theta},\cdot)\in\Sigma_\delta\big\}$, then the following chain of inequalities holds

$$\begin{aligned}R_T(\Sigma_\delta)&\ge\inf_{\bar\vartheta_T\in\mathcal{W}_T}\sup_{S\in\Sigma_\delta}R_T(\bar\vartheta_T,f_S')\ge\inf_{\bar\vartheta_T\in\mathcal{W}_T}\sup_{\boldsymbol{\theta}\in\Theta}R_T(\bar\vartheta_T,f_S')\\&\ge\inf_{\bar\vartheta_T\in\mathcal{W}_T}\int_{\Gamma_T}R_T(\bar\vartheta_T,\boldsymbol{\theta})\,\Lambda(d\boldsymbol{\theta})-\sup_{\bar\vartheta_T\in\mathcal{W}_T}\int_{\Gamma_T\setminus\Theta}R_T(\bar\vartheta_T,\boldsymbol{\theta})\,\Lambda(d\boldsymbol{\theta})\\&\ge B_T(\Lambda)-\sup_{\bar\vartheta_T\in\mathcal{W}_T}\int_{\Gamma_T\setminus\Theta}R_T(\bar\vartheta_T,\boldsymbol{\theta})\,\Lambda(d\boldsymbol{\theta}).\end{aligned}$$

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

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

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

$$R_T(\Sigma_\delta)\ge B_T(\Lambda)-o(T^{-1})\ge(1-\varepsilon)^4(4T^{-1})^{2m\beta}P(m,R)\big(1-o_T(1)\big).$$

Therefore

$$\varliminf_{\delta\to0}\varliminf_{T\to\infty}T^{\frac{2m}{2m+1}}\,R_T(\Sigma_\delta)\ge4^{\frac{2m}{2m+1}}P(m,R)(1-\varepsilon)^4.$$

This completes the proof of the theorem concerning the lower bound, since ε can be taken as small as we want. $\square$

4.3. Asymptotically Efficient Estimator

Now we have a lower bound for the minimax risk. To show that this bound is sharp we have to find an estimator attaining it. In most models where the Pinsker's constant is obtained, the lower bound is attained by the estimators minimizing the linear risk. As it was said in previous chapter, in our problem the analogue of the linear risk is the risk over all kernel-type estimators. These considerations lead us to investigate the behavior of the kernel-type estimator

$$\bar\vartheta_{K,T}(x)=\frac{2}{T}\int_0^T K_T(x-X_t)\,dX_t,\tag{4.6}$$

where $K_T(\cdot)$ is an arbitrary squared integrable function. The corresponding risk is said to be linear because of the linearity of the Fourier transform of $\bar{\vartheta}_{K,T}$ with respect to the kernel-type function's Fourier transform.

Let us denote

$$K_T^*(x) = \frac{b_T^*}{\pi} \int_0^1 (1 - u^{m+\rho_T}) \cos(ub_T^* x)\, du \tag{4.7}$$

with

$$b_T^* = \left(\frac{\pi RT(m+1)(2m+1)}{4m} \right)^{\frac{1}{2m+1}}$$

and ρ_T tending to zero slower than $1/\log(T+1)$. One can choose, for example, $\rho_T = 1/\log\log(T+1)$ or $\rho_T = 1/\sqrt{\log(T+1)}$. We show in this section that the estimator $\bar{\vartheta}_{K^*,T}$ is asymptotically efficient, in the sense that this estimator achieves the lower bound obtained in the previous section. This means that the formula (4.7) gives an asymptotically optimal kernel in the problem of the first derivative of the invariant density estimation.

Before stating and proving the main theorem, we need a technical result. This result is proved in [**38**], but since this book is not yet published we present a sketch of the proof. To formulate this result, we need the following notations:

$$\Phi_S(x) = f_S^{2p}(x) \mathbf{E}_S \left[\frac{\mathbb{1}_{\{\xi > x\}} - F_S(\xi)}{f_S(\xi)} \right]^{2p},$$

$$\Psi_S(x) = f_S^{2p}(x) \mathbf{E}_S \left[\int_0^\xi \frac{\mathbb{1}_{\{y > x\}} - F_S(y)}{f_S(y)}\, \mathrm{d}y \right]^{2p},$$

where p is a number greater than 1. We denote by $V_\delta(S_0)$ the δ-neighborhood of S_0 in the supremum norm.

Lemma 4.1. *If the function S_0 satisfies the conditions (4.1) and (4.2), then 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 true:*

$$\Phi_S(x) \le Ce^{-\gamma|x|}, \tag{4.8}$$

$$\Psi_S(x) \le Ce^{-\gamma|x|}, \tag{4.9}$$

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 (4.1), one can find two positive constants B and γ such that

$$S_0(x)\operatorname{sgn} x < -2\gamma$$

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

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

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

$$\inf_{S\in V_\delta(S_0)} G(S) \ge \int_{|x|\le B} \exp\big\{-2D|x|\big\}\mathrm{d}x > 0. \tag{4.10}$$

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

$$m_B = \inf_{S\in V_\delta} \inf_{|x|\le B} f_S(x) > 0 \quad \text{and} \quad \sup_{S\in V_\delta} f_S(x) \le Ce^{-\gamma|x|}. \tag{4.11}$$

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 inequality (4.8) will be proved.

Using inequalities (4.11), for $x \in [-B, B]$, we obtain

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

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 f_S is a probability density, the last integral is less than one for any S. So the inequality (4.8) 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 $[x, -B]$, 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. □

We are well equipped now in order to state and to prove the main result of this section.

THEOREM 4.2. *If the function S_0 satisfies the conditions (4.1)–(4.3), then the kernel-type estimator $\bar\vartheta_{K^*,T}(\cdot)$ is asymptotically efficient, i. e.,*

$$\lim_{\delta\to 0}\lim_{T\to\infty}\sup_{S\in\Sigma_\delta} T^{\frac{2m}{2m+1}}\, R_T(\bar\vartheta_{K^*,T}, f'_S) = 4^{\frac{2m}{2m+1}} P(m,R).$$

PROOF. We have to find an upper bound of the linear risk

$$R_T(K, f'_S) = \mathbf{E}_S \int_{\mathbb{R}} \left(\bar\vartheta_{K,T}(x) - f'_S(x)\right)^2 dx,$$

where $S(x) \in \Sigma_\delta$ is the unknown trend coefficient. To evaluate this risk we will use the Fourier transformations. Let us denote

$$\varphi_S(\lambda) = \int_{\mathbb{R}} e^{i\lambda x} f'_S(x)\,dx, \qquad \varphi_{K,T}(\lambda) = \int_{\mathbb{R}} e^{i\lambda x}\bar\vartheta_{K,T}(x)\,dx,$$
$$\varphi_K(\lambda) = \int_{\mathbb{R}} e^{i\lambda x} K_T(x)\,dx, \qquad \varphi_T(\lambda) = \frac{1}{T}\int_0^T e^{i\lambda X_t} dX_t.$$

By Parseval's identity

$$R_T(K,S) = \frac{1}{2\pi}\,\mathbf{E}_S \int_{\mathbb{R}} \left|\,\varphi_{K,T}(\lambda) - \varphi_S(\lambda)\right|^2 d\lambda.$$

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

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

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

$$\begin{aligned}R_T(K,S) &= \frac{1}{2\pi}\,\mathbf{E}_S\int_{\mathbb{R}} |2\varphi_K(\lambda)\varphi_T(\lambda) - \varphi_S(\lambda)|^2 d\lambda\\ &= \frac{1}{2\pi}\int_{\mathbb{R}}\big(4|\varphi_K(\lambda)|^2\mathbf{Var}_S\left[\varphi_T(\lambda)\right] + |\varphi_K(\lambda) - 1|^2|\varphi_S(\lambda)|^2\big)d\lambda.\end{aligned}$$

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

$$\varphi_K(\lambda) = \varphi_b(\lambda) = \left(1 - \left|\frac{\lambda}{b}\right|^{m+\rho}\right)_+, \tag{4.12}$$

where $b = b_T$ is a positive number tending to infinity and $\rho = \rho_T > 0$ tends to zero so that

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

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 4.2. *For any $S \in \Sigma_\delta(S_0)$, we have*

$$\int_{\mathbb{R}} |\varphi_b(\lambda)|^2\mathbf{Var}_S\big[\varphi_T(\lambda)\big]\,d\lambda \le \frac{1}{T}\int_{\mathbb{R}} |\varphi_b(\lambda)|^2 d\lambda\big(1 + o_T(1)\big)\ .$$

PROOF. By definition of φ_T we have

$$\varphi_T(\lambda) = \frac{1}{T}\int_0^T e^{i\lambda X_t} S(X_t)\,dt + \frac{1}{T}\int_0^T e^{i\lambda X_t}\,dW_t.$$

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

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

converges to zero faster than

$$\frac{1}{T}\int_{\mathbb{R}} |\varphi_b(\lambda)|^2 d\lambda = \frac{Cb}{T},$$

the lemma will be proved. Let us prove that

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

Indeed, by virtue of formula (1.12), we have

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

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

$$H(x,u) = 2f_S(x)\int_0^u \left(\frac{\mathbb{1}_{\{y>x\}} - F_S(y)}{f_S(y)}\right) \mathrm{d}y, \tag{4.13}$$

$$g(x,u) = 2f_S(x)\left(\frac{\mathbb{1}_{\{u>x\}} - F_S(u)}{f_S(u)}\right). \tag{4.14}$$

Using once more Parseval's identity one can show that

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

where the first inequality is due to the fact that $|\varphi_b(\cdot)| \le 1$. The last two integrals are bounded uniformly on $S \in \Sigma_\delta$ because of Lemma 4.1 and the fact that the trend coefficient S has a polynomial majorant. This completes the proof of Lemma 4.2. □

We return to the proof of Theorem 4.2. By virtue of the last lemma, we have the following upper estimate for the quadratic risk

$$R_T(K, f_S') \le L_T(\varphi_b, \varphi_S)\big(1 + o_T(1)\big),$$

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

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

Since the function $S(x)$ is in the set $\Sigma_\delta(m,R,S_0)$, the Fourier transform φ_S should belong to the set

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

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

$$L_T(\varphi_b,\varphi_S) = \frac{2}{\pi T}\int_{-b}^{b}\Big(1-\Big|\frac{\lambda}{b}\Big|^{m+\rho}\Big)^2 d\lambda + \frac{1}{2\pi}\int_{-b}^{b}\Big|\frac{\lambda}{b}\Big|^{2m+2\rho}|\varphi_S(\lambda)|^2 d\lambda + \frac{1}{2\pi}\int_{|\lambda|>b}|\varphi_S(\lambda)|^2 d\lambda. \quad (4.15)$$

Since φ_S belongs to the set Φ, we have

$$\int_{-b}^{b}\Big|\frac{\lambda}{b}\Big|^{2m+2\rho}|\varphi_S(\lambda)-\varphi_0(\lambda)|^2 d\lambda + \int_{|\lambda|>b}|\varphi_S(\lambda)-\varphi_0(\lambda)|^2 d\lambda \le \frac{2\pi R}{b^{2m}}\ ,$$

and taking into account condition (4.3), for T large enough, we get

$$\int_{-b}^{b}\Big|\frac{\lambda}{b}\Big|^{2m+2\rho}|\varphi_0(\lambda)|^2 d\lambda + \int_{|\lambda|>b}|\varphi_0(\lambda)|^2 d\lambda \le \int_{\mathbb{R}}\Big|\frac{\lambda}{b}\Big|^{2m+2\rho}|\varphi_0(\lambda)|^2 d\lambda$$
$$\le \int_{\mathbb{R}}\frac{1+|\lambda|^{2m+\tau}}{b^{2m+2\rho}}\,|\varphi_0(\lambda)|^2 d\lambda \le \frac{C}{b^{2m+2\rho}}\ .$$

Thus, the sum of the second and the third terms of (4.15) is bounded by

$$\frac{R}{b^{2m}} + \frac{C}{b^{2m+\rho}} = \frac{R}{b^{2m}}\ \big(1+o_T(1)\big),$$

and the first term can be calculated explicitly:

$$\frac{2}{\pi T}\int_{-b}^{b}\Big(1-\Big|\frac{\lambda}{b}\Big|^{m+\rho}\Big)^2 d\lambda = \frac{8bm^2\big(1+o_T(1)\big)}{\pi T(m+1)(2m+1)}\ .$$

It is clear now that b_T should be of order $T^{1/(2m+1)}$ in order that we have the best rate of convergence for the risk of the estimator corresponding to (4.12). It leads us to the inequality

$$\inf_{b>0}\sup_{S\in\Sigma_\delta} L_T(\varphi_b,\varphi_S) \le \big(1+o_T(1)\big)\inf_{b>0}\Big[\frac{8m^2 b}{\pi T(m+1)(2m+1)} + \frac{R}{b^{2m}}\Big]$$
$$= \big(1+o_T(1)\big)\inf_{b>0} G(b).$$

The function $G(b)$ is continuously differentiable and strictly convex, consequently it attains the minimum at the point b^* satisfying the equation

$$\frac{8m^2}{\pi T(m+1)(2m+1)} = \frac{2mR}{(b^*)^{2m+1}},$$

which leads to

$$b^* = \left(\frac{R\pi T(m+1)(2m+1)}{4m} \right)^{\frac{1}{2m+1}}$$

and

$$\inf_{b>0} G(b) = G(b^*) = \frac{(2m+1)R}{(b^*)^{2m}} = 4^{\frac{2m}{2m+1}} P(m,R)\, T^{-\frac{2m}{2m+1}} \ .$$

By an elementary verification it can be shown that the function φ_{b_0} is the Fourier transform of the kernel-type function K_T^* defined by (4.7). We proved so that

$$\sup_{S \in \Sigma_\delta} R_T(K_T^*, S) \le (4T^{-1})^{\frac{2m}{2m+1}} P(m,R)\big(1 + o_T(1)\big).$$

This inequality, associated with the lower bound of Theorem 4.1, completes the proof of Theorem 4.2. □

4.4. Discussion

1. It is clear that to estimate a parameter belonging to a "small" set is easier than to do it for a large space of parameters. In spite of this fact we proved that the asymptotics of local minimax risk coincide with the asymptotics of the global one. Does it mean that the lower bound of Chapter 3 can be obtained as a consequence of Theorem 4.1 ? The answer of this question is positive. Moreover, the approach of this chapter permits to extend the result concerning the global minimax risk to the case $m = 1$ as well.

To prove this affirmation, let us fix a positive number A and denote

$$S_A(x) = \begin{cases} -(m+2)(x-A)^{m+1} & ,\ x > A, \\ 0 & ,\ x \in [-A, A], \\ (m+2)(-x-A)^{m+1} & ,\ x < -A. \end{cases} \tag{4.16}$$

It is evident that this function satisfy the condition (4.1) for any $A > 0$. Following the proof of Lemma 3.2 it can be shown that for sufficiently small values of δ, all the elements of the set $V_\delta(S_A)$ satisfy conditions (3.2)–(3.4) with some constants D, B_1, B_2 do not depending on A. At the same time, we have

$$\big\| f_S^{(m+1)} \big\|_2 \le \big\| f_A^{(m+1)} \big\|_2 + \big\| f_S^{(m+1)} - f_A^{(m+1)} \big\|_2$$

where we used the notation $f_A = f_{S_A}$. Remark now that the normalizing constant $G(S_A)$ is greater than $2A$, and the derivative $f_A^{(m+1)}$ is 0 on $[-A, A]$ and exponentially decreasing elsewhere. Consequently, its L^2-norm is less than CA^{-1}

for a constant C. It is evident that $\sqrt{R} - CA^{-1} = \sqrt{R - o_A(1)}$, where $o_A(1)$ tends to zero as $A \to \infty$. These estimates entail the following inclusion:

$$\Sigma_\delta(m, R - o_A(1), S_A) \subseteq \Sigma(m, R, D, B_1, B_2),$$

which leads to

$$\begin{aligned} \varliminf_{T\to\infty} T^{\frac{2m}{2m+1}} R_T\big(\Sigma(m, R, D, B_i)\big) &\geq \lim_{\delta\to 0} \varliminf_{T\to\infty} T^{\frac{2m}{2m+1}} R_T\big(\Sigma_\delta(m, R - o_A(1), S_A)\big) \\ &= 4^{\frac{2m}{2m+1}} P\big(m, R - o_A(1)\big). \end{aligned}$$

Since this inequality holds for arbitrarily large values of A, it implies the assertion of Theorem 3.1.

2. All the generalizations presented in Section 3.5 are true in the setting of this chapter as well.

CHAPTER 5

Trend Coefficient Estimation

5.1. The Problem

This chapter is devoted to the problem of trend coefficient estimation by continuous path observations. More precisely, we are given a growing bit $x^T = (x_t,\ t \in [0,T])$ of a realization of the random process X satisfying the SDE

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

and our goal is to recover the unknown trend coefficient $S(\cdot)$. To measure the quality of an estimator $\bar{S}_T(\cdot)$ we use the following risk function:

$$\mathcal{R}_T(\bar{S}_T, S) = \int_{\mathbb{R}} \left(\bar{S}_T(x) - S(x)\right)^2 f_S^2(x)\,dx,$$

where $f_S(\cdot)$ is the invariant density of the process X. This is a kind of mean integrated weighted squared error, where the weight function is the square of the invariant density.

We have already mentioned that the problem of trend estimation is closely related with the problems of invariant density and its derivative estimation because of the identity

$$S(x) = \frac{f_S'(x)}{2f_S(x)}. \tag{5.1}$$

Therefore, to estimate $S(x)$ we use asymptotically efficient estimators of $f_S'(x)$ and $f_S(x)$. The first estimator is constructed in the previous chapter, the results concerning asymptotically efficient estimation of invariant density are given in Section 5.3. Throughout this chapter we use the notations introduced in Chapter 4. We establish in Section 5.4 a lower bound for the local minimax risk of all possible estimators. Then we construct an estimator which attains this bound asymptotically. All these results are obtained under the conditions on S_0 introduced in Section 4.1.

The problem of the trend estimation was considered by many authors (see [**1**], [**45**], [**48**], [**18**], [**19**], [**7**], [**54**]). So Pham [**48**] discussed the properties of kernel type estimators and proved their consistency. He obtained also a CLT and studied the asymptotic behavior of the bias for these estimators. In [**45**] Nguyen and Pham proved the consistency of the sieve estimator for linear SDE's, *i. e.,* when the trend coefficient is $\vartheta(t)\, X_t$. Galtchouk and Pergamenschikov [**18**], [**19**] studied the trend estimation by sequential methods when the diffusion process is observed up to a stopping time. The authors constructed an adaptive estimator attaining the optimal adaptive rate of convergence. In a recent paper [**54**], Spokoiny investigated the estimation of trend by locally linear smoothers and proved that the risk of an adaptive procedure is within a factor $\log\log T$ of the risk of "ideal" estimator (corresponding to the optimal choice of smoothing parameter).

Note that there is another way to establish lower bounds for minimax risk. It consists in proving that the statistical experiment we are interested in is asymptotically equivalent to another (simpler) statistical experiment, under which the asymptotic behavior of the minimax risk is known. Some results concerning the equivalence for diffusion process were obtained by Milstein and Nussbaum [**41**], Delattre and Hoffmann [**8**], Genon-Catalot, Laredo and Nussbaum [**20**]. But the results of these papers do not entail a lower bound for our setting, since either they are proved in the asymptotics of small diffusion (*i. e.,* the observation time is fixed and the diffusion coefficient tends to zero) or the asymptotically equivalent experiment is not significantly simpler than our model.

Let us define now the parameter space $\widetilde{\Sigma}_\delta(m, R, S_0)$ we are dealing with. Let $\tilde{V}_\delta(S_0)$ be the set of all m-times continuously differentiable functions $S(\cdot)$ such that

$$\sup_{x\in\mathbb{R}} \left| S^{(i)}(x) - S_0^{(i)}(x) \right| \leq \delta, \tag{5.2}$$

for any $i = 0, 1, \ldots, m-1$. Then we denote

$$\tilde{\Sigma}_\delta(m, R, S_0) = \left\{ S \in \tilde{V}_\delta(S_0) \,\middle|\, \int_{\mathbb{R}} \left[\left(S - S_0\right)^{(m)}(x) \right]^2 f_S^2(x) dx \leq R \right\}.$$

This set is a weighted Sobolev space with weight function $f_S^2(\cdot)$. Very often we will write $\widetilde{\Sigma}_\delta(R)$ or $\widetilde{\Sigma}_\delta$ instead of $\tilde{\Sigma}_\delta(m, R, S_0)$.

In order to use the results of previous chapter in the problem of trend estimation we need to establish some relations between the parameter spaces Σ_δ (see Section 4.1) and $\widetilde{\Sigma}_\delta$. These results are stated and proved in the next section.

We suppose throughout this chapter that the center of localization S_0 is m-times differentiable and its m^{th} derivative has polynomial majorant, *i. e.*,

$$\left|S_0^{(m)}(x)\right| \le C(1+|x|^\nu) \tag{5.3}$$

for a couple of positive constants (C, ν) and for any $x \in \mathbb{R}$. This condition is obviously stronger than the condition (4.2).

5.2. Auxiliary Results

LEMMA 5.1. *If S_0 is m-times continuously differentiable and satisfies conditions (4.1) and (5.3), then the following relations hold:*

$$\sup_{S\in\widetilde{\Sigma}_\delta} G(S) = G(S_0)\big(1+o_\delta(1)\big), \tag{5.4}$$

$$\widetilde{\Sigma}_\delta(m, R, S_0) \subseteq \Sigma_\delta(m, 4R + r_\delta, S_0), \tag{5.5}$$

where r_δ tends to zero as $\delta \to 0$.

PROOF. We start by proving the first assertion of this lemma. Since condition (4.1) is fulfilled, there exist two positive constants B and γ such that

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

for any x such that $|x| > B$, and for any $S \in \tilde{V}_\delta(S_0)$. Let us denote by M the supremum of $S(x)$ when $x \in [-B, B]$ and $S \in \tilde{V}_\delta$. Then, for any positive number $A > B$, we have

$$\begin{aligned} G(S) &= \int_{\mathbb{R}} \exp\Big\{2\int_0^x S(v)\,dv\Big\}dx \\ &\le \int_{-A}^{A} \exp\Big\{2\int_0^x S(v)\,dv\Big\}dx + \int_{|x|>A} \exp\Big\{2\int_0^x S(v)\,dv\Big\}dx \\ &\le e^{2\delta A}\int_{-A}^{A} \exp\Big\{2\int_0^x S_0(v)\,dv\Big\}dx + e^{2MB}\int_{|x|>A} e^{-2\gamma(|x|-B)}dx \\ &\le e^{2\delta A}G(S_0) + Ce^{-2\gamma A}. \end{aligned} \tag{5.6}$$

These inequalities imply that

$$\lim_{\delta\to 0}\sup_{S\in\tilde{V}_\delta(S_0)} \frac{G(S)}{G(S_0)} \le 1 + Ce^{-2\gamma A}.$$

Since A can be chosen as large as we want, we get

$$\lim_{\delta\to 0}\sup_{S\in\tilde V_\delta(S_0)}\frac{G(S)}{G(S_0)}\le 1.$$

It remains to reverse the roles of S and S_0 in (5.6) in order to obtain the first assertion of this lemma. We prove now the inclusion (5.5). Let the function S belong to the set $\tilde\Sigma(m,R,S_0)$ and $P(x_0,\ldots,x_{m-1})$ be a polynomial such that

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

Then, by virtue of conditions (5.2) and (5.3), we have

$$\begin{aligned}\big\|f_S^{(m+1)}(x)-f_0^{(m+1)}\big\|_2 &\le 2\big\|S^{(m)}f_S-S_0^{(m)}f_0\big\|_2\\ &\quad+\big\|P(S,\ldots,S^{(m-1)})f_S(x)-P(S_0,\ldots,S_0^{(m-1)})f_0\big\|_2\\ &\le 2\big\|(S^{(m)}-S_0^{(m)})f_S\big\|_2+2\big\|S_0^{(m)}(f_S-f_0)\big\|_2\\ &\quad+o_\delta(1)+\left[\int_{\mathbb R}(1+|x|^{\nu_1})^2\big(f_S(x)-f_0(x)\big)^2dx\right]^{1/2}. \qquad (5.7)\end{aligned}$$

To finish the proof of this lemma, it remains to verify that for any positive number n, we have

$$\lim_{\delta\to 0}\sup_{S\in\tilde\Sigma_\delta}\int_{\mathbb R}x^{2n}\big(f_S(x)-f_0(x)\big)^2dx=0. \qquad (5.8)$$

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

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

The first integral tends to zero due to the first assertion of this lemma and the fact that f_0 decreases exponentially. To estimate the second integral, we proceed like in (5.6). Thus, for any $A>B$, we have

$$e^{4\int_0^x S(v)dv}-e^{4\int_0^x S_0(v)dv}\le \mathbb{1}_{|x|<A}C(e^{2\delta A}-1)+\mathbb{1}_{|x|\ge A}Ce^{-2\gamma|x|}.$$

This inequality implies immediately (5.8). So the second and the forth terms in (5.7) are $o_\delta(1)$, and the first term is less than $\sqrt{R}$ since $S\in\widetilde\Sigma_\delta(R)$. Consequently

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

This completes the proof of Lemma 5.1. □

LEMMA 5.2. *If the function S_0 satisfies the conditions (4.1)and (5.3), then the following lower bound is true:*

$$\varliminf_{\delta\to 0}\varliminf_{T\to\infty} T^{\frac{2m}{2m+1}}\, R_T(\widetilde{\Sigma}_\delta(R)) \geq 4^{\frac{2m}{2m+1}}\; P(m,4R) = 4P(m,R),$$

where R_T is the minimax risk defined in Section 4.1.

PROOF. The proof of this lemma is the same as the proof of Theorem 4.1. It is enough to remark that $\mathcal{C}_T \subseteq \tilde{\Sigma}_\delta(m, R+\varepsilon, S_0)$ for any $\varepsilon > 0$. □

5.3. Invariant Density Estimation

Now we would like to state some results concerning the nonparametric estimation of invariant density f_S. This problem was widely studied by different authors (cf. [**1**], [**4**], [**36**], [**39**], [**48**],[**49**]). It is well known that the kernel estimators are asymptotically efficient for a large class of kernels (see, for example, [**36**]). But for our purposes it is more convenient to use the local time estimator $f_T^\circ(\cdot)$ introduced in Section 1.4. Remind that it is defined by formula

$$f_T^\circ(x) = \frac{L_T^x}{T}\ ,$$

where L_T^x is the local time of X at instant x. As we have said, in the problem of invariant density estimation by continuous path observations the rate of convergence is $1/\sqrt{T}$, which is not usual for the problems of nonparametric curve estimation. To formulate the exact result, let us denote

$$\mathcal{R}_f(S) = 4\int_{\mathbb{R}} f_S^2(x)\mathbf{E}_S\left(\frac{\mathbb{1}_{\{\xi>x\}} - F_S(\xi)}{f_S(\xi)}\right)^2 dx.$$

As it will be clear from the next proposition, $\mathcal{R}_f(S)$ plays the role of the inverse of Fisher information.

PROPOSITION 2 (cf. [**36**], Prop. 1 and Prop. 4). *If the center of localization S_0 satisfies the condition (4.1), then*

$$\lim_{\delta\to 0}\lim_{T\to\infty}\inf_{\bar{f}_T}\sup_{S\in V_\delta} T\,\mathbf{E}_S\big\|\bar{f}_T - f_S\big\|_2^2 = \mathcal{R}_f(S_0),$$

and this bound is attained by the local time estimator.

This proposition says that the local time estimator is asymptotically efficient over the class of trend coefficients belonging to Σ_0 and satisfying the condition (4.1).

To use this result in trend coefficient estimation we need to control higher order moments of the difference $(f_T^\circ - f_S)$. It can be done using the martingale representation of the local time (see (1.11)) and the estimates of Lemma 4.1. Combining these results we obtain

LEMMA 5.3. *Let S_0 satisfy the conditions (4.1) and (4.2). Then, for any $p \geq 1$ there exist two constants C and γ such that*

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

for any $x \in \mathbb{R}$ and $S \in V_\delta(S_0)$.

5.4. Lower Bound

Actually we know that the density $f_S(x)$ can be estimated with the rate $1/\sqrt{T}$, and its derivative $f_S'(x)$ can be estimated with the rate $T^{-\frac{m}{2m+1}}$, if the trend coefficient S is smooth of order m. Therefore the main contribution to the limit variance of the estimator of $S(x)$ is due to the derivative's estimator. Using these heuristics we establish now a lower bound for the minimax risk in the problem of the trend coefficient estimation using the risk $\mathcal{R}_T$ introduced in the first section of this chapter. In order to do it, we use the lower bound established in Lemma 5.2 and some results concerning the behavior of the local time estimator.

THEOREM 5.1. *Let the function $S_0(\cdot)$ satisfy the conditions (4.1) and (5.3), then*

$$\varliminf_{\delta \to 0} \varliminf_{T \to \infty} \inf_{\bar{S}_T} \sup_{S \in \tilde{\Sigma}_\delta} T^{\frac{2m}{2m+1}} \mathbf{E}_S \int_{\mathbb{R}} \left(\bar{S}_T(x) - S(x) \right)^2 f_S^2(x)\, dx \geq P(m, R),$$

where $P(m, R)$ is Pinsker's constant defined by (2.7).

PROOF. Remark that we can consider only the estimators $\bar{S}_T(\cdot)$ satisfying the condition

$$\left| \bar{S}_T(x) \right| \leq b_T\, e^{x/\ln b_T}. \tag{5.9}$$

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

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

Consequently, the risk of the truncated estimator

$$\tilde{S}_T(x) = \begin{cases} b_T\, e^{x/\ln b_T} &, \text{ if } \quad \bar{S}_T(x) > b_T\, e^{x/\ln b_T}, \\ \bar{S}_T(x) &, \text{ if } \quad \left| \bar{S}_T(x) \right| \leq b_T\, e^{x/\ln b_T}, \\ -b_T\, e^{x/\ln b_T} &, \text{ if } \quad \bar{S}_T(x) < -b_T\, e^{x/\ln b_T}. \end{cases}$$

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 (5.9).

If we denote $\bar{\vartheta}_T(\cdot) = \bar{S}_T(\cdot)\, f_T^\circ(\cdot)$, then using the triangular inequality we get

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

where the following notations are used:

$$A_1 = \mathbf{E}_S \int_{\mathbb{R}} \bar{S}_T^2(x)\big(f_S(x) - f_T^\circ(x)\big)^2 dx,$$
$$A_2 = \mathbf{E}_S \int_{\mathbb{R}} \big(\bar{\vartheta}_T(x) - S(x) f_S(x)\big)^2 dx.$$

The behavior of the term A_2 can be derived from the results of Section 5.2 and 5.3. Indeed, since $\bar{\vartheta}_T$ depends only on observations, it can be considered as a potential estimator of $\frac{1}{2} f_S'$. Consequently, using Lemma 5.2, we get

$$\begin{aligned}\inf_{\bar{\vartheta}_T} \sup_{S \in \widetilde{\Sigma}_\delta(R)} A_2 &= \frac{1}{4}\, R_T(\widetilde{\Sigma}_\delta(R)) \\ &\geq P(m,R) T^{-\frac{2m}{2m+1}} \big(1 + o_T(1)\big)\big(1 + o_\delta(1)\big)\end{aligned}$$

where the term $o_T(1)$ depends on δ. While for A_1, the inequality (5.9) associated with Lemma 5.3 gives the estimate

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

for any $S \in \widetilde{\Sigma}_\delta$ and T large enough. Let us denote

$$\mathcal{R}_T(\widetilde{\Sigma}_\delta) = \inf_{\bar{S}_T} \sup_{S \in \widetilde{\Sigma}_\delta} \mathbf{E}_S \int_{\mathbb{R}} \big(\bar{S}_T(x) - S(x)\big)^2 f_S^2(x)\, dx.$$

Then we have the obvious inequalities:

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

This completes the proof of the theorem. □

5.5. Asymptotically Efficient Estimator

Now we pass to the construction of an estimator of the trend coefficient which is asymptotically efficient in the sense of the lower bound of Theorem 5.1.

DEFINITION 7. *An estimator $\tilde{S}_T(\cdot)$ is said to be asymptotically efficient in the problem of trend coefficient estimation, if*

$$\lim_{\delta\to 0}\lim_{T\to\infty}\sup_{S\in\tilde{\Sigma}_\delta} T^{\frac{2m}{2m+1}}\,\mathbf{E}_S\left[\int_{\mathbb{R}}\big(\tilde{S}_T(x)-S(x)\big)^2 f_S^2(x)\,dx\right] = P(m,R). \tag{5.10}$$

Remind that

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

We have already found an asymptotically efficient estimator $\bar{\vartheta}_{K^*,T}(\cdot)$ of $f_S'(\cdot)$. The invariant density $f_S(\cdot)$ is well estimated by the local time estimator $f_T^\circ(\cdot)$. Thus, it would be very natural to estimate $S(x)$ by

$$\bar{S}_T(x)=\frac{\bar{\vartheta}_{K^*,T}(x)}{2f_T^\circ(x)}\ .$$

The great disadvantage of this estimator is that the denominator part can be very close and even equal to 0. That is why it is practically impossible to prove any property concerning the risk of this estimator. Fortunately, we can get round this difficulty using the following idea. The main feature of the estimator f_T° that we need is the convergence to f_S with a rate faster than $T^{-\frac{m}{(2m+1)}}$. That is why we can replace f_T° by another estimator which converges slower than $1/\sqrt{T}$ but faster than $T^{-m/(2m+1)}$, and which is bounded away from 0.

These heuristics lead us to investigate the behavior of the estimator

$$\hat{S}_T(x)=\frac{\bar{\vartheta}_{K^*,T}(x)}{2f_T^\circ(x)+\varepsilon_T e^{-l_T|x|}} \tag{5.11}$$

where $\bar{\vartheta}_{K^*,T}(\cdot)$ is the asymptotically efficient estimator of the derivative of the invariant density studied in the previous chapter, $f_T^\circ(\cdot)$ is the local time estimator, $\varepsilon_T=T^{-(1-\kappa)/2}$ and $l_T=\big[\ln(T+1)\big]^{-1}$. The positive constant κ is chosen to be strictly smaller than $\beta=1/(2m+1)$. The role of the multiplier $e^{-l_T|x|}$ is of course to ensure the exponential smallness of the tails.

To prove the asymptotic efficiency of this estimator we need an auxiliary result.

LEMMA 5.4. *Let conditions (4.1) and (4.2) be satisfied. Then, there exists a constant C such that*

$$\mathbf{E}_S\big[\bar{\vartheta}^4_{K^*,T}(x)\big] \le CT^{4\beta},$$

for any $x \in \mathbb{R}$ and $S \in \widetilde{\Sigma}_\delta(S_0)$.

PROOF. Remark firstly that the elementary inequality $(a+b)^4 \le 8a^4+8b^4$ implies that

$$\mathbf{E}_S\big[\bar{\vartheta}^4_{K^*,T}(x)\big] \le \frac{C}{T^4}\,\mathbf{E}_S\bigg[\int_0^T K^*_T(x-X_t)S(X_t)dt\bigg]^4 + \frac{C}{T^4}\,\mathbf{E}_S\bigg[\int_0^T K^*_T(x-X_t)\,dW_t\bigg]^4.$$

Now, using BDG (Theorem 1.2) and Cauchy-Schwarz inequalities, one can easily show that

$$\mathbf{E}_S\bigg[\int_0^T K^*_T(x-X_t)\,dW_t\bigg]^4 \le \mathbf{E}_S\bigg[\int_0^T K^*_T(x-X_t)^2\,dt\bigg]^2 \le T^2\mathbf{E}_S\big[K^*_T(x-\xi)^4\big].$$

Similarly, Hölder's inequality leads to

$$\frac{C}{T^4}\,\mathbf{E}_S\bigg[\int_0^T K^*_T(x-X_t)S(X_t)dt\bigg]^4 \le C\mathbf{E}_S\big[K^*_T(x-\xi)^4 S(\xi)^4\big].$$

It follows from (4.7) 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 $\widetilde{\Sigma}_\delta$. These estimates imply the inequality of Lemma 5.4. □

Remark that proceeding like in the proof of Lemma 4.2, one can show that

$$\mathbf{E}_S\big|\bar{\vartheta}_{K^*,T}(x) - f'_S(x)\big|^4 \le CT^{2\beta-2}.$$

This estimate is much sharper than the result of Lemma 5.4, but for the proof of the efficiency of $\hat{S}_T$, the inequality of Lemma 5.4 is amply sufficient.

THEOREM 5.2. *If conditions (4.1), (4.3) and (5.3) 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 introduce the event

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

According to Lemma 5.3 and Chebyshev's inequality, 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 $\gamma_* < \gamma$. Using Cauchy-Schwarz inequality and Lemma 5.4, together with estimate (4.11), one can show that

$$\begin{aligned}\mathbf{E}_S\Big[\big(\hat{S}_T(x)-S(x)\big)^2 f_S^2(x)\mathbb{1}_{\mathbb{B}_T^c(x)}\Big] &\le 2\,\mathbf{E}_S\Big[\hat{S}_T^2(x)f_S^2(x)\mathbb{1}_{\mathbb{B}_T^c(x)}\Big]\\ &\quad +2S^2(x)\,f_S^2(x)\,\mathbf{P}_S\big[\mathbb{B}_T^c(x)\big]\\ &\le \frac{C\varepsilon_T^{-2}\,e^{-2\gamma|x|}}{T^{\kappa p-2\beta}} + CS^2(x)f_S^2(x)T^{-\kappa p}\\ &\le CT^{2-\kappa p}e^{-2\gamma|x|}.\end{aligned}$$

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

$$\int_{\mathbb{R}} \mathbf{E}_S\Big[\big(\hat{S}_T(x)-S(x)\big)^2 f_S^2(x)\mathbb{1}_{\mathbb{B}_T^c(x)}\Big]dx \le CT^{-1}.$$

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

$$\int_{\mathbb{R}} \mathbf{E}_S\Big[\big(\hat{S}_T(x)-S(x)\big)^2 f_S^2(x)\mathbb{1}_{\mathbb{B}_T(x)}\Big]dx \le (\sqrt{A}_1+\sqrt{A}_2)^2,$$

where

$$\begin{aligned}A_1 &= \int_{\mathbb{R}} \mathbf{E}_S\Big[\big(\hat{S}_T(x)-S(x)f_S(x)/2\bar{f}_T(x)\big)^2 f_S^2(x)\mathbb{1}_{\mathbb{B}_T(x)}\Big]dx,\\ A_2 &= \int_{\mathbb{R}} \mathbf{E}_S\Big[\big(S(x)f_S(x)/2\bar{f}_T(x)-S(x)\big)^2 f_S^2(x)\mathbb{1}_{\mathbb{B}_T(x)}\Big]dx.\end{aligned}$$

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:

$$\begin{aligned}4A_1 &\le \int_{\mathbb{R}} \mathbf{E}_S\big(\bar{\vartheta}_{K^*,T}(x)-f_S'(x)\big)^2 dx,\\ 4A_2 &\le \int_{\mathbb{R}} S^2(x)\,\mathbf{E}_S\big(\bar{f}_T(x)-f_S(x)\big)^2 dx.\end{aligned}$$

The condition (5.3) and Lemma 5.3 imply that, uniformly on $S \in \widetilde{\Sigma}_\delta$, the term A_2 is of order $\varepsilon_T^2 l_T^{-2n}$, which is smaller than $T^{-2m/(2m+1)}$ since $\kappa < 1/(2m+1)$. Finally, combining the results of Lemma 5.1 and Theorem 4.2, we obtain

$$\begin{aligned}\sup_{S\in\widetilde{\Sigma}_\delta(R)} T^{\frac{2m}{2m+1}}\,R_T(\hat{S}_T,S) &\le o_T(1) + \sup_{S\in\widetilde{\Sigma}_\delta(R)} \big(T^{\frac{m}{2m+1}}\sqrt{A_1}+o_T(1)\big)^2\\ &\le o_T(1) + \sup_{S\in\Sigma_\delta(4R+r_\delta)} \big(T^{\frac{m}{2m+1}}\sqrt{R_T(\bar{\vartheta}_{K^*,T},f_S')}+o_T(1)\big)^2\\ &\le P(m,R+r_\delta)\,\big(1+o_T(1)\big)\big(1+o_\delta(1)\big),\end{aligned}$$

where the terms $o_T(1)$ may depend on δ and $r_\delta \to 0$ as $\delta \to 0$. This last estimate completes the proof of Theorem 5.2. $\square$

Remark. Throughout this work we supposed that the initial value X_0 of the observed diffusion process X was random with density f_S. We would like to note that the goal of this assumption was to simplify a little bit the proofs of theorems. As a matter of fact, using a uniform version of Theorem 1.6, one can show that all the results of this work are true for ergodic diffusions issuing from any point (possibly random).

Bibliography

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

[2] Belitser, E.N. and Levit, B.Ya. (1995). On minimax filtering over ellipsoids. *Math. Methods Statist.*, Vol. 4, No. 3, 259–273.

[3] Bosq, D. and Davydov, Y. (1999). Local time and density estimation in continuous time. *Math. Methods of Statistics*, Vol. 8, No. 1, 22–45.

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

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

[6] Dalalyan, A.S. and Kutoyants, Yu.A. (2000). Asymptotically efficient trend coefficient estimation for ergodic diffusion. Prepublication 01–2, Université du Maine, Le Mans, submitted.

[7] Delattre, S. and Hoffmann, M. (1999). Mixed Gaussian white noise. Prepublication n° 504, Universités Paris 6 et 7.

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

[9] Donoho, D.L., Liu, R.C. and MacGibbon, K.B. (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.*, Vol. 18, 1416–1437.

[10] Donoho, D.L. and Johnstone, I.M. (1994). Minimax risk over l_p-balls for l_q-error. *Prob. Theor. Rel. Fields*, 99, 277–303.

[11] Donoho, D.L. and Johnstone, I.M. (1996). Neo-classical minimax problems thresholding and adaptive function estimation. *Bernoulli*, 2 (1), 39–62.

[12] Donoho, D.L. and Johnstone, I.M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.*, 26, 879–921.

[13] Donoho, D.L. and Johnstone, I.M. (1999). Asymptotic minimaxity of wavelet estimators with sampled data. *Statistica Sinica*, 9, 1–32.

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

[15] Dynkin, E.B. (1965). *Markov Processe, I, II.* Springer-Verlag, Berlin, Heidelberg and New York.

[16] Efromovich, S.Yu. and Pinsker, M.S. (1981). Estimation of square integrable spectral density based on a sequence of observations. *Problems Inform. Transmission*, 17, 182–196.

[17] Efromovich, S.Yu. and Pinsker, M.S. (1982). Estimation of square integrable probability density of a random variable. *Problems Inform. Transmission*, 18, No. 3, 19–38.

[18] Galtchouk, L. and Pergamenshchikov, S. (2000). Nonparametric sequential minimax estimation of the drift coefficient in diffusion processes. Submitted.

[19] Galtchouk, L. and Pergamenshchikov, S. (2000). Sequential nonparametric adaptive estimation of the drift coefficient in diffusion processes. Accepted in *Proceedings of Lumini Conference*, Ed. Lepski.

[20] Genon-Catalot, V., Laredo, C. and Nussbaum, M. (2001). Asymptotic equivalence of estimating a Poisson intensity and a positive diffusion drift. *To appear in Ann. Statist.*

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

[22] Gill, R.D. and Levit, B. Ya. (1995). Application of the van Trees inequality: a Bayesian Cramer-Rao bound. *Bernoulli*, No. 1, 59–79.

[23] Golubev, G.K. (1991). LAN in problems of non-parametric estimation of functions and lower bounds for quadratic risks. *Theory Probab. Appl.*, Vol. 36, No. 1, 152–157.

[24] Golubev, G.K. (1992). Nonparametric estimation of smooth densities in L^2 (in Russian). *Problems Inform. Transmission*, Vol. 28, No. 1, 52–62.

[25] Golubev, G.K. (1992). Asymptotic minimax estimation of regression in the additive model. *Problems Inform. Transmission*, Vol. 28, No. 2, 101–112.

[26] Golubev, G.K. and Härdle, W. (2000). Second order minimax estimation in partial linear models. *Math. Methods Statist.*, Vol. 9, No. 2, 160–175.

[27] Golubev, G.K. and Khas'minskii, R.Z.(1999). A statistical approach to some inverse problems for partial differential equations. *Problems Inform. Transmission*, Vol. 35, No. 2, 136–149.

[28] Golubev, G.K. and Levit, B. Ya. (1996). On the second order minimax estimation of distribution functions. *Math. Methods Statist.*, Vol. 5, No. 1, 1–31.

[29] Golubev, G.K. and Levit, B. Ya. (1996). Asymptotically efficient estimation for analytic distributions. *Math. Methods Statist.*, Vol. 5, No. 3, 357–368.

[30] Golubev, G.K. and Nussbaum, M. (1990). A risk bound in Sobolev class regression. *Annals of Statistics*, Vol. 18, No. 2, 758–778.

[31] Hasminskii, R.Z. (1980). *Stochastic Stability of Differential Equations.* Sijthoff and Noordhoff, Alpen.

[32] Ibragimov, I.A. and Khas'minskii, R.Z. (1976). Some problems of statistical data processing in communications, Proc. 1975 IEEE-USSR Joint Workshop on Information Theory, N. Y., 86–92.

[33] Ibragimov, I.A. and Khas'minskii, R.Z. (1981). *Statistical Estimation: asymptotic theory.* Springer, New York.

[34] Johnstone, I. *Function Estimation in Gaussian Noise: Sequence Models.* http://www-stat.stanford.edu/~imj.

[35] Kutoyants, Yu.A. (1997). Efficiency of the empirical distribution for ergodic diffusion. *Bernoulli.* Vol. 3, No. 4, 445–456.

[36] Kutoyants, Yu.A. (1998). Efficient density estimation for ergodic diffusion. *Statistical Inference for Stochastic Processes.* Vol. 1, No. 2, 131–155.

[37] Kutoyants, Yu.A. (1998). *Statistical Inference for Spatial Poisson Processes.* Lect. Notes Statist. 134, Springer-Verlag, New York.

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

[39] Leblanc, F. (1997). Density estimation for a class of continuous time process. *Math. Methods of Statistics*, Vol. 6, No. 2, 171–199.

[40] Lucas, A. (1998). Can we estimate the density's derivative with suroptimal rate? *Statist. Inference for Stochastic Processes*, Vol. 1, No. 1, 29–41.

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

[42] Negri, I. (1998). Stationary distribution function estimation for ergodic diffusion processes. *Statistical Inference for Stochastic Processes.* Vol. 1, No. 1, 61–84.

[43] Nemirovski, A. (2000). *Topics in Non-Parametric Statistics.* In: Lectures on Probability Theory and Statistics. École d'Été de Probabilités de St Flour XXVIII - 1998. Lecture Notes in Mathematics, v. 1738., Springer, N.Y.

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

[45] Nguyen, H.T. and Pham, T.D. (1982). Identification of nonstationary diffusion model by the method of sieves. *SIAM J. of control and optimization*, Vol. 20, No. 5, 603–611.

[46] Nussbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L^2. *Annals of Statistics*, Vol. 13, No. 3, 984–997.

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

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

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

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

[51] Revuz, D. and Yor, M. (1991). *Continuous Martingales and Brownian Motion.* Springer-Verlag, Berlin, Heidelberg.

[52] Schipper, M. (1996). Optimal rates and constants in L^2-minimax estimation of probability density functions. *Math. Methods Statist.*, Vol. 5, No. 3, 253–274.

[53] Speckman, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Annals of Statistics*, Vol. 13, No. 3, 970–983.

[54] Spokoiny, V.G. (2000). Adaptive drift estimation for nonparametric diffusion model. *Annals of Statistics*, 28, 815–836.

[55] van Trees, H.L. (1968). *Detection, Estimation and Modulation Theory*, Part 1. New York: Wiley.

[56] Tsybakov, A.B. (1997). Asymptotically Efficient Signal Estimation in l^2 under general loss functions. *Problems Inform. Transmission*, 33, 78–88.